THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Serum diagnostic glycoprotein biomarkers for Barrett's esophagus and esophageal adenocarcinoma

Alok Kishorkumar Shah

Bachelor of Pharmacy, Master of Science (Pharmacy) in Biotechnology

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2015

The University of Queensland Diamantina Institute

Abstract

Esophageal cancer is the eighth most commonly diagnosed cancer and the sixth most common cause of cancer related mortality globally. There are two different types of esophageal cancer, esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC), each accounting for half of the cases. EAC develops in the lower one third of the esophagus as a consequence of chronic gastroesophageal reflux disease (GERD) from precancerous metaplastic condition Barrett's esophagus (BE). Apart from GERD and BE which are major risk factors for EAC, other known risk factors include age, male gender, obesity, Caucasian race, low intake of fruits and vegetables in diet, and *Helicobacter pylori* negative status. Due to changing life style and prevalence of risk factors, the incidences of EAC have been rising for past few decades and now it has become one of the fastest growing malignancies. The survival rate is very poor with only 1 in 5 patients survive more than 5 years after EAC diagnosis, likely due to diagnosis at late stages. To diagnose early treatable dysplastic changes in progression from BE to EAC, BE patients undergo routine endoscopy-biopsies, with the biopsy evaluated by a histopathologist to confirm the dysplastic changes. This current method is invasive and prone to sampling error as well as interobserver variability. Endoscopy requires patient hospitalization and specialist appointment, leading to high expense. Moreover, BE is an asymptomatic condition which means a pool of BE patients are undiagnosed hence not enrolled into the surveillance program. Collectively, it has been shown that current endoscopy-biopsies based diagnostic is impractical and expensive for population wide BE screening or surveillance programs.

In contrast to endoscopy-biopsy, biomarkers from the blood are amenable to population-screening strategies, due to the ease of access and low cost of testing. Moreover, EAC pathogenesis has been associated with changes in the serum glycan profile. However, specific glycoproteins that undergo differential glycosylation are unknown. Therefore, the aims of this thesis were to (i) identify serum diagnostic glycoprotein biomarker candidates for BE and EAC using biomarker discovery pipeline, (ii) develop a targeted proteomics approach to measure biomarker candidates for timely verification, (iii) verify serum glycoprotein candidates in an independent patient cohort, and (iv) test feasibility of using electrochemical detection methodology for the glycoprotein detection.

This translational research project utilizes lectins, naturally occurring proteins with specificity to bind with glycan structures, as affinity agents to isolate glycoproteins with different glycan structures. Our laboratory has previously established lectin magnetic bead array (LeMBA) methodology to identify serum glycoprotein biomarker candidates showing differential lectin

binding (Loo *et al*., *J Proteome Res* 2010 and Choi *et al*., *Electrophoresis* 2011). With the help of a bioinformatician and biostatisticians, GlycoSelector database incorporating statistical analysis pipeline was developed for biomarker discovery using LeMBA platform (http://glycoselector.di.uq.edu.au). Serum samples from 29 patients (healthy - 9, BE - 10 and EAC - 10) were screened using LeMBA-GlycoSelector pipeline. A ranked list of candidate glycoprotein biomarkers that distinguish (i) EAC from BE (ii) BE from healthy and (iii) EAC from healthy group was identified. GlycoSelector analysis resulted in identification of total 183 unique lectin-protein biomarker candidates for targeted verification.

Out of the 20 lectins employed for the biomarker discovery, 6 lectins showing differential binding with glycoprotein candidates were selected for verification. Multiple reaction monitoring-mass spectrometry (MRM-MS) assay was set up for 41 promising glycoprotein candidates. After testing linearity and reproducibility of MRM-MS assay, serum samples from an independent patient cohort were screened using customized LeMBA coupled with MRM-MS. Online web-portal Shiny mixOmics (http://mixomics-projects.di.uq.edu.au/Shiny) was used for statistical analysis. Of the 246 glycoforms measured in the verification stage, 40 glycoforms (as measured by lectin affinity) verified as candidate serum markers. The top candidate for distinguishing healthy from BE was *Narcissus pseudonarcissus* lectin (NPL)-reactive Apolipoprotein B-100; BE vs EAC, *Aleuria aurantia* lectin (AAL)-reactive complement component C9; healthy vs EAC, Erythroagglutinin *Phaseolus vulgaris* (EPHA)-reactive gelsolin. A panel of 8 glycoforms showed an area under receiver operating characteristic curve (AUROC) of 0.94 to discriminate EAC from BE. Two biomarker candidates were independently verified by LeMBA-immunoblotting, confirming the validity of the relative quantitation approach employed.

Mass spectrometry methods employed for biomarker discovery and verification are best suited for research laboratories but not for routine clinical practice whereas electrochemical detection methods have been successfully applied for development of point-of-care diagnostics e.g. glucose biosensor. In this thesis, the feasibility of using electrochemical method for glycoprotein detection has been tested with success using a model glycoprotein ovalbumin with *Sambucus nigra* agglutinin (SNA lectin). A detection limit of 10 pg/mL was demonstrated, in the background of diluted human serum.

Taken together, this study firstly identified and then verified serum diagnostic glycoprotein biomarker candidates using two independent patient cohorts for BE/EAC. The biomarker candidates described here require further clinical evaluation in a large patient cohort including early dysplastic

patient samples. Electrochemical detection method described in the last part of this thesis can be developed further into *in vitro* diagnostic for clinical use employing glycoprotein biomarker candidates.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Alok K. Shah

Publications during candidature

Peer-reviewed paper

- **Shah AK**, Lê Cao KA, Choi E, Chen D, Gautier B, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Joshi V, and Hill MM. Serum glycoprotein biomarker discovery and qualification pipeline reveals diagnostic biomarkers for esophageal adenocarcinoma. *Mol Cell Proteomics*. DOI:10.1074/mcp.M115.050922
- **Shah AK,** Hill MM, Shiddiky MJ and Trau M. Electrochemical detection of glycan and protein epitopes of glycoproteins in serum. *Analyst* 139, 5970-5976 (2014).
- **Shah AK**, Saunders NA, Barbour AP and Hill MM. Early diagnostic biomarkers for esophageal adenocarcinoma − The current state of play. *Cancer Epidemiol Biomarkers Prev* 22, 1185-1209 (2013).

Manuscript under review

- Caragata M, **Shah AK,** Schulz BL, Hill MM, and Punyadeera C. Enrichment and identification of glycoproteins in human saliva using lectin magnetic bead arrays. *Anal Biochem*.

Provisional Patent

- Australian provisional patent application No. 2014904616, Biomarkers and uses thereof.

Selected conference abstracts

- **Shah AK**, Lê Cao KA, Choi E, Chen D, Gautier B, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Hill MM. Towards a blood test for esophageal adenocarcinoma: Discovery of serum glycoprotein diagnostic biomarkers. **The Australian Society for Medical Research (ASMR) Postgraduate Student Conference [Oral Presentation]**, Brisbane, Queensland, Australia. (May 2015)
- **Shah AK**, Lê Cao KA, Choi E, Chen D, Gautier B, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Hill MM. Towards a blood test for esophageal adenocarcinoma. **EACR-OECI Precision Medicine for Cancer Conference,** Neumunster Abbey, Luxembourg. (March 2015)
- **Shah AK**, Lê Cao KA, Choi E, Chen D, Gautier B, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Hill MM. Serum glycoprotein biomarkers for esophageal adenocarcinoma. **20th Lorne Proteomics Symposium [Runner-up student of Australasian Proteomics Society (SOAPS) speaker award]**, Lorne, Victoria, Australia. (Feb 2015)
- **Shah AK**, Chen D, Lê Cao KA, Choi E, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Hill MM. Serum glycoprotein biomarkers for diagnosis of Barrett's esophagus and esophageal

adenocarcinoma. **Princess Alexandra Hospital Health Symposium**, Brisbane, Queensland, Australia. (Aug 2014)

- **Shah AK**, Chen D, Lê Cao KA, Choi E, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Hill MM. Discovery and validation of novel serum glycoprotein biomarkers for Barrett's esophagus and esophageal adenocarcinoma. **American Association for Cancer Research 105th Annual Meeting**, San Diego, California, USA. (April 2014)
- **Shah AK**, Chen D, Lê Cao KA, Choi E, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Hill MM. Novel serum glycoprotein biomarkers for Barrett's esophagus and esophageal adenocarcinoma − Discovery and validation. **26th Lorne Cancer Symposium**, Lorne, Victoria, Australia. (Feb 2014)
- **Shah AK**, Chen D, Lê Cao KA, Choi E, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Hill MM. LeMBA-GlycoSelector pipeline identifies novel serum glyco-biomarkers for Barrett's esophagus and esophageal adenocarcinoma. **International Postgraduate Symposium in Biomedical sciences**, Brisbane, Queensland, Australia. (Oct 2013)
- **Shah AK**, Chen D, Lê Cao KA, Choi E, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Hill MM. Differentially glycosylated circulating protein biomarker discovery for Barrett's esophagus and esophageal adenocarcinoma. **Human Proteome Organization (HUPO) 12th Annual World Congress (Young Investigator Oral Presentation)**, Yokohama, Japan. (Sep 2013)
- **Shah AK**, Chen D, Lê Cao KA, Choi E, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Hill MM. Differential glycosylation of circulating proteins as diagnostic biomarkers for Barrett's esophagus and esophageal adenocarcinoma. **Princess Alexandra Hospital Health Symposium**, Brisbane, Queensland, Australia. (Aug 2013)
- **Shah AK**, Chen D, Lê Cao KA, Choi E, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Hill MM. Circulatory glyco-biomarkers for Barrett's esophagus and esophageal adenocarcinoma. **The Australian Society for Medical Research (ASMR) Postgraduate Student Conference**, Brisbane, Queensland, Australia. (May 2013)
- **Shah AK**, Chen D, Lê Cao KA, Choi E, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Hill MM. Differential glycosylation of circulating proteins as diagnostic biomarkers for Barrett's esophagus and esophageal adenocarcinoma. **18th Lorne Proteomics Symposium**, Lorne, Victoria, Australia. (Feb 2013)

Publications included in this thesis

- **Shah AK**, Saunders NA, Barbour AP and Hill MM. Early diagnostic biomarkers for esophageal adenocarcinoma − The current state of play (Appendix I). *Cancer Epidemiol Biomarkers Prev* 22, 1185-1209 (2013) (1). Partially incorporated in Chapter 1 Introduction.

Contributor	**Statement of contribution**
Alok K. Shah	Design of the article (60%) Compilation of data in literature (100%) Writing, review, and/or revision of the manuscript (65%)
Nicholas A. Saunders	Design of the article (10%) Writing, review, and/or revision of the manuscript (5%)
Andrew P. Barbour	Design of the article (10%) Writing, review, and/or revision of the manuscript (5%)
Michelle M. Hill	Conception of the study (100%) Design of the article (20%) Writing, review, and/or revision of the manuscript (25%)

* All authors revised and approved the final manuscript

- **Shah AK,** Hill MM, Shiddiky MJ and Trau M. Electrochemical detection of glycan and protein epitopes of glycoproteins in serum. *Analyst* 139, 5970-5976 (2014) (2). Incorporated in Chapter 6.

Contributor	**Statement of contribution**
Alok K. Shah	Conception of the study (25%) Design experiment (85%) Acquisition of data (100%) Analysis and interpretation of data (90%) Writing, review, and/or revision of the manuscript (73%)
Michelle M. Hill	Conception of the study (25%) Design experiment (2.5%) Supervisory role (25%) Writing, review, and/or revision of the manuscript (8%)
Muhammad J.A. Shiddiky	Conception of the study (25%) Design experiment (10%)

	Analysis and interpretation of data (10%) Supervisory role (50%) Writing, review, and/or revision of the manuscript (16%)
Matt Trau	Conception of the study (25%) Design experiment (2.5%) Supervisory role (25%) Writing, review, and/or revision of the manuscript (3%)

* All authors revised and approved the final manuscript

- **Shah AK**, Lê Cao KA, Choi E, Chen D, Gautier B, Nancarrow D, Whiteman D, Saunders NA, Barbour AP, Joshi V, and Hill MM. Serum glycoprotein biomarker discovery and qualification pipeline reveals diagnostic biomarkers for esophageal adenocarcinoma. *Mol Cell Proteomics*. DOI:10.1074/mcp.M115.050922 (3). Supplied as appendix II and partially incorporated in Chapter 2, 3, 4, 5 and 7.

Statement of contribution	**Contributor**
Conception of the study	Alok K. Shah (15%), Eunju Choi (5%), Michelle M. Hill (80%)
Design experiment	Alok K. Shah (90%), Eunju Choi (10%)
Collection and selection of patient samples	Derek Nancarrow (50%), David C. Whiteman (50%)
Acquisition of data	Alok K. Shah (95%), Eunju Choi (5% that mainly includes initial optimization experiments)
Development and testing of GlycoSelector	Development: Kim-Anh Lê Cao (50%), David Chen (50%) Testing and user's perspective: Alok K. Shah (25%), Kim-Anh Lê Cao (20%), Eunju Choi (25%), David Chen (20%), Michelle M. Hill (10%)
Development and testing of Shiny mixOmics	Development: Kim-Anh Lê Cao (30%), Benoît Gautier (70%) Testing and user's perspective: Alok K. Shah (20%), Kim-Anh Lê Cao (15%), Benoît Gautier (60%), Michelle M. Hill (5%)
Statistical analyses	Alok K. Shah (70%), Benoît Gautier (20%), Derek Nancarrow (5%), Michelle M. Hill (5%)
Supervisory role	Kim-Anh Lê Cao (15%), Eunju Choi (5%), Nicholas A. Saunders (7.5%), Andrew P. Barbour (7.5%), Michelle M. Hill (65%)
Writing, review, and/or	Alok K. Shah (50%), Kim-Anh Lê Cao (8%), Eunju Choi (15%), Nicholas A. Saunders (3%), Virendra Joshi (1%), Michelle M. Hill

revision of the manuscript	(23%)
Managing and uploading data to public repository	Alok K. Shah (100%)

* All authors revised and approved the final manuscript

Contributions by others to the thesis

The conception, design, guidance and supervision of this project including the support of writing this thesis were provided primarily by Dr. Michelle M. Hill along with contributions by co-supervisors A/Prof Nicholas Saunders and A/Prof Andrew Barbour. Prof Matt Trau and Dr Muhammad Shiddiky were involved in conception, design, guidance and supervision for work presented in Chapter 6.

High-throughput glycoprotein enrichment platform LeMBA (Lectin magnetic bead array) for biomarker discovery and verification described in this thesis had previously been optimized by Dr. Eunju (April) Choi. Dr. Choi optimized buffer compositions and incubation conditions including use of chicken ovalbumin as an internal standard for LeMBA. Her work also included optimization of LeMBA pull-down on Bravo liquid handler (Agilent Technologies) and development of mass spectrometry (QTOF) methods used for biomarker discovery. Her protocols were followed with minor modifications, where applicable. Dr. Choi taught me to perform LeMBA pull-downs including hands-on training to operate Bravo liquid handler and QTOF mass spectrometer.

Statistical analysis scripts for data analysis tools GlycoSelector (http://glycoselector.di.uq.edu.au) and Shiny mixOmics (http://mixomics-projects.di.uq.edu.au/Shiny) were designed and developed by Dr. Kim-Anh Lê Cao with the help of Mr. Benoit Gautier. Dr. David Chen has designed and written GlycoSelector database incorporating statistical analysis pipeline and currently maintains the database. Dr. Choi helped Dr. Chen and Dr. Lê Cao from a user perspective during the initial phase of GlycoSelector development which included features like outlier detection, group binding difference analysis and sparse partial least squares-discriminant analysis (sPLS-DA). Later on I conceptualized normalization method based on individual peptide intensity. Latest GlycoSelector version additionally contains features like stability analysis and group binding differences between three groups. I conceptualized the basis for MRM data analysis including two step data normalization process, correlation analysis to remove outlier peptides and calculation of protein intensity from raw peptide intensity values. Mr. Gautier and Dr. Lê Cao wrote R scripts to automate these data pre-processing steps and to format the data in

compliance with Shiny mixOmics. Shiny mixOmics incorporates tools to perform univariate analysis (Kruskal-Wallis test, ANOVA, ROC analysis) and multivariate analysis (sPLS-DA and stability analysis) to enable selection of complimentary biomarkers. Time and again, my role during the development of GlycoSelector and Shiny mixOmics was to help bioinformatician and statisticians from user perspective, test GlycoSelector and Shiny mixOmics against the data generated by myself in Microsoft Excel and GraphPad Prism®, and suggest improvements for better performance.

Training and technical assistance for mass spectrometric analysis (Chapters 3, 4 and 5) was provided by Mr. Elliot McElroy, Dr. Thomas Hennessy, Mr. Christopher Bowen and Ms. Christine Miller of Agilent Technologies and Ms. Dorothy Loo of The University of Queensland Diamantina Institute.

Work described in Chapter 6 was undertaken in collaboration with Professor Matt Trau and Dr. Muhammad Shiddiky of Australian Institute of Bioengineering and Nanotechnology (AIBN), The University of Queensland, St. Lucia, Australia. Specifically all the experiments were performed by me at AIBN under the direct supervision of Dr. Shiddiky. Dr. Shiddiky helped for data analysis and also helped during manuscript preparation and writing. Discussions held with Dr. Hill and Professor Trau were always fruitful and they also helped during manuscript preparation and writing.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

I have been privileged to carry out journey of my PhD at one of the top institution in the world for biomedical research [The University of Queensland Diamantina Institute (UQDI) located at The Translational Research Institute] and more importantly with such a wonderful people. First and foremost, I would like to thank my supervisor Dr. Michelle Hill. Starting from scholarship application to writing up this thesis, Michelle has always steered me in the right direction and without her continuous support, cooperation, and encouragement this work would have never become possible. Thank you Michelle for being very friendly and patient with me and setting the balance between freedom and supervision from Day 1 to up until now. I would like to extend my gratitude to co-supervisors A/Prof Andrew Barbour and A/Prof Nicholas Saunders who time and again provided feedback on the project and my candidature. I would also like to thank thesis committee members Dr. Gethin Thomas and Dr. Benjamin Schulz for keeping track of my project.

I have been very blessed to work on a collaborative project which gave me an opportunity to work with people from diverse disciplines that extended my horizon about science. Dr. Kim-Anh Lê Cao has always been very helpful with statistical analysis. She has patiently answered all my stupid statistical questions. I am also grateful to Dr. David Chen and Mr. Benoit Gautier for all their help related to statistical analysis and bioinformatics. Professor Matt Trau and Dr. Muhammad Shiddiky willingly accepted me to become a part of their team that allowed me to get insights into whole new world of electrochemistry and point-of-care diagnostics. Although I couldn't continue work with them for a longer period of time, the fundamental concepts I learned from them will last forever with me. I would like to thank Professor David Whiteman and Dr. Derek Nancarrow for providing the necessary patient samples for the project. I would also like to acknowledge all the patients for willingly donating the samples for the research work. Team Agilent (Mr. Elliot McElroy, Dr. Thomas Hennessy, Mr. Christopher Bowen, and Ms. Christine Miller) has provided the technical support for mass spectrometry when required.

Past and present Hill group members have become very good friends and I will definitely miss them once I leave this place. In fact some of them I am already missing. First and foremost, my friend Dr. Eunju (April) Choi has been very instrumental from Day 1 of PhD. She taught me everything starting from OH&S rules to running million dollar babies (mass specs). In fact, her PhD project laid the technical foundation for the work presented in this thesis and her teaching has made start of my PhD very smooth. Outside the lab, April and my buddy Dr. Hyeongsun (Sunny) Moon became the person whom I approached for any kind of advice. Sunny, I will always remember all our scientific discussions and moments that we shared together during our PhD journey. I wish all

the luck to Moon family Sunny, April and little champ Davin. Dr. Kerry Inder, convener of "Fabulous five meeting", is one of the most positive people I have ever met. I bet she will be running one of the best ever school one day, hopefully very soon. Fifth and final member of fabulous five (Kerry, Sunny, April and myself are other four) and one of my best friend Anup has always made his presence felt in past six years of my life. Ms. Jayde Ruelcke and I started our journey in the field of proteomics at the same time and I would like to thank Ms. Dorothy Loo for teaching us first lessons in mass spec/proteomics and for all tricky questions. Jayde has been always willing to extend her help whenever I asked for and I acknowledge her for managing the lab and baking lots of goodies. I appreciate efforts of past and present Hill group members Akash, Amanda, Bastien, Cass, Claire, Deb, Golnoosh, Harriet, Jeff, Kristine, Sharda, Shashank, Tam, and Thomas for making intellectually stimulating lab environment and relaxing Friday morning coffee breaks.

I must acknowledge Michelle once again for not only providing me teaching opportunity for the elective course in proteomics but also for showing faith in me and giving me the responsibility to teach/supervise PhD, summer and visiting students in the lab. I was fortunate to be involved with two SPARQ-ed (Students performing advanced research Queensland) research immersion programs, and for that I thank inspiring teacher/scientist Dr. Peter Darben. I would also like to highlight efforts of UQDI support staff mainly UQDI IT (Marcus, Darren, Peter and Scott) and Ms. Nicole Chandler for helping me with all the travel requests. I acknowledge financial support from Australian Government and The University of Queensland in the form of two scholarships International Postgraduate Research Scholarship (IPRS) and UQ Centennial (UQ Cent). I am extremely delighted to receive travel awards from Cancer Council Queensland and The Jian Zhou Pre-Doctoral Award for attending conferences and visiting labs and companies.

Life here in beautiful Brisbane has become memorable and joyful by having friends Ganesh, Prarthana, Vikas, Anup, Dhirendra, and Naisarg around. I have many friends to thank for always being present in my life; specifically I like to mention Tejas, late Vishank, Jignesh, Bhargav, Dr Anant, and Heli.

Last but not the least; I would like to thank my family including extended family members back home in India. Mom, Dad without your constant love, trust, support and the kind of upbringing I have had, I can't imagine myself to reach at this level in my life. All the efforts and sacrifices you have made for me are incredible and I will always be indebted to you for everything that you have done for me. My lifetime friend, philosopher and guide Dina Dalal (cousin) and Kartik Dalal (cousin-in-law), thank you for always encouraging me to pursue my dreams!

Keywords

Esophageal adenocarcinoma, Barrett's esophagus, diagnostic biomarker, glycoprotein, lectin, multiple reaction monitoring, proteomics, mass spectrometry, faradaic electrochemical impedance spectroscopy, differential pulse voltammetry

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 110106, Medical Biochemistry: Proteins and Peptides (incl. Medical Proteomics), 50%

ANZSRC code: 111202, Cancer Diagnosis, 30%

ANZSRC code: 030104, Immunological and Bioassay Methods, 20%

Fields of Research (FoR) Classification

FoR code: 1101, Medical Biochemistry and Metabolomics, 50%

FoR code: 1112, Oncology and Carcinogenesis, 30%

FoR code: 0301, Analytical Chemistry, 20%

Dedicated to My Beloved Parents (Dad Kishor Shah & Mom Dr. Malti Desai), &

My dear cousin Miss Suhagini Shah (1962 - 2012), who fought with ovarian cancer for 2 years and passed away peacefully at the beginning of my PhD

Table of contents

List of figures

List of tables

List of abbreviations

A1BG	Alpha-1B-glycoprotein
AAL	*Aleuria aurantia* lectin
aCGH	Array-comparative genomic hybridization
AFP	α-fetoprotein
ALM	Antibody overlay lectin microarray
ANOVA	Analysis of variance
AOL	*Aspergillus oryzae* lectin
APOB	Apolipoprotein B-100
APS	Ammonium persulphate
AR	Androgen receptor
Asn	Asparagine
AUROC	Area under receiver operating characteristics curve
BE	Barrett's esophagus
BMI	Body mass index
BPH	Benign prostatic hyperplasia
BPL	*Bauhinia purpurea* lectin
BSA	Bovine serum albumin
C2	Complement C2
C4B	Complement C4-B
C9	Complement component C9
$CaCl_2$	Calcium chloride
CE	Collison energy
CFB	Complement factor B
ConA	Concanavalin A from *Canavalia ensiformis*
COX-2	Cyclooxygenase-2
CV	Co-efficient of variation
DCA	Deoxycholic acid
DNA	Deoxyribonucleic acid
DSA	*Datura stramonium* agglutinin
DTT	Dithiothreitol
EAC	Esophageal adenocarcinoma
ECA	*Erythrina cristagalli* agglutinin
EDRN	Early detection research network
ELISA	Enzyme-linked immunosorbent assay
EPHA	Erythroagglutinin *Phaseolus vulgaris*
ER	Estrogen receptor
ESCC	Esophageal squamous cell carcinoma
ESI	Electrospray ionization
FDA	United Stated Food and Drug Administration
FDR	False discovery rate
FWHM	Full width at half maxima

Gal	Galactose
GalNAc	N-acetylgalactosamine
GERD	Gastroesophageal reflux disease (or disorder)
GlcNAc	N-acetylglucosamine
GNL	*Galanthus nivalis* lectin
GSN	Gelsolin
HAA	*Helix aspersa* agglutinin
HCC	Hepatocellular carcinoma
HCl	Hydrochloric acid
HGD	High-grade dysplasia
HP	Haptoglobin
HPA	*Helix pomatia* agglutinin
HPLC	High pressure or performance liquid chromatography
HRP	Horseradish peroxidase
HSA	Human serum albumin
IB	Immunoblotting
ICAT	Isotope-coded affinity tags
Ig G	Immunoglobulin G
IL	Interleukin
IM	Intestinal metaplasia
iTRAQ	Isobaric tags for relative and absolute quantification
IVD	*In vitro* diagnostic test
JAC	Jacalin from *Artocarpus integrifolia*
LAM	Lectin overlay antibody microarray
LCA	*Lens culinaris* agglutinin
LC-MS/MS	Liquid chromatography-tandem mass spectrometery
LeMBA	Lectin magnetic bead array
LeMBA-MS/MS	Lectin magnetic bead array-tandem mass spectrometry
LGD	Low-grade dysplasia
LOH	Loss of heterozygosity
LPHA	Leukoagglutinating phytohemagglutinin
LR	Likelihood ratio
m/z	Mass-to-charge ratio
MAA	*Maackia amurensis* agglutinin
$MgCl_2$	Magnesium chloride
MHC	Major histocompatibility complex
miRNA	Micro RNA
MMPs	Matrix metalloproteinases
$MnCl_2$	Manganese chloride
MRI	Magnetic resonance imaging
MRM-MS	Multiple reaction monitoring-mass spectrometry
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry

NaCl	Sodium chloride
NPL	*Narcissus pseudonarcissus* lectin
NSAIDs	Non-steroidal anti-inflammatory drugs
PCA	Principal component analysis
PCR	Polymerase chain reaction
PI	Protease inhibitor
PLS-DA	Partial least squares-discriminant analysis
PPI	Proton pump inhibitor
PR	Progesterone receptor
PSA	*Pisum sativum* agglutinin
PVDF	Polyvinylidene difluoride
QC	Quality control
QQQ	Triple quadrupole
QTOF	Quadrupole time of flight
Rb	Retinoblastoma protein
RNA	Ribonucleic acid
ROC curve	Receiver operating characteristic curve
RPM	Revolutions per minute
RT	Room temperature
RTK	Receptor tyrosine kinase
SBA	Soybean agglutinin
SD	Standard deviation
SDS	Sodiumdodecyl sulfate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Ser	Serine
SERM	Selective estrogen receptor modulator
SERPING1	Plasma protease C1 inhibitor
SILAC	Stable isotope labeling with amino acids in cell culture
SIS	Stable isotope standard
SISCAPA	Stable isotope standards and capture by anti-peptide antibodies
SNA	*Sambucus nigra* agglutinin
SNPs	Single nucleotide polymorphisms
sPLS-DA	Sparse partial least squares-discriminant analysis
STL	*Solanum tuberosum* lectin
TBST	Tris-buffered saline-0.1% Tween-20
TEMED	Tetramethylethylenediamine
Thr	Threonine
TMT	Tandem mass tags
Trp	Tryptophan
UEA	*Ulex europeus* agglutinin-I
UHPLC	Ultra high pressure or performance liquid chromatography
VLE	Volumetric laser endomicroscopy
WFA	*Wisteria floribunda* agglutinin

WGA	Wheat germ agglutinin
WHO	World health organization

Chapter 1.

INTRODUCTION

Chapter 1. Introduction

1.1 General overview

The esophagus is 18 to 25 cm long muscular tube that connects pharynx to the stomach and pushes food toward the stomach. The upper and lower esophageal sphincter prevents the backflow of the food. The upper esophageal wall is composed of striated muscle and lower part is composed of smooth muscle while combination of both striated and smooth muscle make up the middle of the tissue. The gross anatomy of the food pipe along with blood supply and lymphatic is illustrated in the Figure 1.1. In particular, arteries, veins and lymph nodes for lower esophagus, where esophageal adenocarcinoma (EAC) develops, are highlighted.

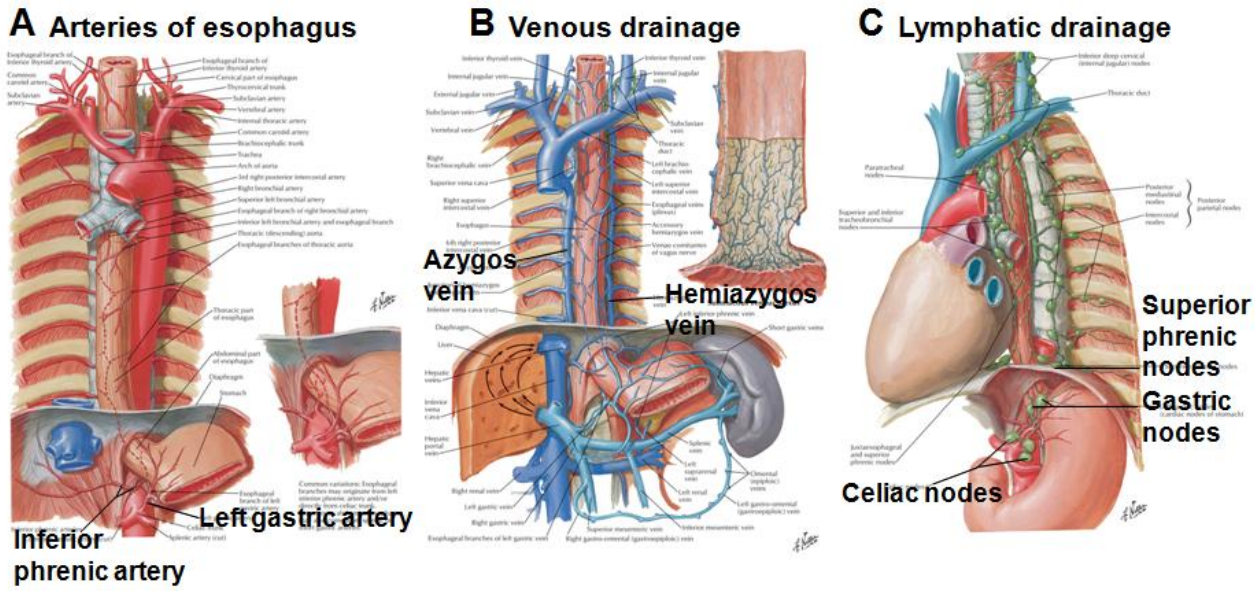


Figure 1.1. (A) Blood supply, (B) venous drainage and (C) lymph drainage of the esophagus. Adapted from Kuo and Daniela (4).

Following heart disease, cancer is the second leading cause of death globally. Four major cancer sites account for half of the cancer related mortalities: lung, colorectal, prostate in men and breast in women. In past two decades, a steady decrease in deaths of these four major site malignancies lead to an overall decrease in cancer related death rates in men and women (5, 6). In contrast, the incidence of EAC is increasing faster than any other cancer type. Before mid 1970s the incidence of EAC represented less than 5% of total esophageal cancer. Over a period of three decades, the incidence rose continuously and now almost half of the esophageal malignancy cases diagnosed are EAC type (7, 8). EAC together with esophageal squamous cell carcinoma (ESCC) is the eighth (tenth in USA) most prevalent cancer and the sixth (eighth in USA) most common cause

of cancer related death globally (6, 9). EAC is generally diagnosed at a late stage, leading to a poor 5 year-survival of less than 20% (6, 10).

EAC arises in the distal one-third of the esophagus as a consequence of gastro-esophageal reflux disease (GERD) and Barrett's esophagus (BE). In response to chronic GERD, normal stratified squamous epithelium of esophagus gets converted into metaplastic columnar epithelium, a condition called Barrett's esophagus (BE) (11-13). BE is a successful adaptation of the distal esophagus in response to chronic GERD. Typically EAC develops through a metaplasia-dysplasia-carcinoma sequence involving genetic and epigenetic modifications leading to uncontrolled cell proliferation. It is characterized by presence of intestinal metaplasia (IM) with low-grade (LGD) to high-grade dysplasia (HGD), the latter of which may develop into invasive carcinoma (14). Figure 1.2 is schematic overview of EAC development.

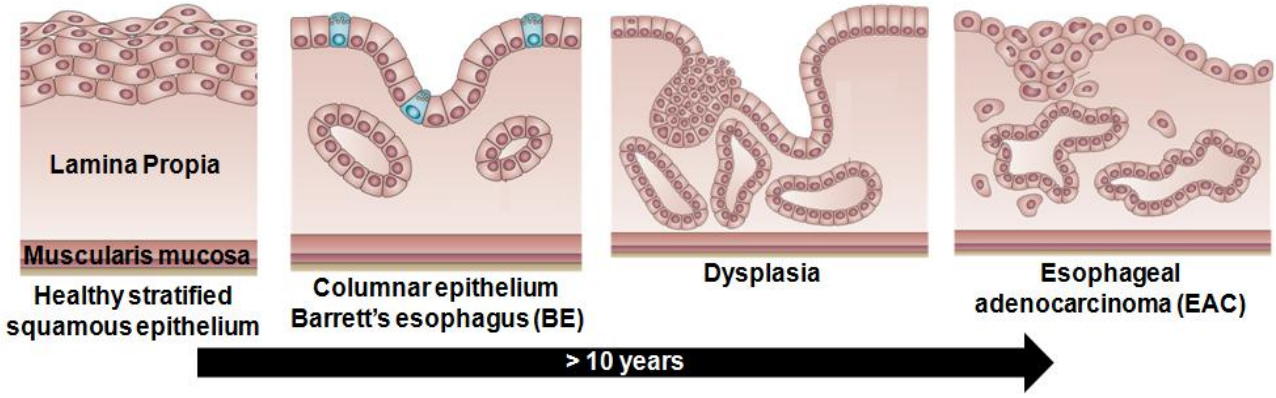


Figure 1.2. Schematic representation of EAC development. In response to chronic GERD, normal stratified squamous epithelium of esophagus is converted into acid resistant columnar epithelium, condition called Barrett's esophagus (BE). Up to 1% of BE patients develop dysplasia and EAC. The blue colored cells represent goblet cells. Adapted from Anaparthy and Sharma (15).

Majority of BE/EAC patients are asymptomatic hence it is difficult to identify early EAC. In advanced stages, patients may present symptoms like dysphagia (difficulty in swallowing), chest pain, weight loss and anemia. Significant number of patients are diagnosed accidentally when they undergo endoscopy for other gastrointestinal abnormalities (16). In majority of cases clinicians have very limited scope for the treatment as tumor has already reached advanced stage at the time of diagnosis. Treatment mainly includes surgery, radiation and chemotherapy either alone or in combination (17). Cisplatin in combination with fluorouracil is the drug of choice as a combination chemotherapy (18). Recent research on EAC has focused on understanding risk factors and identification of early diagnostic biomarkers.

1.2 Barrett's esophagus (BE) and esophageal adenocarcinoma (EAC)

Historically, esophageal ulcers resembling peptic ulcers of the stomach were first described by Albers in the year 1839 (19) and later on by many others including description of the columnar phenotype of the lower esophageal lesions (19-21). In fact until the mid 20th century, confusion exists between the terms "esophagitis" and "peptic ulcer of the esophagus". Australian born physician Norman Rupert Barrett clearly demarcated the two terms in the article published in the year 1950 (22). According to the seminal article, "peptic ulcer of the esophagus" term was mainly used by pathologists and most of these cases were in fact examples of congenital short esophagus. On contrary, esophagitis had become a blunderbuss term which covered many different pathological lesions. So to describe the phenomenon of gastric acid reflux that can give rise to ulceration of esophagus, Dr. Barrett specifically coined the term "reflux esophagitis" (22). Although Dr. Barrett was not the first one to describe the columnar lining of the esophagus, the disease later termed as Barrett's esophagus in honor of contributions made by the pioneer thoracic surgeon (23). Although there is no universally accepted definition of BE, it is formally characterized by presence of metaplastic columnar epithelium in the proximity of the gastroesophageal junction (11). Disagreement still exists between physicians about esophageal intestinal metaplasia to be prerequisite for diagnosis of BE (24). In 1975, the clear link between this columnar lined lower esophagus (or BE) and EAC was established by patient follow-up using repeated esophagoscopies (25). Out of 140 cases of extensive columnar metaplasia followed, 10% of patients developed EAC and this disease progression was irreversible and could not be stopped by an anti-reflux operation (25).

1.2.1 Epidemiology and prevalence

Out of two major types of esophageal cancers, ESCC and EAC, the latter has undergone dramatic epidemiological changes in past few decades. The overall incidence rate of ESCC has remained stable or declined since 1970s (26). On the contrary, the incidence of EAC has risen 7-fold from 3.6 cases per million in 1973 to 25.6 per million in 2006, both in men and women combined (26, 27). This rising incidences of EAC cannot be attributed to overdiagnosis due to improved imaging techniques implied in the screening program (26, 27). According to an estimate, the rate of rise in incidences has slowed down in the past decade and we may have reached a peak but this needs to be carefully monitored over next few years (27). EAC and ESCC showed marked differences in their geographical spread. EAC is more common in developed countries such as the UK (8 in 100,000 individuals) (28), Australia and the USA. Within Europe, southern Europe has the

highest EAC incidence (28). On the other side, ESCC is most common type of esophageal cancer amongst developing Asian countries (29). Racial disparity also occurs between the two types of esophageal cancer. ESCC is more prevalent amongst blacks while EAC is at least twice as common in whites as compared to other ethnic groups (30, 31). Once diagnosed, black patients showed poorer overall survival than whites (32, 33). Taken together, strong genetic and environmental factors relating to ethnicity and geographic distribution seem to be playing critical roles in the incidence of esophageal cancer. Studies also suggest possible link between socioeconomic status and prevalence of esophageal cancer phenotype (29).

As far as prevalence of gastroesophageal reflux disease (GERD) is concerned, it varies considerably according to geography. GERD affects 10-30% of the world population except East Asian countries where its prevalence has reported to be less than 10% (34). Although it is very challenging to screen the general population for the presence of BE using endoscopy, according to a study conducted in a Swedish cohort, around 1.6% of the population harbor either short or long segment BE (35). This number is estimated to be even higher around 5-6% according to mathematical modeling (36) which is very close to the results of endoscopic screening conducted in the patients undergoing colonoscopy (37, 38). Almost 95% and 80% of EAC patients have no prior diagnosis of BE or GERD respectively (39). The conversion rate from BE to EAC is estimated to be 0.12-1.0% per patient-years with large cohort studies suggesting lower conversion rate, contrary to studies conducted in small sample size (40-42). Taken together, incidences of BE/EAC have been rising and show marked differences according to geography.

1.2.2 Risk Factors

Esophageal cancer is unlikely to develop in individuals below age of 40 years, however after that its incidence rises significantly with each decade of life (32). Changing life style and food habits are primarily responsible for the dramatic epidemiological changes in EAC as described in reviews (10, 39, 43). Known EAC risk factors include accumulation of visceral fat in the abdomen (44), male gender, high intake of dietary fat and cholesterol with low intake of fruits and vegetables (45), tobacco smoking (46), reduction in *Helicobacter pylori* infections (47) and Barrett's esophagus (BE), a metaplastic change to the esophageal lining. Individuals with BE carries 30-125 times more risk for EAC development (48).

1.2.2.1 Gastroesophageal reflux disorder (GERD) and Barrett's esophagus (BE)

Patients suffering from recurrent GERD have 7 to 8 fold increased risk for developing specifically EAC without having any effect on development of ESCC, gastric cardia

adenocarcinoma or non-cardia gastric adenocarcinoma (49). The rate of EAC is even higher in individuals with bile reflux. Erosive reflux disease as compared to non-erosive reflux is responsible for esophageal inflammation and ulceration leading to EAC (43). However, GERD alone is not capable of explaining dramatic rise in the EAC incidences. Furthermore, only about 10% to 20% of GERD patients develop BE and this percentage even reduces when we consider EAC suggesting involvement of other risk factors (50).

Barrett's esophagus (BE) is a successful adaptation in response to chronic GERD and characterized by replacement of normal stratified squamous epithelium with metaplastic columnar epithelium (39). Amongst three different types of columnar epithelium named as intestinal, cardia gastric fundic and gastric junctional, only intestinal metaplasia seems to be associated with increased cancer risk and histologically diagnosed by presence of mucous secreting goblet cells (50). Being a strongest risk factor for developing EAC, typically BE patients carry 100 folds increased risk for developing EAC equally for both men and women. The process of EAC development from BE typically involves stages like low-grade dysplasia (LGD), high-grade dysplasia (HGD), early EAC and invasive carcinoma (43, 51). The predominance of BE is almost double in men with typical age of diagnosis 50-59 years (51, 52).

1.2.2.2 Obesity

At the same time while incidences of EAC are increased, obesity has reached epidemic level globally with more than 1.7 billion adults overweight and 300 million people are clinically obese (53). Epidemiological studies have shown strong implication between increasing body mass index (BMI) and risk of EAC, with 2-3 fold increased risk for those with BMI $\geq$ 30 kg/m^2 and 1.5-2 fold in those with BMI = 25.0 to 29.9 kg/m^2 independent of reflux (39, 44, 54). Obesity can cause hiatal hernia and can simply provoke reflux through increasing intra-abdominal pressure. Independent of reflux, obesity can also lead to EAC by other mechanisms (50, 55). It has been realized that rather than only weight, it is actually distribution of the fat that affects the disease progress. Visceral fat is metabolically more active and has been associated with high levels of leptin, interleukin-6 (IL-6), tumor necrosis factor-α (TNF-α) and overall low serum level of adiponectin (56). This chronic low grade inflammation state in obesity can increase cell proliferation leading to cancer (39). Also high serum leptin levels are associated with BE in only men and not in women which may partly explain the male predominance of the disease (57).

1.2.2.3 Age and gender

EAC follows similar age distribution to that of other gastrointestinal malignancies with median age at the time of diagnosis of 60 years (54). Male: female occurrence ratio for EAC is reported to be 3-7:1, making it male predominating disease. The detailed underlying mechanism for this anomaly is yet to be discovered (58). One can obviously speculate a relationship between sex hormones and EAC development. Estrogen may be protecting female in early ages against EAC or it may be testosterone which predisposes men towards the disease.

Three independent studies looking at expression levels of androgen receptor (AR) using immunohistochemistry in healthy, BE and EAC tissue samples (total n = 60) concluded no relationships between AR staining and either disease status or survival of EAC patients (59-61). In agreement with the tissue staining, the esophageal adenocarcinoma cell lines OE19 and OE33 do not express AR (62). However, stromal cells express AR which raises the possibility of indirect role of AR in BE/EAC (62). The paracrine effect of stromal AR and exposure with androgens i.e. serum testosterone can be mediated by fibroblast growth factor (FGF)/fibroblast growth factor receptor-1 (FGFR-1) axis (62). Supporting this idea, OE19 xenografts grew faster in male mice in comparison with female mice along with high intensity staining for FGFR in male mice (62). In a large scale study including prostate cancer patients undergoing antiandrogen therapy it has been concluded that there is no relationship between antiandrogen therapy and secondary esophageal adenocarcinoma development (63). Another clinical study on prostate cancer patients concluded that risk of developing EAC, and not ESCC is lower in prostate cancer patient. This may be due to etiological factors related to prostate cancer or antiandrogen therapy (64). Overall based on the available evidences, testosterone alone cannot explain the gender bias for the disease. It appears that testosterone/AR may not have any direct role to play in EAC development; however, these studies are few in number and limited in scope hence require further confirmation.

There is continuous decrease in male: female incidence ratio of EAC from more than 10:1 in some age groups of less than 50 years of age to 4:1 in mid 80s suggesting age associated factors related to sex may be playing role (58). Importantly role of ovarian hormones cannot be ruled out as ovarian function deteriorates with age in female. Estrogen has protective effect against certain GI tumors while in case of breast and ovarian cancer it promotes tumor progression however there is no such clear information regarding EAC (65, 66). This tissue specific anomaly of estrogen action is determined by distribution of estrogen receptors (ER) namely Estrogen Receptor α (ERα), Estrogen Receptor β (ERβ) and its downstream signaling. ERα is considered to be anti-apoptotic

while ERβ acts as a pro-apoptotic and both the receptors regulate expression of different set of genes in multiple possible ways (65). ERα and progesterone receptor (PR) are not expressed in EAC tissues (67-69). ERβ is present in EAC tissues (67-69) and its expression levels are correlated with tumor stage (69). ERβ localization also differs in BE and EAC tissue samples, in case of former they are present only in nucleus while EAC samples show staining for ERβ throughout cytosol and/or nucleus (68). Estrogen has also been found to protect against chemical carcinogen induced esophageal cancer in rat model (70). In another rat model using nitric oxide (NO), it has been confirmed that male rats showed more tissue damage in response to NO and it was suppressed by presence of 17-β-estradiol (71). *In vitro*, 2-methoxyestradiol, an endogenous by product of 17-β-estradiol decreases cancer cell proliferation and migration by regulation of Bax/Bcl-2 and β-catenin-E-cadherin signaling pathway. However, these effects seem to be independent of ERs (72). A clinical study has shown that post-menopausal women taking tamoxifen as a therapy for breast cancer has 60% statistically non-significant more risk of developing EAC (73). Tamoxifen is a selective estrogen receptor modulator (SERM) with antiestrogenic effect in breast in contrast to its estrogenic effect in bone and endometrium (73). If estrogen actually has a protective role against EAC then post-menopausal hormone use should decrease the risk for developing EAC but actually no correlation was found between post-menopausal hormonal medication and EAC (74).

Apart from regulating reproduction related functions, estrogen is also implicated in the overall physiological metabolism in both sexes (65). High level of estrogen in premenopausal women by interaction with ERs in the brain and with leptin signaling can drive subcutaneous fat distribution which is less harmful in terms of EAC (61, 70). However, in the presence of relatively low levels of estrogen after menopause, visceral fat starts to accumulate predisposing women for increased risk of EAC (63, 70, 71, 73). The scenario for male is less clear as the majority of studies were done in females (65).

In summary, estrogen seems to play a role in EAC. Further studies in this regard may provide useful links to understand the male predominance of the disease. Depending upon the expression of the receptors and its downstream signaling in the tumor, estrogen can play either proliferative or apoptotic role which can be targeted using already available estrogen analogues, SERM or anti-estrogen to establish new supportive therapy for the deadly disease.

1.2.2.4 Alcohol and smoking

Alcohol is a putative risk factor for the ESCC while there is no such correlation for EAC (46, 54, 75, 76). In contrast to alcohol, which is not considered to be a risk factor for the disease,

smokers have almost two times higher risk for EAC. The increased risk remains even a long time after smoking cessation (46, 75, 76).

1.2.2.5 Helicobacter pylori Infection

Plethora of evidences suggests inverse relationship between *H. pylori* infection and EAC (47, 55, 77). The mechanism proposed is due to the ability of *H. pylori* to induce atrophic gastritis which results in a decrease in gastric acid production. *H. pylori* produces ammonia from urea, changing the nature of refluxate which reduces the chances of esophageal damage due to reflux (10, 54). The incidence of infection is declining in the developed countries which may contribute to the increased incidence of EAC (54).

1.2.2.6 Diet

Low intake of antioxidant rich foods, fibers, fruits and vegetables is associated with the incidence of the EAC (45, 78, 79). Furthermore, high intake of dietary fat, dietary cholesterol and animal protein are found to be potential risk factors for EAC (45).

1.2.2.7 Medication

Non-steroidal anti-inflammatory drugs (NSAIDs) consumption is shown to have protective role against development of EAC as well as ESCC. In fact, individuals taking regular Aspirin have shown clinical benefits with respect to esophageal malignancy (80-83). These agents protect against the malignancy by reducing Cyclooxygenase-2 (COX-2) mediated inflammation (84).

Proton-pump inhibitors (PPIs) which reduce gastric acid production may protect against gastric acid mediated damage to esophagus hence show protective role against EAC. However, the data shows mixed results (39).

Overall GERD and related BE, obesity, smoking, change in dietary habits and decrease in *H. pylori* infection are responsible for increase in the EAC cases.

1.2.3 Pathophysiology

Significant numbers of EAC arise as a consequence of BE and GERD. The development of BE is considered to be a two-step process (14). In response to chronic GERD, the first step involves the replacement of normal esophageal squamous mucosa to a simple columnar epithelium called cardiac mucosa (85-87). This first step is very rapid and can occur within 1-2 years. Cardiac mucosa is an unstable epithelium which can express gastric genes leading to gastric differentiation and formation of oxyntocardiac mucosa (88). This is considered to be a favorable change and it is not

premalignant. Alternatively, cardiac mucosa can express intestinal genes which cause the formation of goblet cells making this a pre-malignant mucosa BE. Typically goblet cells formation takes minimum of 5-10 years (89).

Different hypothesis prevail for cellular and molecular changes responsible for making columnar and more acid resistant epithelium in BE (12). During embryonic development, esophagus is initially lined by columnar epithelium which during late embryogenesis is replaced by squamous epithelium through the process of transdifferentiation (90, 91). On this basis it has been believed that columnar epithelium again develops in the esophagus as a consequence of GERD by reversal of the developmental pathway. In fact *in vitro* experiments have confirmed these changes (92-94). Contrary to the transdifferentiation hypothesis, the columnar epithelial cells may originate from the stem cells itself (12). The reprogramming of the stem cell situated in the basal layer of the normal squamous epithelium gives rise to metaplastic columnar type cells (12). Alternatively, stem cells residing in the submucosal glands of the esophagus or bone marrow derived stem cells can also give rise to metaplastic tissue (95, 96).

At the molecular level, the homebox gene CdX2 is essential for intestinal differentiation while CdX1 specify columnar cell phenotype (97, 98). Exposure of esophageal squamous epithelium to acid or bile salts can increase expression of these genes leading to phenotypic changes including formation of crypt like structure and expression of intestinal genes like villin and mucin (99, 100). SOX9 protein is shown to be expressed specifically in BE/EAC which plays a role in the formation of intestinal type goblet cells (101, 102). The role of developmental signaling pathways like Wnt signaling, Notch signaling, Hedgehog and bone morphogenic protein (Bmp) 4 pathways are also described in making Barrett's metaplasia (12).

Development of EAC from BE follows metaplasia-dysplasia-adenoma carcinoma sequence. Disruption of p16/Rb pathway and p53 pathway can drive molecular changes from BE towards EAC (12, 39). The pathogenesis involves development of early and late phase dysplasia resulting into invasive carcinoma (50, 103). The dysplasia is associated with the architectural and cytological changes in the columnar epithelium. Variations in nucleus size and shape, increased nuclear-to-cytoplasm ratio, hyperchromatism, increased numbers of abnormal mitoses, villiform configuration of the mucosal surface and architectural abnormalities like budded, branched, crowded, or irregularly shaped glands are hallmark of dysplasia (103).

1.2.4 Current diagnosis scenario

Esophageal cancer patients do not present any symptoms until the disease is in advanced stages. Currently the majority of EAC patients are diagnosed symptomatically when patients present symptoms like dysphagia, anemia, fatigue, vomiting, and weight loss without trying etc (104). By this time, the tumor is already locally metastasized hence patients often show very poor survival (16). In contrast, more than 90% patients survive beyond 5-years if the diagnosis is made in the early stages of HGD before EAC develops (105).

In order to detect pathological changes leading to EAC development before onset of disease, current clinical practice involves endoscopic screening of high risk GERD patients and to characterize the degree of dysplasia in biopsy samples collected during endoscopy (106, 107). Patient enrollment into an endoscopic screening program may be facilitated by a patient questionnaire of self-evaluated symptoms/complications (108, 109). Once enrolled into the screening program, patient undergoes endoscopy-biopsy every 3 months to 2 years depending on the degree of dysplasia, during which 4 quadrant biopsy samples are taken every 1-2 cm and evaluated for histological changes by expert pathologists (106, 107). As a significant number of patients histologically diagnosed with HGD develop EAC, endoscopic mucosal ablation or esophageal resection (esophagectomy) are options to stop further disease progress in those high risk patients (110, 111). Significantly improved survival is observed in patients diagnosed at an early stage during surveillance endoscopy program as compared to symptomatically diagnosed EAC (112-115).

Although current screening methodology shows promise, outcome of endoscopy-biopsy in many cases is non-reproducible due to interobserver variability and sampling error (111, 116). Furthermore, histological dysplastic changes may be patchy and present heterogeneously in tissue sample. This makes the diagnosis challenging, especially in the early stages of transition to LGD (111, 117). In up to 40% of patients, invasive cancer has been found in resected tissue despite of endoscopic examination was negative for the malignancy (118). Moreover, false positive results also occur, meaning despite intramucosal carcinoma in a biopsy, the subsequently resected tissue has no signs of carcinoma (111). This evidence suggests dysplasia grading is an imperfect measure of cancer risk.

Despite extensive screening with currently available techniques, more than 80% of EAC are diagnosed without any prior diagnosis of BE or GERD (119, 120). According to an estimate more than 80% of BE are undiagnosed hence not getting benefit of the screening program (121). On the

other hand a large proportion of patients undergoing routine biopsy screening do not progress to EAC (39). These suggest inability of current methodologies in screening population to detect high risk patients and to distinguish between disease progressors from non-progressors. In addition, the screening procedure is not very cost-effective (122).

1.2.5 Patient screening, surveillance, management and treatment

Table 1.1 summarizes current scenario for managing patients with GERD, BE, dysplasia and EAC. BE, being an asymptomatic condition by itself, is very difficult to diagnose symptomatically in the patients. So at present, older obese Caucasian males suffering from chronic GERD who has highest likelihood of BE are screened using endoscopic techniques for metaplastic/dysplastic changes (110). However, 44% of BE patients in the Swedish patient cohort did not present any GERD related symptoms and would have been not included in the BE screening according to current guidelines (35). Although the effectiveness of screening to reduce EAC associated mortality is not clearly established, emergence of newer techniques such as esophageal capsule endoscopy could provide a non-invasive and convenient way to diagnose columnar lining of esophagus (123).

For early detection of EAC in patients diagnosed with BE, routine patient surveillance is performed using endoscopy-biopsies method and the degree of dysplasia is determined by an expert pathologist as a measure of disease progression. The degree of dysplasia determines the frequency of screening. Patients with LGD are followed-up every 6 months while HGD requires follow-up every 3 months (110). Patients diagnosed with HGD can be intervened using either endoscopic mucosal resection technique or mucosal ablation therapy such as photodynamic therapy, radiofrequency balloon catheter ablation, or thermal techniques and showed improved survival (124). Once the patient is diagnosed with EAC, treatment depends upon the stage of the disease. For early stage EAC, surgery is the best option. When the disease is locally metastatic, neoadjuvant chemotherapy or neoadjuvant chemoradiation therapy along with surgery has become standard of care. Once EAC is metastasized to distant organs there are not many treatment options available. So patients are managed using chemotherapy and supportive care (17). Once EAC is developed then irrespective of stage and treatment, 5-year survival is less than 20% unfortunately.

Table 1.1. Current strategies to manage and treat patients in the progression from GERD to EAC.

Condition	Patient management and treatment
GERD	Majority of patients are diagnosed symptomatically. Severity and nature of reflux can be determined using impedance-pH measurement. PPIs are prescribed to control reflux.
Barrett's esophagus	High risk patients undergo endoscopy-biopsies screening to diagnose columnar lining of esophagus. Once diagnosed patients are enrolled into surveillance program to detect early dysplastic changes. However around 44% of patients diagnosed with BE have no prior symptomatic representation of GERD which means they don't get benefit of screening. Patients are managed with PPIs and acid control therapies.
Early grade dysplasia	Patients are followed-up with surveillance every 6 months to carefully monitor disease progression.
Late grade dysplasia	Patients undergo endoscopy-biopsies every 3 months to identify early neoplastic changes. Patients can be treated at this stage using ablation techniques.
Esophageal adenocarcinoma	The line of therapy is decided according to TNM staging of the tumor. Early stage patients are treated using surgery alone while locally metastatic patients are treated with surgery and chemotherapy or chemoradiotherapy. For metastatic disease, supportive care is the only option.

Taken together, an ideal strategy will be to identify BE population and monitor their progression for LGD/HGD when patients can be treated to reduce morbidity and mortality associated with EAC. To overcome challenges in the current endoscopy-biopsies based screening program, adjunct use of biomarker has been proposed.

1.3 Biomarkers

According to United States National Institutes of Health a biological marker or a biomarker is "A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (125). Biomarkers have clinical application during all stages of the disease management starting from early diagnosis, monitoring disease progression and predicting therapeutic response (126). A biomarker addresses a clinically relevant question and provides valuable information that can be used for patient management and decision making by the clinician. Ideal biomarker test should be accurate, non-invasive, easy to perform, quick, and informative (127).

Based on sequence of events from exposure to disease development and therapeutic response (126), biomarkers can be classified into exposure (128), susceptibility [*BRCA1/2* mutations in breast cancer] (129), diagnostic [prostate-specific antigen in prostate cancer] (130), predictive [estrogen and progesterone receptor to predict response to endocrine therapy in breast cancer] (131), pharmacodynamic [receptor tyrosine kinases (RTKs) phosphorylation measurement to monitor pharmacological effect of RTK inhibitors] (132), and prognostic [higher CA 19-9 levels are associated with poor survival in pancreatic cancer] (131) markers. Any biomolecule DNA, RNA, protein, metabolite or lipid including circulating tumor cell and imaging measurement can qualify as a biomarker.

1.3.1 Biomarker discovery and development phases

National Cancer Institute Early Detection Research Network (EDRN) guidelines outline biomarker discovery and development to a 5 phase process summarized here (133) and depicted in Figure 1.3. This thesis focused on Phase I and early Phase II of the 5 phase development.

Phase I - Preclinical exploratory study: It compares normal vs. cancer samples (body fluids/tissue) using technologies such as genomics, microarray expression, proteomics, immunohistochemistry or immunoblotting to detect significant changes in proteins/genes/metabolites between the groups.

Phase II - Clinical assay development and validation: It is aimed at developing a clinical assay using a minimally invasive sample collection method. The assay is meant to be robust, reproducible and suitable for stored clinical samples to be used in later phases of development. Sensitivity and specificity are determined at this stage. Sensitivity of an assay is the ability of the test to correctly identify those patients with the disease while specificity of a test refers to the ability of an assay to correctly assign patients without the disease. At the end of this phase one should expect high specificity and sensitivity for the assay. However, it remains to be determined how early the biomarker can predict the disease.

Phase III - Retrospective longitudinal repository studies: The assay is applied on prospectively collected stored samples to determine ability of biomarker to detect the disease before clinical presentation. If so then criteria for positive screening is determined for future use.

Phase IV - Prospective screening: The test is prospectively applied to real population to detect extent and characteristic of disease detected by the biomarker. This phase gives positive predictive value for the test and gives idea about feasibility for last phase of control trials.

Phase V - Cancer control studies: It comprises of large scale clinical trial to determine impact of new screening process on the disease burden in the community.

1.4 Biomarkers for BE/EAC

In transit from IM to LGD to HGD to EAC, cells acquire abilities to become self sufficient for growth, evade apoptosis, proliferate uncontrollably, promote angiogenesis, invade underlined epithelium and start to metastasize. These changes are accompanied with histological changes in tissue architecture, genomic instability, development of tumor microenvironment, modulation of immune response and therefore reflected in body fluids (serum/plasma/mucus/urine) or tissue samples and differentiate in terms of their genome/proteome/metabolome profile (134). Thus, a biomarker can be from any of these sources and reflect underlying pathological or homeostatic changes. Table 1.2 summarizes different classes of biomarkers proposed for BE/EAC.

Table 1.2. Comprehensive summary of different classes of BE/EAC biomarkers.

Biomarker Class	**References**
Tissue Biomarkers	
Genomic abnormalities (Ploidy and LOH)	(135-139)
DNA methylation	Refer to Table 1.3
SNPs/expression array studies	Refer to Table 1.4
Inflammatory Markers COX-2 NF-κB Cytokines MMPs	 (140-146) (147-150) (148, 150-156) (157-163)
Cell cycle abnormalities	(164-166)
miRNA	Refer to Table 1.5
Glycosylation changes	(167-170)
Circulatory Biomarkers	
DNA methylation changes	(171-174)

Biomarker Class	References
Glycan alterations	(175-180)
Metabolic profiling	(181-184)

With respect to EAC, none of the biomarkers, including high grade dysplasia, have been evaluated in phase V while very few are evaluated in phase III and IV. Figure 1.3 summarizes proposed EAC biomarkers and how well they are characterized in the process of biomarker discovery. The following sections will discuss some of the classes of BE/EAC biomarkers.

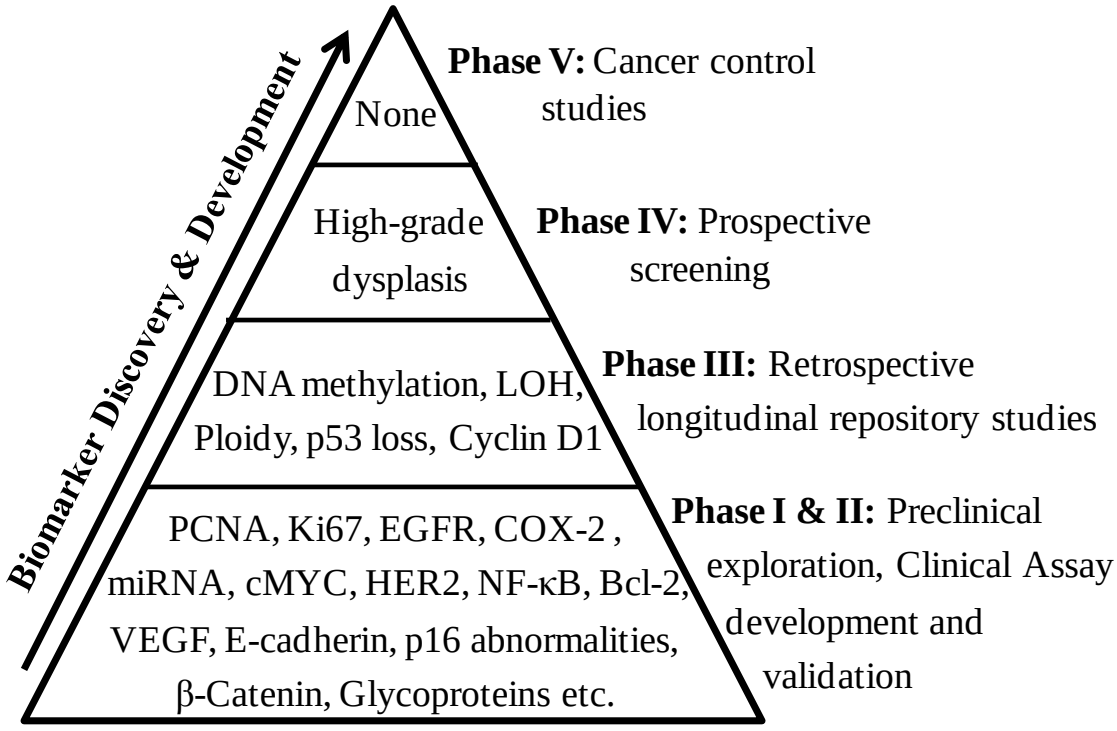


Figure 1.3. Summary of current BE/EAC biomarkers with respect to EDRN clinical phase of development.

1.4.1 Genomic instability

Many groups have studied genomic instability induced by aneuploidy, tetraploidy, DNA methylation, allelic loss and demonstrated some predictive power for these changes. A role for hypermethylation in the promoter regions of tumor-suppressor genes during the development of EAC has also been well-established. Table 1.3 summarizes DNA methylation changes associated with metaplasia-dysplasia-carcinoma development. In the majority of patients, methylation changes are acquired very early during EAC development, hence these alterations could be used as an early diagnostic biomarker. Apart from discriminating patients at different stages of EAC development,

DNA methylation signatures may be useful as predictors for progression from BE to EAC (185, 186), and to monitor response to chemotherapy and survival in EAC patients (187, 188).

Although the individual genomic abnormality has the potential to diagnose disease at different stages, best results are obtained when they are used in combination (135-137). Loss of heterozygosity (LOH) at chromosome 9p and 17p locus are considered to be early events during BE pathogenesis (138). If present with other chromosomal alterations like aneuploidy and tetraploidy, it increases 10-year risk for development of EAC from 12% to ~80% (139). However, with the current flow cytometry technology, it is technically very challenging for clinical laboratories to assess these genomic biomarkers in the patient samples which limits widespread use of these biomarkers in the clinic.

Alternatively, genomic alterations can be detected at the protein level using immunohistochemistry. One of the most common and earliest genomic abnormalities occurs at chromosome 17p which codes for tumor suppressor p53 protein. Loss of p53 protein expression in tissue samples correlates very well with disease progression (189). However, as p53 expression only reflects alterations at one particular gene, it has lower predictive value as compared to techniques monitoring multiple genomic abnormalities. Furthermore, sensitivity drops as mutations or deletions at genomic level may not necessarily be detected at the protein level (190).

In line with genomic abnormalities described above, single nucleotide polymorphisms (SNP) based genotyping can also stratify cancer risk in BE patients. As summarized in Table 1.4, in the past decade, several studies conducted using advanced genomic techniques such as an array-comparative genomic hybridization (aCGH) and SNP arrays confirmed previously reported copy number alterations and identified novel genomic loci undergoing changes during process of metaplasia-dysplasia-carcinoma development (191-197). It has been shown that as the disease progress from early to late stages, SNP abnormalities increases from ~2% to ~30% (191, 194). The total number of SNP alterations in tissue samples is tightly correlated with previously reported DNA abnormalities such as aneuploidy, copy number alterations and LOH highlighting the application of SNP based genotyping to assess genomic abnormalities (191-197). Thus, SNP based genotyping provide an alternative way to assess genomic abnormalities during EAC pathogenesis.

Studies on gene expression changes in EAC have been propelled by recent progress in genomic technologies, each identifying unique sets of gene expression profile which can be used as a biomarker panel for disease diagnosis, prognosis or to predict response to therapy (Table 1.4). Moreover, determination of the gene expression changes has been extremely helpful to understand

detailed pathogenesis and will form basis for developing future therapies. However, future validation using independent sample cohorts will be necessary for the majority of these potential biomarkers.

Apart from genomic abnormalities associated with the disease progression, inheriting genetic factors are also implicated for EAC development. Risk for BE/EAC, GERD is increased by 2-4 fold when a first-degree relative is already affected by any of these conditions (198). A study conducted by The Esophageal Adenocarcinoma Genetics Consortium and The Wellcome Trust Case Control Consortium identified link between SNPs at the MHC locus and chromosome 16q24.1 with risk for BE (199). They also identified SNPs associated with body weight measures were present with more than expected frequency in BE samples supporting epidemiological findings regarding obesity as a risk factor for BE and EAC (199). Wu and colleagues examined the relationship between presence of risk genotypes and onset of EAC. They identified 10 SNPs associated with the age of EAC onset. Genes associated with 5 of the 10 SNPs identified were known to be involved in apoptosis (200).

The comparative genomic analysis between EAC and ESCC reported by Agrawal and colleagues (201) confirmed previously very well described association of p53 gene mutations with esophageal cancer development. The authors also performed comparative genome-wide analysis between matched BE and EAC patient tissue samples and concluded that the majority of genomic changes occur early during EAC development, at the stage of BE (201). Similar conclusions were made by next-generation sequencing of matched biopsy samples obtained from the same patient at the stage of BE and EAC (202). The authors also identified *ARID1A* as novel tumor-suppressor gene and around 15% of EAC patient showed loss of ARID1A protein in tissue samples. *In vitro* studies suggested it to be associated with cell growth, proliferation and invasion (202). Recently published high-resolution methylome analysis has provided first evidence for methylation changes at genomic regions that encode non-coding RNAs. The authors identified long non-coding RNA, AFAP1-AS1 to be severely hypomethylated in BE and EAC tissue samples. Silencing of which significantly reduced aggressiveness of EAC cell lines OE33 and SKGT4 (203).

Recently published two major cancer genome sequencing studies analyzed tissue samples collected from more than 300 patients in total and provided deeper insights into the genomic abnormalities associated with EAC pathogenesis (204, 205). Dulak and colleagues sequenced 149 tumor-normal pairs using exome sequencing including 15 pairs were subjected to whole genome sequencing. They identified high prevalence of A>C transversions at AA dinucleotides in EAC

tissue as compared to paired normal samples. Apart from identifying mutations in *TP53*, *CDKN2A*, *SMAD4*, *ARID1A* and *PIK3CA* genes which is very well-known, they identified 21 potentially novel genes showing mutation in EAC. Functional analyses suggested activation of the RAC1 pathway in EAC condition (204). In order to study the timing of these mutations in progression towards EAC, Weaver and colleagues analyzed mutation status of 26 genes across 66 non-dysplastic BE (NDBE) patients (no signs of dysplasia for median follow-up for 58 months) and 43 HGD samples. Notably, more than half of NDBE patients harbor mutations that are found in EAC samples suggesting genomic abnormalities occur very early during the pathogenesis (205).

Taken together, genomic abnormalities play key roles during each stage of transformation from normal squamous epithelium to EAC. The majority of key mutations are already acquired at the metaplastic stage of BE and only few driver mutations lead to progression of dysplasia and EAC (201, 202, 205). This finding raises the possibility of more functional level changes (e.g. protein expression, protein glycosylation, metabolic changes etc.) driven by early genomic alterations to be associated with development of dysplasia/carcinoma from metaplastic condition.

Table 1.3. Summary of hypermethylated genes during BE/EAC development.

Gene	Location	Function	Method	Number (percentage) of samples showing hypermethylation or study findings					Ref.
				Normal	BE	LGD	HGD	EAC	
p16 (or *CDKN2A* or *INK4A*)	9p21	Cyclin dependent kinase inhibitor	Methylation specific PCR	5/9 (56%)	14/18 (77%)	-	-	18/21 (85%)	(206)
			Methylation sensitive single-strand conformation analysis	0/10 (0%)	4/12 (33%)	3/11 (27%)	3/10 (30%)	18/22 (82%)	(207)
			Methylation specific PCR	0/17 (0%)	14/47 (30%)	9/27 (32%)	10/18 (56%)	22/41 (54%)	(208)
			Methylation specific PCR	2/64 (3%)	14/93 (15%)	-	-	34/76 (45%)	(209)
			Methylation specific PCR	-	3/10 (30%)	-	-	5/11 (45%)	(210)
			Methylation specific PCR	-	27/41(66%)	21/45 (47%)	17/21 (81%)	65/107 (61%)	(211)

			Methylation specific PCR	0%	1/15 (7%)	4/20 (20%)	12/20 (60%)	8/15 (53%)	(212)
			Methylation specific PCR	Separately determined exon 1 and exon 2 methylation. 5/16 (31%) exon-1, 8/16 (50%) exon 2 in EAC patient samples showed hypermethylation. Exon 2 methylation correlates with stage of the tumor (p=0.01)					(213)
O^6-Methylguanine-DNA Methyltransferase (or *MGMT)*	10q26	DNA repair	Methylight technique	2/10 (20%)	8/13 (62%)	-	-	84/132 (64%)	(214)
			Methylation specific PCR	6/29 (21%)	24/27 (89%)	13/13 (100%)		37/47 (79%)	(215)
APC	5q21-q22	Wnt/β-catenin signaling	Methylation specific PCR	0/17 (0%)	24/48 (50%)	14/28 (50%)	14/18 (78%)	20/32 (63%)	(208)
			Methylation-sensitive single-strand conformation analysis and methylation-sensitive dot blot assay	0/16 (0%)	11/11 (100%)	-	-	20/21 (95%)	(216)
				8 out of 14 histologically normal gastric mucosa adjacent to EAC showed significantly different methylation of *APC* promoter.					(217)
GSTM2	1p13.3	Anti-oxidants	Bisulfite	<10%	~50%	~55%		69%	(218)

GSTM3		and protection against DNA damage	pyrosequencing (Sample size: EAC-100, BE-11, Dysplasia-11, Normal esophageal/gastric mucosa-37)	<10%	~13%	~37%	15%	(218)
GPX7	1p32			<10%	~18%	~80%	67%	(218)
GPX3	5q23			<10%	~90%	~88%	62%	(218)
			Methylation specific PCR	2/12 (17%)	13/21(62%)	9/11 (82%)	30/34 (88%)	(219)
TIMP-3	22q12.3	MMP inhibitor	Methylight technique	1/8 (13%)	6/12 (50%)	- -	9/13 (69%)	(220)
Death-associated protein kinase (DAPK)	DAPK1: 9q21.33 DAPK2: 15q22.31 DAPK3: 19p13.3	Tumor-suppresor and mediator of apoptosis	Methylation specific PCR	4/20 (20%)	14/28 (50%)	11/21 (53%)	21/35 (60%)	(221)
Tachykinin-1 (TAC1)	7q21-22	Smooth muscle contractility, epithelial ion transport, vascular permeability	Methylation specific PCR	5/67 (7.5%)	38/60 (63.3%)	12/19 (63.2 %) 11/21 (52.4 %)	41/67 (61.2%)	(222)

		and immune function							
Reprimo	2q23	Regulates p53-mediated cell cycle arrest in G2 phase	Methylation specific PCR	0/19 (0%)	9/25 (36%)	-	7/11 (64%)	47/75 (63%)	(223)
E-Cadherin	16q22.1	Ca^{+2}-dependent intercellular adhesion and maintains normal tissue architecture	Methylation specific PCR	0/4 (0%)	-	-	-	26/31 (84%)	(224)
SOCS-3	17q25.3	Inhibits cytokine signaling	Methylation specific PCR	0%	4/30 (13%)	6/27 (22%)	20/29 (69%)	14/19 (74%)	(225)
SOCS-1	16p13.13			0%	0/30 (0%)	1/27 (4%)	6/29 (21%)	8/19 (42%)	
Secreted frizzled-related proteins (SFRP)									
SFRP1	8p11.21	Wnt antagonist	Methylation specific PCR	7/28 (25%)	30/37 (81%)	-	-	37/40 (93%)	(226)
SFRP2	4q31.3			18/28 (64%)	33/37 (89%)	-	-	33/40 (83%)	

SFRP1	8p11.21		Methylation-sensitive single-strand conformation analysis and methylation-sensitive dot blot assay	1/12 (8%)	6/6 (100%)	-	-	23/24 (96%)	(216)
SFRP2	4q31.3			11/15 (73%)	6/6 (100%)	-	-	19/25 (76%)	
SFRP4	7p14.1		Methylation specific PCR	9/28 (32%)	29/37 (78%)	-	-	29/40 (73%)	(226)
SFRP5	10q24.1			6/28 (21%)	27/37 (73%)	-	-	34/40 (85%)	
Plakophilin-1 (PKP1)	1q32	Cell adhesion and intracellular signaling	Methylation specific PCR	5/55 (9.1%)	5/39 (12.8%)	-	1/4 (25%)	20/60 (33.3)	(227)
GATA-4	8p23.1-p22	Transcription factor and regulate cell differentiation	Methylation specific PCR	0/17 (0%)	-	-	-	31/44 (71%)	(228)
GATA-5	20q13.33			0/17 (0%)	-	-	-	24/44 (55%)	
CDH13 (or H-cadherin or T-cadherin)	16q24	Cell adhesion	Methylation specific PCR	0/66 (0%)	42/60 (70%)	15/19 (78.9 %)	16/21 (76.2)	51/67 (76.1%)	(229)

NELL-1 (nel-like 1)	11p15	Tumor suppressor	Methylation-specific PCR	0/66 (0%)	28/60 (46.7%)	8/19 (42.1 %)	13/21 (61.9%)	32/67 (47.8%)	(230)
Eyes Absent 4	6q23	Apoptosis modulator	Methylation-Specific PCR	2/58 (3%)	27/35 (77%)	-	-	33/40 (83%)	(231)
A-kinase anchoring protein 12 (or Gravin orAKAP12)	6q24-25.2	cell signaling, adhesion, mitogenesis and differentiation	Methylation-Specific PCR	0/66 (0%)	29/60 (48.3%)	10/19 (52.6 %)	11/21 (52.4%)	35/67 (52.2)	(232)
Vimentin	10p13	Cytoskeleton protein	Methylation-specific PCR	0/9 (0%)	10/11 (91%)	-	5/5 (100%)	21/26 (81%)	(233)
RUNX3	1p36	Transcription factor	Methylation-specific PCR	1/63 (2%)	23/93 (25%)	-	-	37/77 (48%)	(209)
HPP1	19pter-p13.1	Tumor-suppressor		2/64 (3%)	41/93 (44%)	-	-	55/77 (71%)	
3-OST-2	16p12	Sulfotransferase enzyme		1/57 (2%)	47/60 (78%)	-	-	28/73 (38%)	
Wnt inhibitory factor-1 (WIF-1)	12q14.3	Wnt antagonist	Methylation specific PCR	81% of BE patients suffering from EAC showed hypermethylated WIF-1 as compared to 20% BE patients without EAC					(234)

CHFR (checkpoint with forkhead associated and ring finger)	12q24	Mitosis check point protein	Bisulfite pyrosequencing	EAC samples 31% (18/58) showed significantly higher CHFR promoter methylation as compared to normal samples (p=0.01).	(235)
Metallothionein 3 (or *MT3*)	16q13	Metal homeostasis and protection against DNA damage	Bisulfite pyrosequencing (Sample size: Normal-33, BE-5, EAC-78)	Identified two regions (R2 and R3) of CpG nucleotides which showed significantly higher methylation in EAC as compared to normal epithelium (FDR<0.001). Increased DNA methylation of MT3 promoter R2 correlates with advanced tumor stage (p=0.005) and lymph node metastasis (p=0.03). DNA methylation of MT3 promoter R3 correlates with tumor staging (p=0.03) but not with lymph node status (p=0.4).	(236)

Methylation marker panel

Sample size	**Method**	**Findings**	**Ref.**
EAC-35 undergoing chemoradiation therapy	Methylation specific PCR	Combined mean of promoter methylation of *p16, Reprimo, p57, p73, RUNX-3, CHFR, MGMT, TIMP-3,* and *HPP1* was lower in patients who responded to chemoradiotherapy (13/35) as compared to patients who didn't respond (22/35). (p=0.003).	(188)
BE-62 (28 BE patients progressed to EAC and remaining 34 BE patients were non-progressors)	Methylation specific PCR	3 tiered stratification model was developed using methylation index (*p16, HPP1,* and *RUNX3*), BE length and pathology. Combined model based on 2 year (AUROC: 0.8386) and 4 year (AUROC: 0.7910) prediction was able to categorize BE patients into low-risk, intermediate-risk and high-risk groups for EAC development.	(186)

BE-195 (145 BE patients progressed to EAC and remaining 50 BE patients were non-progressors)	Methylation specific PCR	*HPP1* (p = 0.0025), *p16* (p=0.0066) and *RUNX3* (p=0.0002) were significantly hypermethylated in progressors as compared to non-progressors. In combination, panel of 8 methylation markers *(p16, HPP1, RUNX3, CDH13, TAC1, NELL1, AKAP12* and *SST)* showed sensitivities of 0.443 and 0.629 at specificity of 0.9 and 0.8 for EAC progression in BE patients using combined model designed based on 2 and 4 years of follow-up.	(185)
EAC-41 (Adjacent normal samples as control)	Methylation specific PCR	Patients having >50% of their genes methylated *(APC, E-cadherin, MGMT, ER,p16, DAP-kinase and TIMP3)* showed significantly poor 2-year survival (p=0.04) and 2-year relapse-free survival (p=0.03) as compared to the patients having <50% methylation.	(187)
BE-18, EAC-38 (Multiple biopsies were taken and classified into normal, BE, HGD and EAC)	Bisulphite modified DNA with PCR	The methylation frequencies of nine genes *(APC, CDKN2A, ID4, MGMT, RBP1, RUNX3, SFRP1, TIMP3, and TMEFF2)* found to be 95%, 59%, 76%, 57%, 70%, 73%, 95%, 74% and 83% respectively in EAC samples while 95%, 28%, 78%, 48%, 58%, 48%, 93%, 88% and 75% respectively in BE samples which was significantly higher as compared to normal squamous epithelium. The methylation frequency for *CDKN2A* and *RUNX3* was significantly higher for EAC as compared to BE biopsy samples.	(237)
Normal-30, BE-29, HGD-8, EAC-29	Illumina GoldenGate methylation bead array	Overall median methylation at the total 706 numbers of most informative CpG sites gradually increased from normal-BE-HGD/EAC (p<0.001). The authors differentiated between EAC vs. normal, HGD vs. normal, BE vs. normal, EAC vs. BE and HGD vs. BE based on 422, 225, 195, 17 and 3 numbers of CpG sites which is showing differential methylation between respective groups.	(238)

Identification phase (BE-22, EAC-24); Retrospective validation phase (BE-60, LGD/HGD-36, EAC-90); Prospective validation phase (98 patients under surveillance).	Identification phase: Illumina Infinium assay; Retrospective/ Prospective validation phase: Pyrosequencing	Based upon initial identification phase, 7 genes *(SLC22A18, ATP2B4, PIGR, GJA12, RIN2, RGN, TCEAL7)* showing most prominent methylation changes were selected for validation. Combination of 4 genes (AUROC 0.988) *SLC22A18, PIGR, GJA12 and RIN2* showed sensitivity of 94% and specificity of 97%. This panel of 4 genes showing differential methylation, stratified patients into low, intermediate and high risk groups for EAC development in prospective validation.	(239)
Non dysplastic BE (Not progressed to EAC)-16, BE mucosa from patients progressed to EAC-12	Methylation-sensitive single-strand conformation analysis and methylation-sensitive dot blot assay	BE samples collected from patients who progressed to EAC in 12 months time period showed 100%, 91% and 92% hypermethylation of *APC, TIMP-3* and *TERT* respectively as compared 36%, 23%, and 17% in BE mucosa collected from patients who didn't progress to EAC.	(240)

Table 1.4. Summary of gene expression profiling studies for BE/EAC. [aCGH:array-comparative genomic hybridization]

Sample Size	**Array Description**	**Outcome**	**Findings**	**External validation**	**Ref.**
BE-21 (Paired normal esophageal and gastric samples as control)	Serial analysis of gene expression	Disease progression	534 tags were significantly differentially expressed between normal esophageal squamous epithelium and BE. The most up-regulated genes in BE as compared to normal epithelium were identified to be trefoil factors, annexin A10 and galectin-4 with each different type of tissue showed an unique cytokeratin expression.	No	(241)
BE and HGD -11	cDNA microarray	Disease	Using 2.5-fold cut-off, identified 131 up-regulated and 16 down-	Real-time	(242)

Sample Size	Array Description	Outcome	Findings	External validation	Ref.
(Matched biopsy samples)		progression	regulated genes in HGD. 24 out of 28 most significantly different genes showed similar changes during validation.	PCR	
EAC-91	Oligo-microarray	Disease progression	A 4-gene panel consists of deoxycytidine kinase, 3'-phosphoadenosine 5'-phosphosulfate synthase 2, sirtuin-2 and tripartite motif-containing 44 predicted 5-year survival.	Immunohistochemistry	(243)
23 paired BE and normal epithelium samples	Transcriptional profiling and proteomics	Disease progression	Identified 2822 genes to be differentially expressed between BE and normal epithelium. Significantly over-expressed genes during BE belonged to cytokines and growth factors, constituents of extracellular matrix, basement membrane and tight junctions, proteins involved in prostaglandin and phosphoinositol metabolism, nitric oxide production and bioenergetics. While genes encoding heat shock protein and various kinases were down-regulated.	No	(244)
Lymph node metastatic (n=55) and non-metastatic (n=22) EAC samples	Oligo-microarray	Disease progression	Lymph node positive samples showed significant down-regulation of Argininosuccinate synthetase as compared to lymph node non-metastatic samples (p=0.048).	No	(245)
EAC-6 and gastric cardia cancer-8	aCGH	Disease progression	Identified *HGF*(45%) and *BCAS1* (27%) to be most frequently over-expressed genes respectively at 7q21 and 20q13 locus.	No	(246)
11 matched sample sets (healthy-BE-EAC	SNP microarray	Disease	60% of BE and 57% of EAC samples contained at least one of the genomic alterations in the form of deletions, duplications,	No	(247)

Sample Size	Array Description	Outcome	Findings	External validation	Ref.
matched-6, normal-BE matched-4 and normal-EAC matched-1)		progression	amplifications, copy-number changes and neutral loss of heterozygosity.		
Normal-39, BE-25,EAC-38 and ESCC-26	cDNA microarray	Disease progression	Clustering showed the separation of samples into 4 distinct groups. 2158 clones were differentially expressed between normal and BE samples while 1306 between BE and EAC. BE/EAC samples showed differential expression of hydrolases, lysozyme, fucosidase, transcription factors, mucins and the trefoil factors.	No	(248)
BE-20, LGD-19, HGD-20 and EAC-42	SNP microarray	Disease progression	Increasing numbers of SNPs and loss of chromosomes with disease progression. Chromosomal disruption was identified in the *FHIT*, *WWOX*, *RUNX1*, *KIF26B*, *MGC48628*, *PDE4D*, *C20orf133*, *GMDS*, *DMD*, and *PARK2* genes in EAC.	No	(249)
EAC-75 specimens from 64 patients, Adjacent paired Normal tissue from EAC patients-28	DNA microarray	Disease progression	Identified *AKR1B10*, *CD93*,*CSPG2*, *DKK3*, *LUM*, *MMP1*, *SOX21*, *SPP1*, *SPARC* and *TWIST1*genes as biomarker based on transcriptomics data. Quantitative real-time PCR identified *SPARC* and *SPP1* genes to be associated with EAC patient survival ($p < 0.024$).	Real-time PCR	(250)
EAC-8, Gastric cardia cancer-3	aCGH and cDNA microarray	Disease progression	Transcriptomics data identified 11 genes to be differentially expressed (*ELF3*, *SLC45A3*, *CLDN12*, *CDK6*, *SMURF1*, *ARPC1B*, *ZKSCAN1*, *MCM7*, *COPS6*, *FDFT1* and *CTSB*). IHC	No	(251)

Sample Size	Array Description	Outcome	Findings	External validation	Ref.
			analysis revealed significant over-expression of *CDK6* a cell-cycle regulator in tumor samples.		
BE-20	aCGH arrays and high density SNP genotyping	Disease progression	Copy number losses were detected at FRA3B (81%), FRA9A/C (71.4%), FRA5E (52.4%) and FRA 4D (52.4%) sites in early BE. Validation study confirmed loss of FRA3B and FRA16D in early BE samples.	Real-time PCR and Pyrosequencing	(252)
BE-11, Gastro-esophageal junction (GEJ) adenocarcinoma-11	aCGH with a whole chromosome 8q contig array	Disease progression	Over-expression of *MYC* and *EXT1* while down-regulation of *MTSS1*, *FAM84B* and *C8orf17* is significantly associated with GEJ adenocarcinoma.		(253)
BE-14, EAC-5, ESCC-3	cDNA microarray	Disease progression	Identified 160 genes that can differentiate between BE and esophageal cancer.	No	(254)
24 paired samples of normal, BE and EAC phenotype	cDNA microarray	Disease progression	214 differentially regulated genes could differentiate between normal, BE and EAC phenotype. Genes involved in epidermal differentiation are under-expressed in EAC as compared to BE. Expression ratio of *GATA6* to *SPRR3* can differentiate between three phenotypes studied.	No	(255)
Pooled biospy samples from BE, esophageal squamous, gastric and	Oligo-microarray	Disease progression	Differentiate different tissue clusters based on gene expression profile. Identified 38 genes that are up-regulated in BE tissue cluster which belong to cell cycle (P1cdc47, PCM-1), cell migration (urokinase-type plasminogen receptor, LUCA-	No	(256)

Sample Size	Array Description	Outcome	Findings	External validation	Ref.
duodenum			1/HYAL1), growth regulation (TGF-β superfamily protein, amphiregulin, Cyr61), stress responses (calcyclin, ATF3, TR3 orphan receptor), epithelial cell surface antigens (epsilon-BP, ESA, integrin β4, mesothelin CAK-1 antigen precursor) and four mucins.		
Normal-24, BE-18, EAC-9	cDNA microarray	Disease progression	Identified 457, 295 and 36 differentially expressed genes respectively between normal-EAC, normal-BE and BE-EAC groups.	No	(257)
89-EAC	cDNA-mediated annealing, selection, extension, and ligation assay with 502 known cancer related genes	Disease progression	Identified differential gene expression between early stages of EAC (T1 and T2) vs. late (T3 and T4). Gene expression profile revealed *ERBB4*, *ETV1*, *TNFSF6*, *MPL* genes to be common between advanced tumor stage and lymph node metastasis.	No	(258)
Normal esophageal mucosa-9, esophagitis-6, BE-10, EAC-5, GEJ adenocarcinoma-9, stomach samples-32 (normal mucosa-11, IM-9, intestinal-type adenocarcinoma-7, and	cDNA microarray	Disease progression	Based on the expression profile, genes associated with the lipid metabolism and cytokine nodule are found to be significantly altered between EAC and other groups.	No	(259)

Sample Size	Array Description	Outcome	Findings	External validation	Ref.
diffuse carcinoma-5)					
17 paired samples of normal, BE/EAC	cDNA microarray	Disease progression	Each tissue type expresses distinct set of genes which can differentiate between their phenotypes. BE and EAC expresses similar set of stromal genes that are different from normal epithelium.	No	(260)
BE-19, EAC-20 (98 tissue specimens were collected and categorized into different groups)	Based on previous microarray studies 23 genes were validated using Real-time PCR	Disease progression	Out of 23 genes, panel of 3 genes (*BFT*, *TSPAN*, *TP*) was able to discriminate between BE and EAC in internal validation with 0% classification error.	N.A.	(261)
Normal-30, BE-31, Gastric mucosa-34, Duodenum-18	Biomarkers for BE were identified using three publically available microarray datasets and validated using Real-time PCR and Immunohistochemistry.	Disease progression	Out of 14 genes identified, dopa decarboxylase (*DDC*) and Trefoil factor 3 (*TFF3*) were validated to be up-regulated in BE.	N.A.	(262)
EAC-56	Oligonucleotide microarray and	Disease progression	Identified four new genes (*EGFR*, *WT1*, *NEIL2* and *MTMR9*) to be over-expressed in 10-25% EAC. Expression levels of these	Immunohistochemistr	(263)

Sample Size	Array Description	Outcome	Findings	External validation	Ref.
	aCGH		four genes differentiated EAC patients into three groups namely good, average and poor depending upon their prognosis ($p < 0.008$)	y	
BE/LGD-72, HGD-11, EAC-15	Bacterial Artificial Chromosome array comparative genomic hybridization	Disease progression	Copy number changes were more common and larger as disease progress to later stages. Patients having copy number alterations involving >70 Mbp were at increased risk of progression to EAC ($p=0.0047$)	No	(197)
EAC-30, BE-6, LGD-9, HGD-10	Genome-wide CGH	Disease progression	Loss of 7q33-q35 was found in HGD as compared to precursor LGD ($p=0.01$). Loss of 16q21-q22 and gain of 20q11.2-q13.1 was significantly different between HGD and EAC ($p=0.02$ and $p=0.03$ respectively).	No	(193)
EAC-30, Lymph node metastasis-8, HGD-11, LGD-8 and BE-6 from 30 EAC patient biopsy samples	CGH	Disease progression	Identified regions undergoing copy number loss and amplification during each stage of transition. Average number of chromosomal imbalance sequentially increased from BE-LGD-HGD-EAC-lymph node metastasis.	No	(191)
42 patients represent different stages of disease	SNP array	Disease progression	SNP abnormalities increases from 2% to more than 30% as the disease progress from BE to EAC. Total number of SNP alterations in tissue samples is tightly correlated with DNA abnormalities such as aneuploidy and LOH.	No	(194)

Sample Size	Array Description	Outcome	Findings	External validation	Ref.
EAC-27 and matched normal-14	SNP array	Disease progression	Confirmed previously described genomic alterations such as amplification on 8q and 20q13 or deletion/LOH on 3p and 9p. Also identified alterations in several novel genes and DNA regions in EAC samples.	No	(195)
EAC-26	SNP array	Disease progression	Confirmed previously reported frequent changes to *FHIT, CDKN2A, TP53* and *MYC* genes in EAC. Identified *PDE4D* and *MGC48628* as tumor-suppressor genes.	No	(196)
EAC-35	cDNA microarray	Response to chemotherapy	Identified 165 differentially expressed genes between poor (n=17) and good outcome (n=18) patient groups. Top functional pathway based on differential gene expression was identified to be TOLL-receptor signaling.	No	(264)
EAC-47 (locally advanced tumor)	cDNA microarray	Response to chemotherapy	Identified 86 genes showing at least 2-fold difference between chemotherapy responders (n=28) and non-responders (n=19). Ephrin B3 receptor, which showed highest difference between the groups, showed strong membrane staining in chemotherapy responding tumors using immunohistochemistry.	No	(265)
EAC-19 patients undergoing chemoradiotherapy	Oligo-microarray	Response to chemoradiotherapy	Reduced expression of *IVL, CRNN, NICE-1*, *S100A2*, and *SPPR3* gense correlated with poor survival and non-response to chemotherapy.	No	(266)
19 patients (EAC-16,	Oligo-microarray	Response to	Lower expression for panel of genes *PERP*, *S100A2*, and *SPRR3*	No	(267)

Sample Size	Array Description	Outcome	Findings	External validation	Ref.
ESCC-2 and adenosquamous carcinoma-1) undergoing chemoradiotherapy		chemoradiotherapy	was associated with non-response to therapy. Pathway analysis identified down-regulation of apoptosis in non-responders.		
EAC-174, ESCC-36	SNPs associated with the chemotherapy drug action pathway	Response to chemoradiotherapy	Identified association between genetic polymorphisms and response to pre-operative chemotherapy (fluorouracil and platinum compounds) and radiotherapy.	No	(268)

1.4.2 Cancer related inflammation

Gastric and bile acid exposure in the esophageal epithelium leads to the development of chronic inflammatory conditions mainly driven by elevated levels of pro-inflammatory cytokines. Chronic inflammatory responses induce cell survival and increase cell proliferation hence play key roles in the development of EAC (156, 269). Expression of various inflammatory molecules like Cyclooxygenase-2 (COX-2), NF-κB, IL-6, IL-8 and Matrix metalloproteinases (MMPs) have been evaluated as prognostic biomarkers for BE/EAC development.

Exposure to gastric/bile acid and cytokines leads to increased COX-2 expression (142). COX-2 is a rate-limiting enzyme that regulates synthesis of prostaglandins from its precursor arachidonic acid. COX-2 directly increases cell proliferation and promote tumor invasion (142), and COX-2 mediated increase in prostaglandins synthesis could result in tumor growth and angiogenesis (270). COX-2 expression has been detected in disease-free esophageal tissue homogenates using immunoblotting (142). In comparison to GERD, patients suffering from erosive reflux show slightly higher gene expressions of this enzyme in tissue samples (271). Several studies have shown significantly increased COX-2 expression correlating with the disease progression from BE to dysplasia and EAC (140-144). Furthermore, expression levels of COX-2 have been demonstrated to have a prognostic value in EAC with higher levels associated with poor survival and increased chances of tumor relapse (145, 146).

Another well studied inflammatory biomarker NF-κB is activated in response to exposure with bile acid and elevated NF-κB expression levels are found in BE, dysplasia, and adenocarcinoma (147-149). Activated NF-κB translocates from cytoplasm to nucleus and up-regulates transcription of the genes involved in inflammatory processes. Moreover, nuclear NF-κB expression has been shown to be correlated with the patient response to chemoradiation therapy. All of the patients who showed complete response to chemoradiation therapy had elevated NF-κB levels pre-treatment and showed lack of active NF-κB post-treatment (150).

In line with NF-κB and COX-2, expression of individual or combinations of pro-inflammatory cytokines IL-1β, IL-6, IL-8 and TNF-α is significantly increased in BE and EAC as compared to squamous epithelium (151, 152, 154). IL-1β and IL-8 expression levels also correlate with the stage of EAC (148). Patients who responded to neoadjuvant chemotherapy treatment showed significantly reduced expressions of IL-8 and IL-1β in post-chemotherapy esophageal tissue sections (150). IL-6 is activated in response to reflux and the IL-6/STAT3 anti-apoptotic pathway may underlie the development of dysplasia and tumor (153). Serum IL-6 levels was reported to

provide 87% sensitivity and 92% specificity for EAC diagnosis in a retrospective study (155). However, the study only compared between healthy and EAC groups. It would be interesting to see how early it can diagnose EAC during the process of metaplasia-dysplasia. Combination of cytokines IFN-γ, IL-1α, IL-8, IL-21, IL-23 along with platelet proteoglycan and miRNA-375 expression profiling has been demonstrated to build an inflammatory risk model which has clinical utility to determine prognosis for EAC patients (156).

MMPs are a family of proteolytic enzymes involved in the degradation of extracellular matrix components. MMPs play a role in both inflammation and tumor metastasis. Immunohistochemical staining for MMP-1, MMP-2, MMP-7 and MMP-9 has been reported to be significantly higher in EAC as compared to healthy individuals (157, 158). Higher level of MMP-1 expression has been associated with the lymph node metastases and possibly poor patient survival (159). Expression of MMP-9 is shown to be an early event during the EAC transformation and its expression levels are correlated with the progression of the disease (160-162). Activity of MMPs is inhibited by a family of proteins called tissue inhibitors of metalloproteinases (TIMPs). Specifically, TIMP-3 gene is methylated in EAC development and its reduced expression is associated with stage of the tumor and patient survival (163). On contrary, Salmela *et al*. described elevated TIMP-1 and TIMP-3 expression in EAC tumor samples (158).

Although the underlying tissue inflammation is very closely associated with EAC development and several inflammatory related biomarkers have been identified, these remain to be validated in large scale biomarker studies.

1.4.3 Cell cycle-related abnormalities

To compensate for the tissue damage induced by gastric/bile acid, the underlying epithelium starts to proliferate rapidly and become uncontrolled resulting in neoplasia. In order to meet the proliferation requirements, the cells have to overcome cell cycle check points. CyclinD1 over-expression is one such means by which cells overcome G1/S-checkpoint, and cyclinD1 immunohistochemical staining has been proposed to identify BE patients with an increased risk for EAC (164). In contrast to cyclinD1, expression of p16 protein results in cell cycle arrest in G1 phase as it has been shown to inhibit cyclin dependent kinase-induced phosphorylation of retinoblastoma protein. Early genomic abnormalities during EAC development significantly affects p16 protein expression which can be determined using immunostaining and implemented as a potential biomarker (165). Further large scale trials are required to confirm cell cycle abnormalities during EAC development to implement them as a biomarker.

Bottom of the pyramid in Figure 1.3 represents list of biomarkers in initial stages of development. Tumors harboring over-expression of growth factor receptors (EGFR and HER-2) are associated with poor patient survival (272, 273), while those over-expressing apoptosis regulator Bcl-2 protein showed prolonged survival (274). Incipient angiogenesis is a marked feature of BE and underlining tissue expresses angiogenesis markers VEGF and its receptors (275). Neovascularization continues as the disease progress from BE to EAC. Measuring degree of neovascularization correlated with histopathological grade of the tumor and associated with the patient survival (276). Expressions of two prominent cell proliferation markers, PCNA and Ki-67 have been described to be altered during BE-EAC development (166).

1.4.4 microRNA

MicroRNA (or miRNA) was first discovered in *Caenohabditis elegans* (277) and from then it has been widely studied in range of biological phenomena. These short stretches of approximately 21 nucleotides do not code for protein but play important roles in gene regulation by either suppressing protein synthesis or cause mRNA cleavage. Unlike siRNA, miRNA can target multiple genes on remote loci hence control diverse group of proteins. Several key properties of carcinogenesis have been demonstrated to be regulated via miRNA, for example, angiogenesis and metastasis (278).

With increased biological understanding of miRNAs and their role in cancer, they have been proposed in several different clinical applications including cancer diagnosis and tumor prognosis, tumor classification and also as a therapeutic target for disease intervention. Differential tissue miRNA expression has been observed in several different malignancies and these changes can be utilized for diagnosis and classification of the tumors (278). MicroRNA arrays were first used to show differential miRNA expression in healthy, BE and EAC tissue samples (279). Since then, a number of different studies have identified miRNA changes associated with the development of the BE/EAC. Table 1.5 summarizes primary findings of miRNA expression profiling studies along with statistical significance and fold change values. Biological significance for some of the miRNA related changes is discussed below.

Smith and colleagues identified reduced expression of miR-200 and miR-141 in BE and EAC tissue samples. They performed bioinformatics analysis and correlated these miRNA expression changes with cellular processes such as cell cycle, cell proliferation, apoptosis and cell migration (280). MiR-196a, which is described as a marker of progression from BE to EAC, can increase cell proliferation and anchorage-independent growth and inhibit apoptosis in EAC cell

lines *in vitro* (281). The down-stream target for miR-196a are verified to be Annexin A1, S100 calcium-binding protein A9, Small proline-rich protein 2C and Keratin 5 which showed reduced expression in EAC patient tissue samples as compared to normal epithelium (281, 282). Several studies described in Table 1.5 report over-expression of miR-192 during EAC carcinogenesis. MiR-192 has been reported to be a target of p53 and has been able to suppress cancer progress in osteosarcoma and colon cancer cell lines through p21 accumulation and cell cycle arrest (283). As shown in Table 1.5 miR-21 is over-expressed during BE/EAC and it can function as an oncogene as shown in tumors of breast, brain, lung, prostate, pancreas, colon, liver and chronic lymphocytic leukemia. It negatively regulates tumor and metastasis suppressor genes PTEN, TPM1, PDCD4 and Sprouty2 (284-287). MiR-194 expression is regulated by hepatocyte nuclear factor (HNF)-1α transcription factor which is induced during BE/EAC and may lead to up-regulation of miR-194 (284). Higher expression of miR-194 is also observed in metastatic pancreatic cell lines (288). Amongst miRNAs found to be down-regulated during EAC development, let-7 family of miRNAs are tumor-suppressor and negatively regulates Ras oncogene. Fassan and colleagues confirmed up-regulation of HMGA2 which is one of the targets of let-7 miRNA using immunohistochemistry in tissue samples (285, 287, 289).

A study published by Wu and colleagues compared miRNA expression profile between healthy (adjacent normal), BE and EAC patient tissue samples (Adjacent normal-35, BE-11, LGD-13, HGD-10, EAC-36) using real-time PCR-based TaqMan Human Micro-RNA array that enabled accurate quantitation of 754 human miRNAs (290). Unsupervised hierarchical clustering according to miRNA expression profiling suggested that there is clear distinction between healthy and BE/EAC phenotype in terms of miRNA expression, whereas BE and EAC samples were indistinguishable from each others as they showed clear overlap. In comparison with healthy samples, 148 and 122 miRNAs were found to be up-regulated in BE and EAC tissue respectively whereas 16 miRNAs were down-regulated in BE and EAC sample as compared to normal epithelium. Amongst handful of miRNAs that were significantly different between BE and EAC include miR-375 (down-regulated in EAC) and miR-106-3b, miR-18, miR-18-3p, miR-20b, and miR-92a-1-3p (up-regulated in EAC). Furthermore, they compared miRNA profiling results with mRNA expression in normal, BE and EAC tissue samples previously published by Nancarrow and colleagues (291). Interestingly, they found that 19 of the top 20 differentially expressed miRNAs targeted one or more of 54 most frequently altered mRNAs from the list published by Nancarrow and colleagues. Moreover, 77.8% (42 of 54) of the differentially expressed mRNAs were potential

targets for the top 20 differential miRNAs which includes miR-203 with 12 potential target mRNAs and let-7c with 11 target mRNAs (290).

Another study by Streppel and colleagues attempted to validate previous published 7 different miRNAs targets in BE/EAC (292). Out of 7 target miRNAs studies, 5 miRNAs (miR-199a/b-3p, -199a-5p, -199b-5p, -200b, and -223) were found to be significantly different between EAC and adjacent normal tissue. Out of 5, only miR-223 showed a stepwise increase during EAC development (292). Functional characterization of miR-223 using miR-223 over-expressing OE33 and JHesoAD1 showed more aggressive phenotype and also showed higher sensitivity for DNA-damaging agents (292).

Further studies in the regards of miRNA and miRNA target genes will improve the biological understanding of EAC pathogenesis and may also provide novel molecular targets for disease intervention. Notably, miRNAs are found to be stable in serum encapsulated in microvesicles hence can be accessed easily (293). In fact circulating miRNA profiling has shown distinct expression patterns in a number of cancers, other than EAC (294). This opens up new avenues for circulating miRNA changes as a potential biomarker for EAC.

Table 1.5. Summary of literature describing miRNA expression changes in BE/EAC. (Wherever needed, fold change values are calculated/adapted from the expression/fold change values described in the original article to have uniform format for the purpose of this literature review).

Sample Size	**Up-regulated in BE/EAC**	**Down-regulated in BE/EAC**	**Ref.**
71 (BE-12, BE without dysplasia-20, LGD-27, EAC/HGD-12)	miR-192 ($p<0.00001$), miR-196a ($p<0.05$): Up-regulated in BE as compared to healthy tissue. miR-196a expression is correlated with progression from IM-LGD-HGD-EAC ($p<0.005$)	miR203 ($p<0.00001$): Down-regulation in BE as compared to healthy tissue	(295)
22 (BE without dysplasia-11, BE with dysplasia-11)	miR-15b (3.3 fold, $p<0.05$), miR-203 (5.7 fold, $p<0.05$): Up-regulated in dysplasia as compared to non-dysplastic BE	miR-486-5p (4.8 fold, $p<0.05$),miR-let-7a (3.3 fold, $p<0.05$): Down-regulated in dysplasia as compared to non-dysplastic BE	(285)
100 (EAC-100, Adjacent normal tissue as control)	miR-21 (~3 fold, $p<0.05$), miR-223 (~2 fold, $p<0.05$), miR-192 (~3.5 fold, $p<0.05$), and miR-194 (~3.5 fold, $p<0.05$): Up-regulated in EAC	miR-203 (~3 fold, $p<0.05$): Down-regulated in EAC as compared to adjacent normal tissue	(286)

Sample Size	Up-regulated in BE/EAC	Down-regulated in BE/EAC	Ref.
	as compared to adjacent normal tissue		
25 (Healthy-9,BE-5, HGD-1, EAC-10)	miR-192 (1.7 fold, FDR<1e-07), miR-194 (2 fold, FDR<1e-07), miR-21 (3.7 fold, FDR=0.0003), miR-200c (1.9 fold, FDR=0.0015), miR-93 (1.3 fold, FDR=0.0108): Up-regulated in EAC as compared to BE	miR-27b (1.43 fold, FDR=0.0003), miR-342 (1.25 fold, FDR=0.0015), miR-125b (2 fold, FDR=0.0108), miR-100 (1.25 fold, FDR=0.011): Down-regulated in EAC as compared to BE	(279)
75 (Healthy-15, BE-15, LGD-15, HGD-15, EAC-15)	miR-215 (62.8 fold, p<1e-07), miR-192 (6.34 fold, p<1e-07): Up-regulated in BE in comparison with normal tissue and remained at similar levels with disease progress	miR-205 (10 fold, p=1.39e-0.5), let-7c (2.04 fold, p=3.11e-05), miR-203 (6.67 fold, p=3.2e-0.5): Down-regulated in BE in comparison with normal tissue and remained at similar levels as disease progresses	(289)
91 (LGD-31, HGD-29, EAC-31, In all cases adjacent normal tissue used as a control)	miR-200a (13.5 fold, p=0.02), miR-513 (1.58 fold, p=0.03), miR-125b (9.2 fold, p=0.04), miR-101 (1.83 fold, p=0.04), miR-197 (1.61 fold, p=0.04): Up-regulated in LGD to HGD transition	miR-23b (1.45 fold, p=0.007), miR-20b (1.56 fold, p=0.01), miR-181b (2.22 fold, p=0.03), miR-203 (1.49 fold, p=0.03), miR-193b (2.70 fold, p=0.04), miR-636 (4.17 fold, p=0.04): Down-regulated in LGD to HGD transition. let-7a (1.75 fold, p=0.01), let-7b (1.59 fold, p=0.009), let-7c (1.69 fold, p=0.03), let-7f (1.69 fold, p=0.03), miR-345 (2 fold, p=0.02), miR-494 (1.72 fold, p=0.03), miR-193a (2.27 fold, p=0.05): Down-regulated in HGD-EAC development process	(287)
48 (BE-19, EAC-29)	miR-21 (~2.8 fold, p<0.05), miR-143 (~11.3 fold, p<0.05), miR-145 (~3.4 fold, p<0.05), miR-194 (~126 fold, p<0.05), miR-215 (~18 fold, p<0.05): Up-regulated in BE as compared to adjacent normal tissue.	miR-203 (~17 fold, p<0.05), miR-205 (~175 fold, p<0.05): Down-regulated in BE as compared to adjacent normal tissue. miR-143 (~3 fold, p<0.05),	(284)

Sample Size	Up-regulated in BE/EAC	Down-regulated in BE/EAC	Ref.
		miR-145 (~1.8 fold, p<0.05), miR-215 (~3.1 fold, p<0.05): Lower expression in EAC as compared to BE	
49 (IM-15, HGD-14, and EAC-20, Adjacent normal tissue)	-	miR-31 (>4 fold, p<0.02), miR-375 (>4 fold, p<0.001): Down-regulated in transition from BE to EAC	(296)
37 (BE-17, EAC-20, 9 adjacent normal tissue samples)	-	miR-141 (~2 fold, p=0.0126), miR-200a (~2.5 fold, p=0.0001), miR-200b (~2.1 fold, p<0.0001), miR-200c (~1.9 fold, p=0.0014), miR-429 (~1.8 fold, p=0.0031): Under-expressed in EAC as compared to BE	(280)
11 (EAC-11, Different lesions were collected from these patients and classified into BE, LGD, HGD, EAC)	miR-196a is over-expressed in early EAC (151 fold)>HGD (62.2 fold, p=0.00002)>LGD (31.1 fold, p=0.0005)>BE (28.9 fold, p=0.00001). Fold changes are calculated as compared to normal epithelium.	-	(282)
45 (EAC patients undergoing surgery)	miR-143 (p=0.0148), miR-199a_3p (p=0.0009), miR-199a_5p (p=0.0129), miR-100 (p=0.0022) and miR-145 (p=0.1176) expression predicted a worse survival followed by esophagectomy. Overexpression of miR-199a_3p/ _5p and miR-99b was associated with lymphnode metastasis	Down-regulation of miR-143 (p=0.0049) and miR-145 (p=0.0069) in EAC as compared to adjacent normal tissue	(297)
24 (BE-24, Progression to EAC-7, Not Progressed to EAC-17 in at least 5 year follow-up)	miR-192 (AUROC=0.61), 194 (AUROC=0.70), 196a (AUROC=0.80), and 196b (AUROC=0.74) showed significantly higher expression in BE samples from patients who	-	(298)

Sample Size	Up-regulated in BE/EAC	Down-regulated in BE/EAC	Ref.
	progressed to EAC as compared with those who did not progress to EAC		
5 (EAC patients undergoing surgery. Adjacent benign tissue as a control)	MiR-296 is over-expressed ~2 fold in EAC as compared to adjacent benign tissue.	-	(299)
22 locally advanced EAC tumor patients undergoing surgery	Negative association between miR-148a expression and tumor differentiation ($p<0.001$). Significantly higher expression of miR-148a in tumors located in the lower esophagus as compared to tumors in the middle esophagus ($p=0.021$).	-	(300)
99 EAC patient tissue samples undergoing surgery	miR-30e ($p=0.002$) and miR-200a ($p=0.044$) expression were associated with poor overall survival. miR-16-2 ($p=0.027$) and miR-30e ($p=0.002$) expression were associated with poor disease-free survival.	-	(301)
60 (Healthy-10, BE-10, Gastric metaplasia-10, LGD-10, HGD-10, EAC-10)	-	miR-125a-5p ($p=0.008$) and miR-125b were progressively down-regulated in lesions from IM to LGD/HGD to EAC.	(302)
32 (EAC-32, Adjacent normal tissue as control)	miR-21 (FDR=0.000067)	miR-203 (FDR=0.000201)	(303)

1.4.5 Glycoproteins

Protein glycosylation is a common post-translational modification with almost half of the proteins synthesized undergoing one of the two major types either N-linked or O-linked glycan modifications. The biosynthetic process of glycosylation is not template driven and occurs co- or post-translationally. N-glycosylation initiates (synthesis of N-glycan precursors) on the cytoplasmic

side of the rough endoplasmic reticulum and targets the asparagine in the sequence of Asn-X-Ser/Thr, where X is any amino acid other than proline. The process continues when protein moves further along the secretory pathway and finally expresses one of the three common glycan types. 1. High Mannose, 2. Complex or 3. Hybrid type (Figure 1.4). In O-type glycosylation, the Ser/Thr residue of the protein is involved in making a glycosidic linkage with the glycan moiety (Figure 1.4) (304). The process of glycosylation is regulated by the expression and localization of glycosyltransferases/glycosidases, protein trafficking, and the availability of substrate glycans (304). The other less common type of glycosylation is known as C-mannosylation where Trp is covalently attached to mannose residues (Figure 1.4). At least 10 different monosaccharides can participate in the process to generate diverse range of glycan structures with numerous possibilities for branching and anomeric linkages which make them difficult to study in comparison with proteins or nucleic acids (Figure 1.4) (304). Despite being heterogenous, glycan modification at each glycosylation site is very specific and stable for a given cell type and physiological state (305).

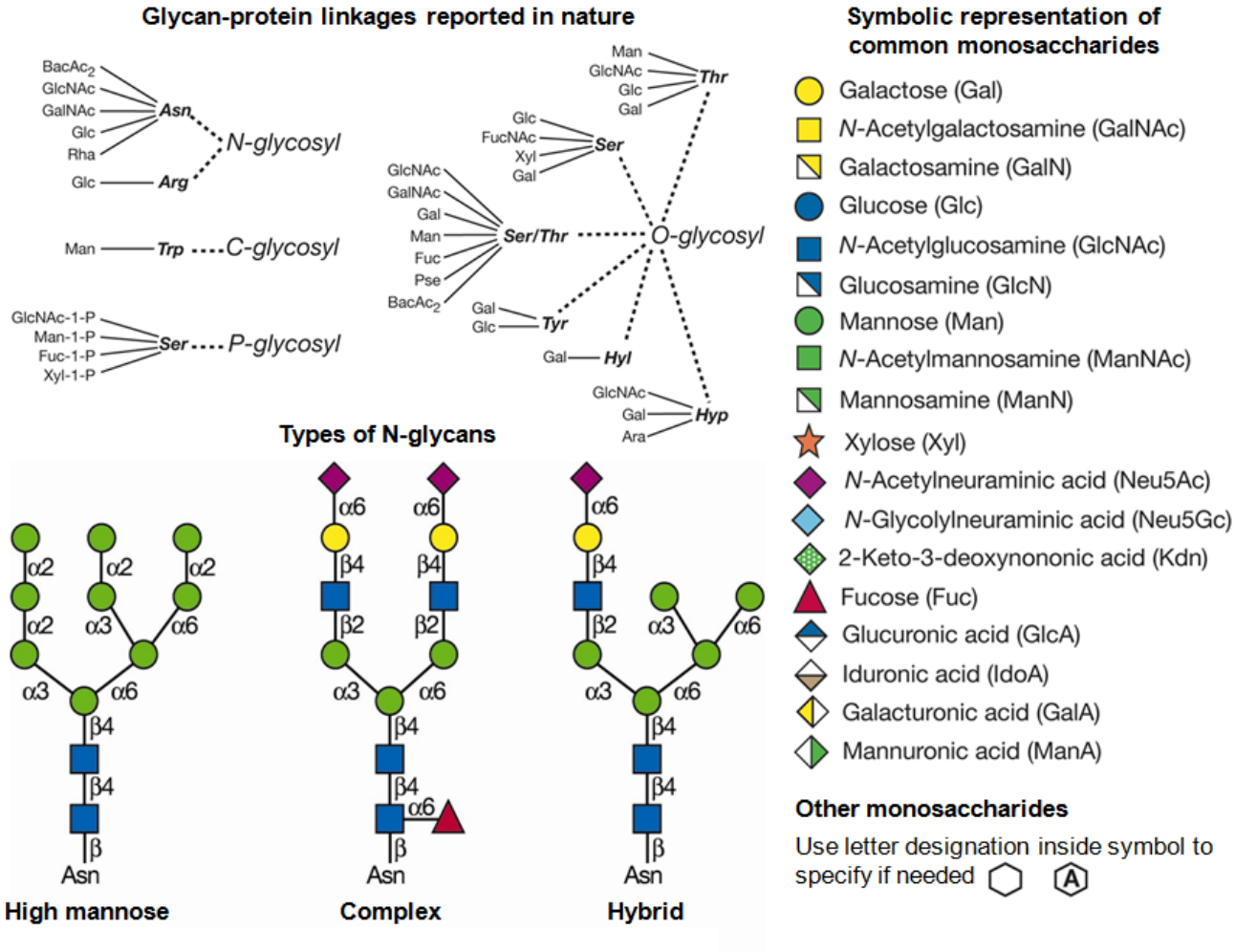


Figure 1.4. Glycan-protein linkages, types of N-glycans and symbolic representation of common monnosaccharides found in nature. Asn, asparagine; Arg, arginine; Trp, tryptophan;

Thr, threonine; Ser, serine; Tyr, tyrosine; Hyl, hydoxylysine; Hyp, hydroxyproline. Adapted from Varki *et al*. (306) and Stanley *et al*. (307).

Aberrant glycosylation changes have previously been reported for several different cancers including breast cancer, prostate cancer, melanoma, pancreatic cancer and ovarian cancer (308, 309). These changes include truncated forms of O-glycans, increased degree of branching in N-glycans, and elevated sialylation, sulfation, and fucosylation with range of other possible variations (309). The differential glycosylation can alter protein interactions, stability, trafficking, immunogenicity and function (308). Tumor specific glycosylation changes are actively involved in neoplastic progression, namely metastasis, as glycoproteins are found abundantly on cell surfaces and extracellular matrices hence play vital role in cellular interactions.

Lectins are a family of glycan-binding proteins used by nature for cell-cell and cell-protein communication in humans and in host-pathogen interactions. Moreover, they are extensively used in glycobiology due to preferential binding of each lectin to recognize specific glycan structures on proteins, lipids and other biomolecules (309, 310). The first effort to identify differential glycosylation in the progression to BE and EAC was made in 1987 by Shimamoto and colleagues using differential binding pattern to 5 lectins in tissue specimens (167). The glycoconjugate expression profile in BE was found to be significantly different from normal esophageal epithelium. Interestingly, glycoconjugate expression between BE and normal duodenum was quite similar. There were minimal glycoconjugate expression changes between BE and LGD. However, EAC tissue samples showed significantly different lectin binding pattern than BE/LGD (167). Using rabbit esophageal epithelium Poorkhalkali and colleagues showed differential lectin binding in response to acid/pepsin exposure suggesting acid exposure can induce cell surface glycosylation changes (311). In 2008 Neumann and colleagues used 4 different lectins to identify pathological mucosal changes (168). They observed two distinct lectin binding patterns. One which is associated with the GERD while the other pattern was characteristic for BE mucosa. Specifically UEA (*Ulex europaeus*) lectin binding was up-regulated in BE tissue sections which suggests possible increase in fucosylation during the disease progress (168). A recently published study has concluded that dysplasia can alter glycan expression hence lectin binding pattern to the tissue samples. Fluorescently labeled WGA (Wheat germ agglutinin) lectin binding intensity was found to be inversely related to the degree of dysplasia (169). Furthermore, the authors used fluorescent capable endoscope *ex vivo* in the study and followed all the protocols in a manner that exactly mimics a clinical study *in vivo.* Followed by topical fluorescein-labeled WGA spray, the authors measured fluorescence in the tissue samples. Measurement of lectin fluorescence was more sensitive approach

to identify dysplastic lesions as compared to white light endoscopic technique. Their data demonstrate clinical utility of such a lectin based endoscopic technique if developed further (169). In a phase III biomarker clinical trial study Bird–Liberman and colleagues combined 3 different abnormalities to predict EAC progression in BE patients. Along with using conventional LGD and DNA content abnormalities they used AOL (*Aspergillus oryzae)* lectin binding to the tissue samples which detects presence of α1-6 fucose on the cell surface (174). Thus, monitoring tissue glycan changes can be combined with existing biomarkers in order to improve the predictive power of the currently used biomarkers.

A potential mechanism responsible for these changes is considered to be bile acid exposure-induced gene expression and secretory pathway changes in esophageal epithelium (312). Using carbohydrate specific lectins that detect N and O-linked glycosylation and core fucosylation, Byrne and colleagues have shown differential lectin binding to the cell surface and differential intracellular localization when normal squamous and barrett's metaplastic cell lines were treated with deoxycholic acid (312). Nancarrow and colleagues profiled mRNA expression in normal squamous esophageal epithelium, BE, and EAC and concluded that BE is a tissue with enhanced glycoprotein synthesis machinery in order to provide strong mucosal defense against acid exposure (291).

1.4.6 Outlook - circulating biomarkers

Last four decades showed continuously increased EAC incidences and similar trend may be expected in future because of rising incidences of obesity and GERD in the population. Current endoscopic screening program might benefit the highest risk population to monitor disease progression. Monitoring dysplasia in the tissue samples has not provided fruitful outcome for early diagnosis, however inclusion of the genomic and cell cycle biomarkers has shown definite improvement in the predictive power over currently used histological technique. Any biomarker requiring tissue samples is going to be difficult to implement for population screening and will not be viable economically. An alternative to tissue-based techniques is to investigate changes in circulating biomarkers. Blood is relatively easy to access and, hence, can be monitored frequently ultimately increasing possibility of detecting early dysplastic changes.

Circulating tumor cells could be one source of biomarkers. Although readily found in the blood, technological advancements are required for sensitive early detection of the low number of tumor cells present in the circulation (313, 314). Alternatives to the detection of circulating tumor cells, Zhai and colleagues applied genome-wide DNA methylation profiling approach to cell free

circulating DNA. They found that cell free circulating DNA methylation profile is a replica of methylation profile found in matched tumor tissue samples and can discriminate between control, BE and EAC conditions (171). Kawakami and colleagues studied methylation of *APC* gene in matched tumor samples and plasma (172). Unlike tumor samples which showed hypermethylation of *APC* DNA early during the EAC development, matched plasma samples from patients suffering from BE and gastritis were found to be negative for *APC* methylation changes. Moreover, as compared to 92% (48/52) of EAC tissue samples, only 25% (13/52) plasma samples were positive for circulatory *APC* methylation changes. However, there was strong correlation between stage of the tumor and plasma positivity for methylated *APC* (172). In combination with *DAPK* methylation, measurement of pre-operative *APC* methylation in peripheral blood was able to discriminate between long (>2.5 years) and short survivors with a sensitivity of 99.9% and specificity of 57.1% (173). Taken together, tracking circulatory DNA methylation changes during EAC development may be an alternative approach to predict early EAC.

Tumor cell moulds the microenvironment to support oncogenesis by releasing soluble and vesicular components, including enzymes, microvesicles, proteoglycans, chemokines and cytokines (315). The tumor microenvironment components are shed into the circulation and may be extremely useful as an early diagnostic biomarker. This concept was demonstrated by Pitteri and colleagues using an inducible HER2/neu mouse model (316). They showed that plasma proteome profiling has ability to detect the cancer before it actually develops. Furthermore, a linear correlation was demonstrated for plasma levels of candidate biomarker proteins with the tumor progression, which were reversed upon tumor regression (316).

Both encapsulated miRNAs and secreted glycoproteins are prime candidates for circulating biomarkers released by the tumor microenvironment. Circulating miRNAs are secreted in nanometer-sized vesicles called exosomes or microvesicles. An advantage of circulating miRNA over protein biomarkers is the ability for amplification, increasing the sensitivity of detection. Comparative analysis of circulating miRNA can be performed using miRNA microarray and quantitative real-time PCR (294). Future studies should aim to discover and validate circulating miRNA changes associated with EAC development and progression.

1.4.6.1 Glycan profiling

For BE and EAC, serum glycan profiling using mass spectrometry has identified differential expression of glycan structures in different disease states. Mechref and colleagues analysed N-linked glycan diversity present in 84 patient serum samples (Healthy-18, BE-5, HGD-11, EAC-50)

(175). They identified 98 glycan features with different intensities in disease onsets and 26 of them correspond to known glycan structures. They demonstrated statistically significant glycan changes between 4 different conditions (Healthy/BE/HGD/EAC) with three of the known potential N-glycan biomarker predicted EAC with 94% sensitivity and 60% specificity (175). Another similar study used microchip electrophoresis with laser-induced fluorescence detection for N-glycan profiling and able to differentiate between the healthy, BE, HGD, and EAC conditions (176). Similar to above mentioned N-glycan profiling studies, very recently, Gaye and colleagues showed that ion mobility-mass spectrometric analysis of serum N-glycan can also distinguish between normal and EAC phenotype (177). All of these studies unanimously suggest circulatory N-linked glycan changes during EAC pathogenesis. Mann and colleagues enriched fucosylated serum glycoproteins using lectins and then used shot gun proteomics to identify protein in different physiological states, including healthy, BE, and EAC (178). Although the study showed promising trends, the statistical power was not achieved due to the very low number of samples. To improve the throughput of glyco-centric proteomics studies, we have developed lectin magnetic bead array-mass spectrometry (LeMBA-MS), a high-throughput platform where a panel of lectins individually immobilized the magnetic beads is used to capture glycoproteins followed by on-bead trypsin digest and liquid-chromatography-tandem mass spectrometry for protein identification (317, 318). Parallel screening of a panel of lectins may be helpful to identify differentially glycosylated circulating proteins during EAC pathogenesis.

1.4.6.2 Metabolic profiling

In recent past, efforts have been made to profile metabolic changes associated with EAC pathogenesis. Metabolic profiling studies have identified changes associated with nucleoside metabolism, tri-carboxylic acid cycle, fatty acid and amino acid metabolism during EAC development in tissue samples and more importantly using easily accessible bio-fluids, blood and urine. Early metabolic changes in the histologically normal epithelium were observed, particularly for phosphocholine, glutamate, myo-inositol, adenosine-containing compounds, uridine-containing compounds and inosine (319). Djukovic and colleagues used targeted approach to profile eight different serum nucleosides between healthy subjects (n = 12) and EAC patients (n = 14) using high-performance liquid chromatography (HPLC) coupled with triple quadrupole mass spectrometer. Amongst eight nucleosides they profiled, five were significantly different between the two groups. 3 out of 5 significantly different nucleosides namely 1-methyladenosine, N^2,N^2-dimethylguanosine, N^2-methylguanosine were methylated nucleosides indicating increased tRNA methylation, similar to DNA hypermethylation in EAC condition (181). Zhang and colleagues

studied serum metabolomic changes using NMR alone and NMR in combination with LC-MS in EAC (n = 67), HGD (n = 9), BE (n = 3) and healthy volunteers (n = 34). Their model based on PLS-DA was being able to distinguish between different phenotypes by achieving area under receiver operating characteristics curve (AUROC) of as high as 0.95. Based on candidate metabolites they identified altered pathways associated with EAC development to be energy metabolism, fatty acid metabolism and amino acid metabolism (182, 183). Urine metabolomics could also distinguish between healthy, BE and EAC phenotypes. Davis and colleagues generated urine metabolic signatures which was able to discriminate between healthy, BE and EAC phenotypes, as well as distinguish EAC from pancreatic cancer (184). These metabolic profiling studies open up new avenue to detect early EAC using circulatory biomarkers.

1.5 Glycoproteins as cancer biomarkers

The majority of United States Food and Drug Administration (FDA) approved cancer protein biomarkers are glycoproteins (130). Currently total glycoprotein levels are monitored in circulation to make the informed diagnostic decisions e.g. serum carbohydrate antigen 125 (CA125) level > 35 units/mL indicates gynecological abnormalities in women hence patients are referred for further investigation (320). Similarly, serum prostate-specific antigen levels > 4 ng/mL indicate prostate related abnormalities in male (321). Positive test results for total prostate-specific antigen measurements may be due to either prostate cancer or other non-cancerous abnormalities such as benign prostatic hyperplasia (BPH) or prostate infection/inflammation hence its implementation as a screening tool lead to overdiagnosis and overtreatment of prostate cancer (130). Recent studies show that monitoring aberrant glycosylation of prostate-specific antigen (fucosylation and sialylation) is a more specific, better predictor of prostate cancer and showed better correlation with Gleason score (322-326). In the case of hepatocellular carcinoma (HCC), increase in tumor size is closely linked to circulatory α-fetoprotein (AFP) levels. But AFP levels remain unchanged during early pathogenesis of HCC making total AFP measurement to be impractical as an early diagnostic marker. However, fucosylated AFP level changes in the circulation due to increased expression of α1-6 fucosyltransferase during early HCC pathogenesis (327, 328). Fucosylated AFP measurement test for early detection of primary HCC has been approved by the FDA and produced into a microfluidic based clinical immunoanalyzer test μTASWako®i30 (commercialized by Wako diagnostics) by measuring *Lens culinaris* agglutinin (LCA) (Fucose specific) bound fraction of AFP (AFP-L3, Figure 1.5) (329). Increased ratio of AFP-L3/total AFP indicates increased risk for HCC development hence patients are monitored extensively using ultrasound for further confirmation. This example suggests monitoring specific glycosylation change can serve as an early diagnostic

marker even before imaging modalities (330-332). For several other glycoprotein biomarkers, studies have shown that monitoring specific glycosylation changes are better biomarker candidates as compared to measuring total glycoprotein levels e.g. fucosylated haptoglobin (pancreatic cancer), elevated levels of glycoforms containing sialyl Lewis X on haptoglobin, α1-acid glycoprotein, α1-antichymotrypsin and immunoglobulin G (Ig G) (ovarian cancer) and fucosylated Ig G (stomach cancer) (327). Taken together, aberrant glycosylation changes may turn out to be more specific and sensitive biomarkers as compared to monitoring total glycoprotein levels.

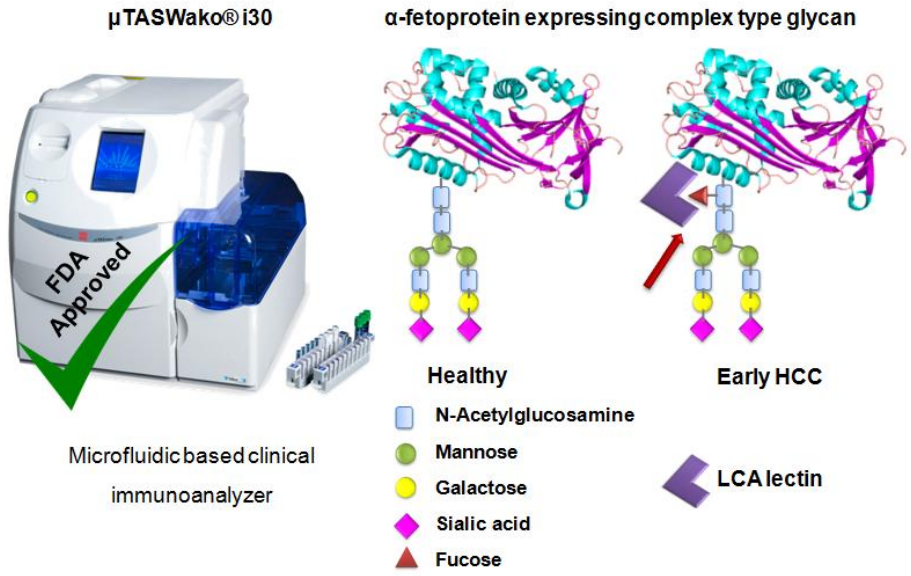


Figure 1.5. FDA approved μTASWako®i30 immunoanalyzer measures AFP-L3 to diagnose early HCC. LCA, *Lens culinaris* agglutinin; HCC, Hepatocellular carcinoma.

Three potential mechanisms may contribute to tumor specific glycan modifications. (i) Altered expression of glycan processing enzymes, (ii) tumor microenvironment and (iii) local or systemic activation of differential cell types which otherwise are not active. Over 400 proteins are estimated to be involved in carbohydrate binding and metabolism (333). Genomic abnormalities during tumor progression can alter expression of these proteins resulting in aberrant glycosylation (322). Several of these glycosyltransferase/glucosidase enzymes themselves are considered as potential biomarker. For example N-acetylglucosamine transferase V (GlcNAcT-V) is responsible for β1-6 branching of N-glycans and this particular glycan alteration in target proteins such as cadherin, integrin, and other cytokine receptors is responsible for tumor metastasis. Furthermore mice lacking GlcNAcT-V have shown reduced polyomavirus antigen induced tumor growth and metastasis which demonstrate GlcNAcT-V as a potential therapeutic target (334). Another enzyme N-acetylgalactosaminyltransferase is responsible for initial steps of mucin O-glycosylation. SNPs associated with this enzyme are inversely related to risk for developing ovarian cancer (335). Last

but not the least α-L-fucosidase activity in serum has shown to predict early development of hepatocellular carcinoma in liver cirrhosis patients (336). α-L-fucosidase activity has also been correlated with progression-free survival in patients undergoing trastuzumab monotherapy. This study extends use of enzyme activity of α-L-fucosidase in blood circulation as a predictive biomarker for treatment of breast cancer with monoclonal antibody therapy (337).

The other potential source for circulatory tumor specific glycoprotein is the tumor microenvironment. The stromal components co-evolve with the cancer cells and help tumor to metastasize, get nutrients, evade immune response and to attain other hallmark characteristics (338). The microenvironment consists of cancer associated fibroblast, extracellular matrix, secretome, inflammatory cells, glial cells, innate/adaptive immune cells, adipocytes, vasculature and specialized mesenchymal cells (339). The secretome component consists of proteins, receptors, proteoglycans, cytokines, chemokines, growth factors, angiogenesis factors, proteases etc (315). Aberrant angiogenesis leads to accumulation of tissue fluids in the tumor environment and this fluid extravasate from the tumor site and mixes with the circulation. This fluid is rich in proteins secreted through classical or non-classical secretory pathway of the tumor cells and may be extremely useful as a diagnostic purpose as these changes can be monitored in circulation and tumor specific. The microenvironment and their secretions vary with tumor progression and identification of stage specific biomarkers can be useful to monitor the disease progress and predict therapeutic outcome. In fact using inducible HER2/neu mouse model Pitteri and colleagues showed that plasma proteome profiling has ability to detect the cancer before it actually develops (316). They have shown linear correlation between plasma levels of the candidate proteins with the tumor progression and these changes are reversible as they return to the original level with tumor regression. The majority of those candidate plasma proteins were acute-phase proteins, immune cell proteins, cytoskeletal, extracellular matrix proteins and quite a few amongst them were glycoproteins (316).

Different cell types express different glycosylation machinery. For example, haptoglobin (HP) is an acute phase protein which is mainly synthesized in liver and its main function is clearance of hemoglobin at the site of inflammation. The β-chain of HP undergoes N-linked glycosylation and has a molecular weight of about 39 kDa. However, neutrophils during their specific stage of development also synthesize HP but with different glycosylation leading to an apparent molecular weight of 45-65 kDa on western blot (340). Furthermore, upon activation this differentially glycosylated HP is released along with all other granular contents (340). This explains one more possibility for specific glycan changes in the circulatory glycoproteins.

1.5.1 Methodologies for glycoprotein biomarker discovery

The common aims in glycoprotein studies are to identify substrate protein undergoing glycosylation, to determine the actual site of glycan modification and to elucidate the glycan structure for each site. HPLC coupled mass spectrometry is the most commonly used technique to study glycan modifications. Advancements in the bioinformatic tools and development of relevant databases such as GlycoGene database, lectin frontier database, GlycoProtein database, Glycan mass spectral database have facilitated the glycoprotein research (341). Overall glycoprotein studies are mainly driven by advancements in the chromatographic, mass spectrometric and bioinformatics techniques.

Biological samples are fairly complex and significant pre-processing is required before they can be used for glycoprotein analysis. In order to detect medium or low abundance glycoprotein, various enrichment techniques are developed. Followed by glycoprotein enrichment, there are two main approaches for glycoprotein detection, top-down workflow (glycoprotein based) or bottom-up workflow (glycopeptide based) analysis. As the name suggests the top-down workflow initially enrich glycoprotein followed by trypsin digest while for glycopeptide based techniques, the protein is typsin-digested first followed by glycopeptide enrichment. Tryptic peptides can be directly analyzed by LC-MS to identify the underlined proteins. An additional enzymatic glycan release step (i.e. PNGase F treatment for N-glycan release) is required before mass spectrometric analysis and glycan structural elucidation (342). Alternatively, glycosylation can be detected on the glycopeptide level to get the connectivity between the glycan and the peptide carrier (343). The following section briefly discusses glycoprotein/glycopeptide enrichment techniques.

1.5.1.1 Using lectins in the glycoprotein enrichment workflows

The glycoprotein enrichment workflows can be broadly categorized into two different types. (i) Lectin based, and (ii) non-lectin based. Lectins are carbohydrate binding proteins and occur abundantly in nature. They generally recognize glycan structures with low affinity but with high avidity mainly through hydrogen bonding, hydrophobic interactions and van der Waals forces (344). A range of lectins for selective enrichment of glycan moiety are commercially available in different formats to suit specific experimental requirements, these have been reviewed previously (309, 310). To name few, lectins are available as immobilized on solid support such as agarose, silica or even magnetic beads. They are applied in different chromatographic settings such as tubes, columns and microfluidic channels. One single lectin can be used in isolation to enrich narrower range of glycoepitopes or multiple lectins can be used in combination in order to enrich broad range

of glycan structures (345). Multiple lectins with different carbohydrate specificity is used serially to enrich selective glycan structures in a process called serial lectin affinity chromatography. This particular approach was successfully applied particularly to isolate O-linked glycans (346). In order to facilitate high-throughput glycan analysis lectin microarrays have emerged where series of lectins are spotted on the solid support, samples are tagged with fluorophore and detection is carried out using fluorescence microarray scanners (347). Recently our laboratory developed lectin magnetic bead array-tandem mass spectrometry (LeMBA-MS/MS), a high-throughput platform where array of lectins are immobilized on the magnetic beads to capture glycoproteins followed by on bead trypsin digest in line with mass spectrometry (317, 318). These lectin based high-throughput techniques have been playing key role in glycobiomarker research.

Apart from using lectin biology for enrichment, chromatographic technique such as hydrophilic interaction chromatography (HILIC) (348) and chemical modification techniques have been developed for isolating glycopeptides. Based on hydrazide chemistry, Zhang and colleagues in 2003 developed solid phase extraction method for isolating glycoprotein (349) and later modified it to make it suitable for glycopeptides (350). Utilizing fact that Boronic acid form stable cyclic esters with cis-diol containing carbohydrates at higher pH, Zhou and colleagues developed magnetic nanoparticles to capture glycoprotein/glycopeptides and being a reversible reaction, captured glycoprotein/glycopeptides can be released at lower pH simply by acid elution (351). The main difference between lectin-based and non-lectin based protocols is that there are many lectins which bind to different glycan groups but chemical methods are usually not selective for a particular class of glycan and enrich samples for glycoproteins as a whole.

1.5.1.2 LeMBA

To improve the throughput of lectin based glycoprotein biomarker discovery platforms, our laboratory previously developed lectin magnetic bead array (LeMBA) workflow (Figure 1.6) (317, 318). Unlike single, serial and multi lectin affinity chromatography, LeMBA uses 20 individual lectin coated magnetic beads in microplate format to enrich a sub-population of serum glycoproteins having high-affinity to bind with a particular lectin. Table 1.6 contains details of lectins used in LeMBA workflow along with their general glycan specificity. LeMBA incorporates following key steps: (i) coupling lectins with MyOne™ Tosylactivated Dynabeads® (magnetic beads), (ii) Spike serum samples with internal standard ovalbumin and sample denaturation, (iii) glycoprotein capture, (iv) removal of unbound proteins by several washing steps and (v) elution of bound proteins for western immunoblotting or on-bead trypsin digest for mass spectrometric

analysis. With the use of magnetic beads, LeMBA allows single-step isolation of serum glycoprotein using 20 individual lectin-coated magnetic beads. Composition of different buffers used during LeMBA protocol namely denaturation buffer, binding buffer and washing buffer were carefully optimized for salts and detergents to avoid co-isolation of protein complexes without adversely affecting lectin pull-down efficiency. Internal standard chicken ovalbumin can account for minor variations during sample processing and pull-down steps between different time points. Glycoprotein capture, washing and on-bead trypsin digestion steps are performed on liquid handler which makes LeMBA semi-automated and reproducible over several weeks of sample preparation. LeMBA is a versatile platform and can be coupled with diverse protein identification, characterization and quantitation techniques. When coupled with nano-HPLC-MS/MS to identify non-glycosylated peptides from the isolated glycoproteins with lectin exclusion list for protein identification, it has demonstrated nanomolar sensitivity and linearity. Last but not the least it is applicable across different species.

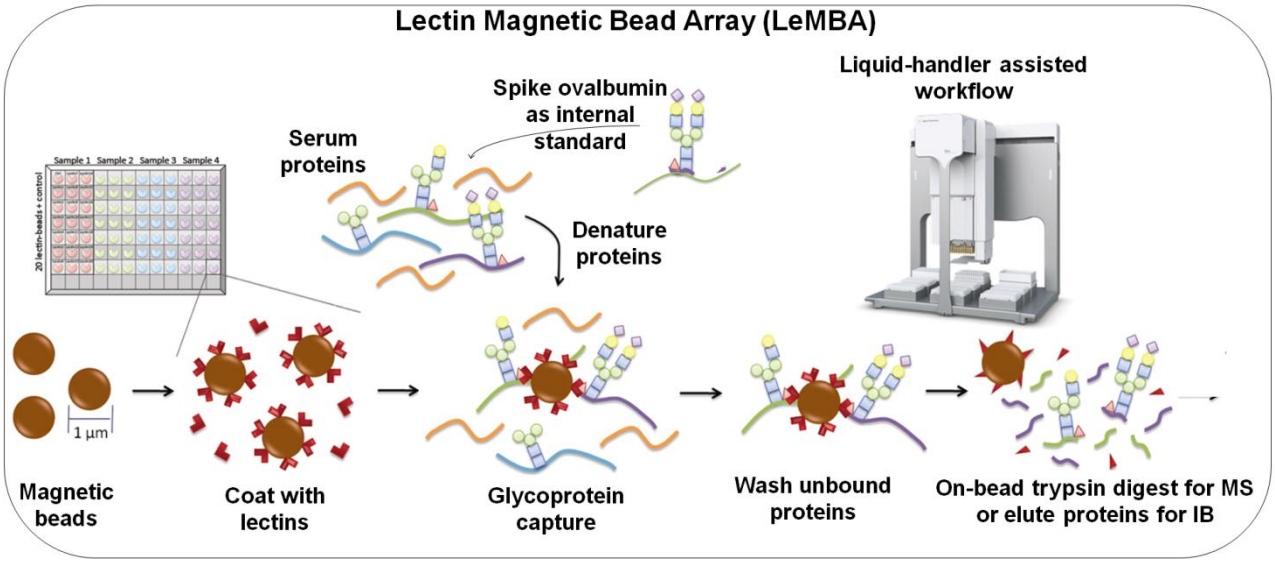


Figure 1.6. Lectin magnetic bead array (LeMBA) workflow for serum glycoprotein biomarker discovery and development. Adapted from Choi *et al.* (317).

Table 1.6. Glycan specificity of lectins used in LeMBA. Adapted from Choi et al (317).

Carbohydrate	Abbreviations	Source	Ligand motif	Supplier
Mannose	ConA	Jack bean	α-Man, α-Glc, α-GlcNAc	Sigma
	GNL	*Galanthus nivalis*	Manα1-3Man terminal	Vector
	NPL	Daffodil	Manα1-6Man	Vector
N-acetylglucosamine	DSA	*Datura stramonium*	β1-4GlcNAc oligomers	Vector
	HAA	*Helix aspersa*	α-GlcNAc and α-GalNAc	Sigma
	STL	Potato	GlcNAcβ1-4GlcNAc oligomers	Vector
	WGA	Wheat germ	GlcNAcβ1-4GlcNAc , Neu5Ac	Sigma

Carbohydrate	Abbreviations	Source	Ligand motif	Supplier
Galactose	BPL	*Bauhinia purpurea*	Galβ1-3GalNAc	Vector
	ECA	*Erythrina cristagalli*	Galβ1-4GlcNAc	Sigma
	JAC	Jackfruit (Jacalin)	Galβ1-3GalNAc	Vector
N-acetylgalactosamine	SBA	Soy bean	GalNAcα1-3Gal	Sigma
	HPA	*Helix pomatia*	α-GalNAc	Sigma
	WFA	*Wisteria floribunda*	GalNAcα1-6Gal/α1-3GalNAc	Sigma
Fucose	AAL	*Aleuria aurantia*	Fucα1-6GlcNAc	Vector
	PSA	*Pisum sativum*	Fucα1-6GlcNAc	Sigma
	UEA	*Ulex europaeus*	Fucα1-2Galβ1-4GlcNAc	Vector
Sialic acid	MAA	*Maackia amurensis*	Neu5Acα2-3	Vector
	SNA	Elderberry	Neu5Acα2-6	Vector
Others	EPHA	Erythroagglutinin *Phaseolus vulgaris*	Bisecting GlcNAc	Vector
	LPHA	Leucoagglutinin *Phaseolus vulgaris*	Tri/Tetra-antennary β1-6GlcNAc	Vector

1.6 Thesis aims and significance

Current endoscopy-biopsy based clinical practice for diagnosis and management of EAC pathogenesis hasn't been able to curb morbidity and mortality related to this lethal malignancy. There is an urgent need to identify circulatory biomarkers which can be developed further into a blood based diagnostic test (1). Out of different classes of circulatory biomarkers, serum glycoproteins are potential candidates. In fact, serum glycan profile differs between healthy, BE, dysplastic and EAC patients (167-169, 174, 311). However, specific glycoprotein biomarker candidates showing differential glycosylation are not known. This study hypothesized that EAC pathogenesis is associated with changes in the glycosylation of serum proteins hence serum glycoprotein can act as a potential diagnostic biomarkers to monitor EAC pathogenesis. Based on this working hypothesis, this thesis has the following aims.

(i) To discover serum glycoprotein biomarker candidates using LeMBA workflow that distinguish between healthy, EAC and metaplastic BE phenotype.

(ii) To develop a targeted proteomics approach to measure biomarker candidates for a timely verification.

(iii) To verify serum glycoprotein candidates identified in Aim 1 (Chapter 3) using targeted proteomics approach developed in Aim 2 (Chapter 4) in an independent patient cohort.

(iv) The mass spectrometric approach employed to address Aim 1 to 3 is best suited for research laboratory but not for routine clinical diagnostic. So final aim of this project was to test feasibility of using electrochemical detection methodology for the glycoprotein detection.

Chapter 3 describes extension of LeMBA with development of GlycoSelector database incorporating statistical analysis pipeline. Using this pipeline, lectin-glycoprotein biomarker candidates that distinguish between healthy, BE and EAC phenotype are reported.

Chapter 4 addresses Aim 2, which include development of targeted proteomics assay for a timely verification of biomarker candidates described in Chapter 3. It also describes testing linearity and reproducibility of the targeted proteomics assay.

By screening an independent cohort of serum samples using targeted proteomics approach described in Chapter 4, verified list of serum glycoprotein biomarker candidate has been described in Chapter 5.

The publication of Aim 4, described in Chapter 6, has reported the proof-of-concept work to use electrochemical detection method for glycoprotein detection.

Collectively, this thesis describes discovery and verification of serum glycoprotein biomarker candidates for BE/EAC diagnosis. The electrochemical method has been successfully applied for glycoprotein detection/quantitation and can be developed further as a clinically useful diagnostic platform.

Chapter 2.

MATERIALS AND METHODS

Chapter 2. Materials and methods

Materials and methods used frequently across different thesis chapters are described in this chapter. For methods belong to specific chapter please refer to respective chapter.

2.1 Reagents

MyOne™ Tosyl activated Dynabeads® were from life technologies. Lectins *Aleuria aurantia* lectin (AAL), *Bauhinia purpurea* lectin (BPL), *Datura stramonium* agglutinin (DSA), Erythroagglutinin Phaseolus vulgaris (E-PHA), *Galanthus nivalis* lectin (GNL), Jacalin (JAC), Leucoagglutinin Phaseolus vulgaris (L-PHA), *Maackia amurensis* agglutinin-II (MAA), *Narcissus pseudonarcissus* lectin (NPL), *Sambucus nigra* agglutinin (SNA), *Solanum tuberosum* lectin (STL) and *Ulex europaeus* agglutinin-I (UEA) were from Vector Laboratories. Modified sequencing grade trypsin was from Promega. Protein assay dye reagent (Bradford reagent), Triton X-100 and SDS solution were from Bio-rad. Tris base, glycine, sodium chloride and acrylamide/bis-acrylamide solution 40% w/v 29:1 were from Amresco. Glycerol, disodium hydrogen phosphate dihydrate, sodium dihydrogen phosphate dihydrate, calcium chloride dihydrate ($CaCl_2$) and tween-20 were from Ajax Finechem. Magnesium chloride ($MgCl_2$) and manganese chloride ($MnCl_2$) were from Univar. Immobilon-P Polyvinylidene difluoride (PVDF) was from Millipore and SuperSignal West Pico Chemiluscent Substrate was from Thermo Scientific. Monoclonal anti-gelsolin antibody was from Epitomics (EP1940Y), polyclonal anti-haptoglobin was from Gen Way Biotech (GWB-16A7EA) and HRP labeled anti-rabbit HRP was from Invitrogen (A10547). The developer and fixer for western blot development were from Kodak. For quadrupole time of flight runs, acetonitrile, isocratic HPLC grade was from Scharlau and for triple quadrupole runs, acetonitrile CHROMASOLV® gradient grade was from Sigma. Mass spectrometry reagents were from Agilent Technologies. All other reagents, including lectins not listed above, were from Sigma unless otherwise specified.

2.2 Buffers and solutions

2.2.1 LeMBA

Protease inhibitor cocktail (PI) 1000x stock (diluted to 1x in final solution) – 1 μg/μL Aprotinin, 1 μg/μL Antipain, 1 μg/μL Pepstatin A, 1 μg/μL Leupeptin and 250 mM Benzamidine

Lectin resuspension buffer – 20 mM sodium dihydrogen phosphate monohydrate, 80 mM disodium hydrogen phosphate dihydrate, 0.1 mM $CaCl_2$, 0.1 mM $MnCl_2$

Bead activation buffer – 3M ammonium sulfate (pH 7.9)

Bead blocking buffer – 1 M Tris-HCl (pH 7.4), 1% w/v glycine

Bead storage buffer – 20 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 0.5% v/v Triton, 1x PI

Denaturation buffer – 20 mM Tris-HCl (pH 7.4), 20 mM DTT, 1% w/v SDS and 5% v/v Triton

Binding buffer A (for lectins EPHA, SNA and STL) – 20 mM Tris-HCl (pH 7.4), 0.2% w/v SDS, 1 mM DTT, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 1% v/v Triton, 1x PI

Binding buffer B (for remaining all lectins) – 20 mM Tris-HCl (pH 7.4), 0.05% w/v SDS, 1 mM DTT, 300 mM NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 1% v/v Triton, 1x PI

Washing buffer A (for lectins EPHA, SNA and STL) – 20 mM Tris-HCl (pH 7.4), 0.2% w/v SDS, 1 mM DTT, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 1% v/v Triton

Washing buffer B (for remaining all lectins) – 20 mM Tris-HCl (pH 7.4), 0.05% w/v SDS, 1 mM DTT, 300 mM NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 1% v/v Triton

50 mM ammonium bicarbonate for washing and trypsin resuspension

2.2.2 SDS-PAGE and western immunoblotting

Laemmli sample buffer 5x stock (diluted to 1x for use) - 0.2 M Tris-HCl (pH 6.8), 10% w/v SDS, 50% v/v glycerol, 0.25 M DTT, 0.1% w/v bromophenol blue

Buffer A for resolving gel – 750 mM Tris-HCl (pH 8.8), 0.2% w/v SDS

Resolving gel –

Ingredients	**Gel percentage**				
	6%	**8%**	**10%**	**12%**	**15%**
MiliQ water	34.00%	29.00%	24.00%	19.00%	11.50%
Acrylamide/Bis-acrylamide solution (40%)	15.00%	20.00%	25.00%	30.00%	37.50%
Buffer A	50.00%	50.00%	50.00%	50.00%	50.00%
Ammonium persulphate (APS) solution (10%)	1.00%	1.00%	1.00%	1.00%	1.00%
Tetramethylethylenediamine (TEMED)	0.05%	0.05%	0.05%	0.05%	0.05%

Buffer B for stacking gel – 250 mM Tris-HCl (pH 6.8), 0.2% w/v SDS

Stacking gel – 39% MiliQ water, 10% v/v Acrylamide/Bis-acrylamide solution (40% w/v), 50% v/v Buffer B, 1% v/v of 10% w/v Ammonium persulphate (APS) solution, 0.1% v/v Tetramethylethylenediamine (TEMED)

Gel running buffer 10x stock (diluted to 1x for use) – 250 mM Tris, 1.92 M glycine, 1% w/v SDS

Transfer buffer 10x stock (wet transfer) – 1.92 M glycine, 250 mM Tris

Transfer buffer 1x (wet transfer) – 10% v/v of 10x transfer buffer, 20% v/v methanol

Tris buffered saline (TBS) for western blotting (pH 7.4) 10x stock – 500 mM Tris, 1.5 M NaCl, approximately 0.3% v/v HCl to adjust pH 7.4

Tris buffered saline Tween 20 (TBST) for western blotting (pH 7.4) – 10% v/v of 10x TBS, 0.1% v/v Tween 20

Strip solution – 1M glycine (or 5M sodium hydroxide), 5% v/v HCl, and 1% w/v SDS

Developer – 130 mL developer solution diluted in 270 ml of tap water

Fixer – 130 mL fixer solution diluted in 270 mL of tap water

Coomassie Brilliant Blue for membrane staining – 0.25% w/v Coomassie Brilliant Blue R-250, 7% v/v acetic acid, 40% v/v methanol

Destain for membrane – 40% v/v methanol, 7% v/v acetic acid

Colloidal Coomassie blue for gel staining – 10% w/v ammonium sulfate, 10% v/v orthophosphoric acid, 0.12% w/v Coomassie Blue G-150, 20% v/v methanol

Destain for gel – 1% v/v acetic acid

2.2.3 Mass spectrometry

Buffer A – 0.1% v/v mass spectrometry grade formic acid in MilliQ water

Buffer B – 90% v/v isocratic grade acetonitrile in 0.1% v/v mass spectrometry grade formic acid in MilliQ water for Chapter 3. 0.1% v/v mass spectrometry grade formic acid in gradient grade acetonitrile for Chapter 4 and Chapter 5.

Injector needle wash – 20% v/v methanol in 0.1% v/v formic acid made up in MiliQ water

2.3 Sample collection

The study was approved by The University of Queensland Human and Animal Ethics Committees. Samples were randomized prior to all experiments. Serum samples from healthy, BE and EAC patients were acquired through the Australian Cancer Study (ACS) (352) and Study of Digestive Health (SDH) (353). All 29 serum samples (Healthy-9, BE-10 and EAC-10) used for

biomarker discovery phase and 79 serum samples (Healthy-20, BE-20, EAC-20 and population control-19) used for biomarker qualification study were age and gender matched. Healthy controls were individuals with no history of esophageal cancer and no evidence of esophageal histological abnormality at the time of sample collection. BE patients had a histologically confirmed diagnosis of Barrett's mucosa. EAC patients had histologically confirmed adenocarcinoma within the distal esophagus or gastro-esophageal junction. Although cancer staging for EAC patients was available for verification cohort, patients were not stratified according to disease progression due to relatively small sample size. The EAC cohort consists of patients from early to late stages of EAC. EAC patient sera were collected prior to the commencement of cancer treatment. Population controls were volunteers with no self-reported history of EAC or BE. All subjects signed written informed consent as a part of sample collection process. For categorical and numerical variables related to patient information, *P* values were calculated using Fisher's exact test and Kruskal-Wallis test respectively. 10 ml of serum was collected from each patient during the trial. The samples (50 µl for biomarker discovery and 500 µl for biomarker verification) were received from QIMR Berghofer Medical Research Institute and stored at -80 °C until use. Typically, samples were thawed once for protein estimation and at the same time samples were denatured. Particularly for biomarker verification, samples were aliquoted into two different tubes (50 µl each) during first freeze thaw for future use.

2.4 Protein methods

2.4.1 Bradford protein assay

Serum samples were diluted 1 in 100, 1 in 150 or 1 in 200 in MiliQ water. Bovine serum albumin (BSA) standard solutions at the concentration of 0 mg/mL (blank), 0.2 mg/mL, 0.4 mg/mL, 0.6 mg/mL, 0.8 mg/mL and 1 mg/mL were made. 5 µL of BSA standard or 5 µL of diluted sera were arrayed at least in duplicate in a clear 96 well plate. Bradford reagent was diluted 5 times in MiliQ water and 200 µL of this dilution was added to each well. Whenever required, solution in 96 well plate was mixed using liquid handler. Absorbance was measured at 595 nm using FLUOstar OPTIMA microplate reader (BMG Labtech). Raw absorbance values were exported to Microsoft excel. Standard curve using absorbance values of known BSA concentrations was plotted and protein concentration of samples was determined using slope and y-intercept of the standard curve.

2.4.2 SDS-PAGE and western immunoblotting

Mini-PROTEAN® tetra cell system coupled with wet transfer assembly from Bio-Rad was used. Protein samples were boiled in 1x Laemmli sample buffer at 95 °C for 5 min (354). For

LeMBA-immunoblotting, 2x or 3x concentration of sample buffer was used. The samples were resolved on appropriate SDS-PAGE gels and transferred to Immobilon-P polyvinylidene difluoride (PVDF) membranes using wet transfer. The membranes were blocked with either 5% w/v non-fat milk or 5% w/v BSA for 30 min at room temperature. The membranes were probed with primary antibody overnight at 4 °C. On next day, after washing membranes 5 times with 1x TBST for 5 min each, the membranes were probed with appropriate secondary antibody conjugated with horseradish peroxidase (HRP) for 2 hr at room temperature. After washing membranes at least 3 times with 1x TBST for 5 min each, the membranes were developed using SuperSignal West Pico chemiluminescence, and captured on film. Densitometric analysis was performed using ImageJ (NIH, USA) (355).

2.5 Lectin magnetic bead array (LeMBA)

LeMBA was performed as described previously, with modifications (317, 318).

2.5.1 Coupling of lectins with Dynabeads®

Lectins are supplied as lyophilized powder which may or may not contain necessary salts in the vials. Lectins were reconstituted using either MiliQ water (lectins supplied containing necessary salts) or lectin resuspension buffer (lectin supplied without any salts) and stored at -80 °C in aliquots. Lectins were covalently attached with Dynabeads via primary amine or sulphydryl group (https://tools.thermofisher.com/content/sfs/manuals/dynabeads_myone_tosylactivated_man.pdf).
For each lectin, 100 µL of 100 mg/mL MyOne tosyl activated Dynabeads® were washed three times with lectin resuspension buffer using magnetic tube holder. To the beads, 100 µL of 3 M ammonium sulphate (pH 7.9) was added and beads were mixed using vortex. 100 µL of 5 mg/mL lectin was added to the activated beads and tubes were incubated rotating at 20 RPM for 24 hr at 37 °C. On the next day, the supernatant was removed and non-reacted sites of the beads were blocked by incubation with 1 mL of 1 M Tris containing 1% w/v glycine for 16 hr at 37 °C. On third day, supernatant was removed and lectin-beads were washed three times using bead storage buffer. The beads were resuspended using 1 mL of bead storage buffer and stored at cold temperature. 50 µL of this lectin-bead conjugate is used for one pull-down experiment with 50 µg of denatured serum protein. To minimize experimental variation due to batch effect, lectin-beads sufficient for each phase of biomarker discovery and verification experiments were made at once and used within 3 months. The lectin to Dynabeads coupling ratio is 10 mg of beads per 0.5 mg of lectin which is sufficient for 20 pull-downs after conjugation.

2.5.2 Serum sample preparation for LeMBA

The Bradford protein assay was performed to measure serum protein concentration. The serum samples were spiked with 10 picomole ovalbumin per reaction as an internal standard i.e. 50 µg of serum proteins were spiked with 10 picomole of chicken ovalbumin. The serum protein mixture was denatured and reduced in denaturation buffer at 60 °C for 30 min followed by alkylation with 100 mM iodoacetamide for 1 hr at room temperature maintaining dark condition.

2.5.3 Liquid handler assisted LeMBA pull-down and trypsin digestion

For each pull down experiment, lectins [all lectins (20 lectins plus control bead) for biomarker discovery or specified lectins for biomarker verification] were arrayed in each well of a 96 well plate. The 20 different lectins were chosen to accommodate natural diversity of glycans. The lectins were selected according to glycan recognition epitopes as per literature and described by Choi *et al* (317). The Bradford protein assay was performed to measure serum protein concentration. The serum samples were spiked with 10 pmol ovalbumin per reaction as an internal standard. The serum protein mixture was denatured and reduced using denaturing buffer at 60°C for 30 min followed by alkylation with 100 mM iodoacetamide for 1 hr at 37°C maintaining dark condition. 50 µg alkylated serum sample per reaction was incubated with lectin conjugated beads in 100 µl binding buffer at 4°C for 1 hr on the plate shaker. Following glycoprotein capture, beads were washed (i) with binding buffer for 3 times then washed with (ii) 50 mM ammonium bicarbonate for seven times including changing plates three times in-between washes. For on-bead trypsin digest, 0.95 µg of sequencing grade trypsin in 20 µl of 50 mM ammonium bicarbonate was added to each reaction mixture and incubated at 37°C overnight. The next day, digested peptides were transferred to a new plate. Beads were washed with an equivalent volume of 50 mM ammonium bicarbonate, and supernatant was combined with digested peptides. Pooled peptide samples were dried under the vacuum and plates were stored at -80°C until further use. Bravo liquid handler (Agilent Technologies) was used to make the platform high throughput.

2.5.4 LeMBA-western immunoblotting

LeMBA pull-down until glycoprotein capture step was performed as described above in section 2.5.3. After glycoprotein capture, beads were washed only with binding buffer for 3 times. All further washing steps with ammonium bicarbonate and trypsin digest were not required. Beads were directly boiled in 2x Laemmli sample buffer to elute captured glycoproteins, run on SDS-PAGE and proteins were transferred to PVDF membrane using wet transfer. Western immunoblotting using antibody raised against target glycoprotein was performed as described in

section 2.4.2. To compare results across membranes, pull-down sample from one healthy patient serum (unrelated to samples used in screen) was loaded in equal amounts on all gels. Raw densitometric values were normalized using an internal control sample loaded onto each gel.

2.6 Statistical analysis

Statistical analysis was performed using Microsoft Excel 2007, GraphPad Prism 6, R statistical programming language and web-based tools GlycoSelector for biomarker discovery (http://glycoselector.di.uq.edu.au/) and Shiny mixOmics for biomarker verification (http://mixomics-projects.di.uq.edu.au/Shiny). The details about steps followed for statistical analysis is explained in the respective result chapters. Both biomarker discovery and verification employed relative quantitation using internal standard chicken ovalbumin spiked at the step of sample denaturation. The common steps for GlycoSelector and Shiny mixOmics include, (i) uploading data, (ii) outlier detection, (iii) group binding difference analysis for GlycoSelector OR univariate statistical analysis including determination of AUROC for Shiny mixOmics, and (iv) Sparse partial least squares-discriminant analysis (sPLS-DA) for multivariate feature selection including stability assessment of a candidate using leave-one-out validation for GlycoSelector OR cross-validation for Shiny mixOmics.

Chapter 3.

SERUM GLYCOPROTEIN BIOMARKER DISCOVERY USING LEMBA−GLYCOSELECTOR PIPELINE

Chapter 3. Serum glycoprotein biomarker discovery using LeMBA−GlycoSelector pipeline

3.1 Introduction

Biomarker discovery is the first important step in the paradigm of biomarker discovery and development process (133, 356). In clinical practice, biomarkers can aid in prediction, cause, diagnosis, staging, regression, selection/monitoring of treatment or outcome of treatment for a disease (126, 357). The following considerations should be taken into account prior to biomarker discovery screen: (i) Unmet clinical need, (ii) rationale of using a particular class of biomolecule to be screened for biomarker purpose, (iii) biological sample to be used for screening, (iv) expected outcome of the screen. Usually clinical decisions are made not just based on one biomarker but it takes into consideration patient history, nature of illness and results from other available noninvasive medical imaging techniques, such as magnetic resonance imaging (MRI) and ultrasound (358). In the central dogma of molecular biology, DNA, RNA, protein or metabolite can serve as a biomarker candidate. The particular interest of using protein candidate as a biomarker candidate lies in the diversity of proteins found in the biology. The estimates suggest there are 20,300 genes (359), 41,993 metabolites (360, 361) and 100,000 mRNA transcripts (358) present in a human being. Even with modest estimates considering the variety of post-translational modifications on proteins, the number of different protein species can easily be more than a million (362, 363). This enormous diversity of protein variants possesses an immense opportunity to identify biomarker candidates. At the same time, it is a challenge for available analytical methodologies to correctly identify and quantitate a specific protein variant in biological samples (358).

It is not surprising to know that protein biomarkers are already widely used in pathology laboratories to assist clinicians in decision making leading to better patient management hence improved health outcomes. Human plasma/serum has been described as the most comprehensive and complex proteome (362). It is a circulating representation of all body tissues and reflects both physiological and pathological processes. It is mainly composed of proteins secreted by solid tissues (mainly liver and intestine) that act in plasma, immunoglobulins, "long distance" receptor ligands that include classical peptide and protein hormones, "local receptor" ligands that include cytokines, temporary passengers e.g. lysosomal proteins that are secreted and then taken up via a receptor for sequestration in the lysosomes. It also contains tissue leakage products such as cardiac troponins, creatinine kinase, or myoglobin used in the diagnosis of myocardial infarction, secretions from blood or immune cells and foreign proteins e.g. protein originated from infectious organisms

or parasites (362). The remarkable progress in mass spectrometry, high-throughput antibody production, bioinformatics and biostatistics algorithms in the past couple of decades now enables the study of this complex plasma/serum proteome in a holistic manner (358). Almost half of the blood proteins are glycoproteins and it is a major contributor to the diversity of protein species observed/predicted in the serum/plasma proteome (362). As described in section 1.5.1.1 of the thesis, our laboratory recently developed LeMBA workflow which uses 20 naturally occurring glycan binding proteins lectins to enrich sub-glycoproteome from a serum sample. Here, LeMBA was combined with discovery proteomics platform and GlycoSelector data analysis pipeline for BE/EAC diagnostic biomarker discovery.

3.1.1 Relative quantitation in proteomics based biomarker discovery pipeline

Three main steps in biomarker discovery pipelines are (i) sample preparation, (ii) mass spectrometric analysis for protein identification and quantitation, and (iii) data analysis. Sample preparation for glycoprotein biomarker discovery using LeMBA protocol is semi-automated, reproducible, and high-throughput hence suitable for screening enough number of patient samples required for biomarker discovery (317). Moreover, it uses serum sample as a source of biomarker discovery from beginning of the study (317, 318).

For mass spectrometry based discovery proteomics methods, the main objective is to identify as many candidates as possible in the sample with reliable relative quantitation approach to compare between different disease states. Two main approaches for protein identification exist based on mass spectrometry analysis (i) top-down proteomics (364-366), and (ii) bottom-up proteomics (367). Top-down proteomics aims for intact proteins and retain a lot of information about protein sequence, protein isoforms and post-translational modifications (364-366). Top-down proteomics of biological samples (e.g. serum) result in complex spectrum which is difficult to annotate hence limits the protein identification. On contrary, bottom-up proteomics is based on the identification of protein cleavage products (mainly peptides), and provide high sensitivity resulting into very high number of protein identifications in biological samples and remain the method of choice. Bottom-up proteomics involves the following steps: (i) protein digestion by proteolytic enzyme with known sequence specificity e.g. trypsin, (ii) peptide separation by LC, (iii) peptide ionization, (iv) peptide fragmentation, and (v) detection of mass-to-charge ratios (*m/z*) and abundance of peptide ions and their fragment ions (367). Two main types of modern day mass spectrometers are used for proteomics purposes. (i) Time-of-flight (TOF) analyzers combined with quadrupole for ion selection and electrospray ionization (ESI) source can provide high sensitivity,

high mass accuracy (2 to 5 ppm), high resolution (10,000 to 40,000 in MS1 and MS/MS modes respectively), and fast scan time (368). (ii) Combination of an Orbitrap analyzer with ion-trap for ion selection and ESI source provides high mass accuracy (1 to 5 ppm) in MS and MS/MS modes, resolution up to 240,000, and relatively fast scan speeds (369, 370). Latest instrument based on Fourier transform ion cyclotron resonance (LTQ-FTICR), offers capabilities of an Orbitrap with resolutions up to 750,000 (371, 372).

There are three main approaches for incorporating quantitation into biomarker discovery pipeline: (i) Metabolic and enzymatic labeling, (ii) chemical labeling, and (iii) Label-free quantitation. Stable isotope labeling with amino acids in cell culture (SILAC) is a common form of metabolic labeling whereby treated (experiment group) cells are grown in media containing heavy isotope-labeled ($^{13}C_6$ and $^{15}N_7$) amino acids as compared to control cells which are grown in light isotope-lebaled ($^{12}C_6$ and $^{14}N_7$) amino acids or vice versa. Upon more than 5 cell divisions in the respective media, these amino acids are incorporated into protein sequence during translation in the cells. Equimolar mixture of these cell lysates is mixed and analyzed by discovery proteomics methods using mass spectrometric analysis. The peptides from heavy labeled amino acids differ from light counterpart in terms of mass hence show mass shift in precursor ion *m/z*. The final result is reported in the form of ratio between heavy and light peptide levels (373). When ^{18}O containing water ($H_2{}^{18}O$) instead of normal water ($H_2{}^{16}O$) is used during enzyme digestion, enzymatic labeling takes place. Exchange of two ^{16}O atoms for two ^{18}O atoms on C-terminal of peptides occurs which is observed in the form of mass shift of 4 Da (374). Chemical labeling can be accomplished using heavy or light isotope-labeled and chemically reactive tags. Example includes isotope-coded affinity tags (ICAT) that allow for labeling of cysteine residues in proteins with either heavy or light version, followed by affinity purification and mass spectrometric analysis to determine ratio of heavy and light isotope-labeled tags for quantification (375). Isobaric tags for relative and absolute quantification (iTRAQ) (376) or tandem mass tags (TMT) (377) produce reporter ions upon peptide fragmentation for quantitation and also allow multiplexing up to some extent. The limitation of metabolic or chemical labeling procedure is addition of one more step to already complex sample preparation workflow, with more chance for sample loss, inaccuracies and mix-ups. Moreover, labeling techniques are not plausible for screening many patient samples (358). Label-free quantitation offers a cheap alternative as opposed to labeling method, allows wider coverage of dynamic range, and suitable for screening large number of patient samples (378). Spectral counting or extracted ion chromatograms are two main options to implement label-free proteomics quantitation. It has also been demonstrated that addition of internal standard protein in complex

biological sample results into accurate relative quantitation using label-free proteomics workflow (378).

Most proteomics workflow generates list of peptides/proteins identified for each sample with some quantitative information. Many commonly used software tools can handle conventional proteomics data sets and spits out list of candidates that differentiate two groups being compared. Inclusion of 20 different lectin pull-down per sample to capture glycosylation differences adds one more dimension to LeMBA data sets hence development of appropriate data analysis platform is warranted. Generally multivariate approaches such as principal component analysis (PCA) and partial least squares (PLS) are employed for analyzing omics data sets (379). It is advisable to validate the multivariate analysis model using ideally an independent data set. In case when independent data is not available then internal validation such as leave-one-out or cross-validation approaches can be utilized (379, 380).

The aim of this chapter is to extend LeMBA workflow with development of GlycoSelector database incorporating statistical analysis pipeline. Using this LeMBA-GlycoSelector workflow, serum samples from healthy, BE and EAC patients were screened to identify diagnostic glycoprotein biomarker candidates.

3.2 Experimental procedures

3.2.1 Sample information

Serum samples from 29 patients (healthy - 9, BE - 10 and EAC - 10) were randomized across 8 plates in the given order for LeMBA pull-down and mass spectrometry analyses (Plate 1 - 2, 30, 16, 24; Plate 2 - 14, 22, 8, 27; Plate 3 - 23, 3, 12, 9; Plate 4 - 10, 17, 28, 18; Plate 5 - 29, 6, 19; Plate 6 - 13, 26, 4, 25; Plate 7 - 15, 5, 21, 1; and Plate 8 - 7, 11). Details regarding source of the serum samples is described in section 2.3. Table 3.1 contains information of the samples used for biomarker discovery. For categorical and numerical variables related to patient information, *P* values were calculated using Fisher's exact test and Kruskal-Wallis test respectively. Table 3.2 summarizes clinical characteristics of the patient cohort for biomarker discovery.

Table 3.1. Details of samples used for biomarker discovery.

Sample number	Mass spec run round	GlycoSelector sample ID	GlycoSelector run ID	Study ID	Patient phenotype	Gender	Ref age	Protein (mg/mL)
2	1	31	35	47024	SDH control	M	43.02	80.63

Sample number	Mass spec run round	GlycoSelector sample ID	GlycoSelector run ID	Study ID	Patient phenotype	Gender	Ref age	Protein (mg/mL)
3	1	32	36	45180	SDH control	M	68.23	107.58
4	3	52	56	43187	SDH control	M	56.76	133.68
5	2	45	49	92365	Population control	M	76.46	67.92
6	3	53	67	96191	Population control	M	71.42	106.31
7	3	54	58	94429	Population control	M	56.72	94.80
8	2	46	50	93411	Population control	M	62.07	79.90
9	1	33	37	95091	Population control	M	66.35	88.01
10	1	34	38	96190	Population control	M	70.93	97.02
1	2	44	48	43084	SDH control (later on developed BE so classified as BE)	M	67.81	65.57
11	3	55	59	45004	BE no dysplasia	M	76.07	87.17
12	1	35	39	43026	BE no dysplasia	M	72.19	99.81
13	3	56	60	43004	BE no dysplasia	M	72.81	106.76
14	2	47	51	45052	BE no dysplasia	M	78.07	83.86
15	2	48	52	45050	BE no dysplasia	M	33.91	86.92
16	1	36	40	43113	BE no dysplasia	M	55.95	107.58
17	1	37	41	45137	BE no dysplasia	M	39.07	82.68
18	1	38	42	47007	BE no dysplasia	M	55.10	83.46
19	3	57	61	43115	BE no dysplasia	M	56.64	77.16
21	2	49	53	33100	EAC patient	M	55.86	73.42
22	2	50	54	33072	EAC patient	M	56.53	106.63
23	1	39	43	25017	EAC patient	M	78.97	84.98
24	1	40	44	61043	EAC patient	M	65.25	81.84
25	3	59	63, 68	25011	EAC patient	M	65.97	67.40
26	3	60	64	61040	EAC patient	M	75.77	95.19
27	2	51	55	21139	EAC patient	M	69.74	97.71
28	1	41	45	21113	EAC patient	M	60.35	78.35
29	3	61	65	40259	EAC patient	M	66.78	70.72
30	1	42	46	21233	EAC patient	M	54.58	82.24

Table 3.2. Clinical characteristics of the patient cohort for biomarker discovery.

Variables	Healthy	BE	EAC	*P* value (Healthy vs BE vs EAC)
Sample size	9	10	10	
Age (Median ± SD)	66 ± 10	62 ± 15	66 ± 8	0.9311

Variables	Healthy	BE	EAC	*P* value (Healthy vs BE vs EAC)
Gender	All male	All male	All male	
Protein concentration (µg/µL)	95 ± 19	85 ± 13	82 ± 13	0.3641
Gastritis*	1 (11.1%)	1 (11.1%)	1 (10.0%)	1.0000
Peptic ulcer	3 (33.3%)	2 (20.0%)	3 (30.0%)	0.8792
Hiatus hernia	0 (0.0%)	4 (40.0%)	6 (60.0%)	0.0217
Other malignancy	1 (11.1%)	2 (20.0%)	2 (20.0%)	1.0000

*All the analyses were performed based on available patient information. Gastritis status for one BE patient was missing.

3.2.2 LeMBA

LeMBA was performed as described in section 2.5. Figure 3.1 below display typical layout of a plate for LeMBA pull-down and mass spectrometry runs. The samples were run in following order on mass spectrometer: A1 to G1, A2 to G2, ..., and A12 to G12.

	1	2	3	4	5	6	7	8	9	10	11	12
	Sample 1			**Sample 2**			**Sample 3**			**Sample 4**		
A	Ctrl	WFA	PSA	Ctrl	WFA	PSA	Ctrl	WFA	PSA	Ctrl	WFA	PSA
B	NPL	MAA	SNA	NPL	MAA	SNA	NPL	MAA	SNA	NPL	MAA	SNA
C	STL	DSA	LPHA	STL	DSA	LPHA	STL	DSA	LPHA	STL	DSA	LPHA
D	UEA	WGA	JAC	UEA	WGA	JAC	UEA	WGA	JAC	UEA	WGA	JAC
E	HAA	SBA	AAL	HAA	SBA	AAL	HAA	SBA	AAL	HAA	SBA	AAL
F	HPA	ECA	EPHA	HPA	ECA	EPHA	HPA	ECA	EPHA	HPA	ECA	EPHA
G	GNL	BPL	ConA	GNL	BPL	ConA	GNL	BPL	ConA	GNL	BPL	ConA
H	Empty row											

Figure 3.1. Typical plate layout for LeMBA pull-down and mass spectrometric run for biomarker discovery.

3.2.3 Nano-HPLC-MS/MS for biomarker discovery

After LeMBA pull-down and on-bead trypsin digestion, the peptide samples were resuspended in 20 µl of 0.1% v/v formic acid for HPLC-MS/MS (Agilent 6520 quadrupole time of flight [QTOF] coupled with a Chip Cube and 1200 HPLC). Optimal volume of sample injection for HPLC-MS/MS analysis was previously optimized: 9 µL were loaded for HAA, HPA and UEA, 6 µL for NPL, STL, GNL, 5 µL for BPL, DSA, ECA, MAA, SBA, WFA, and WGA, 4 µL for AAL, SNA, LPHA, PSA and JAC, 1 µL for EPHA and ConA. The nano pump was set at 0.3 µL/min and the capillary pump at 4 µL/min. The HPLC-chip used contains 160 nl C18 trapping column, and 75

μm x 150 mm 300 Å C18 analytical column (G4240-62010 Agilent Technologies). Buffer A was 0.1% v/v formic acid and Buffer B was 90% v/v acetonitrile containing 0.1% v/v formic acid. Peptides were eluted from the column using a gradient from 6% B to 46% B at 45 min. Nano pump %B was increased to 95 %B at 45.5 min and plateaued till 55.5 min, then decreased to 6% B at 58.5 min. The mass spectrometer was operated in 2 GHz extended dynamic range and programmed to acquire 8 precursor MS1 spectra per second and 4 MS/MS spectra for each MS1 spectra. Dynamic exclusion was applied after 2 MS/MS within 0.25 min. Exclusion for lectin peptides was applied as reported previously (317). The QTOF was tuned and calibrated prior to analysis. One hundred femtomole/μL of pre-digested bovine serum albumin peptides were used as quality control, before and after each plate. Levels of reference ions 299.2945 and 1221.9906 were maintained at minimum 5000 and 1000 counts respectively. Blank injection was run after each sample injection to minimize sample carry over. After running each patient sample on mass spectrometer, long column clean-up was performed. The samples were run on mass spectrometer at three different time points. At each time point, similar number of patients from 3 patient groups were run to avoid possible bias arising from mass spectrometry analysis.

3.2.4 Database search

The raw data was extracted and searched using Spectrum Mill MS proteomics workbench (Agilent Technologies, Rev.B.04.00.127) against Swissprot human database containing total 20,242 entries (release 3rd Jan 2012). Similar MS/MS spectra acquired on the precursor *m/z* within ± 1.4 m/z and within ± 15 sec were merged. The following parameters were used for the search: 2 maximum missed cleavages, minimum matched peak intensity of 50%, precursor mass tolerance of ± 20 ppm, product mass tolerance of ± 50 ppm, calculate reversed database scores enabled and dynamic peak thresholding enabled. Carbamidomethylation was selected as fixed modification and oxidized methionine was selected as a variable modification. Precursor mass shift range from -17.0 Da to 177.0 Da was allowed for variable modification. Results were filtered by protein score > 15, peptide score > 6, and % scored peak intensity (% SPI) > 60. Automatic validation was used to validate proteins and peptides with default settings and false discovery rates (FDRs) were calculated using reversed hits. The same data was searched once again against chicken ovalbumin sequence.

3.2.5 Data normalization using internal standard chicken ovalbumin

To account for experimental variations between pull-downs and during the mass spectrometric analysis, 10 pmol ovalbumin per lectin pull-down was spiked in as an internal standard in the first step at the stage of sample denaturation. The intensities of at least three out of

seven ovalbumin peptides listed in Table 3.3 were used for normalization. Peptides were selected based on their ability to meet the following criteria: (i) peptide sequence should be unique to chicken ovalbumin, (ii) peptides being compared across samples should have the same charge, (iii) peptides being compared should have scores higher than 10 based on Spectrum Mill, and (iv) PrecursorAveragineChiSquared should be higher than 0.85 as suggested by Spectrum Mill. The intensities of ovalbumin peptides were extracted from Spectrum Mill. Lectins for which Spectrum Mill failed to identify at least 3 consistent peptides across all samples, ovalbumin peptides (mass over charge *m/z*) were extracted at MS1 level from the raw data acquisition files using Mass Hunter Qualitative Analysis B.05.00 and manually integrated to obtain abundance values. As intensities given from Spectrum Mill and the Qualitative software varied, only one out of two methods was chosen for each peptide. For each sample, all ovalbumin peptide intensities for all lectins were compiled into a single comma-separated value (.csv) file and uploaded to GlycoSelector. Normalization for each peptide was performed within GlycoSelector. Since each lectin binds to ovalbumin with different affinity, normalization was performed for each lectin separately across all samples. Two different normalization approaches were examined; (i) based upon total protein intensity whereby individual ovalbumin peptide intensity was summed to get the total protein intensity. This total protein intensity was then utilized to calculate the normalization factor or (ii) using individual ovalbumin peptide intensity whereby a normalization factor is calculated individually based on each peptide and averaged to derive the final normalization factor. Both methods gave comparable results, the second approach based on individual peptide intensity was considered for further analysis.

Table 3.3. List of ovalbumin peptides selected for data normalization.

Peptide sequence	**Charge**	***m/z***	**Delta *m/z***	**Lectin Name**
(K)ISQAVHAAHAEINEAGR(E)	3	591.9737	0.007	All lectins
(K)AFKDEDTQAMPFR(V)	2	778.3669	0.008	DSA, GNL, HAA, JAC, MAA, NPL, PSA, SNA, STL, WFA, WGA
(K)LTEWTSSNVMEER(K)	2	791.3667	0.011	AAL, BPL, ConA, ECA, EPHA, GNL, HPA, JAC, MAA, NPL, PSA, SBA, SNA, UEA, WFA, WGA
(R)GGLEPINFQTAADQAR(E)	2	844.4261	0.019	AAL, BPL, ConA, DSA, ECA, EPHA, GNL, HPA, JAC, LPHA, MAA, NPL, PSA, SBA, SNA, STL, WFA, WGA

Peptide sequence	Charge	*m/z*	Delta *m/z*	Lectin Name
(R)NVLQPSSVDSQTAMVLVNAIVFK(G)	3	820.782	0.01	SBA, STL, WGA
(R)VTEQESKPVQMMYQIGLFR(V)	3	762.0587	0.011	EPHA, LPHA, MAA, NPL, PSA, STL, WFA, WGA
(K)ISQAVHAAHAEINEAGR(E)	4	444.2327	0.004	All lectins

3.2.6 GlycoSelector analysis

GlycoSelector (http://glycoselector.di.uq.edu.au), a customized web-based portal to store and analyze multidimensional LeMBA-MS/MS discovery data was developed to select a list of biomarkers for verification phase. The workflow of using GlycoSelector for data analysis is presented in Figure 3.2. GlycoSelector stores patient information including age, gender, unique hospital reference number and categorizes each patient according to their phenotype. For each patient, multiple sample runs could be stored. Each run is given a unique run ID and stored with informative details such as patient phenotype (normal/esophageal etc.), source of sample (tissue/serum/urine), and diagnosis (normal/benign/malignant etc.). This information is used to define patient groups for the downstream comparison/statistical analyses. As an input data file, GlycoSelector stores protein list file obtained through a Spectrum Mill search. This file contains the lectin name, list of proteins identified, along with the total intensity for each protein identified for every individual patient sample. An internal standard file is stored together with protein list file for each sample run. It contains abundance values for individual ovalbumin peptide to be used for data normalization.

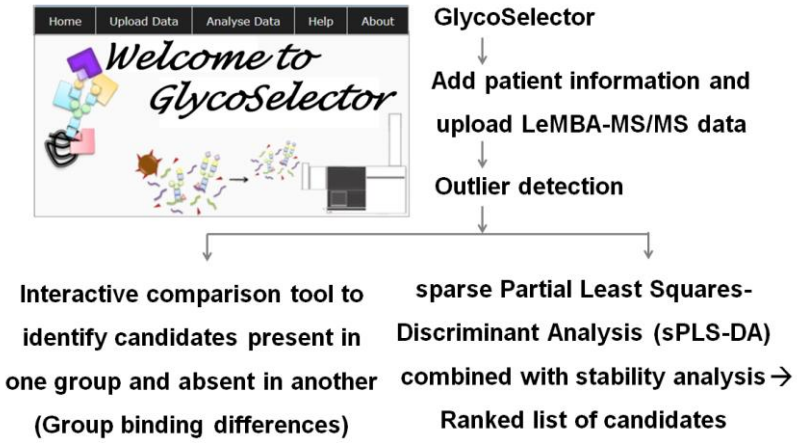


Figure 3.2. Steps followed for data analysis using GlycoSelector (http://glycoselector.di.uq.edu.au) platform for biomarker discovery.

3.2.6.1 Group binding difference analysis

GlycoSelector was programmed with a feature called "group binding difference" to compare proteins identified in one group of patient samples and absent in another group for each lectin pull-down. As the main purpose of the biomarker discovery phase is to identify candidate biomarkers for

future verification, the criteria chosen for group comparisons were not very stringent. Lectin-protein combinations that were present in 60% of one patient group and absent in 40% of another group were selected. A comparison was made between healthy vs BE, BE vs EAC and healthy vs EAC patient cohorts.

An exclusion list of proteins was applied before group binding difference and statistical analysis in GlycoSelector. The exclusion list included common contaminant proteins like keratin, proteins like immunoglobulin and serum albumin which cannot be considered as biomarkers for clinical use hence eliminated from the analysis.

3.2.6.2 Workflow for statistical analysis

The statistical tools for discovery and verification, namely GlycoSelector and Shiny mixOmics follow the same workflow. The first step is outlier detection, which aims to identify and therefore remove samples showing abnormally high or low protein intensities across many lectin pull-downs, thereby preventing any detrimental effect on downstream statistical analysis. Outliers may be due to improper sample handling, technical difficulties during mass spectrometric analysis, or due to batch/plate effect. Graphical outputs using unsupervised approaches, which do not take into account the patient phenotypes, were generated using lectin-protein intensities to visualize any potential outliers. Principal component analysis (PCA) (381) is a multivariate approach which highlights samples contributing to a large variance. Hierarchical clustering (using Euclidian distance and Ward agglomeration method) produces clusters amongst variables and samples that can be visually represented through a dendrogram on the left hand side (lectin-protein intensities) and the top (samples) of the heatmap, with red (green) color indicating low (high) intensities of the lectin-proteins. Boxplots of the scaled data enable to visualize the variability on each sample, and coefficient of variation for each sample across all lectin-proteins were calculated and represented in the form of barplots.

A supervised multivariate approach sparse partial least squares-discriminant analysis (sPLS-DA) (382) was applied to select discriminative lectin-protein candidates able to classify patients into two different phenotypes. The approach outputs a selected ranked list of candidates per component along with sample representation to visualize the patient phenotypes. The method seeks for the best linear combination of lectin-proteins that can classify the sample into their respective groups. The model parameters (number of components in the model and number of variables to select on each component) are tuned using leave-one-out cross-validation (GlycoSelector) or 5-fold cross-validation repeated 1000 times (Shiny mixOmics). A stability measure that records the

frequency of each feature being selected across several data sets generated during the cross-validation procedure. Candidates with high stability value (close to 100%) are robust biomarkers, as they are repeatedly selected across slight perturbations of the original data set. Only the robust lectin-protein candidates which passed an arbitrary cut-off of 70% were reported. Lectin-protein candidates that were not identified in more than 2/3rd of the patient samples in at least one of the two groups being compared using sPLS-DA feature in GlycoSelector were eliminated from the analysis.

3.3 Results

3.3.1 LeMBA-GlycoSelector biomarker discovery pipeline

Prior to this study, Dr. Eunju Choi optimized LeMBA-nano-HPLC-MS/MS incorporating internal standard chicken ovalbumin at the very first step of sample preparation. As illustrated in the LeMBA workflow (Figure 1.6), serum samples were spiked with 10 picomole ovalbumin per lectin pull-down during sample denaturation. Internal standard chicken ovalbumin experimentally showed binding with all 20 lectins used in LeMBA. It was envisaged that total chicken ovalbumin intensity (sum of measured peptide intensity) can be used for normalization. Out of total 7 ovalbumin peptides listed in Table 3.3, anywhere between 3 and 7 peptides per lectin were selected for normalization e.g. 3 peptides for HAA while all 7 peptides for WGA. It was observed that amongst ovalbumin peptides qualified for using as an internal standard, raw peptide intensity varied in the magnitude up to 10 folds. This means that if total chicken ovalbumin intensity (sum of measured peptide intensity) is used for normalization then not all the peptides are given equal importance. In order to give equal weight to all the qualifying chicken ovalbumin peptide for purpose of internal standard normalization, second approach based on calculation of normalization factor using individual peptide was tried. As illustrated in Figure 3.3, normalization factor calculated by both approaches (i) based on total ovalbumin intensity and (ii) based on individual ovalbumin peptide intensity showed high correlation. The second approach based on individual peptide intensity was considered for further analysis as it gave equal weigh to individual peptides.

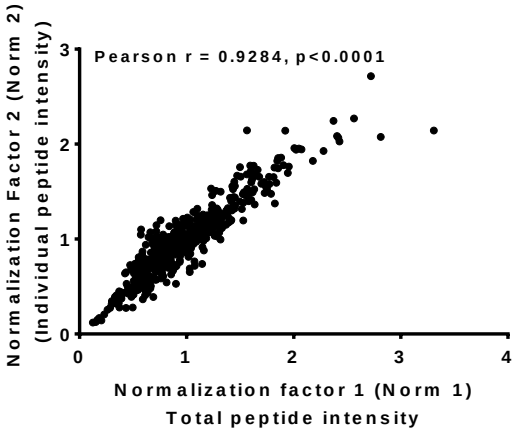


Figure 3.3. Comparison of two normalization methods for biomarker discovery screen data. Normalization method 1 (Norm1) used the total protein intensity of ovalbumin protein (sum of minimum 3 individual peptide intensity) divided by the average total ovalbumin protein intensity of all samples for a particular lectin as normalization factor. Normalization method 2 (Norm 2) divided each ovalbumin peptide intensity with the average

peptide intensity across all samples for each lectin and has taken the average of at least 3 such normalization factors calculated for different ovalbumin peptides to determine final normalization factor.

To facilitate data analysis and biomarker candidate selection using LeMBA-MS/MS data, GlycoSelector database incorporating statistical analysis pipeline was developed. Prior to this thesis, Dr. Kim-Anh Lê Cao and Dr. David Chen with inputs from Dr. Eunju Choi created GlycoSelector which incorporated features like uploading raw LeMBA-MS/MS data including patient information, data normalization, outlier detection, group binding difference analysis and sPLS-DA. It was realized that along with multivariate feature selection using sPLS-DA, internal validation is required. So after several discussions with Dr. Lê Cao, leave-one-out cross-validation was incorporated into the statistical analysis pipeline and new GlycoSelector version was released. The leave-one-out cross-validation is repeated 1000 times by perturbation of original data set by leaving one sample out each time. The frequency of same feature selection across validation process is reported as a stability proportion where 100% (or stability value of 1) indicates selection of biomarker candidates all the time during cross-validation. Arbitrary cut-off of 70% was selected for the data analysis i.e. lectin-protein candidates showing stability proportion of more than 70% were selected for verification. On contrary to older version which used total ovalbumin intensity for data normalization, new GlycoSelector version uses normalization approach based on individual ovalbumin peptide intensity. Figure 3.2 illustrates steps followed to select lectin-protein biomarker candidates using GlycoSelector analysis.

3.3.2 BE/EAC biomarker discovery

After inclusion of new features into GlycoSelector, serum samples from 29 patient samples were screened using LeMBA-GlycoSelector pipeline. Clinical characteristics of the patient cohort for biomarker discovery are described in Table 3.2. All samples were age matched and collected from male patients. BE and EAC patient groups had significantly higher proportion of patients with hiatus hernia compared to the healthy group, as has previously been reported (383), suggesting hiatus hernia to be a risk factor for BE/EAC. A total of 195 unique proteins were identified from the LeMBA-MS/MS screen. The glycoproteins bound several lectins suggesting heterogeneity and multiplicity of glycosylation. On average, 40 proteins per lectin pull-down were identified. Total number of proteins identified per lectin pull-down varied reflecting specificity of lectin-glycan interactions. HAA showed least binding (average number of proteins identified = 12) as compared to WGA which bound to maximum number of proteins (average number of proteins identified =

59). There was no difference between total number of proteins identified between healthy, BE and EAC patient groups (Figure 3.4).

3.3.3 Glycoselector analysis

The results of main three steps followed using GlycoSelector data analysis platform are mentioned below.

3.3.3.1 Outlier detection

The first step in GlycoSelector analysis is outlier detection using four different graphical visualization tools namely PCA, boxplots, hierarchical clustering, and barplots for coefficient of variation. Outlier may be due to improper sample collection and handling, differential sample preparation and LeMBA pull-down, technical difficulties during mass spectrometric analysis, or due to batch/plate effect. The main purpose of this analysis is to identify potentially an outlier sample run and to prevent its detrimental consequences on final statistical analysis.

When the outlier detection analysis was performed on biomarker discovery data set, sample run ID 63 was considered to be an outlier due to consistent anomalous results in all 4 graphical outputs (Figure 3.5A to 3.5D). This result was coincided with lower than cut-off scores and coverage for bovine serum albumin pre-digested quality control (QC) sample ran on mass spectrometer after the sample was run. The mass spectrometer was re-calibrated resulting into scores and coverage for QC above the threshold. The sample was analyzed again (run ID 68), resulting in no outliers being identified (Figure 3.5E to 3.5H).

3.3.3.2 Group binding difference

A group binding difference feature identifies lectin-protein candidates present in one patient group and absent in the other. This feature was particularly useful for mass spectrometry data with many zeroes which either indicate a true absence of a protein in a sample, or the concentration was below the detection limit of the mass spectrometer. Lectin-protein candidates present in more than 60% of one sample group but less than 40% of the other group and vice versa were identified. The analyses were performed comparing healthy vs BE, BE vs EAC and healthy vs EAC patient groups. The reason for choosing non-stringent cut-off for group binding difference analysis was to select as many candidates as possible for verification stage. Table 3.4 contains the list of biomarker candidates identified using group binding difference analysis. Total 37 unique lectin protein candidates were identified by this analysis. Except HAA, all 19 lectins appeared in the list and showed differential binding with one or more of 26 unique protein candidates.

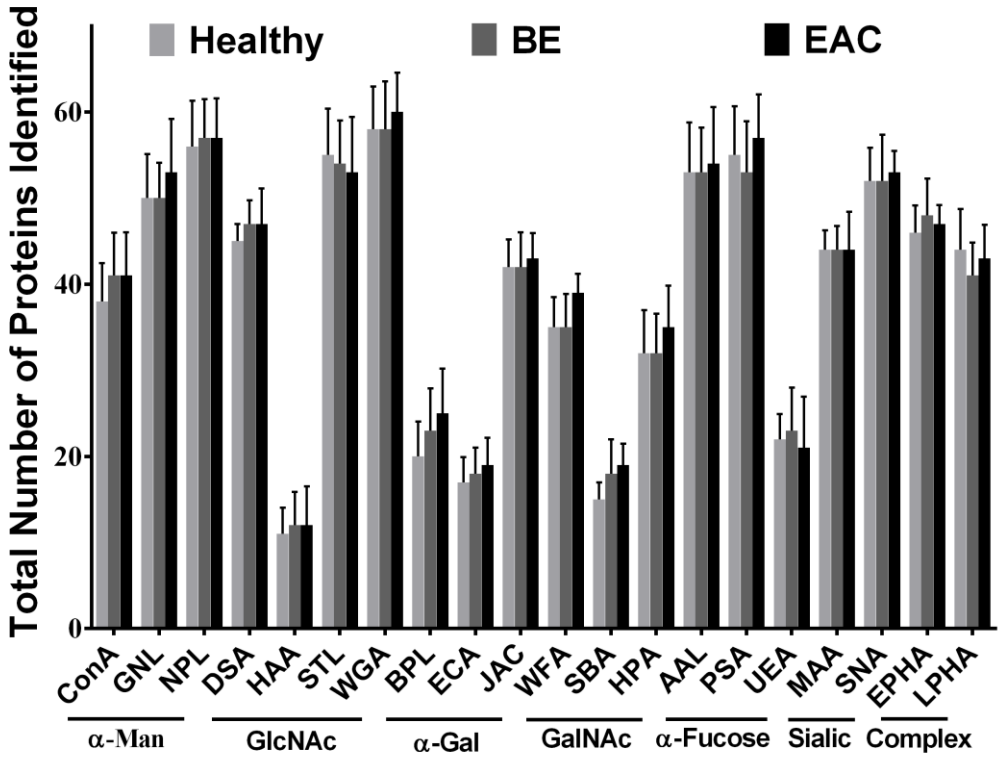


Figure 3.4. Total number of proteins identified per individual lectin pull-down for each patient group for biomarker discovery screen. There was no statistical difference between the number of proteins identified between different phenotypes. The bar graph represents average (±SD) of total number of proteins identified for a particular lectin pull-down across patient samples.

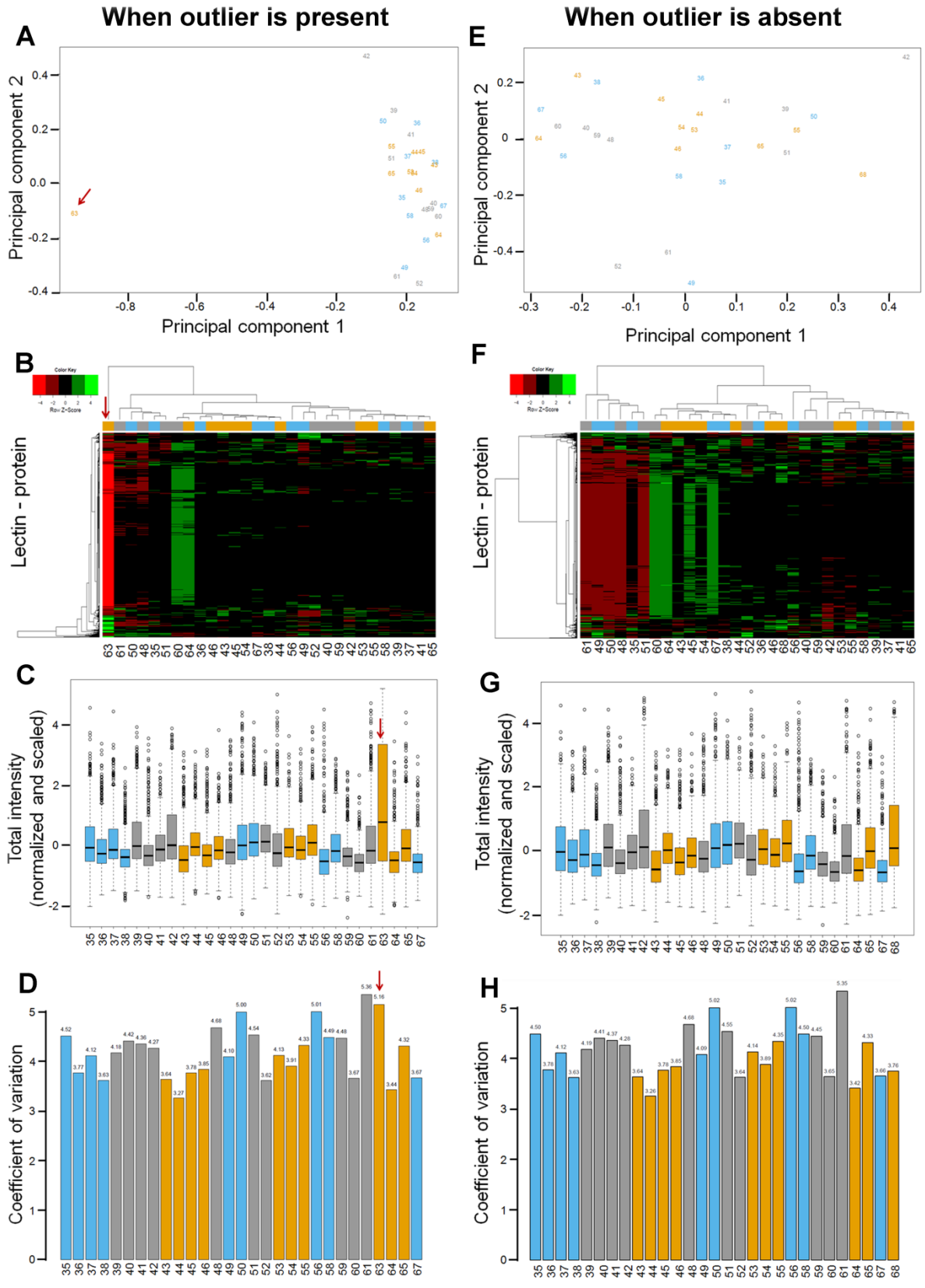


Figure 3.5. Outlier detection feature of GlycoSelector allows the visualization of experimental errors using four different statistical tools. (A and E) Principal component analysis, (B and F) hierarchical clustering, (C and G) boxplots and (D and H) barplots representing the coefficient of variation. Unique numbers on the graph indicate the individual sample run identifier. Run number

63 (red arrow) in panel A to D was flagged as an outlier based on the visualization tools. The sample was re-analyzed on the mass spectrometer and outlier detection was performed again (panel E to H).

Table 3.4. List of lectin-protein candidates identified using group binding difference analysis. The proteins are denoted using Uniprot accession numbers.

Healthy vs BE	BE vs EAC		Healthy vs EAC	
AAL_P10909	AAL_P06396	GNL_P10643	AAL_P06396	EPHA_P02748
AAL_P02747	AAL_O75636	HPA_P01042	JAC_P06396	GNL_P02746
LPHA_P02774	GNL_P06396	LPHA_P01031	LPHA_P05090	HPA_P01042
STL_P08519	JAC_P06396	LPHA_P02748	LPHA_P02774	HPA_P00450
STL_O75636	JAC_P00748	SNA_P02748	PSA_P06396	HPA_P00747
STL_P02765	LPHA_P05090	STL_P08519	STL_P06396	HPA_P00751
AAL_O75636	PSA_P06396		STL_O75636	LPHA_P01031
BPL_P0C0L5	SNA_P08697		UEA_P19823	LPHA_P02748
EPHA_P02748	STL_P06396		WGA_P02746	MAA_P02748
HPA_P00751	AAL_P02747		BPL_P0C0L5	NPL_P01008
JAC_P00748	AAL_P10909		ConA_P02760	SNA_P02748
NPL_P01008	ConA_P02760		DSA_P02748	WFA_P05546
SBA_P04003	ECA_P00450		DSA_P04217	
SNA_P02743	GNL_P02746		ECA_P00450	

3.3.3.3 Statistical analysis

sPLS-DA combined with stability analysis based on leave-one-out validation was employed for multivariate feature selection. The methodology is unable to handle the dataset including many zero values. Hence lectin-protein candidates that were not identified in more than 2/3rd of the patient samples in at least one of the two groups being compared were eliminated from sPLS-DA statistical analysis as they were already taken into consideration for group binding difference analysis. sPLS-DA combined with stability analysis was performed to identify candidates that differentiate BE from healthy, EAC from BE and EAC from healthy phenotype. As illustrated in the sPLS-DA sample representation in Figure 3.6, top 100 lectin-protein candidates in the model, showed distinct clusters of samples according to their phenotype. To select the most consistent candidates across patients for taking to the second verification stage, stability analysis was employed, which utilizes a leave-one-out strategy to assess the robustness of each candidate biomarker. A relatively non-stringent cut-off of 70% was chosen for this purpose. Out of the top 100 lectin-protein pairs, 57 candidates passed the stability cut-off of 70% between healthy vs BE, 72 candidates passed for BE

vs EAC, and 76 candidates, passed for healthy vs EAC analysis (Figure 3.6). List of candidates identified using sPLS-DA along with stability values are presented in Table 3.5.

The summary of biomarker discovery screen results using GlycoSelector analysis is presented in Figure 3.7. Using sPLS-DA/stability analysis and group binding difference feature, the discovery screen identified 54 serum proteins with differential binding to one or more lectins between healthy, BE and EAC serum samples, resulting in a total of 183 unique lectin-protein combinations. Candidates identified using sPLS-DA and the group binding differences feature were complementary and showed no overlap between lectin-protein candidates, justifying the use of two different approaches for candidate selection. Each of the 20 lectins used in the biomarker discovery phase showed differential binding with at least one candidate protein glycoform, endorsing the use of multiple lectins for biomarker discovery (Figure 3.7A). There was considerable overlap between lectin protein candidates identified between healthy vs BE, BE vs EAC and healthy vs EAC patient groups (Figure 3.7B).

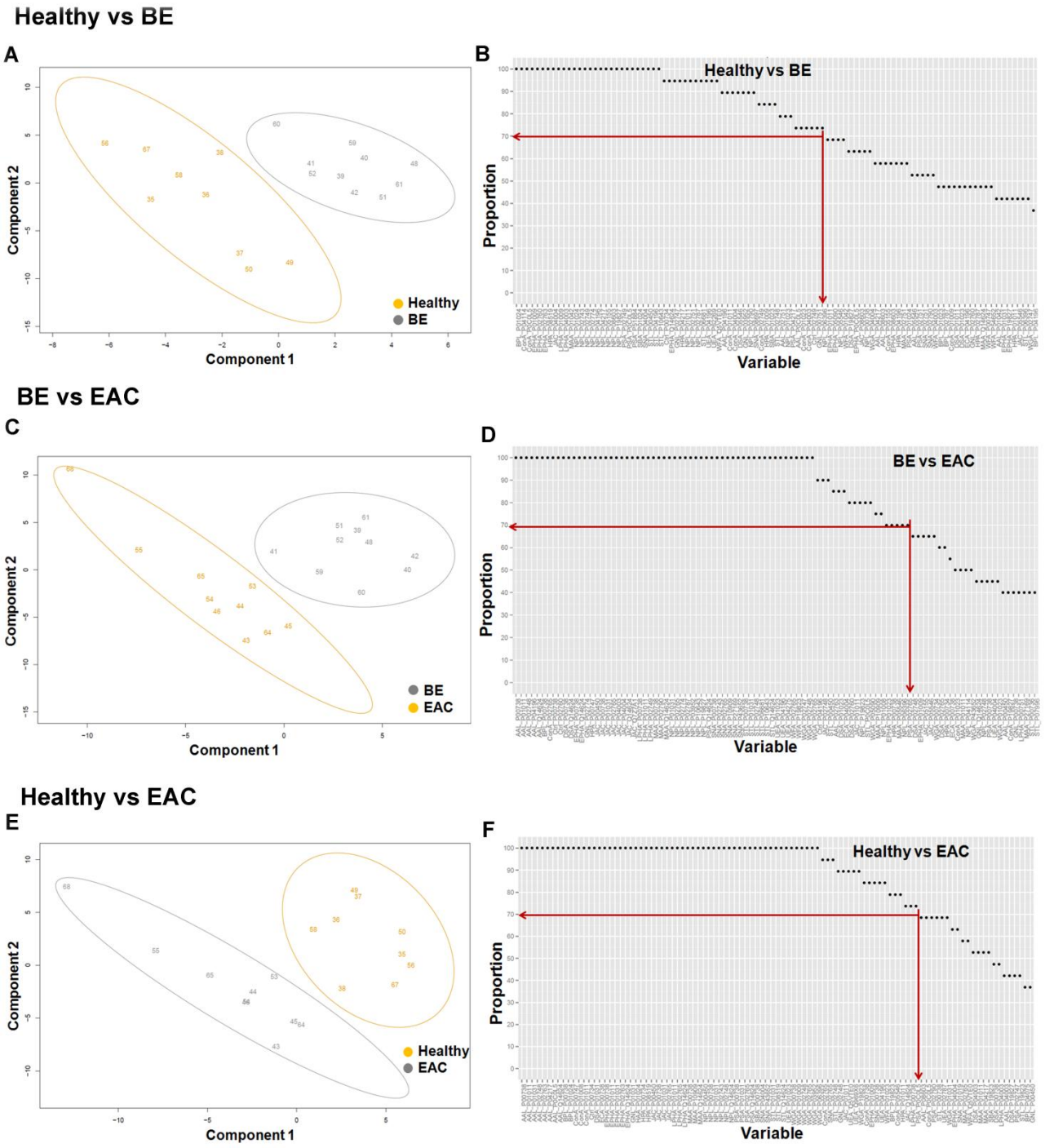


Figure 3.6. Results for sPLS-DA combined with stability analysis. (A), **(C)** and **(E)** Represent sPLS-DA plot for top 100 lectin-protein candidates. sPLS-DA differentiated **(A)** BE from healthy, **(C)** EAC from BE and **(E)** EAC from healthy phenotypes. Amongst these top 100 ranked lectin-protein combinations identified using sPLS-DA, **(B)** 57 candidates for healthy vs BE, **(D)** 72 candidates for BE vs EAC and **(F)** 76 candidates for healthy vs EAC passed the stability cut-off of 70% based on leave-one-out cross-validation.

Table 3.5. Ranked list of lectin-protein candidates identified using sPLS-DA along with stability values. Proteins are represented using Uniprot accession numbers.

Healthy vs BE			BE vs EAC			Healthy vs EAC		
sPLS-DA rank	Candidate name	Stability	sPLS-DA rank	Candidate name	Stability	sPLS-DA rank	Candidate name	Stability
1	BPL_P01024	100.0%	1	STL_P02748	100.0%	1	EPHA_Q14624	100.0%
2	NPL_P05155	100.0%	2	JAC_P00450	100.0%	2	MAA_Q14624	100.0%
3	NPL_P19827	100.0%	3	PSA_Q14624	100.0%	3	EPHA_P00738	100.0%
4	STL_P00450	100.0%	4	EPHA_Q14624	100.0%	4	JAC_P01031	100.0%
5	EPHA_P01009	100.0%	5	WGA_P02765	100.0%	5	AAL_Q14624	100.0%
6	STL_P04196	100.0%	6	AAL_Q14624	100.0%	6	WGA_P02765	100.0%
7	JAC_P04004	100.0%	7	EPHA_P00738	100.0%	7	STL_P01031	100.0%
8	PSA_P51884	100.0%	8	STL_P06681	100.0%	8	AAL_P02748	100.0%
9	NPL_P01024	100.0%	9	HPA_P00747	100.0%	9	LPHA_P02765	100.0%
10	NPL_P02774	100.0%	10	SNA_P02765	100.0%	10	SBA_P00738	100.0%
11	ConA_P04114	100.0%	12	JAC_P00738	100.0%	11	STL_Q14624	100.0%
12	NPL_P04217	100.0%	13	NPL_P02774	100.0%	12	STL_P08519	100.0%
13	MAA_P01042	100.0%	14	NPL_P01031	100.0%	13	HPA_P08519	100.0%
14	LPHA_P01009	100.0%	15	MAA_P00738	100.0%	14	HPA_P00734	100.0%
15	STL_P10643	100.0%	16	SNA_P27169	100.0%	15	AAL_P01031	100.0%
16	SNA_P04004	100.0%	17	BPL_P00738	100.0%	16	WGA_P03952	100.0%
17	EPHA_P02760	100.0%	18	STL_P01031	100.0%	17	BPL_P00738	100.0%
18	PSA_P0C0L5	100.0%	19	LPHA_P02749	100.0%	18	UEA_P01023	100.0%
19	HPA_P08519	100.0%	20	AAL_P02748	100.0%	19	PSA_Q14624	100.0%
20	NPL_P08603	100.0%	21	WGA_P02747	100.0%	20	NPL_P01031	100.0%
21	EPHA_P02763	100.0%	22	WFA_P19823	100.0%	21	JAC_P00738	100.0%
22	NPL_P02743	100.0%	23	LPHA_P00738	100.0%	22	PSA_P01031	100.0%
23	LPHA_P04003	100.0%	24	JAC_P04003	100.0%	23	HAA_P01024	100.0%
24	ConA_P0C0L5	100.0%	25	NPL_P19827	100.0%	25	SNA_P43652	100.0%
25	WFA_Q5VTE0	94.7%	26	UEA_P01023	100.0%	26	LPHA_P01011	100.0%
26	NPL_P02749	100.0%	27	NPL_P10643	100.0%	27	WGA_P02748	100.0%
27	PSA_P02749	100.0%	28	UEA_P04196	100.0%	28	JAC_P00450	100.0%
28	UEA_P62269	94.7%	29	JAC_P04004	100.0%	29	AAL_P02763	100.0%
29	SNA_P08697	89.5%	30	JAC_P02765	100.0%	30	NPL_P02743	100.0%
30	UEA_P04196	94.7%	31	LPHA_P01011	100.0%	31	AAL_P00738	100.0%
31	NPL_P04196	100.0%	32	MAA_Q14624	100.0%	32	ECA_P00738	100.0%
32	EPHA_Q14624	94.7%	33	NPL_P03952	100.0%	33	BPL_P01024	100.0%
33	NPL_P02751	94.7%	34	WGA_P02748	100.0%	34	NPL_P02748	100.0%
34	GNL_P03952	89.5%	35	SNA_P02743	100.0%	35	WGA_P06396	100.0%
35	GNL_P02747	94.7%	36	SNA_P00738	100.0%	36	JAC_P02765	100.0%
36	HPA_P04217	94.7%	37	NPL_P02765	100.0%	37	MAA_P10909	100.0%
37	SBA_P01024	100.0%	38	JAC_Q14624	100.0%	38	MAA_P00738	100.0%
38	NPL_P02787	94.7%	39	STL_Q14624	100.0%	39	ConA_P01008	100.0%
40	NPL_P02790	89.5%	40	GNL_P02774	100.0%	40	SNA_P04004	100.0%
41	NPL_P01011	94.7%	41	JAC_P01031	100.0%	41	NPL_P00450	100.0%
42	STL_P06681	94.7%	42	JAC_Q7Z7A1	100.0%	42	NPL_Q14624	100.0%
43	AAL_P04196	89.5%	43	STL_P10643	100.0%	43	DSA_P01031	100.0%
44	HPA_P01009	84.2%	44	STL_P00738	100.0%	44	AAL_P0C0L5	100.0%
45	DSA_P05546	89.5%	45	DSA_P00738	100.0%	45	GNL_P01031	100.0%
46	ConA_P04004	89.5%	46	ConA_P02765	100.0%	46	LPHA_Q14624	100.0%
47	STL_P02748	84.2%	47	UEA_P04275	100.0%	47	STL_P02748	94.7%
49	PSA_P04217	79.0%	48	NPL_P02787	100.0%	48	ConA_P01011	100.0%

Healthy vs BE			BE vs EAC			Healthy vs EAC		
sPLS-DA rank	Candidate name	Stability	sPLS-DA rank	Candidate name	Stability	sPLS-DA rank	Candidate name	Stability
50	ConA_P01023	73.7%	49	AAL_P01011	100.0%	49	UEA_Q969X6	89.5%
51	AAL_P02751	79.0%	50	AAL_P00738	100.0%	50	AAL_P01011	100.0%
52	AAL_P0C0L5	73.7%	51	AAL_P04196	100.0%	51	EPHA_P01011	100.0%
53	ConA_P04003	73.7%	52	SNA_P10909	100.0%	52	WGA_P19823	89.5%
54	ConA_P01011	89.5%	53	DSA_Q14624	100.0%	53	SNA_P02765	94.7%
55	SBA_P01023	84.2%	54	MAA_P02790	100.0%	54	ConA_P02765	94.7%
60	NPL_P01023	79.0%	55	STL_P00450	90.0%	55	PSA_P00738	100.0%
61	GNL_P00751	73.7%	56	WFA_P02765	100.0%	56	PSA_P02765	100.0%
65	ConA_P02749	84.2%	57	SNA_P43652	100.0%	57	EPHA_P02763	100.0%
66	NPL_P06396	73.7%	60	JAC_P02748	90.0%	58	WGA_P00738	100.0%
			61	AAL_P02763	85.0%	59	WGA_P01009	100.0%
			62	DSA_P01023	85.0%	60	EPHA_P01031	100.0%
			63	ECA_P04004	85.0%	61	ConA_P00747	100.0%
			64	NPL_P01008	75.0%	62	AAL_P04217	100.0%
			65	DSA_P02760	80.0%	63	WFA_P01023	84.2%
			66	NPL_P19823	80.0%	64	UEA_P04196	84.2%
			67	WGA_P05090	80.0%	65	UEA_Q5VTE0	89.5%
			69	EPHA_P01023	70.0%	66	NPL_P00738	100.0%
			70	PSA_P02765	70.0%	67	EPHA_P01009	84.2%
			73	MAA_P05546	70.0%	69	PSA_P0C0L5	73.7%
			74	STL_P0C0L5	80.0%	70	BPL_P19823	79.0%
			76	MAA_P10909	75.0%	71	ConA_P00738	84.2%
			77	NPL_P06396	70.0%	72	GNL_P02748	89.5%
			79	JAC_P01011	80.0%	73	SNA_P00738	84.2%
			81	HPA_P00738	70.0%	75	LPHA_P00738	73.7%
						76	ConA_P05546	79.0%
						79	JAC_Q14624	73.7%
						84	JAC_P01011	89.5%
						86	HPA_P01011	79.0%

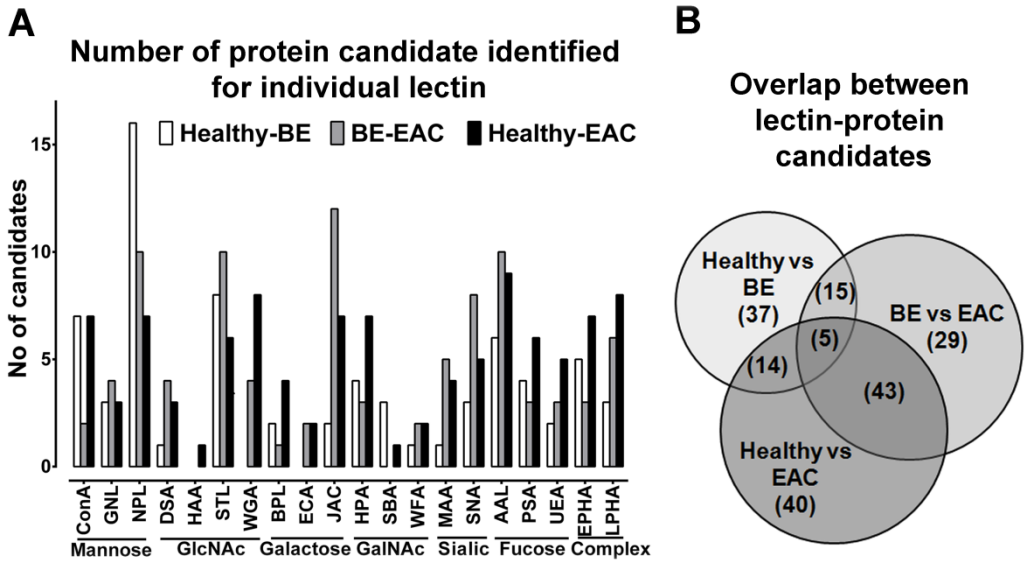


Figure 3.7. Summary of biomarker discovery results. (A) Number of unique candidate proteins identified for each lectin in LeMBA-GlycoSelector analysis. All 20 lectins used for screening identified at least one protein candidate. **(B)** Overlap between lectin-protein candidates that differentiate BE from healthy, EAC from BE, and EAC from healthy phenotype.

3.3.4 LeMBA-western immunoblotting validation for top two candidates

Immunoblotting was used for orthogonal protein level confirmation of the LeMBA-MS/MS screen. Two protein candidates which showed altered binding to *Aleuria aurantia* lectin (AAL) and had antibodies available were chosen for protein level validation. AAL-haptoglobin (HP; Uniprot entry: P00738) was one of the top ranked candidates in sPLS-DA analysis for healthy vs EAC and BE vs EAC, while AAL-gelsolin (GSN; Uniprot entry: P06396) was identified using the group binding difference feature of GlycoSelector as on-off change between BE vs EAC and healthy vs EAC. Using the same set of discovery serum samples, LeMBA pull-down using AAL was performed. Haptoglobin and gelsolin binding was measured by immunoblotting in AAL pull-down. A control serum sample was loaded on every blot as a normalizer between membranes. LeMBA-immunoblotting confirmed the MS/MS results (Figure 3.8), and showed higher sensitivity as it detected low levels of gelsolin in all patient samples, when some were undetectable by MS/MS [AAL-HP: label-free proteomics *P* value = 0.0868, western immunoblotting *P* value = 0.0267; AAL-GSN: label-free proteomics *P* value = 0.0254, western immunoblotting *P* value = 0.0019].

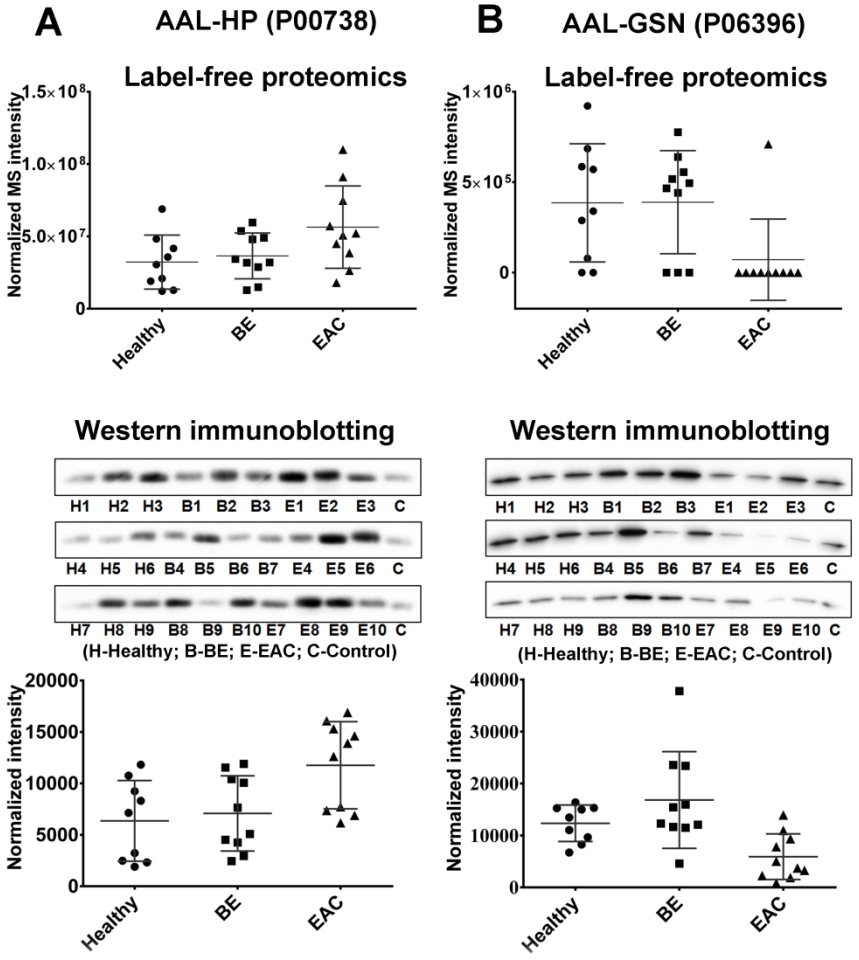


Figure 3.8. Protein level validation for top two candidates using LeMBA-western immunoblotting. **(A)** AAL-HP and **(B)** AAL-GSN were the top two candidates identified using sPLS-DA and group binding difference tool, respectively. **(A and B, top panel)** Label-free proteomics relative quantitation results for AAL-HP and AAL-GSN respectively. **(A and B, lower panel)** Normalized intensity for AAL-HP and AAL-GSN using immunoblotting.

3.3.5 Lectin binding profile of candidates under different scenarios

Changes in the total glycoprotein candidate levels and/or a subtle change in the glycan structure may directly/indirectly induce a change in the binding pattern of lectins and create a lectin binding profile for each glycoprotein biomarker candidate. Figure 3.9 illustrates lectin binding profile of three such candidates namely haptoglobin (Figure 3.9A), gelsolin (Figure 3.9B) and α-2-macroglobulin (A2M; Uniprot entry: P01023) (Figure 3.9C). For simplicity, only 8 out of total 20 lectins are mentioned in the Figure 3.9. Haptoglobin was identified by sPLS-DA combined with stability analysis and showed differential binding with multiple lectins for BE vs EAC (EPHA, JAC, MAA, BPL, LPHA, SNA, STL, DSA, AAL, and HPA) and healthy vs EAC (EPHA, SBA, BPL, JAC, AAL, ECA, MAA, PSA, WGA, NPL, ConA, SNA, and LPHA) comparisons (Figure 3.9A). For all the lectin pull-downs, as compared to healthy and BE patient groups, EAC samples showed higher levels of haptoglobin. This may be due to changes in the total protein levels of haptoglobin. Gelsolin was identified using the group binding difference analysis. It showed differential lectin binding for BE vs EAC (AAL, GNL, JAC, PSA, and STL) and healthy vs EAC (AAL, JAC, PSA, and STL) analysis (Figure 3.9B). Unlike haptoglobin which showed increased lectin binding in EAC samples, gelsolin showed reduced binding. In fact, gelsolin was not identified during the mass spectrometric analysis specifically in EAC patient sample lectin pull-downs which suggest that gelsolin was either absent in the samples or more likely it was present just below the detection limit of the mass spectrometer. Like haptoglobin, α-2-macroglobulin was identified using sPLS-DA combined with stability analysis. It showed differential binding with multiple lectins for all three, healthy vs BE (ConA, SBA, and NPL), BE vs EAC (UEA, DSA, and EPHA) and healthy vs EAC (UEA and WFA) comparisons (Figure 3.9C). The lectins with which it showed differential binding are different between all three comparisons suggesting progressive changes in glycosylation from healthy-BE-EAC development. A2M showed no difference between healthy, BE and EAC patient groups for AAL, JAC, and MAA lectin pull-downs. DSA-A2M was statistically significantly different between BE and EAC patient groups while UEA-A2M was significantly different in EAC patients as compared to healthy and BE patients. For the remaining lectin pull-downs using EPHA, NPL and WFA, A2M was different according to sPLS-DA analysis but could not achieve statistical significance. It is noteworthy that although UEA and AAL lectin bind to similar glycan structure i.e. fucose, only UEA showed statistically significant differential binding with A2M while AAL showed no difference suggesting differential specificity of the lectins from the similar class to recognize the glycan structures.

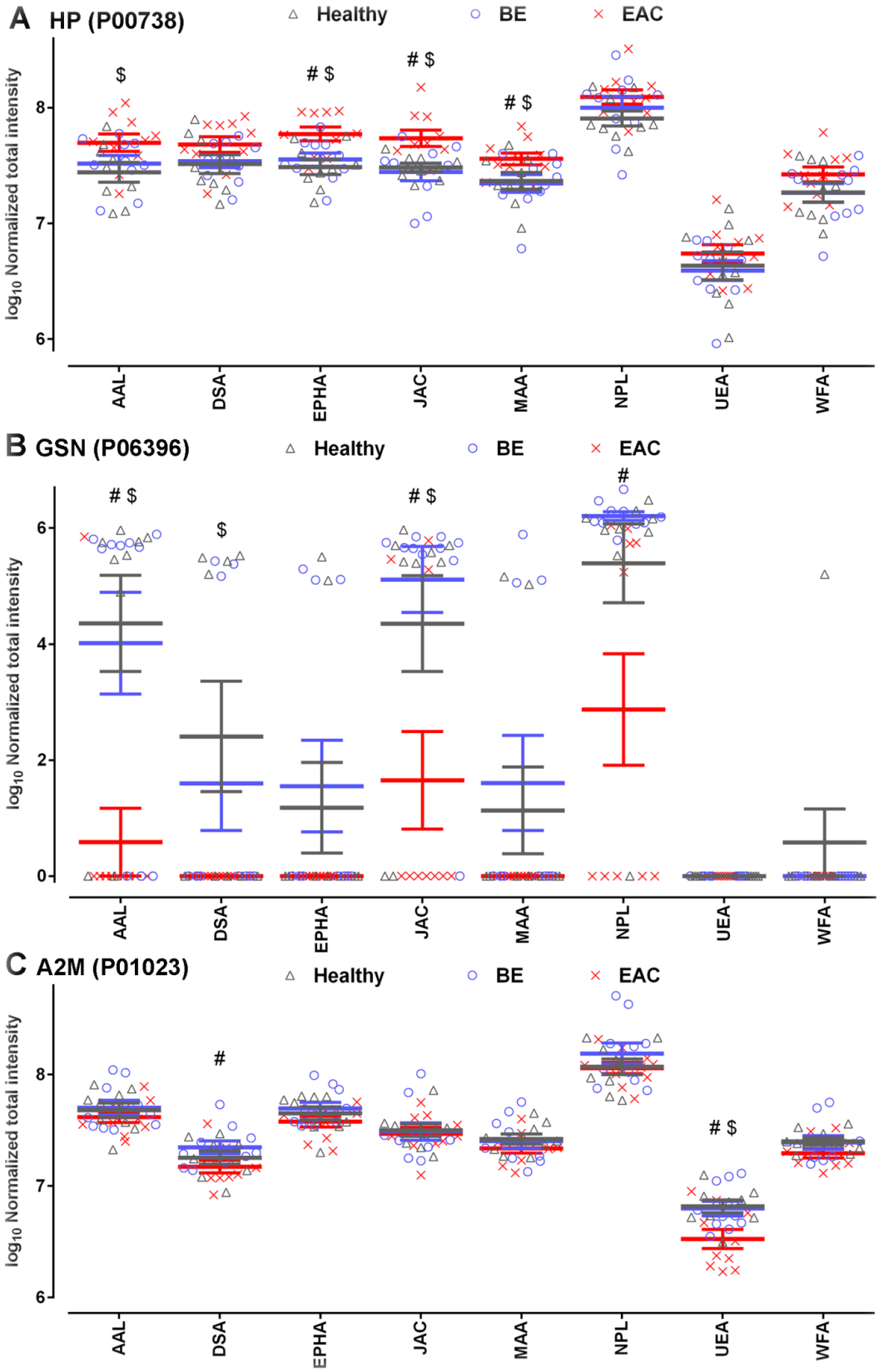


Figure 3.9. Lectin binding signature of putative glycoprotein biomarker candidates. Log transformed normalized total intensity from LeMBA-GlycoSelector biomarker discovery screen for three candidates **(A)** haptoglobin, **(B)** gelsolin, and **(C)** α-2-macroglobulin is plotted against 8

different lectin pull-downs for all 29 samples belong to healthy, BE and EAC patient groups ($^{\#}P < 0.05$ for BE vs EAC comparison; $^{\$}P < 0.05$ for healthy vs EAC comparison).

3.4 Discussion

3.4.1 Overview

This chapter describes identification of serum glycoproteins that show differential binding with one or more lectins between healthy, BE and EAC patient groups. Serum samples from healthy, BE and EAC patient samples were processed using LeMBA-LC-MS/MS workflow. GlycoSelector was upgraded with new normalization method and statistical analysis pipeline was modified to include internal validation in the form of stability analysis. GlycoSelector analysis identified total 183 unique lectin-protein pairs as potential diagnostic biomarker candidates. Mass spectrometric quantitation at peptide level was validated for top two candidates at protein level using western-immunoblotting.

3.4.2 LeMBA-GlycoSelector biomarker discovery pipeline

Label-free quantitation was chosen for quantifying LeMBA results (317). Out of two main label-free approaches, (i) protein-based methods rely upon spectral count and related indices, and (ii) peptide-based methods rely upon ion intensities and protein correlation profiling (384), the latter was chosen for LeMBA workflow (317). Protein-based quantitation approaches using spectral count can be universally applied for diverse proteomics data sets. The main limitation of this approach is its limited applicability to relatively quantify proteins with few numbers of spectra. This situation arises in following two scenarios. (i) Very low abundance proteins yield only a few spectra, and (ii) Low molecular weight proteins result in fewer tryptic peptides hence only a few spectra (384, 385). On the contrary, peptide-based methods such as total ion intensities can cover extended dynamic range of quantitation (385). Peptide-based methods use MS/MS spectra from the discovery proteomics experiments to assign/identify a peptide sequences. The height or volume of a peak for each peptide ion is calculated by extracting a given *m/z* from corresponding MS spectrums and used for quantitation (384). Using this peptide-based intensities, Spectrum Mill search engine calculates mean and total protein intensities for each protein species. Earlier work by Dr. Choi and Dr. Lê Cao concluded that total protein intensity offers better quantitation over mean protein intensity (Dr. Eunju Choi, PhD thesis).

GlycoSelector data analysis platform uses sPLS-DA for multivariate analysis and to rank lectin-protein biomarker candidates. sPLS-DA cannot perform at its best when there are lots of zero values (when missing values are considered to be 0) which is the case for LeMBA-LC-MS/MS data. So during sPLS-DA analysis, cut-off was set to remove variables with many zero values. The removed variables with many zero values were separately analyzed using group binding difference

feature. Biomarker exploratory studies such as the one described in the present chapter analyze large number of variables in limited number of patient samples (133, 356). Considering its design, biomarker discovery studies are prone to false discovery due to low sample size and large number of variables analyzed. To keep the false discovery rate to minimum and to find out stable variables, leave-one-out cross-validation (386) was employed during sPLS-DA.

The result described in this chapter demonstrates the successful application of LeMBA-GlycoSelector pipeline to discover diagnostic biomarker candidates for BE/EAC. Firstly, the assay is robust with % CV for the internal standard ovalbumin peptides below or around 40% for all the lectin pull-downs. Similar levels of variations have been observed by Plavina and colleagues in the multi-lectin chromatography platform they developed for plasma protein biomarker discovery (387). Compared to other established assays, the variation for internal standard appears to be high mainly because mass spectrometry technique employed was semi-quantitative. Secondly, the assay was consistent and reproducible, as demonstrated by similar number of protein candidates identified for each lectin pull-down across different patient samples. Thirdly, the outlier detection analysis successfully identified the outlier present in the dataset due to technical problem during the mass spectrometric run. Collectively, these three steps/results ensure that data acquired is of good enough quality for subsequent statistical analysis.

The tools incorporated in the GlycoSelector aimed to identify as many candidates as possible hence the criteria chosen to select a list of biomarkers were lenient. Out of total 54 unique glycoprotein candidates, 3 were identified by group binding difference analysis only, 28 identified by sPLS-DA analysis only and 23 by both sPLS-DA and group binding difference analysis. The complementary results justify the use of two methods for biomarker identification.

3.4.3 Lectin binding signature of biomarker candidate

The glycoprotein biomarker candidates showed differential binding with one or more lectins under one of three possible scenarios. (i) Total glycoprotein level changes in the serum lead to overall increased/decreased binding with multiple lectins e.g. haptoglobin (Figure 3.9A). (ii) Changes in the glycan occupancy at a particular glycosylation site lead to differential binding with multiple lectins e.g. gelsolin (Figure 3.9B). (iii) Differential expression of a specific glycan structure altered binding of a glycoprotein to a particular lectin or a group of lectins (Figure 3.9C). Proof-of-concept LeMBA work published earlier suggested changes in the binding of glycoproteins with multiple lectins by neuraminidase treatment (317). Neuraminidase is an enzyme which cleaves the glycosidic linkages involving sialic acid residues. As expected, neuraminidase treatment led to

reduced binding of serum proteins with sialic acid binding lectins namely SNA and MAA. In addition, neuraminidase treatment led to considerable increase in binding of serum proteins with mannose, fucose and complex glycan structures binding lectins (317). Based on this it was proposed that disease related glycosylation changes will result in a differential binding of candidate glycoprotein with multiple lectins (lectin signature) due to impaired three-dimensional structure (317). Although many glycoprotein candidates identified in the biomarker discovery showed differential binding with multiple lectins, none of the candidates showed similar changes observed in the proof-of-concept work i.e. none of the candidate identified showed increased binding with one lectin and reduced binding with other lectins when comparing two different phenotypes. This suggests that the phenomena observed in the proof-of-concept work requires dramatic changes in the glycosylation, possible only with controlled experiments such as neuraminidase treatment as compared to subtle changes observed in the lectin binding between healthy, BE and EAC phenotypes.

Haptoglobin is a positive acute-phase hemoglobin scavenging protein primarily produced by liver in the body. It is a heterotetramer and consists of two α and two β chains. Haptoglobin binds free hemoglobin (388). The haptoglobin-hemoglobin complex is rapidly cleared by monocytes and macrophages via CD163 receptors present on their cell surface (389). The β Chain of haptoglobin harbor four sites for N-linked glycosylation and mainly express complex type glycans (390). Aberrant fucose and sialic acid expressing haptoglobin in serum has been demonstrated as potential biomarker for various cancers such as colon cancer (391), hepatocellular carcinoma (392), prostate cancer (393), and pancreatic cancer (394, 395). It would be interesting to study impact of this differential glycosylation on its functions mainly hemoglobin binding. The detailed mechanism of increased fucosylation of haptoglobin in pancreatic cancer is well studied. Apart from hepatocytes, pancreatic cancer cells themselves and infiltrating lymphocytes around pancreatic cancers express fucosylated haptoglobin (394). Mutations in oncogenic Ras are associated with pancreatic cancer (396). Activation of oncogenic Ras leads to expression of pro-inflammatory cytokine IL-6 (397) which exhibits two fold effects. IL-6 can induce expression of haptoglobin and it also up-regulates machinery related to fucosylation process resulting in increased serum fucosylated haptoglobin levels (394).

Gelsolin was not known to be glycosylated until very recently when Ma and colleagues using novel glycosite profiling strategy identified Asn at position 118 in the gelsolin sequence to be glycosylated (343). However, the detail glycan structure of gelsolin is remained to be characterized. In this biomarker discovery screen, gelsolin was identified in multiple lectin pull-downs (AAL,

DSA, EPHA, GNL, JAC, LPHA, MAA, NPL, PSA, SNA, STL, WFA, and WGA). Based on this result, it is very likely that gelsolin express complex type glycan structure. Gelsolin is a villin family member and exists in two forms namely plasma (mainly expressed by muscle cells) and cytoplasmic (ubiquitously expressed). It is a Ca^{+2} regulated actin filament severing, capping, and nucleating protein (398). Serum gelsolin levels are found to be low in patients with acute liver failure, myocardial infarction, sepsis, myonecrosis (399), and rheumatoid arthritis (400). Some of these pathological conditions involve tissue necrosis which causes release of actin. Free actin is neutralized by binding with gelsolin and these actin-gelsolin complexes are cleared by reticuloendothelial system leading to low levels of gelsolin in circulation (399, 401-403). Irrespective of the disease condition, low gelsolin levels is a marker of poorer patient prognosis (403). Cytoplasmic gelsolin has been found to be down-regulated in variety of cancers namely breast, colorectal, gastric, bladder, lung, prostate, kidney, ovarian, pancreatic, and oral cancers (398). As compared to cytoplasmic gelsolin, much less is known about plasma gelsolin. So once gelsolin is verified in an independent patient cohort, it would be interesting to characterize gelsolin glycan structure and its impact on function.

α-2-macroglobulin is a relatively large tetrameric molecule containing multiple glycosylation and cross-linking sites (404, 405). It is expressed by multiple cell types such as lung fibroblasts, monocytes, macrophages, hepatocytes, astrocytes and adrenocortical cells (404). Apart from rapid neutralization of proteinases released during tissue injury which is a primary function of A2M, presence of multiple reactive sites enable it to carry out secondary functions like binding, transportation and targeting of many biomolecules such as cytokines, hormones, and lipids (406). Due to its very complex physiological roles, it is not surprising to know that A2M level is altered in a variety of pathological conditions such as pancreatic cancer, rheumatoid arthritis, chronic liver disease, inflammatory joint disease, multiple sclerosis, myocardial infarction, pancreatitis and nephrotic syndrome etc (406, 407). From the A2M lectin binding signature between healthy, BE and EAC patients it is plausible that A2M is undergoing differential glycosylation without major changes at the total protein levels.

Collectively, lectin binding signatures for these three representative glycoprotein candidates suggest possible scenarios of differential lectin binding for the biomarker candidates identified. It is important to note that all three protein candidates mentioned vary significantly in terms of their molecular size with haptoglobin being smallest (42 kDa) and α-2-macroglobulin being largest (750 kDa) while gelsolin being intermediate (90 kDa). Furthermore, gelsolin, haptoglobin and α-2-macroglobulin contain 1, 4, and 8 known glycosylation sites suggesting LeMBA workflow is

applicable across heterogeneous species. Albeit at this moment it remained to be determined what are the exact glycosylation structures responsible for differential binding with the lectin-beads used in the LeMBA workflow.

3.4.4 Protein level validation of mass spectrometric data

Many protein biomarker discovery workflows identify low abundant proteins as potential biomarker candidates using sophisticated sample preparation and proteomics technologies. Even though the candidates identified exhibit very high performance during the initial discovery stages, they are not suitable for further stages of biomarker development due to lack of high quality capture/detection affinity reagents such as antibodies required for assay development (408). Proteins identified in the LeMBA pull-down are medium to high abundant serum proteins for which well characterized antibodies are readily available commercially. Using the antibodies against two proteins, haptoglobin and gelsolin, two lectin-protein biomarker candidates were successfully validated. For AAL-HP, protein level validation using western immunoblotting showed very similar results as compared to mass spectrometric quantitation based on peptide level. AAL-GSN was identified using group binding difference analysis meaning gelsolin was not identified for many patient samples in AAL pull-down. Western immunoblotting results showed the presence of gelsolin in AAL pull-down for all patient samples although at lower levels and confirmed significant difference between EAC samples as compared to healthy and BE phenotypes. The total protein level changes in the glycoprotein candidates, if any, can be easily tested using unenriched/non-lectin bound sample approaches using techniques like LC-MS/MS, enzyme-linked immunosorbent assay (ELISA) or western immunoblotting.

In summary, this chapter describes identification of a list of lectin-protein diagnostic biomarker candidates using LeMBA-GlycoSelector pipeline. The next phase of this project went on to verify these candidates using targeted proteomics approach in an independent patient cohort.

Chapter 4.

DEVELOPMENT AND VALIDATION OF MULTIPLE REACTION MONITORING-MASS SPECTROMETRY (MRM-MS) ASSAY

Chapter 4. Development and validation of multiple reaction monitoring-mass spectrometry (MRM-MS) assay

4.1 Introduction

Cancer biomarker discovery and development pipeline is formally divided into five phases (133). The goal of initial stages of biomarker discovery is to identify as many candidates as possible by screening relatively few numbers of clinical samples. Moving forward to later stages of development, the aim is to monitor a panel of biomarkers, packaged in the form of an *in vitro* diagnostic test (IVD), in large scale multi-center clinical trial to evaluate actual diagnostic performance in clinical setting (133, 356, 409). In the previous chapter, serum samples from 29 patient samples were screened using LeMBA-GlycoSelector pipeline with a total 183 lectin-protein candidates discovered as potential biomarker candidates to distinguish between healthy, BE and EAC phenotype. The discovery proteomics technique employed in Chapter 3 is semi-quantitative. For biomarker verification, the aim is to screen relatively larger cohort of patient samples using targeted method which is quantitative, reproducible, rapid, and cost-effective. Multiple reaction monitoring-mass spectrometry (MRM-MS) has emerged as a preferred methodology for precise and accurate quantification of 10s to 100s of proteins in very short duration (410-413). The aim of this chapter is therefore to develop and validate MRM-MS assay for the glycoprotein candidates identified in Chapter 3.

Traditionally, antibody based methodologies have been extensively used for protein quantitation in complex backgrounds. As an example, enzyme-linked immunosorbent assay (ELISA) relies mainly on specificity of antibodies to recognize a particular protein epitope. According to Antibodypedia (http://www.antibodypedia.com/) (414), more than a million antibodies are now commercially available against 92% of the human genome. However, not all of these antibodies are of quality to be used for biomarker verification or in a diagnostic assay and the cost to screen the 100s of biomarker candidates using antibody based assay will be enormous. To monitor glycosylation status of a glycoprotein biomarker candidate, either capture or detection antibody in ELISA can be replaced with a particular lectin of interest. The lectin-antibody based assays are developed in a wide variety of formats using diverse chemistries for read-out e.g. antibody-overlay lectin microarray (ALM) (415), lectin-overlay antibody microarray (LAM) (416), lectin immunosorbent assay (417, 418), or AlphaLISA assay (419). Irrespective of chemistry of detection or assay platform, antibodies based assays lack multiplex capabilities and are unable to quantify 10s to 100s of protein analytes in a single assay, which is a primary requirement for

biomarker verification screen. Hence although antibody based methods are very precise, sensitive, and high-throughput, they are best suited for clinical laboratory but not for purpose of biomarker verification in research laboratories which requires quantitation of 10s to 100s of proteins (356). Earlier, it has been established that LeMBA platform is sensitive, high-throughput, semi-automated and able to meet the demand for sample preparation required for biomarker verification phase (317). When combine with powerful targeted proteomics technique such as MRM-MS, it can be used for biomarker verification.

4.1.1 Multiple reaction monitoring-mass spectrometry (MRM-MS)

Multiple reaction monitoring-mass spectrometry (MRM-MS) has emerged from selected reaction monitoring (SRM) which was developed as targeted mass spectrometry technique to monitor mainly small molecules such as xenobiotics, metabolites or drugs (413). MRM-MS is usually coupled with liquid chromatography for best results whereby the chromatographic system is directly connected in-line with electrospray ionization (ESI) end of the mass spectrometer (Figure 4.1). In a typical workflow (Figure 4.1), proteolytic digest of the complex biological sample is separated using liquid chromatography followed by ionization of the analytes. The ionized peptides undergo selection and fragmentation inside triple quadrupole mass spectrometer and fragmented ions are guided to electron multiplier detector which records the signal in digital format. The selection of ions occurs at two levels to provide assay specificity. (i) Precursor ions are gated in the first mass analyzer (Q1) which then undergoes fragmentation inside the collision cell (Q2) and (ii) resulting product ions are specifically selected inside the second quadrupole (Q3). The pair of precursor-product ion is called transition. The mass spectrometric data acquisition technology has improved remarkably in past few years. Modern day instruments allow repeated and sequential monitoring of several transitions that is fast as compared to chromatographic elution of an analyte (up to 200 transitions for 6490 triple quadrupole mass spectrometer, Agilent Technologies). These results in counts of number of fragment ions (also called intensity) in chromatographic time scale for several transitions that allow quantification of multiple analytes.

The major steps for configuring MRM-MS assays include (i) selection of peptides which are unique for the protein candidates and not shared with other proteins, (ii) selection of transitions for each peptide that do not show any interference (generally multiple product ions from same precursor ion are monitored), and (iii) determination of retention time to allow mass spectrometer to scan for a particular set of transitions in defined retention time window for improved multiplexing. Development of open-source software tool Skyline (http://skyline.maccosslab.org/) by MacLean

and colleagues of University of Washington is considered to be one of the key advancement in the field of targeted proteomics (420). Skyline is vendor neutral software that provides support not only for selecting peptides and optimization of transitions but also for downstream data analysis (420).

This chapter describes development of MRM-MS assay for candidates identified in Chapter 3 for biomarker verification, including characterization of assay linearity and reproducibility.

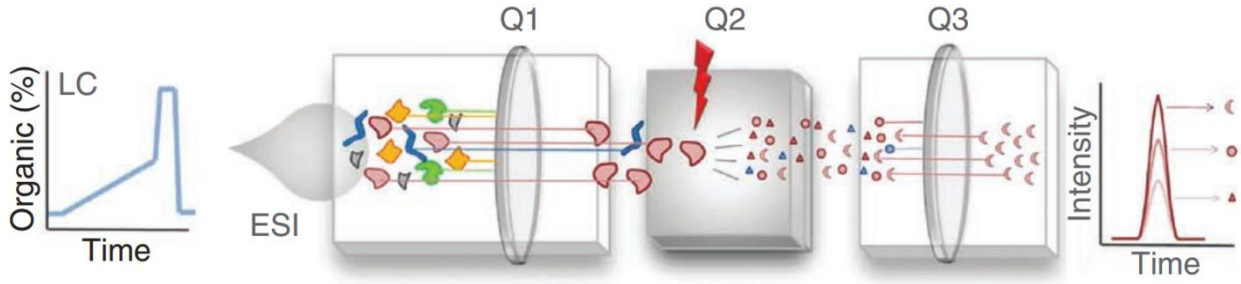


Figure 4.1. A typical workflow of LC-MRM-MS. The complex peptide mixture resulting from proteolytic digest of biological sample undergoes separation using reverse phase liquid chromatography. The LC is connected in-line with triple quadrupole mass spectrometer with electrospray ionization (ESI) interface. The molecular ions of a peptide of interest are selected in Q1 and fragmented in Q2 (Collision cell). The resulting product ions undergo selection in Q3 and travel to the electron multiplier detector which counts number of target fragments over a time resulting in MRM trace for each transition. Adapted from Picotti P. and Aebersold R. (411).

4.2 Experimental procedures

4.2.1 Comparison between nano-flow and standard-flow MRM-MS

Two instruments 1260 HPLC coupled with 6490 triple quadrupole mass spectrometer (Agilent Technologies, nano-flow HPLC-MRM-MS) and 1290 UHPLC coupled with 6490 triple quadrupole mass spectrometer (Agilent Technologies, standard-flow UHPLC-MRM-MS) were compared by running human serum albumin (HSA) peptides standard mix (Agilent Technologies, #G2455-85001). Peptide sequence LVNEVTEFAK (Transitions 575.3 → 937.5 and 575.3 → 694.4; @ Collision energy (CE) 20 eV) was monitored. LC and mass spectrometer parameters for nano-flow and standard-flow MRM-MS are mentioned below.

4.2.1.1 Nano-flow HPLC-MRM-MS parameters

Chip detail: Polaris-HR-Chip-3C18 (G4240-62030, Agilent Technologies) contained 360 nl enrichment column and 75 µm x 150 mm analytical column packed with Polaris C18-A, 180 Å, 3 µm stationary phase.

LC parameters:

Buffer A = 0.1% formic acid, Buffer B = 90% acetonitrile containing 0.1% formic acid

Capillary (loading) pump flow rate = 2.0 µL/min, Nano (analytical) pump flow rate = 0.3 µL/min

LC gradient:

Capillary pump		**Nano pump**	
Time (min)	**Solvent % B**	**Time (min)**	**Solvent %B**
0.0 min	5%	0.0 min	3%
0.5 min	5%	7.0 min	65%
6.5 min	50%	7.5 min	95%
7.0 min	5%	9.5 min	95%
Stop time = 15 min	5%	10.5 min	3%
		Stop time = 15.0 min	3%

Chip cube parameter:

Change inner valve position to enrichment = 11.0 min

Source parameters:

Gas temperature = 150 ˚C, Gas flow rate = 11 L/min, Capillary voltage = 1900 V

4.2.1.2 Standard-flow UHPLC-MRM-MS parameters

Column detail: ZORBAX Rapid resolution high definition Eclipse plus C18 2.1 x 50 mm, 1.8 µm (959757-902, Agilent Technologies)

LC parameters:

Buffer A = 0.1% formic acid, Buffer B = 0.1% formic acid in acetonitrile

Flow rate = 400 µL/min, Column temperature = 50 ˚C

LC gradient:

Time (min)	**Solvent % B**
0.0 min	3%
7.0 min	50%
7.5 min	95%
10.5 min	95%
11.0 min	3%
Stop time = 15 min	3%

Source parameters:

Gas temperature = 150 ˚C, Gas flow = 15 L/min, Nebulizer = 30 psi, Sheath gas heater = 250 ˚C, Sheath gas flow = 11 L/min, Capillary voltage = 3500 V, Nozzle voltage = 300 V

HSA peptides standard mix was serially diluted 10 fold across 7 different concentrations ranging from 1 picomole/µL up to 1 attomole/µL. For nano-flow-HPLC-MRM-MS, HSA peptide standard mix ranging from 100 femtomole up to 1 attomole was injected in triplicate. While for standard-flow-UHPLC-MRM-MS, HSA peptide standard mix ranging from 1 picomole up to 10 attomole was injected in triplicate. The peak area was extracted using Mass Hunter QQQ quantitative analysis software version B.05.02/Build 5.2.365.0. The retention time was compared using Skyline version 2.1.0.4936 (http://skyline.maccosslab.org/).

4.2.2 MRM-MS assay development

MRM-MS assay was set up on the Agilent Technologies 6490 triple quadrupole mass spectrometer coupled with 1290 standard-flow Infinity UHPLC fitted with an ESI source (ESI Jet Stream). The following section describes details about assay development and validation. Six lectins (AAL, EPHA, JAC, NPL, PSA, and WGA) were chosen for verification. MRM-MS assay was developed for 41 glycoprotein candidates identified in biomarker discovery.

4.2.2.1 Selection of peptides and transitions

MRM selector function of Spectrum Mill was used to get a list of the top ten peptides per protein for MRM method development. A few runs from the LeMBA-QTOF discovery data set was used for this purpose. The parameters specified included 10 peptides per protein with a score of above 10 and % score peak intensity of 70%. The top four product y-ions for each precursor ion greater than precursor *m/z* were selected for MRM method development. The formula Collision energy (CE) = 0.036 *m/z* - 4.8 was used to calculate CE for each precursor. Multiple MRM methods consisting of maximum 200 transitions were created as a first step of method development. All methods were transferred across to Skyline for ease of data visualization and analysis (420). Subsequent steps of method refinement were performed using Skyline. Using LeMBA-MS/MS discovery data (.mzxml and .pepxml files), a reference spectral library was built in Skyline. This reference library was used to compare the peptide fragmentation pattern in the MRM method as compared to QTOF data, and also to rank transitions. LeMBA pull-down of multiple lectins was combined and run for each method to identify best MRM transitions. Each method incorporated transitions for internal standard chicken ovalbumin. Retention time prediction calculator iRT-C18 of Skyline was used to increase confidence of peptide identification (421). iRT scale was calibrated using the known retention time of the peptides listed in Table 4.1. Based on the calibration plot, retention time for the peptides of interest was predicted.

MRM transitions showing good response at the correct retention time without any interference were selected for the next step. After the first round of method development, three MRM methods were created and each of these methods was tested in triplicates to find transitions showing stable responses. Some product y-ions (greater than precursor *m/z*) showed considerably low response. So to find out transition with better response, up to five b- and y-ions less than precursor *m/z* were tried. Only transitions showing stable response during multiple runs were selected. Using retention time information for each peptide, one final dynamic MRM method was created incorporating a total of 145 peptides and 465 transitions with delta retention times of 2.5, 3 or 4 min, to quantify 41 proteins. Table 4.2 contains a detailed list of transitions used in the method.

Table 4.1. Peptides used as standards to plot retention time prediction calibration curve.

Peptide sequence	iRT value
VASMASEK	0.00
ISQAVHAAHAEINEAGR	33.91
AVEVLPK	62.07
GGLEPINFQTAADQAR	119.54
LTEWTSSNVMEER	93.10
VTSIQDWVQK	100.00

4.2.2.2 LC and mass spectrometer parameters

The UHPLC system consisted of a reverse phase chromatographic column AdvanceBio Peptide Mapping (150 x 2.1 mm i.d., 2.7 µm, part number 653750-902, Agilent Technologies) with a 5 mm long guard column, maintained at 60 °C temperature. Mobile phase A consisted of 0.1% formic acid, and mobile phase B consisted of 99.9% acetonitrile and 0.1% formic acid. The UHPLC system was operated at a flow rate of 0.4 mL/min. The gradient used for peptide separation was as follows: 3% B at 0 min; 30% B at 20 min; 40% B at 24 min; 95% B at 24.5 min; 95% B at 28.5 min; 3% B at 29 min; followed by conditioning of column for 5 min at 3% B before injecting the next sample.

Agilent 6490 triple quadrupole mass spectrometer was operated in positive ion mode and controlled by MassHunter Workstation software (version B.06.00 build 6.0.6025.4 SP4, Agilent Technologies). The MRM acquisition parameters were 150 V high pressure RF, 60 V low pressure RF, 4000 V capillary voltage, 300 V nozzle voltage, 11 L/min sheath gas flow at a temperature of 250 °C, 15 L/min drying gas flow at a temperature of 250 °C, 30 psi nebulizer gas flow, unit resolution [0.7 Da full width at half maximum in the first quadrupole (Q1) and the third quadrupole

(Q3)], and 200 V delta EMV (+). Fragmentor was set at 380 V and cell accelerator voltage was set at 5 V.

4.2.3 Selection of heavy labeled internal standards and two-step normalization approach

Consistent with biomarker discovery, 10 pmol chicken ovalbumin was spiked-in as an internal standard per lectin pull-down. With the additional use of heavy labeled SIS ovalbumin peptide for MRM analysis, a two step normalization approach became feasible. These two steps separately accounted for variation due to mass spectrometry analysis and LeMBA pull-down. In step 1, the intensity of spiked-in internal standard chicken ovalbumin peptide was normalized using heavy labeled SIS ovalbumin peptide to account for mass spectrometric variation. In step 2, normalized intensity for all peptides for each sample was calculated based on the normalized intensity of spiked-in internal standard chicken ovalbumin, to account for variation during sample preparation steps. The two most consistent chicken ovalbumin peptides, based on discovery screen data, were selected for performing normalization (i) ISQAVHAAHAEINEAGR without methionine and (ii) VASMASEK which contains methionine. The reason for choosing the two separate peptides, one which contains methionine and another which does not, was to account for batch effect of methionine oxidation. To get normalized response of six methionine containing peptides in the final MRM-MS assay, ovalbumin peptide VASMASEK which includes methionine was used; while for normalization of the rest non-methionine containing peptides, ovalbumin peptide ISQAVHAAHAEINEAGR was used. C-terminal isotopic [$^{13}C_6$, $^{15}N_7$] lysine or [$^{13}C_6$, $^{15}N_7$] arginine labeled peptide for ISQAVHAAHAEINEAG**R** and VASMASE**K** with > 95% purity were obtained from Sigma. Two charge states +3 and +4 were monitored for natural and SIS ISQAVHAAHAEINEAGR peptide. The sum of both charge states was used for normalization. C-terminal isotopic [$^{13}C_6$, $^{15}N_7$] lysine labeled peptide AVEVLP**K**, which belongs to Gelsolin, and VTSIQDWVQ**K**, which belongs to Haptoglobin were also incorporated as internal standard peptides.

4.2.4 Determination of loading capacity for each lectin pull-down

Loading capacity for individual lectin pull-down was determined by injecting varying amounts of LeMBA pull-down and monitoring peptide responses using MRM-MS assay. Each LeMBA pull-down sample was resuspended in 20 µL 0.1% formic acid. 5 µL, 10 µL, and 15 µL of the LeMBA pull-down was injected into mass spectrometer for each lectin except EPHA (for EPHA 2 µL, 4 µL, 6 µL, 8 µL, and 10 µL was injected). Using Skyline, individual peptide responses between different injection volumes were monitored carefully to look for peptides showing

saturated responses. Loading capacity for individual lectin pull-down was determined based on overall response for all the peptides monitored across different injection volumes.

4.2.5 Determination of linearity and reproducibility of MRM-MR assay

Linearity of the MRM-MS method was determined by injecting varying concentrations of SIS peptides spiked-into combined LeMBA pull-down sample of multiple lectins. The amount of SIS peptide spiked-in for each of four peptides was adjusted in such a manner that the response from the 1X labeled peptide mix fell within a 5-fold range of the cognate natural peptide. The concentration of spiked-in SIS peptide varied from 0.008X to 25X covering 3125 fold linear range where 1X concentration indicates mixture of 150 femtomole of ISQAVHAAHAEINEAG**R** and VASMASE**K** each, 300 femtomole of VTSIQDWVQ**K**, and 30 femtomole of AVEVLP**K**. All dilutions were run in triplicate on each day for three consecutive days (n = 9). The ratio of SIS peptide response/natural peptide response was plotted.

Reproducibility of MRM-MS assay was determined by injecting 8 µL of combined LeMBA pull-down sample of multiple lectins in quadruplicate on each day for four consecutive days (n = 16). Percent coefficient of variation (% CV) between runs was calculated using peptide responses normalized with respect to ovalbumin peptide.

Table 4.2. List of transitions included in the MRM-MS assay.

Compound Name	**Precursor ion *(m/z)***	**Product ion *(m/z)***	**Ret Time (min)**	**Delta Ret Time**	**Collision Energy**
O75882_ALYVHGGYK	336.52	660.35	6.09	3	7.3
O75882_ALYVHGGYK	336.52	561.28	6.09	3	7.3
O75882_ALYVHGGYK	336.52	412.21	6.09	3	7.3
O75882_GVKGDECQLCEVENR	598.27	806.35	6.94	3	16.7
O75882_GVKGDECQLCEVENR	598.27	646.32	6.94	3	16.7
O75882_GVKGDECQLCEVENR	598.27	517.27	6.94	3	16.7
O75882_SEAACLAAGPGIR	636.82	641.37	9.46	3	18.1
O75882_SEAACLAAGPGIR	636.82	499.30	9.46	3	18.1
O75882_SEAACLAAGPGIR	636.82	288.20	9.46	3	18.1
O75882_SVNNVVVR	443.76	700.41	6.37	3	11.2
O75882_SVNNVVVR	443.76	586.37	6.37	3	11.2
O75882_SVNNVVVR	443.76	472.32	6.37	3	11.2
P00734_GQPSVLQVVNLPIVERPVCK	744.76	1024.09	18.14	3	22
P00734_GQPSVLQVVNLPIVERPVCK	744.76	882.51	18.14	3	22
P00734_GQPSVLQVVNLPIVERPVCK	744.76	683.06	18.14	3	22
P00734_HQDFNSAVQLVENFCR	655.31	824.37	16.13	3	18.8
P00734_HQDFNSAVQLVENFCR	655.31	800.33	16.13	3	18.8
P00734_HQDFNSAVQLVENFCR	655.31	725.30	16.13	3	18.8
P00734_SGIECQLWR	574.78	891.41	11.9	2.5	15.9

Compound Name	Precursor ion *(m/z)*	Product ion *(m/z)*	Ret Time (min)	Delta Ret Time	Collision Energy
P00734_SGIECQLWR	574.78	762.37	11.9	2.5	15.9
P00734_SGIECQLWR	574.78	602.34	11.9	2.5	15.9
P00738_FTDHLK	380.70	613.33	4.66	4	8.9
P00738_FTDHLK	380.70	512.28	4.66	4	8.9
P00738_FTDHLK	380.70	397.26	4.66	4	8.9
P00738_ILGGHLDAK	308.52	349.19	6.79	3	6.3
P00738_ILGGHLDAK	308.52	218.15	6.79	3	6.3
P00738_ILGGHLDAK	308.52	109.58	6.79	3	6.3
P00738_VGYVSGWGR	490.75	881.43	9.79	3	12.9
P00738_VGYVSGWGR	490.75	661.34	9.79	3	12.9
P00738_VGYVSGWGR	490.75	562.27	9.79	3	12.9
P00738_VTSIQDWVQK	602.32	1003.52	11.8	2.5	16.9
P00738_VTSIQDWVQK	602.32	803.40	11.8	2.5	16.9
P00738_VTSIQDWVQK	602.32	675.35	11.8	2.5	16.9
P00738_VTSIQDWVQK.heavy	606.33	1011.53	11.8	2.5	17
P00738_VTSIQDWVQK.heavy	606.33	811.42	11.8	2.5	17
P00738_VTSIQDWVQK.heavy	606.33	683.36	11.8	2.5	17
P00747_LSSPAVITDK	515.79	917.49	9.02	3	13.8
P00747_LSSPAVITDK	515.79	830.46	9.02	3	13.8
P00747_LSSPAVITDK	515.79	743.43	9.02	3	13.8
P00747_NLDENYCR	542.23	856.33	5.04	3	14.7
P00747_NLDENYCR	542.23	741.30	5.04	3	14.7
P00747_NLDENYCR	542.23	498.21	5.04	3	14.7
P00747_VIPACLPSPNYVVADR	885.96	1117.56	14.89	2.5	27.1
P00747_VIPACLPSPNYVVADR	885.96	933.48	14.89	2.5	27.1
P00747_VIPACLPSPNYVVADR	885.96	779.89	14.89	2.5	27.1
P00751_CLVNLIEK	494.78	715.43	13.16	2.5	13
P00751_CLVNLIEK	494.78	616.37	13.16	2.5	13
P00751_CLVNLIEK	494.78	389.24	13.16	2.5	13
P00751_LEDSVTYHCSR	456.21	823.35	5.51	3	11.6
P00751_LEDSVTYHCSR	456.21	627.26	5.51	3	11.6
P00751_LEDSVTYHCSR	456.21	559.24	5.51	3	11.6
P00751_YGLVTYATYPK	638.33	942.49	12.29	2.5	18.2
P00751_YGLVTYATYPK	638.33	843.42	12.29	2.5	18.2
P00751_YGLVTYATYPK	638.33	742.38	12.29	2.5	18.2
P00751_YGQTIRPICLPCTEGTTR	708.35	921.41	12.13	2.5	20.7
P00751_YGQTIRPICLPCTEGTTR	708.35	887.95	12.13	2.5	20.7
P00751_YGQTIRPICLPCTEGTTR	708.35	837.43	12.13	2.5	20.7
P01009_DTEEEDFHVDQVTTVK	631.29	838.39	10.51	2.5	17.9
P01009_DTEEEDFHVDQVTTVK	631.29	773.87	10.51	2.5	17.9
P01009_DTEEEDFHVDQVTTVK	631.29	709.35	10.51	2.5	17.9
P01009_LSITGTYDLK	555.81	910.49	12.15	2.5	15.2
P01009_LSITGTYDLK	555.81	797.40	12.15	2.5	15.2
P01009_LSITGTYDLK	555.81	696.36	12.15	2.5	15.2
P01009_SVLGQLGITK	508.31	829.51	13.39	2.5	13.5
P01009_SVLGQLGITK	508.31	716.43	13.39	2.5	13.5
P01009_SVLGQLGITK	508.31	531.35	13.39	2.5	13.5

Compound Name	Precursor ion *(m/z)*	Product ion *(m/z)*	Ret Time (min)	Delta Ret Time	Collision Energy
P01011_EIGELYLPK	531.30	819.46	13.44	2.5	14.3
P01011_EIGELYLPK	531.30	633.40	13.44	2.5	14.3
P01011_EIGELYLPK	531.30	520.31	13.44	2.5	14.3
P01011_ITLLSALVETR	608.37	888.51	18.44	3	17.1
P01011_ITLLSALVETR	608.37	775.43	18.44	3	17.1
P01011_ITLLSALVETR	608.37	688.40	18.44	3	17.1
P01011_NLAVSQVVHK	547.82	867.50	8.14	3	14.9
P01011_NLAVSQVVHK	547.82	796.47	8.14	3	14.9
P01011_NLAVSQVVHK	547.82	697.40	8.14	3	14.9
P01012_AFKDEDTQAMPFR	778.36	850.42	10.21	2.5	23.2
P01012_AFKDEDTQAMPFR	778.36	669.31	10.21	2.5	23.2
P01012_AFKDEDTQAMPFR	778.36	419.24	10.21	2.5	23.2
P01012_GGLEPINFQTAADQAR	844.42	1121.53	13.55	2.5	25.6
P01012_GGLEPINFQTAADQAR	844.42	860.42	13.55	2.5	25.6
P01012_GGLEPINFQTAADQAR	844.42	666.34	13.55	2.5	25.6
P01012_ISQAVHAAHAEINEAGR	591.97	830.91	6.07	3	16.5
P01012_ISQAVHAAHAEINEAGR	591.97	638.31	6.07	3	16.5
P01012_ISQAVHAAHAEINEAGR	591.97	546.26	6.07	3	16.5
P01012_ISQAVHAAHAEINEAGR	444.23	859.43	6.07	3	11.2
P01012_ISQAVHAAHAEINEAGR	444.23	638.31	6.07	3	11.2
P01012_ISQAVHAAHAEINEAGR	444.23	546.26	6.07	3	11.2
P01012_ISQAVHAAHAEINEAGR.heavy	595.31	835.92	6.07	3	16.6
P01012_ISQAVHAAHAEINEAGR.heavy	595.31	643.32	6.07	3	16.6
P01012_ISQAVHAAHAEINEAGR.heavy	595.31	556.27	6.07	3	16.6
P01012_ISQAVHAAHAEINEAGR.heavy	446.73	869.44	6.07	3	11.3
P01012_ISQAVHAAHAEINEAGR.heavy	446.73	643.32	6.07	3	11.3
P01012_ISQAVHAAHAEINEAGR.heavy	446.73	556.27	6.07	3	11.3
P01012_LTEWTSSNVMEER	791.36	1052.47	11.22	2.5	23.7
P01012_LTEWTSSNVMEER	791.36	951.42	11.22	2.5	23.7
P01012_LTEWTSSNVMEER	791.36	564.24	11.22	2.5	23.7
P01012_NVLQPSSVDSQTAMVLVNAIVFK	820.78	903.57	22.55	4	24.7
P01012_NVLQPSSVDSQTAMVLVNAIVFK	820.78	790.48	22.55	4	24.7
P01012_NVLQPSSVDSQTAMVLVNAIVFK	820.78	393.25	22.55	4	24.7
P01012_VASMASEK	411.70	723.33	3.08	4	10
P01012_VASMASEK	411.70	652.30	3.08	4	10
P01012_VASMASEK	411.70	434.22	3.08	4	10
P01012_VASMASEK.heavy	415.71	731.35	3.08	4	10.2
P01012_VASMASEK.heavy	415.71	660.31	3.08	4	10.2
P01012_VASMASEK.heavy	415.71	442.24	3.08	4	10.2
P01012_YPILPEYLQCVK	761.90	1149.60	16.62	3	22.6
P01012_YPILPEYLQCVK	761.90	1036.51	16.62	3	22.6
P01012_YPILPEYLQCVK	761.90	324.18	16.62	3	22.6
P01019_ALQDQLVLVAAK	634.88	956.58	14.34	2.5	18.1
P01019_ALQDQLVLVAAK	634.88	600.41	14.34	2.5	18.1
P01019_ALQDQLVLVAAK	634.88	501.34	14.34	2.5	18.1
P01019_DPTFIPAPIQAK	649.36	837.52	14.57	2.5	18.6
P01019_DPTFIPAPIQAK	649.36	724.44	14.57	2.5	18.6

Compound Name	Precursor ion *(m/z)*	Product ion *(m/z)*	Ret Time (min)	Delta Ret Time	Collision Energy
P01019_DPTFIPAPIQAK	649.36	556.35	14.57	2.5	18.6
P01019_FMQAVTGWK	534.27	789.43	12.09	2.5	14.4
P01019_FMQAVTGWK	534.27	661.37	12.09	2.5	14.4
P01019_FMQAVTGWK	534.27	491.26	12.09	2.5	14.4
P01019_LDTEDKLR	330.51	531.32	5.02	3	7.1
P01019_LDTEDKLR	330.51	438.72	5.02	3	7.1
P01019_LDTEDKLR	330.51	416.30	5.02	3	7.1
P01019_SLDFTELDVAAEK	719.36	975.50	15.22	3	21.1
P01019_SLDFTELDVAAEK	719.36	745.41	15.22	3	21.1
P01019_SLDFTELDVAAEK	719.36	316.67	15.22	3	21.1
P01023_LPPNVVEESAR	605.82	1000.51	8.98	3	17
P01023_LPPNVVEESAR	605.82	903.45	8.98	3	17
P01023_LPPNVVEESAR	605.82	690.34	8.98	3	17
P01023_SLFTDLEAENDVLHCVAFAVPK	825.75	1064.52	21.06	4	24.9
P01023_SLFTDLEAENDVLHCVAFAVPK	825.75	1014.00	21.06	4	24.9
P01023_SLFTDLEAENDVLHCVAFAVPK	825.75	899.94	21.06	4	24.9
P01023_SLFTDLEAENDVLHCVAFAVPK	825.75	835.42	21.06	4	24.9
P01023_YSDASDCHGEDSQAFCEK	702.60	971.86	5.58	3	20.5
P01023_YSDASDCHGEDSQAFCEK	702.60	928.35	5.58	3	20.5
P01023_YSDASDCHGEDSQAFCEK	702.60	835.31	5.58	3	20.5
P01024_AAVYHHFISDGVR	491.25	665.34	8.21	3	12.9
P01024_AAVYHHFISDGVR	491.25	534.27	8.21	3	12.9
P01024_AAVYHHFISDGVR	491.25	533.27	8.21	3	12.9
P01024_EVVADSVWVDVK	673.35	1018.52	14.06	2.5	19.4
P01024_EVVADSVWVDVK	673.35	646.36	14.06	2.5	19.4
P01024_EVVADSVWVDVK	673.35	246.18	14.06	2.5	19.4
P01024_LLPVGR	327.72	428.26	8.4	3	7
P01024_LLPVGR	327.72	331.21	8.4	3	7
P01024_LLPVGR	327.72	214.63	8.4	3	7
P01024_SGIPIVTSPYQIHFTK	596.66	1033.55	15.88	3	16.7
P01024_SGIPIVTSPYQIHFTK	596.66	765.92	15.88	3	16.7
P01024_SGIPIVTSPYQIHFTK	596.66	660.85	15.88	3	16.7
P01031_IDTQDIEASHYR	483.23	990.46	7.75	3	12.6
P01031_IDTQDIEASHYR	483.23	762.35	7.75	3	12.6
P01031_IDTQDIEASHYR	483.23	633.31	7.75	3	12.6
P01031_IDTQDIEASHYR	483.23	562.27	7.75	3	12.6
P01031_IVACASYKPSR	417.89	737.39	5.75	3	10.2
P01031_IVACASYKPSR	417.89	520.25	5.75	3	10.2
P01031_IVACASYKPSR	417.89	484.73	5.75	3	10.2
P01031_TLLPVSKPEIR	418.26	729.43	11.55	2.5	10.3
P01031_TLLPVSKPEIR	418.26	514.30	11.55	2.5	10.3
P01031_TLLPVSKPEIR	418.26	463.28	11.55	2.5	10.3
P01042_DIPTNSPELEETLTHTITK	713.70	955.99	15.21	3	20.9
P01042_DIPTNSPELEETLTHTITK	713.70	856.94	15.21	3	20.9
P01042_DIPTNSPELEETLTHTITK	713.70	756.40	15.21	3	20.9
P01042_ENFLFLTPDCK	692.33	993.51	16.6	3	20.1
P01042_ENFLFLTPDCK	692.33	880.42	16.6	3	20.1

Compound Name	Precursor ion *(m/z)*	Product ion *(m/z)*	Ret Time (min)	Delta Ret Time	Collision Energy
P01042_ENFLFLTPDCK	692.33	733.35	16.6	3	20.1
P01042_IASFSQNCDIYPGK	800.38	966.43	10.84	2.5	24
P01042_IASFSQNCDIYPGK	800.38	464.25	10.84	2.5	24
P01042_IASFSQNCDIYPGK	800.38	301.19	10.84	2.5	24
P01042_ICVGCPR	431.21	748.32	6.16	3	10.7
P01042_ICVGCPR	431.21	588.29	6.16	3	10.7
P01042_ICVGCPR	431.21	489.22	6.16	3	10.7
P02748_AIEDYINEFSVR	728.36	1142.55	15.78	3	21.4
P02748_AIEDYINEFSVR	728.36	864.46	15.78	3	21.4
P02748_AIEDYINEFSVR	728.36	751.37	15.78	3	21.4
P02748_FTPTETNKAEQCCEETASSISLHGK	707.07	859.72	9.18	3	20.7
P02748_FTPTETNKAEQCCEETASSISLHGK	707.07	828.46	9.18	3	20.7
P02748_FTPTETNKAEQCCEETASSISLHGK	707.07	741.43	9.18	3	20.7
P02748_LSPIYNLVPVK	621.88	1042.63	16.82	3	17.6
P02748_LSPIYNLVPVK	621.88	832.49	16.82	3	17.6
P02748_LSPIYNLVPVK	621.88	521.82	16.82	3	17.6
P02748_RPWNVASLIYETK	526.29	653.35	15.46	3	14.1
P02748_RPWNVASLIYETK	526.29	540.27	15.46	3	14.1
P02748_RPWNVASLIYETK	526.29	248.16	15.46	3	14.1
P02749_TCPKPDDLPFSTVVPLK	638.67	927.42	16.51	3	18.2
P02749_TCPKPDDLPFSTVVPLK	638.67	743.47	16.51	3	18.2
P02749_TCPKPDDLPFSTVVPLK	638.67	743.47	16.51	3	18.2
P02749_TCPKPDDLPFSTVVPLK	638.67	665.87	16.51	3	18.2
P02749_TCPKPDDLPFSTVVPLK	638.67	665.87	16.51	3	18.2
P02749_TFYEPGEEITYSCKPGYVSR	795.04	1067.99	11.88	2.5	23.8
P02749_TFYEPGEEITYSCKPGYVSR	795.04	986.46	11.88	2.5	23.8
P02749_TFYEPGEEITYSCKPGYVSR	795.04	921.94	11.88	2.5	23.8
P02749_VCPFAGILENGAVR	751.89	928.52	16.53	3	22.3
P02749_VCPFAGILENGAVR	751.89	758.42	16.53	3	22.3
P02749_VCPFAGILENGAVR	751.89	622.34	16.53	3	22.3
P02749_WSPELPVCAPIICPPPSIPTFATLR	940.49	805.46	22.63	4	29.1
P02749_WSPELPVCAPIICPPPSIPTFATLR	940.49	648.87	22.63	4	29.1
P02749_WSPELPVCAPIICPPPSIPTFATLR	940.49	600.34	22.63	4	29.1
P02751_SYTITGLQPGTDYK	772.39	978.49	11.72	2.5	23
P02751_SYTITGLQPGTDYK	772.39	808.38	11.72	2.5	23
P02751_SYTITGLQPGTDYK	772.39	680.32	11.72	2.5	23
P02751_VDVIPVNLPGEHGQR	543.96	893.46	13.56	2.5	14.8
P02751_VDVIPVNLPGEHGQR	543.96	780.37	13.56	2.5	14.8
P02751_VDVIPVNLPGEHGQR	543.96	602.32	13.56	2.5	14.8
P02751_VTWAPPPSIDLTNFLVR	642.69	977.54	21.95	4	18.3
P02751_VTWAPPPSIDLTNFLVR	642.69	749.43	21.95	4	18.3
P02751_VTWAPPPSIDLTNFLVR	642.69	749.40	21.95	4	18.3
P02751_VTWAPPPSIDLTNFLVR	642.69	734.91	21.95	4	18.3
P02751_VTWAPPPSIDLTNFLVR	642.69	686.39	21.95	4	18.3
P02765_CNLLAEK	424.22	573.36	7.38	3	10.5
P02765_CNLLAEK	424.22	460.28	7.38	3	10.5
P02765_CNLLAEK	424.22	347.19	7.38	3	10.5

Compound Name	Precursor ion *(m/z)*	Product ion *(m/z)*	Ret Time (min)	Delta Ret Time	Collision Energy
P02765_EHAVEGDCDFQLLK	554.26	763.43	11.99	2.5	15.2
P02765_EHAVEGDCDFQLLK	554.26	738.31	11.99	2.5	15.2
P02765_EHAVEGDCDFQLLK	554.26	648.41	11.99	2.5	15.2
P02765_FSVVYAK	407.23	666.38	9.54	3	9.9
P02765_FSVVYAK	407.23	579.35	9.54	3	9.9
P02765_FSVVYAK	407.23	480.28	9.54	3	9.9
P02765_HTLNQIDEVK	598.82	1059.57	6.91	3	16.8
P02765_HTLNQIDEVK	598.82	958.52	6.91	3	16.8
P02765_HTLNQIDEVK	598.82	822.41	6.91	3	16.8
P02774_SCESNSPFPVHPGTAECCTK	755.65	1009.40	9.4	3	22.4
P02774_SCESNSPFPVHPGTAECCTK	755.65	944.92	9.4	3	22.4
P02774_SCESNSPFPVHPGTAECCTK	755.65	800.87	9.4	3	22.4
P02774_THLPEVFLSK	390.89	578.29	12.66	2.5	9.3
P02774_THLPEVFLSK	390.89	494.30	12.66	2.5	9.3
P02774_THLPEVFLSK	390.89	352.20	12.66	2.5	9.3
P02774_VLEPTLK	400.25	700.42	9.03	3	9.6
P02774_VLEPTLK	400.25	587.34	9.03	3	9.6
P02774_VLEPTLK	400.25	458.30	9.03	3	9.6
P02787_FDEFFSEGCAPGSK	789.33	1039.45	12.43	2.5	23.6
P02787_FDEFFSEGCAPGSK	789.33	892.38	12.43	2.5	23.6
P02787_FDEFFSEGCAPGSK	789.33	805.35	12.43	2.5	23.6
P02787_IECVSAETTEDCIAK	863.39	1224.54	8.92	3	26.3
P02787_IECVSAETTEDCIAK	863.39	1066.47	8.92	3	26.3
P02787_IECVSAETTEDCIAK	863.39	937.43	8.92	3	26.3
P02787_KPVEEYANCHLAR	529.60	841.41	6.11	3	14.3
P02787_KPVEEYANCHLAR	529.60	770.37	6.11	3	14.3
P02787_KPVEEYANCHLAR	529.60	729.84	6.11	3	14.3
P02787_SAGWNIPIGLLYCDLPEPR	724.37	1049.47	22.07	4	21.3
P02787_SAGWNIPIGLLYCDLPEPR	724.37	1009.55	22.07	4	21.3
P02787_SAGWNIPIGLLYCDLPEPR	724.37	886.41	22.07	4	21.3
P02787_SAGWNIPIGLLYCDLPEPR	724.37	771.90	22.07	4	21.3
P02790_EVGTPHGIILDSVDAAFICPGSSR	833.42	1180.54	17.7	3	25.2
P02790_EVGTPHGIILDSVDAAFICPGSSR	833.42	994.48	17.7	3	25.2
P02790_EVGTPHGIILDSVDAAFICPGSSR	833.42	923.44	17.7	3	25.2
P02790_LLQDEFPGIPSPLDAAVECHR	788.73	1351.64	18.29	3	23.6
P02790_LLQDEFPGIPSPLDAAVECHR	788.73	957.42	18.29	3	23.6
P02790_LLQDEFPGIPSPLDAAVECHR	788.73	883.44	18.29	3	23.6
P02790_LLQDEFPGIPSPLDAAVECHR	788.73	809.90	18.29	3	23.6
P02790_NFPSPVDAAFR	610.81	959.49	14.13	2.5	17.2
P02790_NFPSPVDAAFR	610.81	862.44	14.13	2.5	17.2
P02790_NFPSPVDAAFR	610.81	775.41	14.13	2.5	17.2
P02790_YYCFQGNQFLR	748.34	1169.55	13.6	2.5	22.1
P02790_YYCFQGNQFLR	748.34	1009.52	13.6	2.5	22.1
P02790_YYCFQGNQFLR	748.34	862.45	13.6	2.5	22.1
P03952_DSVTGTLPK	459.25	715.43	7.56	3	11.7
P03952_DSVTGTLPK	459.25	616.37	7.56	3	11.7
P03952_DSVTGTLPK	459.25	515.32	7.56	3	11.7

Compound Name	Precursor ion *(m/z)*	Product ion *(m/z)*	Ret Time (min)	Delta Ret Time	Collision Energy
P03952_GVNVCQETCTK	648.29	926.37	5.22	3	18.5
P03952_GVNVCQETCTK	648.29	766.34	5.22	3	18.5
P03952_GVNVCQETCTK	648.29	509.24	5.22	3	18.5
P03952_VLTPDAFVCR	589.31	864.40	13.04	2.5	16.4
P03952_VLTPDAFVCR	589.31	483.23	13.04	2.5	16.4
P03952_VLTPDAFVCR	589.31	432.71	13.04	2.5	16.4
P04003_FSAICQGDGTWSPR	791.36	1163.49	11.3	2.5	23.7
P04003_FSAICQGDGTWSPR	791.36	1003.46	11.3	2.5	23.7
P04003_FSAICQGDGTWSPR	791.36	875.40	11.3	2.5	23.7
P04003_LSLEIEQLELQR	735.91	1028.57	16.36	3	21.7
P04003_LSLEIEQLELQR	735.91	915.49	16.36	3	21.7
P04003_LSLEIEQLELQR	735.91	786.45	16.36	3	21.7
P04004_DVWGIEGPIDAAFTR	823.91	1076.54	19.12	3	24.9
P04004_DVWGIEGPIDAAFTR	823.91	947.49	19.12	3	24.9
P04004_DVWGIEGPIDAAFTR	823.91	890.47	19.12	3	24.9
P04004_FEDGVLDPDYPR	711.83	1146.54	12.33	2.5	20.8
P04004_FEDGVLDPDYPR	711.83	875.43	12.33	2.5	20.8
P04004_FEDGVLDPDYPR	711.83	762.34	12.33	2.5	20.8
P04004_RVDTVDPPYPR	438.90	629.34	8.19	3	11
P04004_RVDTVDPPYPR	438.90	532.29	8.19	3	11
P04004_RVDTVDPPYPR	438.90	472.25	8.19	3	11
P04004_VDTVDPPYPR	579.79	744.37	9.04	3	16.1
P04004_VDTVDPPYPR	579.79	629.34	9.04	3	16.1
P04004_VDTVDPPYPR	579.79	532.29	9.04	3	16.1
P04114_GFEPTLEALFGK	654.85	975.55	19.33	3	18.8
P04114_GFEPTLEALFGK	654.85	664.37	19.33	3	18.8
P04114_GFEPTLEALFGK	654.85	535.32	19.33	3	18.8
P04114_ILGEELGFASLHDLQLLGK	685.05	913.99	20.3	4	19.9
P04114_ILGEELGFASLHDLQLLGK	685.05	756.43	20.3	4	19.9
P04114_ILGEELGFASLHDLQLLGK	685.05	699.89	20.3	4	19.9
P04114_ILGEELGFASLHDLQLLGK	685.05	317.22	20.3	4	19.9
P04114_SPAFTDLHLR	386.21	653.37	11.84	2.5	9.1
P04114_SPAFTDLHLR	386.21	538.35	11.84	2.5	9.1
P04114_SPAFTDLHLR	386.21	425.26	11.84	2.5	9.1
P04196_ALDLINKR	471.79	758.45	8.54	3	12.2
P04196_ALDLINKR	471.79	643.42	8.54	3	12.2
P04196_ALDLINKR	471.79	530.34	8.54	3	12.2
P04196_DGYLFQLLR	562.81	789.50	19.07	3	15.5
P04196_DGYLFQLLR	562.81	676.41	19.07	3	15.5
P04196_DGYLFQLLR	562.81	529.35	19.07	3	15.5
P04196_DSPVLIDFFEDTER	841.90	1284.61	20.25	4	25.5
P04196_DSPVLIDFFEDTER	841.90	1171.53	20.25	4	25.5
P04196_DSPVLIDFFEDTER	841.90	1058.44	20.25	4	25.5
P04196_GGEGTGYFVDFSVR	745.85	1089.54	15.15	3	22.1
P04196_GGEGTGYFVDFSVR	745.85	1032.51	15.15	3	22.1
P04196_GGEGTGYFVDFSVR	745.85	869.45	15.15	3	22.1
P04217_ATWSGAVLAGR	544.80	730.42	11.2	2.5	14.8

Compound Name	Precursor ion *(m/z)*	Product ion *(m/z)*	Ret Time (min)	Delta Ret Time	Collision Energy
P04217_ATWSGAVLAGR	544.80	643.39	11.2	2.5	14.8
P04217_ATWSGAVLAGR	544.80	586.37	11.2	2.5	14.8
P04217_CEGPIPDVTFELLR	823.42	1299.73	19.7	3	24.8
P04217_CEGPIPDVTFELLR	823.42	1089.59	19.7	3	24.8
P04217_CEGPIPDVTFELLR	823.42	877.51	19.7	3	24.8
P05090_NILTSNNIDVK	615.84	1003.54	11.02	2.5	17.4
P05090_NILTSNNIDVK	615.84	890.46	11.02	2.5	17.4
P05090_NILTSNNIDVK	615.84	789.41	11.02	2.5	17.4
P05090_VLNQELR	436.25	772.43	6.59	3	10.9
P05090_VLNQELR	436.25	659.35	6.59	3	10.9
P05090_VLNQELR	436.25	545.30	6.59	3	10.9
P05090_WYEIEK	434.22	681.35	9.17	3	10.8
P05090_WYEIEK	434.22	518.28	9.17	3	10.8
P05090_WYEIEK	434.22	389.24	9.17	3	10.8
P05090_WYEIEK	434.22	350.15	9.17	3	10.8
P05155_FQPTLLTLPR	593.35	910.57	16.56	3	16.6
P05155_FQPTLLTLPR	593.35	712.47	16.56	3	16.6
P05155_FQPTLLTLPR	593.35	599.39	16.56	3	16.6
P05155_GVTSVSQIFHSPDLAIR	609.66	835.95	16.26	3	17.1
P05155_GVTSVSQIFHSPDLAIR	609.66	785.42	16.26	3	17.1
P05155_GVTSVSQIFHSPDLAIR	609.66	692.37	16.26	3	17.1
P05155_HRLEDMEQALSPSVFK	472.49	664.37	13.77	2.5	12.2
P05155_HRLEDMEQALSPSVFK	472.49	651.32	13.77	2.5	12.2
P05155_HRLEDMEQALSPSVFK	472.49	577.33	13.77	2.5	12.2
P05155_LLDSLPSDTR	558.80	890.42	10.34	2.5	15.3
P05155_LLDSLPSDTR	558.80	775.39	10.34	2.5	15.3
P05155_LLDSLPSDTR	558.80	575.28	10.34	2.5	15.3
P05546_QFPILLDFK	560.82	845.51	19.79	3	15.4
P05546_QFPILLDFK	560.82	635.38	19.79	3	15.4
P05546_QFPILLDFK	560.82	522.29	19.79	3	15.4
P05546_TLEAQLTPR	514.79	814.44	9.24	3	13.7
P05546_TLEAQLTPR	514.79	685.40	9.24	3	13.7
P05546_TLEAQLTPR	514.79	486.30	9.24	3	13.7
P05546_YEITTIHNLFR	469.59	549.31	14.52	2.5	12.1
P05546_YEITTIHNLFR	469.59	501.28	14.52	2.5	12.1
P05546_YEITTIHNLFR	469.59	343.69	14.52	2.5	12.1
P06396_AVEVLPK	378.24	585.36	8.52	3	8.8
P06396_AVEVLPK	378.24	456.32	8.52	3	8.8
P06396_AVEVLPK	378.24	244.17	8.52	3	8.8
P06396_AVEVLPK.heavy	382.24	593.37	8.52	3	9
P06396_AVEVLPK.heavy	382.24	464.33	8.52	3	9
P06396_AVEVLPK.heavy	382.24	252.18	8.52	3	9
P06396_DSQEEEKTEALTSAK	555.93	732.36	5.81	3	15.2
P06396_DSQEEEKTEALTSAK	555.93	668.34	5.81	3	15.2
P06396_DSQEEEKTEALTSAK	555.93	590.35	5.81	3	15.2
P06396_QTQVSVLPEGGETPLFK	915.49	1373.73	15.81	3	28.2
P06396_QTQVSVLPEGGETPLFK	915.49	1187.63	15.81	3	28.2

Compound Name	Precursor ion *(m/z)*	Product ion *(m/z)*	Ret Time (min)	Delta Ret Time	Collision Energy
P06396_QTQVSVLPEGGETPLFK	915.49	1074.55	15.81	3	28.2
P06396_TGAQELLR	444.25	729.43	8.2	3	11.2
P06396_TGAQELLR	444.25	658.39	8.2	3	11.2
P06396_TGAQELLR	444.25	530.33	8.2	3	11.2
P06396_VPFDAATLHTSTAMAAQHGMDDDGTGQK	719.08	893.06	11.79	2.5	21.1
P06396_VPFDAATLHTSTAMAAQHGMDDDGTGQK	719.08	844.04	11.79	2.5	21.1
P06396_VPFDAATLHTSTAMAAQHGMDDDGTGQK	719.08	360.66	11.79	2.5	21.1
P06681_AVISPGFDVFAK	625.84	967.49	16.45	3	17.7
P06681_AVISPGFDVFAK	625.84	880.46	16.45	3	17.7
P06681_AVISPGFDVFAK	625.84	783.40	16.45	3	17.7
P06681_DFHINLFR	354.52	549.31	15.45	3	8
P06681_DFHINLFR	354.52	435.27	15.45	3	8
P06681_DFHINLFR	354.52	400.23	15.45	3	8
P08519_LFLEPTQADIALLK	786.46	1311.75	18.99	3	23.5
P08519_LFLEPTQADIALLK	786.46	1198.67	18.99	3	23.5
P08519_LFLEPTQADIALLK	786.46	1069.63	18.99	3	23.5
P08519_NPDAVAAPYCYTR	749.34	1001.45	9.48	3	22.2
P08519_NPDAVAAPYCYTR	749.34	930.41	9.48	3	22.2
P08519_NPDAVAAPYCYTR	749.34	859.38	9.48	3	22.2
P08603_IEGDEEMHCSDDGFWSK	681.27	964.86	16.76	3	19.7
P08603_IEGDEEMHCSDDGFWSK	681.27	900.34	16.76	3	19.7
P08603_IEGDEEMHCSDDGFWSK	681.27	685.27	16.76	3	19.7
P08603_IEGDEEMHCSDDGFWSK	681.27	420.22	16.76	3	19.7
P08603_LSYTCEGGFR	595.27	989.41	8.89	3	16.6
P08603_LSYTCEGGFR	595.27	826.35	8.89	3	16.6
P08603_LSYTCEGGFR	595.27	725.30	8.89	3	16.6
P08603_SITCIHGVWTQLPQCVAIDK	776.06	930.47	15.93	3	23.1
P08603_SITCIHGVWTQLPQCVAIDK	776.06	465.74	15.93	3	23.1
P08603_SITCIHGVWTQLPQCVAIDK	776.06	262.14	15.93	3	23.1
P08603_TGDEITYQCR	621.77	727.32	6.07	3	17.6
P08603_TGDEITYQCR	621.77	463.21	6.07	3	17.6
P08603_TGDEITYQCR	621.77	175.12	6.07	3	17.6
P08603_VSVLCQENYLIQEGEEITCK	804.72	1206.57	16.93	3	24.2
P08603_VSVLCQENYLIQEGEEITCK	804.72	1093.50	16.93	3	24.2
P08603_VSVLCQENYLIQEGEEITCK	804.72	408.19	16.93	3	24.2
P0C0L5_GLQDEDGYR	526.74	754.30	5.25	3	14.2
P0C0L5_GLQDEDGYR	526.74	639.27	5.25	3	14.2
P0C0L5_GLQDEDGYR	526.74	395.20	5.25	3	14.2
P0C0L5_GSFEFPVGDAVSK	670.33	919.49	14.13	2.5	19.3
P0C0L5_GSFEFPVGDAVSK	670.33	772.42	14.13	2.5	19.3
P0C0L5_GSFEFPVGDAVSK	670.33	576.30	14.13	2.5	19.3
P0C0L5_QGSFQGGFR	492.24	798.39	7.33	3	12.9
P0C0L5_QGSFQGGFR	492.24	564.29	7.33	3	12.9
P0C0L5_QGSFQGGFR	492.24	436.23	7.33	3	12.9
P10643_ELSHLPSLYDYSAYR	605.30	774.34	14.25	2.5	17
P10643_ELSHLPSLYDYSAYR	605.30	659.31	14.25	2.5	17
P10643_ELSHLPSLYDYSAYR	605.30	496.25	14.25	2.5	17

Compound Name	Precursor ion *(m/z)*	Product ion *(m/z)*	Ret Time (min)	Delta Ret Time	Collision Energy
P10643_ILPLTVCK	472.29	717.40	12.18	2.5	12.2
P10643_ILPLTVCK	472.29	507.26	12.18	2.5	12.2
P10643_ILPLTVCK	472.29	359.20	12.18	2.5	12.2
P10643_LIDQYGTHYLQSGSLGGEYR	753.03	925.44	12.35	2.5	22.3
P10643_LIDQYGTHYLQSGSLGGEYR	753.03	894.42	12.35	2.5	22.3
P10643_LIDQYGTHYLQSGSLGGEYR	753.03	581.27	12.35	2.5	22.3
P10643_LTPLYELVK	538.32	861.51	16.05	3	14.6
P10643_LTPLYELVK	538.32	651.37	16.05	3	14.6
P10643_LTPLYELVK	538.32	431.26	16.05	3	14.6
P10909_FMETVAEK	477.73	807.39	7.5	3	12.4
P10909_FMETVAEK	477.73	676.35	7.5	3	12.4
P10909_FMETVAEK	477.73	547.31	7.5	3	12.4
P10909_LFDSDPITVTVPVEVSR	937.50	1086.62	17.67	3	28.9
P10909_LFDSDPITVTVPVEVSR	937.50	886.50	17.67	3	28.9
P10909_LFDSDPITVTVPVEVSR	937.50	686.38	17.67	3	28.9
P10909_VTTVASHTSDSDVPSGVTEVVVK	772.06	1201.53	11.55	2.5	23
P10909_VTTVASHTSDSDVPSGVTEVVVK	772.06	1014.58	11.55	2.5	23
P10909_VTTVASHTSDSDVPSGVTEVVVK	772.06	917.53	11.55	2.5	23
P19823_FLHVPDTFEGHFDGVPVISK	747.72	872.93	16.75	3	22.1
P19823_FLHVPDTFEGHFDGVPVISK	747.72	543.35	16.75	3	22.1
P19823_FLHVPDTFEGHFDGVPVISK	747.72	398.22	16.75	3	22.1
P19823_FYNQVSTPLLR	669.36	785.49	13.6	2.5	19.3
P19823_FYNQVSTPLLR	669.36	686.42	13.6	2.5	19.3
P19823_FYNQVSTPLLR	669.36	498.34	13.6	2.5	19.3
P19823_IQPSGGTNINEALLR	791.93	1341.71	12.26	2.5	23.7
P19823_IQPSGGTNINEALLR	791.93	1244.66	12.26	2.5	23.7
P19823_IQPSGGTNINEALLR	791.93	1157.63	12.26	2.5	23.7
P19827_AAISGENAGLVR	579.32	902.47	8.41	3	16.1
P19827_AAISGENAGLVR	579.32	815.44	8.41	3	16.1
P19827_AAISGENAGLVR	579.32	629.37	8.41	3	16.1
P19827_EVAFDLEIPK	580.81	932.51	16.04	3	16.1
P19827_EVAFDLEIPK	580.81	861.47	16.04	3	16.1
P19827_EVAFDLEIPK	580.81	714.40	16.04	3	16.1
P19827_FAHYVVTSQVVNTANEAR	669.34	874.44	10.78	2.5	19.3
P19827_FAHYVVTSQVVNTANEAR	669.34	775.37	10.78	2.5	19.3
P19827_FAHYVVTSQVVNTANEAR	669.34	661.33	10.78	2.5	19.3
P27169_EVQPVELPNCNLVK	819.93	1086.56	13.39	2.5	24.7
P27169_EVQPVELPNCNLVK	819.93	957.52	13.39	2.5	24.7
P27169_EVQPVELPNCNLVK	819.93	844.43	13.39	2.5	24.7
P27169_IQNILTEEPK	592.83	943.51	10.74	2.5	16.5
P27169_IQNILTEEPK	592.83	716.38	10.74	2.5	16.5
P27169_IQNILTEEPK	592.83	603.30	10.74	2.5	16.5
P27169_YVYIAELLAHK	440.58	781.46	16.24	3	11.1
P27169_YVYIAELLAHK	440.58	581.38	16.24	3	11.1
P27169_YVYIAELLAHK	440.58	468.29	16.24	3	11.1
P43652_AIPVTQYLK	516.81	848.49	12.99	2.5	13.8
P43652_AIPVTQYLK	516.81	751.43	12.99	2.5	13.8

Compound Name	Precursor ion *(m/z)*	Product ion *(m/z)*	Ret Time (min)	Delta Ret Time	Collision Energy
P43652_AIPVTQYLK	516.81	652.37	12.99	2.5	13.8
P43652_ESLLNHFLYEVAR	530.95	637.33	18.46	3	14.3
P43652_ESLLNHFLYEVAR	530.95	574.80	18.46	3	14.3
P43652_ESLLNHFLYEVAR	530.95	517.77	18.46	3	14.3
P51884_FNALQYLR	512.78	763.45	14.78	2.5	13.7
P51884_FNALQYLR	512.78	692.41	14.78	2.5	13.7
P51884_FNALQYLR	512.78	579.32	14.78	2.5	13.7
P51884_ILGPLSYSK	489.29	864.48	11.63	2.5	12.8
P51884_ILGPLSYSK	489.29	694.38	11.63	2.5	12.8
P51884_ILGPLSYSK	489.29	597.32	11.63	2.5	12.8
P51884_SVPMVPPGIK	512.80	610.39	12.41	2.5	13.7
P51884_SVPMVPPGIK	512.80	511.32	12.41	2.5	13.7
P51884_SVPMVPPGIK	512.80	419.75	12.41	2.5	13.7
Q14624_EKAEAQAQYSAAVAK	522.27	908.48	5.54	3	14
Q14624_EKAEAQAQYSAAVAK	522.27	709.39	5.54	3	14
Q14624_EKAEAQAQYSAAVAK	522.27	546.32	5.54	3	14
Q14624_EKAEAQAQYSAAVAK	522.27	459.29	5.54	3	14
Q14624_ILDDLSPR	464.76	815.43	10.13	2.5	11.9
Q14624_ILDDLSPR	464.76	702.34	10.13	2.5	11.9
Q14624_ILDDLSPR	464.76	472.29	10.13	2.5	11.9
Q14624_LGVYELLLK	524.33	934.56	17.84	3	14.1
Q14624_LGVYELLLK	524.33	877.54	17.84	3	14.1
Q14624_LGVYELLLK	524.33	778.47	17.84	3	14.1

4.2.6 Determination of linearity of LeMBA pull-down

Serum sample (50 µg each) was spiked with 0.1 picomole, 0.5 picomole, 1 picomole, 10 picomole, 100 picomole and 200 picomole of chicken ovalbumin per pull-down. Using this spiked-in serum sample, LeMBA-MRM-MS was performed using NPL and JAC.

4.2.7 Data processing

Raw data from MRM-MS experiment was processed using Skyline. All peaks were manually checked for correct integration, and peak area for each peptide (sum of all transitions) was exported for further analysis. For linearity experiments, the ratio of SIS:Natural peptide was calculated and plotted against SIS peptide spiked-in concentration. Median normalization was performed. Natural ovalbumin peptide peak intensity was first normalized with respective SIS labeled ovalbumin peptides. Next, using normalized intensity of natural ovalbumin peptide, the intensity of all other peptides was normalized. As mentioned in the methods above, methionine and non-methionine containing peptides were dealt with separately during normalization steps, to account for batch effects in methionine oxidation. For reproducibility experiments, the normalized

response with respect to ovalbumin peptide was calculated for each run, and the % CV of 16 injections of the same sample run over a period of four days calculated.

4.3 Results

4.3.1 Nano-flow vs standard-flow LC-MRM-MS

The biomarker discovery was performed using LeMBA coupled with nano-HPLC-MS/MS (nano-HPLC-QTOF) system. For biomarker verification using LC-MRM-MS assay, two different systems were available. (i) nano-flow-HPLC-triple quadrupole mass spectrometer and (ii) standard-flow-UHPLC-triple quadrupole mass spectrometer. The analytical performance of two instruments was compared for the MRM-MS based quantitation of a mix of 7 HSA peptide sample ran across range of dilutions. The chromatographic conditions for both instruments were optimized separately for this comparison. The result of one peptide is shown in Figure 4.2 as an example. Similar results were observed for the remaining 6 peptides, in terms of quantitation between nano-flow and standard-flow systems. Overall, the nano-flow system was found to be as much as 10-fold more sensitive as compared to standard-flow system.

Chromatographic peak shape and peak width are important analytical parameter to consider when comparing different chromatographic platforms. The chromatogram for peptide LVNEVTEFAK monitored using nano-flow and standard-flow MRM-MS is shown in Figure 4.3A and 4.3B respectively. The full width at half maxima (FWHM) for standard-flow chromatographic separation was 0.0333 min (2.00 sec) as compared to nano-flow chromatographic elution which was 0.0961 min (5.77 sec) (Figure 4.3C). This means, in a complex biological sample such as human serum which contains ~10,000 proteins with a range of more than 10 order-of-magnitudes, standard-flow UHPLC will have better peak-to-peak resolution over nano-flow system. For the constant amount of sample injected, narrow and sharper peak over broader peak can be easy to distinguish from the background noise as well. It is proven that standard-flow UHPLC can handle 5 to 10 times more samples in a single run as compared to nano-flow which means loss in analytical sensitivity can be compensated partly by injecting more amount of sample on column (422). It has also been demonstrated that along with narrower and sharper peak shapes, standard-flow UHPLC over nano-flow platform is less prone to interferences with the MRM transitions. Taken together, the appropriate system should be chosen based on sample availability i.e. if one is not limited by amount of sample than standard-flow UHPLC can provide equal sensitivity as compared to nano-flow HPLC with additional benefits of robust and reproducible chromatography, better resolution and very low interferences for monitoring MRM transitions (422). LeMBA platform is not limited

by sample amount, so considering benefits of standard-flow system over nano-flow system; it was decided to set up MRM-MS assay on standard-flow-UHPLC coupled triple quadrupole mass spectrometer for biomarker verification.

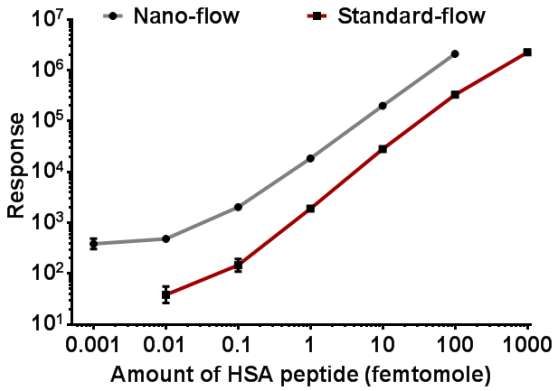


Figure 4.2. Comparison of nano-flow vs standard-flow coupled triple quadrupole mass spectrometer based on MRM-MS quantitation of HSA peptide LVNEVTEFAK.

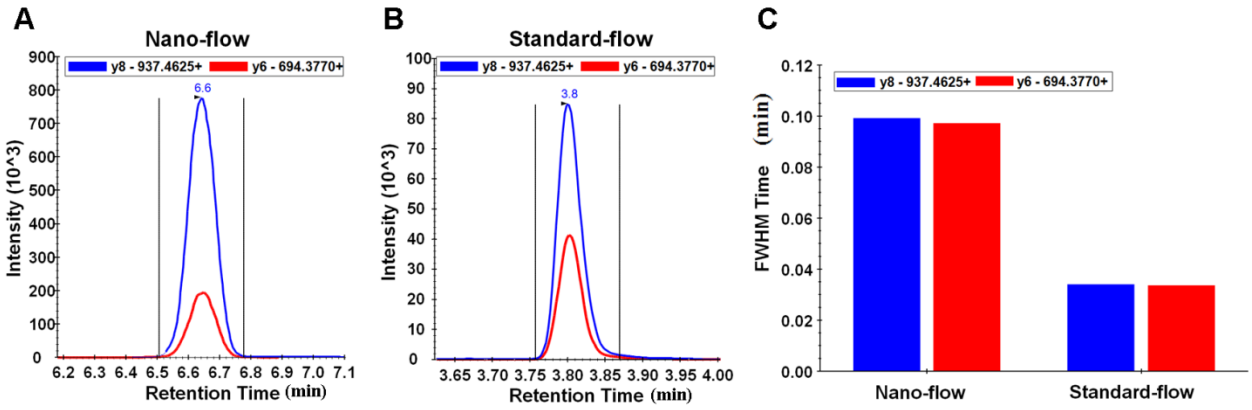


Figure 4.3. Chromatographic elution profile for nano-flow vs standard-flow LC-MRM-MS. Elution of HSA peptide LVNEVTEFAK (Transitions 575.3 → 937.5 and 575.3 → 694.4; @ Collision energy (CE) 20 eV) was monitored using MRM-MS for **(A)** nano-flow-HPLC and **(B)** standard-flow-UHPLC. **(C)** Demonstrates comparison between FWHM of 100 femtomole injection of HSA peptide mix on nano-flow and standard-flow LC-MRM-MS.

4.3.2 MRM-MS assay development

Six lectins (AAL, EPHA, JAC, NPL, PSA, and WGA) were chosen for verification. The following three criteria were considered for choosing lectins. (i) Total number of candidates identified with each lectin in the biomarker discovery screen (Chapter 3), (ii) Previously published serum glycan profiling and lectin histochemistry studies that compared between healthy, BE and EAC phenotypes (169, 170, 175, 179), and (iii) Glycan reactivity group for each of the lectin. MRM-MS assay was developed for 41 glycoprotein candidates identified in biomarker discovery screen. Figure 4.4 illustrates the steps followed for developing MRM-MS assay. The discovery

proteomics data described earlier in Chapter 3 was used to select best possible peptides and transitions for developing MRM-MS assay.

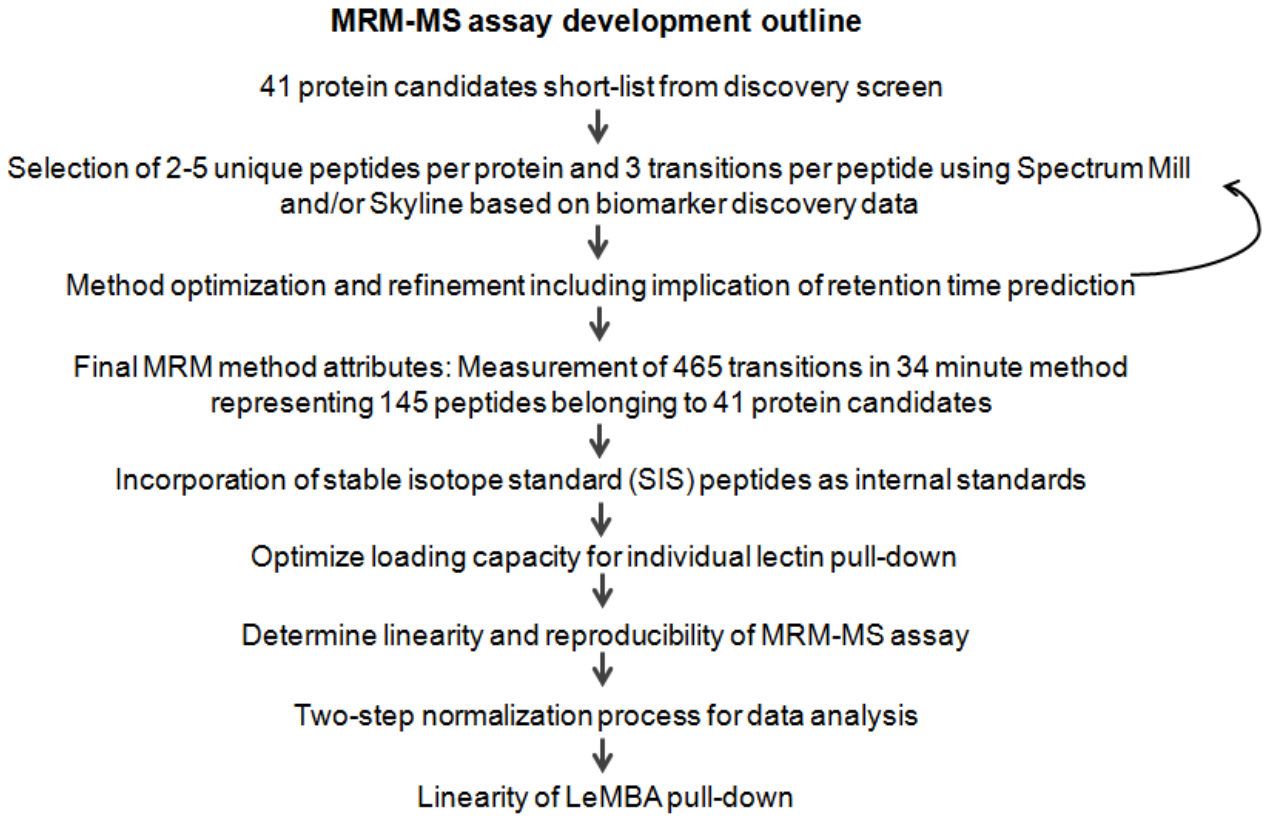


Figure 4.4. Steps followed for development of MRM-MS assay development.

To select 2-5 unique peptides per protein and 3 transitions per peptide for MRM-MS assay, a maximum of 10 peptides per protein and 5 transitions per peptide were imported from biomarker discovery data using MRM selector function of Spectrum Mill. Agilent triple quadrupole mass spectrometer can handle a maximum 200 transitions per MRM method. Collectively, 6 methods were created incorporating transitions to monitor peptides used as retention time prediction standards. LC gradient was kept constant throughout different stages of MRM method development. As a first step of method development, a combined LeMBA pull-down sample from multiple lectins was injected and around 1200 transitions in 6 different MRM methods were monitored. Peptides eluting at a retention time predicted by iRT and transitions showing no interferences were selected for the next round. In the second stage, 3 MRM methods were run in triplicates to qualify transitions showing reproducible responses and to determine retention time of each peptide. In the third and final step, MRM transitions were scheduled whereby retention time and delta retention time for each peptide was specified to create ultimate dynamic (schedule) MRM method. The

dynamic MRM method monitors 465 transitions in 34 min to quantify 145 peptides representing 41 protein candidates.

Peptides with methionine in their sequence (which can undergo oxidation as variable modification) were excluded from MRM-MS assay as far as possible. For six of the proteins, peptides containing methionine showed stable response at least during the method development stages. So in final MRM-MS assay these six peptides containing methionine were chosen. It is important to highlight that for all of these six proteins at least two non-methionine containing peptides were monitored. In addition, four methionine containing peptides monitored were belonging to chicken ovalbumin.

4.3.3 Retention time prediction

In a chromatographic separation, a peptide elutes at a particular retention time, adding one more dimension to the mass spectrometric data recordings. The peptide retention time parameter can be used as an independent way to qualify a peptide for MRM-MS assay development. The retention time of a peptide for a particular LC set up is determined by 3 factors (421). (i) Intrinsic properties of a peptide. It is determined by peptide sequence and structure along with the physicochemical interaction of the peptide with stationary phase and the solvent used for elution (423, 424), (ii) The set up of LC system which affects all peptides consistently and affected by parameters like column length, column temperature, mobile phase gradient, dead volume in the system (425), and (iii) variability in the LC system caused by varying amount of sample loading, variations due to pump pressure or column aging. In this thesis, the biomarker discovery was performed using nano-flow-HPLC-QTOF instrument while MRM-MS assay for biomarker verification was developed on standard-flow-UHPLC-triple quadrupole mass spectrometer. Considering the two LC platforms are different, it was anticipated that peptide retention time for the same peptide would vary between two systems. iRT retention time prediction tool (421) built in Skyline (420) was used as an independent parameter to qualify peptides and to increase peak identification confidence for MRM-MS assay development. The iRT scale was calibrated using known retention time of 6 peptides mentioned in Table 4.1. Transitions for each of these 6 peptides were monitored across all stages of MRM-MS assay development.

The serum/plasma proteome is complex hence it is possible to see interferences and similar looking peaks while selecting transitions for MRM-MS assay development. Figure 4.5A demonstrates practical application of retention time prediction tool. As can be seen from chromatogram, there are two identical looking peaks for peptide YGLVTYATYPK at retention

time 12.4 min and 13.7 min respectively. The predicted retention time according to iRT is 12.4 min suggesting response seen at 12.4 min is due to actual peptide YGLVTYATYPK while peak seen at 13.7 min is due to interference. In the scenario where peptide retention time prediction tool would not be employed, this kind of transitions (or even peptides) would have been dropped from the MRM assay due to lack of confidence. Figure 4.5B and 4.5C shows regression analysis between iRT value and observed retention time of peptides for 1st step of MRM-MS assay development and final MRM-MS assay. The majority of the peptides observed in 1st stage of MRM-MS assay showed response at the retention time predicted by iRT. Peptides which didn't elute at the correct retention time were eliminated from the MRM-MS assay. The resulting dynamic MRM-MS assay showed a perfect correlation between iRT value and observed retention time in experiments.

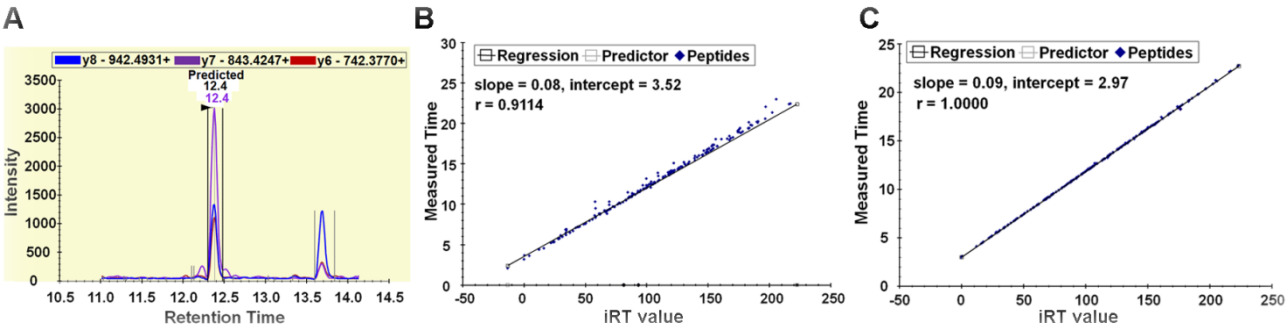


Figure 4.5. Implication of retention time prediction tool iRT on MRM-MS assay development. **(A)** Illustrates chromatographic elution profile of peptide YGLVTYATYPK [Transitions 638.3 → 942.5, 843.4 and 742.4; @ Collision energy (CE) 18.2 eV]. Out of two peaks observed at retention time 12.4 min and 13.7 min, peak at 12.4 min corresponds with retention time predicted by iRT. Regression analysis between observed retention time and iRT value **(B)** for 1st stage of MRM-MS assay development and, **(C)** for final dynamic MRM-MS assay.

4.3.4 Incorporation of heavy labeled internal standard peptides

The nature of LC-MRM-MS assay allows incorporation of stable isotope labeled peptides (heavy amino acid labeled peptide bearing same sequence as natural peptide) as an internal standard. While synthesizing stable isotope standard (SIS) peptides, one or more amino acid in the natural peptide sequence is replaced with isotopic Carbon ($^{13}C_6$) and Nitrogen ($^{15}N_7$) atom counter parts. The resulting heavy labeled peptide carries same chemical properties as the natural peptide hence it elutes at the same time as the natural peptide during chromatography. However, SIS peptide differs in terms of its mass because of isotopic labeling. Trypsin is most commonly used proteolytic enzyme used in the proteomics workflows which digest proteins into peptides containing either lysine (K) or arginine (R) at C-terminus. Hence most common labeling for internal standard peptide uses either isotopically labeled lysine (K) or Arginine (R). In terms of mass, isotopic

Carbon and Nitrogen labeled lysine and arginine differ from natural peptide by 8 Da and 10 Da respectively.

SIS peptides are generally spiked-into the sample after the trypsin digest step and before mass spectrometric analysis. This allows absolute quantitation of peptides present in the samples using known concentration of spiked-in internal standard. Here, the purpose of incorporating SIS peptides into the workflow is to account for mass spectrometric variations over a period of time while screening patient samples for biomarker verification. Native chicken ovalbumin is spiked-in at the sample preparation step to account for variations due to sample processing and LeMBA pull-down. In biomarker discovery (Chapter 3), ovalbumin peptides showed %CV around 40% which is accounted for variations due to sample preparation and LeMBA pull-down plus variations due to mass spectrometric analysis. For biomarker verification, the goal is to reduce this variation to allow better quantitation hence SIS peptides for chicken ovalbumin were incorporated into the workflow just before mass spectrometric analysis to separately account for mass spectrometric variation from variability arising due to sample preparation and LeMBA pull-down. Two ovalbumin peptides VASMASE**K** which contains methionine and ISQAVHAAHAEINEAG**R** without methionine were chosen. The methionine containing peptide was chosen to account for batch effect of methionine oxidation. Sum of two charge states +3 and +4 was used for normalization. In Chapter 3, gelsolin and haptoglobin were successfully validated using orthogonal technique LeMBA-western immunoblotting. So to be more confident, isotopic [$^{13}C_6$, $^{15}N_7$] lysine labeled peptide AVEVLP**K** (belongs to gelsolin), and VTSIQDWVQ**K** (belongs to haptoglobin) were also incorporated as internal standard peptides. Ideally, the concentration of spiked-in SIS peptide should be adjusted such that its response falls within 10-fold range of natural peptide (426). The optimized spiked-in amount was found to be 150 femtomole for ISQAVHAAHAEINEAG**R** and VASMASE**K** each, 300 femtomole for VTSIQDWVQ**K**, and 30 femtomole for AVEVLP**K.** Figure 4.6 illustrates examples of co-elution of SIS labeled peptides within 10-fold response of natural peptide.

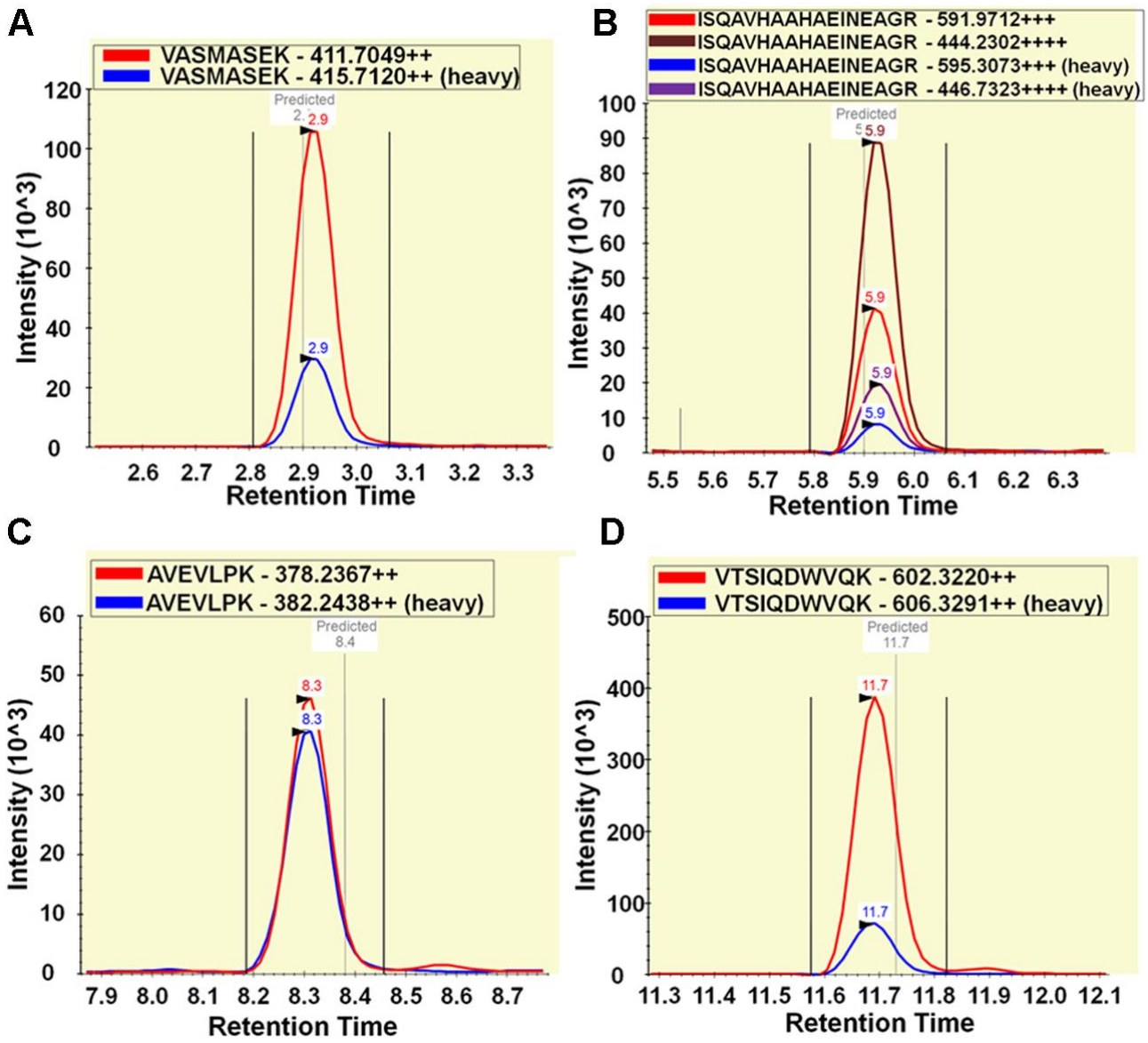


Figure 4.6. Incorporation of stable isotope standard (SIS) peptides into MRM-MS assay. Co-elution of heavy isotope labeled peptide and natural peptide for **(A)** VASMASEK, **(B)** ISQAVHAAHAEINEAGR, **(C)** AVEVLPK and **(D)** VTSIQDWVQK. Peak areas of heavy peptides fell within 10-fold of natural peptide levels.

4.3.5 Linearity and reproducibility of MRM-MS assay

The linearity of the MRM-MS assay was evaluated by spiking a range of dilutions of 4 SIS peptides, spanning a 3125 fold dilution range, into a constant amount of LeMBA pull-down sample. As shown in Figure 4.7A, all 4 SIS peptides showed linear response from 25X dilution up to 0.008X. The reproducibility of the MRM method was determined by running the same sample in quadruplicate for four consecutive days. As illustrated in Figure 4.7B, 86% of the peptides measured using MRM method showed percent coefficient of variation (%CV) below 10%, while 9% of peptides showed %CV between 10-20%, and only 5% of the peptides were above 20%. Out

of 8 peptides that showed %CV greater than 20%, 7 of them were methionine containing peptides. The remaining peptide showed % CV of 20.1% which is just above empirical cut-off of 20%. 3 out of 10 methionine containing peptides showed %CV of 0.96% (VASMASEK), 3.84% (HRLEDMEQALSPSVFK) and 16.75% (NVLQPSSVDSQTAMVLVNAIVFK) suggesting variable oxidation of methionine may depend upon the peptide sequence. This experiment also determined the stability of the sample resuspended in 0.1% formic acid under the storage condition in the auto sampler. It was anticipated that once samples were resuspended in 96 well plates, they would be run within three days. Hence reproducibility was checked for four consecutive days after reconstituting samples. Taken together, MRM-MS assay developed was linear and reproducible.

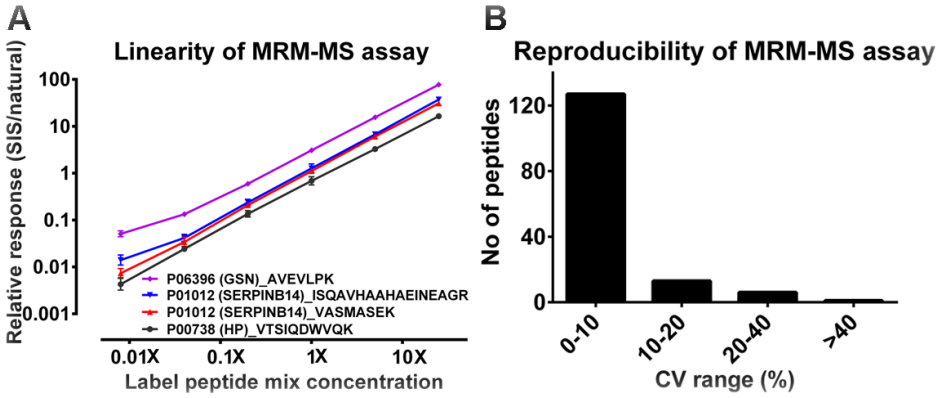


Figure 4.7. Determination of linearity and reproducibility of MRM-MS assay. **(A)** Linearity of MRM-MS assay was confirmed using SIS labeled peptide mix of 4 peptides diluted across 3125 fold and spiked-into a constant amount of LeMBA pull-down sample. **(B)** Reproducibility of MRM-MS assay was determined for 16 replicate injections ran over 4 days period.

4.3.6 Optimization of loading capacity for each lectin pull-down

Loading capacity (amount of peptide to be loaded on LC column) for each lectin pull-down was optimized individually. With increase in amount of peptide loaded on the column, the peak area from MRM-MS measurement show corresponding increase until certain threshold is reached. The optimal loading capacity is reached when the increase in sample amount no longer results in increase in the MRM-MS response (422). To determine the loading capacity for each lectin pull-down, varying amount of LeMBA pull-down was mixed with SIS peptides and injected into the mass spectrometer. Peak areas across all peptides between different injection volumes were monitored to determine loading capacity for each lectin. Replicate comparison tool of Skyline was very useful to compare peak areas across different injection volumes. Based on this comparison, following protocol was used for final experiments. 10 µL of the LeMBA pull-down reconstituted in 0.1% formic acid was mixed with 6 µL SIS peptide mixture containing 150 femtomole of ISQAVHAAHAEINEAG**R** and VASMASE**K** each, 300 femtomole of VTSIQDWVQ**K**, and 30

femtomole of AVEVLP**K**. Out of the total 16 µL mixture, the optimized loading for AAL, JAC, NPL and PSA lectin was 13 µL, EPHA lectin was 11.5 µL and WGA lectin was found to be 12.5 µL.

4.3.7 Linearity of LeMBA pull-down

Earlier work from our laboratory determined linearity of LeMBA pull-down by spiking-in different amount of chicken ovalbumin into human serum, followed by pull-down using ConA lectin and nano-flow HPLC-MS/MS for quantitation (317). Here, the linearity of LeMBA pull-down was tested by spiking 50 µg of serum proteins with 0.1 picomole (4.43 ng), 0.5 picomole (22.14 ng), 1 picomole (44.29 ng), 10 picomole (442.87 ng), 100 picomole (4.43 µg) and 200 picomole (8.86 µg) of chicken ovalbumin at the time of sample denaturation. LeMBA pull-down was performed in triplicate using NPL and JAC lectin, followed by quantitation of ovalbumin peptide using standard-flow UHPLC-MRM-MS assay. As can be seen from graph in Figure 4.8, there was linear increase in MRM-MS response with increasing amount of ovalbumin titrated in the background of serum proteins. Below 0.5 picomole, the pull-down showed non-reproducible results while above 100 picomole the beads were getting saturated hence no more increase in response was observed.

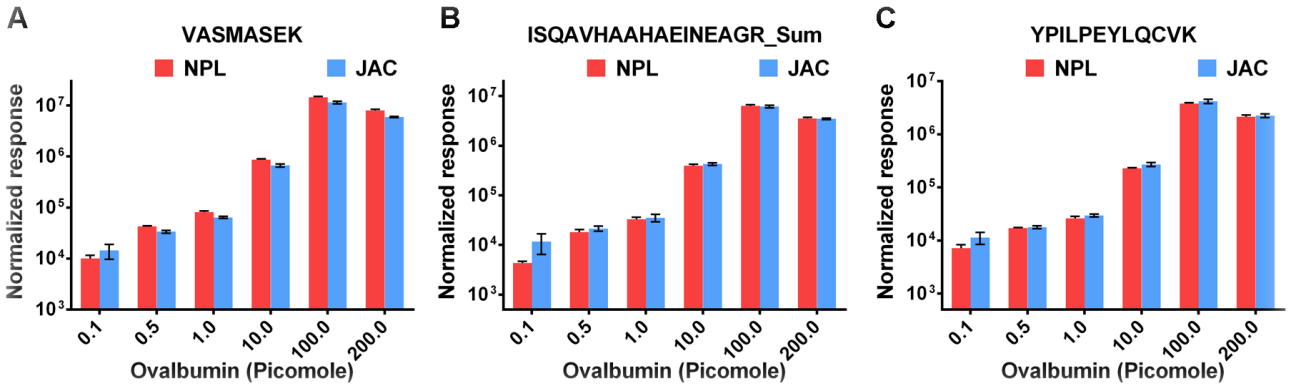


Figure 4.8. Linearity of LeMBA pull-down. The linearity of LeMBA pull-down was tested by titrating chicken ovalbumin in the background of serum protein for NPL and JAC LeMBA pull-down. Quantitation was performed using standard-flow UHPLC-MRM-MS by monitoring transitions for peptides (A) VASMASEK, (B) ISQAVHAAHAEINEAGR, and (C) YPILPEYLQCVK.

4.4 Discussion

4.4.1 Overview

While LeMBA-MS/MS using QTOF and GlycoSelector successfully identified candidate biomarkers (Chapter 3), QTOF is not optimal for biomarker verification as the measurements are semi-quantitative. Therefore, the aim of present chapter was to establish a targeted assay using MRM-MS for biomarker candidates identified in Chapter 3. MRM-MS assay was set up on UHPLC-coupled triple quadrupole mass spectrometer. Later on, SIS peptides were incorporated in the assay. The established MRM-MS assay demonstrated good linearity and reproducibility. After determining optimal loading for each lectin pull-down, linearity of LeMBA pull-down was also confirmed.

4.4.2 Nano-flow vs standard-flow HPLC: Does it matter?

Nano-flow HPLC coupled mass spectrometry enables protein discovery using very minute quantity of biological sample. Now it has almost become a method of choice in protein biomarker discovery using limited quantity of available biological samples (427, 428). Compared with capillary (~50 μL/min) and standard-flow (~500 μL/min) HPLC, nano-flow HPLC (~500 nL/min) offers substantial increase in sensitivity and detection capabilities but at the cost of ease of use and system robustness (429). In addition, this technology is not mature enough to consistently analyze hundreds of samples in reproducible manner hence not ideal for biomarker verification purposes (429). Recently, it has been demonstrated that when sample amount is not limited, standard-flow UHPLC coupled triple quadrupole mass spectrometer can provide comparable sensitivity against nano-flow system with the additional benefits of increase in dynamic range (422). In my experience, standard-flow UHPLC coupled triple quadrupole mass spectrometer is very stable over a long period of time, requires less maintenance hence yield less system down time as compared to nano-flow coupled triple quadrupole mass spectrometer. Moreover, in agreement with observation made by Percy and colleagues (422), the result described in this chapter demonstrates narrow chromatographic peak width using standard-flow UHPLC over nano-flow HPLC and offers better quantitation and less interference while developing MRM-MS assay (422).

Running and maintaining sophisticated instruments like a mass spectrometer is very costly hence all possible efforts should be made during assay development to keep total run time of the method and overall screening as minimal as possible. With nano-flow HPLC coupled QTOF mass spectrometer used for biomarker discovery, blank had to be run after running each sample to reduce the carry-over of the sample which resulted in almost 30% increase in total time spent on the

instrument. Initial experiments using standard-flow UHPLC coupled mass spectrometer resulted in no carry-over of the sample during subsequent runs hence blank injection after each sample run was not required. Furthermore, over a long period of time, retention time reproducibility of the standard-flow system is superior compared to the nano-flow system. This means the delta retention time for each transition can be minimized allowing better multiplexing, longer dwell time (amount of time spent to analyze and detect one ion) for each transition, and lower time for overall method. Taken together, as LeMBA methodology was not limited by sample size, it was decided to use standard-flow UHPLC over nano-flow HPLC platform for biomarker verification study.

4.4.3 MRM-MS assay

Multiple reaction monitoring has emerged as a practical alternative to antibody based platforms for accurate, rapid, reproducible and timely verification of protein biomarker candidates (410). This chapter demonstrates successful development of MRM-MS assay to monitor 41 glycoprotein biomarker candidates in very short time duration of 34 min. iRT, a retention time prediction tool was successfully employed to predict peptide retention time and to increase confidence for peptide identification (421). Each peptide based on its sequence elutes at a specific retention time which can be predicted based upon its hydrophobicity. The most widely used algorithm to predict retention time was SSRCalc (430) until iRT was developed. iRT showed almost 4 times improvement over SSRCalc to predict retention time of the peptides (421). Skyline software version 1.2 and beyond incorporates iRT. Previous studies used a peptide mix of commercially available synthetic standards with different hydrophobicities (hence different retention times) to calibrate iRT scale (431). In contrast, the result described in this chapter used peptides from internal standard chicken ovalbumin that covered entire gradient and two other heavy labeled peptides to calibrate iRT scale, and lead to successful prediction of retention time. This result demonstrates an additional application of using internal standard such as chicken ovalbumin for retention time prediction. With the use of very powerful retention time prediction algorithm such as iRT it is very easy to transfer MRM methods across different instruments even from different vendors. This is very useful in later stages of biomarker development to conduct multi-center evaluation of biomarker candidates in laboratories across the world with different instrument configurations (432). The iRT can also be applied to newly emerging data independent acquisition platforms such as SWATH-MS for biomarker discovery studies (433, 434).

4.4.3.1 Incorporation of relative quantitation in MRM-MS assay using stable isotope labeled peptides

Almost a decade after the entire human genome was sequenced, mass spectrometry based draft of the entire human proteome is now available (435, 436). With availability of this wealth of information and recent advancements in targeted proteomics techniques, it is now feasible to perform absolute or relative quantitation of proteins by accurately measuring the peptide constituents. Stable isotope labeled peptide standards (SIS) are widely used for the mass spectrometric quantitation purposes (437). One of the amino acid constituent of SIS peptide is labeled with isotopic carbon ($^{13}C_6$) and/or nitrogen ($^{15}N_7$) atom. This causes SIS peptide to share exactly same physicochemical properties as its light (natural) counterpart, with the exception of overall mass. Workflow such as stable isotope standards and capture by anti-peptide antibodies (SISCAPA) involves addition of SIS peptide to the sample as internal standard and uses a specific antibody to capture a target peptide from a tryptic digest of plasma (438). Although SISCAPA requires expertise to generate anti-peptide antibody of high quality, it is very sensitive, high-throughput and has potential to overcome current limitations of immunoassays (439). In this chapter, SIS peptides for internal standard chicken ovalbumin were used to separately account for mass spectrometric variation from variation due to LeMBA pull-down. SIS peptides were spiked into the LeMBA pull-down at the time of reconstitution of the sample for mass spectrometric analysis. For best results, response of SIS peptide should fall within 10-fold of response observed for natural peptide (426). Each peptide based on its physicochemical properties has different ionization efficiency and column retention. This means, injecting same amount of different peptides sequences will result in varying responses hence the amount of each SIS peptide to be spiked-in was individually optimized with success as mentioned in the result section.

4.4.3.2 Linearity and reproducibility of the assay

Plasma proteome is the most complex proteome and it extends over quantitative dynamic range of 10^{10} fold. High abundant proteins such as albumin is present at the level of 35 to 50 mg/mL ($3\text{-}5\times10^{10}$ pg/mL) while on the lower side interleukin-6 is present in the range of undetectable to 5 pg/mL levels (362). For the biomarker verification assay, it is necessary to cover at least part of this extended dynamic range. The result described in this chapter confirms linearity of the MRM-MS assay over 4 $\log_{10}$ range.

The acceptable %CV for a mass spectrometry based biomarker verification assay is less than 20% (422). As mentioned in the result, the majority of the measured peptides show %CV of less

than 20% suggesting robust performance of MRM-MS. Figure 4.9 below shows further breakdown of the results of the MRM-MS reproducibility experiment. It is evident that normalizing raw peptide intensities using SIS peptide measurements resulted in lower % CV (better reproducibility) as compared to raw intensities itself. This means that small variations during mass spectrometry analysis are accounted for by inclusion of SIS peptides into the workflow. Over periods of time this will reduce the technical variations and improve the quantitation.

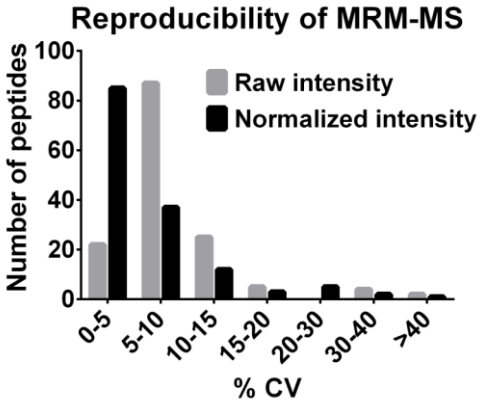


Figure 4.9. Reproducibility of MRM-MS assay before and after normalization using SIS peptides.

Earlier work by Dr Choi optimized the amount of serum to be used for LeMBA pull-down while keeping amount of lectin-beads for pull-down constant. The amount of serum protein was optimized such that it falls within linear range of pull-down efficiency i.e. increase in amount of serum protein resulted into increased binding with the lectin-beads. This earlier work also determined linearity of LeMBA by spiking serum with increasing amount of chicken ovalbumin followed by pull-down using ConA lectin and nano-LC-MS/MS analysis (317). In this chapter, the linearity of LeMBA pull-down using NPL and JAC lectin has been tested. Chicken ovalbumin was spiked from 4.43 ng to 8.86 µg into 50 µg of serum proteins. A linear increase in response was observed up to 4.43 µg (100 picomole) of chicken ovalbumin amount, beyond which there was no increase in response suggesting saturation of lectin-beads at such a high concentration. According to an estimate, up to 50% of the proteins synthesized undergo glycosylation (327, 440). Except albumin which constitute ~50% of the proteome (362), if 50% of the remaining serum proteins are glycoproteins then approximately 12.5 µg out of 50 µg of serum proteins used per lectin pull-down are glycosylated. 8.86 µg of chicken ovalbumin is more than 50% of total serum glycoprotein used for the pull-down hence it was expected to see saturation of lectin-beads when ovalbumin was spiked-in at very high levels.

In conclusion, this chapter describes successful development of MRM-MS assay for 41 glycoprotein candidates. The assay measures 2-5 peptides for each of 41 glycoprotein candidates in

a 34 min MRM-MS method. The assay showed linear response and was found to be reproducible. This targeted proteomics assay will be used for verification of biomarker candidates in an independent patient cohort.

Chapter 5.

VERIFICATION OF LECTIN−GLYCOPROTEIN BIOMARKER CANDIDATES USING LEMBA−COUPLED MRM-MS

Chapter 5. Verification of lectin−glycoprotein biomarker candidates using LeMBA−coupled MRM-MS

5.1 Introduction

The translation of candidate biomarker discovered in research laboratory based discovery screen to clinical use involves several stages of validation. As the potential candidates move forward from initial discovery stages to later validation stages, the list of number of candidate biomarkers screened gets smaller while the number of patient samples screened increases (133, 409). Generally, anywhere from dozens up to hundreds of protein candidate biomarkers are selected as potential biomarkers after completion of biomarker discovery phase. The sample size of the biomarker discovery is small hence interindividual and intraindividual variations are overlooked. Moreover, the distribution of physiological levels of candidate biomarkers in healthy population may result in a bias during biomarker discovery. The presence of confounding risk factors add another layer of complexity and result in identification of false-positive candidates at this initial stage (358). In the recent past, literature has been flooded with publications describing outcome of biomarker discovery studies but evidence of verified list of biomarker candidates ready for clinical validation are seldom (358, 362, 441). Hence, for translational outcome it is essential to verify the biomarker candidates identified during biomarker discovery in an independent patient cohort before conducting large scale clinical validation.

The main goal of biomarker verification is to evaluate performance of the biomarker candidates in an independent patient cohort and establish sensitivity, specificity, area under receiver operating characteristics curve (AUROC) and statistical significance of an individual candidate. The sensitivity of a biomarker is defined as the proportion of patients with disease who will have a positive readout for the biomarker measurement (442). The specificity of a test is defined as the percentage of cohort without underlying disease who will have negative results when tested for a particular biomarker candidate (442). For continuous variables such as mass spectrometric measurements, sensitivity and specificity values vary according to the cut-off (threshold) set for classifying patients into disease or healthy phenotypes. The ROC curve is generated by plotting sensitivity of all possible cut-off points for the test on y-axis against 1-specificity on x-axis (443). The ROC curve is a graphical representation for assessing the ability of a test to discriminate between disease and healthy patient groups (444). It is widely accepted as a method of choice to decide test cut-off for diagnostic purposes (444, 445). AUROC is a reflection of how good the test is at distinguishing between healthy and disease condition and can be used as a single measure to

measure the discriminative ability of a test (446). A perfect biomarker will show AUROC of 1.0 while AUROC of 0.5 indicates biomarker of no use. The biomarker with AUROC > 0.9 has a high accuracy, between 0.7-0.9 has moderate accuracy while between 0.5-0.7 indicates low accuracy (446). Quiet often, a panel of complementary biomarker candidates is identified to improve the diagnostic value of a single biomarker candidate. Similar to single biomarker candidate, sensitivity, specificity, and AUROC value can be determined for a panel of biomarker candidates.

The present chapter describes verification of lectin-protein diagnostic biomarker candidates identified in Chapter 3 using targeted MRM-MS assay developed in Chapter 4 in an independent patient cohort. Statistical significance and AUROC values were calculated for individual biomarker candidates. Furthermore, multivariate analysis was performed to identify a biomarker panel which discriminate between BE and EAC.

5.2 Experimental procedures

5.2.1 Sample information

Serum samples from 61 patients (healthy-20, BE-21 and EAC-20) were randomized for LeMBA pull-down. 19 Population control samples were not randomized. Total 80 serum samples were divided across 7 plates for LeMBA pull down and mass spectrometry analyses (Plate 1 - A1, A2, A3, A4, A5, A6, A7, A8, A9, B1, I7; Plate 2 - E9, E4, C2, F2, E5, C3, F3, E6, C4, F5, E7, C5; Plate 3 - E8, C6, F9, F1, C7, G1, F4, C8, G2; Plate 4 - D1, G6, F8, D2, G7, G3, D3, G9, G5, D4, H1, G8; Plate 5 - D5, H3, H2, D6, H5, H4, D7, H6, H7, D8, I1, H8; Plate 6 - I2, H9, D9, I3, I5, E1, I4, I8, E2, I6, I9, E3; and Plate 7 - F6, C9, G4, B2, B3, B4, B5, B6, B7, B8, B9, C1). Details regarding source of the serum samples is described in section 2.3. Table 5.1 contains details of the samples used for biomarker verification. All samples were collected from male patients and matched for age. Sample I7 was found to be an outlier after LeMBA-MRM-MS data analysis so now onwards remaining 79 patient samples are considered for data analysis. For categorical and numerical variables related to patient information, *P* values were calculated using Fisher's exact test and Kruskal-Wallis test respectively. Table 5.2 summarizes clinical characteristics of the patient cohort for biomarker verification.

Table 5.1. Details of samples used for biomarker verification.

Sample	ID	Phenotype	Age*	Body mass index[#]	Smoking status	Cumulative smoking history	Alcohol consumption[$]	Reflux frequency[&]
F3	43168	BE negative	50.43	Obese I (-<35)	Current	30+ pack yrs	21+ std	Daily

					smoker		drinks/wk	
G7	43316	BE negative	56.41	Obese I (-<35)	Never Smoker	Never smoked	7-20 std drinks/wk	Never
G2	43280	BE negative	57.11	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	21+ std drinks/wk	<Once/month
G6	48024	BE negative	58.53	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	<1 std drink/wk	Monthly (few times/mo)
H4	43302	BE negative	58.55	Overweight(-<30)	Never Smoker	Never smoked	1-6 std drinks/wk	Monthly (few times/mo)
H1	45295	BE negative	58.66	Obese I (-<35)	Never Smoker	Never smoked	7-20 std drinks/wk	Never
H9	48092	BE negative	62.77	Obese I (-<35)	Ex-Smoker	30+ pack yrs	21+ std drinks/wk	Daily
G1	45257	BE negative	64.04	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	7-20 std drinks/wk	Monthly (few times/mo)
H8	48077	BE negative	64.25	Obese I (-<35)	Never Smoker	Never smoked	7-20 std drinks/wk	Never
F9	43202	BE negative	64.31	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	1-6 std drinks/wk	Never
H7	48072	BE negative	64.39	Obese I (-<35)	Ex-Smoker	1-29.9 pack yrs	7-20 std drinks/wk	Weekly (few times/wk)
G4	43291	BE negative	70.57	Healthy wt (-<25)	Never Smoker	Never smoked	1-6 std drinks/wk	<Once/month
F2	45203	BE negative	71.85	Healthy wt (-<25)	Ex-Smoker	1-29.9 pack yrs	21+ std drinks/wk	<Once/month
G9	43325	BE negative	71.93	Healthy wt (-<25)	.	1-29.9 pack yrs	21+ std drinks/wk	.
I9	43668	BE negative	73.11	Healthy wt (-<25)	Never Smoker	Never smoked	7-20 std drinks/wk	Monthly (few times/mo)
I5	43552	BE negative	73.55	Obese I (-<35)	Ex-Smoker	30+ pack yrs	21+ std drinks/wk	Weekly (few times/wk)
H2	48051	BE negative	73.72	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	<1 std drink/wk	Monthly (few times/mo)
I8	43636	BE negative	74.66	Overweight(-<30)	Never Smoker	Never smoked	7-20 std drinks/wk	Never
E9	46070	BE negative	74.77	Overweight(-<30)	Current smoker	30+ pack yrs	7-20 std drinks/wk	Monthly (few times/mo)
F5	43211	BE negative	74.97	Healthy wt (-<25)	Never Smoker	Never smoked	None	Never
F8	47034	BE positive	51.54	Overweight(-<30)	Ex-Smoker	30+ pack yrs	21+ std drinks/wk	Daily
H3	45303	BE positive	52.22	Overweight(-<30)	Never Smoker	Never smoked	21+ std drinks/wk	Monthly (few times/mo)
E7	43123	BE positive	55.16	Obese I (-<35)	Ex-Smoker	1-29.9 pack yrs	1-6 std drinks/wk	Never
E5	43059	BE positive	55.89	Overweight(-<30)	Never Smoker	Never smoked	7-20 std drinks/wk	Daily
E4	43014	BE positive	55.92	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	<1 std drink/wk	Monthly (few times/mo)
F1	46010	BE positive	57.25	Overweight(-<30)	Current smoker	30+ pack yrs	21+ std drinks/wk	Never
F4	46077	BE positive	57.77	Healthy wt (-<25)	Ex-Smoker	1-29.9 pack yrs	21+ std drinks/wk	Weekly (few times/wk)
E6	45062	BE positive	58.79	Overweight(-<30)	Never Smoker	Never smoked	7-20 std drinks/wk	Monthly (few times/mo)
E8	45159	BE positive	58.84	Overweight(-<30)	Never	Never smoked	21+ std	Daily

					Smoker		drinks/wk	
I6	43563	BE positive	60.22	Healthy wt (-<25)	Ex-Smoker	1-29.9 pack yrs	21+ std drinks/wk	Monthly (few times/mo)
H5	45328	BE positive	60.33	Obese I (-<35)	Never Smoker	Never smoked	21+ std drinks/wk	<Once/month
I2	43477	BE positive	63.45	Overweight(-<30)	Current smoker	1-29.9 pack yrs	21+ std drinks/wk	Daily
I1	48094	BE positive	64.06	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	1-6 std drinks/wk	Monthly (few times/mo)
H6	43380	BE positive	64.72	Overweight(-<30)	Never Smoker	Never smoked	1-6 std drinks/wk	Never
G5	48015	BE positive	70.47	Obese I (-<35)	Ex-Smoker	30+ pack yrs	21+ std drinks/wk	Never
I4	45421	BE positive	74	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	<1 std drink/wk	<Once/month
F6	45210	BE positive	74.34	Healthy wt (-<25)	Never Smoker	Never smoked	1-6 std drinks/wk	Daily
G8	43315	BE positive	74.46	Healthy wt (-<25)	Ex-Smoker	1-29.9 pack yrs	7-20 std drinks/wk	<Once/month
I3	45435	BE positive	74.75	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	7-20 std drinks/wk	Monthly (few times/mo)
G3	43269	BE positive	74.98	Healthy wt (-<25)	Never Smoker	Never smoked	None	Daily
I7	46216	BE positive	74.65	Healthy wt (-<25)	Never Smoker	Never smoked	1-6 std drinks/wk	Monthly (few times/mo)
A4	93136	Control	53.45	Overweight(-<30)	Never Smoker	Never smoked	21+ std drinks/wk	Monthly (few times/mo)
A3	94236	Control	53.95	Obese I (-<35)	Never Smoker	Never smoked	1-6 std drinks/wk	<Once/month
B4	94424	Control	56.24	Obese I (-<35)	Never Smoker	Never smoked	21+ std drinks/wk	<Once/month
C1	95133	Control	56.61	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	<1 std drink/wk	Never
B9	94429	Control	56.72	Healthy wt (-<25)	Current smoker	30+ pack yrs	7-20 std drinks/wk	<Once/month
B5	95083	Control	58.84	Healthy wt (-<25)	Never Smoker	Never smoked	1-6 std drinks/wk	Monthly (few times/mo)
B8	95101	Control	59.8	Overweight(-<30)	Never Smoker	Never smoked	7-20 std drinks/wk	Never
A1	96061	Control	60.66	Obese I (-<35)	Ex-Smoker	30+ pack yrs	21+ std drinks/wk	Daily
A6	94259	Control	61.31	Obese II (-<40)	Ex-Smoker	1-29.9 pack yrs	7-20 std drinks/wk	Never
A2	94126	Control	62.29	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	1-6 std drinks/wk	Never
A5	94261	Control	62.82	Healthy wt (-<25)	Ex-Smoker	30+ pack yrs	21+ std drinks/wk	<Once/month
A7	94432	Control	63.11	Healthy wt (-<25)	Ex-Smoker	30+ pack yrs	1-6 std drinks/wk	<Once/month
B3	95154	Control	70.57	Healthy wt (-<25)	Ex-Smoker	30+ pack yrs	21+ std drinks/wk	Never
A9	94449	Control	70.79	Overweight(-<30)	Never Smoker	Never smoked	7-20 std drinks/wk	<Once/month
B7	96189	Control	72.76	Overweight(-<30)	Ex-Smoker	30+ pack yrs	7-20 std drinks/wk	Monthly (few times/mo)
B6	92358	Control	73.05	Overweight(-<30)	Ex-Smoker	1-29.9 pack	21+ std	Monthly (few

						yrs	drinks/wk	times/mo)
B2	94452	Control	73.42	Healthy wt (-<25)	Ex-Smoker	1-29.9 pack yrs	7-20 std drinks/wk	Never
B1	94458	Control	74.18	Overweight(-<30)	Never Smoker	Never smoked	None	<Once/month
A8	94461	Control	74.19	Healthy wt (-<25)	Ex-Smoker	1-29.9 pack yrs	1-6 std drinks/wk	Never
C2	42094	EAC	52.52	Obese I (-<35)	Ex-Smoker	1-29.9 pack yrs	7-20 std drinks/wk	Daily
C3	61017	EAC	54.96	Obese I (-<35)	Ex-Smoker	1-29.9 pack yrs	7-20 std drinks/wk	Weekly (few times/wk)
E1	50231	EAC	55.67	Healthy wt (-<25)	Current smoker	1-29.9 pack yrs	21+ std drinks/wk	Daily
D8	61189	EAC	57.29	Obese II (-<40)	Ex-Smoker	1-29.9 pack yrs	21+ std drinks/wk	Weekly (few times/wk)
D2	61093	EAC	57.51	Obese I (-<35)	Ex-Smoker	1-29.9 pack yrs	1-6 std drinks/wk	Weekly (few times/wk)
D6	21416	EAC	57.59	Obese I (-<35)	Ex-Smoker	30+ pack yrs	21+ std drinks/wk	Daily
C9	21283	EAC	58.97	Obese I (-<35)	Ex-Smoker	30+ pack yrs	21+ std drinks/wk	Daily
D3	35022	EAC	60.09	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	21+ std drinks/wk	Daily
D9	21420	EAC	60.54	Obese I (-<35)	Current smoker	30+ pack yrs	1-6 std drinks/wk	Daily
D4	40270	EAC	60.88	Overweight(-<30)	Never Smoker	Never smoked	1-6 std drinks/wk	Weekly (few times/wk)
D7	33168	EAC	61.72	Obese III (>=40)	Ex-Smoker	30+ pack yrs	1-6 std drinks/wk	Weekly (few times/wk)
E3	33189	EAC	63.1	Overweight(-<30)	Never Smoker	Never smoked	7-20 std drinks/wk	Weekly (few times/wk)
E2	50272	EAC	70.14	Overweight(-<30)	Ex-Smoker	30+ pack yrs	1-6 std drinks/wk	Weekly (few times/wk)
C4	42112	EAC	70.45	Obese I (-<35)	Ex-Smoker	1-29.9 pack yrs	7-20 std drinks/wk	Never
C8	21252	EAC	70.46	Obese I (-<35)	Ex-Smoker	1-29.9 pack yrs	21+ std drinks/wk	Weekly (few times/wk)
C5	40173	EAC	70.72	Obese I (-<35)	Ex-Smoker	1-29.9 pack yrs	1-6 std drinks/wk	Monthly (few times/mo)
D5	40304	EAC	71.11	Obese I (-<35)	Ex-Smoker	30+ pack yrs	21+ std drinks/wk	Daily
D1	21330	EAC	73.62	Obese I (-<35)	Never Smoker	Never smoked	7-20 std drinks/wk	Monthly (few times/mo)
C6	61065	EAC	73.76	Obese I (-<35)	Never Smoker	Never smoked	None	Monthly (few times/mo)
C7	33078	EAC	74.16	Overweight(-<30)	Ex-Smoker	1-29.9 pack yrs	7-20 std drinks/wk	<Once/month

*For cases, age represent age at the time of diagnosis. For controls, age represent age at 1st letter sent.

#Body mass index is according to data available one year before diagnosis. It is categorized into 6 categories according to WHO.

$Alcohol consumption represents average number of standard drinks per week.

[&]Reflux frequency is according to 10 years before diagnosis.

Table 5.2. Clinical characteristics of the patient cohort for biomarker verification.

Variables	Healthy	BE	EAC	*P* value (Healthy vs BE vs EAC)	Population Control	*P* value (Healthy vs Pop. Control)
Sample size	20	20	20		19	
Gender	All male	All male	All male		All male	
Age in years (Median ± SD)	64 ± 8	60 ± 8	61 ± 7	0.4283	62 ± 7	0.2793
Protein concentration (μg/μl)	83 ± 10	78 ± 12	85 ± 13	0.6486	89 ± 13	0.0785
Reflux frequency* (10 years before diagnosis)				0.0108		0.2155
<Once/month	9 (47.4%)	7 (35.0%)	2 (10.0%)		14 (73.7%)	
Monthly (few times/month)	6 (31.6%)	6 (30.0%)	3 (15.0%)		4 (21.1%)	
Weekly or daily	4 (21.0%)	7 (35.0%)	15 (75.0%)		1 (5.3%)	
Body mass index				0.0076		0.6090
Healthy (<25)	5 (25.0%)	5 (25.0%)	1 (5.0%)		7 (36.8%)	
Overweight (25-30)	8 (40.0%)	12 (60.0%)	5 (25.0%)		8 (42.1%)	
Obese (>=30)	7 (35.0%)	3 (15.0%)	14 (70.0%)		4 (21.1%)	
Smoking history				0.6116		0.7813
Never smoked	8 (40.0%)	8 (40.0%)	4 (20.0%)		7 (36.8%)	
1-29.9 pack per year	8 (40.0%)	9 (45.0%)	10 (50.0%)		6 (31.6%)	
30+ pack per year	4 (20.0%)	3 (15.0%)	6 (30.0%)		6 (31.6%)	
Alcohol consumption				0.6637		0.8379
<1 standard drink/week	3 (15.0%)	3 (15.0%)	1 (5.0%)		2 (10.5%)	
1-6 standard drink/week	3 (15.0%)	4 (20.0%)	6 (30.0%)		5 (26.3%)	
7-20 standard drink/week	8 (40.0%)	4 (20.0%)	6 (30.0%)		6 (31.6%)	
21+ standard drink/week	6 (30.0%)	9 (45.0%)	7 (35.0%)		6 (31.6%)	

*All the analyses were performed based on available patient information. Reflux frequency for one healthy patient was missing.

5.2.2 LeMBA-UHPLC-MRM-MS

LeMBA using 6 lectins (AAL, EPHA, JAC, NPL, PSA, and WGA) was performed as described in section 2.5. Each plate for LeMBA pull-down could handle 12 patient samples, with a

total of 7 plates required to process 80 patient samples. To reduce the variability due to sample preparation step, all 80 serum samples were denatured at the same time and spiked with 10 picomole of internal standard chicken ovalbumin per lectin pull-down. Sufficient volume of serum was available to allow processing of double quantity of sample than actual requirement. After denaturation, samples were aliquoted in half and stored at -80 ˚C for later use. Lectin-beads conjugated in single batch were used across LeMBA pull-downs to reduce variability. On the day of LeMBA pull-down, the denatured serum samples were diluted in binding buffer. Lectins were arranged in an alphabetical order in 96 well-plate where each raw represent an individual lectin. After trypsin digestion, peptides were quantified using UHPLC-MRM-MS assay mentioned in Chapter 4. The samples were run according to lectin pull-down on mass spectrometer i.e. LeMBA pull-down samples for the same lectin on each plate were run one after another. After running every 6 LeMBA pull-down samples, a column flush was performed for cleaning and maintenance. Column flush method contains a mixture of gradient and isocratic flow of solvents as mentioned in the table below. 100 femtomole of HSA peptide standard mix was injected as QC after each column flush run to monitor instrument performance.

Time (min)	**Solvent % B**
0.0 min	3%
2.0 min	15%
4.5 min	15%
6.5 min	25%
9.0 min	25%
11.0 min	35%
13.5 min	35%
15.5 min	45%
19.5 min	45%
21.5 min	60%
24.0 min	60%
26.0 min	95%
32.0 min	95%
34.0 min	3%
Stop time = 40.0 min	3%

5.2.3 Shiny mixOmics analysis

The raw data processing and normalization was performed as mentioned earlier in sections 4.2.3 and 4.2.7. Briefly, peak area for each peptide (sum of all transitions) was extracted using Skyline. All peaks were manually checked for correct integration. Median normalization of native chicken ovalbumin peptide ISQAVHAAHAEINEAGR and VASMASEK was performed using

their isotopically labeled counterparts. Using normalized intensity of natural ovalbumin peptides, median normalization was performed for all the peptides measured in MRM-MS assay. The major steps for statistical analysis include: (i) converting peptide intensity to protein intensity, (ii) outlier detection (for details see section 3.2.6.2), (iii) univariate analysis using non-parametric tests such as Kruskal-Wallis tests (significance level set to 0.05) and ROC analysis for each lectin-protein biomarker candidate, and (iv) multivariate analysis using sPLS-DA combined with stability analysis. Normalized peptide response was calculated as mentioned above, using Microsoft excel. Further downstream statistical analysis was performed with a dedicated web application Shiny mixOmics (http://mixomics-projects.di.uq.edu.au/Shiny) using the R statistical software (447) and the R package mixOmics (448) implementing all the statistical data analysis steps described.

The first step in the data analysis pipeline is to infer protein measurements from normalized peptide intensities. 2 - 5 peptides per proteins were measured for each protein. In order to qualify for quantitation, more than 50% of the measured peptides from the same protein must have a Pearson correlation coefficient of more than 0.6. The normalized intensity for different peptides belonging to the same protein varied in the magnitude of more than 10 fold, suggesting simply summing up peptide intensity values to determine protein intensity values would lead to biased results. To overcome this problem, equal weight to each peptide was given irrespective of its absolute intensity when calculating a normalization factor. After converting peptide intensity into protein intensity, outlier detection was performed using the same approach as described previously during GlycoSelector analysis for biomarker discovery in section 3.2.6.2.

After removal of an outlier I7, univariate statistical analysis using non-parametric tests such as Kruskal-Wallis (significance level set to 0.05) and ROC analysis was performed between healthy vs BE, BE vs EAC, and healthy vs EAC patient groups, for each lectin-protein candidate. AUROC value determines the diagnostic ability of each individual lectin-protein candidate to differentiate between two different phenotypes. Sensitivity, specificity and likelihood ratio were calculated at a specified cut-off value for normalized protein response. A likelihood ratio of 10 indicates that the patients with the disease are 10 times more likely to be diagnosed compared to those without the disease (449).

Supervised multivariate approach sPLD-DA was applied to select a panel of biomarker candidate to distinguish patients into two different phenotypes. Stability analysis was carried out that measures frequency of each feature being selected across 5-fold cross-validation repeated 1000 times. Only the robust lectin-protein candidates were reported. Besides the candidate selection and

the stability analysis, another numerical output from sPLS-DA was the classification error rate resulting from the cross-validation procedure. Receiver Operating Characteristic curve (ROC) was also reported for a combination of lectin-protein candidates selected by sPLS-DA. The results of the multivariate analysis were summarized in the form of sPLS-DA sample representation on the first two components using 95% confidence interval ellipses (450). The efforts were made to find out a multimarker panel that distinguish BE from healthy and EAC from BE. Unfortunately, the cross-validation error-rate for healthy vs. BE analysis was very high meaning failure to identify a panel of biomarker candidates with high diagnostic potential.

5.2.4 Analysis for confounders

To check the impact of confounding covariates [reflux frequency, body mass index (BMI), smoking, and alcohol consumption] on biomarker candidates, an additional 19 population control (electoral roll) serum samples were measured using LeMBA-MRM-MS, to achieve sufficient number of disease-free samples for statistical analysis. Healthy and population control sample groups were merged and categorized according to reflux frequency, BMI, cumulative smoking history and alcohol consumption. Kruskal-Wallis test was applied to all the qualified candidates for each confounding factor. Candidates that showed $P < 0.05$ for BMI, reflux, cumulative smoking history or alcohol consumption were considered as false positives and removed prior to multivariate analysis.

5.3 Results

Based on biomarker discovery results, six lectins (AAL, EPHA, JAC, NPL, PSA, and WGA) and 41 protein candidates were selected for biomarker verification. As described in Chapter 4, targeted proteomics assay MRM-MS was developed for quantitation of potential biomarker candidates using LeMBA pull-down. Serum samples from 80 subjects (healthy-20, BE-21, EAC-20, and population control-19) were screened using LeMBA-UHPLC-MRM-MS assay. Table 5.1 and 5.2 depicts patient information used in the screen. The prevalence of reflux and obesity was higher in BE/EAC patient groups as compared to the healthy group, which reiterates fact that reflux and obesity are major risk factors for BE/EAC (451). Age matched electoral roll control and healthy groups were very similar across all measured covariates (Table 5.2).

5.3.1 LeMBA-UHPLC-MRM-MS screen: Quality check

When using modern day technology like LC-mass spectrometer which involves very complex and sophisticated instrumentation, it is necessary to perform enough quality checks to

build confidence in the actual data that is being acquired. This was achieved in following two ways. (i) 100 femtomole of HSA peptide mix was run routinely (after every 6 samples) to monitor instrument performance, and (ii) %CV for SIS and natural chicken ovalbumin peptides was calculated after the screen to check reproducibility of LeMBA pull-down and mass spectrometric analysis.

Figure 5.1 depicts response of 100 femtomole HSA peptide mix analyzed routinely during biomarker verification screen. Total 7 peptides were monitored and showed reproducible response throughout the time period of biomarker verification. In Figure 5.1, response for only one out of seven peptides is displayed for simplicity. Secondly, % CV of the entire LeMBA-MRM-MS screen was calculated based on response of heavy labeled SIS peptides and internal standard natural chicken ovalbumin peptides before and after normalization. All three SIS peptides, except methionine containing heavy labeled peptides, showed a % CV of less than 20% suggesting reproducibility of MRM-MS over entire period of biomarker verification. Initially, % CV for internal standard chicken ovalbumin peptide was found to be around 40-60% which was much higher than the expected 20%. To find out the reason for this very high % CV, internal standard ovalbumin responses were compared using replicate comparison tool in Skyline across all samples. Interestingly, ovalbumin response from samples belongs to one particular plate was much higher as compared to remaining plates suggesting technical error during LeMBA pull-down. Later on, using the second remaining half of the frozen denatured samples, LeMBA pull-down was performed again for the 12 samples. After repeating LeMBA pull-down for samples because of which ovalbumin % CV was found to be high, %CV for normalized as well as non-normalized intensity of natural internal standard ovalbumin peptide was around or below 20% (Table 5.3), suggesting robust performance (422) of LeMBA-MRM-MS screen over several weeks. Interestingly, normalized intensity of natural methionine containing peptide VASMASEK showed less % CV as compared to non-normalized intensity, suggesting SIS peptide VASMASE**K** containing methionine was able to correct for batch effects in methionine oxidation.

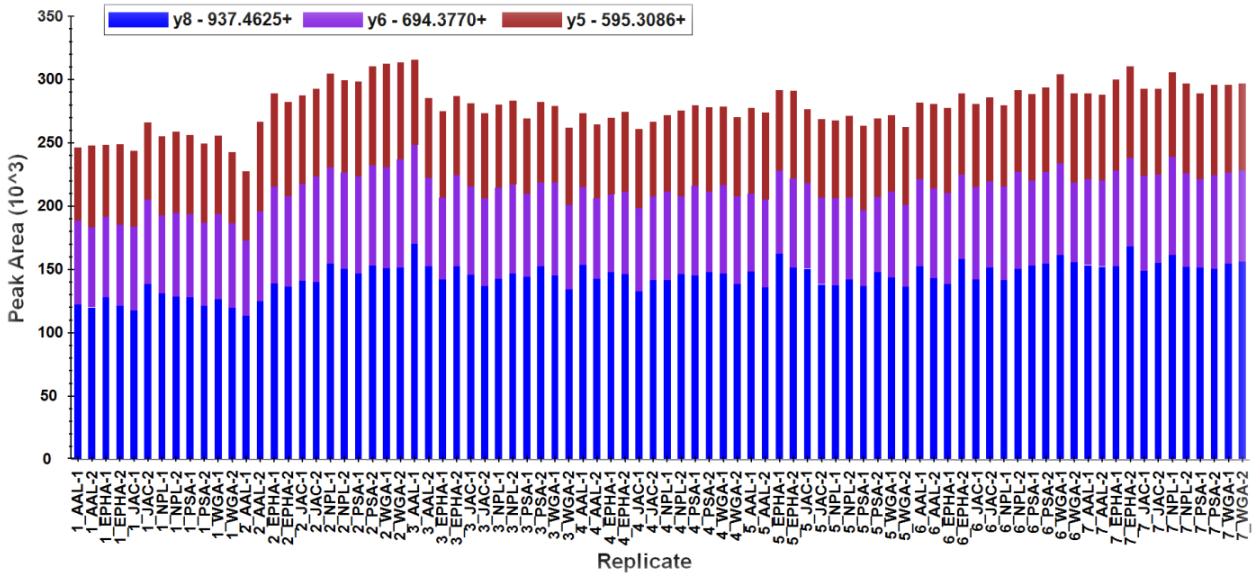


Figure 5.1. Response of HSA peptide mix (QC) during biomarker verification screen. The figure displays peak area for peptide LVNEVTEFAK (Transitions 575.3 m/z → 937.5 m/z, 694.4 m/z, and 595.3 m/z; @ Collision energy (CE) 20 eV) monitored during biomarker verification screen. Replicate nomenclature indicates plate number_lectin name_run number.

Table 5.3. % CV for SIS and natural chicken ovalbumin peptides for BE/EAC biomarker verification screen.

AAL		**NPL**	
Protein name_peptide sequence	**%CV**	**Protein name_peptide sequence**	**%CV**
P00738_VTSIQDWVQK_Heavy	9.98	P00738_VTSIQDWVQK_Heavy	16.46
P01012_ISQAVHAAHAEINEAGR_Sum	15.92	P01012_VASMASEK	24.77
P01012_ISQAVHAAHAEINEAGR_Sum_Heavy	12.92	P01012_VASMASEK_Heavy	25.09
P01012_VASMASEK	21.51	P01012_ISQAVHAAHAEINEAGR_Sum	13.12
P01012_VASMASEK_Heavy	23.66	P01012_ISQAVHAAHAEINEAGR_Sum_Heavy	10.92
P06396_AVEVLPK_Heavy	9.83	P06396_AVEVLPK_Heavy	5.88
Normalized intensity of natural		**Normalized intensity of natural**	
P01012_ISQAVHAAHAEINEAGR_Sum	20.75	P01012_ISQAVHAAHAEINEAGR_Sum	17.23
P01012_VASMASEK	16.99	P01012_VASMASEK	14.31
EPHA		**PSA**	
Protein name_peptide sequence	**%CV**	**Protein name_peptide sequence**	**%CV**
P00738_VTSIQDWVQK_Heavy	14.39	P00738_VTSIQDWVQK_Heavy	13.58
P01012_VASMASEK	25.41	P01012_VASMASEK	26.15
P01012_VASMASEK_Heavy	25.45	P01012_VASMASEK_Heavy	23.97

P01012_ISQAVHAAHAEINEAGR_Sum	16.23	P01012_ISQAVHAAHAEINEAGR_Sum	13.27
P01012_ISQAVHAAHAEINEAGR_Heavy_Sum	13.34	P01012_ISQAVHAAHAEINEAGR_Heavy_Sum	12.62
P06396_AVEVLPK_Heavy	5.38	P06396_AVEVLPK_Heavy	5.43
Normalized intensity of natural		**Normalized intensity of natural**	
P01012_ISQAVHAAHAEINEAGR_Sum	20.96	P01012_ISQAVHAAHAEINEAGR_Sum	18.06
P01012_VASMASEK	18.62	P01012_VASMASEK	14.28
JAC		**WGA**	
Protein name_peptide sequence	**%CV**	**Protein name_peptide sequence**	**%CV**
P00738_VTSIQDWVQK_Heavy	16.50	P00738_VTSIQDWVQK_Heavy	20.50
P01012_VASMASEK	23.62	P01012_VASMASEK	32.97
P01012_VASMASEK_Heavy	25.82	P01012_VASMASEK_Heavy	31.04
P01012_ISQAVHAAHAEINEAGR_Sum	11.78	P01012_ISQAVHAAHAEINEAGR_Sum	22.81
P01012_ISQAVHAAHAEINEAGR_Sum_Heavy	12.42	P01012_ISQAVHAAHAEINEAGR_Sum_Heavy	17.57
P06396_AVEVLPK_Heavy	6.86	P06396_AVEVLPK_Heavy	14.44
Normalized intensity of natural		**Normalized intensity of natural**	
P01012_ISQAVHAAHAEINEAGR_Sum	18.17	P01012_ISQAVHAAHAEINEAGR_Sum	17.05
P01012_VASMASEK	16.00	P01012_VASMASEK	17.61

5.3.2 Shiny mixOmics analysis

So far, it has been demonstrated that LeMBA-MRM-MS showed very high linearity and reproducibility and passed all the quality checks which means the data acquired is of high quality. For each protein, 2-5 peptides were measured for quantitation. Ideally, all the peptides belong to same protein should give similar results when used for quantitation. To test this, Pearson correlation was performed between peptide responses from same proteins across all patient samples. In order to qualify for quantitation, more than 50% of the measured peptides from the same protein must have a Pearson correlation coefficient of more than 0.6. Figure 5.2 depicts the percentage of lectin-protein candidates quantified with decreasing cut-off for Pearson correlation co-efficient. As shown in the figure, using a correlation cut-off of 0.9, 0.8, 0.7 and 0.6, respectively, 55%, 85%, 95% and 96% of the lectin-protein candidates were quantified. The remaining 4% of the candidates which did not meet the criteria were discarded from further analysis. The normalized intensity for different peptides belonging to the same protein varied in the magnitude of more than 10 fold, suggesting simply summing up peptide intensity values to determine protein intensity values would lead to biased results. To overcome this problem, equal weight was given to each peptide irrespective of its absolute intensity when calculating protein intensity. After converting peptide intensity into protein

intensity, outlier detection was performed using the same approach as described previously during GlycoSelector analysis for biomarker discovery. Sample I7 was found to be an outlier and removed from further analysis.

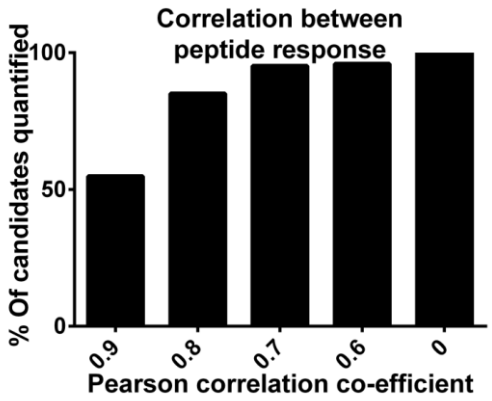


Figure 5.2. Correlation between peptide responses for individual proteins as measured by Pearson correlation, showing % of quantifiable candidates.

5.3.2.1 Candidate selection: Univariate analysis

Two sequential steps were used to evaluate and select candidate biomarkers from the verification data; first, Kruskal-Wallis non-parametric test to assess statistical significance of each individual candidate, then AUROC value was used to measure the diagnostic potential of each marker. Comparisons were made between healthy vs BE, BE vs EAC and healthy vs EAC phenotypes. Out of total 246 lectin-protein candidates, 45 candidates were significantly different between two groups (*P* value < 0.05) (Table 5.4). Amongst them, 26 lectin-protein candidates showed AUROC of more than 0.7 in at least one of the three phenotype comparisons. Boxplots and ROC curves of the top candidate for healthy vs BE, BE vs EAC and healthy vs EAC are shown in Figure 5.3A to 5.3F respectively. Apolipoprotein B-100 (APOB; Uniprot entry: P04114) showed differential binding with NPL lectin between healthy and BE patient groups (*P* value = 0.0231, AUROC = 0.71). It showed sensitivity of 30%, specificity of 95%, and likelihood ratio of 6.0 at cut-off value of 569508. Complement component C9 (C9; Uniprot entry: P02748) was the top candidate to differentiate EAC from BE phenotype (*P* value = 0.0001, AUROC = 0.85) with sensitivity of 55%, specificity of 95%, and likelihood ratio of 11 at cut-off value > 420932. EPHA-Gelsolin was statistically significantly different between healthy and EAC patient groups (*P* value = 0.0014, AUROC = 0.80, Sensitivity = 35%, Specificity = 95%, Likelihood ratio = 7.0, and Cut-off = 110686). Figure 5.4 depicts Venn diagram of number of candidates that can differentiate between healthy vs BE, BE vs EAC and healthy vs EAC patient groups. Out of total 45 candidates that showed statistically significant difference, 16 candidates overlapped between healthy vs EAC and BE vs EAC analysis and might be of greatest interest as they can differentiate EAC from healthy as

well as BE phenotype. α-2-Macroglobulin (A2M; Uniprot entry: P01023) was statistically significantly different in both healthy vs BE and healthy vs EAC analysis.

Table 5.4. Verified list of candidates shown by lectin affinity-protein ID that were significantly different for either healthy vs BE or BE vs EAC or healthy vs EAC analysis. Proteins are denoted using gene symbol; number in the bracket denotes Uniprot accession number. AUROC values of more than 0.7 are highlighted in bold.

Lectin-Protein	**Healthy vs BE**		**BE vs EAC**		**Healthy vs EAC**	
	***P* value**	**AUROC**	***P* value**	**AUROC**	***P* value**	**AUROC**
AAL-APOB (P04114)	0.1368	0.6375	0.0453	0.6850	0.9569	0.4950
AAL-C5 (P01031)	0.6073	0.5475	0.0483	0.6825	0.2340	0.6100
AAL-C7 (P10643)	0.2793	0.6000	**0.0063**	**0.7525**	0.3169	0.5925
AAL-C9 (P02748)	0.2793	0.6000	**0.0001**	**0.8525**	**0.0161**	**0.7225**
AAL-GSN (P06396)	0.7455	0.5300	**0.0087**	**0.7425**	**0.0265**	**0.7050**
AAL-HP (P00738)	0.8711	0.4850	0.0398	0.6900	0.0583	0.6750
EPHA-A2M (P01023)	**0.0248**	**0.7075**	0.9138	0.4900	**0.0186**	**0.7175**
EPHA-AHSG (P02765)	0.5162	0.5600	0.1941	0.6200	0.0483	0.6825
EPHA-C7 (P10643)	0.1368	0.6375	0.0398	0.6900	0.6849	0.5375
EPHA-C9 (P02748)	0.0583	0.6750	**0.0003**	**0.8375**	**0.0265**	**0.7050**
EPHA-GSN (P06396)	0.2036	0.6175	**0.0200**	**0.7150**	**0.0014**	**0.7950**
EPHA-HP (P00738)	0.7455	0.5300	**0.0200**	**0.7150**	0.0305	0.7000
EPHA-SERPINA3 (P01011)	0.4171	0.5750	**0.0265**	**0.7050**	0.0620	0.6725
EPHA-TF (P02787)	0.7455	0.4700	0.0326	0.6975	0.0935	0.6550
JAC-A1BG (P04217)	0.6263	0.5450	0.0483	0.6825	0.1231	0.6425
JAC-APOB (P04114)	0.0305	0.7000	0.0699	0.6675	0.5700	0.5525
JAC-C4BPA (P04003)	0.7251	0.5325	0.0935	0.6550	**0.0128**	**0.7300**
JAC-C5 (P01031)	0.6073	0.4525	0.0425	0.6875	0.0483	0.6825
JAC-C7 (P10643)	0.2914	0.5975	**0.0094**	**0.7400**	0.0834	0.6600
JAC-C9 (P02748)	0.2914	0.5975	**0.0007**	**0.8125**	**0.0029**	**0.7750**
JAC-CFB (P00751)	0.9353	0.5075	0.0373	0.6925	0.0373	0.6925
JAC-GSN (P06396)	0.8498	0.5175	0.0305	0.7000	**0.0215**	**0.7125**
JAC-HP (P00738)	0.9569	0.5050	0.0483	0.6825	0.0583	0.6750
JAC-HPX (P02790)	0.7868	0.5250	0.0742	0.6650	**0.0200**	**0.7150**
JAC-SERPINA1 (P01009)	0.3040	0.5950	0.0453	0.6850	0.2448	0.6075

JAC-SERPINA3 (P01011)	0.9569	0.5050	**0.0102**	**0.7375**	0.0305	0.7000
JAC-SERPIND1 (P05546)	0.1368	0.6375	0.4819	0.5650	0.0483	0.6825
JAC-SERPING1 (P05155)	0.5518	0.5550	0.2559	0.6050	**0.0200**	**0.7150**
NPL-AFM (P43652)	0.5338	0.5575	0.0483	0.6825	0.1762	0.6250
NPL-APOB (P04114)	**0.0231**	**0.7100**	**0.0231**	**0.7100**	0.8924	0.5125
NPL-C4BPA (P04003)	0.0989	0.6525	0.6849	0.5375	**0.0231**	**0.7100**
NPL-C9 (P02748)	0.5885	0.5500	**0.0049**	**0.7600**	**0.0074**	**0.7475**
NPL-GSN (P06396)	0.8924	0.5125	**0.0173**	**0.7200**	0.0583	0.6750
NPL-HP (P00738)	0.8077	0.5225	0.0884	0.6575	0.0326	0.6975
NPL-SERPINA3 (P01011)	0.5518	0.5550	0.0989	0.6525	0.0305	0.7000
PSA-C5 (P01031)	0.4017	0.5775	0.0453	0.6850	0.3040	0.5950
PSA-C7 (P10643)	0.2914	0.5975	**0.0019**	**0.7875**	0.0742	0.6650
PSA-C9 (P02748)	0.2036	0.6175	**0.0008**	**0.8100**	**0.0161**	**0.7225**
PSA-GSN (P06396)	0.3577	0.5850	0.0483	0.6825	**0.0110**	**0.7350**
PSA-HP (P00738)	0.8498	0.4825	0.0483	0.6825	0.0425	0.6875
PSA-SERPINA3 (P01011)	0.8077	0.5225	0.0425	0.6875	0.0834	0.6600
WGA-C9 (P02748)	0.4819	0.5650	**0.0032**	**0.7725**	**0.0053**	**0.7575**
WGA-GSN (P06396)	0.7868	0.5250	**0.0119**	**0.7325**	0.0742	0.6650
WGA-HP (P00738)	0.7455	0.5300	0.0483	0.6825	**0.0215**	**0.7125**
WGA-SERPINA3 (P01011)	0.4819	0.5650	0.0989	0.6525	**0.0063**	**0.7525**

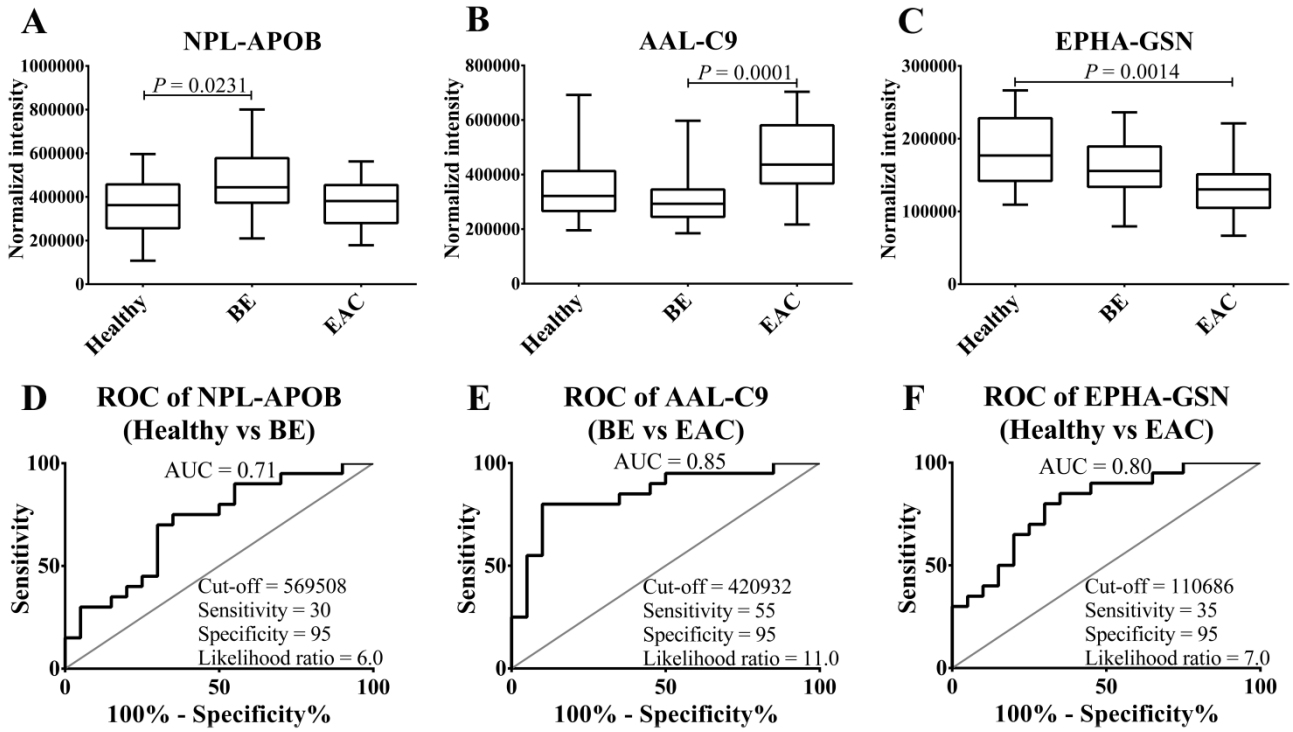


Figure 5.3. Boxplots and ROC curves of top biomarker candidate for healthy vs BE, BE vs EAC and healthy vs EAC comparison respectively. (A to C) Boxplots and **(D to F)** ROC curves of NPL-APOB, AAL-C9 and EPHA-GSN which were top biomarker candidate for healthy vs BE, BE vs EAC, and healthy vs EAC comparison, respectively.

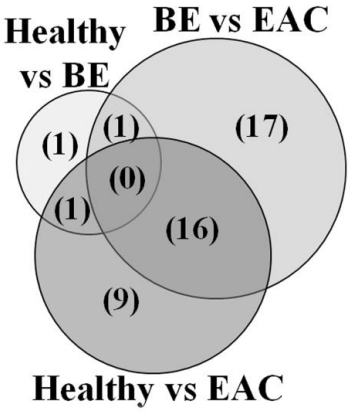


Figure 5.4. Overlap between lectin-protein candidates that differentiate BE from healthy, EAC from BE, and EAC from healthy phenotype. *P* values were calculated using Kruskal-Wallis test and $P < 0.05$ was considered to be statistically significant.

5.3.3 Comparison between LeMBA-MRM-MS and LeMBA-western immunoblotting

Orthogonal qualification at protein level using LeMBA-immunoblotting (IB) was performed for AAL-HP and AAL-GSN using samples from the qualification cohort. The relative quantitation for natural (light) peptide was performed with respect to SIS heavy labeled peptide for LeMBA-MRM-MS data. Once again, there was agreement between peptide level quantitation using MRM-MS and protein level quantitation using western immunoblotting (Figure 5.5)**,** validating the LeMBA-MRM-MS workflow [AAL-HP: MRM-MS *P* value = 0.0235, western immunoblotting *P*

value = 0.1037, MRM-MS AUROC = 0.69, western immunoblotting AUROC = 0.69; AAL-GSN: MRM-MS *P* value = 0.0120, western immunoblotting *P* value = 0.0203, MRM-MS AUROC = 0.70, western immunoblotting AUROC = 0.73].

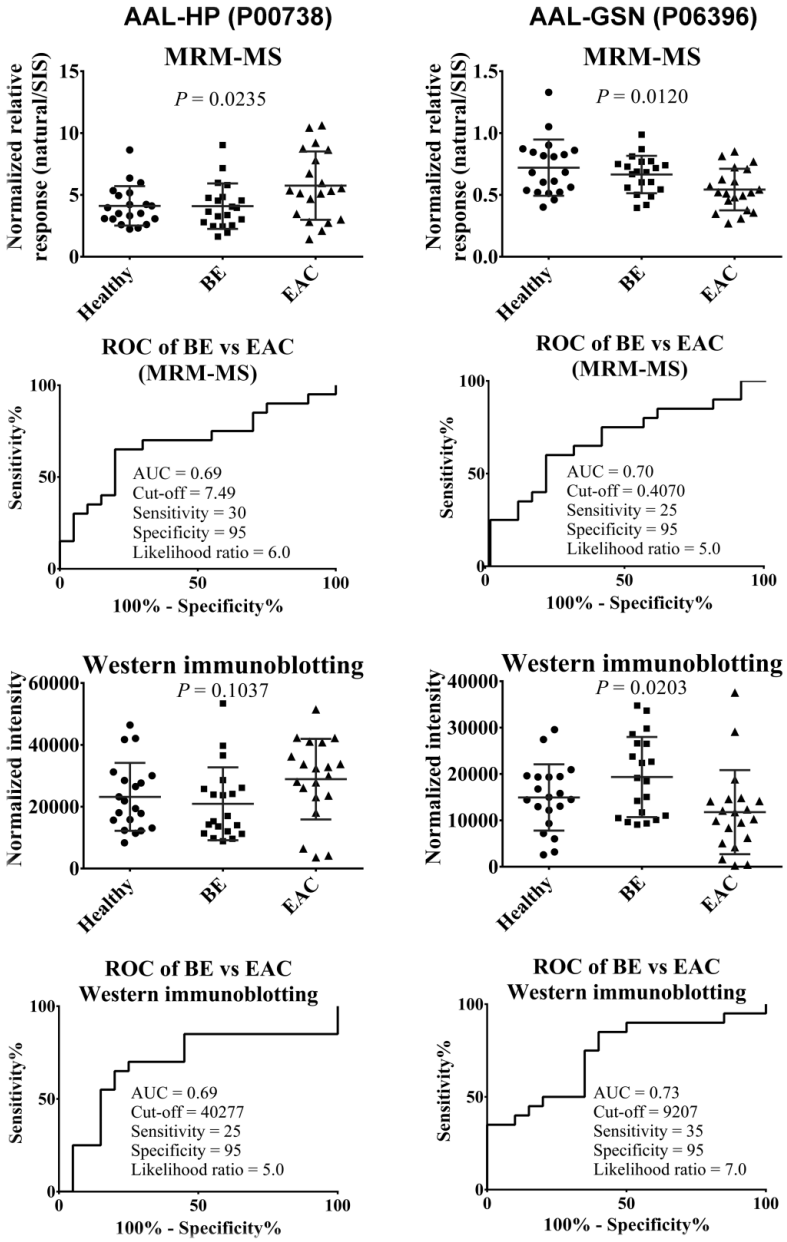


Figure 5.5. Comparison between LeMBA-MRM-MS and LeMBA-western-immunoblotting quantitation for AAL-HP (left column) and AAL-GSN (right column).

5.3.4 Identification of candidates affected by confounding covariates

As expected from the known risk factors, healthy, BE and EAC patient groups significantly differ according to BMI and reflux frequency (Table 5.2). In comparison with healthy patients, BE and EAC patient groups had a higher proportion of patients experiencing frequent GERD and of patients with obesity. Therefore, it may be possible that some of the candidates identified are due to confounding covariates rather than the actual disease phenotype. To evaluate this hypothesis, firstly

the cohort size of healthy phenotype was increased by LeMBA-MRM-MS measurement of an additional 19 serum samples collected as electoral roll control samples. These disease-free patient samples were then classified according to potential confounding variables (reflux frequency, BMI, cumulative smoking history and alcohol consumption). The statistical significance of each 45 lectin-protein candidates for each of the four covariates was assessed using a Kruskal-Wallis test (Table 5.5). Most of the candidate biomarkers were not significantly correlated with the covariates. As examples, boxplots of the data for the top 3 biomarker candidates of the disease-free cohort classified according to covariates are shown in Figure 5.6. Out of the four covariates studied, reflux frequency is perhaps the most important factor to be considered in the context of BE/EAC. Notably, none of the candidates were affected by reflux frequency, suggesting specificity of the candidates to diagnose disease phenotype. Five candidates significantly correlated with covariates (Figure 5.7). APOB showed differential binding with lectins AAL, JAC, and NPL according to BMI classification. This is most likely due to increased levels of total APOB with increase in BMI, suggesting underlying changes in the lipoprotein metabolism (452). Plasma protease C1 inhibitor (SERPING1; Uniprot entry: P05155) showed significantly reduced binding with JAC lectin in samples classified as overweight and obese as compared to healthy while JAC-alpha-1B-glycpoprotein (A1BG; Uniprot entry: P04217) varied according to alcohol consumption. This covariate analysis leads to eliminate 5 candidates from the qualified biomarker list, leaving 40 putative biomarker candidates for future studies. Out of the 5 candidates that were eliminated, JAC-APOB was identified in healthy vs BE analysis, AAL-APOB and JAC-A1BG were identified in BE vs EAC analysis, JAC-SERPING1 was identified in healthy vs EAC analysis while NPL-APOB was significantly different in heatlhy vs BE and BE vs EAC analysis. Notably, none of the 16 lectin-protein candidates that distinguish EAC from BE and healthy phenotype were identified as confounding candidates.

Table 5.5. Effect of covariates reflux frequency, BMI, cumulative smoking history and alcohol consumption on lectin-protein candidates (Boxplots and ROC curves in Appendix III).

Candidates	**Kruskal-Wallis test *P* value**			
	Reflux	**BMI**	**Smoking**	**Alcohol**
AAL-HP (P00738)	0.4452	0.6956	0.9445	0.8134
AAL-C5 (P01031)	0.659	0.269	0.9029	0.3398
AAL-C9 (P02748)	0.5822	0.4072	0.4915	0.4829
AAL-APOB (P04114)	0.6243	**0.0186**	0.8123	0.9948
AAL-GSN (P06396)	0.4123	0.8594	0.1478	0.6315
AAL-C7 (P10643)	0.6758	0.0739	0.0501	0.6401
EPHA-HP (P00738)	0.431	0.4362	0.9714	0.9552

Candidates	Kruskal-Wallis test *P* value			
	Reflux	**BMI**	**Smoking**	**Alcohol**
EPHA-SERPINA3 (P01011)	0.0579	0.2047	0.9103	0.535
EPHA-A2M (P01023)	0.397	0.5303	0.5812	0.7236
EPHA-C9 (P02748)	0.4827	0.4094	0.9293	0.4872
EPHA-AHSG (P02765)	0.2708	0.511	0.7131	0.4888
EPHA-TF (P02787)	0.563	0.452	0.9709	0.4174
EPHA-GSN (P06396)	0.5456	0.5501	0.4477	0.8568
EPHA-C7 (P10643)	0.864	0.3541	0.1908	0.3607
JAC-HP (P00738)	0.6961	0.7152	0.9893	0.9481
JAC-CFB (P00751)	0.5286	0.7001	0.4301	0.2691
JAC-SERPINA1 (P01009)	0.4182	0.5653	0.5733	0.3476
JAC-SERPINA3 (P01011)	0.2744	0.2339	0.621	0.7416
JAC-C5 (P01031)	0.9924	0.4983	0.2901	0.226
JAC-C9 (P02748)	0.4713	0.2519	0.4643	0.8035
JAC-HPX (P02790)	0.5113	0.9854	0.2554	0.4262
JAC-C4BPA (P04003)	0.3084	0.8952	0.6967	0.8589
JAC-APOB (P04114)	0.4739	**0.0125**	0.9127	0.9694
JAC-A1BG (P04217)	0.3763	0.8237	0.7572	**0.0232**
JAC-SERPING1 (P05155)	0.3262	**0.0059**	0.8609	0.5831
JAC-SERPIND1 (P05546)	0.7649	0.9412	0.6738	0.9743
JAC-GSN (P06396)	0.986	0.9915	0.2739	0.3497
JAC-C7 (P10643)	0.8268	0.1254	0.1019	0.7597
NPL-HP (P00738)	0.7445	0.5826	0.9835	0.8753
NPL-SERPINA3 (P01011)	0.4498	0.629	0.6827	0.3815
NPL-C9 (P02748)	0.6141	0.2782	0.6845	0.8707
NPL-C4BPA (P04003)	0.3259	0.9849	0.369	0.9051
NPL-APOB (P04114)	0.2792	**0.022**	0.822	0.9886
NPL-GSN (P06396)	0.9189	0.9762	0.2807	0.5319
NPL-AFM (P43652)	0.0915	0.2721	0.6962	0.694
PSA-HP (P00738)	0.6289	0.4362	0.8631	0.9842
PSA-SERPINA3 (P01011)	0.2951	0.2194	0.9087	0.2117
PSA-C5 (P01031)	0.554	0.6758	0.7538	0.3677
PSA-C9 (P02748)	0.425	0.0947	0.839	0.5174
PSA-GSN (P06396)	0.7811	0.8824	0.4982	0.4654
PSA-C7 (P10643)	0.8628	0.0744	0.0784	0.6652
WGA-HP (P00738)	0.623	0.472	0.9861	0.5777
WGA-SERPINA3 (P01011)	0.266	0.3595	0.6288	0.6383
WGA-C9 (P02748)	0.5639	0.0635	0.6448	0.7164
WGA-GSN (P06396)	0.8165	0.9591	0.4524	0.5606

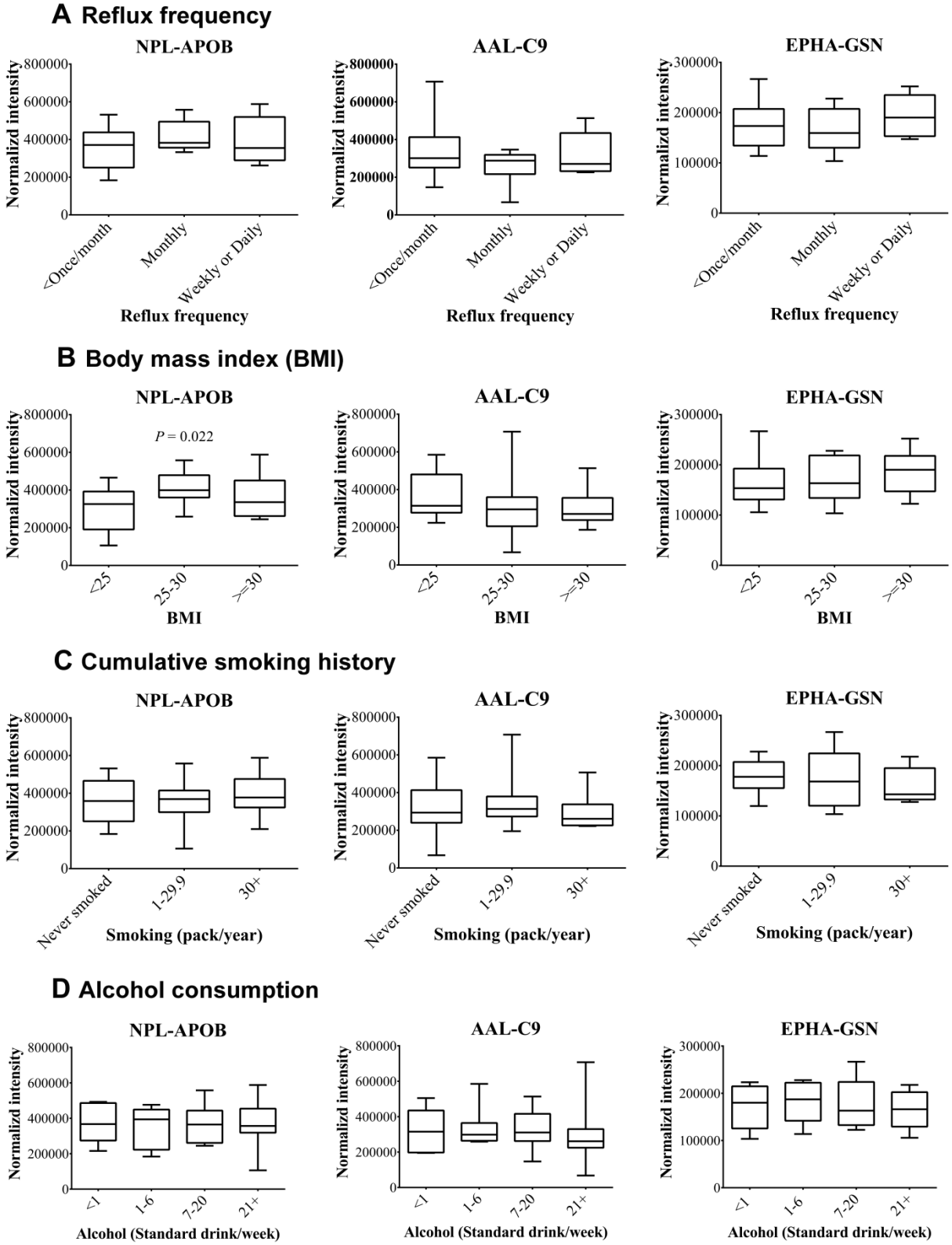


Figure 5.6. Assessing effect of confounding covariates on the top 3 biomarker candidates. Levels of NPL-APOB, AAL-C9 and EPHA-GSN were monitored in 39 serum samples (healthy-20 and population control-19) using MRM-MS. Samples were categorized according to **(A)** reflux frequency, **(B)** BMI, **(C)** smoking history and **(D)** alcohol consumption. $P < 0.05$ using Kruskal-

Wallis test was considered to be statistically significant. Out of the top 3 candidates, only NPL-APOB was significantly different according BMI categorization.

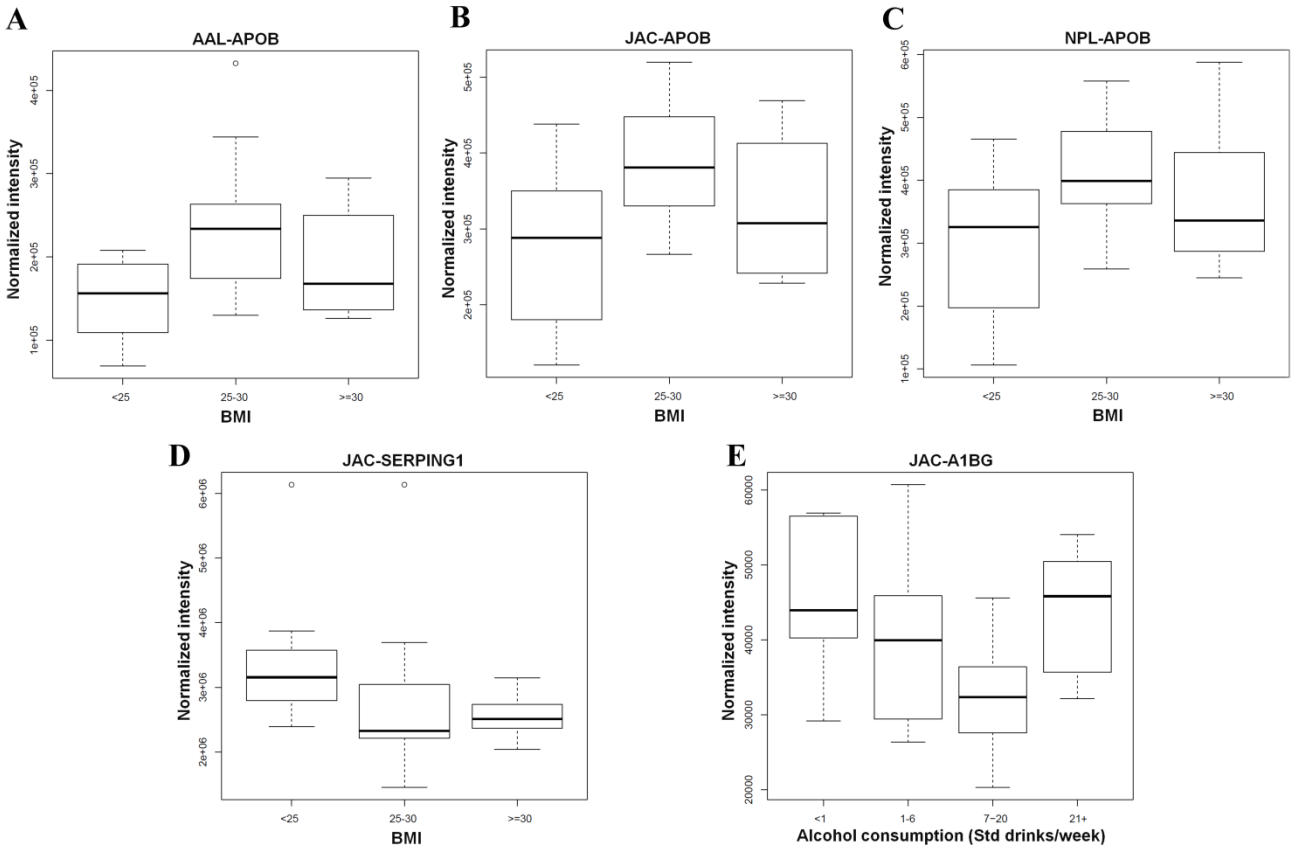


Figure 5.7. Boxplots of candidates identified as false positives according to covariates analysis. (A) AAL-APOB (*P* value = 0.0186), **(B)** JAC-APOB (*P* value = 0.0125), **(C)** NPL-APOB (*P* value = 0.022), and **(D)** JAC-SERPING1 (*P* value = 0.0059) were significantly different according to BMI status. **(E)** JAC-A1BG levels varied according to alcohol consumption (*P* value = 0.0232). *P* values are calculated using Kruskal-Wallis test.

5.3.5 Multimarker panel to distinguish EAC from BE

An individual lectin-protein biomarker candidate showed maximum AUROC of 0.85 to differentiate between EAC and BE phenotypes. For clinical requirement, a panel of biomarker with AUROC close to 1.0 is required. So next, multivariate analysis was performed to examine the potential of protein glycoforms as complementary biomarkers, focusing on differential diagnosis of EAC and BE, as the most urgent clinical need. After removal of confounding candidates, sPLS-DA was used to derive a multimarker panel that distinguish BE and EAC (Figure 5.8). The biomarker panel (BE vs EAC) included four unique proteins namely complement component C9 (C9; Uniprot entry: P02748), alpha-1B-glycoprotein (A1BG; Uniprot entry: P04217), complement C4-B (C4B; Uniprot entry: P0C0L5) and complement C2 (C2; Uniprot entry: P06681) with each of the six lectins appearing at least once in the panel. Using 5-fold cross-validation repeated 1000 times on

this multimarker panel, the model showed cross-validation error rate of 37.47% and moderate separation of the BE and EAC sample representations (Figure 5.8A). The combined signature of the eight candidates gave an AUROC of 0.9425 with 95% specificity and 80% sensitivity (Figure 5.8B).

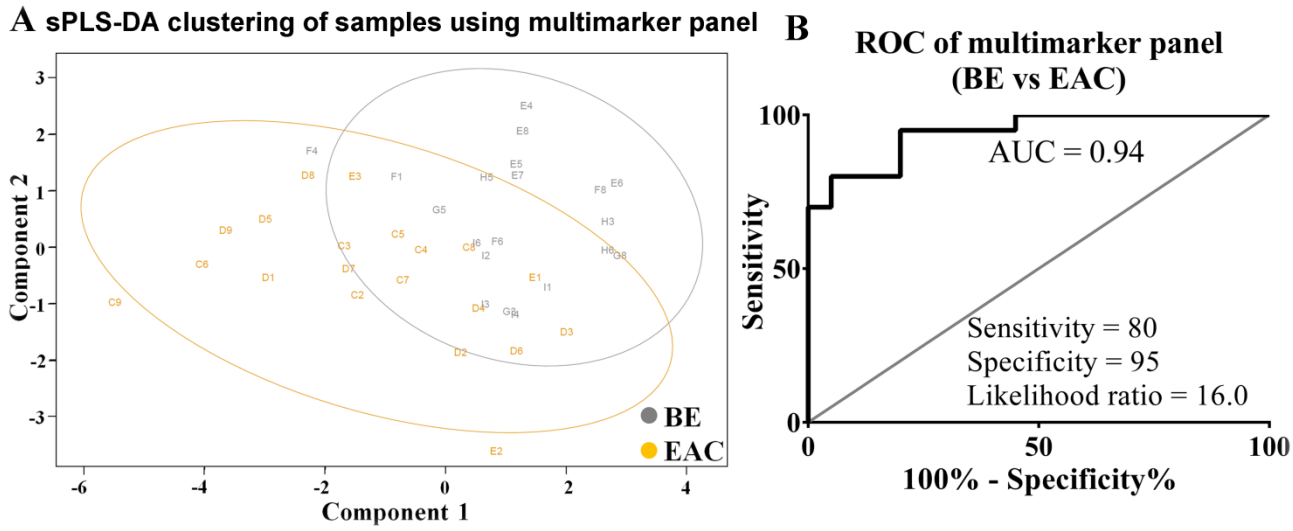


Figure 5.8. Multimarker panel to distinguish EAC from BE. (A) sPLS-DA and **(B)** ROC curve analysis of a multimarker panel consists of AAL-C9, EPHA-A1BG, EPHA-C9, JAC-C9, NPL-C2, NPL-C4B, PSA-C9, and WGA-C9.

5.4 Discussion

5.4.1 Overview

The results in this chapter describe verification of lectin-protein diagnostic biomarker candidates in an independent patient cohort. LeMBA-MRM-MS-Shiny mixOmics workflow has been utilized for biomarker verification. Candidates affected by confounding covariates were also identified. Multivariate analysis revealed a biomarker panel with very high diagnostic potential to distinguish between EAC and BE.

5.4.2 Quality check for biomarker verification dataset

For any sophisticated analytical instrument such as mass spectrometer, it is important to perform enough quality checks to consistently monitor its performance while running the samples. The triple quadrupole mass spectrometer used for the targeted proteomics approach can isolate/quantify peptide/fragment ions up to 1400 *m/z* ratio. All the precursor ions in the MRM-MS assay were below 1000 *m/z* ratio while some of the fragment ions were between 1000 *m/z* to 1400 *m/z*. So collectively, precursor and fragment ions cover entire mass range of the instrument. This means that QC sample should ideally consist of mixture of peptides (or small molecules) to cover

the entire mass range to allow holistic evaluation of the instrument performance. Commercially available HSA synthetic peptide mix consist of peptides which can cover this entire mass range hence it was used as a QC and showed very consistent response throughout.

The internal standard ovalbumin and corresponding SIS labeled peptide showed minimal possible variation for such a high-throughput, sensitive and multiplexed workflow (422). In addition, correlation analysis was performed amongst the peptide responses from the same protein to allow better quantitation. In line with the observation made by Domanski and colleagues (453), absolute response for individual peptide from the same protein varied significantly. Despite these large differences between peptide responses, 85% of the lectin-protein candidates showed Pearson correlation co-efficient of more than 0.8 amongst peptides emerging from the same protein suggesting reproducibility in the trypsin digestion across the entire assay. Furthermore, western immunoblotting quantitation gave comparable results with mass spectrometric quantitation which further validate the quantitation approached employed here. Taken together, the biomarker verification data is of sufficiently high quality to provide reliable results.

5.4.3 Differential glycosylation in EAC

Univariate analysis using Kruskal-Wallis tests discovered lectin-protein candidates between the three phenotypes compared. It is evident from the results that EAC phenotype is associated with glycosylation changes as many serum proteins showed differential lectin binding between EAC vs healthy/BE phenotype. Only three candidates showed statistically significant difference between healthy and BE patient groups. Overall the results raise the possibility of major glycosylation changes taking place during progression of BE to EAC but not from healthy to BE. The findings from genomic sequencing studies concluded that except few key mutations, the majority of mutations observed in EAC lesions are already present in BE condition suggesting common mutations are acquired early during the pathogenesis. At this stage these two results seem to contrast each other raising possibility that more functional level changes [e.g. gene/protein expression (241, 256, 454-456), protein glycosylation, metabolic changes etc.] driven by environmental factors and/or early genomic alterations to be associated with development of dysplasia/carcinoma from metaplastic condition. In line with this, studies have shown differential expression of glycan structures in tissue and serum samples during metaplasia-dysplasia-carcinoma sequence (167-170, 174-180, 457).

5.4.4 Confounding covariates

Ideally, biomarker discovery and development should be performed using human biological materials which differ according to disease of interest and matched for all other confounding covariates (356, 458). Practically, it is very difficult to match the cohorts being compared across all the co-morbidities. In addition, some of the confounding covariates such as GERD and obesity are risk factors for BE/EAC hence will always be present. So it is inevitable that some of the candidates verified are actually due to confounding covariates and not associated with the disease phenotypes. It is very important that these false positive candidates are identified early during the biomarker development. In this chapter, 5 such false positive candidates were identified by screening additional serum samples from population control. The population control patients were not endoscopically confirmed negative for columnar lined esophagus i.e. BE, but it was assumed that all the population control samples, otherwise healthy, are negative for BE as its prevalence is low (35). Identification of APOB as a false positive candidate affected by BMI further validate the results of biomarker verification as it is well known that APOB level is associated with obesity/BMI (452).

5.4.5 Multivariate analysis

Preferably, biomarker verification datasets should be divided into training and test sets. The biomarker panel identified using the training set should be tested using the independent test dataset to accurately estimate the error-rates and performance of the biomarker panel (459, 460). Like the biomarker verification study described here, it is not always possible to get enough number of patient samples to categorize into two separate datasets namely test and training. In this scenario, another popular validation method is n-fold cross-validation (461-463). In cross-validation, the entire verification dataset is randomly divided into n mutually exclusive equal subsets and the model is trained multiple times (100s or even 1000s) each time keeping one of the folds out as an validation set (386). This method is useful in small sample size because the entire data are used both as training and validation purposes without wasting of the data in a holdout validation set (386). In comparison to existing approaches (statistical methods and machine learning workflows) for selection of biomarkers, sPLS-DA is superior in terms of graphical outputs hence offer easy interpretation of the results and it has demonstrated similar classification performance to other workflows (382). So collectively, sPLS-DA was employed for multivariate analysis with 5-fold cross-validation. The multimarker panel showed AUROC of 0.94 showing improved performance of a combination of biomarkers over individual lectin-protein candidates.

In conclusion, LeMBA-MRM-MS-Shiny mixOmics biomarker verification pipeline was able to successfully verify the candidates identified using LeMBA-GlycoSlector biomarker discovery pipeline. The diagnostic ability of an individual lectin-protein biomarker candidate was determined. The panel of biomarker showed very high diagnostic potential to differentiate between EAC and BE. The biomarkers verified require further evaluation in an independent large patient cohort including dysplastic samples.

Chapter 6.

ELECTROCHEMICAL DETECTION OF GLYCAN AND PROTEIN EPITOPES OF GLYCOPROTEINS IN SERUM

Chapter 6. Electrochemical detection of glycan and protein epitopes of glycoproteins in serum

6.1 Manuscript information

Improvements in the lectin/non-lectin based glycoprotein enrichment methodologies along with rapid technological advancements in the chromatographic and mass spectrometry platforms led to identification of novel glycoprotein biomarker candidates for various diseases in the recent past including the work described so far in the thesis. However, these workflows are poorly suited for routine clinical use due to high complexity and cost. Alternatively, electrochemical detection methods can achieve rapid, cost-effective, sensitive, selective and accurate quantification of biomolecules (464-468). In fact commercially available glucose biosensors use the principle of electrochemical detection (468-470). Lectin based biosensors developed so far can monitor overall changes in the glycosylation profile but do not give any information about underlying glycoprotein to which glycan is attached (471-474). To increase the specificity and diagnostic applicability of lectin based biosensors, we developed label-free electrochemical detection method to monitor specific glycan event on a target glycoprotein (2). This chapter describes proof-of-concept study for interrogating glycan epitope of a model glycoprotein chicken ovalbumin using *Sambucus nigra* agglutinin (SNA lectin) and protein epitope by anti-ovalbumin antibody followed with label-free electrochemical detection in the background of diluted human serum. Chicken ovalbumin mainly express N-linked high mannose and hybrid type glycan (475, 476), with expression of terminal sialic acid modified complex glycan at very low levels (477, 478). SNA lectin used in the assay preferentially binds with α2-6 linked sialic acid residues (479). So only a fraction of total ovalbumin is anticipated to bind with SNA lectin. The ovalbumin and SNA lectin were chosen for this proof-of-concept study to test the ability of electrochemical detection to monitor minimal interaction between ovalbumin and SNA lectin.

The following manuscript was published in Analyst (2014), volume 139, issue 22, pages 5970-5976. The candidate and first author on the paper, Alok K. Shah, was mainly responsible for conducting the research and writing the manuscript. The detailed contributions from co-authors namely Michelle M. Hill, Muhammad J. A. Shiddiky and Matt Trau is listed on page vii & viii of the thesis.

Analyst

PAPER

Cite this: *Analyst*, 2014, **139**, 5970

Electrochemical detection of glycan and protein epitopes of glycoproteins in serum

Alok K. Shah,[a] Michelle M. Hill,[a] Muhammad J. A. Shiddiky*[b] and Matt Trau*[b]

Received 2nd May 2014
Accepted 29th August 2014

DOI: 10.1039/c4an00781f

www.rsc.org/analyst

Aberrant protein glycosylation is associated with a range of pathological conditions including cancer and possesses diagnostic importance. Translation of glycoprotein biomarkers will be facilitated by the development of a rapid and sensitive analytical platform that simultaneously interrogates both the glycan and protein epitopes of glycoproteins in body fluids such as serum or saliva. To this end, we developed an electrochemical biosensor based on the immobilization of a lectin on the gold electrode surface to recognize/capture a target glycan epitope conjugated to glycoproteins, followed by detection of the protein epitope using a target protein-specific antibody. Electrochemical signals are generated by label-free voltammetric or impedimetric interrogation of a ferro/ferricyanide redox couple (*e.g.* $[Fe(CN)_6]^{3-/4-}$) on the sensing surface, where the change in voltammetric current or interfacial electron transfer resistance was measured. The detection system was demonstrated using the model glycoprotein chicken ovalbumin with *Sambucus nigra* agglutinin type I (SNA lectin), and exhibits femtomolar sensitivity in the background of diluted human serum. The results obtained in this proof-of-concept study demonstrate the possibility of using electrochemical detection for developing cheap point-of-care diagnostics with high specificity and sensitivity for blood glycoprotein biomarkers.

Introduction

Differential protein glycosylation is a feature of diverse pathological conditions such as atherosclerosis, ulcerative colitis, rheumatoid arthritis, microbial infection, Alzheimer's disease and cancer.[1] Monitoring specific glycosylation changes for glycoprotein biomarker candidates such as fucosylated haptoglobin (pancreatic cancer) and sialylated prostate-specific antigen (prostate cancer) provides higher diagnostic power compared to changes in total glycoprotein levels.[2] In fact, measurement of fucosylated α-fetoprotein in blood (AFP-L3 test) has been approved by the FDA for early detection of hepatocellular carcinoma.[3,4] Over the past few years, various glycoprotein enrichment platforms have been coupled with mass spectrometric and nuclear magnetic resonance spectroscopic techniques to uncover disease specific glycosylation changes for candidate glycoprotein biomarkers. These include single, serial and multi-lectin affinity chromatography,[5,6] lectin magnetic bead array (LeMBA),[7,8] boronic acid[9] and hydrazide chemistry[10] based extraction methods. While all these methods have excellent analytical performance in detecting candidate cancer glyco-biomarkers, they are poorly suited for routine clinical use due to high maintenance/running cost, requirement of technical expertise, long assay time and complicated data analysis procedures. The development of detection methods that are user-friendly, robust, sensitive, quick and cheaper than currently available methods is warranted to fulfill important needs for developing future diagnostics.

Electrochemical (EC) detection methods offer elegant ways for interfacing bio-recognition and transduction events and represent substantial drivers to achieve rapid, cost-effective, sensitive, selective and accurate quantification of biomolecules.[11–15] Among many EC methods, faradaic electrochemical impedance spectroscopy (F-EIS) is one of the most effective methods for the label-free detection of biomolecules and for probing the build-up of the biomaterial sensing film on the electrodes.[16–18] In F-EIS, the binding of a target protein with its ligand (*i.e.*, a specific antibody) on the electrode surface can be detected, where the change in impedance of the electrode surface and its interface to the electrolyte solution containing a redox probe (*e.g.* $[Fe(CN)_6]^{3-/4-}$) is measured in the form of its electron transfer resistance (R_{et}).[19–21] This interfacial electron transfer reaction of the redox process can also be measured *via* a voltammetric technique, where the presence or absence of the target proteins will alter the voltammetric current of the redox process at the sensing surface. More recently, the differential pulse voltammetric (DPV) interrogation of the interfacial electron transfer reaction of the $[Fe(CN)_6]^{3-/4-}$ process generated upon protein binding has been used as an effective label-free tool for protein detection.[22,23]

[a]*The University of Queensland Diamantina Institute, The University of Queensland, Translational Research Institute, QLD 4102, Australia. E-mail: m.shiddiky@uq.edu.au; m.trau@uq.edu.au; Fax: +61-7-33463973; Tel: +61-7-33464178*

[b]*Australian Institute for Bioengineering and Nanotechnology (AIBN), Corner College and Cooper Roads (Bldg 75), The University of Queensland, QLD 4072, Australia*

In the past few years, much attention has been focused on development of label-free electrochemical biosensors to monitor glycosylation changes (*e.g.* mannosylation, galactosylation, sialylation and fucosylation) in complex biological fluids[24–27] using naturally occurring lectin as a glycan recognition element. These biosensors detect overall changes in the glycan profile without monitoring a specific glycoprotein to which glycan is attached. To increase the specificity and diagnostic value, analytical methods are required to detect not only the overall status of glycosylation in the sample but also aberrant glycosylation for a specific glycoprotein. To overcome this limitation, we investigated label-free electrochemical detection methods for the simultaneous interrogation of specific glycan on a target glycoprotein.

Here we describe the development of a simple method that can efficiently and reproducibly detect both glycan and protein epitopes of a glycoprotein in the background of a diluted serum sample at a femtomolar concentration. This method consists of a gold macrodisk electrode that is biochemically functionalized with *Sambucus nigra* agglutinin type I lectin (SNA lectin) [specific to recognize terminal sialic acid attached to galactose through α-(2-6) linkage].[28] The attachment of the target glycoprotein chicken egg albumin (ovalbumin) on the SNA lectin functionalized electrode and complexation with a polyclonal anti-ovalbumin antibody (Fig. 1 scheme) were followed by the F-EIS and DPV measurements. Each bimolecular layer on the sensor surface acts as a barrier for the interfacial electron transfer reaction of the $[Fe(CN)_6]^{3-/4-}$ process, resulting in an increase in R_{ct} or a decrease in DPV current response. The presence of target glycoprotein ovalbumin and subsequent complexation with the polyclonal anti-ovalbumin antibody appear to further block the $[Fe(CN)_6]^{3-/4-}$ process from accessing the electrode surface effectively. When we monitored R_{ct} or DPV current responses generated from the $[Fe(CN)_6]^{3-/4-}$ process before and after ovalbumin capture, there was a clear correlation between the presence of the target ovalbumin and changes in R_{ct} or DPV current response. To mimic clinical scenario, we spiked ovalbumin into serum and showed linear changes in current readout with a concentration range from 10 pg mL^{-1} to 500 pg mL^{-1}. To the best of our knowledge, this is the first electrochemical method that simultaneously interrogates specific glycans and the target protein on which the glycan is attached. Moreover, we believe that the simplicity of this technology could facilitate the translation of the current glycoprotein biomarker research.

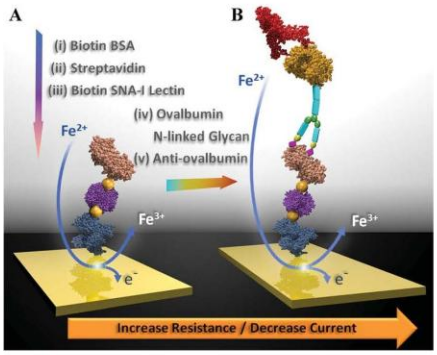


Fig. 1 (A) Schematic illustration of the preparation of an SNA lectin immunosensing layer. (B) Schematic view of glycoprotein ovalbumin capture and label-free detection using an anti-ovalbumin antibody which results in increased charge transfer resistance (R_{ct}) and a corresponding decrease in DPV peak current.

Materials and methods

Materials

Biotin labeled *Sambucus nigra* agglutinin type I (SNA-I) lectin was purchased from Vector Laboratories (USA). Lyophilized chicken egg albumin (ovalbumin) (#A5503), polyclonal anti-chicken egg albumin antibody produced in rabbit (#C6534), potassium ferrocyanide, potassium ferricyanide, and KCl were purchased from Sigma (Australia). Biotinylated BSA was purchased from Thermo Scientific (USA) while multivalent streptavidin was purchased from Invitrogen (USA). Human blood sample was collected from a healthy volunteer with consent and ethics approved by the University of Queensland Human Ethics Committee.

Cleaning gold disk electrodes

Gold disk working electrodes (diameter = 3 mm) were purchased from CH Instruments (Austin, USA). The electrodes were reused after cleaning and regeneration. The electrodes were cleaned using piranha ($H_2SO_4 : H_2O_2 = 3 : 1$) on a sonicator water-bath for 30 s to 1 min (note: piranha is a highly toxic and hazardous chemical. It has to be handled in a fume-hood with adequate personal protective equipment. Universal safety guidelines should be followed for its disposal). The electrodes were then physically polished with 1 micron alumina and subsequently with a 0.05 micron alumina slurry. Prior to electrochemical cleaning, electrodes were sonicated in acetone for 20 min. Electrochemical cleaning was performed in 0.5 M H_2SO_4 until reproducible characteristic gold electrode profiles were achieved.

Construction of biosensor

Immediately before functionalization, thoroughly cleaned gold disk electrodes were dried under a flow of nitrogen gas. The electrodes were incubated with 500 μg mL^{-1} biotin-BSA in 1× PBS (137 mM NaCl, 2 mM KCl, 10 mM phosphate buffer, pH 7.4) on a thermo shaker set at 25 °C, 300 rpm for 45 min. The electrodes were washed with 1× PBS after each incubation step 3 times. The electrodes were then incubated with multivalent streptavidin (500 μg mL^{-1} in 1× PBS, at 25 °C, 300 rpm for 45 min), followed by biotinylated SNA lectin (500 μg mL^{-1} in 150 mM NaCl containing 0.1 mM Ca^{2+}, at 25 °C, 300 rpm for 45 min) to form the bio-recognition layer of lectin on the electrode surface. Ovalbumin stock solution was made at a concentration of 2 mg mL^{-1} and stored at −30 °C in aliquots. The designated concentration of ovalbumin was freshly made (in 150 mM NaCl

solution containing 0.1 mM Ca^{2+}) prior to use and protein capture was performed for 45 min by gentle agitation using an intelli-mixer (PCOD Scientific). 1 in 1000 dilution of the polyclonal anti-ovalbumin antibody in 1× PBS for 15 min was used for detection of captured chicken egg albumin. Designated concentrations of ovalbumin were spiked in diluted serum (1 in 1000 dilution of serum in 150 mM NaCl solution containing 0.1 mM Ca^{2+}).

After modification of the sensor surfaces (up to SNA lectin attachment), the actual detection steps include capturing the ovalbumin antigen followed by anti-ovalbumin detection. Both these steps are directly followed by the label-free impedance (or DPV) technique. Thus once the electrode is modified with the SNA lectin, the overall detection process takes approximately 2 h.

Electrochemical measurement

All electrochemical experiments were conducted at room temperature (25 ± 1 °C) in a standard three-electrode electrochemical cell arrangement using an electrochemical workstation CHI 650D, CH Instruments (Austin, USA). The electrochemical cell consisted of a gold disk electrode sensor as a working electrode, a platinum wire counter electrode and a Ag/AgCl (3 M KCl) reference electrode. Electrochemical signals were measured in 1× PBS buffer (pH 7.4) containing 2.5 mM $[Fe(CN)_6]^{3-}$/$[Fe(CN)_6]^{4-}$ (1 : 1) and 0.1 M KCl. DP signals were obtained with a potential step of 4 mV, a pulse amplitude of 50 mV, a pulse width of 200 ms, a sampling width of 16.7 ms and a pulse period of 500 ms. The F-EIS spectra were recorded in 1× PBS buffer (pH 7.4) containing 2.5 mM $[Fe(CN)_6]^{3-}$/$[Fe(CN)_6]^{4-}$ (1 : 1) and 0.1 M KCl using an alternating current voltage of 5 mV, with the frequency range of 0.1 (or 1) Hz to 100 kHz. During the run, the bias DC current was applied below a frequency of 100 Hz. The faradaic current generated by the $K_3[Fe(CN)_6]$/$K_4[Fe(CN)_6]$ probe accounts for the presence of a biomolecule. The current changes corresponding to ovalbumin detection were calculated as follows:

$$\%\ \text{Decrease of DPV peak current} = (I_{\text{before}} - I_{\text{after}}) \times 100/I_{\text{before}} \quad (1)$$

where I_{before} = peak current at the ovalbumin capture step [*e.g.*, current recorded at biotin-BSA/multivalent streptavidin/biotinylated SNA lectin/designated concentration of ovalbumin (either in buffer or spiked into 1 in 1000 serum)] and I_{after} = peak current at the detection step [*e.g.*, current recorded at biotin-BSA/multivalent streptavidin/biotinylated SNA lectin/designated concentration of ovalbumin (either in buffer or spiked into 1 in 1000 serum)/1 in 1000 anti-ovalbumin polyclonal antibody].

For each electrode, the peak current was normalized with a DPV response obtained at the initial step when electrodes were clean.

Results and Discussion

Construction of biosensor

Fig. 1 describes an assembly scheme of the biosensor. As reported previously,[18] biotin labeled BSA is used for coating the gold electrode surface. Subsequently, multivalent streptavidin was used as a linker to immobilize the biotin labeled SNA lectin to form a bio-recognition layer on the electrode surface. The formation of the bio-recognition layer on the electrode surface affects analytical performance of the biosensor and in our experiments this formation is controlled by attachment of biotin-BSA, streptavidin and biotinylated SNA lectin. To achieve maximal analytical performance, we optimized binding conditions for all three biomolecules one after another by incubating electrodes with three different concentrations of biotin-BSA, multivalent streptavidin and biotinylated SNA lectin.

To determine the optimal biotin-BSA concentration at first, we incubated thoroughly cleaned gold electrodes with 100 μg mL^{-1}, 500 μg mL^{-1} and 1000 μg mL^{-1} solution of biotin-BSA for 45 min. The electrodes were washed three times with 1× PBS, followed by F-EIS measurements to determine the optimal biotin-BSA concentration. As compared to 100 μg mL^{-1}, incubation of electrodes with 500 μg mL^{-1} concentration of biotin-BSA showed an increase in size of the semicircle on the Nyquist plot (Fig. 2A, i *vs.* ii) suggesting an increased electron transfer resistance (R_{ct}). There was no further increase in R_{ct} at 1000 μg mL^{-1} biotin-BSA concentration as compared to 500 μg mL^{-1}

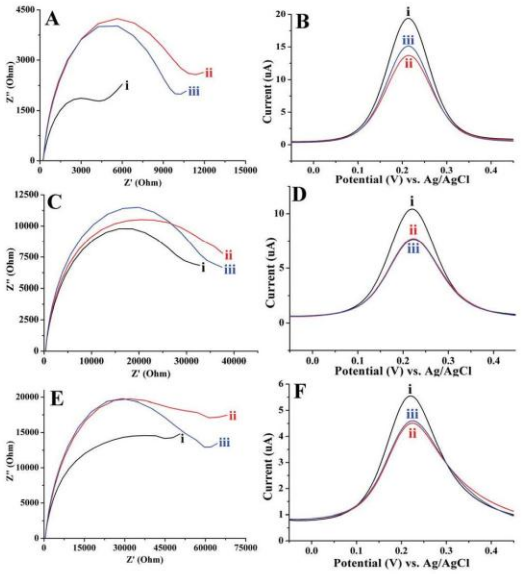


Fig. 2 Optimization of biosensor construction. Nyquist plots of a gold electrode modified with (A) biotin-BSA, (C) biotin-BSA/streptavidin, and (E) biotin-BSA/streptavidin/SNA lectin in 1× PBS buffer containing 2.5 mM $K_3[Fe(CN)_6]$, 2.5 mM $K_4[Fe(CN)_6]\cdot 3H_2O$ and 0.1 M KCl. (B, D, and F) Corresponding DPV responses obtained at a gold electrode modified with (B) biotin-BSA, (D) biotin-BSA/streptavidin, and (F) biotin-BSA/streptavidin/SNA lectin, respectively. Concentrations of biotin-BSA in Fig. A & B, multivalent streptavidin in Fig. C & D (following incubation with 500 μg mL^{-1} of biotin-BSA), and SNA lectin in Fig. E & F (following incubation with 500 μg mL^{-1} biotin-BSA and 500 μg mL^{-1} streptavidin each) were (i) 100 μg mL^{-1}, (ii) 500 μg mL^{-1}, and (iii) 1000 μg mL^{-1}.

(Fig. 2A, ii *vs.* iii). As 500 μg mL^{-1} biotin-BSA showed best F-EIS *i.e.* maximum R_{ct} response, the same binding condition was used for subsequent experiments. Next, F-EIS for three different concentrations 100 μg mL^{-1}, 500 μg mL^{-1} and 1000 μg mL^{-1} of multivalent streptavidin under optimal biotin-BSA binding conditions was obtained. As shown in Fig. 2C, incubation of the electrode with a 500 μg mL^{-1} concentration of multivalent streptavidin showed a maximum R_{ct} under optimal biotin-BSA binding conditions. Similarly, as shown in Figure 2E, 500 μg mL^{-1} of biotinylated SNA lectin showed the best F-EIS response under optimal biotin-BSA and multivalent streptavidin binding conditions. To confirm F-EIS results we conducted parallel DPV measurements and as shown in Fig. 2B, D and F, there was a decrease in peak DPV current corresponding to an increase in R_{ct} suggesting reduction in actual electron transfer between the electrode/redox electrolyte double layer with increased resistance. For all three biomolecules, the maximum R_{ct} and minimal DPV peak current was observed at 500 μg mL^{-1} concentration. Moreover, there was no further increase in R_{ct} at a higher concentration *i.e.* 1000 μg mL^{-1}. This suggests that the electrode surface is getting saturated when incubated with 500 μg mL^{-1} concentration of biomolecules and there is no non-specific binding of the protein directly on the electrode surface at a higher concentration. As the protein binding to the electrode surface is showing saturation without any non-specific binding, it is anticipated that there will be minimal or no non-specific binding of biomolecule of interest directly to the electrode surface during further stages of the experiment.

Glycoprotein capture and detection

After optimizing conditions for SNA lectin immobilization on the electrode surface, next we captured glycoprotein ovalbumin in 150 mM NaCl solution containing 0.1 mM Ca^{2+}. The formation of bio-recognition layers and capture of ovalbumin was followed by electrochemical detection. As shown in Fig. 3A, the biotin-BSA coated electrode gave rise to the smallest semicircle (Fig. 3Ai), followed by streptavidin (Fig. 3Aii), SNA lectin (Fig. 3Aiii) and ovalbumin (Fig. 3Aiv) which showed a gradual increase in semicircle indicating an increase in electron transfer resistance. In line with changes observed in impedance, we observed a corresponding decrease in peak DPV current (Fig. 3B). These results confirm the results shown in a previous section for successful formation of bio-recognition layers on the electrode surface. Next, it demonstrates successful capture of target glycoprotein ovalbumin on a SNA biosensor with electrochemical detection.

Lectins generally recognize glycan structures with low affinity but with high avidity mainly through hydrogen bonding, hydrophobic interactions and van der Waals forces.[29] In fact, glycan structure specific interactions of a glycoprotein with lectins are very well demonstrated in the recent past mainly using sialic acid binding lectin biosensors.[25–27,30] Incubation of the SNA lectin biosensor with asialofetuin, a non-sialic acid expressing glycoform variant of glycoprotein fetuin, showed very minimal change in the baseline impedance.[25–27,30] In contrast, a linear increase from baseline impedance was

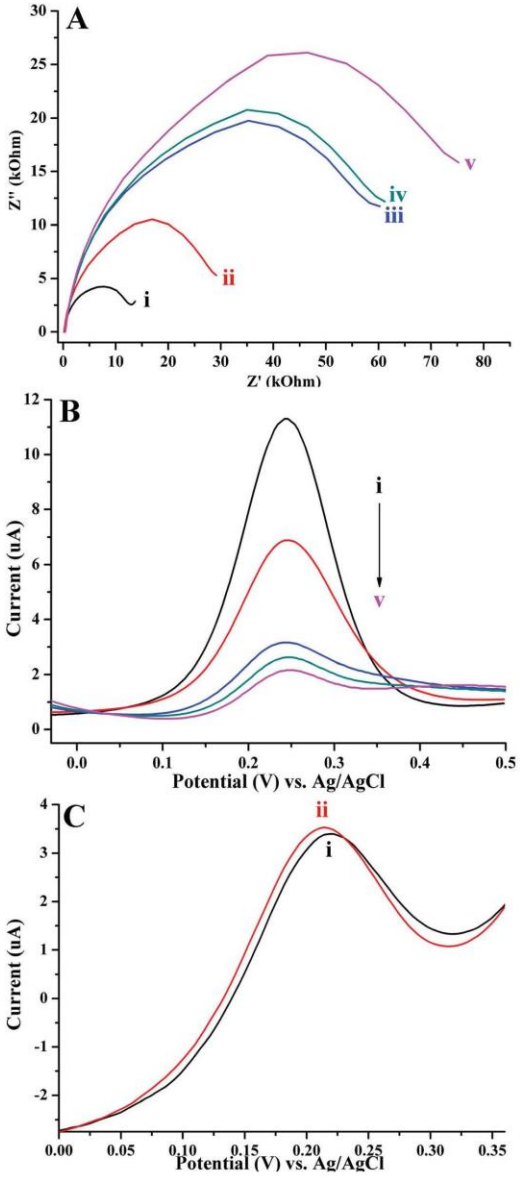


Fig. 3 (A) Nyquist plots and (B) differential pulse voltammetric responses for monitoring (i) biotin-BSA (ii) biotin-BSA/multivalent streptavidin (iii) biotin-BSA/multivalent streptavidin/SNA lectin (iv) biotin-BSA/multivalent streptavidin/SNA lectin/1 ng mL^{-1} ovalbumin and (v) biotin-BSA/multivalent streptavidin/SNA lectin/1 ng mL^{-1} ovalbumin/anti-ovalbumin antibody in 1× PBS buffer containing 2.5 mM $K_3[Fe(CN)_6]$, 2.5 mM $K_4[Fe(CN)_6]\cdot 3H_2O$ and 0.1 M KCl. (C) Differential pulse voltammetric responses to detect any non-specific binding of anti-ovalbumin antibody with SNA lectin (i) biotin-BSA/multivalent streptavidin/SNA lectin and (ii) biotin-BSA/multivalent streptavidin/SNA lectin/anti-ovalbumin antibody.

observed when the SNA lectin biosensor was incubated with an increasing concentration of fetuin expressing a correct glycan epitope.[25–27,30] Apart from showing very high binding specificity towards the glycan epitope, these SNA lectin biosensors have demonstrated sensitivity for glycoprotein detection (*e.g.* fetuin) up to attomolar concentration.[25–27,30] SNA along with other electrochemical lectin biosensors developed so far has been applied to glycan profiling of healthy and rheumatoid arthritis human serum samples,[27] monitor glycosylation changes between normal and cancerous pancreatic cell extracts,[25] dengue diagnosis[31] and microorganism recognition,[32] without detecting specific glycoproteins. Detection of both glycan and protein epitopes of a glycoprotein is required to provide the specificity to measure the glycosylation-specific biomarker, for example fucosylated alpha-fetoprotein. Therefore, following the capture of ovalbumin glycoprotein using a SNA lectin biosensor, we interrogated the ovalbumin protein backbone using a polyclonal anti-ovalbumin antibody. Upon capturing ovalbumin on a SNA biosensor, electrodes were incubated with 1 in 1000 dilution of anti-ovalbumin antibody, followed by F-EIS and DPV measurements. As compared to the ovalbumin capture step (Fig. 3Aiv) incubation of the electrode with an anti-ovalbumin antibody (Fig. 3Av) showed an increase in impedance, with a corresponding decrease in DPV peak current (Fig. 3B). These results indicate stepwise construction of the bio-recognition layer, capture of ovalbumin on the SNA biosensor and detection of ovalbumin using a polyclonal anti-ovalbumin antibody. As F-EIS and DPV techniques showed an excellent agreement between the results obtained for stepwise binding of biomolecules on the sensor surface, we presented only DPV responses for remaining all experiments.

As antibodies are glycoproteins themselves, they may directly interact with lectins which may lead to false positive results. Therefore, contribution of a false positive response due to direct interaction between the SNA lectin and a polyclonal anti-ovalbumin antibody was tested. The gold electrode modified with biotin-BSA/streptavidin/SNA lectin was directly incubated with an anti-ovalbumin antibody to allow any possible direct interaction between them. As shown in Fig. 3C we did not observe any change in the DPV response between the biotin-BSA/streptavidin/SNA lectin and the biotin-BSA/streptavidin/SNA lectin incubated with an anti-ovalbumin antibody. This finding indicates that there was no direct interaction between the ovalbumin antibody with the SNA lectin.

Thus, we demonstrate stepwise formation of a bio-recognition layer; capture of glycoprotein ovalbumin on the SNA biosensor surface followed by detection of ovalbumin using an anti-ovalbumin antibody and label-free electrochemical detection using F-EIS and DPV measurements. Moreover, we did not observe any direct interaction between the lectin and the antibody which excludes possibility for false positive results. To the best of our knowledge, this is the first electrochemical biosensor which interrogates both the glycan epitope of a glycoprotein using a lectin and protein epitope using an anti-protein antibody with label-free electrochemical monitoring.

Detection of ovalbumin in serum

To evaluate the biosensor performance under more physiological conditions, we conducted a series of experiments in which ovalbumin was spiked into diluted human serum collected from a healthy donor. To mimic biomarker detection conditions, ovalbumin at a final concentration of 1 ng mL $^{-1}$ was spiked into the 1 : 1000 serum sample. Similar dilution of the serum sample has been recommended to study the overall changes in glycan profiles using electrochemical detection.[27] Blood is very heterogeneous in terms of its protein composition, consisting of proteins with abundance varying by several orders of magnitude. Top six abundant plasma proteins account for 85% of the total plasma proteome.[33] Out of them, five are found to be glycosylated (transferrin, haptoglobin, α1-antitrypsin, IgG, and IgA) and show varying degree of binding to the SNA lectin. This means that one would expect a significant amount of glycoprotein capture when the electrode immobilized with the SNA lectin is incubated with the serum/plasma sample. As can be seen in Fig. 4A, the DPV peak current was significantly reduced when the SNA lectin modified electrode was incubated with the serum sample (Fig. 4A, i *vs.* ii). However, there was no further decrease in the DPV response with anti-ovalbumin antibody incubation suggesting that the antibody did not bind non-specifically with any SNA lectin-bound human serum proteins (Fig. 4A, ii *vs.* iii).

When ovalbumin was spiked into serum, incubation of the electrode with the polyclonal anti-ovalbumin antibody (Fig. 4Bii) showed significant reduction in the peak DPV response as compared to the DPV peak current resulted at the ovalbumin capture step in serum (Fig. 4Bi). These data clearly demonstrate that our method is highly effective for specific detection of ovalbumin glycoprotein in the presence of large excess of several other SNA lectin binding proteins in serum.

To assess the dynamic range and lower limit of detection of our method, a dilution series of ovalbumin protein (1 pg mL $^{-1}$ to 10 ng mL $^{-1}$) was spiked in diluted serum. For each ovalbumin concentration, the electrochemical response was measured at the glycoprotein capture step and detection step using an anti-ovalbumin antibody. The percentage change in the peak current between these two measurements is attributed to the amount of ovalbumin present in the sample. As shown in Fig. 4C, there was a gradual increase in the percentage change in DPV peak current with increasing concentrations of ovalbumin spiked in serum ranging from 10 pg mL $^{-1}$ to 500 pg mL $^{-1}$.

Beyond 500 pg mL $^{-1}$, there was no further increase in DPV response suggesting saturation of the glycoprotein capture on the electrode surface. The lower limit of detection of 10 pg mL $^{-1}$ described here is comparable with previous electrochemical glycan profiling studies.[25,27] This 10 pg mL $^{-1}$ lower limit of detection in a 1000 fold diluted serum sample will actually translate into 10 ng mL $^{-1}$ in the actual undiluted sample. Previously published lectin electrochemical biosensors showed femtomolar sensitivity and outperformed conventional lectin-ELISA and lectin-microarrays by several fold in terms of lower limit of detection.[25,27] Unlike previous studies which only

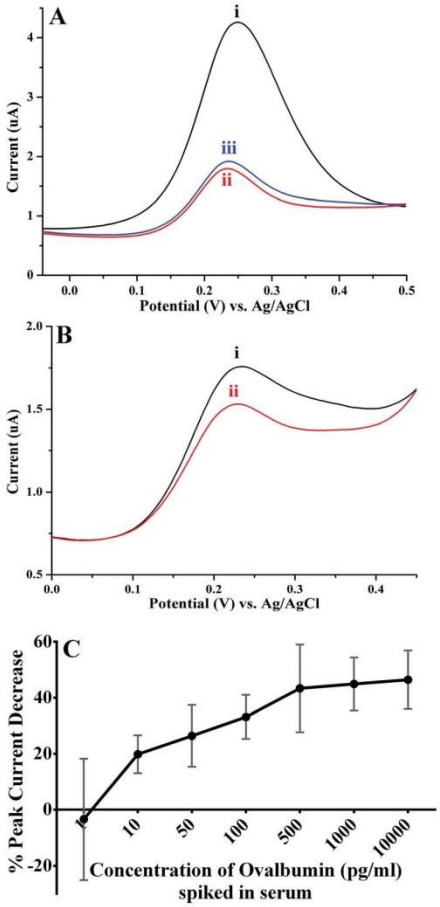


Fig. 4 Differential pulse voltammetric responses for (A) detection of SNA binding proteins present in human serum (i) biotin-BSA/multivalent streptavidin/SNA lectin (ii) capturing SNA lectin binding glycoproteins present in human serum (1 in 1000 dilution) and (iii) detection of any non-specific recognition of SNA bound serum glycoprotein using an anti-ovalbumin antibody. (B) Detection of ovalbumin spiked in human serum (i) capturing 1 ng mL^{-1} ovalbumin in human serum (1 in 1000 dilution) using SNA lectin and (ii) its detection using an anti-ovalbumin antibody. (C) Calibration plot for % change in the differential pulse voltammetric response for detection of a designated concentration of ovalbumin spiked in human serum (1 in 1000 dilution) using an anti-ovalbumin antibody. Each data point represents average ± SD (standard deviation) of 3 independent experiments.

interrogated glycan, here we interrogated both glycan and protein epitopes of a glycoprotein in the background of diluted serum with a detection limit of 10 pg mL^{-1}. The detection limit of our assay could be further improved *via* incorporation and optimization of the device geometry and experimental parameters.[34] Moreover, designing a multiplexed electrochemical biosensor[25,34] would help to monitor the panel of clinically relevant glycoprotein biomarkers in parallel.

Conclusions

In this proof-of-concept study we demonstrated interrogation of glycan along with a protein epitope of a glycoprotein ovalbumin using SNA lectin and an anti-ovalbumin antibody respectively followed by label-free electrochemical detection in the background of diluted serum. Our lectin-antibody biosensor detects specific glycans and the glycoprotein to which the glycan is attached, hence provides increased specificity over existing lectin biosensors. Furthermore, with a detection limit of 10 pg mL^{-1}, the lectin-antibody electrochemical biosensor described here could be developed further to achieve point-of-care diagnosis for clinically relevant glycoprotein biomarkers.

Conflict of interest

The authors declare no competing financial interest.

Acknowledgements

AKS is supported by International Postgraduate Research Scholarship (IPRS) and UQ Centennial (UQ Cent) Scholarship. This work was supported by the ARC DECRA (DE120102503) to MJAS. We also acknowledge funding received by our laboratories from the National Breast Cancer Foundation of Australia (CG-12-07) to MT, NC-14-022 to MH and MT. These grants have significantly contributed to the environment to stimulate the research described here.

References

1 A. Varki and H. H. Freeze, in *Essentials of Glycobiology*, ed. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E.Etzler, Cold Spring Harbor (NY), 2nd edn, 2009.
2 B. Adamczyk, T. Tharmalingam and P. M. Rudd, *Biochim. Biophys. Acta*, 2012, **1820**, 1347–1353.
3 K. Oda, A. Ido, T. Tamai, M. Matsushita, K. Kumagai, S. Mawatari, A. Saishoji, T. Kure, K. Ohno, E. Toyokura, D. Imanaka, A. Moriuchi, H. Uto, M. Oketani, T. Hashiguchi and H. Tsubouchi, *Oncol. Rep.*, 2011, **26**, 1227–1233.
4 C. Kagebayashi, I. Yamaguchi, A. Akinaga, H. Kitano, K. Yokoyama, M. Satomura, T. Kurosawa, M. Watanabe, T. Kawabata, W. Chang, C. Li, L. Bousse, H. G. Wada and S. Satomura, *Anal. Biochem.*, 2009, **388**, 306–311.
5 R. D. Cummings and S. Kornfeld, *J. Biol. Chem.*, 1982, **257**, 11235–11240.
6 Z. Yang and W. S. Hancock, *J. Chromatogr. A*, 2004, **1053**, 79–88.
7 D. Loo, A. Jones and M. M. Hill, *J. Proteome Res.*, 2010, **9**, 5496–5500.
8 E. Choi, D. Loo, J. W. Dennis, C. A. O'Leary and M. M. Hill, *Electrophoresis*, 2011, **32**, 3564–3575.
9 K. Sparbier, T. Wenzel and M. Kostrzewa, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2006, **840**, 29–36.

10 H. Zhang, X. J. Li, D. B. Martin and R. Aebersold, *Nat. Biotechnol.*, 2003, **21**, 660–666.
11 K. Chuah, L. M. Lai, I. Y. Goon, S. G. Parker, R. Amal and J. Justin Gooding, *Chem. Commun.*, 2012, **48**, 3503–3505.
12 Y. Du, C. Chen, B. Li, M. Zhou, E. Wang and S. Dong, *Biosens. Bioelectron.*, 2010, **25**, 1902–1907.
13 J. F. Rusling, *Anal. Chem.*, 2013, **85**, 5304–5310.
14 M. J. Shiddiky, M. A. Rahman and Y. B. Shim, *Anal. Chem.*, 2007, **79**, 6886–6890.
15 J. Wang, *Biosens. Bioelectron.*, 2006, **21**, 1887–1892.
16 E. Katz and I. Willner, *Electroanalysis*, 2003, **15**, 913–947.
17 M. Zhou and S. Dong, *Acc. Chem. Res.*, 2011, **44**, 1232–1243.
18 Y. S. Grewal, M. J. Shiddiky, S. A. Gray, K. M. Weigel, G. A. Cangelosi and M. Trau, *Chem. Commun.*, 2013, **49**, 1551–1553.
19 A. J. Bard and L. R. Faulkner, *Electrochemical Methods: Fundamentals and Applications*, John Wiley & Sons, Inc, 2nd edn, 2001.
20 F. Lisdat and D. Schafer, *Anal. Bioanal. Chem.*, 2008, **391**, 1555–1567.
21 J. S. Daniels and N. Pourmand, *Electroanalysis*, 2007, **19**, 1239–1257.
22 J. Das and S. O. Kelley, *Anal. Chem.*, 2011, **83**, 1167–1172.
23 Y. S. Grewal, M. J. Shiddiky, L. J. Spadafora, G. A. Cangelosi and M. Trau, *Biosens. Bioelectron.*, 2014, **55**, 417–422.
24 H. Yang, Z. Li, X. Wei, R. Huang, H. Qi, Q. Gao, C. Li and C. Zhang, *Talanta*, 2013, **111**, 62–68.
25 V. J. Nagaraj, S. Aithal, S. Eaton, M. Bothara, P. Wiktor and S. Prasad, *Nanomedicine*, 2010, **5**, 369–378.
26 J. T. La Belle, J. Q. Gerlach, S. Svarovsky and L. Joshi, *Anal. Chem.*, 2007, **79**, 6959–6964.
27 T. Bertok, L. Klukova, A. Sediva, P. Kasak, V. Semak, M. Micusik, M. Omastova, L. Chovanova, M. Vlcek, R. Imrich, A. Vikartovska and J. Tkac, *Anal. Chem.*, 2013, **85**, 7324–7332.
28 N. Shibuya, I. J. Goldstein, W. F. Broekaert, M. Nsimba-Lubaki, B. Peeters and W. J. Peumans, *J. Biol. Chem.*, 1987, **262**, 1596–1601.
29 N. Sharon, *Trends Biochem. Sci.*, 1993, **18**, 221–226.
30 T. Bertok, A. Sediva, J. Katrlik, P. Gemeiner, M. Mikula, M. Nosko and J. Tkac, *Talanta*, 2013, **108**, 11–18.
31 M. D. L. Oliveira, M. L. Nogueira, M. T. S. Correia, L. C. B. B. Coelho and C. A. S. Andrade, *Sens. Actuators, B*, 2011, **155**, 789–795.
32 M. Gamella, S. Campuzano, C. Parrado, A. J. Reviejo and J. M. Pingarron, *Talanta*, 2009, **78**, 1303–1309.
33 I. Musselman and D. W. Speicher, *Curr Protoc Protein Sci*, 2005, Ch. 24, Unit 24.1.
34 E. J. Wee, S. Rauf, K. M. Koo, M. J. Shiddiky and M. Trau, *Lab Chip*, 2013, **13**, 4385–4391.

Chapter 7.

DISCUSSION AND FUTURE DIRECTIONS

Chapter 7. Discussion and future directions

7.1.1 Overview of the thesis

This thesis has made significant contribution in the following areas of research. (i) It has filled a gap in the proteomics biomarker research with development of integrated biomarker discovery and verification pipeline using LeMBA platform. (ii) With the help of LeMBA-GlycoSelector and LeMBA-Shiny mixOmics pipeline, lectin-glycoprotein diagnostic biomarker candidates for BE/EAC were discovered and verified. This study demonstrates utility of the pipeline including statistical methods for discovery and verification of biomarker candidates. (iii) Provided proof-of-concept demonstration for using electrochemical detection techniques for glycoprotein interrogation. Collectively, this thesis describes innovative translational biomarker research for diagnosis of BE/EAC.

7.1.2 Biomarker discovery and verification pipeline

Blood based *in vitro* diagnostics employing cancer biomarkers hold promise for early diagnosis and improved patient outcomes, hence there is extensive research on the identification and development of novel cancer biomarkers (480). Discovery and development of new biomarkers is a long and challenging process requiring multi-disciplinary collaborations. Despite intensive efforts, the rate of introduction of new candidate biomarker into clinical practice is on the decline due to multitude of reasons (362, 481, 482). One of the main reasons for failure of cancer biomarker research to deliver clinically applicable diagnostic tests is attributed to unavailability of in-depth biomarker discovery and validation pipeline (482). In addition, large number of false positives, unavailability of high quality patient samples, poor study design, inappropriate statistical analysis, patient heterogeneity, limited understanding of cancer pathogenesis at molecular level and lack of follow-up verification studies have been adversely affecting outcome of cancer biomarker research (130, 441, 482, 483).

Recent technological advancements in proteomics have accelerated the discovery of biomarkers, leading to numerous biomarker publications over the years (441, 484). However, only a handful of new cancer biomarkers have completed the long journey allowing translation into clinical practice (130). One possible way to overcome this challenge is to develop biomarker discovery and development pipeline that allows seamless transition for biomarkers identified through biomarker discovery to further stages of development (485-490). To avoid the high complexity of serum/plasma proteome and the associated requisite multi-dimensional sample separation, most of these workflows focused on tissues or proximal fluids during the discovery

phase, with the goal of extending the findings to plasma. This approach often leads to failure when the candidates are not detected in plasma due to the limited sensitivity of the available analytical methods, or the absence of candidates in the plasma (483).

In this thesis, an alternative workflow to identify glycosylation changes in medium to high abundant glycoproteins has been presented which uses serum as the sample source throughout discovery and verification stages (Figure 7.1). The workflow was designed to enhance the feasibility of glycoprotein biomarker discovery and translation, through scientific rigor while managing the experimental cost. Firstly, serum is used as the sample source throughout discovery and verification, hence eliminating the risk of switching tissue type during biomarker development. Secondly, single step enrichment using liquid handler assisted LeMBA-system reduces sample processing variability. Thirdly, comparatively inexpensive approach of label-free proteomics has been employed using relative quantitation with respect to a spiked-in internal standard chicken ovalbumin. This approach achieved the necessary analytical linearity and reproducibility throughout the more than 2000 hr of total mass spectrometer run time performed in the study. This cost-effective approach can be applied across other existing proteomics platforms to not only account for variations during sample processing but also to reduce the cost of at least initial stages of biomarker studies when many candidates are selected for verification and absolute quantification using SIS labeled peptide would be costly. Fourthly, a sequential filtering approach (459) has been applied in which many candidates were selected from biomarker discovery proteomics data, and verified using MRM-MS with increasing sample size in a cost-effective manner.

Using the biomarker discovery workflow, a total of 183 unique lectin-protein candidates were identified that can distinguish between healthy, BE, and EAC phenotypes. All 20 lectins used in the discovery phase showed differential binding with anywhere between one [e.g. *Helix aspersa* agglutinin (HAA)] to twenty five [e.g. NPL] glycoprotein candidates for pairwise comparison between patient phenotypes suggesting widespread changes in the serum glycosylation profile between healthy, BE and EAC samples. These results are in clear agreement with previous serum glycan profiling studies which identified widespread changes in serum glycan structures between healthy, BE, dysplastic and EAC phenotypes (175-177, 179, 457). Many of 41 protein candidates selected for verification were not identified during biomarker discovery in all 6 lectin-pull downs due to low sensitivity of QTOF mass spectrometer. With increased sensitivity of MRM-MS assay, all 41 protein candidates were identified in all 6 lectin pull-downs for all the patient samples. So out of the total 246 lectin-protein candidates measured for verification, 45 candidates (18.3%) were verified in an independent cohort of patients. Interestingly, only 3 out of 45 candidates were

significantly different between healthy vs BE comparison, with the other 42 candidates differentially present in EAC as compared to either healthy or BE samples. This suggests that EAC phenotype is significantly different from BE and healthy in terms of serum glycoprotein profile.

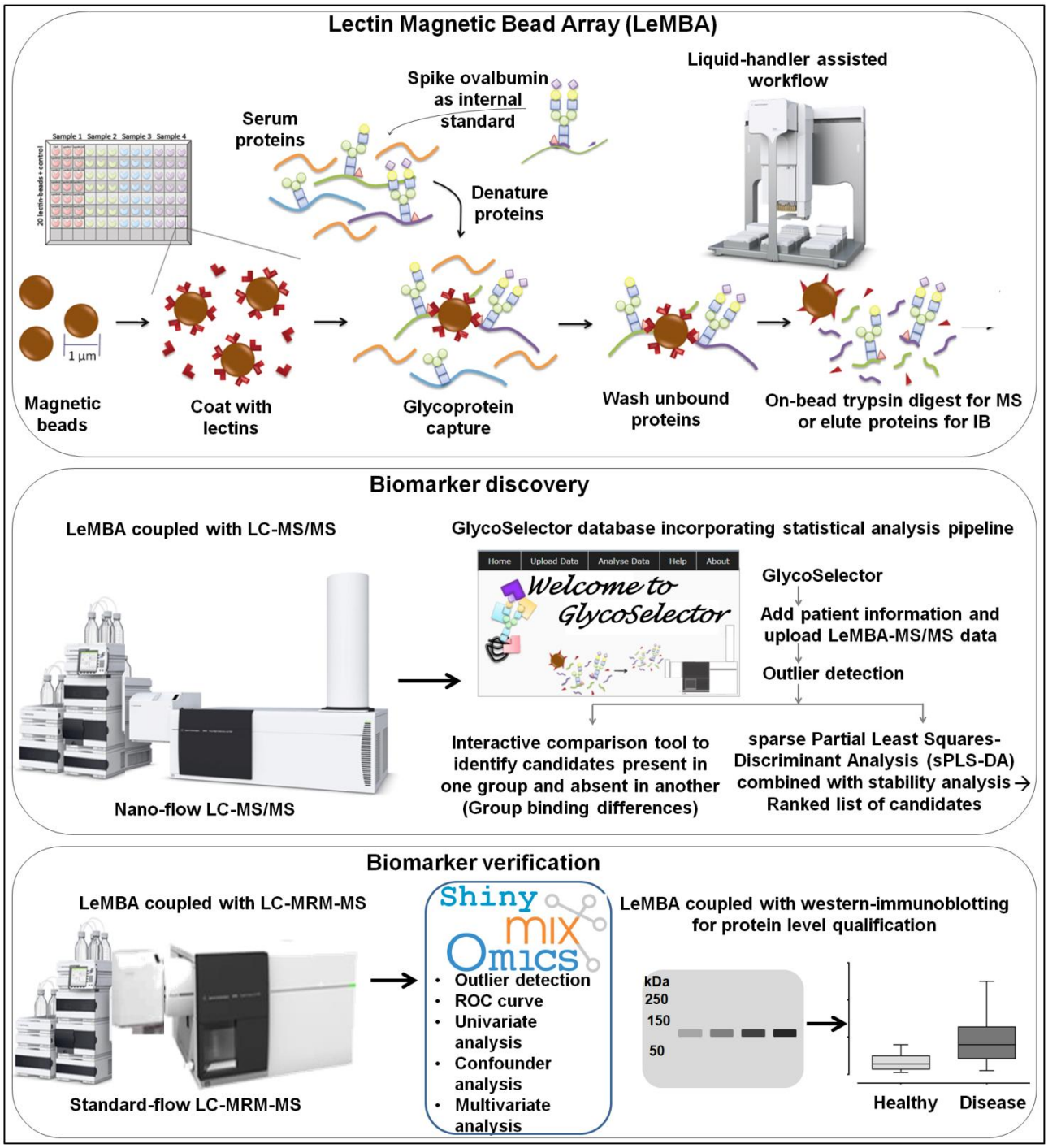


Figure 7.1. Generalized workflow schematic for serum glycoprotein biomarker discovery and qualification. Serum samples from healthy, BE and EAC patient groups were enriched for sub-glycoproteomes using 20 individual lectin coated magnetic beads, followed by on-bead trypsin digest and tandem mass spectrometry for label-free quantitation referencing to internal standard

chicken ovalbumin. In-house database and statistical analysis pipeline "GlycoSelector" (http://glycoselector.di.uq.edu.au/) identified lectin-protein pairs present in one patient group and absent in the other. Sparse partial least squares-discriminant analysis (sPLS-DA) combined with stability analysis was used to generate ranked lists of lectin-protein candidates. For biomarker verification, selected candidates were measured using multiple reaction monitoring-mass spectrometry (MRM-MS) in an independent patient cohort for a subgroup of lectin pull-downs. Dedicated statistical analysis tool "Shiny mixOmics" (http://mixomics-projects.di.uq.edu.au/Shiny) was developed incorporating features to plot ROC curve and to perform univariate/multivariate statistical analyses. LeMBA-immunoblotting was used as an orthogonal method to verify peptide level MS data for selected candidates at the protein level.

7.1.3 Getting biological insights from the candidate biomarkers

Although the main aim of the present study was to discover and verify the glycoprotein biomarker candidates with promising diagnostic ability, the verified biomarker proteins can provide novel insights on pathophysiology of the disease.

7.1.3.1 Functional annotation analysis

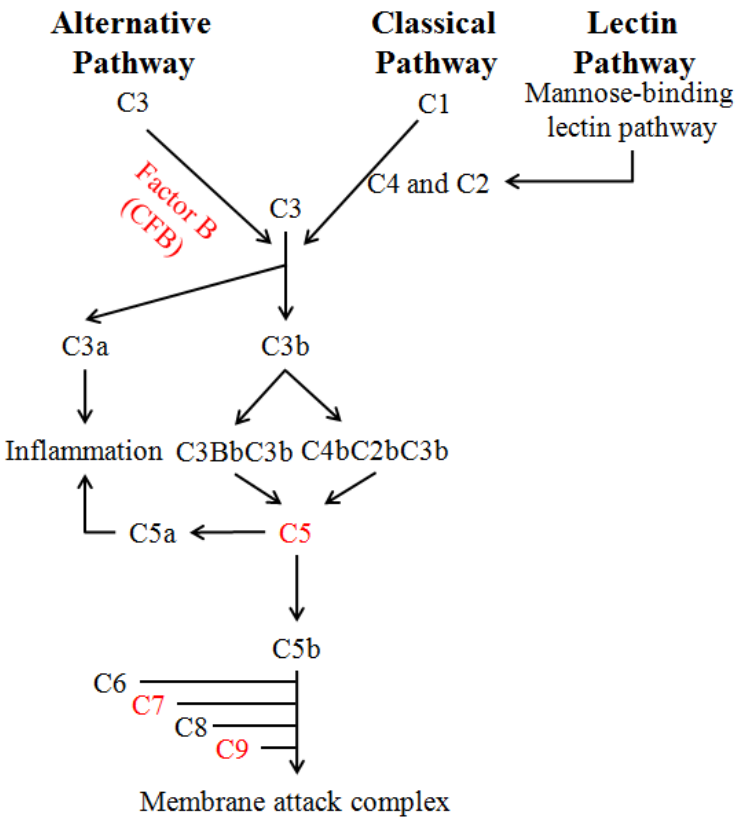
To begin to assess the main functional differences between BE and EAC, functional annotation analysis using online bioinformatics tool DAVID (the database for annotation, visualization and integrated discovery) (http://david.abcc.ncifcrf.gov/) (491, 492) was performed on a list of candidates that differentiated EAC from BE in biomarker verification dataset. The final candidates were selected using univariate Kruskal-Wallis test ($P < 0.05$) and multivariate analysis using sPLS-DA (stability > 70%), and consisted of 17 unique proteins (P00738, P00751, P01009, P01011, P01024, P01031, P02748, P02787, P04114, P04217, P05155, P05546, P06396, P06681, P0C0L5, P10643, P43652) from 59 lectin-protein pairs. Plasma proteome gene list (493) was converted to DAVID IDs and used as a background. Setting the adjusted FDR *P*-values < 0.05 for ontology categories and cluster scores over 3 using functional annotation clustering function, the top Annotation Cluster with an Enrichment Score of 10.4 was found to be SP_PIR_KEYWORD glycoprotein (P = 1.82E-08) and the UP_SEQ_FEATURE glycosylation site:N-linked (GlcNAc...) (P = 2.32E-06), in agreement with the glycoprotein enrichment strategy.

Additional clusters related to acute inflammation, complement cascade pathway, and endopeptidase inhibition, were over-represented within the 17 genes that discriminated BE and EAC. KEGG “Complement and coagulation cascades” pathway (hsa04610) was significantly over-represented (P = 4.6E-18).

7.1.3.2 Complement pathway in EAC

In early 2000s, the link between inflammation and cancer was proposed including novel roles of complement proteins in the malignancy (494). Like other cancer, EAC is no different in the sense that chronic inflammation plays central role in development of EAC (495-497) which is in agreement with the result of functional annotation analysis, and points to alterations in the complement cascade during EAC development. Classically, complement proteins have been considered as a central part of the innate immune response and it serves as first line defense of the body (498). Figure 7.2A illustrates the cascade of events during activation of complement system. The proteins highlighted in red are found as candidate proteins from the biomarker verification results.

A Complement cascade



B Statistical significance of lectin-complement proteins biomarker candidates for BE vs EAC comparison

	CFB	C5	C7	C9
AAL	N.S.	0.0483	0.0063	0.0001
EPHA	N.S.	N.S.	0.0398	0.0003
JAC	0.0373	0.0425	0.0094	0.0007
NPL	N.S.	N.S.	N.S.	0.0049
PSA	N.S.	0.0453	0.0019	0.0008
WGA	N.S.	N.S.	N.S.	0.0032

Figure 7.2. Overview of the complement cascade. (A) The complement system gets activated in tightly regulated fashion by one of the three pathways, alternative pathway, classical pathway or lectin pathway. All three pathways activate C3 by different mechanisms that lead to cleavage and activation of C5 into C5a and C5b. C5a is a potent anaphylactic peptide while C5b can initiate the terminal pathway of membrane attack complex formation which begins with the non-enzymatic assembly of the components C5b-C9. Adapted from Pio *et al.* (494) and Zipfel and Skerka (499). The complement proteins denoted in red [C5, C7, C9 and complement factor B (or CFB)] are found

to be up-regulated in one or more lectin pull-downs in EAC as compared to BE phenotype. **(B)** Denotes Kruskal-Wallis test *P* values of lectin-complement cascade protein biomarker candidates for BE vs EAC comparison.

There are three well-established pathways of complement activation namely the classical pathway, the lectin pathway and the alternative pathway (Figure 7.2A). All three pathways converge at complement C3 (C3) which eventually causes activation of complement C5 (C5) that leads to formation of membrane attack complex (MAC) consisting of complement C5b (C5b), complement component C6 (C6), complement component C7 (C7), complement component C8 (C8) and complement component C9 (C9) and results in destruction of target antigen. The process of complement activation is kept under check by complement regulators namely plasma protease C1 inhibitor, factor I, C4b-binding protein, complement factor H, vitronectin, clusterin, complement receptor type 1, membrane cofactor protein (CD46), decay-accelerating factor (CD55), and CD59. It is very well accepted that complement proteins carry out effective immune surveillance to check tumor initiation and progression (494). Increased levels of complement proteins or activation of complement pathways in tumor tissue, tissue fluid and/or serum/plasma have been observed in lung cancer, ovarian cancer, colorectal cancer, neuroblastoma, myeloma, lymphoma and leukemia suggesting capacity of complement system to recognize malignant cells (494). At the same time, tumor cells express/secrete complement regulator molecules that help them to evade the complement surveillance (500). Many recently developed monoclonal antibodies based immunotherapy target the complement regulator proteins to facilitate complement mediated clearance of the tumor and also to increase effectiveness of conventional treatments (500).

In line with results described in this thesis, Narayanasamy and colleagues identified fucosylated complement component C9 in serum as a biomarker for squamous cell lung cancer (501). Very recently Song and colleagues (180) also identified changes in the glycosylation of complement proteins for EAC and HGD compared to healthy phenotype. They used lectin-affinity chromatography (a mix of fucose and sialic acid binding lectin) and hydrazide chemistry-based glycoprotein enrichment methods and identified complement C3 and complement C1r subcomponent as differentially present in HGD and EAC samples respectively, as compared to healthy blood serum. The differences between the complement proteins identified by Song and colleagues, and those that discriminate BE and EAC in this thesis may be the result of divergent sample processing steps. For example, Song and colleagues (180) used serum sample after depletion of the seven most abundant proteins as compared to the workflow mentioned in this thesis where the serum samples were denatured to break protein complexes without depletion of abundant

proteins. In the LeMBA workflow, an individual lectin has been used for enrichment of a particular type of glycan while Song and colleagues used a mixture of sialic acid binding SNA lectin and fucose specific AAL lectin for glycoprotein enrichment. Nonetheless, changes observed in the glycosylation or total protein levels of complement proteins may suggest a mechanism for inflammation and complement cascades in EAC pathogenesis (495).

Figure 7.2B summarizes statistical significance of all lectin-complement protein biomarker candidates for BE vs EAC analysis. It is interesting to observe that proteins belong to terminal complement cascade are showing differential binding with many lectins as compared to proteins involved in the initial steps of complement activation i.e. CFB showed differential binding with only one lectin whereas C5, C7, and C9 showed differential binding with 3, 4, and all 6 lectins respectively. The diverse cell/tissue types such as liver, blood monocytes, tissue macrophages, epithelial cells, cancer cells and fibroblasts can synthesize and secrete components of complement cascade (502). With activation of the complement cascade, it may be possible that same complement protein is now synthesized and secreted by a different cell/tissue type in addition to normal cell/tissue type with potentially different glycan expression resulting into differential lectin binding between two phenotypes. However, possibility of changes in total levels of the complement proteins between BE and EAC cannot be ruled out at this stage. Collectively, inflammation and complement cascade seems to be playing role in EAC development. Further studies in this regard may help understand the disease pathogenesis and it may allow targeting of EAC using currently available immunotherapy. Although altered glycosylation of complement proteins has been observed in quite a few cancers (180, 501, 503, 504) including EAC, the exact impact of glycosylation on complement cascade and function of complement proteins is yet to be determined.

7.1.3.3 Bile acids, microbiome, diet and glycan: Is there a link?

Exposure of esophageal epithelium with bile acids due to gastroesophageal reflux can be a causative factor for changes in the expression of glycan. When Nehra and colleagues profiled esophageal aspirates in the patients suffering from erosive esophagitis and Barrett's esophagus and compared with that of healthy individuals, they found significantly increased proportion of secondary bile acids, deoxycholic acid and taurodeoxycholic acids to be present in patients with erosive esophagitis and BE (505). Bile acids (especially secondary bile acids) are known for their causative roles for many gastrointestinal cancers (506, 507). At physiological levels repeated and prolonged exposure of bile acids in esophageal tissue, which is not meant to deal with gastric/bile acid, can cause oxidative/nitrosative stress, DNA damage leading to generation of genomic

instability, apoptosis, and ultimately cancer (508, 509). In line with this fact, Byrne and colleagues demonstrated that secondary bile acid deoxycholic acid (DCA) exposure in cell lines derived from normal squamous epithelium (HET-1A) and Barrett's metaplastic epithelium (QH) leads to disruption of Golgi structures and affects protein glycosylation suggesting possible mechanism for differential glycosylation at least at the tissue level (312). However, its impact on serum/plasma glycoproteome remains to be determined.

Bile acids are physiological detergents and enable absorption of lipids, cholesterol, and fat-soluble vitamins from the gastrointestinal tract (507). Cholic acid and chenodeoxycholic acid are primary bile acids which are derived from cholesterol by a sequence of enzymatic reactions that involve 17 different enzymes and occur mainly in the liver (510). After synthesis they are conjugated with either glycine or taurine and excreted and then stored in the gall bladder. When dietary fat enters the proximal intestine, they are released into the intestinal tract and play a critical role for lipid absorption in the ileum. The majority of bile acids are reabsorbed in the terminal ileum by an active bile salt re-absorptive mechanism leaving less than 5% of bile acids to enter the colon (511). In the colon, they are converted into secondary bile acids namely deoxycholic acid and lithocholic acid by bacterial flora (512). Out of the two secondary bile acids, lithocholic acid is fairly insoluble and little of it is re-absorbed while deoxycholic acid is partly re-absorbed in the colon and enters the enterohepatic circulation (513). In the liver, it undergoes conjugation and is secreted again in the bile. So normally, the circulating bile acid pool is composed of 30-40% cholic and chenodeoxycholic acid (primary bile acid), 20-30% deoxycholic acid and 5% lithocholic acid (514). All the factors that affect bile composition leading to increased levels of secondary bile acid can create pro-cancerous phenotype. One of the well-known factors that affect the bile composition is diet. High fat diet can cause increased levels of bile acids in the gastrointestinal tract. Conversely, diets rich in fibers (fruits and vegetables) aid in excretion of secondary bile acid and lower the risk of cancer (515). In line with the fact that the diets rich in fruits and vegetables lower the risk for BE/EAC, it has been experimentally demonstrated that high animal fat intake enhances the development of BE/EAC and changes the composition of bile acid in an animal model (516).

Since decades ago when no "omics" techniques existed, it has been known that the microorganisms present in the large intestine are responsible for the conversion of primary bile acids into secondary (517). So any changes in this microbial flora can potentially affect bile composition and may have implications in EAC pathogenesis. Recent evidences suggest that food and obesity have huge impact on the intestinal microbiome and bile acid metabolism (518-521). However, the majority of studies published and cited in this section were conducted in the context

of colon cancer. So the implications of microbiome mediated changes in the bile composition on BE/EAC pathogenesis remains to be determined. Furthermore, the possible impact of changes in the bile composition on glycosylation changes both at tissue level and in the circulation has not been investigated.

7.1.3.4 Diagnosis of BE/EAC: What does future look like?

The success of cancer screening programs in improving outcome for many cancer types in the past decades emphasize the importance of early diagnosis and the development of suitable screening/surveillance tools (6). Indeed, the lack of cost-effective screening/surveillance methodology to facilitate early diagnosis of EAC is one of the main reasons for the high mortality (522). Current endoscopy based screening is costly, requires specialist appointment, is not suitable for frequent large scale at risk population monitoring, hence not very effective (523). While white light endoscopy with biopsy is most widely used for screening/surveillance (524), several innovative methods are being evaluated for more economic screening, including advanced imaging, non-endoscopic sampling, and blood biomarkers. Novel advanced imaging techniques may improve yield of biopsies for dysplastic Barrett's by targeting suspicious areas and may be more effective than random biopsies (525-527). These real time techniques like probe based confocal endomicroscopy (pCLE) (526, 527) with resolution up to 1 micron can potentially decrease sampling errors associated with endoscopic biopsies. Another potentially useful imaging technique is volumetric laser endomicroscopy (VLE). Based on the principle of backscattering of near infrared wavelength light to obtain an image of esophageal wall layers, which can scan 6 cm of esophagus in 90 seconds to a depth of 2 mm with a resolution of 7 microns (525), VLE can potentially help target suspicious areas harbouring subsquamous glands which are proposed to be a risk for EAC (528). These imaging modalities can be performed during ongoing endoscopy and potentially improve surveillance and detection of dysplasia and EAC. While biomarkers could potentially risk stratify patient population, these invasive and expensive "red flagging" techniques could potentially benefit and improve screening and surveillance.

To enable safe, cheap, and minimally invasive diagnosis of BE in primary care setting, Lao-Sirieix and colleagues developed a novel device called the capsule sponge, or Cytosponge (262, 529). The gelatin capsule contains a polyurethane sponge and attached to a string. The capsule is swallowed which dissolves in the stomach within 5 min. The sponge is then retrieved by pulling out the string that samples cells from the stomach to the oropharynx. The cell samples obtained using Cytosponge can be subjected to histological characterization (524). Alternatively, Cytosponge

sample can be evaluated for more objective markers like expression level changes in TFF3 gene to diagnose underlying BE (262). In fact, Cytosponge sampling with immunohistochemical evaluation of trefoil factor 3 can diagnose BE > 2 cm in length with sensitivity of 90.0% and specificity of 93.5% as compared to gastroscopy (524). As far as application of Cytosponge sampling in diagnosing dysplasia or EAC is concerned, because the majority of recurrently mutated genes in EAC are also mutated in BE (201, 204, 205), the opportunity to distinguish between BE and dysplasia/EAC based on genomic profiling alone might prove to be very limited. Validation of existing protein biomarkers in samples collected using Cytosponge may provide an alternative opportunity for diagnosis of dysplasia/EAC.

An ideal way to carry out routine screening/surveillance for BE/dysplasia/EAC will be by developing an *in vitro* diagnostic tool that uses blood as a sample type. Cell free circulating DNA, miRNA, glycan and metabolic changes have shown differential presence between healthy, BE, dysplastic and EAC phenotype in serum or plasma and all of them qualify for development of diagnostics (1, 530). In this thesis, serum glycoproteins showing differential binding with lectins have been identified to differentiate mainly between EAC and healthy/BE conditions. Now, the efforts are required to validate these changes using clinically applicable diagnostic tests in multicenter clinical trial to establish clear comparison between these circulatory biomarkers and existing endoscopy-biopsy screening. If found comparable, the blood based diagnostic can replace endoscopy-biopsy as a 1st line screening/surveillance tool while there is no doubt that endoscopy-biopsy will remain gold standard for confirmation and accurate staging of the disease.

Like other common cancers (531), it is likely that introduction of any new screening program for EAC will lead to false-positive identifications. However, blood based *in vitro* diagnostic will avoid endoscopy-biopsy in the patients who are currently monitored using these costly and inconvenient procedures multiple times per year. The cost-savings by avoiding frequent endoscopy-biopsy for monitoring purposes may compensate for an additional endoscopy-biopsy referred due to false-positive identifications. In this manner, using the same amount of funding, wider patient cohort can be screened that may lead to better patient outcomes and health benefits.

7.1.4 Biomarker translation using electrochemical biosensor

Electrochemical methods were initially applied for accurate and sensitive detection of mainly redox analytes (532). The most successful application so far developed using electrochemical detection is glucose biosensor which is widely used in self-use point-of-care settings (533). Later on, electrochemical biosensors were modified for measurements of non-redox

analytes including protein biomarkers but their use is mainly confined to research laboratories (532, 534). In this thesis, an electrochemical detection technique, faradaic electrochemical impedance spectroscopy (F-EIS), was successfully used for interrogation of a model glycoprotein chicken ovalbumin (2). While the study demonstrates the principle of electrochemical detection of a glycoprotein, several technological advancements and follow-up validation studies are required for using electrochemical biosensor as an *in vitro* diagnostic test in clinical application. (i) In the current format, macrodisk electrodes were reused after cleaning for all the experiments. This method worked for a proof-of-principle study but when it comes to testing actual biomarker candidates using real clinical sample, disposable electrodes are more suitable. (ii) The lectins and antibodies to monitor actual biomarker candidates need to be tested for interaction between glycans present on antibody surface and lectins. If there is any interaction, couple of strategies can be tried. Either un-reacted free lectin sites after glycoprotein capture can be blocked using unrelated IgG molecules (415) or alternatively, antibody glycans can be chemically oxidized to inhibit their interaction with the lectins (416). (iii) In the current format, biotin labeled bovine serum albumin (BSA) and multivalent streptavidin were used for capture of biotinylated lectins. Instead of BSA which is very large biomolecule, smaller chemical species such as thiol-PEG-biotin should be tested for enhanced performance of the biosensor. (iv) The electrochemical biosensor should be designed to allow multiplexing whereby multiple candidate biomarkers can be simultaneously monitored. Collectively, once the electrochemical detection technique is validated by monitoring actual biomarker using clinical sample and show comparable results with the mass spectrometric quantitation, they can replace the mass spectrometric quantitation in later phases of biomarker development.

7.1.5 Limitations of the study

While lectins are very useful tool for glycoprotein studies, they are criticized for lack of specificity. Lectins generally recognize glycan structures with low affinity but with high avidity mainly through hydrogen bonding, hydrophobic interactions and van der Waals forces with a dissociation constant (K_d) value lies in the range of 10^{-3} M to 10^{-7} M as compared to antigen-antibody interactions which shows K_d value in the range of 10^{-6} M to 10^{-9} M (344, 535). So only based on differential lectin binding it is difficult to predict structural changes in the glycosylation. The differential binding between serum glycoprotein biomarker candidates and lectins is indirect measure of glycosylation change. The lectin-based LeMBA workflow utilized in this thesis does not provide any direct evidence of changes in the expression of glyco-epitopes associated with the glycoprotein biomarker candidates. Moreover, the focus throughout this thesis was to identify

glycosylation changes in the medium to high abundant proteins meaning low abundant glycoprotein candidates could have been easily missed out.

A total of 29 and 60 samples were screened respectively from healthy, BE and EAC groups in biomarker discovery and verification phases. Although the number of samples is within a range of proposed sample size for biomarker discovery and verification (133), the verification phase did not include enough number of samples to divide them separately into training and validation sets for statistical analysis. Furthermore, samples from early or late dysplastic patients were not available for biomarker discovery and verification so at the moment relative timing of the changes in the biomarker levels in EAC development is not clear. The healthy along with population control samples (total n = 39) were combined to find out candidates affected by confounding covariates. Some of the subgroups for the statistical analyses had only few samples (e.g. n = 5 for patients with daily or weekly symptoms of reflux and n = 5 for patients with alcohol consumption with <1 standard drink/week) resulting in limited statistical power. Even though reflux frequency was assessed as a confounding covariate, there was a lack of true inflammatory patient control group to confidently identify true EAC biomarker candidates from closely related inflammatory conditions.

Although electrochemical detection platform demonstrated applicability to monitor glycoprotein biomarker candidates in a proof-of-concept work, the reproducibility of measurements was far away from what is actually required for clinical assay. The biosensor construction relied upon reproducible adsorption of biomolecule bovine serum albumin that is relatively larger in size as compared to chemical based modifications. In addition, non-reacted electrode surface was not blocked so the possibility of non-specific interaction of proteins with the electrode surface in subsequent stages of the assay cannot be ruled out.

7.1.6 Future opportunities

7.1.6.1 Validation of lectin-protein biomarker candidates in an independent patient cohort including dysplastic samples

The findings of this study should be replicated in an independent patient cohort that include dysplastic patient samples to determine exact timing of change in the level of biomarker candidates during EAC development. Ideally, the patient cohort should be well characterized and findings of endoscopy-biopsy should be available for each patient to allow comparison between changes in the lectin-protein biomarker candidates in circulation with actual disease progression. Once the findings are replicated, the final biomarker panel can be evaluated in multi-centre prospective trial to determine actual positive predictive value of the test.

7.1.6.2 MRM-MS assay for biomarker discovery

The MRM-MS assay used for verification showed better reproducibility over discovery proteomics workflow employed in biomarker discovery. The MRM-MS assay includes transitions for proteins identified as candidates in biomarker discovery and excludes any additional proteins which are generally identified in LeMBA pull-down. The current MRM-MS assay can be optimized further to include transitions of these common proteins identified in LeMBA pull-down in addition to existing list of potential candidates and can be used for even biomarker discovery purposes. This may cause slight increase in the total duration of MRM-MS assay, but it will allow monitoring of many more candidates from LeMBA pull-down as compared to only significant ones at no extra cost. When an independent patient cohort that includes dysplastic patient samples will be screened using this proposed MRM-MS assay, it will widen the scope of the validation. This may lead to identification of new candidates specific for dysplasia which may not be present in the current list of candidates identified between healthy, BE and EAC phenotypes.

7.1.6.3 Glycan/glycosite characterization of biomarker candidates

The biomarker candidate identified should be subjected to full glycomic characterization to determine the changes in the glycan structure and/or site of glycosylation between different disease states. This may help understand basis of the differential lectin binding between different phenotypes.

7.1.6.4 Determine biological basis for changes in the biomarker levels

Along with screening serum samples, tissue samples from the matched patients (independent validation cohort) should be available to perform staining for candidate biomarkers in matched tissue sections. This will help to determine possible source of the biomarker candidates in the circulation i.e. tumor tissue or stromal components. This may also aid understand pathological basis of the cancer-associated glycosylation changes.

7.1.7 Conclusions

The work presented in this thesis identified serum glycoprotein diagnostic biomarker candidates for Barrett's esophagus and esophageal adenocarcinoma using LeMBA-GlycoSelector pipeline. The targeted proteomics assay developed for candidate biomarkers demonstrated linearity and reproducibility for biomarker verification. The selected lectin-glycoprotein biomarker candidates were verified using LeMBA-MRM-MS-Shiny mixOmics workflow in an independent patient cohort leading to identification of a panel of serum glycoprotein biomarker candidate to

distinguish between BE and EAC. Lastly, results obtained using electrochemical detection methodology demonstrated the possibility of using electrochemical detection for developing cheap point-of-care diagnostics with high specificity and sensitivity for blood glycoprotein biomarkers. It is anticipated that this work will be extended further to develop an *in vitro* diagnostic test to screen and monitor patients at risk of EAC.

Chapter 8.

REFERENCES

Chapter 8. References

1. Shah AK, Saunders NA, Barbour AP, et al. Early diagnostic biomarkers for esophageal adenocarcinoma--the current state of play. Cancer Epidemiol. Biomarkers Prev. 2013;22:1185-209.
2. Shah AK, Hill MM, Shiddiky MJ, et al. Electrochemical detection of glycan and protein epitopes of glycoproteins in serum. Analyst 2014;139:5970-6.
3. Shah AK, Le Cao K, Choi E, et al. Serum glycoprotein biomarker discovery and qualification pipeline reveals novel diagnostic biomarker candidates for esophageal adenocarcinoma. Mol. Cell. Proteomics 2015.
4. Kuo B, Daniela U. Esophagus-anatomy and development. GI Motility online 2006.
5. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J. Clin. 2012;62:10-29.
6. Siegel R, Ma J, Zou Z, et al. Cancer statistics, 2014. CA Cancer J. Clin. 2014;64:9-29.
7. Cook MB, Chow WH, Devesa SS. Oesophageal cancer incidence in the United States by race, sex, and histologic type, 1977-2005. Br. J. Cancer 2009;101:855-9.
8. Demeester SR. Epidemiology and biology of esophageal cancer. Gastrointest. Cancer Res. 2009;3:S2-5.
9. Ferlay J, Shin HR, Bray F, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int. J. Cancer 2010;127:2893-917.
10. Holmes RS, Vaughan TL. Epidemiology and pathogenesis of esophageal cancer. Semin. Radiat. Oncol. 2007;17:2-9.
11. Vakil N, van Zanten SV, Kahrilas P, et al. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. Am. J. Gastroenterol. 2006;101:1900-20.
12. Wang DH, Souza RF. Biology of Barrett's esophagus and esophageal adenocarcinoma. Gastrointest. Endosc. Clin. N. Am. 2011;21:25-38.
13. Spechler SJ, Souza RF. Barrett's esophagus. N. Engl. J. Med. 2014;371:836-45.
14. Oh DS, Demeester SR. Pathophysiology and treatment of Barrett's esophagus. World J. Gastroenterol. 2010;16:3762-72.

15. Anaparthy R, Sharma P. Progression of Barrett oesophagus: role of endoscopic and histological predictors. Nat. Rev. Gastroenterol. Hepatol. 2014;11:525-34.

16. Bird-Lieberman EL, Fitzgerald RC. Early diagnosis of oesophageal cancer. Br. J. Cancer 2009;101:1-6.

17. Tougeron D, Richer JP, Silvain C. Management of esophageal adenocarcinoma. J. Visc. Surg. 2011;148:e161-70.

18. Ilson DH. Esophageal cancer chemotherapy: recent advances. Gastrointest. Cancer Res. 2008;2:85-92.

19. Tileston W. Peptic ulcer of the oesophagus. The American Journal of the Medical Sciences 1906;132:240-265.

20. Lyall A. Chronic peptic ulcer of the œsophagus: A report of eight cases. Br. J. Surg. 1937;24:534-547.

21. Stewart MJ, Hartfall SJ. Chronic peptic ulcer of the œsophagus. The Journal of Pathology and Bacteriology 1929;32:9-14.

22. Barrett NR. Chronic peptic ulcer of the oesophagus and 'oesophagitis'. Br. J. Surg. 1950;38:175-82.

23. Lord RV. Norman Barrett, "doyen of esophageal surgery". Ann. Surg. 1999;229:428-39.

24. Sharma P, McQuaid K, Dent J, et al. A critical review of the diagnosis and management of Barrett's esophagus: the AGA Chicago Workshop. Gastroenterology 2004;127:310-30.

25. Naef AP, Savary M, Ozzello L. Columnar-lined lower esophagus: an acquired lesion with malignant predisposition. Report on 140 cases of Barrett's esophagus with 12 adenocarcinomas. J. Thorac. Cardiovasc. Surg. 1975;70:826-35.

26. Brown LM, Devesa SS, Chow WH. Incidence of adenocarcinoma of the esophagus among white Americans by sex, stage, and age. J. Natl. Cancer Inst. 2008;100:1184-7.

27. Pohl H, Sirovich B, Welch HG. Esophageal adenocarcinoma incidence: are we reaching the peak? Cancer Epidemiol. Biomarkers Prev. 2010;19:1468-70.

28. Bollschweiler E, Wolfgarten E, Gutschow C, et al. Demographic variations in the rising incidence of esophageal adenocarcinoma in white males. Cancer 2001;92:549-55.

29. Hongo M, Nagasaki Y, Shoji T. Epidemiology of esophageal cancer: Orient to Occident. Effects of chronology, geography and ethnicity. J. Gastroenterol. Hepatol. 2009;24:729-35.

30. Baquet CR, Commiskey P, Mack K, et al. Esophageal cancer epidemiology in blacks and whites: racial and gender disparities in incidence, mortality, survival rates and histology. J. Natl. Med. Assoc. 2005;97:1471-8.

31. Sonnenberg A. Effects of environment and lifestyle on gastroesophageal reflux disease. Dig. Dis. 2011;29:229-34.

32. Pickens A, Orringer MB. Geographical distribution and racial disparity in esophageal cancer. Ann. Thorac. Surg. 2003;76:S1367-9.

33. Revels SL, Morris AM, Reddy RM, et al. Racial Disparities in Esophageal Cancer Outcomes. Ann. Surg. Oncol. 2012;20:1136-1141.

34. El-Serag HB, Sweet S, Winchester CC, et al. Update on the epidemiology of gastro-oesophageal reflux disease: a systematic review. Gut 2014;63:871-80.

35. Ronkainen J, Aro P, Storskrubb T, et al. Prevalence of Barrett's esophagus in the general population: an endoscopic study. Gastroenterology 2005;129:1825-31.

36. Hayeck TJ, Kong CY, Spechler SJ, et al. The prevalence of Barrett's esophagus in the US: estimates from a simulation model confirmed by SEER data. Dis. Esophagus 2010;23:451-7.

37. Rex DK, Cummings OW, Shaw M, et al. Screening for Barrett's esophagus in colonoscopy patients with and without heartburn. Gastroenterology 2003;125:1670-7.

38. Gerson LB, Banerjee S. Screening for Barrett's esophagus in asymptomatic women. Gastrointest. Endosc. 2009;70:867-73.

39. Reid BJ, Li X, Galipeau PC, et al. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. Nat. Rev. Cancer 2010;10:87-101.

40. Shaheen NJ, Crosby MA, Bozymski EM, et al. Is there publication bias in the reporting of cancer risk in Barrett's esophagus? Gastroenterology 2000;119:333-8.

41. Hvid-Jensen F, Pedersen L, Drewes AM, et al. Incidence of adenocarcinoma among patients with Barrett's esophagus. N. Engl. J. Med. 2011;365:1375-83.

42. Bhat S, Coleman HG, Yousef F, et al. Risk of malignant progression in Barrett's esophagus patients: results from a large population-based study. J. Natl. Cancer Inst. 2011;103:1049-57.

43. Rutegard M, Lagergren P, Nordenstedt H, et al. Oesophageal adenocarcinoma: the new epidemic in men? Maturitas 2011;69:244-8.

44. Kubo A, Corley DA. Body mass index and adenocarcinomas of the esophagus or gastric cardia: a systematic review and meta-analysis. Cancer Epidemiol. Biomarkers Prev. 2006;15:872-8.

45. Mayne ST, Risch HA, Dubrow R, et al. Nutrient intake and risk of subtypes of esophageal and gastric cancer. Cancer Epidemiol. Biomarkers Prev. 2001;10:1055-62.

46. Wu AH, Wan P, Bernstein L. A multiethnic population-based study of smoking, alcohol and body size and risk of adenocarcinomas of the stomach and esophagus (United States). Cancer Causes Control 2001;12:721-32.

47. Weston AP, Badr AS, Topalovski M, et al. Prospective evaluation of the prevalence of gastric Helicobacter pylori infection in patients with GERD, Barrett's esophagus, Barrett's dysplasia, and Barrett's adenocarcinoma. Am. J. Gastroenterol. 2000;95:387-94.

48. Tischoff I, Tannapfel A. Barrett's esophagus: can biomarkers predict progression to malignancy? Expert Rev. Gastroenterol. Hepatol. 2008;2:653-63.

49. Lagergren J, Bergstrom R, Lindgren A, et al. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N. Engl. J. Med. 1999;340:825-31.

50. Wild CP, Hardie LJ. Reflux, Barrett's oesophagus and adenocarcinoma: burning questions. Nat. Rev. Cancer 2003;3:676-84.

51. Phillips WA, Lord RV, Nancarrow DJ, et al. Barrett's esophagus. J. Gastroenterol. Hepatol. 2011;26:639-48.

52. Cook MB, Wild CP, Forman D. A systematic review and meta-analysis of the sex ratio for Barrett's esophagus, erosive reflux disease, and nonerosive reflux disease. Am. J. Epidemiol. 2005;162:1050-61.

53. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ. Tech. Rep. Ser. 2000;894:i-xii, 1-253.

54. Lagergren J. Adenocarcinoma of oesophagus: what exactly is the size of the problem and who is at risk? Gut 2005;54 Suppl 1:i1-5.

55. Chow WH, Blaser MJ, Blot WJ, et al. An inverse relation between cagA+ strains of Helicobacter pylori infection and risk of esophageal and gastric cardia adenocarcinoma. Cancer Res. 1998;58:588-90.

56. Ryan AM, Duong M, Healy L, et al. Obesity, metabolic syndrome and esophageal adenocarcinoma: epidemiology, etiology and new targets. Cancer Epidemiol. 2011;35:309-19.

57. Kendall BJ, Macdonald GA, Hayward NK, et al. Leptin and the risk of Barrett's oesophagus. Gut 2008;57:448-54.

58. Rutegard M, Shore R, Lu Y, et al. Sex differences in the incidence of gastrointestinal adenocarcinoma in Sweden 1970-2006. Eur. J. Cancer 2010;46:1093-100.

59. Nordenstedt H, Younes M, El-Serag HB. Expression of androgen receptors in barrett esophagus. J. Clin. Gastroenterol. 2012;46:251-2.

60. Tiffin N, Suvarna SK, Trudgill NJ, et al. Sex hormone receptor immunohistochemistry staining in Barrett's oesophagus and adenocarcinoma. Histopathology 2003;42:95-6.

61. Tihan T, Harmon JW, Wan X, et al. Evidence of androgen receptor expression in squamous and adenocarcinoma of the esophagus. Anticancer Res. 2001;21:3107-14.

62. Awan AK, Iftikhar SY, Morris TM, et al. Androgen receptors may act in a paracrine manner to regulate oesophageal adenocarcinoma growth. Eur. J. Surg. Oncol. 2007;33:561-8.

63. Lagergren J, Nyren O. Do sex hormones play a role in the etiology of esophageal adenocarcinoma? A new hypothesis tested in a population-based cohort of prostate cancer patients. Cancer Epidemiol. Biomarkers Prev. 1998;7:913-5.

64. Cooper SC, Croft S, Day R, et al. Patients with prostate cancer are less likely to develop oesophageal adenocarcinoma: could androgens have a role in the aetiology of oesophageal adenocarcinoma? Cancer Causes Control 2009;20:1363-8.

65. Yang H, Sukocheva OA, Hussey DJ, et al. Estrogen, male dominance and esophageal adenocarcinoma: Is there a link? World J. Gastroenterol. 2012;18:393-400.

66. Chandanos E, Rubio CA, Lindblad M, et al. Endogenous estrogen exposure in relation to distribution of histological type and estrogen receptors in gastric adenocarcinoma. Gastric Cancer 2008;11:168-74.

67. Akgun H, Lechago J, Younes M. Estrogen receptor-beta is expressed in Barrett's metaplasia and associated adenocarcinoma of the esophagus. Anticancer Res. 2002;22:1459-61.

68. Liu L, Chirala M, Younes M. Expression of estrogen receptor-beta isoforms in Barrett's metaplasia, dysplasia and esophageal adenocarcinoma. Anticancer Res. 2004;24:2919-24.

69. Kalayarasan R, Ananthakrishnan N, Kate V, et al. Estrogen and progesterone receptors in esophageal carcinoma. Dis. Esophagus 2008;21:298-303.

70. Kobayashi K. Effect of sex hormone on the experimental induction of esophageal cancer. Nihon Geka Gakkai zasshi 1985;86:280-9.

71. Masaka T, Iijima K, Endo H, et al. Gender differences in oesophageal mucosal injury in a reflux oesophagitis model of rats. Gut 2012;62:6-14.

72. Kambhampati S, Banerjee S, Dhar K, et al. 2-methoxyestradiol inhibits Barrett's esophageal adenocarcinoma growth and differentiation through differential regulation of the beta-catenin-E-cadherin axis. Mol. Cancer Ther. 2010;9:523-34.

73. Chandanos E, Lindblad M, Jia C, et al. Tamoxifen exposure and risk of oesophageal and gastric adenocarcinoma: a population-based cohort study of breast cancer patients in Sweden. Br. J. Cancer 2006;95:118-22.

74. Bodelon C, Anderson GL, Rossing MA, et al. Hormonal factors and risks of esophageal squamous cell carcinoma and adenocarcinoma in postmenopausal women. Cancer Prev. Res. (Phila.) 2011;4:840-50.

75. Gammon MD, Schoenberg JB, Ahsan H, et al. Tobacco, alcohol, and socioeconomic status and adenocarcinomas of the esophagus and gastric cardia. J. Natl. Cancer Inst. 1997;89:1277-84.

76. Vaughan TL, Davis S, Kristal A, et al. Obesity, alcohol, and tobacco as risk factors for cancers of the esophagus and gastric cardia: adenocarcinoma versus squamous cell carcinoma. Cancer Epidemiol. Biomarkers Prev. 1995;4:85-92.

77. Henrik Siman J, Forsgren A, Berglund G, et al. Helicobacter pylori infection is associated with a decreased risk of developing oesophageal neoplasms. Helicobacter 2001;6:310-6.

78. Terry P, Lagergren J, Hansen H, et al. Fruit and vegetable consumption in the prevention of oesophageal and cardia cancers. Eur. J. Cancer Prev. 2001;10:365-9.

79. Terry P, Lagergren J, Ye W, et al. Antioxidants and cancers of the esophagus and gastric cardia. Int. J. Cancer 2000;87:750-4.

80. Funkhouser EM, Sharp GB. Aspirin and reduced risk of esophageal carcinoma. Cancer 1995;76:1116-9.

81. Thun MJ, Namboodiri MM, Calle EE, et al. Aspirin use and risk of fatal cancer. Cancer Res. 1993;53:1322-7.

82. Farrow DC, Vaughan TL, Hansten PD, et al. Use of aspirin and other nonsteroidal anti-inflammatory drugs and risk of esophageal and gastric cancer. Cancer Epidemiol. Biomarkers Prev. 1998;7:97-102.

83. Anderson LA, Johnston BT, Watson RG, et al. Nonsteroidal anti-inflammatory drugs and the esophageal inflammation-metaplasia-adenocarcinoma sequence. Cancer Res. 2006;66:4975-82.

84. Souza RF, Shewmake K, Beer DG, et al. Selective inhibition of cyclooxygenase-2 suppresses growth and induces apoptosis in human esophageal adenocarcinoma cells. Cancer Res. 2000;60:5767-72.

85. Chandrasoma PT, Der R, Ma Y, et al. Histologic classification of patients based on mapping biopsies of the gastroesophageal junction. Am. J. Surg. Pathol. 2003;27:929-36.

86. Oberg S, Peters JH, DeMeester TR, et al. Inflammation and specialized intestinal metaplasia of cardiac mucosa is a manifestation of gastroesophageal reflux disease. Ann. Surg. 1997;226:522-30; discussion 530-2.

87. Bremner CG, Lynch VP, Ellis FH, Jr. Barrett's esophagus: congenital or acquired? An experimental study of esophageal mucosal regeneration in the dog. Surgery 1970;68:209-16.

88. Chandrasoma PT, Der R, Dalton P, et al. Distribution and significance of epithelial types in columnar-lined esophagus. Am. J. Surg. Pathol. 2001;25:1188-93.

89. DeMeester SR, DeMeester TR. Columnar mucosa and intestinal metaplasia of the esophagus: fifty years of controversy. Ann. Surg. 2000;231:303-21.

90. Guillem PG. How to make a Barrett esophagus: pathophysiology of columnar metaplasia of the esophagus. Dig. Dis. Sci. 2005;50:415-24.

91. Johns BA. Developmental changes in the oesophageal epithelium in man. J. Anat. 1952;86:431-42.

92. Tosh D, Slack JM. How cells change their phenotype. Nat. Rev. Mol. Cell Biol. 2002;3:187-94.

93. Yu WY, Slack JM, Tosh D. Conversion of columnar to stratified squamous epithelium in the developing mouse oesophagus. Dev. Biol. 2005;284:157-70.

94. Sawhney RA, Shields HM, Allan CH, et al. Morphological characterization of the squamocolumnar junction of the esophagus in patients with and without Barrett's epithelium. Dig. Dis. Sci. 1996;41:1088-98.

95. Sarosi G, Brown G, Jaiswal K, et al. Bone marrow progenitor cells contribute to esophageal regeneration and metaplasia in a rat model of Barrett's esophagus. Dis. Esophagus 2008;21:43-50.

96. Korbling M, Katz RL, Khanna A, et al. Hepatocytes and epithelial cells of donor origin in recipients of peripheral-blood stem cells. N. Engl. J. Med. 2002;346:738-46.

97. Silberg DG, Swain GP, Suh ER, et al. Cdx1 and cdx2 expression during intestinal development. Gastroenterology 2000;119:961-71.

98. Gao N, White P, Kaestner KH. Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. Dev. Cell 2009;16:588-99.

99. Liu T, Zhang X, So CK, et al. Regulation of Cdx2 expression by promoter methylation, and effects of Cdx2 transfection on morphology and gene expression of human esophageal epithelial cells. Carcinogenesis 2007;28:488-96.

100. Hu Y, Jones C, Gellersen O, et al. Pathogenesis of Barrett esophagus: deoxycholic acid up-regulates goblet-specific gene MUC2 in concert with CDX2 in human esophageal cells. Arch. Surg. 2007;142:540-4; discussion 544-5.

101. Blache P, van de Wetering M, Duluc I, et al. SOX9 is an intestine crypt transcription factor, is regulated by the Wnt pathway, and represses the CDX2 and MUC2 genes. J. Cell Biol. 2004;166:37-47.

102. Bastide P, Darido C, Pannequin J, et al. Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. J. Cell Biol. 2007;178:635-48.

103. Reid BJ, Haggitt RC, Rubin CE, et al. Observer variation in the diagnosis of dysplasia in Barrett's esophagus. Hum. Pathol. 1988;19:166-78.

104. Behrens A, Pech O, Graupe F, et al. Barrett's adenocarcinoma of the esophagus: better outcomes through new methods of diagnosis and treatment. Dtsch Arztebl Int 2011;108:313-9.

105. Wang VS, Hornick JL, Sepulveda JA, et al. Low prevalence of submucosal invasive carcinoma at esophagectomy for high-grade dysplasia or intramucosal adenocarcinoma in Barrett's esophagus: a 20-year experience. Gastrointest. Endosc. 2009;69:777-83.

106. American Gastroenterological A, Spechler SJ, Sharma P, et al. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. Gastroenterology 2011;140:1084-91.

107. Playford RJ. New British Society of Gastroenterology (BSG) guidelines for the diagnosis and management of Barrett's oesophagus. Gut 2006;55:442.

108. Locke GR, Zinsmeister AR, Talley NJ. Can symptoms predict endoscopic findings in GERD? Gastrointest. Endosc. 2003;58:661-70.

109. Gerson LB, Edson R, Lavori PW, et al. Use of a simple symptom questionnaire to predict Barrett's esophagus in patients with symptoms of gastroesophageal reflux. Am. J. Gastroenterol. 2001;96:2005-12.

110. Wang KK, Sampliner RE, Practice Parameters Committee of the American College of G. Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett's esophagus. Am. J. Gastroenterol. 2008;103:788-97.

111. Spechler SJ. Dysplasia in Barrett's esophagus: limitations of current management strategies. Am. J. Gastroenterol. 2005;100:927-35.

112. Incarbone R, Bonavina L, Saino G, et al. Outcome of esophageal adenocarcinoma detected during endoscopic biopsy surveillance for Barrett's esophagus. Surg. Endosc. 2002;16:263-6.

113. Ferguson MK, Durkin A. Long-term survival after esophagectomy for Barrett's adenocarcinoma in endoscopically surveyed and nonsurveyed patients. J. Gastrointest. Surg. 2002;6:29-35; discussion 36.

114. Fountoulakis A, Zafirellis KD, Dolan K, et al. Effect of surveillance of Barrett's oesophagus on the clinical outcome of oesophageal cancer. Br. J. Surg. 2004;91:997-1003.

115. Corley DA, Levin TR, Habel LA, et al. Surveillance and survival in Barrett's adenocarcinomas: a population-based study. Gastroenterology 2002;122:633-40.

116. Downs-Kelly E, Mendelin JE, Bennett AE, et al. Poor interobserver agreement in the distinction of high-grade dysplasia and adenocarcinoma in pretreatment Barrett's esophagus biopsies. Am. J. Gastroenterol. 2008;103:2333-2340.

117. Moyes LH, Going JJ. Still waiting for predictive biomarkers in Barrett's oesophagus. J. Clin. Pathol. 2011;64:742-50.

118. Collard JM. High-grade dysplasia in Barrett's esophagus. The case for esophagectomy. Chest Surg. Clin. N. Am. 2002;12:77-92.

119. Dulai GS, Guha S, Kahn KL, et al. Preoperative prevalence of Barrett's esophagus in esophageal adenocarcinoma: a systematic review. Gastroenterology 2002;122:26-33.

120. Bytzer P, Christensen PB, Damkier P, et al. Adenocarcinoma of the esophagus and Barrett's esophagus: a population-based study. Am. J. Gastroenterol. 1999;94:86-91.

121. Kadri S, Lao-Sirieix P, Fitzgerald RC. Developing a nonendoscopic screening test for Barrett's esophagus. Biomark. Med. 2011;5:397-404.

122. Hirst NG, Gordon LG, Whiteman DC, et al. Is endoscopic surveillance for non-dysplastic Barrett's esophagus cost-effective? Review of economic evaluations. J. Gastroenterol. Hepatol. 2011;26:247-54.

123. Lin OS, Schembre DB, Mergener K, et al. Blinded comparison of esophageal capsule endoscopy versus conventional endoscopy for a diagnosis of Barrett's esophagus in patients with chronic gastroesophageal reflux. Gastrointest. Endosc. 2007;65:577-83.

124. American Society for Gastrointestinal Endoscopy Technology C. Mucosal ablation devices. Gastrointest. Endosc. 2008;68:1031-42.

125. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin. Pharmacol. Ther. 2001;69:89-95.

126. Mayeux R. Biomarkers: potential uses and limitations. NeuroRx 2004;1:182-8.

127. Bensalah K, Montorsi F, Shariat SF. Challenges of cancer biomarker profiling. Eur. Urol. 2007;52:1601-9.

128. Silins I, Hogberg J. Combined toxic exposures and human health: biomarkers of exposure and effect. Int. J. Environ. Res. Public Health 2011;8:629-47.

129. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 1994;266:66-71.

130. Pavlou MP, Diamandis EP, Blasutig IM. The long journey of cancer biomarkers from the bench to the clinic. Clin. Chem. 2013;59:147-57.

131. Mehta S, Shelling A, Muthukaruppan A, et al. Predictive and prognostic molecular markers for cancer medicine. Ther. Adv. Med. Oncol. 2010;2:125-48.

132. Gainor JF, Longo DL, Chabner BA. Pharmacodynamic biomarkers: falling short of the mark? Clin. Cancer Res. 2014;20:2587-94.

133. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. J. Natl. Cancer Inst. 2001;93:1054-61.

134. Sund M, Kalluri R. Tumor stroma derived biomarkers in cancer. Cancer Metastasis Rev. 2009;28:177-83.

135. Galipeau PC, Cowan DS, Sanchez CA, et al. 17p (p53) allelic losses, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. Proc. Natl. Acad. Sci. U. S. A. 1996;93:7081-4.

136. Fang M, Lew E, Klein M, et al. DNA abnormalities as marker of risk for progression of Barrett's esophagus to adenocarcinoma: image cytometric DNA analysis in formalin-fixed tissues. Am. J. Gastroenterol. 2004;99:1887-94.

137. Reid BJ, Levine DS, Longton G, et al. Predictors of progression to cancer in Barrett's esophagus: baseline histology and flow cytometry identify low- and high-risk patient subsets. Am. J. Gastroenterol. 2000;95:1669-76.

138. Wu TT, Watanabe T, Heitmiller R, et al. Genetic alterations in Barrett esophagus and adenocarcinomas of the esophagus and esophagogastric junction region. Am. J. Pathol. 1998;153:287-94.

139. Galipeau PC, Li X, Blount PL, et al. NSAIDs modulate CDKN2A, TP53, and DNA content risk for progression to esophageal adenocarcinoma. PLoS Med. 2007;4:e67.

140. Morris CD, Armstrong GR, Bigley G, et al. Cyclooxygenase-2 expression in the Barrett's metaplasia-dysplasia-adenocarcinoma sequence. Am. J. Gastroenterol. 2001;96:990-6.

141. Lagorce C, Paraf F, Vidaud D, et al. Cyclooxygenase-2 is expressed frequently and early in Barrett's oesophagus and associated adenocarcinoma. Histopathology 2003;42:457-65.

142. Shirvani VN, Ouatu-Lascar R, Kaur BS, et al. Cyclooxygenase 2 expression in Barrett's esophagus and adenocarcinoma: Ex vivo induction by bile salts and acid exposure. Gastroenterology 2000;118:487-96.

143. Allameh A, Rasmi Y, Nasseri-Moghaddam S, et al. Immunohistochemical analysis of selected molecular markers in esophagus precancerous, adenocarcinoma and squamous cell carcinoma in Iranian subjects. Cancer Epidemiol. 2009;33:79-84.

144. Wilson KT, Fu S, Ramanujam KS, et al. Increased expression of inducible nitric oxide synthase and cyclooxygenase-2 in Barrett's esophagus and associated adenocarcinomas. Cancer Res. 1998;58:2929-34.

145. Prins MJ, Verhage RJ, ten Kate FJ, et al. Cyclooxygenase isoenzyme-2 and vascular endothelial growth factor are associated with poor prognosis in esophageal adenocarcinoma. J. Gastrointest. Surg. 2012;16:956-66.

146. Bhandari P, Bateman AC, Mehta RL, et al. Prognostic significance of cyclooxygenase-2 (COX-2) expression in patients with surgically resectable adenocarcinoma of the oesophagus. BMC Cancer 2006;6:134.

147. Jenkins GJ, Harries K, Doak SH, et al. The bile acid deoxycholic acid (DCA) at neutral pH activates NF-kappaB and induces IL-8 expression in oesophageal cells in vitro. Carcinogenesis 2004;25:317-23.

148. O'Riordan JM, Abdel-latif MM, Ravi N, et al. Proinflammatory cytokine and nuclear factor kappa-B expression along the inflammation-metaplasia-dysplasia-adenocarcinoma sequence in the esophagus. Am. J. Gastroenterol. 2005;100:1257-64.

149. Abdel-Latif MM, O'Riordan J, Windle HJ, et al. NF-kappaB activation in esophageal adenocarcinoma: relationship to Barrett's metaplasia, survival, and response to neoadjuvant chemoradiotherapy. Ann. Surg. 2004;239:491-500.

150. Abdel-Latif MM, O'Riordan JM, Ravi N, et al. Activated nuclear factor-kappa B and cytokine profiles in the esophagus parallel tumor regression following neoadjuvant chemoradiotherapy. Dis. Esophagus 2005;18:246-52.

151. Jenkins GJ, Mikhail J, Alhamdani A, et al. Immunohistochemical study of nuclear factor-kappaB activity and interleukin-8 abundance in oesophageal adenocarcinoma; a useful strategy for monitoring these biomarkers. J. Clin. Pathol. 2007;60:1232-7.

152. Milano F, Jorritsma T, Rygiel AM, et al. Expression pattern of immune suppressive cytokines and growth factors in oesophageal adenocarcinoma reveal a tumour immune escape-promoting microenvironment. Scand. J. Immunol. 2008;68:616-23.

153. Dvorak K, Chavarria M, Payne CM, et al. Activation of the interleukin-6/STAT3 antiapoptotic pathway in esophageal cells by bile acids and low pH: relevance to barrett's esophagus. Clin. Cancer Res. 2007;13:5305-13.

154. Deans DA, Wigmore SJ, Gilmour H, et al. Elevated tumour interleukin-1beta is associated with systemic inflammation: A marker of reduced survival in gastro-oesophageal cancer. Br. J. Cancer 2006;95:1568-75.

155. Lukaszewicz-Zajac M, Mroczko B, Kozlowski M, et al. Higher importance of interleukin 6 than classic tumor markers (carcinoembryonic antigen and squamous cell cancer antigen) in the diagnosis of esophageal cancer patients. Dis. Esophagus 2012;25:242-9.

156. Nguyen GH, Schetter AJ, Chou DB, et al. Inflammatory and microRNA gene expression as prognostic classifier of Barrett's-associated esophageal adenocarcinoma. Clin. Cancer Res. 2010;16:5824-34.

157. Murray GI, Duncan ME, O'Neil P, et al. Matrix metalloproteinase-1 is associated with poor prognosis in oesophageal cancer. J. Pathol. 1998;185:256-61.

158. Salmela MT, Karjalainen-Lindsberg ML, Puolakkainen P, et al. Upregulation and differential expression of matrilysin (MMP-7) and metalloelastase (MMP-12) and their inhibitors TIMP-1 and TIMP-3 in Barrett's oesophageal adenocarcinoma. Br. J. Cancer 2001;85:383-92.

159. Grimm M, Lazariotou M, Kircher S, et al. MMP-1 is a (pre-)invasive factor in Barrett-associated esophageal adenocarcinomas and is associated with positive lymph node status. J. Transl. Med. 2010;8:99.

160. El-Kenawy Ael M, Lotfy M, El-Kott A, et al. Significance of matrix metalloproteinase 9 and CD34 expressions in esophageal carcinoma: correlation with DNA content. J. Clin. Gastroenterol. 2005;39:791-4.

161. Herszenyi L, Hritz I, Pregun I, et al. Alterations of glutathione S-transferase and matrix metalloproteinase-9 expressions are early events in esophageal carcinogenesis. World J. Gastroenterol. 2007;13:676-82.

162. Mroczko B, Kozlowski M, Groblewska M, et al. Expression of matrix metalloproteinase-9 in the neoplastic and interstitial inflammatory infiltrate cells in the different histopathological types of esophageal cancer. Folia Histochem. Cytobiol. 2008;46:471-8.

163. Darnton SJ, Hardie LJ, Muc RS, et al. Tissue inhibitor of metalloproteinase-3 (TIMP-3) gene is methylated in the development of esophageal adenocarcinoma: loss of expression correlates with poor prognosis. Int. J. Cancer 2005;115:351-8.

164. Bani-Hani K, Martin IG, Hardie LJ, et al. Prospective study of cyclin D1 overexpression in Barrett's esophagus: association with increased risk of adenocarcinoma. J. Natl. Cancer Inst. 2000;92:1316-21.

165. Shi XY, Bhagwandeen B, Leong AS. p16, cyclin D1, Ki-67, and AMACR as markers for dysplasia in Barrett esophagus. Appl. Immunohistochem. Mol. Morphol. 2008;16:447-52.

166. Kerkhof M, Kusters JG, van Dekken H, et al. Biomarkers for risk stratification of neoplastic progression in Barrett esophagus. Cell. Oncol. 2007;29:507-17.

167. Shimamoto C, Weinstein WM, Boland CR. Glycoconjugate expression in normal, metaplastic, and neoplastic human upper gastrointestinal mucosa. J. Clin. Invest. 1987;80:1670-8.

168. Neumann H, Wex T, Monkemuller K, et al. Lectin UEA-I-binding proteins are specifically increased in the squamous epithelium of patients with Barrett's esophagus. Digestion 2008;78:201-7.

169. Bird-Lieberman EL, Neves AA, Lao-Sirieix P, et al. Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Barrett's esophagus. Nat. Med. 2012;18:315-21.

170. Iwaya Y, Hasebe O, Koide N, et al. Reduced expression of alphaGlcNAc in Barrett's oesophagus adjacent to Barrett's adenocarcinoma--a possible biomarker to predict the malignant potential of Barrett's oesophagus. Histopathology 2014;64:536-46.

171. Zhai R, Zhao Y, Su L, et al. Genome-wide DNA methylation profiling of cell-free serum DNA in esophageal adenocarcinoma and Barrett esophagus. Neoplasia 2012;14:29-33.

172. Kawakami K, Brabender J, Lord RV, et al. Hypermethylated APC DNA in plasma and prognosis of patients with esophageal adenocarcinoma. J. Natl. Cancer Inst. 2000;92:1805-11.

173. Hoffmann AC, Vallbohmer D, Prenzel K, et al. Methylated DAPK and APC promoter DNA detection in peripheral blood is significantly associated with apparent residual tumor and outcome. J. Cancer Res. Clin. Oncol. 2009;135:1231-7.

174. Bird-Lieberman EL, Dunn JM, Coleman HG, et al. Population-Based Study Reveals New Risk-Stratification Biomarker Panel for Barrett's Esophagus. Gastroenterology 2012;143:927-35.

175. Mechref Y, Hussein A, Bekesova S, et al. Quantitative serum glycomics of esophageal adenocarcinoma and other esophageal disease onsets. J. Proteome Res. 2009;8:2656-66.

176. Mitra I, Zhuang Z, Zhang Y, et al. N-glycan profiling by microchip electrophoresis to differentiate disease states related to esophageal adenocarcinoma. Anal. Chem. 2012;84:3621-7.

177. Gaye MM, Valentine SJ, Hu Y, et al. Ion mobility-mass spectrometry analysis of serum N-linked glycans from esophageal adenocarcinoma phenotypes. J. Proteome Res. 2012;11:6102-10.

178. Mann B, Madera M, Klouckova I, et al. A quantitative investigation of fucosylated serum glycoproteins with application to esophageal adenocarcinoma. Electrophoresis 2010;31:1833-41.

179. Hammoud ZT, Mechref Y, Hussein A, et al. Comparative glycomic profiling in esophageal adenocarcinoma. J. Thorac. Cardiovasc. Surg. 2010;139:1216-23.

180. Song E, Zhu R, Hammoud ZT, et al. LC-MS/MS quantitation of esophagus disease blood serum glycoproteins by enrichment with hydrazide chemistry and lectin affinity chromatography. J. Proteome Res. 2014;13:4808−4820.

181. Djukovic D, Baniasadi HR, Kc R, et al. Targeted serum metabolite profiling of nucleosides in esophageal adenocarcinoma. Rapid Commun. Mass Spectrom. 2010;24:3057-62.

182. Zhang J, Bowers J, Liu L, et al. Esophageal cancer metabolite biomarkers detected by LC-MS and NMR methods. PLoS One 2012;7:e30181.

183. Zhang J, Liu L, Wei S, et al. Metabolomics study of esophageal adenocarcinoma. J. Thorac. Cardiovasc. Surg. 2011;141:469-75, e1-4.

184. Davis VW, Schiller DE, Eurich D, et al. Urinary metabolomic signature of esophageal cancer and Barrett's esophagus. World J. Surg. Oncol. 2012;10:271.

185. Jin Z, Cheng Y, Gu W, et al. A multicenter, double-blinded validation study of methylation biomarkers for progression prediction in Barrett's esophagus. Cancer Res. 2009;69:4112-5.

186. Sato F, Jin Z, Schulmann K, et al. Three-tiered risk stratification model to predict progression in Barrett's esophagus using epigenetic and clinical features. PLoS One 2008;3:e1890.

187. Brock MV, Gou M, Akiyama Y, et al. Prognostic importance of promoter hypermethylation of multiple genes in esophageal adenocarcinoma. Clin. Cancer Res. 2003;9:2912-9.

188. Hamilton JP, Sato F, Greenwald BD, et al. Promoter methylation and response to chemotherapy and radiation in esophageal cancer. Clin. Gastroenterol. Hepatol. 2006;4:701-8.

189. Casson AG, Evans SC, Gillis A, et al. Clinical implications of p53 tumor suppressor gene mutation and protein expression in esophageal adenocarcinomas: results of a ten-year prospective study. J. Thorac. Cardiovasc. Surg. 2003;125:1121-31.

190. Murray L, Sedo A, Scott M, et al. TP53 and progression from Barrett's metaplasia to oesophageal adenocarcinoma in a UK population cohort. Gut 2006;55:1390-7.

191. Walch AK, Zitzelsberger HF, Bruch J, et al. Chromosomal imbalances in Barrett's adenocarcinoma and the metaplasia-dysplasia-carcinoma sequence. Am. J. Pathol. 2000;156:555-66.

192. Lai LA, Paulson TG, Li X, et al. Increasing genomic instability during premalignant neoplastic progression revealed through high resolution array-CGH. Genes Chromosomes Cancer 2007;46:532-42.

193. Riegman PH, Vissers KJ, Alers JC, et al. Genomic alterations in malignant transformation of Barrett's esophagus. Cancer Res. 2001;61:3164-70.

194. Li X, Galipeau PC, Sanchez CA, et al. Single nucleotide polymorphism-based genome-wide chromosome copy change, loss of heterozygosity, and aneuploidy in Barrett's esophagus neoplastic progression. Cancer Prev. Res. (Phila.) 2008;1:413-23.

195. Wiech T, Nikolopoulos E, Weis R, et al. Genome-wide analysis of genetic alterations in Barrett's adenocarcinoma using single nucleotide polymorphism arrays. Lab. Invest. 2009;89:385-97.

196. Nancarrow DJ, Handoko HY, Smithers BM, et al. Genome-wide copy number analysis in esophageal adenocarcinoma using high-density single-nucleotide polymorphism arrays. Cancer Res. 2008;68:4163-72.

197. Paulson TG, Maley CC, Li X, et al. Chromosomal instability and copy number alterations in Barrett's esophagus and esophageal adenocarcinoma. Clin. Cancer Res. 2009;15:3305-14.

198. Robertson EV, Jankowski JA. Genetics of gastroesophageal cancer: paradigms, paradoxes, and prognostic utility. Am. J. Gastroenterol. 2008;103:443-9.

199. The Esophageal Adenocarcinoma Genetics C, The Wellcome Trust Case Control C, Su Z, et al. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. Nat. Genet. 2012;44:1131-1136.

200. Wu IC, Zhao Y, Zhai R, et al. Association between polymorphisms in cancer-related genes and early onset of esophageal adenocarcinoma. Neoplasia 2011;13:386-92.

201. Agrawal N, Jiao Y, Bettegowda C, et al. Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. Cancer Discov. 2012;2:899-905.

202. Streppel MM, Lata S, Delabastide M, et al. Next-generation sequencing of endoscopic biopsies identifies ARID1A as a tumor-suppressor gene in Barrett's esophagus. Oncogene 2014;33:347-57.

203. Wu W, Bhagat TD, Yang X, et al. Hypomethylation of Noncoding DNA Regions and Overexpression of the Long Noncoding RNA, AFAP1-AS1, in Barrett's Esophagus and Esophageal Adenocarcinoma. Gastroenterology 2013;144:956-966.e4.

204. Dulak AM, Stojanov P, Peng S, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. Nat. Genet. 2013;45:478-86.

205. Weaver JM, Ross-Innes CS, Shannon N, et al. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. Nat. Genet. 2014;46:837-43.

206. Hardie LJ, Darnton SJ, Wallis YL, et al. p16 expression in Barrett's esophagus and esophageal adenocarcinoma: association with genetic and epigenetic alterations. Cancer Lett. 2005;217:221-30.

207. Bian YS, Osterheld MC, Fontolliet C, et al. p16 inactivation by methylation of the CDKN2A promoter occurs early during neoplastic progression in Barrett's esophagus. Gastroenterology 2002;122:1113-21.

208. Wang JS, Guo M, Montgomery EA, et al. DNA promoter hypermethylation of p16 and APC predicts neoplastic progression in Barrett's esophagus. Am. J. Gastroenterol. 2009;104:2153-60.

209. Schulmann K, Sterian A, Berki A, et al. Inactivation of p16, RUNX3, and HPP1 occurs early in Barrett's-associated neoplastic progression and predicts progression risk. Oncogene 2005;24:4138-48.

210. Wong DJ, Barrett MT, Stoger R, et al. p16INK4a promoter is hypermethylated at a high frequency in esophageal adenocarcinomas. Cancer Res. 1997;57:2619-22.

211. Wong DJ, Paulson TG, Prevo LJ, et al. p16(INK4a) lesions are common, early abnormalities that undergo clonal expansion in Barrett's metaplastic epithelium. Cancer Res. 2001;61:8284-9.

212. Vieth M, Schneider-Stock R, Rohrich K, et al. INK4a-ARF alterations in Barrett's epithelium, intraepithelial neoplasia and Barrett's adenocarcinoma. Virchows Arch. 2004;445:135-41.

213. Kempster S, Phillips WA, Baindur-Hudson S, et al. Methylation of exon 2 of p16 is associated with late stage oesophageal cancer. Cancer Lett. 2000;150:57-62.

214. Baumann S, Keller G, Puhringer F, et al. The prognostic impact of O6-Methylguanine-DNA Methyltransferase (MGMT) promotor hypermethylation in esophageal adenocarcinoma. Int. J. Cancer 2006;119:264-8.

215. Kuester D, El-Rifai W, Peng D, et al. Silencing of MGMT expression by promoter hypermethylation in the metaplasia-dysplasia-carcinoma sequence of Barrett's esophagus. Cancer Lett. 2009;275:117-26.

216. Clement G, Braunschweig R, Pasquier N, et al. Alterations of the Wnt signaling pathway during the neoplastic progression of Barrett's esophagus. Oncogene 2006;25:3084-92.

217. Clement G, Bosman FT, Fontolliet C, et al. Monoallelic methylation of the APC promoter is altered in normal gastric mucosa associated with neoplastic lesions. Cancer Res. 2004;64:6867-73.

218. Peng DF, Razvi M, Chen H, et al. DNA hypermethylation regulates the expression of members of the Mu-class glutathione S-transferases and glutathione peroxidases in Barrett's adenocarcinoma. Gut 2009;58:5-15.

219. Lee OJ, Schneider-Stock R, McChesney PA, et al. Hypermethylation and loss of expression of glutathione peroxidase-3 in Barrett's tumorigenesis. Neoplasia 2005;7:854-61.

220. Gu P, Xing X, Tanzer M, et al. Frequent loss of TIMP-3 expression in progression of esophageal and gastric adenocarcinomas. Neoplasia 2008;10:563-72.

221. Kuester D, Dar AA, Moskaluk CC, et al. Early involvement of death-associated protein kinase promoter hypermethylation in the carcinogenesis of Barrett's esophageal adenocarcinoma and its association with clinical progression. Neoplasia 2007;9:236-45.

222. Jin Z, Olaru A, Yang J, et al. Hypermethylation of tachykinin-1 is a potential biomarker in human esophageal cancer. Clin. Cancer Res. 2007;13:6293-300.

223. Hamilton JP, Sato F, Jin Z, et al. Reprimo methylation is a potential biomarker of Barrett's-Associated esophageal neoplastic progression. Clin. Cancer Res. 2006;12:6637-42.

224. Corn PG, Heath EI, Heitmiller R, et al. Frequent hypermethylation of the 5' CpG island of E-cadherin in esophageal adenocarcinoma. Clin. Cancer Res. 2001;7:2765-9.

225. Tischoff I, Hengge UR, Vieth M, et al. Methylation of SOCS-3 and SOCS-1 in the carcinogenesis of Barrett's adenocarcinoma. Gut 2007;56:1047-53.

226. Zou H, Molina JR, Harrington JJ, et al. Aberrant methylation of secreted frizzled-related protein genes in esophageal adenocarcinoma and Barrett's esophagus. Int. J. Cancer 2005;116:584-91.

227. Kaz AM, Luo Y, Dzieciatkowski S, et al. Aberrantly methylated PKP1 in the progression of Barrett's esophagus to esophageal adenocarcinoma. Genes Chromosomes Cancer 2012;51:384-93.

228. Guo M, House MG, Akiyama Y, et al. Hypermethylation of the GATA gene family in esophageal cancer. Int. J. Cancer 2006;119:2078-83.

229. Jin Z, Cheng Y, Olaru A, et al. Promoter hypermethylation of CDH13 is a common, early event in human esophageal adenocarcinogenesis and correlates with clinical risk factors. Int. J. Cancer 2008;123:2331-6.

230. Jin Z, Mori Y, Yang J, et al. Hypermethylation of the nel-like 1 gene is a common and early event and is associated with poor prognosis in early-stage esophageal adenocarcinoma. Oncogene 2007;26:6332-40.

231. Zou H, Osborn NK, Harrington JJ, et al. Frequent methylation of eyes absent 4 gene in Barrett's esophagus and esophageal adenocarcinoma. Cancer Epidemiol. Biomarkers Prev. 2005;14:830-4.

232. Jin Z, Hamilton JP, Yang J, et al. Hypermethylation of the AKAP12 promoter is a biomarker of Barrett's-associated esophageal neoplastic progression. Cancer Epidemiol. Biomarkers Prev. 2008;17:111-7.

233. Moinova H, Leidner RS, Ravi L, et al. Aberrant vimentin methylation is characteristic of upper gastrointestinal pathologies. Cancer Epidemiol. Biomarkers Prev. 2012;21:594-600.

234. Clement G, Guilleret I, He B, et al. Epigenetic alteration of the Wnt inhibitory factor-1 promoter occurs early in the carcinogenesis of Barrett's esophagus. Cancer Sci. 2008;99:46-53.

235. Soutto M, Peng D, Razvi M, et al. Epigenetic and genetic silencing of CHFR in esophageal adenocarcinomas. Cancer 2010;116:4033-42.

236. Peng D, Hu TL, Jiang A, et al. Location-specific epigenetic regulation of the metallothionein 3 gene in esophageal adenocarcinomas. PLoS One 2011;6:e22009.

237. Smith E, De Young NJ, Pavey SJ, et al. Similarity of aberrant DNA methylation in Barrett's esophagus and esophageal adenocarcinoma. Mol. Cancer 2008;7:75.

238. Kaz AM, Wong CJ, Luo Y, et al. DNA methylation profiling in Barrett's esophagus and esophageal adenocarcinoma reveals unique methylation signatures and molecular subclasses. Epigenetics 2011;6:1403-12.

239. Alvi MA, Liu X, O'Donovan M, et al. DNA Methylation as an Adjunct to Histopathology to Detect Prevalent, Inconspicuous Dysplasia and Early-Stage Neoplasia in Barrett's Esophagus. Clin. Cancer Res. 2013;19:878-88.

240. Clement G, Braunschweig R, Pasquier N, et al. Methylation of APC, TIMP3, and TERT: a new predictive marker to distinguish Barrett's oesophagus patients at risk for malignant transformation. J. Pathol. 2006;208:100-7.

241. van Baal JW, Milano F, Rygiel AM, et al. A comparative analysis by SAGE of gene expression profiles of Barrett's esophagus, normal squamous esophagus, and gastric cardia. Gastroenterology 2005;129:1274-81.

242. Sabo E, Meitner PA, Tavares R, et al. Expression analysis of Barrett's esophagus-associated high-grade dysplasia in laser capture microdissected archival tissue. Clin. Cancer Res. 2008;14:6440-8.

243. Peters CJ, Rees JR, Hardwick RH, et al. A 4-gene signature predicts survival of patients with resected adenocarcinoma of the esophagus, junction, and gastric cardia. Gastroenterology 2010;139:1995-2004 e15.

244. Ostrowski J, Mikula M, Karczmarski J, et al. Molecular defense mechanisms of Barrett's metaplasia estimated by an integrative genomics. J. Mol. Med. (Berl.) 2007;85:733-43.

245. Lagarde SM, Ver Loren van Themaat PE, Moerland PD, et al. Analysis of gene expression identifies differentially expressed genes and pathways associated with lymphatic dissemination in patients with adenocarcinoma of the esophagus. Ann. Surg. Oncol. 2008;15:3459-70.

246. van Dekken H, Vissers K, Tilanus HW, et al. Genomic array and expression analysis of frequent high-level amplifications in adenocarcinomas of the gastro-esophageal junction. Cancer Genet. Cytogenet. 2006;166:157-62.

247. Akagi T, Ito T, Kato M, et al. Chromosomal abnormalities and novel disease-related regions in progression from Barrett's esophagus to esophageal adenocarcinoma. Int. J. Cancer 2009;125:2349-59.

248. Greenawalt DM, Duong C, Smyth GK, et al. Gene expression profiling of esophageal cancer: comparative analysis of Barrett's esophagus, adenocarcinoma, and squamous cell carcinoma. Int. J. Cancer 2007;120:1914-21.

249. Gu J, Ajani JA, Hawk ET, et al. Genome-wide catalogue of chromosomal aberrations in barrett's esophagus and esophageal adenocarcinoma: a high-density single nucleotide polymorphism array analysis. Cancer Prev. Res. (Phila.) 2010;3:1176-86.

250. Kim SM, Park YY, Park ES, et al. Prognostic biomarkers for esophageal adenocarcinoma identified by analysis of tumor transcriptome. PLoS One 2010;5:e15074.

251. van Dekken H, Tilanus HW, Hop WC, et al. Array comparative genomic hybridization, expression array, and protein analysis of critical regions on chromosome arms 1q, 7q, and 8p in adenocarcinomas of the gastroesophageal junction. Cancer Genet. Cytogenet. 2009;189:37-42.

252. Lai LA, Kostadinov R, Barrett MT, et al. Deletion at fragile sites is a common and early event in Barrett's esophagus. Mol. Cancer Res. 2010;8:1084-94.

253. van Duin M, van Marion R, Vissers KJ, et al. High-resolution array comparative genomic hybridization of chromosome 8q: evaluation of putative progression markers for gastroesophageal junction adenocarcinomas. Cytogenet Genome Res 2007;118:130-7.

254. Xu Y, Selaru FM, Yin J, et al. Artificial neural networks and gene filtering distinguish between global gene expression profiles of Barrett's esophagus and esophageal cancer. Cancer Res. 2002;62:3493-7.

255. Kimchi ET, Posner MC, Park JO, et al. Progression of Barrett's metaplasia to adenocarcinoma is associated with the suppression of the transcriptional programs of epidermal differentiation. Cancer Res. 2005;65:3146-54.

256. Barrett MT, Yeung KY, Ruzzo WL, et al. Transcriptional analyses of Barrett's metaplasia and normal upper GI mucosae. Neoplasia 2002;4:121-8.

257. Wang S, Zhan M, Yin J, et al. Transcriptional profiling suggests that Barrett's metaplasia is an early intermediate stage in esophageal adenocarcinogenesis. Oncogene 2006;25:3346-56.

258. Hammoud ZT, Badve S, Zhao Q, et al. Differential gene expression profiling of esophageal adenocarcinoma. J. Thorac. Cardiovasc. Surg. 2009;137:829-34.

259. Gomes LI, Esteves GH, Carvalho AF, et al. Expression profile of malignant and nonmalignant lesions of esophagus and stomach: differential activity of functional modules related to inflammation and lipid metabolism. Cancer Res. 2005;65:7127-36.

260. Hao Y, Triadafilopoulos G, Sahbaie P, et al. Gene expression profiling reveals stromal genes expressed in common between Barrett's esophagus and adenocarcinoma. Gastroenterology 2006;131:925-33.

261. Brabender J, Marjoram P, Salonga D, et al. A multigene expression panel for the molecular diagnosis of Barrett's esophagus and Barrett's adenocarcinoma of the esophagus. Oncogene 2004;23:4780-8.

262. Lao-Sirieix P, Boussioutas A, Kadri SR, et al. Non-endoscopic screening biomarkers for Barrett's oesophagus: from microarray analysis to the clinic. Gut 2009;58:1451-9.

263. Goh XY, Rees JR, Paterson AL, et al. Integrative analysis of array-comparative genomic hybridisation and matched gene expression profiling data reveals novel genes with prognostic significance in oesophageal adenocarcinoma. Gut 2011;60:1317-26.

264. Rao S, Welsh L, Cunningham D, et al. Correlation of overall survival with gene expression profiles in a prospective study of resectable esophageal cancer. Clin. Colorectal Cancer 2011;10:48-56.

265. Schauer M, Janssen KP, Rimkus C, et al. Microarray-based response prediction in esophageal adenocarcinoma. Clin. Cancer Res. 2010;16:330-7.

266. Luthra MG, Ajani JA, Izzo J, et al. Decreased expression of gene cluster at chromosome 1q21 defines molecular subgroups of chemoradiotherapy response in esophageal cancers. Clin. Cancer Res. 2007;13:912-9.

267. Luthra R, Wu TT, Luthra MG, et al. Gene expression profiling of localized esophageal carcinomas: association with pathologic response to preoperative chemoradiation. J. Clin. Oncol. 2006;24:259-67.

268. Wu X, Gu J, Wu TT, et al. Genetic variations in radiation and chemotherapy drug action pathways predict clinical outcomes in esophageal cancer. J. Clin. Oncol. 2006;24:3789-98.

269. Majka J, Rembiasz K, Migaczewski M, et al. Cyclooxygenase-2 (COX-2) is the key event in pathophysiology of Barrett's esophagus. Lesson from experimental animal model and human subjects. J. Physiol. Pharmacol. 2010;61:409-18.

270. Uefuji K, Ichikura T, Mochizuki H. Cyclooxygenase-2 expression is related to prostaglandin biosynthesis and angiogenesis in human gastric cancer. Clin. Cancer Res. 2000;6:135-8.

271. Kuramochi H, Vallbohmer D, Uchida K, et al. Quantitative, tissue-specific analysis of cyclooxygenase gene expression in the pathogenesis of Barrett's adenocarcinoma. J. Gastrointest. Surg. 2004;8:1007-16.

272. Brien TP, Odze RD, Sheehan CE, et al. HER-2/neu gene amplification by FISH predicts poor survival in Barrett's esophagus-associated adenocarcinoma. Hum. Pathol. 2000;31:35-9.

273. Wang KL, Wu TT, Choi IS, et al. Expression of epidermal growth factor receptor in esophageal and esophagogastric junction adenocarcinomas: association with poor outcome. Cancer 2007;109:658-67.

274. Raouf AA, Evoy DA, Carton E, et al. Loss of Bcl-2 expression in Barrett's dysplasia and adenocarcinoma is associated with tumor progression and worse survival but not with response to neoadjuvant chemoradiation. Dis. Esophagus 2003;16:17-23.

275. Auvinen MI, Sihvo EI, Ruohtula T, et al. Incipient angiogenesis in Barrett's epithelium and lymphangiogenesis in Barrett's adenocarcinoma. J. Clin. Oncol. 2002;20:2971-9.

276. Mobius C, Stein HJ, Becker I, et al. Vascular endothelial growth factor expression and neovascularization in Barrett's carcinoma. World J. Surg. 2004;28:675-9.

277. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 1993;75:843-54.

278. Metias SM, Lianidou E, Yousef GM. MicroRNAs in clinical oncology: at the crossroads between promises and problems. J. Clin. Pathol. 2009;62:771-6.

279. Feber A, Xi L, Luketich JD, et al. MicroRNA expression profiles of esophageal cancer. J. Thorac. Cardiovasc. Surg. 2008;135:255-60.

280. Smith CM, Watson DI, Leong MP, et al. miR-200 family expression is downregulated upon neoplastic progression of Barrett's esophagus. World J. Gastroenterol. 2011;17:1036-44.

281. Luthra R, Singh RR, Luthra MG, et al. MicroRNA-196a targets annexin A1: a microRNA-mediated mechanism of annexin A1 downregulation in cancers. Oncogene 2008;27:6667-78.

282. Maru DM, Singh RR, Hannah C, et al. MicroRNA-196a is a potential marker of progression during Barrett's metaplasia-dysplasia-invasive adenocarcinoma sequence in esophagus. Am. J. Pathol. 2009;174:1940-8.

283. Braun CJ, Zhang X, Savelyeva I, et al. p53-Responsive micrornas 192 and 215 are capable of inducing cell cycle arrest. Cancer Res. 2008;68:10094-104.

284. Wijnhoven BP, Hussey DJ, Watson DI, et al. MicroRNA profiling of Barrett's oesophagus and oesophageal adenocarcinoma. Br. J. Surg. 2010;97:853-61.

285. Bansal A, Lee IH, Hong X, et al. Feasibility of microRNAs as biomarkers for Barrett's Esophagus progression: a pilot cross-sectional, phase 2 biomarker study. Am. J. Gastroenterol. 2011;106:1055-63.

286. Mathe EA, Nguyen GH, Bowman ED, et al. MicroRNA expression in squamous cell carcinoma and adenocarcinoma of the esophagus: associations with survival. Clin. Cancer Res. 2009;15:6192-200.

287. Yang H, Gu J, Wang KK, et al. MicroRNA expression signatures in Barrett's esophagus and esophageal adenocarcinoma. Clin. Cancer Res. 2009;15:5744-52.

288. Mees ST, Mardin WA, Wendel C, et al. EP300--a miRNA-regulated metastasis suppressor gene in ductal adenocarcinomas of the pancreas. Int. J. Cancer 2010;126:114-24.

289. Fassan M, Volinia S, Palatini J, et al. MicroRNA expression profiling in human Barrett's carcinogenesis. Int. J. Cancer 2011;129:1661-70.

290. Wu X, Ajani JA, Gu J, et al. MicroRNA expression signatures during malignant progression from Barrett's esophagus to esophageal adenocarcinoma. Cancer Prev. Res. (Phila.) 2013;6:196-205.

291. Nancarrow DJ, Clouston AD, Smithers BM, et al. Whole genome expression array profiling highlights differences in mucosal defense genes in Barrett's esophagus and esophageal adenocarcinoma. PLoS One 2011;6:e22513.

292. Streppel MM, Pai S, Campbell NR, et al. MicroRNA 223 is upregulated in the multistep progression of Barrett's esophagus and modulates sensitivity to chemotherapy by targeting PARP1. Clin. Cancer Res. 2013;19:4067-78.

293. D'Souza-Schorey C, Clancy JW. Tumor-derived microvesicles: shedding light on novel microenvironment modulators and prospective cancer biomarkers. Genes Dev. 2012;26:1287-99.

294. Zen K, Zhang CY. Circulating microRNAs: a novel class of biomarkers to diagnose and monitor human cancers. Med. Res. Rev. 2012;32:326-48.

295. Luzna P, Gregar J, Uberall I, et al. Changes of microRNAs-192, 196a and 203 correlate with Barrett's esophagus diagnosis and its progression compared to normal healthy individuals. Diagn. Pathol. 2011;6:114.

296. Leidner RS, Ravi L, Leahy P, et al. The microRNAs, MiR-31 and MiR-375, as candidate markers in Barrett's esophageal carcinogenesis. Genes Chromosomes Cancer 2012;51:473-9.

297. Feber A, Xi L, Pennathur A, et al. MicroRNA prognostic signature for nodal metastases and survival in esophageal adenocarcinoma. Ann. Thorac. Surg. 2011;91:1523-30.

298. Revilla-Nuin B, Parrilla P, Lozano J, et al. Predictive Value of MicroRNAs in the Progression of Barrett Esophagus to Adenocarcinoma in a Long-Term Follow-Up Study. Ann. Surg. 2012.

299. Hong L, Han Y, Zhang H, et al. The prognostic and chemotherapeutic value of miR-296 in esophageal squamous cell carcinoma. Ann. Surg. 2010;251:1056-63.

300. Hummel R, Hussey DJ, Michael MZ, et al. MiRNAs and their association with locoregional staging and survival following surgery for esophageal carcinoma. Ann. Surg. Oncol. 2011;18:253-60.

301. Catenacci DV, Cervantes G, Yala S, et al. RON (MST1R) is a novel prognostic marker and therapeutic target for gastroesophageal adenocarcinoma. Cancer Biol. Ther. 2011;12:9-46.

302. Fassan M, Pizzi M, Realdon S, et al. The HER2-miR125a5p/miR125b loop in gastric and esophageal carcinogenesis. Hum. Pathol. 2013;44:1804-10.

303. Wang HB, Jiang ZB, Li M. Research on the Typical miRNA and Target Genes in Squamous Cell Carcinoma and Adenocarcinoma of Esophagus Cancer with DNA Microarray. Pathol. Oncol. Res. 2014.

304. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. Nat. Rev. Mol. Cell Biol. 2012;13:448-62.

305. Pan S, Chen R, Aebersold R, et al. Mass spectrometry based glycoproteomics--from a proteomics perspective. Mol. Cell. Proteomics 2011;10:R110 003251.

306. Varki A, Sharon N. Historical Background and Overview. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY); 2009.

307. Stanley P, Schachter H, Taniguchi N. N-Glycans. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY); 2009.

308. Kim EH, Misek DE. Glycoproteomics-based identification of cancer biomarkers. Int J Proteomics 2011;2011:601937.

309. Fanayan S, Hincapie M, Hancock WS. Using lectins to harvest the plasma/serum glycoproteome. Electrophoresis 2012;33:1746-54.

310. Ongay S, Boichenko A, Govorukhina N, et al. Glycopeptide enrichment and separation for protein glycosylation analysis. J Sep Sci 2012;35:2341-72.

311. Poorkhalkali N, Jacobson I, Helander HF. Lectin histochemistry of the esophagus in several mammalian species. Anat. Embryol. (Berl.) 1999;200:541-9.

312. Byrne AM, Sharma R, Duggan G, et al. Deoxycholic acid impairs glycosylation and fucosylation processes in esophageal epithelial cells. Glycobiology 2012;22:638-48.

313. Allard WJ, Matera J, Miller MC, et al. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. Clin. Cancer Res. 2004;10:6897-904.

314. Devriese LA, Voest EE, Beijnen JH, et al. Circulating tumor cells as pharmacodynamic biomarker in early clinical oncological trials. Cancer Treat. Rev. 2011;37:579-89.

315. Chen ST, Pan TL, Juan HF, et al. Breast tumor microenvironment: proteomics highlights the treatments targeting secretome. J. Proteome Res. 2008;7:1379-87.

316. Pitteri SJ, Kelly-Spratt KS, Gurley KE, et al. Tumor microenvironment-derived proteins dominate the plasma proteome response during breast cancer induction and progression. Cancer Res. 2011;71:5090-100.

317. Choi E, Loo D, Dennis JW, et al. High-throughput lectin magnetic bead array-coupled tandem mass spectrometry for glycoprotein biomarker discovery. Electrophoresis 2011;32:3564-75.

318. Loo D, Jones A, Hill MM. Lectin magnetic bead array for biomarker discovery. J. Proteome Res. 2010;9:5496-500.

319. Yakoub D, Keun HC, Goldin R, et al. Metabolic profiling detects field effects in nondysplastic tissue from esophageal cancer patients. Cancer Res. 2010;70:9129-36.

320. Buys SS, Partridge E, Greene MH, et al. Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: findings from the initial screen of a randomized trial. Am. J. Obstet. Gynecol. 2005;193:1630-9.

321. Barry MJ. Clinical practice. Prostate-specific-antigen testing for early diagnosis of prostate cancer. N. Engl. J. Med. 2001;344:1373-7.

322. Meany DL, Chan DW. Aberrant glycosylation associated with enzymes as cancer biomarkers. Clin. Proteomics 2011;8:7.

323. Peracaula R, Tabares G, Royle L, et al. Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. Glycobiology 2003;13:457-70.

324. Ohyama C, Hosono M, Nitta K, et al. Carbohydrate structure and differential binding of prostate specific antigen to Maackia amurensis lectin between prostate cancer and benign prostate hypertrophy. Glycobiology 2004;14:671-9.

325. Fukushima K, Satoh T, Baba S, et al. alpha1,2-Fucosylated and beta-N-acetylgalactosaminylated prostate-specific antigen as an efficient marker of prostatic cancer. Glycobiology 2010;20:452-60.

326. Meany DL, Zhang Z, Sokoll LJ, et al. Glycoproteomics for prostate cancer detection: changes in serum PSA glycosylation patterns. J. Proteome Res. 2009;8:613-9.

327. Adamczyk B, Tharmalingam T, Rudd PM. Glycans as cancer biomarkers. Biochim. Biophys. Acta 2012;1820:1347-53.

328. Noda K, Miyoshi E, Uozumi N, et al. Gene expression of alpha1-6 fucosyltransferase in human hepatoma tissues: a possible implication for increased fucosylation of alpha-fetoprotein. Hepatology 1998;28:944-52.

329. Kagebayashi C, Yamaguchi I, Akinaga A, et al. Automated immunoassay system for AFP-L3% using on-chip electrokinetic reaction and separation by affinity electrophoresis. Anal. Biochem. 2009;388:306-11.

330. Taketa K, Endo Y, Sekiya C, et al. A collaborative study for the evaluation of lectin-reactive alpha-fetoproteins in early detection of hepatocellular carcinoma. Cancer Res. 1993;53:5419-23.

331. Shiraki K, Takase K, Tameda Y, et al. A clinical study of lectin-reactive alpha-fetoprotein as an early indicator of hepatocellular carcinoma in the follow-up of cirrhotic patients. Hepatology 1995;22:802-7.

332. Oda K, Ido A, Tamai T, et al. Highly sensitive lens culinaris agglutinin-reactive alpha-fetoprotein is useful for early detection of hepatocellular carcinoma in patients with chronic liver disease. Oncol. Rep. 2011;26:1227-33.

333. Muthana SM, Campbell CT, Gildersleeve JC. Modifications of glycans: biological significance and therapeutic opportunities. ACS Chem. Biol. 2012;7:31-43.

334. Granovsky M, Fata J, Pawling J, et al. Suppression of tumor growth and metastasis in Mgat5-deficient mice. Nat. Med. 2000;6:306-12.

335. Sellers TA, Huang Y, Cunningham J, et al. Association of single nucleotide polymorphisms in glycosylation genes with risk of epithelial ovarian cancer. Cancer Epidemiol. Biomarkers Prev. 2008;17:397-404.

336. Ishizuka H, Nakayama T, Matsuoka S, et al. Prediction of the development of hepato-cellular-carcinoma in patients with liver cirrhosis by the serial determinations of serum alpha-L-fucosidase activity. Intern. Med. 1999;38:927-31.

337. Matsumoto K, Shimizu C, Arao T, et al. Identification of predictive biomarkers for response to trastuzumab using plasma FUCA activity and N-glycan identified by MALDI-TOF-MS. J. Proteome Res. 2009;8:457-62.

338. Mbeunkui F, Johann DJ, Jr. Cancer and the tumor microenvironment: a review of an essential relationship. Cancer Chemother. Pharmacol. 2009;63:571-82.

339. Celis JE, Gromov P, Cabezon T, et al. Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. Mol. Cell. Proteomics 2004;3:327-44.

340. Theilgaard-Monch K, Jacobsen LC, Nielsen MJ, et al. Haptoglobin is synthesized during granulocyte differentiation, stored in specific granules, and released by neutrophils in response to activation. Blood 2006;108:353-61.

341. Narimatsu H, Sawaki H, Kuno A, et al. A strategy for discovery of cancer glyco-biomarkers in serum using newly developed technologies for glycoproteomics. FEBS J 2010;277:95-105.

342. Wei X, Li L. Comparative glycoproteomics: approaches and applications. Brief Funct Genomic Proteomic 2009;8:104-13.

343. Ma C, Zhao X, Han H, et al. N-linked glycoproteome profiling of human serum using tandem enrichment and multiple fraction concatenation. Electrophoresis 2013;34:2440-50.

344. Sharon N. Lectin-carbohydrate complexes of plants and animals: an atomic view. Trends Biochem. Sci. 1993;18:221-6.

345. Yang Z, Hancock WS. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. J. Chromatogr. A 2004;1053:79-88.

346. Durham M, Regnier FE. Targeted glycoproteomics: serial lectin affinity chromatography in the selection of O-glycosylation sites on proteins from the human blood proteome. J. Chromatogr. A 2006;1132:165-73.

347. Gupta G, Surolia A, Sampathkumar SG. Lectin microarrays for glycomic analysis. OMICS 2010;14:419-36.

348. Ruhaak LR, Huhn C, Waterreus WJ, et al. Hydrophilic interaction chromatography-based high-throughput sample preparation method for N-glycan analysis from total human plasma glycoproteins. Anal. Chem. 2008;80:6119-26.

349. Zhang H, Li XJ, Martin DB, et al. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat. Biotechnol. 2003;21:660-6.

350. Zhou Y, Aebersold R, Zhang H. Isolation of N-linked glycopeptides from plasma. Anal. Chem. 2007;79:5826-37.

351. Zhou W, Yao N, Yao G, et al. Facile synthesis of aminophenylboronic acid-functionalized magnetic nanoparticles for selective separation of glycopeptides and glycoproteins. Chem. Commun. (Camb.) 2008:5577-9.

352. Whiteman DC, Sadeghi S, Pandeya N, et al. Combined effects of obesity, acid reflux and smoking on the risk of adenocarcinomas of the oesophagus. Gut 2008;57:173-80.

353. Smith KJ, O'Brien SM, Smithers BM, et al. Interactions among smoking, obesity, and symptoms of acid reflux in Barrett's esophagus. Cancer Epidemiol. Biomarkers Prev. 2005;14:2481-6.

354. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227:680-5.

355. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012;9:671-5.

356. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat. Biotechnol. 2006;24:971-83.

357. Majewski IJ, Bernards R. Taming the dragon: genomic biomarkers to individualize the treatment of cancer. Nat. Med. 2011;17:304-12.

358. Drabovich AP, Pavlou MP, Batruch I, et al. Chapter 2 - Proteomic and Mass Spectrometry Technologies for Biomarker Discovery. In: Veenstra HJID, editor. Proteomic and Metabolomic Approaches to Biomarker Discovery. Boston: Academic Press; 2013. p. 17-37.

359. Legrain P, Aebersold R, Archakov A, et al. The human proteome project: current state and future direction. Mol. Cell. Proteomics 2011;10:M111 009993.

360. Wishart DS, Knox C, Guo AC, et al. HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res. 2009;37:D603-10.

361. Wishart DS, Jewison T, Guo AC, et al. HMDB 3.0--The Human Metabolome Database in 2013. Nucleic Acids Res. 2013;41:D801-7.

362. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol. Cell. Proteomics 2002;1:845-67.

363. Jensen ON. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. Curr. Opin. Chem. Biol. 2004;8:33-41.

364. Kelleher NL. Top-down proteomics. Anal. Chem. 2004;76:197A-203A.

365. Patrie SM, Roth MJ, Zhang J. Chapter 20 - Top-Down Mass Spectrometry for Protein Molecular Diagnostics and Biomarker Discovery. In: Veenstra HJID, editor. Proteomic and Metabolomic Approaches to Biomarker Discovery. Boston: Academic Press; 2013. p. 313-332.

366. Tran JC, Zamdborg L, Ahlf DR, et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature 2011;480:254-8.

367. Zhang Y, Fonslow BR, Shan B, et al. Protein analysis by shotgun/bottom-up proteomics. Chem. Rev. 2013;113:2343-94.

368. Andrews GL, Simons BL, Young JB, et al. Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). Anal. Chem. 2011;83:5442-6.

369. Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal. Chem. 2000;72:1156-62.

370. Hu Q, Noll RJ, Li H, et al. The Orbitrap: a new mass spectrometer. J. Mass Spectrom. 2005;40:430-43.

371. Bogdanov B, Smith RD. Proteomics by FTICR mass spectrometry: top down and bottom up. Mass Spectrom. Rev. 2005;24:168-200.

372. Scigelova M, Hornshaw M, Giannakopulos A, et al. Fourier transform mass spectrometry. Mol. Cell. Proteomics 2011;10:M111 009431.

373. Ong SE, Blagoev B, Kratchmarova I, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics 2002;1:376-86.

374. Yao X, Freas A, Ramirez J, et al. Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus. Anal. Chem. 2001;73:2836-42.

375. Gygi SP, Rist B, Gerber SA, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 1999;17:994-9.

376. Ross PL, Huang YN, Marchese JN, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics 2004;3:1154-69.

377. Thompson A, Schafer J, Kuhn K, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 2003;75:1895-904.

378. Trudgian DC, Ridlova G, Fischer R, et al. Comparative evaluation of label-free SINQ normalized spectral index quantitation in the central proteomics facilities pipeline. Proteomics 2011;11:2790-7.

379. Long FH. Chapter 19 - Multivariate Analysis for Metabolomics and Proteomics Data. In: Veenstra HJID, editor. Proteomic and Metabolomic Approaches to Biomarker Discovery. Boston: Academic Press; 2013. p. 299-311.

380. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J. Clin. Epidemiol. 2001;54:774-81.

381. Jolliffe I. Principal component analysis. In: Principal component analysis. Encyclopedia of statistics in behavioral science: John Wiley & Sons, Ltd; 2005.

382. Le Cao KA, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. BMC Bioinformatics 2011;12:253.

383. Andrici J, Tio M, Cox MR, et al. Hiatal hernia and the risk of Barrett's esophagus. J. Gastroenterol. Hepatol. 2013;28:415-31.

384. Schulze WX, Usadel B. Quantitation in mass-spectrometry-based proteomics. Annu. Rev. Plant Biol. 2010;61:491-516.

385. Asara JM, Christofk HR, Freimark LM, et al. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. Proteomics 2008;8:994-9.

386. Broadhurst D, Kell D. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2006;2:171-196.

387. Plavina T, Wakshull E, Hancock WS, et al. Combination of abundant protein depletion and multi-lectin affinity chromatography (M-LAC) for plasma protein biomarker discovery. J. Proteome Res. 2007;6:662-71.

388. Sadrzadeh SM, Bozorgmehr J. Haptoglobin phenotypes in health and disorders. Am. J. Clin. Pathol. 2004;121 Suppl:S97-104.

389. Kristiansen M, Graversen JH, Jacobsen C, et al. Identification of the haemoglobin scavenger receptor. Nature 2001;409:198-201.

390. Ratanasopa K, Chakane S, Ilyas M, et al. Trapping of human hemoglobin by haptoglobin: molecular mechanisms and clinical applications. Antioxid Redox Signal 2013;18:2364-74.

391. Park SY, Lee SH, Kawasaki N, et al. alpha1-3/4 fucosylation at Asn 241 of beta-haptoglobin is a novel marker for colon cancer: a combinatorial approach for development of glycan biomarkers. Int. J. Cancer 2012;130:2366-76.

392. Ang IL, Poon TC, Lai PB, et al. Study of serum haptoglobin and its glycoforms in the diagnosis of hepatocellular carcinoma: a glycoproteomic approach. J. Proteome Res. 2006;5:2691-700.

393. Yoon SJ, Park SY, Pang PC, et al. N-glycosylation status of beta-haptoglobin in sera of patients with prostate cancer vs. benign prostate diseases. Int. J. Oncol. 2010;36:193-203.

394. Miyoshi E, Nakano M. Fucosylated haptoglobin is a novel marker for pancreatic cancer: detailed analyses of oligosaccharide structures. Proteomics 2008;8:3257-62.

395. Okuyama N, Ide Y, Nakano M, et al. Fucosylated haptoglobin is a novel marker for pancreatic cancer: a detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation. Int. J. Cancer 2006;118:2803-8.

396. Hruban RH, van Mansfeld AD, Offerhaus GJ, et al. K-ras oncogene activation in adenocarcinoma of the human pancreas. A study of 82 carcinomas using a combination of mutant-enriched polymerase chain reaction analysis and allele-specific oligonucleotide hybridization. Am. J. Pathol. 1993;143:545-54.

397. Ancrile B, Lim KH, Counter CM. Oncogenic Ras-induced secretion of IL6 is required for tumorigenesis. Genes Dev. 2007;21:1714-9.

398. Li GH, Arora PD, Chen Y, et al. Multifunctional roles of gelsolin in health and diseases. Med. Res. Rev. 2012;32:999-1025.

399. Suhler E, Lin W, Yin HL, et al. Decreased plasma gelsolin concentrations in acute liver failure, myocardial infarction, septic shock, and myonecrosis. Crit. Care Med. 1997;25:594-8.

400. Osborn TM, Verdrengh M, Stossel TP, et al. Decreased levels of the gelsolin plasma isoform in patients with rheumatoid arthritis. Arthritis Res. Ther. 2008;10:R117.

401. Stossel TP. From signal to pseudopod. How cells control cytoplasmic actin assembly. J. Biol. Chem. 1989;264:18261-4.

402. DiNubile MJ. Plasma gelsolin as a biomarker of inflammation. Arthritis Res. Ther. 2008;10:124.

403. Peddada N, Sagar A, Ashish, et al. Plasma gelsolin: a general prognostic marker of health. Med. Hypotheses 2012;78:203-10.

404. Chen X, Kong X, Zhang Z, et al. Alpha-2-macroglobulin as a radioprotective agent: a review. Chin. J. Cancer Res. 2014;26:611-21.

405. Lin Z, Lo A, Simeone DM, et al. An N-glycosylation Analysis of Human Alpha-2-Macroglobulin Using an Integrated Approach. J Proteomics Bioinform 2012;5:127-134.

406. Rehman AA, Ahsan H, Khan FH. alpha-2-Macroglobulin: a physiological guardian. J. Cell. Physiol. 2013;228:1665-75.

407. Lin Z, Yin H, Lo A, et al. Label-free relative quantification of alpha-2-macroglobulin site-specific core-fucosylation in pancreatic cancer by LC-MS/MS. Electrophoresis 2014;35:2108-15.

408. Hanash SM, Pitteri SJ, Faca VM. Mining the plasma proteome for cancer biomarkers. Nature 2008;452:571-9.

409. Choi E, Hill MM. Targeted High-Throughput Glycoproteomics for Glyco-Biomarker Discovery. In: Leung H-C, editor. Integrative Proteomics. Shanghai, China: InTech; 2012. p. 159-182.

410. Method of the Year 2012. Nat Methods 2013;10:1.

411. Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods 2012;9:555-66.

412. Liebler DC, Zimmerman LJ. Targeted quantitation of proteins by mass spectrometry. Biochemistry 2013;52:3797-806.

413. Doerr A. Mass spectrometry-based targeted proteomics. Nat Methods 2013;10:23.

414. Alm T, von Feilitzen K, Lundberg E, et al. A chromosome-centric analysis of antibodies directed toward the human proteome using Antibodypedia. J. Proteome Res. 2014;13:1669-76.

415. Kuno A, Kato Y, Matsuda A, et al. Focused differential glycan analysis with the platform antibody-assisted lectin profiling for glycan-related biomarker verification. Mol. Cell. Proteomics 2009;8:99-108.

416. Chen S, LaRoche T, Hamelinck D, et al. Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays. Nat Methods 2007;4:437-44.

417. Li D, Chiu H, Zhang H, et al. Analysis of serum protein glycosylation by a differential lectin immunosorbant assay (dLISA). Clin. Proteomics 2013;10:12.

418. Li D, Chiu H, Chen J, et al. Integrated analyses of proteins and their glycans in a magnetic bead-based multiplex assay format. Clin. Chem. 2013;59:315-24.

419. Wen CL, Chen KY, Chen CT, et al. Development of an AlphaLISA assay to quantify serum core-fucosylated E-cadherin as a metastatic lung adenocarcinoma biomarker. J. Proteomics 2012;75:3963-76.

420. MacLean B, Tomazela DM, Shulman N, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 2010;26:966-8.

421. Escher C, Reiter L, MacLean B, et al. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics 2012;12:1111-21.

422. Percy AJ, Chambers AG, Yang J, et al. Comparison of standard- and nano-flow liquid chromatography platforms for MRM-based quantitation of putative plasma biomarker proteins. Anal. Bioanal. Chem. 2012;404:1089-101.

423. Shibue M, Mant CT, Hodges RS. Effect of anionic ion-pairing reagent concentration (1-60 mM) on reversed-phase liquid chromatography elution behaviour of peptides. J. Chromatogr. A 2005;1080:58-67.

424. Chen Y, Mant CT, Hodges RS. Temperature selectivity effects in reversed-phase liquid chromatography due to conformation differences between helical and non-helical peptides. J. Chromatogr. A 2003;1010:45-61.

425. Snyder LR, Dolan JW. High-performance gradient elution: the practical application of the linear-solvent-strength model: John Wiley & Sons; 2007.

426. Kuzyk MA, Smith D, Yang J, et al. Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Mol. Cell. Proteomics 2009;8:1860-77.

427. Ishihama Y. Proteomic LC-MS systems using nanoscale liquid chromatography with tandem mass spectrometry. J. Chromatogr. A 2005;1067:73-83.

428. Vollmer M, van de Goor T. HPLC-Chip/MS technology in proteomic profiling. Methods Mol. Biol. 2009;544:3-15.

429. Gonzalez Fernandez-Nino SM, Smith-Moritz AM, Chan LJ, et al. Standard flow liquid chromatography for shotgun proteomics in bioenergy research. Front Bioeng Biotechnol 2015;3:44.

430. Krokhin OV, Craig R, Spicer V, et al. An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. Mol. Cell. Proteomics 2004;3:908-19.

431. Liu Y, Huttenhain R, Surinova S, et al. Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. Proteomics 2013;13:1247-56.

432. Abbatiello SE, Mani DR, Schilling B, et al. Design, implementation and multisite evaluation of a system suitability protocol for the quantitative assessment of instrument performance in liquid chromatography-multiple reaction monitoring-MS (LC-MRM-MS). Mol. Cell. Proteomics 2013;12:2623-39.

433. Sajic T, Liu Y, Aebersold R. Using data-independent, high-resolution mass spectrometry in protein biomarker research: Perspectives and clinical applications. Proteomics Clin. Appl. 2015;9:307-21.

434. Gillet LC, Navarro P, Tate S, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell. Proteomics 2012;11:O111 016717.

435. Wilhelm M, Schlegl J, Hahne H, et al. Mass-spectrometry-based draft of the human proteome. Nature 2014;509:582-7.

436. Kim MS, Pinto SM, Getnet D, et al. A draft map of the human proteome. Nature 2014;509:575-81.

437. Keshishian H, Addona T, Burgess M, et al. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. Mol. Cell. Proteomics 2007;6:2212-29.

438. Anderson NL, Anderson NG, Haines LR, et al. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). J. Proteome Res. 2004;3:235-44.

439. Anderson L. Six decades searching for meaning in the proteome. J. Proteomics 2014;107:24-30.

440. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim. Biophys. Acta 1999;1473:4-8.

441. Anderson NL, Ptolemy AS, Rifai N. The riddle of protein diagnostics: future bleak or bright? Clin. Chem. 2013;59:194-7.

442. Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. Acta Paediatr. 2007;96:338-41.

443. Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. Acta Paediatr. 2007;96:644-7.

444. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ 1994;309:188.

445. Obuchowski NA, Lieber ML, Wians FH, Jr. ROC curves in clinical chemistry: uses, misuses, and possible solutions. Clin. Chem. 2004;50:1118-25.

446. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. Intensive Care Med. 2003;29:1043-51.

447. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria (2014). http://www.R-project.org/.

448. Le Cao, K.A., Gonzalez, I., Dejean, S., with contributions from Monget, P., Coquery, J., Yao, FZ., Liquet, B. and Rohart, F. mixOmics: Omics data integration project (2014). R package version 5.0-3. http://CRAN.R-project.org/package=mixOmics.

449. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004;329:168-9.

450. Murdoch DJ, Chow ED. A graphical display of large correlation matrices. The American Statistician 1996;50:178-180.

451. Thrift AP, Shaheen NJ, Gammon MD, et al. Obesity and risk of esophageal adenocarcinoma and Barrett's esophagus: a mendelian randomization study. J. Natl. Cancer Inst. 2014;106.

452. Riches FM, Watts GF, Naoumova RP, et al. Hepatic secretion of very-low-density lipoprotein apolipoprotein B-100 studied with a stable isotope technique in men with visceral obesity. Int. J. Obes. Relat. Metab. Disord. 1998;22:414-23.

453. Domanski D, Percy AJ, Yang J, et al. MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. Proteomics 2012;12:1222-43.

454. van Baal JW, Diks SH, Wanders RJ, et al. Comparison of kinome profiles of Barrett's esophagus with normal squamous esophagus and normal gastric cardia. Cancer Res. 2006;66:11605-12.

455. Peng D, Sheta EA, Powell SM, et al. Alterations in Barrett's-related adenocarcinomas: a proteomic approach. Int. J. Cancer 2008;122:1303-10.

456. Zhao J, Chang AC, Li C, et al. Comparative proteomics analysis of Barrett metaplasia and esophageal adenocarcinoma using two-dimensional liquid mass mapping. Mol. Cell. Proteomics 2007;6:987-99.

457. Hu Y, Desantos-Garcia JL, Mechref Y. Comparative glycomic profiling of isotopically permethylated N-glycans by liquid chromatography/electrospray ionization mass spectrometry. Rapid Commun. Mass Spectrom. 2013;27:865-77.

458. Mischak H, Allmaier G, Apweiler R, et al. Recommendations for biomarker identification and qualification in clinical proteomics. Sci. Transl. Med. 2010;2:46ps42.

459. Feng Z, Prentice R, Srivastava S. Research issues and strategies for genomic and proteomic biomarker discovery and validation: a statistical perspective. Pharmacogenomics 2004;5:709-19.

460. Wu B, Abbott T, Fishman D, et al. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. Bioinformatics 2003;19:1636-43.

461. Martens H, Geladi P. Multivariate calibration. Encyclopedia of statistical sciences.: John Wiley & Sons, Inc; 2004.

462. Eriksson L, Johansson E, Kettaneh-Wold N, et al. Multi- and megavariate data analysis: Principles and applications: Umetrics Academy, Umea; 2001.

463. Li J, Zhang Z, Rosenzweig J, et al. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clin. Chem. 2002;48:1296-304.

464. Chuah K, Lai LM, Goon IY, et al. Ultrasensitive electrochemical detection of prostate-specific antigen (PSA) using gold-coated magnetic nanoparticles as 'dispersible electrodes'. Chem. Commun. (Camb.) 2012;48:3503-5.

465. Du Y, Chen C, Li B, et al. Layer-by-layer electrochemical biosensor with aptamer-appended active polyelectrolyte multilayer for sensitive protein determination. Biosens. Bioelectron. 2010;25:1902-7.

466. Rusling JF. Multiplexed electrochemical protein detection and translation to personalized cancer diagnostics. Anal. Chem. 2013;85:5304-10.

467. Shiddiky MJ, Rahman MA, Shim YB. Hydrazine-catalyzed ultrasensitive detection of DNA and proteins. Anal. Chem. 2007;79:6886-90.

468. Wang J. Electrochemical biosensors: towards point-of-care cancer diagnostics. Biosens. Bioelectron. 2006;21:1887-92.

469. Wang J. Electrochemical glucose biosensors. Chem. Rev. 2008;108:814-25.

470. Gubala V, Harris LF, Ricco AJ, et al. Point of care diagnostics: status and future. Anal. Chem. 2012;84:487-515.

471. Yang H, Li Z, Wei X, et al. Detection and discrimination of alpha-fetoprotein with a label-free electrochemical impedance spectroscopy biosensor array based on lectin functionalized carbon nanotubes. Talanta 2013;111:62-8.

472. Nagaraj VJ, Aithal S, Eaton S, et al. NanoMonitor: a miniature electronic biosensor for glycan biomarker detection. Nanomedicine (Lond) 2010;5:369-78.

473. La Belle JT, Gerlach JQ, Svarovsky S, et al. Label-free impedimetric detection of glycan-lectin interactions. Anal. Chem. 2007;79:6959-64.

474. Bertok T, Klukova L, Sediva A, et al. Ultrasensitive impedimetric lectin biosensors with efficient antifouling properties applied in glycoprofiling of human serum samples. Anal. Chem. 2013;85:7324-32.

475. Thaysen-Andersen M, Mysling S, Hojrup P. Site-specific glycoprofiling of N-linked glycopeptides using MALDI-TOF MS: strong correlation between signal strength and glycoform quantities. Anal. Chem. 2009;81:3933-43.

476. Harvey DJ, Wing DR, Kuster B, et al. Composition of N-linked carbohydrates from ovalbumin and co-purified glycoproteins. J. Am. Soc. Mass Spectrom. 2000;11:564-71.

477. Yamashita K, Tachibana Y, Hitoi A, et al. Sialic acid-containing sugar chains of hen ovalbumin and ovomucoid. Carbohydr. Res. 1984;130:271-88.

478. Yang Y, Barendregt A, Kamerling JP, et al. Analyzing protein micro-heterogeneity in chicken ovalbumin by high-resolution native mass spectrometry exposes qualitatively and semi-quantitatively 59 proteoforms. Anal. Chem. 2013;85:12037-45.

479. Shibuya N, Goldstein IJ, Broekaert WF, et al. The elderberry (Sambucus nigra L.) bark lectin recognizes the Neu5Ac(alpha 2-6)Gal/GalNAc sequence. J. Biol. Chem. 1987;262:1596-601.

480. Taguchi A, Hanash SM. Unleashing the power of proteomics to develop blood-based cancer markers. Clin. Chem. 2013;59:119-26.

481. Anderson NL. The roles of multiple proteomic platforms in a pipeline for new diagnostics. Mol. Cell. Proteomics 2005;4:1441-4.

482. Hanash SM, Baik CS, Kallioniemi O. Emerging molecular biomarkers--blood-based strategies to detect and monitor cancer. Nat. Rev. Clin. Oncol. 2011;8:142-50.

483. Paulovich AG, Whiteaker JR, Hoofnagle AN, et al. The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. Proteomics Clin. Appl. 2008;2:1386-1402.

484. Diamandis EP. Cancer biomarkers: can we turn recent failures into success? J. Natl. Cancer Inst. 2010;102:1462-7.

485. Addona TA, Shi X, Keshishian H, et al. A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. Nat. Biotechnol. 2011;29:635-43.

486. Keshishian H, Burgess MW, Gillette MA, et al. Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury. Mol. Cell. Proteomics 2015; DOI 10.1074/mcp.M114.046813.

487. Whiteaker JR, Zhang H, Zhao L, et al. Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. J. Proteome Res. 2007;6:3962-75.

488. Liu NQ, Braakman RB, Stingl C, et al. Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. J. Mammary Gland Biol. Neoplasia 2012;17:155-64.

489. Whiteaker JR, Lin C, Kennedy J, et al. A targeted proteomics-based pipeline for verification of biomarkers in plasma. Nat. Biotechnol. 2011;29:625-34.

490. Ademowo OS, Hernandez B, Collins E, et al. Discovery and confirmation of a protein biomarker panel with potential to predict response to biological therapy in psoriatic arthritis. Ann. Rheum. Dis. 2014; DOI 10.1136/annrheumdis-2014-205417.

491. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 2009;4:44-57.

492. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1-13.

493. Nanjappa V, Thomas JK, Marimuthu A, et al. Plasma Proteome Database as a resource for proteomics research: 2014 update. Nucleic Acids Res. 2014;42:D959-65.

494. Pio R, Corrales L, Lambris JD. The role of complement in tumor growth. Adv. Exp. Med. Biol. 2014;772:229-62.

495. Hardikar S, Onstad L, Song X, et al. Inflammation and oxidative stress markers and esophageal adenocarcinoma incidence in a Barrett's esophagus cohort. Cancer Epidemiol. Biomarkers Prev. 2014;23:2393-403.

496. Picardo SL, Maher SG, O'Sullivan JN, et al. Barrett's to oesophageal cancer sequence: a model of inflammatory-driven upper gastrointestinal cancer. Dig. Surg. 2012;29:251-60.

497. Kaakoush NO, Castano-Rodriguez N, Man SM, et al. Is Campylobacter to esophageal adenocarcinoma as Helicobacter is to gastric adenocarcinoma? Trends Microbiol. 2015; DOI 10.1016/j.tim.2015.03.009.

498. Walport MJ. Complement. First of two parts. N. Engl. J. Med. 2001;344:1058-66.

499. Zipfel PF, Skerka C. Complement regulators and inhibitory proteins. Nat. Rev. Immunol. 2009;9:729-40.

500. Kolev M, Towner L, Donev R. Complement in cancer and cancer immunotherapy. Arch. Immunol. Ther. Exp. (Warsz.) 2011;59:407-19.

501. Narayanasamy A, Ahn JM, Sung HJ, et al. Fucosylated glycoproteomic approach to identify a complement component 9 associated with squamous cell lung cancer (SQLC). J. Proteomics 2011;74:2948-58.

502. Perlmutter DH, Colten HR. Molecular immunobiology of complement biosynthesis: a model of single-cell control of effector-inhibitor balance. Annu. Rev. Immunol. 1986;4:231-51.

503. Qiu Y, Patwa TH, Xu L, et al. Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot. J. Proteome Res. 2008;7:1693-703.

504. Wu J, Xie X, Nie S, et al. Altered expression of sialylated glycoproteins in ovarian cancer sera using lectin-based ELISA assay and quantitative glycoproteomics analysis. J. Proteome Res. 2013;12:3342-52.

505. Nehra D, Howell P, Williams CP, et al. Toxic bile acids in gastro-oesophageal reflux disease: influence of gastric acidity. Gut 1999;44:598-602.

506. Bernstein H, Bernstein C, Payne CM, et al. Bile acids as carcinogens in human gastrointestinal cancers. Mutat. Res. 2005;589:47-65.

507. Ajouz H, Mukherji D, Shamseddine A. Secondary bile acids: an underrecognized cause of colon cancer. World J. Surg. Oncol. 2014;12:164.

508. Payne CM, Bernstein C, Dvorak K, et al. Hydrophobic bile acids, genomic instability, Darwinian selection, and colon carcinogenesis. Clin. Exp. Gastroenterol. 2008;1:19-47.

509. Dvorak K, Payne CM, Chavarria M, et al. Bile acids in combination with low pH induce oxidative stress and oxidative DNA damage: relevance to the pathogenesis of Barrett's oesophagus. Gut 2007;56:763-71.

510. Bjorkhem I, Eggertsen G. Genes involved in initial steps of bile acid synthesis. Curr. Opin. Lipidol. 2001;12:97-103.

511. Nagengast FM, Grubben MJ, van Munster IP. Role of bile acids in colorectal carcinogenesis. Eur. J. Cancer 1995;31A:1067-70.

512. Hill MJ. Bile flow and colon cancer. Mutat. Res. 1990;238:313-20.

513. Hofmann AF, Cravetto C, Molino G, et al. Simulation of the metabolism and enterohepatic circulation of endogenous deoxycholic acid in humans using a physiologic pharmacokinetic model for bile acid metabolism. Gastroenterology 1987;93:693-709.

514. Hofmann AF. The continuing importance of bile acids in liver and intestinal disease. Arch. Intern. Med. 1999;159:2647-58.

515. Jenkins DJ, Wolever TM, Rao AV, et al. Effect on blood lipids of very high intakes of fiber in diets low in saturated fat and cholesterol. N. Engl. J. Med. 1993;329:21-6.

516. Chen KH, Mukaisho K, Sugihara H, et al. High animal-fat intake changes the bile-acid composition of bile juice and enhances the development of Barrett's esophagus and esophageal adenocarcinoma in a rat duodenal-contents reflux model. Cancer Sci. 2007;98:1683-8.

517. Rosenberg IH. Influence of intestinal bacteria on bile acid metabolism and fat absorption. Contributions from studies of blind-loop syndrome. Am. J. Clin. Nutr. 1969;22:284-91.

518. O'Keefe SJ, Li JV, Lahti L, et al. Fat, fibre and cancer risk in African Americans and rural Africans. Nat Commun 2015;6:6342.

519. Daniel H, Moghaddas Gholami A, Berry D, et al. High-fat diet alters gut microbiota physiology in mice. ISME J 2014;8:295-308.

520. Sears CL, Garrett WS. Microbes, microbiota, and colon cancer. Cell Host Microbe 2014;15:317-28.

521. Ha CW, Lam YY, Holmes AJ. Mechanistic links between gut microbial community dynamics, microbial functions and metabolic health. World J. Gastroenterol. 2014;20:16498-517.

522. Verbeek RE, Leenders M, Ten Kate FJ, et al. Surveillance of Barrett's esophagus and mortality from esophageal adenocarcinoma: a population-based cohort study. Am. J. Gastroenterol. 2014;109:1215-22.

523. Spechler SJ. Barrett esophagus and risk of esophageal cancer: a clinical review. JAMA 2013;310:627-36.

524. Kadri SR, Lao-Sirieix P, O'Donovan M, et al. Acceptability and accuracy of a non-endoscopic screening test for Barrett's oesophagus in primary care: cohort study. BMJ 2010;341:c4372.

525. Suter MJ, Gora MJ, Lauwers GY, et al. Esophageal-guided biopsy with volumetric laser endomicroscopy and laser cautery marking: a pilot clinical study. Gastrointest. Endosc. 2014;79:886-96.

526. Falk GW. Probe-based confocal endomicroscopy in Barrett's esophagus: the real deal or another tease? Gastrointest. Endosc. 2011;74:473-6.

527. Leggett CL, Gorospe EC. Application of confocal laser endomicroscopy in the diagnosis and management of Barrett's esophagus. Ann Gastroenterol 2014;27:193-199.

528. Anders M, Lucks Y, El-Masry MA, et al. Subsquamous extension of intestinal metaplasia is detected in 98% of cases of neoplastic Barrett's esophagus. Clin. Gastroenterol. Hepatol. 2014;12:405-10.

529. Lao-Sirieix P, Rous B, O'Donovan M, et al. Non-endoscopic immunocytological screening test for Barrett's oesophagus. Gut 2007;56:1033-4.

530. Weaver JM, Ross-Innes CS, Fitzgerald RC. The '-omics' revolution and oesophageal adenocarcinoma. Nat. Rev. Gastroenterol. Hepatol. 2014;11:19-27.

531. Croswell JM, Kramer BS, Kreimer AR, et al. Cumulative incidence of false-positive results in repeated, multimodal cancer screening. Ann. Fam. Med. 2009;7:212-22.

532. Han KN, Li CA, Seong GH. Microfluidic chips for immunoassays. Annu. Rev. Anal. Chem. (Palo Alto Calif.) 2013;6:119-41.

533. Turner AP. Biosensors: sense and sensibility. Chem. Soc. Rev. 2013;42:3184-96.

534. Luo X, Davis JJ. Electrical biosensors and the label free detection of protein disease biomarkers. Chem. Soc. Rev. 2013;42:5944-62.

535. Hirabayashi J. Concept, strategy and realization of lectin-based glycan profiling. J. Biochem. 2008;144:139-47.

Chapter 9.

APPENDIX

Chapter 9. Appendices

9.1 Appendix I: Review article entitled "Early diagnostic biomarkers for esophageal adenocarcinoma—The current state of play"

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Cancer Epidemiology, Biomarkers & Prevention

Early Diagnostic Biomarkers for Esophageal Adenocarcinoma—The Current State of Play

Alok Kishorkumar Shah, Nicholas A. Saunders, Andrew P. Barbour, et al.

Cancer Epidemiol Biomarkers Prev 2013;22:1185-1209. Published OnlineFirst April 10, 2013.

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Cancer Epidemiology, Biomarkers & Prevention

Review

Early Diagnostic Biomarkers for Esophageal Adenocarcinoma—The Current State of Play

Alok Kishorkumar Shah[1], Nicholas A. Saunders[1], Andrew P. Barbour[2], and Michelle M. Hill[1]

Abstract

Esophageal adenocarcinoma (EAC) is one of the two most common types of esophageal cancer with alarming increase in incidence and very poor prognosis. Aiming to detect EAC early, currently high-risk patients are monitored using an endoscopic-biopsy approach. However, this approach is prone to sampling error and interobserver variability. Diagnostic tissue biomarkers related to genomic and cell-cycle abnormalities have shown promising results, although with current technology these tests are difficult to implement in the screening of high-risk patients for early neoplastic changes. Differential miRNA profiles and aberrant protein glycosylation in tissue samples have been reported to improve performance of existing tissue-based diagnostic biomarkers. In contrast to tissue biomarkers, circulating biomarkers are more amenable to population-screening strategies, due to the ease and low cost of testing. Studies have already shown altered circulating glycans and DNA methylation in BE/EAC, whereas disease-associated changes in circulating miRNA remain to be determined. Future research should focus on identification and validation of these circulating biomarkers in large-scale trials to develop *in vitro* diagnostic tools to screen population at risk for EAC development. *Cancer Epidemiol Biomarkers Prev; 22(7); 1185–209. ©2013 AACR.*

Introduction

After heart disease, cancer is the second leading cause of death globally. Four major cancer sites account for half of the cancer-related mortalities: lung, colorectal, prostate in men, and breast in women. In past 2 decades, a steady decrease in deaths of these 4 major site malignancies led to an overall decrease in cancer-related death rates in men and women (1). In contrast, the incidence of esophageal adenocarcinoma (EAC) is increasing faster than any other cancer type. EAC together with esophageal squamous cell carcinoma (ESCC) is the eighth most-common cancer by prevalence and sixth most-common cause of cancer-related death globally (2). In 1970s, the incidence of EAC represented less than 5% of total esophageal cancer, and a majority of esophageal cancer cases diagnosed were ESCC. Over a period of 3 decades, EAC incidences have been increasing continuously, especially in western countries among Caucasians. Now almost half of the esophageal malignancy cases diagnosed are EAC (3, 4). EAC and ESCC show marked differences in their geographic spread. EAC is more common in developed countries such as the United Kingdom (8 in 100,000 individuals; ref. 5), Australia, and the United States. Within Europe, southern Europe has the highest EAC incidence (5). On the other side, ESCC is the most common type of esophageal cancer among developing Asian countries (6). Racial disparity also occurs between the 2 types of esophageal cancer. ESCC is more prevalent among Blacks, whereas EAC is at least twice as common in Whites as compared with other ethnic groups (7, 8). Once diagnosed, Black patients showed poorer overall survival than Whites (9, 10). Taken together, strong genetic and environmental factors relating to ethnicity and geographic distribution seem to be playing critical roles in the incidence of esophageal cancer. Studies also suggest possible links between socioeconomic status and the prevalence of esophageal cancer phenotype (6).

Risk Factors

In the majority of cases, EAC is diagnosed at a late stage, leading to a poor 5-year survival of less than 15% (11). Hence, recent research for EAC has focused on understanding risk factors and the identification of early diagnostic biomarkers.

Esophageal cancer is unlikely to develop in individuals younger than 40 years of age; however, after that the incidence increases significantly with each decade of life (9). Changing lifestyle and food habits are primarily responsible for the dramatic epidemiologic changes in EAC as described in recent reviews (11–13). Known EAC risk factors include accumulation of visceral fat in the abdomen (14), male gender, high intake of dietary fat and cholesterol with low intake of fruits and vegetables (15), tobacco smoking (16), reduction in *Helicobacter pylori*

Authors' Affiliations: [1]The University of Queensland Diamantina Institute; and [2]School of Medicine, The University of Queensland, Woolloongabba, Queensland, Australia

Corresponding Author: Michelle M. Hill, The University of Queensland Diamantina Institute, Level 5, Translational Research Institute, 37 Kent Street, Woolloongabba, QLD 4102. Phone: 61-7-3443-7049; Fax: 61-7-3443-6966; E-mail: m.hill2@uq.edu.au

doi: 10.1158/1055-9965.EPI-12-1415

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

infections (17), and Barrett's esophagus (BE), a metaplastic change to the esophageal lining. Individuals with Barrett's esophagus carry almost 30 to 125 times more risk for EAC development, and 0.5% to 1% of patients with Barrett's esophagus are estimated to develop EAC each year (18). Barrett's esophagus is characterized by replacement of normal stratified squamous epithelium with metaplastic columnar epithelium and is considered to be a successful adaptation of the distal esophagus in response to chronic gastroesophageal reflux disorder (GERD; ref. 19).

GERD is a very common condition in the western population with around 20% reporting weekly symptoms of heartburn and acid regurgitation (20). Refluxate-containing bile acid, along with gastric acid, is considered to be more harmful, leading to inflammation, ulceration, Barrett's esophagus, and ultimately EAC. Development of Barrett's esophagus is a slow process and distinctive mucus-secreting goblet cell formation can take 5 to 10 years (21, 22). Typically, EAC develops through metaplasia–dysplasia–carcinoma sequence involving genetic and epigenetic modifications, leading to uncontrolled cell proliferation, and is characterized by the presence of intestinal metaplasia with low-grade (LGD) to high-grade dysplasia (HGD), which eventually may progress to invasive carcinoma (20).

Current Diagnosis Scenario

To detect pathologic changes leading to EAC development before onset of disease, current clinical practice involves endoscopic screening of patients with high-risk GERD and to characterize the degree of dysplasia in biopsy samples collected during endoscopy (23, 24). Enrollment of patients into an endoscopic screening program may be facilitated by a patient questionnaire of self-evaluated symptoms/complications (25, 26). Once enrolled into the screening program, a patient undergoes endoscopy-biopsy every 3 months to 2 years depending on the degree of dysplasia, during which 4 quadrant biopsy samples are taken every 1 to 2 cm and evaluated for histologic changes by expert pathologists (23, 24). As a significant number of patients histologically diagnosed with HGD develop EAC, endoscopic mucosal ablation or esophageal resection (esophagectomy) are options to stop further disease progression in those high-risk patients (27, 28). Significantly improved survival is observed in patients diagnosed at an early stage during surveillance endoscopy program as compared with symptomatically diagnosed EAC (29–32).

Although current screening methodology shows promise, outcome of endoscopy-biopsy in many cases is nonreproducible due to interobserver variability and sampling error (28, 33). Furthermore, histologic dysplastic changes may be patchy and present heterogeneously in the tissue sample. This makes the diagnosis challenging, especially in the early stages of transition to LGD (28, 34). In up to 40% of patients, invasive cancer has been found in resected tissue despite negative endoscopic examination for the malignancy (35). Moreover, false-positive results also occur, meaning despite intramucosal carcinoma in a biopsy, the subsequently resected tissue has no signs of carcinoma (28). These evidence suggest dysplasia grading is an imperfect measure of cancer risk.

Despite extensive screening with currently available techniques, more than 80% of EACs are diagnosed without any prior diagnosis of Barrett's esophagus or GERD (36, 37). According to an estimate, more than 80% of Barrett's esophagus cases are undiagnosed and therefore are not getting the benefit of the screening program (38). On the other hand, a large proportion of patients undergoing routine biopsy screening do not progress to EAC (13). These suggest inability of current methodologies in screening population to detect high-risk patients and to distinguish between disease progressors from nonprogressors. In addition, the screening procedure is not very cost-effective (39). To overcome these challenges, adjunct use of biomarker has been proposed to stratify the risk associated with EAC development.

Biomarkers in EAC

According to United States' NIH, a biomarker is "a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (40)."

In transit from intestinal metaplasia to LGD to HGD to EAC, cells acquire abilities to become self-sufficient for growth, evade apoptosis, proliferate uncontrollably, promote angiogenesis, invade underlined epithelium, and start to metastasize. These changes are accompanied with histologic changes in tissue architecture, genomic instability, development of tumor microenvironment, modulation of immune response, and are therefore reflected in body fluids (serum/plasma/mucus/urine) or tissue samples and differentiate in terms of their genome/proteome/metabolome profile (41). Thus, a biomarker can be from any of these sources and reflect underlying pathologic or homeostatic changes. Table 1 summarizes different classes of biomarkers proposed for BE/EAC.

National Cancer Institute Early Detection Research Network (EDRN) guidelines outline biomarker discovery and development to a 5-phase process summarized below (42) and depicted in Fig. 1.

Phase I—Preclinical exploratory study: it compares normal versus cancer samples (body fluids/tissue) using technologies such as genomics, microarray expression, proteomics, immunohistochemistry, or immunoblotting to detect significant changes in proteins/genes/metabolites between the groups.

Phase II—Clinical assay development and validation: it is aimed at developing a clinical assay using a minimally invasive sample collection method. The assay is meant to be robust, reproducible, and suitable for stored clinical samples to be used in later phases of development. At the end of this phase, one should

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Table 1. Comprehensive summary of different classes of BE/EAC biomarkers

Biomarker class	Ref.
Tissue biomarkers	
Genomic abnormalities (ploidy and LOH)	(47–51)
DNA methylation	Refer to Table 2
SNPs/expression array studies	Refer to Table 3
Inflammatory markers	
COX-2	(69, 72–77)
NF-κB	(78–81)
Cytokines	(67, 79, 81–86)
MMPs	(87–93)
Cell-cycle abnormalities	(94, 95, 101)
miRNA	Refer to Table 4
Glycosylation changes	(121, 123–125)
Circulatory biomarkers	
DNA methylation changes	(130–132)
Glycan alterations	(135–138)
Metabolic profiling	(142–145)

expect high specificity and sensitivity for the assay. However, it remains to be determined how early the biomarker can predict the disease.

Phase III—Retrospective longitudinal repository studies: the assay is applied on prospectively collected stored samples to determine the ability of the biomarker to detect the disease before clinical presentation. If so, then criteria for positive screening is determined for future use.

Phase IV—Prospective screening: the test is prospectively applied to real population to detect the extent and characteristic of disease detected by the biomarker. This phase gives positive predictive value for the test and gives an idea about feasibility for last phase of control trials.

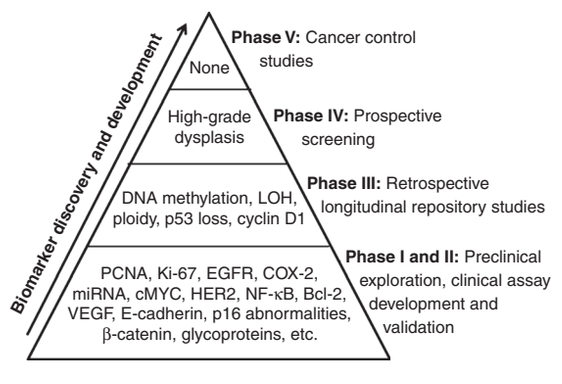


Figure 1. Summary of current BE/EAC Biomarkers with respect to EDRN clinical phase of development.

Phase V—Cancer control studies: it comprises large-scale clinical trial to determine the impact of new screening process on the disease burden in the community.

With respect to EAC, none of the biomarkers, including high-grade dysplasia, have been evaluated in phase V, whereas very few are evaluated in phase III and IV. Figure 1 summarizes proposed EAC biomarkers and how well they are characterized in the process of biomarker discovery. The following sections will discuss some of the classes of BE/EAC biomarkers.

Genomic Instability

Many groups have studied genomic instability induced by aneuploidy, tetraploidy, DNA methylation, allelic loss and shown some predictive power for these changes. A role for hypermethylation in the promoter regions of tumor-suppressor genes during the development of EAC has also been well established. Table 2 summarizes DNA methylation changes associated with metaplasia–dysplasia–carcinoma development. In the majority of patients, methylation changes are acquired very early during EAC development, hence these alterations could be used as an early diagnostic biomarker. Apart from discriminating patients at different stages of EAC development, DNA methylation signatures may be useful as predictors for progression from Barrett's esophagus to EAC (43, 44) and for response to chemotherapy and survival in patients with EAC (45, 46).

Although the individual genomic abnormality has the potential to diagnose disease at different stages, best results are obtained when they are used in combination (47–49). LOH at chromosome 9p and 17p locus are considered to be early events during Barrett's esophagus pathogenesis (50). If present with other chromosomal alterations such as aneuploidy and tetraploidy, it increases the 10-year risk for development of EAC from 12% to approximately 80% (51). However, with the current flow cytometry technology, it is technically very challenging for clinical laboratories to assess these genomic biomarkers in the samples, which limits widespread use of these biomarkers in the clinic.

Alternatively, genomic alterations can be detected at the protein level using immunohistochemistry. One of the most common and earliest genomic abnormality occurs at chromosome 17p, which codes for tumor-suppressor p53 protein. Loss of p53 protein expression in tissue samples correlates with disease progression (52). However, as p53 expression only reflects alterations at one particular gene, it has lower predictive value as compared with techniques monitoring multiple genomic abnormalities. Furthermore, sensitivity drops as mutations or deletions at genomic level may not necessarily be detected at the protein level (53).

In line with the genomic abnormalities described earlier, single-nucleotide polymorphism (SNP)–based genotyping can also stratify cancer risk in patients with Barrett's esophagus. As summarized in Table 3, in the past

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Table 2. Summary of hypermethylated genes during BE/EAC development

Gene	Location	Function	Method	Number (%) of samples showing hypermethylation or study findings					Ref.
				Normal	BE	LGD	HGD	EAC	
p16 (or *CDKN2A* or *INK4A*)	*9p21*	Cyclin-dependent kinase inhibitor	Methylation-specific PCR	5/9 (56%)	14/18 (77%)	—	—	18/21 (85%)	(146)
			Methylation-sensitive single-strand conformation analysis	0/10 (0%)	4/12 (33%)	3/11 (27%)	3/10 (30%)	18/22 (82%)	(147)
			Methylation-specific PCR	0/17 (0%)	14/47 (30%)	9/27 (32%)	10/18 (56%)	22/41 (54%)	(148)
			Methylation-specific PCR	2/64 (3%)	14/93 (15%)	—	—	34/76 (45%)	(149)
			Methylation-specific PCR	—	3/10 (30%)	—	—	5/11 (45%)	(150)
			Methylation-specific PCR	—	27/41(66%)	21/45 (47%)	17/21 (81%)	65/107 (61%)	(151)
			Methylation-specific PCR	0%	1/15 (7%)	4/20 (20%)	12/20 (60%)	8/15 (53%)	(152)
			Methylation-specific PCR	Separately determined exon 1 and 2 methylation. Five of 16 (31%) exon-1, 8/16 (50%) exon 2 in EAC patient samples showed hypermethylation. Exon 2 methylation correlates with stage of the tumor ($P = 0.01$)					(153)
O^6-*Methylguanine-DNA Methyltransferase* (or *MGMT*)	*10q26*	DNA repair	Methylight technique	2/10 (20%)	8/13 (62%)	—	—	84/132 (64%)	(154)
			Methylation-specific PCR	6/29 (21%)	24/27 (89%)	13/13 (100%)		37/47 (79%)	(155)
APC	*5q21-q22*	Wnt/β-catenin signaling	Methylation-specific PCR	0/17 (0%)	24/48 (50%)	14/28 (50%)	14/18 (78%)	20/32 (63%)	(148)
			Methylation-sensitive single-strand conformation analysis and methylation-sensitive dot blot assay	0/16 (0%)	11/11 (100%)	—	—	20/21 (95%)	(156)
				Eight of 14 histologically normal gastric mucosa adjacent to EAC showed significantly different methylation of *APC* promoter.					(157)
GSTM2	*1p13.3*	Antioxidants and protection against DNA damage	Bisulfite pyrosequencing (sample size: EAC-100, BE-11, dysplasia-11, normal esophageal/gastric mucosa-37)	<10%	∼50%	∼55%		69%	(158)
GSTM3				<10%	∼13%	∼37%		15%	(158)
GPX7	*1p32*			<10%	∼18%	∼80%		67%	(158)
GPX3	*5q23*			<10%	∼90%	∼88%		62%	(158)
			Methylation-specific PCR	2/12 (17%)	13/21(62%)	9/11 (82%)		30/34 (88%)	(159)
TIMP-3	*22q12.3*	MMP inhibitor	Methylight technique	1/8 (13%)	6/12 (50%)	—	—	9/13 (69%)	(160)
Death-associated protein kinase (*DAPK*)	*DAPK1: 9q21.33 DAPK2: 15q22.31 DAPK3: 19p13.3*	Tumor-suppressor and mediator of apoptosis	Methylation-specific PCR	4/20 (20%)	14/28 (50%)	11/21 (53%)		21/35 (60%)	(161)

(Continued on the following page)

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Table 2. Summary of hypermethylated genes during BE/EAC development (Cont'd)

Gene	Location	Function	Method	Number (%) of samples showing hypermethylation or study findings					Ref.
				Normal	BE	LGD	HGD	EAC	
Tachykinin-1 (*TAC1*)	*7q21-22*	Smooth muscle contractility, epithelial ion transport, vascular permeability and immune function	Methylation-specific PCR	5/67 (7.5%)	38/60 (63.3%)	12/19 (63.2%)	11/21 (52.4%)	41/67 (61.2%)	(162)
Reprimo	*2q23*	Regulates p53-mediated cell-cycle arrest in G_2-phase	Methylation-specific PCR	0/19 (0%)	9/25 (36%)	—	7/11 (64%)	47/75 (63%)	(163)
E-Cadherin	*16q22.1*	Ca^{+2}-dependent intercellular adhesion and maintains normal tissue architecture	Methylation-specific PCR	0/4 (0%)	—	—	—	26/31 (84%)	(164)
SOCS-3	*17q25.3*	Inhibits cytokine signaling	Methylation-specific PCR	0%	4/30 (13%)	6/27 (22%)	20/29 (69%)	14/19 (74%)	(165)
SOCS-1	*16p13.13*			0%	0/30 (0%)	1/27 (4%)	6/29 (21%)	8/19 (42%)	
Secreted frizzled-related proteins (*SFRP*)									
SFRP1	*8p11.21*	Wnt antagonist	Methylation-specific PCR	7/28 (25%)	30/37 (81%)	—	—	37/40 (93%)	(166)
SFRP2	*4q31.3*			18/28 (64%)	33/37 (89%)	—	—	33/40 (83%)	
SFRP1	*8p11.21*		Methylation-sensitive single-strand conformation analysis and methylation-sensitive dot blot assay	1/12 (8%)	6/6 (100%)	—	—	23/24 (96%)	(156)
SFRP2	*4q31.3*			11/15 (73%)	6/6 (100%)	—	—	19/25 (76%)	
SFRP4	*7p14.1*		Methylation-specific PCR	9/28 (32%)	29/37 (78%)	—	—	29/40 (73%)	(166)
SFRP5	*10q24.1*			6/28 (21%)	27/37 (73%)	—	—	34/40 (85%)	
Plakophilin-1 (*PKP1*)	*1q32*	Cell adhesion and intracellular signaling	Methylation-specific PCR	5/55 (9.1%)	5/39 (12.8%)	—	1/4 (25%)	20/60 (33.3)	(167)
GATA-4	*8p23.1-p22*	Transcription factor and regulate cell differentiation	Methylation-specific PCR	0/17 (0%)	—	—	—	31/44 (71%)	(168)
GATA-5	*20q13.33*			0/17 (0%)	—	—	—	24/44 (55%)	
CDH13 (or *H-cadherin* or *T-cadherin*)	*16q24*	Cell adhesion	Methylation-specific PCR	0/66 (0%)	42/60 (70%)	15/19 (78.9%)	16/21 (76.2)	51/67 (76.1%)	(169)

(Continued on the following page)

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Table 2. Summary of hypermethylated genes during BE/EAC development (Cont'd)

				Number (%) of samples showing hypermethylation or study findings					
Gene	**Location**	**Function**	**Method**	**Normal**	**BE**	**LGD**	**HGD**	**EAC**	**Ref.**
NELL-1 (*nel-like 1*)	*11p15*	Tumor suppressor	Methylation-specific PCR	0/66 (0%)	28/60 (46.7%)	8/19 (42.1%)	13/21 (61.9%)	32/67 (47.8%)	(170)
Eyes Absent 4	*6q23*	Apoptosis modulator	Methylation-specific PCR	2/58 (3%)	27/35 (77%)	—	—	33/40 (83%)	(171)
A-kinase anchoring protein 12 (or *Gravin* or *AKAP12*)	*6q24-25.2*	Cell-signaling, adhesion, mitogenesis, and differentiation	Methylation-specific PCR	0/66 (0%)	29/60 (48.3%)	10/19 (52.6%)	11/21 (52.4%)	35/67 (52.2)	(172)
Vimentin	*10p13*	Cytoskeleton protein	Methylation-specific PCR	0/9 (0%)	10/11 (91%)	—	5/5 (100%)	21/26 (81%)	(173)
RUNX3	*1p36*	Transcription factor	Methylation-specific PCR	1/63 (2%)	23/93 (25%)	—	—	37/77 (48%)	(149)
HPP1	*19pter-p13.1*	Tumor suppressor		2/64 (3%)	41/93 (44%)	—	—	55/77 (71%)	
3-OST-2	*16p12*	Sulfotransferase enzyme		1/57 (2%)	47/60 (78%)	—	—	28/73 (38%)	
Wnt inhibitory factor-1 (*WIF-1*)	*12q14.3*	Wnt antagonist	Methylation-specific PCR	81% of patients with Barrett's esophagus suffering from EAC showed hypermethylated WIF-1 as compared with 20% of patients with Barrett's esophagus without EAC					(174)
CHFR (*checkpoint with forkhead associated and ring finger*)	*12q24*	Mitosis check point protein	Bisulfite pyrosequencing	EAC samples 31% (18/58) showed significantly higher *CHFR* promoter methylation as compared with normal samples ($P = 0.01$).					(175)
Metallothionein 3 (or *MT3*)	*16q13*	Metal homeostasis and protection against DNA damage	Bisulfite pyrosequencing (sample size: normal-33, BE-5, EAC-78)	Identified 2 regions (R2 and R3) of CpG nucleotides, which showed significantly higher methylation in EAC as compared with normal epithelium (FDR < 0.001). Increased DNA methylation of MT3 promoter R2 correlates with advanced tumor stage ($P = 0.005$) and lymph node metastasis ($P = 0.03$). DNA methylation of MT3 promoter R3 correlates with tumor staging ($P = 0.03$) but not with lymph node status ($P = 0.4$).					(176)

Methylation marker panel

Sample size	**Method**	**Findings**	**Ref.**
EAC-35 undergoing chemoradiotherapy	Methylation-specific PCR	Combined mean of promoter methylation of *p16*, *Reprimo*, *p57*, *p73*, *RUNX-3*, *CHFR*, *MGMT*, *TIMP-3*, and *HPP1* was lower in patients who responded to chemoradiotherapy (13/35) as compared with patients who did not respond (22/35; $P = 0.003$).	(46)
BE-62 (28 patients with Barrett's esophagus progressed to EAC and remaining 34 patients with Barrett's esophagus were nonprogressors)	Methylation-specific PCR	Three-tiered stratification model was developed using methylation index (*p16*, *HPP1*, and *RUNX3*), Barrett's esophagus length and pathology. Combined model based on 2- (ROC: 0.8386) and 4-year (ROC: 0.7910) prediction was able to categorize patients with Barrett's esophagus into low-risk, intermediate-risk, and high-risk groups for EAC development.	(44)

(Continued on the following page)

Table 2. Summary of hypermethylated genes during BE/EAC development (Cont'd)

Methylation marker panel

Sample size	Method	Findings	Ref.
BE-195 (145 patients with Barrett's esophagus progressed to EAC and remaining 50 patients with Barrett's esophagus were nonprogressors)	Methylation-specific PCR	*HPP1* ($P = 0.0025$), *p16* ($P = 0.0066$), and *RUNX3* ($P = 0.0002$) were significantly hypermethylated in progressors as compared with nonprogressors. In combination, panel of 8 methylation markers (*p16*, *HPP1*, *RUNX3*, *CDH13*, *TAC1*, *NELL1*, *AKAP12*, and *SST*) showed sensitivities of 0.443 and 0.629 at specificity of 0.9 and 0.8 for EAC progression in patients with Barrett's esophagus using combined model designed on the basis of 2 and 4 years of follow-up.	(43)
EAC-41 (adjacent normal samples as control)	Methylation-specific PCR	Patients having more than 50% of their genes methylated (*APC*, *E-cadherin*, *MGMT*, *ER*, *p16*, *DAP-kinase*, and *TIMP3*) showed significantly poor 2-year survival ($P = 0.04$) and 2-year relapse-free survival ($P = 0.03$) as compared with the patients having less than 50% methylation.	(45)
BE-18, EAC-38 (multiple biopsies were taken and classified into normal, Barrett's esophagus, HGD, and EAC)	Bisulfite-modified DNA with PCR	The methylation frequencies of 9 genes (*APC*, *CDKN2A*, *ID4*, *MGMT*, *RBP1*, *RUNX3*, *SFRP1*, *TIMP3*, and *TMEFF2*) found to be 95%, 59%, 76%, 57%, 70%, 73%, 95%, 74%, and 83%, respectively, in EAC samples, whereas 95%, 28%, 78%, 48%, 58%, 48%, 93%, 88%, and 75%, respectively, in Barrett's esophagus samples, which was significantly higher as compared with normal squamous epithelium. The methylation frequency for *CDKN2A* and *RUNX3* was significantly higher for EAC as compared with Barrett's esophagus biopsy samples.	(177)
Normal-30, BE-29, HGD-8, EAC-29	Illumina GoldenGate methylation bead array	Overall median methylation at the total 706 numbers of most informative CpG sites gradually increased from normal-BE-HGD/EAC ($P < 0.001$). The authors differentiated between EAC vs. normal, HGD vs. normal, Barrett's esophagus vs. normal, EAC vs. Barrett's esophagus, and HGD vs. Barrett's esophagus based on 422, 225, 195, 17, and 3 numbers of CpG sites, which is showing differential methylation between respective groups.	(178)
Identification phase (BE-22, EAC-24); retrospective validation phase (BE-60, LGD/HGD-36, EAC-90); prospective validation phase (98 patients under surveillance).	Identification phase: Illumina Infinium assay; retrospective/prospective validation phase: pyrosequencing	On the basis of initial identification phase, 7 genes (*SLC22A18*, *ATP2B4*, *PIGR*, *GJA12*, *RIN2*, *RGN*, and *TCEAL7*) showing most prominent methylation changes were selected for validation. Combination of 4 genes (ROC 0.988) *SLC22A18*, *PIGR*, *GJA12*, and *RIN2* showed sensitivity of 94% and specificity of 97%. This panel of 4 genes showing differential methylation, stratified patients into low-, intermediate-, and high-risk groups for EAC development in prospective validation.	(179)
Nondysplastic Barrett's esophagus (not progressed to EAC)-16, Barrett's esophagus mucosa from patients progressed to EAC-12	Methylation-sensitive single-strand conformation analysis and methylation-sensitive dot blot assay	Barrett's esophagus samples collected from patients who progressed to EAC in 12 months time period showed 100%, 91%, and 92% hypermethylation of *APC*, *TIMP-3*, and *TERT*, respectively, as compared with 36%, 23%, and 17% in Barrett's esophagus mucosa collected from patients who did not progress to EAC.	(180)

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Table 3. Summary of gene expression profiling studies for BE/EAC

Sample size	Array description	Outcome	Findings	External validation	Ref.
BE-21 (paired normal esophageal and gastric samples as control)	Serial analysis of gene expression, PCR and immunoblotting	Disease progression	Of note, 534 tags were significantly differentially expressed between normal esophageal squamous epithelium and Barrett's esophagus. The most upregulated genes in Barrett's esophagus as compared with normal epithelium were identified to be trefoil factors, annexin A10 and galectin-4 with each different type of tissue showed an unique cytokeratin expression.	No	(181)
Barrett's esophagus and HGD-11 (matched biopsy samples)	cDNA microarray	Disease progression	Using 2.5-fold cutoff, identified 131 upregulated and 16 downregulated genes in HGD. Twenty-four of 28 most significantly different genes showed similar changes during validation.	Real-time PCR	(182)
EAC-91	Oligo-microarray	Disease progression	A 4-gene panel consists of deoxycytidine kinase, 3′-phosphoadenosine 5′-phosphosulfate synthase 2, sirtuin-2, and tripartite motif-containing 44 predicted 5-year survival.	Immunohistochemistry	(183)
Twenty-three paired Barrett's esophagus and normal epithelium samples	Transcriptional profiling and proteomics	Disease progression	Identified 2,822 genes to be differentially expressed between Barrett's esophagus and normal epithelium. Significantly overexpressed genes during Barrett's esophagus belonged to cytokines and growth factors, constituents of extracellular matrix, basement membrane and tight junctions, proteins involved in prostaglandin and phosphoinositol metabolism, nitric oxide production and bioenergetics. While genes encoding HSP and various kinases were downregulated.	No	(184)
Lymph node metastatic ($n = 55$) and nonmetastatic ($n = 22$) EAC samples	Oligo-microarray	Disease progression	Lymph node–positive samples showed significant downregulation of argininosuccinate synthetase as compared with lymph node nonmetastatic samples ($P = 0.048$).	No	(185)
EAC-6 and gastric cardia cancer-8	aCGH	Disease progression	Identified *HGF* (45%) and *BCAS1* (27%) to be most frequently overexpressed genes respectively at 7q21 and 20q13 locus.	No	(186)
Eleven matched sample sets (healthy-BE-EAC matched-6, normal-BE matched-4 and normal-EAC matched-1)	SNP microarray	Disease progression	60% of Barrett's esophagus and 57% of EAC samples contained at least one of the genomic alterations in the form of deletions, duplications, amplifications, copy number changes, and neutral LOH.	No	(187)

(Continued on the following page)

Table 3. Summary of gene expression profiling studies for BE/EAC (Cont'd)

Sample size	Array description	Outcome	Findings	External validation	Ref.
Normal-39, BE-25, EAC-38, and ESCC-26	cDNA microarray	Disease progression	Clustering showed the separation of samples into 4 distinct groups. Of note, 2,158 clones were differentially expressed between normal and Barrett's esophagus samples, whereas 1,306 between Barrett's esophagus and EAC. BE/EAC samples showed differential expression of hydrolases, lysozyme, fucosidase, transcription factors, mucins, and the trefoil factors.	No	(188)
BE-20, LGD-19, HGD-20 and EAC-42	SNP microarray	Disease progression	Increasing numbers of SNPs and loss of chromosomes with disease progression. Chromosomal disruption was identified in the *FHIT*, *WWOX*, *RUNX1*, *KIF26B*, *MGC48628*, *PDE4D*, *C20orf133*, *GMDS*, *DMD*, and *PARK2* genes in EAC.	No	(189)
EAC-75 specimens from 64 patients, adjacent paired normal tissue from patients with EAC-28	DNA microarray	Disease progression	Identified *AKR1B10*, *CD93*, *CSPG2*, *DKK3*, *LUM*, *MMP1*, *SOX21*, *SPP1*, *SPARC*, and *TWIST1* genes as biomarker based on transcriptomics data. Quantitative real-time PCR identified *SPARC* and *SPP1* genes to be associated with EAC patient survival ($P < 0.024$).	Real-time PCR	(190)
EAC-8, gastric cardia cancer-3	aCGH and cDNA microarray	Disease progression	Transcriptomics data identified 11 genes to be differentially expressed (*ELF3*, *SLC45A3*, *CLDN12*, *CDK6*, *SMURF1*, *ARPC1B*, *ZKSCAN1*, *MCM7*, *COPS6*, *FDFT1*, and *CTSB*). IHC analysis revealed significant overexpression of *CDK6* a cell-cycle regulator in tumor samples.	No	(191)
BE-20	aCGH arrays and high density SNP genotyping	Disease progression	Copy number losses were detected at FRA3B (81%), FRA9A/C (71.4%), FRA5E (52.4%), and FRA 4D (52.4%) sites in early Barrett's esophagus. Validation study confirmed loss of FRA3B and FRA16D in early Barrett's esophagus samples.	Real-time PCR and pyrosequencing	(192)
BE-11, gastroesophageal junction (GEJ) adenocarcinoma-11	aCGH with a whole chromosome 8q contig array	Disease progression	Overexpression of *MYC* and *EXT1*, while downregulation of *MTSS1*, *FAM84B*, and *C8orf17* is significantly associated with GEJ adenocarcinoma.		(193)
BE-14, EAC-5, ESCC-3	cDNA microarray	Disease progression	Identified 160 genes that can differentiate between Barrett's esophagus and esophageal cancer.	No	(194)
Twenty-four paired samples of normal, Barrett's esophagus, and EAC phenotype	cDNA microarray	Disease progression	Of note, 214 differentially regulated genes could differentiate between normal, Barrett's esophagus, and EAC phenotype. Genes involved in epidermal	No	(195)

(Continued on the following page)

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Table 3. Summary of gene expression profiling studies for BE/EAC (Cont'd)

Sample size	Array description	Outcome	Findings	External validation	Ref.
			differentiation are underexpressed in EAC as compared with Barrett's esophagus. Expression ratio of *GATA6* to *SPRR3* can differentiate between 3 phenotypes studied.		
Pooled biospy samples from Barrett's esophagus, esophageal squamous, gastric, and duodenum	Oligo-microarray	Disease progression	Differentiate different tissue clusters based on gene expression profile. Identified 38 genes that are upregulated in Barrett's esophagus tissue cluster, which belong to cell cycle (P1cdc47, PCM-1), cell migration (urokinase-type plasminogen receptor, LUCA-1/HYAL1), growth regulation (TGF-β superfamily protein, amphiregulin, Cyr61), stress responses (calcyclin, ATF3, TR3 orphan receptor), epithelial cell surface antigens (epsilon-BP, ESA, integrin β4, mesothelin CAK-1 antigen precursor), and 4 mucins.	No	(196)
Normal-24, BE-18, EAC-9	cDNA microarray	Disease progression	Identified 457, 295, and 36 differentially expressed genes, respectively, between normal-EAC, normal-Barrett's esophagus, and BE–EAC groups.	No	(197)
89-EAC	cDNA-mediated annealing, selection, extension, and ligation assay with 502 known cancer-related genes	Disease progression	Identified differential gene expression between early stages of EAC (T1 and T2) vs. late (T3 and T4). Gene expression profile revealed *ERBB4*, *ETV1*, *TNFSF6*, *MPL* genes to be common between advanced tumor stage and lymph node metastasis.	No	(198)
Normal esophageal mucosa-9, esophagitis-6, BE-10, EAC-5, GEJ adenocarcinoma-9, stomach samples-32 (normal mucosa-11, IM-9, intestinal-type adenocarcinoma-7, and diffuse carcinoma-5)	cDNA microarray	Disease progression	On the basis of the expression profile, genes associated with the lipid metabolism and cytokine nodule are found to be significantly altered between EAC and other groups.	No	(199)
Seventeen paired samples of normal, BE/EAC	cDNA microarray	Disease progression	Each tissue type expresses distinct set of genes, which can differentiate between their phenotypes. Barrett's esophagus and EAC expresses similar set of stromal genes that are different from normal epithelium.	No	(200)
BE-19, EAC-20 (98 tissue specimens were collected and categorized into different groups)	On the basis of previous microarray studies 23 genes were validated using real-time PCR	Disease progression	Out of 23 genes, panel of 3 genes (*BFT*, *TSPAN*, and *TP*) was able to discriminate between Barrett's esophagus and EAC in internal validation with 0% classification error.	N.A.	(201)

(Continued on the following page)

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Table 3. Summary of gene expression profiling studies for BE/EAC (Cont'd)

Sample size	Array description	Outcome	Findings	External validation	Ref.
Normal-30, BE-31, gastric mucosa-34, duodenum-18	Biomarkers for Barrett's esophagus were identified using 3 publicly available microarray datasets and validated using real-time PCR and immunohistochemistry.	Disease progression	Out of 14 genes identified, dopa decarboxylase (*DDC*) and Trefoil factor 3 (*TFF3*) were validated to be upregulated in Barrett's esophagus.	N.A.	(202)
EAC-56	Oligonucleotide microarray and aCGH	Disease progression	Identified 4 new genes (*EGFR*, *WT1*, *NEIL2*, and *MTMR9*) to be overexpressed in 10% to 25% EAC. Expression levels of these 4 genes differentiated patients with EAC into 3 groups namely good, average, and poor depending upon their prognosis ($P < 0.008$)	Immunohistochemistry	(203)
BE/LGD-72, HGD-11, EAC-15	Bacterial artificial chromosome aCGH	Disease progression	Copy number changes were more common and larger as disease progress to later stages. Patients having copy number alterations involving more than 70 Mbp were at increased risk of progression to EAC ($P = 0.0047$)	No	(60)
EAC-30, BE-6, LGD-9, HGD-10	Genome-wide CGH	Disease progression	Loss of 7q33-q35 was found in HGD as compared with precursor LGD ($P = 0.01$). Loss of 16q21-q22 and gain of 20q11.2-q13.1 was significantly different between HGD and EAC ($P = 0.02$ and 0.03, respectively).	No	(56)
EAC-30, lymph node metastasis-8, HGD-11, LGD-8, and BE-6 from 30 EAC patient biopsy samples	CGH	Disease progression	Identified regions undergoing copy number loss and amplification during each stage of transition. Average number of chromosomal imbalance sequentially increased from BE–LGD–HGD–EAC–lymph node metastasis.	No	(54)
Forty-two patients represent different stages of disease	SNP array	Disease progression	SNP abnormalities increases from 2% to more than 30% as the disease progress from Barrett's esophagus to EAC. Total number of SNP alterations in tissue samples is tightly correlated with DNA abnormalities such as aneuploidy and LOH.	No	(57)
EAC-27 and matched normal-14	SNP array	Disease progression	Confirmed previously described genomic alterations such as amplification on 8q and 20q13 or deletion/LOH on 3p and 9p. Also identified alterations in several novel genes and DNA regions in EAC samples.	No	(58)

(Continued on the following page)

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Table 3. Summary of gene expression profiling studies for BE/EAC (Cont'd)

Sample size	Array description	Outcome	Findings	External validation	Ref.
EAC-26	SNP array	Disease progression	Confirmed previously reported frequent changes to *FHIT*, *CDKN2A*, *TP53*, and *MYC* genes in EAC. Identified *PDE4D* and *MGC48628* as tumor-suppressor genes.	No	(59)
EAC-35	cDNA microarray	Response to chemotherapy	Identified 165 differentially expressed genes between poor ($n = 17$) and good outcome ($n = 18$) patient groups. Top functional pathway based on differential gene expression was identified to be Toll-receptor signaling.	No	(204)
EAC-47 (locally advanced tumor)	cDNA microarray	Response to chemotherapy	Identified 86 genes showing at least 2-fold difference between chemotherapy responders ($n = 28$) and nonresponders ($n = 19$). Ephrin B3 receptor, which showed highest difference between the groups, showed strong membrane staining in chemotherapy responding tumors using immunohistochemistry.	No	(205)
Patients with EAC-19 undergoing chemoradiotherapy	Oligo-microarray	Response to chemoradiotherapy	Reduced expression of *IVL*, *CRNN*, *NICE-1*, *S100A2*, and *SPPR3* genes correlated with poor survival and nonresponse to chemotherapy.	No	(206)
19 patients (EAC-16, ESCC-2 and adenosquamous carcinoma-1) undergoing chemoradiotherapy	Oligo-microarray	Response to chemoradiotherapy	Lower expression for panel of genes *PERP*, *S100A2*, and *SPRR3* was associated with nonresponse to therapy. Pathway analysis identified downregulation of apoptosis in nonresponders.	No	(207)
EAC-174, ESCC-36	SNPs associated with the chemotherapy drug action pathway	Response to chemoradiotherapy	Identified association between genetic polymorphisms and response to preoperative chemotherapy (fluorouracil and platinum compounds) and radiotherapy.	No	(208)

NOTE: Majority of studies described in this table include validation of results in the same patient cohort as used in discovery phase. Only studies that included validation using independent patient cohort are described as external validation.

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

decade, several studies conducted using advanced genomic techniques such as array-comparative genomic hybridization (aCGH) and SNP arrays confirmed previously reported copy number alterations and identified novel genomic loci undergoing changes during process of metaplasia–dysplasia–carcinoma development (54–60). It has been shown that as the disease progresses from early to late stages, SNP abnormalities increase from approximately 2% to 30% (54, 57). The total number of SNP alterations in tissue samples is tightly correlated with previously reported DNA abnormalities such as aneuploidy, copy number alterations, and LOH highlighting the application of SNP-based genotyping to assess genomic abnormalities (54–60). Thus, SNP-based genotyping provides an alternative way to assess genomic abnormalities during EAC pathogenesis.

Studies on gene expression changes in EAC have been propelled by recent progress in genomic technologies, each identifying unique sets of gene expression profile, which can be used as a biomarker panel for disease diagnosis, prognosis, or to predict response to therapy (Table 3). Moreover, determination of the gene expression changes has been extremely helpful to understand detailed pathogenesis and will form basis for developing future therapies. However, future validation using independent sample cohorts will be necessary for the majority of these potential biomarkers.

Apart from genomic abnormalities associated with the disease progression, inheriting genetic factors are also implicated for EAC development. Risk for BE/EAC and GERD is increased by 2- to 4-fold when a first-degree relative is already affected by any of these conditions (61). Recently, a study conducted by The Esophageal Adenocarcinoma Genetics Consortium and The Wellcome Trust Case Control Consortium identified link between SNPs at the MHC locus and chromosome 16q24.1 with risk for Barrett's esophagus (62). They also identified SNPs associated with body weight measures that were present with more than expected frequency in Barrett's esophagus samples supporting epidemiologic findings about obesity as a risk factor for Barrett's esophagus and EAC (62). Wu and colleagues examined the relationship between presence of risk genotypes and the onset of EAC. They identified 10 SNPs associated with the age of EAC onset. Genes associated with 5 of 10 SNPs identified were known to be involved in apoptosis (63).

Recently, published cancer genome–sequencing studies have given deeper insights into the genomic abnormalities associated with the EAC pathogenesis. The comparative genomic analysis between EAC and ESCC reported by Agrawal and colleagues (64) confirmed previously very well-described association of *p53* gene mutations with esophageal cancer development. The authors also conducted comparative genome-wide analysis between matched Barrett's esophagus and EAC patient tissue samples and concluded that the majority of genomic changes occur early during EAC development, at the stage of Barrett's esophagus (64). Similar conclusions were made by next-generation sequencing of biopsy samples obtained from the same patient at the stage of Barrett's esophagus and EAC (65). The authors also identified *ARID1A* as novel tumor-suppressor gene and around 15% of patients with EAC showed loss of ARID1A protein in tissue samples. *In vitro* studies suggested it to be associated with cell growth, proliferation, and invasion (65). Very recently published high-resolution methylome analysis has provided first evidence for methylation changes at genomic regions that encode noncoding RNAs. The authors identified long noncoding RNA, AFAP1-AS1, to be severely hypomethylated in Barrett's esophagus and EAC tissue samples, silencing of which significantly reduced aggressiveness of EAC cell lines OE33 and SKGT4 (66).

Taken together, genomic abnormalities play key roles during each stage of transformation from normal squamous epithelium to EAC.

Cancer-Related Inflammation

Gastric and bile acid exposure in the esophageal epithelium leads to the development of chronic inflammatory conditions mainly driven by elevated levels of proinflammatory cytokines. Chronic inflammatory responses induce cell survival and increase cell proliferation, hence play key roles in the development of EAC (67, 68). Expressions of various inflammatory molecules such as COX-2, NF-κB, interleukin (IL)-6, IL-8, and matrix metalloproteinases (MMP) have been evaluated as prognostic biomarkers for BE/EAC development.

Exposure to gastric/bile acid and cytokines leads to increased COX-2 expression (69). COX-2 is a rate-limiting enzyme that regulates synthesis of prostaglandins from arachidonic acid. COX-2 directly increases cell proliferation and promotes tumor invasion (69), and COX-2–mediated increase in prostaglandin synthesis could result in tumor growth and angiogenesis (70). COX-2 expression has been detected in disease-free esophageal tissue homogenates using immunoblotting (69). In comparison with GERD, patients suffering from erosive reflux show slightly higher gene expressions of this enzyme in tissue samples (71). Several studies have shown significantly increased COX-2 expression correlating with the disease progression from Barrett's esophagus to dysplasia and EAC (69, 72–75). Furthermore, expression levels of COX-2 have been shown to have a prognostic value in EAC with higher levels associated with poor survival and increased chances of tumor relapse (76, 77).

Another well-studied inflammatory biomarker NF-κB is activated in response to exposure with bile acid and elevated NF-κB expression levels are found during Barrett's esophagus, dysplasia, and adenocarcinoma (78–80). Activated NF-κB translocates from cytoplasm to nucleus and upregulates transcription of the genes involved in inflammatory processes. Moreover, nuclear NF-κB expression has been shown to be correlated with the patient response to chemoradiotherapy. All of the patients who showed complete response to chemoradiotherapy

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

had elevated NF-κB levels pretreatment and showed lack of active NF-κB posttreatment (81).

In line with NF-κB and COX-2, expression of individual or combinations of proinflammatory cytokines IL-1β, IL-6, IL-8, and TNF-α is significantly increased in Barrett's esophagus and EAC as compared with squamous epithelium (82–84). IL-1β and IL-8 expression levels also correlate with the stage of EAC (79). Patients who responded to neoadjuvant chemotherapy treatment showed significantly reduced expressions of IL-8 and IL-1β in postchemotherapy esophageal tissue sections (81). IL-6 is activated in response to reflux and the IL-6/STAT3 antiapoptotic pathway may underlie the development of dysplasia and tumor (85). Serum IL-6 levels were reported to provide 87% sensitivity and 92% specificity for EAC diagnosis in a recent retrospective study (86). However, the study only compared between healthy and EAC groups. It would be interesting to see how early it can diagnose EAC during the process of metaplasia–dysplasia. Combination of cytokines IFN-γ, IL-1α, IL-8, IL-21, and IL-23 along with platelet proteoglycan and miRNA-375 expression profiling has been shown to build an inflammatory risk model, which has clinical use to determine prognosis for patients with EAC (67).

MMPs are a family of proteolytic enzymes involved in the degradation of extracellular matrix components. MMPs play a role in both inflammation and tumor metastasis. Immunohistochemical staining for MMP-1, MMP-2, MMP-7, and MMP-9 has been reported to be significantly higher in EAC as compared with healthy individuals (87, 88). Higher level of MMP-1 expression has been associated with the lymph node metastases and possibly poor patient survival (89). Expression of MMP-9 is shown to be an early event during the EAC transformation and its expression levels are correlated with the progression of the disease (90–92). Activity of MMPs is inhibited by a family of proteins called tissue inhibitors of metalloproteinases (TIMP). Specifically, *TIMP-3* gene is methylated in EAC development and its reduced expression is associated with stage of the tumor and patient survival (93). On contrary, Salmela and colleagues described elevated TIMP-1 and TIMP-3 expression in EAC tumor samples (88).

Although the underlying tissue inflammation is very closely associated with EAC development and several inflammation-related biomarkers have been identified, these remain to be validated in large-scale biomarker studies.

Cell Cycle–Related Abnormalities

To compensate for the tissue damage induced by gastric/bile acid, the underlying epithelium starts to proliferate rapidly and become uncontrolled resulting in neoplasia. To meet the proliferation requirements, the cells have to overcome cell-cycle checkpoints. Cyclin D1 overexpression is one such means by which cells overcome G_1–S checkpoint, and cyclin D1 immunohistochemical staining has been proposed to identify patients with Barrett's esophagus with an increased risk for EAC (94). In contrast to cyclin D1, expression of p16 protein results in cell-cycle arrest in G_1 phase as it has been shown to inhibit cyclin-dependent kinase–induced phosphorylation of retinoblastoma protein. Early genomic abnormalities during EAC development significantly affect p16 protein expression, which can be determined using immunostaining and implemented as a potential biomarker (95). Further large-scale trials are required to confirm cell-cycle abnormalities during EAC development to implement them as a biomarker.

Bottom of the pyramid in Fig. 1 represents list of biomarkers in the initial stages of development. Tumors harboring overexpression of growth factor receptors [EGF receptor (EGFR) and HER-2] are associated with poor patient survival (96, 97), whereas those overexpressing apoptosis regulator Bcl-2 protein showed prolonged survival (98). Incipient angiogenesis is a marked feature of Barrett's esophagus and underlining tissue expresses angiogenesis markers VEGF and its receptors (99). Neovascularization continues as the disease progresses from Barrett's esophagus to EAC. Measuring the degree of neovascularization correlated with histopathologic grade of the tumor and associated with the patient survival (100). Expression of 2 prominent cell proliferation markers, PCNA and Ki-67, has been described to be altered during BE–EAC development (101).

miRNA

miRNA was first discovered in *Caenorhabditis elegans* (102) and since then it has been widely studied in a variety of biologic phenomena. These short stretches of approximately 21 nucleotides do not code for protein but play important roles in gene regulation by either suppressing protein synthesis or causing mRNA cleavage. Unlike siRNA, miRNA can target multiple genes on remote loci and therefore control diverse group of proteins. Several key properties of carcinogenesis have been shown to be regulated via miRNA, for example, angiogenesis and metastasis (103).

With increased biologic understanding of miRNAs and their role in cancer, they have been proposed in several different clinical applications including cancer diagnosis and tumor prognosis, tumor classification, and also as a therapeutic target for disease intervention. Differential tissue miRNA expression has been observed in several different malignancies and these changes can be used for diagnosis and classification of the tumors (103). miRNA bioarrays were first used to show differential miRNA expression in healthy, Barrett's esophagus, and EAC tissue samples (104). Since then, a number of different studies have identified miRNA changes associated with the development of the BE–EAC. Table 4 summarizes primary findings of miRNA expression profiling studies along with statistical significance and fold-change values. Biologic significance for some of the miRNA-related changes is discussed later.

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Smith and colleagues identified reduced expression of miR-200 and miR-141 in Barrett's esophagus and EAC tissue samples. They conducted bioinformatics analysis and correlated these miRNA expression changes with cellular processes such as cell cycle, cell proliferation, apoptosis, and cell migration (105). miR-196a, which is described as a marker of progression from Barrett's esophagus to EAC, can increase cell proliferation and anchorage-independent growth and inhibit apoptosis in EAC cell lines *in vitro* (106). The downstream targets for miR-196a are verified to be Annexin A1, S100 calcium-binding protein A9, small proline-rich protein 2C, and Keratin 5, which showed reduced expression in EAC patient tissue samples as compared with normal epithelium (106, 107). Several studies described in Table 4 report overexpression of miR-192 during EAC carcinogenesis. miR-192 has been reported to be a target of p53 and has been able to suppress cancer progression in osteosarcoma and colon cancer cell lines through p21 accumulation and cell-cycle arrest (108). As shown in Table 4, miR-21 is overexpressed during BE/EAC and it can function as an oncogene as shown in tumors of breast, brain, lung, prostate, pancreas, colon, liver, and chronic lymphocytic leukemia. It negatively regulates tumor- and metastasis-suppressor genes *PTEN*, *TPM1*, *PDCD4*, and *Sprouty2* (109–112). miR-194 expression is regulated by hepatocyte nuclear factor (HNF)-1α transcription factor, which is induced during BE/EAC and may lead to upregulation of miR-194 (109). Higher expression of miR-194 is also observed in metastatic pancreatic cell lines (113). Among miRNAs found to be downregulated during EAC development, let-7 family of miRNAs is tumor-suppressive and negatively regulates *Ras* oncogene. Fassan and colleagues confirmed upregulation of *HMGA2*, which is one of the target of let-7 miRNA, using immunohistochemistry in tissue samples (110, 112, 114). Further studies in the regards of miRNA and miRNA target genes will improve the biologic understanding of EAC pathogenesis and may also provide novel molecular targets for disease intervention.

Notably, miRNAs are found to be stable in serum encapsulated in microvesicles and can be accessed easily (115). In fact, circulating miRNA profiling has shown distinct expression patterns in a number of cancers, other than EAC (116). This opens up new avenues for circulating miRNA changes as a potential biomarker for EAC.

Glycoproteins

Protein glycosylation is a common posttranslational modification with almost half of the proteins synthesized undergoing 1 of the 2 major types either N-linked or O-linked glycan modifications. The biosynthetic process of glycosylation is regulated by the expression and localization of glycosyltransferases/glycosidases and the availability of substrate glycans (117).

Aberrant glycosylation changes have previously been reported in several different cancers namely breast cancer, prostate cancer, melanoma, pancreatic cancer, ovarian cancer, etc. (118, 119). These changes include truncated forms of O-glycans, increased degree of branching in N-glycans, and elevated sialylation, sulfation, and fucosylation with a range of other possible variations (119). The differential glycosylation can alter protein interactions, stability, trafficking, immunogenicity, and function (118). Tumor-specific glycosylation changes are actively involved in neoplastic progression, namely metastasis, as glycoproteins are found abundantly on cell surfaces and extracellular matrices and therefore play a vital role in cellular interactions.

Lectins are a family of glycan-binding proteins extensively used in glycobiology due to preferential binding of each lectin to recognize specific glycan structures (119, 120). The first effort to identify differential glycosylation in the progression to Barrett's esophagus and EAC was made in 1987 by Shimamoto and colleagues using differential binding pattern to 5 lectins in tissue specimens (121). The glycoconjugate expression profile in Barrett's esophagus was found to be significantly different from normal esophageal epithelium. Interestingly, glycoconjugate expression between Barrett's esophagus and normal duodenum was quite similar. There were minimal glycoconjugate expression changes between Barrett's esophagus and LGD. However, EAC tissue samples showed significantly different lectin-binding pattern than BE/LGD (121). Using rabbit esophageal epithelium, Poorkhalkali and colleagues showed differential lectin binding in response to acid/pepsin exposure suggesting acid exposure can induce cell surface glycosylation changes (122). In 2008, Neumann and colleagues used 4 different lectins to identify pathologic mucosal changes (123). They observed 2 distinct lectin-binding patterns. One was associated with the GERD, whereas the other pattern was characteristic for Barrett's esophagus mucosa. Specifically, UEA (*Ulex europaeus*) lectin binding was upregulated in Barrett's esophagus tissue sections, which suggests possible increase in fucosylation during the disease progress (123). A recently published study has concluded that dysplasia can alter glycan expression and lectin binding to the tissue samples. Fluorescently labeled WGA (wheat germ agglutinin) lectin-binding intensity was found to be inversely related to the degree of dysplasia (124). Furthermore, the authors used fluorescent-capable endoscope *ex vivo* in the study and followed all the protocols in a manner that exactly mimics a clinical study *in vivo*. Followed by topical fluorescein-labeled WGA spray, the authors measured fluorescence in the tissue samples. Measurement of lectin fluorescence was a more sensitive approach to identify dysplastic lesions as compared with white light endoscopic technique. Their data show clinical use of such a lectin-based endoscopic technique if developed further (124). In a phase III biomarker clinical trial study, Bird-Liberman and colleagues combined 3 different abnormalities to predict EAC progression in patients with Barrett's esophagus. Along with using conventional LGD and DNA content abnormalities they used AOL (*Aspergillus oryzae)* lectin binding to the tissue samples, which detects presence of α1-6 fucose on the cell surface

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Table 4. Summary of literature describing miRNA expression changes in BE/EAC

Sample size	Upregulated in BE/EAC	Downregulated in BE/EAC	Ref.
71 (BE-12, Barrett's esophagus without dysplasia-20, LGD-27, EAC/HGD-12)	miR-192 ($P < 0.00001$), miR-196a ($P < 0.05$): upregulated in Barrett's esophagus as compared with healthy tissue.miR-196a expression is correlated with progression from IM-LGD-HGD-EAC ($P < 0.005$).	miR203 ($P < 0.00001$): downregulation in Barrett's esophagus as compared with healthy tissue.	(209)
22 (Barrett's esophagus without dysplasia-11, Barrett's esophagus with dysplasia-11)	miR-15b (3.3-fold; $P < 0.05$), miR-203 (5.7-fold; $P < 0.05$): upregulated in dysplasia as compared with nondysplastic Barrett's esophagus.	miR-486-5p (4.8-fold; $P < 0.05$), miR-let-7a (3.3-fold; $P < 0.05$): downregulated in dysplasia as compared with nondysplastic Barrett's esophagus.	(110)
100 (EAC-100, adjacent normal tissue as control)	miR-21 (~3-fold; $P < 0.05$), miR-223 (~2-fold; $P < 0.05$), miR-192 (~3.5-fold; $P < 0.05$), and miR-194 (~3.5-fold; $P < 0.05$): upregulated in EAC as compared with adjacent normal tissue.	miR-203 (~3-fold; $P < 0.05$): downregulated in EAC as compared with adjacent normal tissue.	(111)
25 (Healthy-9, BE-5, HGD-1, EAC-10)	miR-192 (1.7-fold; FDR < 1 e−07), miR-194 (2-fold; FDR < 1e−07), miR-21 (3.7-fold; FDR = 0.0003), miR-200c (1.9-fold; FDR = 0.0015), miR-93 (1.3-fold; FDR = 0.0108): upregulated in EAC as compared with Barrett's esophagus.	miR-27b (1.43-fold; FDR = 0.0003), miR-342 (1.25-fold; FDR = 0.0015), miR-125b (2-fold; FDR = 0.0108), miR-100 (1.25-fold; FDR = 0.011): downregulated in EAC as compared with Barrett's esophagus.	(104)
75 (Healthy-15, BE-15, LGD-15, HGD-15, EAC-15)	miR-215 (62.8-fold; $P <$ 1e−07), miR-192 (6.34-fold; $P <$ 1e−07): upregulated in Barrett's esophagus in comparison with normal tissue and remained at similar levels with disease progress.	miR-205 (10-fold; $P =$ 1.39e−0.5), let-7c (2.04-fold; $P =$ 3.11e−05), miR-203 (6.67-fold; $P =$ 3.2e−0.5): downregulated in Barrett's esophagus in comparison with normal tissue and remained at similar levels as disease progresses.	(114)
91 (LGD-31, HGD-29, EAC-31, In all cases adjacent normal tissue used as a control)	miR-200a (13.5-fold; $P = 0.02$), miR-513 (1.58-fold; $P = 0.03$), miR-125b (9.2-fold; $P = 0.04$), miR-101 (1.83-fold; $P = 0.04$), miR-197 (1.61-fold; $P = 0.04$): upregulated in LGD to HGD transition.	miR-23b (1.45-fold; $P = 0.007$), miR-20b (1.56-fold; $P = 0.01$), miR-181b (2.22-fold; $P = 0.03$), miR-203 (1.49-fold; $P = 0.03$), miR-193b (2.70-fold; $P = 0.04$), miR-636 (4.17-fold; $P = 0.04$): downregulated in LGD to HGD transition. let-7a (1.75-fold; $P = 0.01$), let-7b (1.59-fold; $P = 0.009$), let-7c (1.69-fold; $P = 0.03$), let-7f (1.69-fold; $P = 0.03$), miR-345 (2-fold; $P = 0.02$), miR-494 (1.72-fold; $P = 0.03$), miR-193a (2.27-fold; $P = 0.05$): downregulated in HGD-EAC development process.	(112)
48 (BE-19, EAC-29)	miR-21 (~2.8-fold; $P < 0.05$), miR-143 (~11.3-fold; $P < 0.05$), miR-145 (~3.4-fold; $P < 0.05$), miR-194 (~126-fold; $P < 0.05$), miR-215 (~18-fold; $P < 0.05$): upregulated in Barrett's esophagus as compared with adjacent normal tissue.	miR-203 (~17-fold; $P < 0.05$), miR-205 (~175-fold; $P < 0.05$): downregulated in Barrett's esophagus as compared with adjacent normal tissue.miR-143 (~3-fold; $P < 0.05$), miR-145 (~1.8-fold; $P < 0.05$), miR-215 (~3.1-fold; $P < 0.05$): Lower expression in EAC as compared with Barrett's esophagus.	(109)
49 (IM-15, HGD-14, and EAC-20, adjacent normal tissue)	—	miR-31 (>4-fold; $P < 0.02$), miR-375 (>4-fold; $P < 0.001$): downregulated in transition from Barrett's esophagus to EAC.	(210)
37 (BE-17, EAC-20, 9 adjacent normal tissue samples)	—	miR-141 (~2-fold; $P = 0.0126$), miR-200a (~2.5-fold; $P = 0.0001$), miR-200b (~2.1-fold; $P < 0.0001$), miR-200c (~1.9-fold; $P = 0.0014$), miR-429 (~1.8-fold; $P = 0.0031$): underexpressed in EAC as compared with Barrett's esophagus.	(105)

(Continued on the following page)

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

Table 4. Summary of literature describing miRNA expression changes in BE/EAC (Cont'd)

Sample size	Upregulated in BE/EAC	Downregulated in BE/EAC	Ref.
11 (EAC-11, different lesions were collected from these patients and classified into Barrett's esophagus, LGD, HGD, and EAC)	miR-196a is overexpressed in early EAC (151-fold) > HGD (62.2-fold; $P = 0.00002$) > LGD (31.1-fold; $P = 0.0005$) > Barrett's esophagus (28.9-fold; $P = 0.00001$). Fold changes are calculated as compared with normal epithelium.	—	(107)
45 (patients with EAC undergoing surgery)	miR-143 ($P = 0.0148$), miR-199a_3p ($P = 0.0009$), miR-199a_5p ($P = 0.0129$), miR-100 ($P = 0.0022$) and miR-145 ($P = 0.1176$) expression predicted a worse survival followed by esophagectomy. Overexpression of miR-199a_3p/_5p and miR-99b was associated with lymphnode metastasis.	Downregulation of miR-143 ($P = 0.0049$) and miR-145 ($P = 0.0069$) in EAC as compared with adjacent normal tissue.	(211)
24 (BE-24, progression to EAC-7, not progressed to EAC-17 in at least 5-y follow-up)	miR-192 (ROC AUC = 0.61), 194 (ROC AUC = 0.70), 196a (ROC AUC = 0.80), and 196b (ROC AUC = 0.74) showed significantly higher expression in Barrett's esophagus samples from patients who progressed to EAC as compared with those who did not progress to EAC.	—	(212)
5 (patients with EAC undergoing surgery. Adjacent benign tissue as a control)	miR-296 is overexpressed ~2-fold in EAC as compared with adjacent benign tissue.	—	(213)
22 patients with locally advanced EAC tumor undergoing surgery	Negative association between miR-148a expression and tumor differentiation ($P < 0.001$). Significantly higher expression of miR-148a in tumors located in the lower esophagus as compared with tumors in the middle esophagus ($P = 0.021$).	—	(214)
99 EAC patient tissue samples undergoing surgery	miR-30e ($P = 0.002$) and miR-200a ($P = 0.044$) expression were associated with poor overall survival. miR-16-2 ($P = 0.027$) and miR-30e ($P = 0.002$) expression were associated with poor disease-free survival.	—	(215)

NOTE: Wherever needed, fold-change values are calculated/adapted from the expression/fold-change values described in the original article to have uniform format for the purpose of this review.

(125). Thus, monitoring tissue glycan changes can be combined with existing biomarkers to improve the predictive power of the currently used biomarkers.

A potential mechanism responsible for these changes is considered to be bile acid exposure-induced gene expression and secretory pathway changes in esophageal epithelium. Using carbohydrate-specific lectins that detect N- and O-linked glycosylation and core fucosylation, Byrne and colleagues have shown differential lectin binding to the cell surface and differential intracellular localization when normal squamous and Barrett's metaplastic cell lines were treated with deoxycholic acid (126). Nancarrow and colleagues profiled whole-genome expression in normal squamous esophageal epithelium, Barrett's esophagus, and EAC and concluded that Barrett's esophagus is a tissue with enhanced glycoprotein synthesis machinery to provide strong mucosal defense against acid exposure (127).

Outlook—Circulating Biomarkers

Last 3 decades showed continuously increased EAC incidences and similar trend is expected in future because of rising incidences of obesity and GERD in the population. Current endoscopic screening program might benefit the highest risk population to monitor disease progression. Monitoring dysplasia in the tissue samples has not provided fruitful outcome for early diagnosis; however, inclusion of the genomic and cell-cycle biomarkers has shown definite improvement in the predictive power over currently used histologic technique. Any biomarker

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

requiring tissue samples is going to be difficult to implement for population screening and will not be economically viable. An alternative to tissue-based techniques is to investigate changes in circulating biomarkers. Blood is relatively easy to access and can be monitored frequently, ultimately increasing the possibility of detecting early dysplastic changes.

Circulating tumor cells could be one source of biomarkers. Although readily found in the blood, technological advancements are required for sensitive early detection of the low number of tumor cells present in the circulation (128, 129). Alternative to the detection of circulating tumor cells, Zhai and colleagues applied genome-wide DNA methylation profiling approach to cell-free circulating DNA. They found that cell-free circulating DNA methylation profile is a replica of methylation profile found in matched tumor tissue samples and can discriminate between healthy, Barrett's esophagus, and EAC conditions (130). Kawakami and colleagues (131) studied methylation of *APC* gene in matched tumor samples and plasma. Unlike tumor samples that showed hypermethylation of *APC* DNA early during the EAC development, matched plasma samples from patients suffering from Barrett's esophagus and gastritis were found to be negative for *APC* methylation changes. Moreover, as compared with 92% (48 of 52) of EAC tissue samples, only 25% (13 of 52) of plasma samples were positive for circulatory *APC* methylation changes. However, there was a strong correlation between stage of the tumor and plasma positivity for methylated *APC* (131). In combination with *DAPK* methylation, measurement of preoperative *APC* methylation in peripheral blood was able to discriminate between long (>2.5 years) and short survivors with a sensitivity of 99.9% and specificity of 57.1% (132). Taken together, tracking circulatory DNA methylation changes during EAC development may be an alternative approach to predict early EAC.

Tumor cell moulds the microenvironment to support oncogenesis by releasing soluble and vesicular components, including enzymes, microvesicles, proteoglycans, chemokines, and cytokines (133). The tumor microenvironment components are shed into the circulation and may be extremely useful as an early diagnostic biomarker. This concept was recently showed by Pitteri and colleagues using an inducible HER2/neu mouse model (134). They showed that plasma proteome profiling has the ability to detect cancer before it actually develops. Furthermore, a linear correlation was shown for plasma levels of candidate biomarker proteins with the tumor progression, which were reversed upon tumor regression (134).

Both encapsulated miRNAs and secreted glycoproteins are prime candidates for circulating biomarkers released by the tumor microenvironment. Circulating miRNA is secreted in nanometer-sized vesicles called exosomes or microvesicles. An advantage of circulating miRNA over protein biomarkers is the ability for amplication, increasing the sensitivity of detection. Comparative analysis of circulating miRNA can be conducted using miRNA microarray and quantitative real-time PCR (116). Future studies should aim to discover and validate circulating miRNA changes associated with EAC development and progression.

Glycan Profiling

For Barrett's esophagus and EAC, serum glycan profiling using mass spectrometry has identified differential expression of glycan structures in different disease states. Mechref and colleagues analyzed N-linked glycan diversity present in 84 patient serum samples (Healthy-18, BE-5, HGD-11, and EAC-50; ref. 135). They identified 98 glycan features with different intensities in disease onsets and 26 of them correspond to known glycan structures. They showed statistically significant glycan changes between 4 different conditions (Healthy/BE/HGD/EAC) with 3 of the known potential N-glycan biomarkers predicting EAC with 94% sensitivity and 60% specificity (135). Another study used microchip electrophoresis with laser-induced fluorescence detection for N-glycan profiling and were able to differentiate between the healthy, Barrett's esophagus, HGD, and EAC conditions (136). Similar to abovementioned N-glycan profiling studies, very recently, Gaye and colleagues showed that ion mobility-mass spectrometric analysis of serum N-glycan can also distinguish between normal and EAC phenotype (137). All of these studies unanimously suggest circulatory N-linked glycan changes during EAC pathogenesis. Mann and colleagues enriched fucosylated serum glycoproteins using lectins and then used shot gun proteomics to identify protein in different physiologic states, including healthy samples, Barrett's esophagus, and EAC (138). Although the study showed promising trends, the statistical power was not achieved because of the very low number of samples. To improve the throughput of glycoproteomics studies, we developed lectin magnetic bead array-mass spectrometry (LeMBA-MS), a high-throughput platform where a panel of lectins individually immobilized on magnetic beads is used to capture glycoproteins followed by on-bead trypsin digest and liquid chromatography–tandem mass spectrometry for protein identification (139, 140). Parallel screening of a panel of lectins may be helpful to identify differentially glycosylated circulating proteins during EAC pathogenesis.

Metabolic Profiling

In recent past, efforts have been made to profile metabolic changes associated with EAC pathogenesis. Metabolic profiling studies have identified changes associated with nucleoside metabolism, tricarboxylic acid cycle, fatty acid, and amino acid metabolism during EAC development in tissue samples and more importantly using easily accessible biofluids, blood and urine. Early metabolic changes in the histologically normal epithelium were observed, particularly for phosphocholine, glutamate, myo-inositol, adenosine-containing compounds, uridine-containing compounds, and inosine (141). Djukovic and colleagues used targeted approach to profile 8

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

different serum nucleosides between healthy subjects ($n = 12$) and patients with EAC ($n = 14$) using high-performance liquid chromatography coupled with triple quadrupole mass spectrometer. Among 8 nucleosides they profiled, 5 were significantly different between the 2 groups. Three of 5 significantly different nucleosides, 1-methyladenosine, N^2,N^2-dimethylguanosine, and N^2-methylguanosine, were methylated nucleosides indicating increased tRNA methylation, similar to DNA hypermethylation in EAC condition (142). Zhang and colleagues studied serum metabolomic changes using nuclear magnetic resonance (NMR) alone and NMR in combination with liquid chromatography/mass spectrometry (LC/MS) in EAC ($n = 67$), HGD ($n = 9$), Barrett's esophagus ($n = 3$), and healthy volunteers ($n = 34$). Their model based on Partial Least Square Discrimination Analysis was able to distinguish between different phenotypes by achieving area under receiver operating characteristics curve (AUROC) as high as 0.95. On the basis of candidate metabolites, they identified altered pathways associated with EAC development to be energy metabolism, fatty acid metabolism, and amino acid metabolism (143, 144). Urine metabolomics could also distinguish between healthy, Barrett's esophagus, and EAC phenotypes. Davis and colleagues generated urine metabolic signatures, which were able to discriminate between healthy, Barrett's esophagus, and EAC phenotypes, as well as distinguish EAC from pancreatic cancer (145). These metabolic profiling studies open up new avenues to detect early EAC using circulatory biomarkers.

Improved biologic understanding, in combination with technical advancements in the field of genomics, proteomics, glycomics, and metabolomics, has played key roles in the identification and validation of circulatory biomarkers for EAC. Development of an assay platform, which can be clinically used for these circulatory biomarkers, will help to conduct the large scale multicentered trials and transform the circulatory biomarkers into clinical use.

Summary and Future Perspectives

Clinical advancements in endoscopy and new tissue sampling techniques such as brush cytology can improve the endoscopic-biopsy management of BE/EAC in near future. Genomic abnormalities and cell-cycle biomarkers have already shown their potential use to diagnose early pathologic changes using tissue samples. However, wider clinical application will depend on the technical ability of individual clinical pathology laboratories. As these changes are detected in the tissue samples, it would be difficult to implement them in large-scale high-risk population screening to identify early neoplastic changes. Recent advancements in RNA sequencing, circulatory DNA methylation profiling, metabolic profiling, and glycoproteomics may provide ways for the development of noninvasive *in vitro* diagnostic biomarker for routine monitoring and identification of patients with nonsymptomatic BE/EAC. Future studies should focus to combine different classes of circulatory biomarkers in large-scale trials to improve the predictive power of the individual marker. Development of novel cost-effective assay platforms that can transform discoveries from research laboratories to the clinics require equal emphasis for the widespread benefit from the circulatory biomarkers.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: A.K. Shah, N.A. Saunders, M.M. Hill
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A.K. Shah
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.K. Shah
Writing, review, and/or revision of the manuscript: A.K. Shah, N.A. Saunders, A. Barbour, M.M. Hill

Received December 18, 2012; revised March 6, 2013; accepted March 31, 2013; published OnlineFirst April 10, 2013.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin 2012;62:10–29.
2. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010;127:2893–917.
3. Demeester SR. Epidemiology and biology of esophageal cancer. Gastrointest Cancer Res 2009;3:S2–5.
4. Cook MB, Chow WH, Devesa SS. Oesophageal cancer incidence in the United States by race, sex, and histologic type, 1977–2005. Br J Cancer 2009;101:855–9.
5. Bollschweiler E, Wolfgarten E, Gutschow C, Holscher AH. Demographic variations in the rising incidence of esophageal adenocarcinoma in White males. Cancer 2001;92:549–55.
6. Hongo M, Nagasaki Y, Shoji T. Epidemiology of esophageal cancer: orient to occident. Effects of chronology, geography and ethnicity. J Gastroenterol Hepatol 2009;24:729–35.
7. Baquet CR, Commiskey P, Mack K, Meltzer S, Mishra SI. Esophageal cancer epidemiology in Blacks and Whites: racial and gender disparities in incidence, mortality, survival rates and histology. J Natl Med Assoc 2005;97:1471–8.
8. Sonnenberg A. Effects of environment and lifestyle on gastroesophageal reflux disease. Dig Dis 2011;29:229–34.
9. Pickens A, Orringer MB. Geographical distribution and racial disparity in esophageal cancer. Ann Thorac Surg 2003;76:S1367–9.
10. Revels SL, Morris AM, Reddy RM, Akateh C, Wong SL. Racial disparities in esophageal cancer outcomes. Ann Surg Oncol 2013;20:1136–41.
11. Holmes RS, Vaughan TL. Epidemiology and pathogenesis of esophageal cancer. Semin Radiat Oncol 2007;17:2–9.
12. Rutegard M, Lagergren P, Nordenstedt H, Lagergren J. Oesophageal adenocarcinoma: the new epidemic in men? Maturitas 2011;69:244–8.
13. Reid BJ, Li X, Galipeau PC, Vaughan TL. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. Nat Rev Cancer 2010;10:87–101.
14. Kubo A, Corley DA. Body mass index and adenocarcinomas of the esophagus or gastric cardia: a systematic review and meta-analysis. Cancer Epidemiol Biomarkers Prev 2006;15:872–8.
15. Mayne ST, Risch HA, Dubrow R, Chow WH, Gammon MD, Vaughan TL, et al. Nutrient intake and risk of subtypes of esophageal and gastric cancer. Cancer Epidemiol Biomarkers Prev 2001;10:1055–62.

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

16. Wu AH, Wan P, Bernstein L. A multiethnic population-based study of smoking, alcohol and body size and risk of adenocarcinomas of the stomach and esophagus (United States). Cancer Causes Control 2001;12:721–32.
17. Weston AP, Badr AS, Topalovski M, Cherian R, Dixon A, Hassanein RS. Prospective evaluation of the prevalence of gastric *Helicobacter pylori* infection in patients with GERD, Barrett's esophagus, Barrett's dysplasia, and Barrett's adenocarcinoma. Am J Gastroenterol 2000;95:387–94.
18. Tischoff I, Tannapfel A. Barrett's esophagus: can biomarkers predict progression to malignancy? Expert Rev Gastroenterol Hepatol 2008; 2:653–63.
19. Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R, Global Consensus G. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. Am J Gastroenterol 2006;101:1900–20.
20. Oh DS, Demeester SR. Pathophysiology and treatment of Barrett's esophagus. World J Gastroenterol 2010;16:3762–72.
21. DeMeester SR, DeMeester TR. Columnar mucosa and intestinal metaplasia of the esophagus: fifty years of controversy. Ann Surg 2000;231:303–21.
22. Oberg S, Johansson J, Wenner J, Walther B. Metaplastic columnar mucosa in the cervical esophagus after esophagectomy. Ann Surg 2002;235:338–45.
23. American Gastroenterological AssociationSpechler SJ, Sharma P, Souza RF, Inadomi JM, Shaheen NJ. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. Gastroenterology 2011;140:1084–91.
24. Playford RJ. New British Society of Gastroenterology (BSG) guidelines for the diagnosis and management of Barrett's oesophagus. Gut 2006;55:442.
25. Locke GR, Zinsmeister AR, Talley NJ. Can symptoms predict endoscopic findings in GERD? Gastrointest Endosc 2003;58:661–70.
26. Gerson LB, Edson R, Lavori PW, Triadafilopoulos G. Use of a simple symptom questionnaire to predict Barrett's esophagus in patients with symptoms of gastroesophageal reflux. Am J Gastroenterol 2001; 96:2005–12.
27. Wang KK, Sampliner REPractice Parameters Committee of the American College of G. Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett's esophagus. Am J Gastroenterol 2008;103:788–97.
28. Spechler SJ. Dysplasia in Barrett's esophagus: limitations of current management strategies. Am J Gastroenterol 2005;100:927–35.
29. Incarbone R, Bonavina L, Saino G, Bona D, Peracchia A. Outcome of esophageal adenocarcinoma detected during endoscopic biopsy surveillance for Barrett's esophagus. Surg Endosc 2002;16: 263–6.
30. Ferguson MK, Durkin A. Long-term survival after esophagectomy for Barrett's adenocarcinoma in endoscopically surveyed and nonsurveyed patients. J Gastrointest Surg 2002;6:29–35.
31. Fountoulakis A, Zafirellis KD, Dolan K, Dexter SP, Martin IG, Sue-Ling HM. Effect of surveillance of Barrett's oesophagus on the clinical outcome of oesophageal cancer. Br J Surg 2004;91:997–1003.
32. Corley DA, Levin TR, Habel LA, Weiss NS, Buffler PA. Surveillance and survival in Barrett's adenocarcinomas: a population-based study. Gastroenterology 2002;122:633–40.
33. Downs-Kelly E, Mendelin JE, Bennett AE, Castilla E, Henricks WH, Schoenfield L, et al. Poor interobserver agreement in the distinction of high-grade dysplasia and adenocarcinoma in pretreatment Barrett's esophagus biopsies. Am J Gastroenterol 2008;103:2333–40.
34. Moyes LH, Going JJ. Still waiting for predictive biomarkers in Barrett's oesophagus. J Clin Pathol 2011;64:742–50.
35. Collard JM. High-grade dysplasia in Barrett's esophagus. The case for esophagectomy. Chest Surg Clin N Am 2002;12:77–92.
36. Dulai GS, Guha S, Kahn KL, Gornbein J, Weinstein WM. Preoperative prevalence of Barrett's esophagus in esophageal adenocarcinoma: a systematic review. Gastroenterology 2002;122:26–33.
37. Bytzer P, Christensen PB, Damkier P, Vinding K, Seersholm N. Adenocarcinoma of the esophagus and Barrett's esophagus: a population-based study. Am J Gastroenterol 1999;94:86–91.
38. Kadri S, Lao-Sirieix P, Fitzgerald RC. Developing a nonendoscopic screening test for Barrett's esophagus. Biomark Med 2011;5: 397–404.
39. Hirst NG, Gordon LG, Whiteman DC, Watson DI, Barendregt JJ. Is endoscopic surveillance for non-dysplastic Barrett's esophagus cost-effective? Review of economic evaluations. J Gastroenterol Hepatol 2011;26:247–54.
40. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001;69:89–95.
41. Sund M, Kalluri R. Tumor stroma derived biomarkers in cancer. Cancer Metastasis Rev 2009;28:177–83.
42. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;93:1054–61.
43. Jin Z, Cheng Y, Gu W, Zheng Y, Sato F, Mori Y, et al. A multicenter, double-blinded validation study of methylation biomarkers for progression prediction in Barrett's esophagus. Cancer Res 2009;69: 4112–5.
44. Sato F, Jin Z, Schulmann K, Wang J, Greenwald BD, Ito T, et al. Three-tiered risk stratification model to predict progression in Barrett's esophagus using epigenetic and clinical features. PLoS ONE 2008;3:e1890.
45. Brock MV, Gou M, Akiyama Y, Muller A, Wu TT, Montgomery E, et al. Prognostic importance of promoter hypermethylation of multiple genes in esophageal adenocarcinoma. Clin Cancer Res 2003;9: 2912–9.
46. Hamilton JP, Sato F, Greenwald BD, Suntharalingam M, Krasna MJ, Edelman MJ, et al. Promoter methylation and response to chemotherapy and radiation in esophageal cancer. Clin Gastroenterol Hepatol 2006;4:701–8.
47. Galipeau PC, Cowan DS, Sanchez CA, Barrett MT, Emond MJ, Levine DS, et al. 17p (p53) allelic losses, 4N (G_2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. Proc Natl Acad Sci U S A 1996;93:7081–4.
48. Fang M, Lew E, Klein M, Sebo T, Su Y, Goyal R. DNA abnormalities as marker of risk for progression of Barrett's esophagus to adenocarcinoma: image cytometric DNA analysis in formalin-fixed tissues. Am J Gastroenterol 2004;99:1887–94.
49. Reid BJ, Levine DS, Longton G, Blount PL, Rabinovitch PS. Predictors of progression to cancer in Barrett's esophagus: baseline histology and flow cytometry identify low- and high-risk patient subsets. Am J Gastroenterol 2000;95:1669–76.
50. Wu TT, Watanabe T, Heitmiller R, Zahurak M, Forastiere AA, Hamilton SR. Genetic alterations in Barrett esophagus and adenocarcinomas of the esophagus and esophagogastric junction region. Am J Pathol 1998;153:287–94.
51. Galipeau PC, Li X, Blount PL, Maley CC, Sanchez CA, Odze RD, et al. NSAIDs modulate CDKN2A, TP53, and DNA content risk for progression to esophageal adenocarcinoma. PLoS Med 2007;4:e67.
52. Casson AG, Evans SC, Gillis A, Porter GA, Veugelers P, Darnton SJ, et al. Clinical implications of p53 tumor suppressor gene mutation and protein expression in esophageal adenocarcinomas: results of a ten-year prospective study. J Thorac Cardiovasc Surg 2003;125: 1121–31.
53. Murray L, Sedo A, Scott M, McManus D, Sloan JM, Hardie LJ, et al. TP53 and progression from Barrett's metaplasia to oesophageal adenocarcinoma in a UK population cohort. Gut 2006; 55:1390–7.
54. Walch AK, Zitzelsberger HF, Bruch J, Keller G, Angermeier D, Aubele MM, et al. Chromosomal imbalances in Barrett's adenocarcinoma and the metaplasia–dysplasia–carcinoma sequence. Am J Pathol 2000;156:555–66.
55. Lai LA, Paulson TG, Li X, Sanchez CA, Maley C, Odze RD, et al. Increasing genomic instability during premalignant neoplastic progression revealed through high resolution array-CGH. Genes Chromosomes Cancer 2007;46:532–42.
56. Riegman PH, Vissers KJ, Alers JC, Geelen E, Hop WC, Tilanus HW, et al. Genomic alterations in malignant transformation of Barrett's esophagus. Cancer Res 2001;61:3164–70.

57. Li X, Galipeau PC, Sanchez CA, Blount PL, Maley CC, Arnaudo J, et al. Single nucleotide polymorphism-based genome-wide chromosome copy change, loss of heterozygosity, and aneuploidy in Barrett's esophagus neoplastic progression. Cancer Prev Res 2008;1:413–23.
58. Wiech T, Nikolopoulos E, Weis R, Langer R, Bartholome K, Timmer J, et al. Genome-wide analysis of genetic alterations in Barrett's adenocarcinoma using single nucleotide polymorphism arrays. Lab Invest 2009;89:385–97.
59. Nancarrow DJ, Handoko HY, Smithers BM, Gotley DC, Drew PA, Watson DI, et al. Genome-wide copy number analysis in esophageal adenocarcinoma using high-density single-nucleotide polymorphism arrays. Cancer Res 2008;68:4163–72.
60. Paulson TG, Maley CC, Li X, Li H, Sanchez CA, Chao DL, et al. Chromosomal instability and copy number alterations in Barrett's esophagus and esophageal adenocarcinoma. Clin Cancer Res 2009;15:3305–14.
61. Robertson EV, Jankowski JA. Genetics of gastroesophageal cancer: paradigms, paradoxes, and prognostic utility. Am J Gastroenterol 2008;103:443–9.
62. The Esophageal Adenocarcinoma Genetics C, The Wellcome Trust Case Control CSu Z, Gay LJ, Strange A, Palles C, et al. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. Nat Genet 2012;44:1131–6.
63. Wu IC, Zhao Y, Zhai R, Liu G, Ter-Minassian M, Asomaning K, et al. Association between polymorphisms in cancer-related genes and early onset of esophageal adenocarcinoma. Neoplasia 2011;13:386–92.
64. Agrawal N, Jiao Y, Bettegowda C, Hutfless SM, Wang Y, David S, et al. Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. Cancer Discov 2012;2:899–905.
65. Streppel MM, Lata S, Delabastide M, Montgomery EA, Wang JS, Canto MI, et al. Next-generation sequencing of endoscopic biopsies identifies ARID1A as a tumor-suppressor gene in Barrett's esophagus. Oncogene. 2013 Jan 14. [Epub ahead of print].
66. Wu W, Bhagat TD, Yang X, Song JH, Cheng Y, Agarwal R, et al. Hypomethylation of noncoding DNA regions and overexpression of the long noncoding RNA, AFAP1-AS1, in Barrett's esophagus and esophageal adenocarcinoma. Gastroenterology. 2013 Jan 16. [Epub ahead of print].
67. Nguyen GH, Schetter AJ, Chou DB, Bowman ED, Zhao R, Hawkes JE, et al. Inflammatory and microRNA gene expression as prognostic classifier of Barrett's-associated esophageal adenocarcinoma. Clin Cancer Res 2010;16:5824–34.
68. Majka J, Rembiasz K, Migaczewski M, Budzynski A, Ptak-Belowska A, Pabianczyk R, et al. Cyclooxygenase-2 (COX-2) is the key event in pathophysiology of Barrett's esophagus. Lesson from experimental animal model and human subjects. J Physiol Pharmacol 2010;61:409–18.
69. Shirvani VN, Ouatu-Lascar R, Kaur BS, Omary MB, Triadafilopoulos G. Cyclooxygenase 2 expression in Barrett's esophagus and adenocarcinoma: *ex vivo* induction by bile salts and acid exposure. Gastroenterology 2000;118:487–96.
70. Uefuji K, Ichikura T, Mochizuki H. Cyclooxygenase-2 expression is related to prostaglandin biosynthesis and angiogenesis in human gastric cancer. Clin Cancer Res 2000;6:135–8.
71. Kuramochi H, Vallbohmer D, Uchida K, Schneider S, Hamoui N, Shimizu D, et al. Quantitative, tissue-specific analysis of cyclooxygenase gene expression in the pathogenesis of Barrett's adenocarcinoma. J Gastrointest Surg 2004;8:1007–16.
72. Morris CD, Armstrong GR, Bigley G, Green H, Attwood SE. Cyclooxygenase-2 expression in the Barrett's metaplasia–dysplasia–adenocarcinoma sequence. Am J Gastroenterol 2001;96:990–6.
73. Lagorce C, Paraf F, Vidaud D, Couvelard A, Wendum D, Martin A, et al. Cyclooxygenase-2 is expressed frequently and early in Barrett's oesophagus and associated adenocarcinoma. Histopathology 2003;42:457–65.
74. Allameh A, Rasmi Y, Nasseri-Moghaddam S, Tavangar SM, Sharifi R, Sadreddini M. Immunohistochemical analysis of selected molecular markers in esophagus precancerous, adenocarcinoma and squamous cell carcinoma in Iranian subjects. Cancer Epidemiol 2009;33:79–84.
75. Wilson KT, Fu S, Ramanujam KS, Meltzer SJ. Increased expression of inducible nitric oxide synthase and cyclooxygenase-2 in Barrett's esophagus and associated adenocarcinomas. Cancer Res 1998;58:2929–34.
76. Prins MJ, Verhage RJ, ten Kate FJ, van Hillegersberg R. Cyclooxygenase isoenzyme-2 and vascular endothelial growth factor are associated with poor prognosis in esophageal adenocarcinoma. J Gastrointest Surg 2012;16:956–66.
77. Bhandari P, Bateman AC, Mehta RL, Stacey BS, Johnson P, Cree IA, et al. Prognostic significance of cyclooxygenase-2 (COX-2) expression in patients with surgically resectable adenocarcinoma of the oesophagus. BMC Cancer 2006;6:134.
78. Jenkins GJ, Harries K, Doak SH, Wilmes A, Griffiths AP, Baxter JN, et al. The bile acid deoxycholic acid (DCA) at neutral pH activates NF-kappaB and induces IL-8 expression in oesophageal cells *in vitro*. Carcinogenesis 2004;25:317–23.
79. O'Riordan JM, Abdel-latif MM, Ravi N, McNamara D, Byrne PJ, McDonald GS, et al. Proinflammatory cytokine and nuclear factor kappa-B expression along the inflammation–metaplasia–dysplasia–adenocarcinoma sequence in the esophagus. Am J Gastroenterol 2005;100:1257–64.
80. Abdel-Latif MM, O'Riordan J, Windle HJ, Carton E, Ravi N, Kelleher D, et al. NF-kappaB activation in esophageal adenocarcinoma: relationship to Barrett's metaplasia, survival, and response to neoadjuvant chemoradiotherapy. Ann Surg 2004;239:491–500.
81. Abdel-Latif MM, O'Riordan JM, Ravi N, Kelleher D, Reynolds JV. Activated nuclear factor-kappa B and cytokine profiles in the esophagus parallel tumor regression following neoadjuvant chemoradiotherapy. Dis Esophagus 2005;18:246–52.
82. Jenkins GJ, Mikhail J, Alhamdani A, Brown TH, Caplin S, Manson JM, et al. Immunohistochemical study of nuclear factor-kappaB activity and interleukin-8 abundance in oesophageal adenocarcinoma; a useful strategy for monitoring these biomarkers. J Clin Pathol 2007;60:1232–7.
83. Milano F, Jorritsma T, Rygiel AM, Bergman JJ, Sondermeijer C, Ten Brinke A, et al. Expression pattern of immune suppressive cytokines and growth factors in oesophageal adenocarcinoma reveal a tumour immune escape-promoting microenvironment. Scand J Immunol 2008;68:616–23.
84. Deans DA, Wigmore SJ, Gilmour H, Paterson-Brown S, Ross JA, Fearon KC. Elevated tumour interleukin-1beta is associated with systemic inflammation: a marker of reduced survival in gastro-oesophageal cancer. Br J Cancer 2006;95:1568–75.
85. Dvorak K, Chavarria M, Payne CM, Ramsey L, Crowley-Weber C, Dvorakova B, et al. Activation of the interleukin-6/STAT3 antiapoptotic pathway in esophageal cells by bile acids and low pH: relevance to barrett's esophagus. Clin Cancer Res 2007;13:5305–13.
86. Lukaszewicz-Zajac M, Mroczko B, Kozlowski M, Niklinski J, Laudanski J, Szmitkowski M. Higher importance of interleukin 6 than classic tumor markers (carcinoembryonic antigen and squamous cell cancer antigen) in the diagnosis of esophageal cancer patients. Dis Esophagus 2012;25:242–9.
87. Murray GI, Duncan ME, O'Neil P, McKay JA, Melvin WT, Fothergill JE. Matrix metalloproteinase-1 is associated with poor prognosis in oesophageal cancer. J Pathol 1998;185:256–61.
88. Salmela MT, Karjalainen-Lindsberg ML, Puolakkainen P, Saarialho-Kere U. Upregulation and differential expression of matrilysin (MMP-7) and metalloelastase (MMP-12) and their inhibitors TIMP-1 and TIMP-3 in Barrett's oesophageal adenocarcinoma. Br J Cancer 2001;85:383–92.
89. Grimm M, Lazariotou M, Kircher S, Stuermer L, Reiber C, Hofelmayr A, et al. MMP-1 is a (pre-)invasive factor in Barrett-associated esophageal adenocarcinomas and is associated with positive lymph node status. J Transl Med 2010;8:99.
90. El-Kenawy Ael M, Lotfy M, El-Kott A, El-Shahat M. Significance of matrix metalloproteinase 9 and CD34 expressions in esophageal

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

carcinoma: correlation with DNA content. J Clin Gastroenterol 2005;39:791–4.
91. Herszenyi L, Hritz I, Pregun I, Sipos F, Juhasz M, Molnar B, et al. Alterations of glutathione S-transferase and matrix metalloproteinase-9 expressions are early events in esophageal carcinogenesis. World J Gastroenterol 2007;13:676–82.
92. Mroczko B, Kozlowski M, Groblewska M, Lukaszewicz M, Niklinski J, Laudanski J, et al. Expression of matrix metalloproteinase-9 in the neoplastic and interstitial inflammatory infiltrate cells in the different histopathological types of esophageal cancer. Folia Histochem Cytobiol 2008;46:471–8.
93. Darnton SJ, Hardie LJ, Muc RS, Wild CP, Casson AG. Tissue inhibitor of metalloproteinase-3 (TIMP-3) gene is methylated in the development of esophageal adenocarcinoma: loss of expression correlates with poor prognosis. Int J Cancer 2005;115:351–8.
94. Bani-Hani K, Martin IG, Hardie LJ, Mapstone N, Briggs JA, Forman D, et al. Prospective study of cyclin D1 overexpression in Barrett's esophagus: association with increased risk of adenocarcinoma. J Natl Cancer Inst 2000;92:1316–21.
95. Shi XY, Bhagwandeen B, Leong AS. p16, cyclin D1, Ki-67, and AMACR as markers for dysplasia in Barrett esophagus. Appl Immunohistochem Mol Morphol 2008;16:447–52.
96. Brien TP, Odze RD, Sheehan CE, McKenna BJ, Ross JS. HER-2/neu gene amplification by FISH predicts poor survival in Barrett's esophagus-associated adenocarcinoma. Hum Pathol 2000;31:35–9.
97. Wang KL, Wu TT, Choi IS, Wang H, Resetkova E, Correa AM, et al. Expression of epidermal growth factor receptor in esophageal and esophagogastric junction adenocarcinomas: association with poor outcome. Cancer 2007;109:658–67.
98. Raouf AA, Evoy DA, Carton E, Mulligan E, Griffin MM, Reynolds JV. Loss of Bcl-2 expression in Barrett's dysplasia and adenocarcinoma is associated with tumor progression and worse survival but not with response to neoadjuvant chemoradiation. Dis Esophagus 2003;16:17–23.
99. Auvinen MI, Sihvo EI, Ruohtula T, Salminen JT, Koivistoinen A, Siivola P, et al. Incipient angiogenesis in Barrett's epithelium and lymphangiogenesis in Barrett's adenocarcinoma. J Clin Oncol 2002;20:2971–9.
100. Mobius C, Stein HJ, Becker I, Feith M, Theisen J, Gais P, et al. Vascular endothelial growth factor expression and neovascularization in Barrett's carcinoma. World J Surg 2004;28:675–9.
101. Kerkhof M, Kusters JG, van Dekken H, Kuipers EJ, Siersema PD. Biomarkers for risk stratification of neoplastic progression in Barrett esophagus. Cell Oncol 2007;29:507–17.
102. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 1993;75:843–54.
103. Metias SM, Lianidou E, Yousef GM. MicroRNAs in clinical oncology: at the crossroads between promises and problems. J Clin Pathol 2009;62:771–6.
104. Feber A, Xi L, Luketich JD, Pennathur A, Landreneau RJ, Wu M, et al. MicroRNA expression profiles of esophageal cancer. J Thorac Cardiovasc Surg 2008;135:255–60.
105. Smith CM, Watson DI, Leong MP, Mayne GC, Michael MZ, Wijnhoven BP, et al. miR-200 family expression is downregulated upon neoplastic progression of Barrett's esophagus. World J Gastroenterol 2011;17:1036–44.
106. Luthra R, Singh RR, Luthra MG, Li YX, Hannah C, Romans AM, et al. MicroRNA-196a targets annexin A1: a microRNA-mediated mechanism of annexin A1 downregulation in cancers. Oncogene 2008;27:6667–78.
107. Maru DM, Singh RR, Hannah C, Albarracin CT, Li YX, Abraham R, et al. MicroRNA-196a is a potential marker of progression during Barrett's metaplasia–dysplasia–invasive adenocarcinoma sequence in esophagus. Am J Pathol 2009;174:1940–8.
108. Braun CJ, Zhang X, Savelyeva I, Wolff S, Moll UM, Schepeler T, et al. p53-Responsive microRNAs 192 and 215 are capable of inducing cell cycle arrest. Cancer Res 2008;68:10094–104.
109. Wijnhoven BP, Hussey DJ, Watson DI, Tsykin A, Smith CM, Michael MZ, et al. MicroRNA profiling of Barrett's oesophagus and oesophageal adenocarcinoma. Br J Surg 2010;97:853–61.
110. Bansal A, Lee IH, Hong X, Anand V, Mathur SC, Gaddam S, et al. Feasibility of microRNAs as biomarkers for Barrett's Esophagus progression: a pilot cross-sectional, phase 2 biomarker study. Am J Gastroenterol 2011;106:1055–63.
111. Mathe EA, Nguyen GH, Bowman ED, Zhao Y, Budhu A, Schetter AJ, et al. MicroRNA expression in squamous cell carcinoma and adenocarcinoma of the esophagus: associations with survival. Clin Cancer Res 2009;15:6192–200.
112. Yang H, Gu J, Wang KK, Zhang W, Xing J, Chen Z, et al. MicroRNA expression signatures in Barrett's esophagus and esophageal adenocarcinoma. Clin Cancer Res 2009;15:5744–52.
113. Mees ST, Mardin WA, Wendel C, Baeumer N, Willscher E, Senninger N, et al. EP300—a miRNA-regulated metastasis suppressor gene in ductal adenocarcinomas of the pancreas. Int J Cancer 2010;126:114–24.
114. Fassan M, Volinia S, Palatini J, Pizzi M, Baffa R, De Bernard M, et al. MicroRNA expression profiling in human Barrett's carcinogenesis. Int J Cancer 2011;129:1661–70.
115. D'Souza-Schorey C, Clancy JW. Tumor-derived microvesicles: shedding light on novel microenvironment modulators and prospective cancer biomarkers. Genes Dev 2012;26:1287–99.
116. Zen K, Zhang CY. Circulating microRNAs: a novel class of biomarkers to diagnose and monitor human cancers. Med Res Rev 2012;32:326–48.
117. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol 2012;13:448–62.
118. Kim EH, Misek DE. Glycoproteomics-based identification of cancer biomarkers. Int J Proteomics 2011;2011:601937.
119. Fanayan S, Hincapie M, Hancock WS. Using lectins to harvest the plasma/serum glycoproteome. Electrophoresis 2012;33:1746–54.
120. Ongay S, Boichenko A, Govorukhina N, Bischoff R. Glycopeptide enrichment and separation for protein glycosylation analysis. J Sep Sci 2012;35:2341–72.
121. Shimamoto C, Weinstein WM, Boland CR. Glycoconjugate expression in normal, metaplastic, and neoplastic human upper gastrointestinal mucosa. J Clin Invest 1987;80:1670–8.
122. Poorkhalkali N, Jacobson I, Helander HF. Lectin histochemistry of the esophagus in several mammalian species. Anat Embryol (Berl) 1999;200:541–9.
123. Neumann H, Wex T, Monkemuller K, Vieth M, Fry LC, Malfertheiner P. Lectin UEA-I–binding proteins are specifically increased in the squamous epithelium of patients with Barrett's esophagus. Digestion 2008;78:201–7.
124. Bird-Lieberman EL, Neves AA, Lao-Sirieix P, O'Donovan M, Novelli M, Lovat LB, et al. Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Barrett's esophagus. Nat Med 2012;18:315–21.
125. Bird-Lieberman EL, Dunn JM, Coleman HG, Lao-Sirieix P, Oukrif D, Moore CE, et al. Population-based study reveals new risk-stratification biomarker panel for barrett's esophagus. Gastroenterology 2012;143:927–35.
126. Byrne AM, Sharma R, Duggan G, Kelleher D, Long A. Deoxycholic acid impairs glycosylation and fucosylation processes in esophageal epithelial cells. Glycobiology 2012;22:638–48.
127. Nancarrow DJ, Clouston AD, Smithers BM, Gotley DC, Drew PA, Watson DI, et al. Whole genome expression array profiling highlights differences in mucosal defense genes in Barrett's esophagus and esophageal adenocarcinoma. PLoS ONE 2011;6:e22513.
128. Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, et al. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. Clin Cancer Res 2004;10:6897–904.
129. Devriese LA, Voest EE, Beijnen JH, Schellens JH. Circulating tumor cells as pharmacodynamic biomarker in early clinical oncological trials. Cancer Treat Rev 2011;37:579–89.

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

130. Zhai R, Zhao Y, Su L, Cassidy L, Liu G, Christiani DC. Genome-wide DNA methylation profiling of cell-free serum DNA in esophageal adenocarcinoma and Barrett esophagus. Neoplasia 2012;14:29–33.
131. Kawakami K, Brabender J, Lord RV, Groshen S, Greenwald BD, Krasna MJ, et al. Hypermethylated APC DNA in plasma and prognosis of patients with esophageal adenocarcinoma. J Natl Cancer Inst 2000;92:1805–11.
132. Hoffmann AC, Vallbohmer D, Prenzel K, Metzger R, Heitmann M, Neiss S, et al. Methylated DAPK and APC promoter DNA detection in peripheral blood is significantly associated with apparent residual tumor and outcome. J Cancer Res Clin Oncol 2009;135: 1231–7.
133. Chen ST, Pan TL, Juan HF, Chen TY, Lin YS, Huang CM. Breast tumor microenvironment: proteomics highlights the treatments targeting secretome. J Proteome Res 2008;7:1379–87.
134. Pitteri SJ, Kelly-Spratt KS, Gurley KE, Kennedy J, Buson TB, Chin A, et al. Tumor microenvironment-derived proteins dominate the plasma proteome response during breast cancer induction and progression. Cancer Res 2011;71:5090–100.
135. Mechref Y, Hussein A, Bekesova S, Pungpapong V, Zhang M, Dobrolecki LE, et al. Quantitative serum glycomics of esophageal adenocarcinoma and other esophageal disease onsets. J Proteome Res 2009;8:2656–66.
136. Mitra I, Zhuang Z, Zhang Y, Yu CY, Hammoud ZT, Tang H, et al. N-glycan profiling by microchip electrophoresis to differentiate disease states related to esophageal adenocarcinoma. Anal Chem 2012;84: 3621–7.
137. Gaye MM, Valentine SJ, Hu Y, Mirjankar N, Hammoud ZT, Mechref Y, et al. Ion mobility-mass spectrometry analysis of serum N-linked glycans from esophageal adenocarcinoma phenotypes. J Proteome Res 2012;11:6102–10.
138. Mann B, Madera M, Klouckova I, Mechref Y, Dobrolecki LE, Hickey RJ, et al. A quantitative investigation of fucosylated serum glycoproteins with application to esophageal adenocarcinoma. Electrophoresis 2010;31:1833–41.
139. Choi E, Loo D, Dennis JW, O'Leary CA, Hill MM. High-throughput lectin magnetic bead array-coupled tandem mass spectrometry for glycoprotein biomarker discovery. Electrophoresis 2011;32: 3564–75.
140. Loo D, Jones A, Hill MM. Lectin magnetic bead array for biomarker discovery. J Proteome Res 2010;9:5496–500.
141. Yakoub D, Keun HC, Goldin R, Hanna GB. Metabolic profiling detects field effects in nondysplastic tissue from esophageal cancer patients. Cancer Res 2010;70:9129–36.
142. Djukovic D, Baniasadi HR, Kc R, Hammoud Z, Raftery D. Targeted serum metabolite profiling of nucleosides in esophageal adenocarcinoma. Rapid Commun Mass Spectrom 2010;24:3057–62.
143. Zhang J, Bowers J, Liu L, Wei S, Gowda GA, Hammoud Z, et al. Esophageal cancer metabolite biomarkers detected by LC–MS and NMR methods. PLoS ONE 2012;7:e30181.
144. Zhang J, Liu L, Wei S, Nagana Gowda GA, Hammoud Z, Kesler KA, et al. Metabolomics study of esophageal adenocarcinoma. J Thorac Cardiovasc Surg 2011;141:469–75.
145. Davis VW, Schiller DE, Eurich D, Sawyer MB. Urinary metabolomic signature of esophageal cancer and Barrett's esophagus. World J Surg Oncol 2012;10:271.
146. Hardie LJ, Darnton SJ, Wallis YL, Chauhan A, Hainaut P, Wild CP, et al. p16 expression in Barrett's esophagus and esophageal adenocarcinoma: association with genetic and epigenetic alterations. Cancer Lett 2005;217:221–30.
147. Bian YS, Osterheld MC, Fontolliet C, Bosman FT, Benhattar J. p16 inactivation by methylation of the CDKN2A promoter occurs early during neoplastic progression in Barrett's esophagus. Gastroenterology 2002;122:1113–21.
148. Wang JS, Guo M, Montgomery EA, Thompson RE, Cosby H, Hicks L, et al. DNA promoter hypermethylation of p16 and APC predicts neoplastic progression in Barrett's esophagus. Am J Gastroenterol 2009;104:2153–60.
149. Schulmann K, Sterian A, Berki A, Yin J, Sato F, Xu Y, et al. Inactivation of p16, RUNX3, and HPP1 occurs early in Barrett's-associated neoplastic progression and predicts progression risk. Oncogene 2005;24:4138–48.
150. Wong DJ, Barrett MT, Stoger R, Emond MJ, Reid BJ. p16INK4a promoter is hypermethylated at a high frequency in esophageal adenocarcinomas. Cancer Res 1997;57:2619–22.
151. Wong DJ, Paulson TG, Prevo LJ, Galipeau PC, Longton G, Blount PL, et al. p16(INK4a) lesions are common, early abnormalities that undergo clonal expansion in Barrett's metaplastic epithelium. Cancer Res 2001;61:8284–9.
152. Vieth M, Schneider-Stock R, Rohrich K, May A, Ell C, Markwarth A, et al. INK4a-ARF alterations in Barrett's epithelium, intraepithelial neoplasia and Barrett's adenocarcinoma. Virchows Arch 2004;445: 135–41.
153. Kempster S, Phillips WA, Baindur-Hudson S, Thomas RJ, Dow C, Rockman SP. Methylation of exon 2 of p16 is associated with late stage oesophageal cancer. Cancer Lett 2000;150:57–62.
154. Baumann S, Keller G, Puhringer F, Napieralski R, Feith M, Langer R, et al. The prognostic impact of O6-methylguanine-DNA methyltransferase (MGMT) promotor hypermethylation in esophageal adenocarcinoma. Int J Cancer 2006;119:264–8.
155. Kuester D, El-Rifai W, Peng D, Ruemmele P, Kroeckel I, Peters B, et al. Silencing of MGMT expression by promoter hypermethylation in the metaplasia—dysplasia–carcinoma sequence of Barrett's esophagus. Cancer Lett 2009;275:117–26.
156. Clement G, Braunschweig R, Pasquier N, Bosman FT, Benhattar J. Alterations of the Wnt signaling pathway during the neoplastic progression of Barrett's esophagus. Oncogene 2006;25:3084–92.
157. Clement G, Bosman FT, Fontolliet C, Benhattar J. Monoallelic methylation of the APC promoter is altered in normal gastric mucosa associated with neoplastic lesions. Cancer Res 2004;64:6867–73.
158. Peng DF, Razvi M, Chen H, Washington K, Roessner A, Schneider-Stock R, et al. DNA hypermethylation regulates the expression of members of the Mu-class glutathione S-transferases and glutathione peroxidases in Barrett's adenocarcinoma. Gut 2009;58:5–15.
159. Lee OJ, Schneider-Stock R, McChesney PA, Kuester D, Roessner A, Vieth M, et al. Hypermethylation and loss of expression of glutathione peroxidase-3 in Barrett's tumorigenesis. Neoplasia 2005;7:854–61.
160. Gu P, Xing X, Tanzer M, Rocken C, Weichert W, Ivanauskas A, et al. Frequent loss of TIMP-3 expression in progression of esophageal and gastric adenocarcinomas. Neoplasia 2008;10:563–72.
161. Kuester D, Dar AA, Moskaluk CC, Krueger S, Meyer F, Hartig R, et al. Early involvement of death-associated protein kinase promoter hypermethylation in the carcinogenesis of Barrett's esophageal adenocarcinoma and its association with clinical progression. Neoplasia 2007;9:236–45.
162. Jin Z, Olaru A, Yang J, Sato F, Cheng Y, Kan T, et al. Hypermethylation of tachykinin-1 is a potential biomarker in human esophageal cancer. Clin Cancer Res 2007;13:6293–300.
163. Hamilton JP, Sato F, Jin Z, Greenwald BD, Ito T, Mori Y, et al. Reprimo methylation is a potential biomarker of Barrett's-associated esophageal neoplastic progression. Clin Cancer Res 2006;12:6637–42.
164. Corn PG, Heath EI, Heitmiller R, Fogt F, Forastiere AA, Herman JG, et al. Frequent hypermethylation of the 5′ CpG island of E-cadherin in esophageal adenocarcinoma. Clin Cancer Res 2001;7:2765–9.
165. Tischoff I, Hengge UR, Vieth M, Ell C, Stolte M, Weber A, et al. Methylation of SOCS-3 and SOCS-1 in the carcinogenesis of Barrett's adenocarcinoma. Gut 2007;56:1047–53.
166. Zou H, Molina JR, Harrington JJ, Osborn NK, Klatt KK, Romero Y, et al. Aberrant methylation of secreted frizzled-related protein genes in esophageal adenocarcinoma and Barrett's esophagus. Int J Cancer 2005;116:584–91.
167. Kaz AM, Luo Y, Dzieciatkowski S, Chak A, Willis JE, Upton MP, et al. Aberrantly methylated PKP1 in the progression of Barrett's esophagus to esophageal adenocarcinoma. Genes Chromosomes Cancer 2012;51:384–93.
168. Guo M, House MG, Akiyama Y, Qi Y, Capagna D, Harmon J, et al. Hypermethylation of the GATA gene family in esophageal cancer. Int J Cancer 2006;119:2078–83.
169. Jin Z, Cheng Y, Olaru A, Kan T, Yang J, Paun B, et al. Promoter hypermethylation of CDH13 is a common, early event in human

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

esophageal adenocarcinogenesis and correlates with clinical risk factors. Int J Cancer 2008;123:2331–6.
170. Jin Z, Mori Y, Yang J, Sato F, Ito T, Cheng Y, et al. Hypermethylation of the nel-like 1 gene is a common and early event and is associated with poor prognosis in early-stage esophageal adenocarcinoma. Oncogene 2007;26:6332–40.
171. Zou H, Osborn NK, Harrington JJ, Klatt KK, Molina JR, Burgart LJ, et al. Frequent methylation of eyes absent 4 gene in Barrett's esophagus and esophageal adenocarcinoma. Cancer Epidemiol Biomarkers Prev 2005;14:830–4.
172. Jin Z, Hamilton JP, Yang J, Mori Y, Olaru A, Sato F, et al. Hypermethylation of the AKAP12 promoter is a biomarker of Barrett's-associated esophageal neoplastic progression. Cancer Epidemiol Biomarkers Prev 2008;17:111–7.
173. Moinova H, Leidner RS, Ravi L, Lutterbaugh J, Barnholtz-Sloan JS, Chen Y, et al. Aberrant vimentin methylation is characteristic of upper gastrointestinal pathologies. Cancer Epidemiol Biomarkers Prev 2012;21:594–600.
174. Clement G, Guilleret I, He B, Yagui-Beltran A, Lin YC, You L, et al. Epigenetic alteration of the Wnt inhibitory factor-1 promoter occurs early in the carcinogenesis of Barrett's esophagus. Cancer Sci 2008;99:46–53.
175. Soutto M, Peng D, Razvi M, Ruemmele P, Hartmann A, Roessner A, et al. Epigenetic and genetic silencing of CHFR in esophageal adenocarcinomas. Cancer 2010;116:4033–42.
176. Peng D, Hu TL, Jiang A, Washington MK, Moskaluk CA, Schneider-Stock R, et al. Location-specific epigenetic regulation of the metallothionein 3 gene in esophageal adenocarcinomas. PLoS ONE 2011;6:e22009.
177. Smith E, De Young NJ, Pavey SJ, Hayward NK, Nancarrow DJ, Whiteman DC, et al. Similarity of aberrant DNA methylation in Barrett's esophagus and esophageal adenocarcinoma. Mol Cancer 2008;7:75.
178. Kaz AM, Wong CJ, Luo Y, Virgin JB, Washington MK, Willis JE, et al. DNA methylation profiling in Barrett's esophagus and esophageal adenocarcinoma reveals unique methylation signatures and molecular subclasses. Epigenetics 2011;6:1403–12.
179. Alvi MA, Liu X, O'Donovan M, Newton R, Wernisch L, Shannon NB, et al. DNA methylation as an adjunct to histopathology to detect prevalent, inconspicuous dysplasia and early-stage neoplasia in Barrett's esophagus. Clin Cancer Res 2013;19:878–88.
180. Clement G, Braunschweig R, Pasquier N, Bosman FT, Benhattar J. Methylation of APC, TIMP3, and TERT: a new predictive marker to distinguish Barrett's oesophagus patients at risk for malignant transformation. J Pathol 2006;208:100–7.
181. van Baal JW, Milano F, Rygiel AM, Bergman JJ, Rosmolen WD, van Deventer SJ, et al. A comparative analysis by SAGE of gene expression profiles of Barrett's esophagus, normal squamous esophagus, and gastric cardia. Gastroenterology 2005;129:1274–81.
182. Sabo E, Meitner PA, Tavares R, Corless CL, Lauwers GY, Moss SF, et al. Expression analysis of Barrett's esophagus-associated high-grade dysplasia in laser capture microdissected archival tissue. Clin Cancer Res 2008;14:6440–8.
183. Peters CJ, Rees JR, Hardwick RH, Hardwick JS, Vowler SL, Ong CA, et al. A 4-gene signature predicts survival of patients with resected adenocarcinoma of the esophagus, junction, and gastric cardia. Gastroenterology 2010;139:1995–2004.
184. Ostrowski J, Mikula M, Karczmarski J, Rubel T, Wyrwicz LS, Bragoszewski P, et al. Molecular defense mechanisms of Barrett's metaplasia estimated by an integrative genomics. J Mol Med (Berl) 2007;85:733–43.
185. Lagarde SM, Ver Loren van Themaat PE, Moerland PD, Gilhuijs-Pederson LA, Ten Kate FJ, Reitsma PH, et al. Analysis of gene expression identifies differentially expressed genes and pathways associated with lymphatic dissemination in patients with adenocarcinoma of the esophagus. Ann Surg Oncol 2008;15:3459–70.
186. van Dekken H, Vissers K, Tilanus HW, Kuo WL, Tanke HJ, Rosenberg C, et al. Genomic array and expression analysis of frequent high-level amplifications in adenocarcinomas of the gastro-esophageal junction. Cancer Genet Cytogenet 2006;166:157–62.
187. Akagi T, Ito T, Kato M, Jin Z, Cheng Y, Kan T, et al. Chromosomal abnormalities and novel disease-related regions in progression from Barrett's esophagus to esophageal adenocarcinoma. Int J Cancer 2009;125:2349–59.
188. Greenawalt DM, Duong C, Smyth GK, Ciavarella ML, Thompson NJ, Tiang T, et al. Gene expression profiling of esophageal cancer: comparative analysis of Barrett's esophagus, adenocarcinoma, and squamous cell carcinoma. Int J Cancer 2007;120:1914–21.
189. Gu J, Ajani JA, Hawk ET, Ye Y, Lee JH, Bhutani MS, et al. Genome-wide catalogue of chromosomal aberrations in barrett's esophagus and esophageal adenocarcinoma: a high-density single nucleotide polymorphism array analysis. Cancer Prev Res 2010;3:1176–86.
190. Kim SM, Park YY, Park ES, Cho JY, Izzo JG, Zhang D, et al. Prognostic biomarkers for esophageal adenocarcinoma identified by analysis of tumor transcriptome. PLoS ONE 2010;5:e15074.
191. van Dekken H, Tilanus HW, Hop WC, Dinjens WN, Wink JC, Vissers KJ, et al. Array comparative genomic hybridization, expression array, and protein analysis of critical regions on chromosome arms 1q, 7q, and 8p in adenocarcinomas of the gastroesophageal junction. Cancer Genet Cytogenet 2009;189:37–42.
192. Lai LA, Kostadinov R, Barrett MT, Peiffer DA, Pokholok D, Odze R, et al. Deletion at fragile sites is a common and early event in Barrett's esophagus. Mol Cancer Res 2010;8:1084–94.
193. van Duin M, van Marion R, Vissers KJ, Hop WC, Dinjens WN, Tilanus HW, et al. High-resolution array comparative genomic hybridization of chromosome 8q: evaluation of putative progression markers for gastroesophageal junction adenocarcinomas. Cytogenet Genome Res 2007;118:130–7.
194. Xu Y, Selaru FM, Yin J, Zou TT, Shustova V, Mori Y, et al. Artificial neural networks and gene filtering distinguish between global gene expression profiles of Barrett's esophagus and esophageal cancer. Cancer Res 2002;62:3493–7.
195. Kimchi ET, Posner MC, Park JO, Darga TE, Kocherginsky M, Karrison T, et al. Progression of Barrett's metaplasia to adenocarcinoma is associated with the suppression of the transcriptional programs of epidermal differentiation. Cancer Res 2005;65:3146–54.
196. Barrett MT, Yeung KY, Ruzzo WL, Hsu L, Blount PL, Sullivan R, et al. Transcriptional analyses of Barrett's metaplasia and normal upper GI mucosae. Neoplasia 2002;4:121–8.
197. Wang S, Zhan M, Yin J, Abraham JM, Mori Y, Sato F, et al. Transcriptional profiling suggests that Barrett's metaplasia is an early intermediate stage in esophageal adenocarcinogenesis. Oncogene 2006;25:3346–56.
198. Hammoud ZT, Badve S, Zhao Q, Li L, Saxena R, Thorat MA, et al. Differential gene expression profiling of esophageal adenocarcinoma. J Thorac Cardiovasc Surg 2009;137:829–34.
199. Gomes LI, Esteves GH, Carvalho AF, Cristo EB, Hirata R Jr, Martins WK, et al. Expression profile of malignant and nonmalignant lesions of esophagus and stomach: differential activity of functional modules related to inflammation and lipid metabolism. Cancer Res 2005;65:7127–36.
200. Hao Y, Triadafilopoulos G, Sahbaie P, Young HS, Omary MB, Lowe AW. Gene expression profiling reveals stromal genes expressed in common between Barrett's esophagus and adenocarcinoma. Gastroenterology 2006;131:925–33.
201. Brabender J, Marjoram P, Salonga D, Metzger R, Schneider PM, Park JM, et al. A multigene expression panel for the molecular diagnosis of Barrett's esophagus and Barrett's adenocarcinoma of the esophagus. Oncogene 2004;23:4780–8.
202. Lao-Sirieix P, Boussioutas A, Kadri SR, O'Donovan M, Debiram I, Das M, et al. Non-endoscopic screening biomarkers for Barrett's oesophagus: from microarray analysis to the clinic. Gut 2009;58:1451–9.
203. Goh XY, Rees JR, Paterson AL, Chin SF, Marioni JC, Save V, et al. Integrative analysis of array-comparative genomic hybridisation and matched gene expression profiling data reveals novel genes with prognostic significance in oesophageal adenocarcinoma. Gut 2011;60:1317–26.
204. Rao S, Welsh L, Cunningham D, te-Poele RH, Benson M, Norman A, et al. Correlation of overall survival with gene expression profiles in a

Published OnlineFirst April 10, 2013; DOI: 10.1158/1055-9965.EPI-12-1415

prospective study of resectable esophageal cancer. Clin Colorectal Cancer 2011;10:48–56.

205. Schauer M, Janssen KP, Rimkus C, Raggi M, Feith M, Friess H, et al. Microarray-based response prediction in esophageal adenocarcinoma. Clin Cancer Res 2010;16:330–7.

206. Luthra MG, Ajani JA, Izzo J, Ensor J, Wu TT, Rashid A, et al. Decreased expression of gene cluster at chromosome 1q21 defines molecular subgroups of chemoradiotherapy response in esophageal cancers. Clin Cancer Res 2007;13:912–9.

207. Luthra R, Wu TT, Luthra MG, Izzo J, Lopez-Alvarez E, Zhang L, et al. Gene expression profiling of localized esophageal carcinomas: association with pathologic response to preoperative chemoradiation. J Clin Oncol 2006;24:259–67.

208. Wu X, Gu J, Wu TT, Swisher SG, Liao Z, Correa AM, et al. Genetic variations in radiation and chemotherapy drug action pathways predict clinical outcomes in esophageal cancer. J Clin Oncol 2006;24:3789–98.

209. Luzna P, Gregar J, Uberall I, Radova L, Prochazka V, Ehrmann J Jr. Changes of microRNAs-192, 196a and 203 correlate with Barrett's esophagus diagnosis and its progression compared to normal healthy individuals. Diagn Pathol 2011;6:114.

210. Leidner RS, Ravi L, Leahy P, Chen Y, Bednarchik B, Streppel M, et al. The microRNAs, MiR-31 and MiR-375, as candidate markers in Barrett's esophageal carcinogenesis. Genes Chromosomes Cancer 2012;51:473–9.

211. Feber A, Xi L, Pennathur A, Gooding WE, Bandla S, Wu M, et al. MicroRNA prognostic signature for nodal metastases and survival in esophageal adenocarcinoma. Ann Thorac Surg 2011;91:1523–30.

212. Revilla-Nuin B, Parrilla P, Lozano J, Martinez de Haro LF, Ortiz A, Martinez C, et al. Predictive value of MicroRNAs in the progression of Barrett esophagus to adenocarcinoma in a long-term follow-up study. Ann Surg 2013;257:886–93.

213. Hong L, Han Y, Zhang H, Li M, Gong T, Sun L, et al. The prognostic and chemotherapeutic value of miR-296 in esophageal squamous cell carcinoma. Ann Surg 2010;251:1056–63.

214. Hummel R, Hussey DJ, Michael MZ, Haier J, Bruewer M, Senninger N, et al. MiRNAs and their association with locoregional staging and survival following surgery for esophageal carcinoma. Ann Surg Oncol 2011;18:253–60.

215. Hu Y, Correa AM, Hoque A, Guan B, Ye F, Huang J, et al. Prognostic significance of differentially expressed miRNAs in esophageal cancer. Int J Cancer 2011;128:132–43.

9.2 Appendix II: Research article entitled "Serum glycoprotein biomarker discovery and qualification pipeline reveals novel diagnostic biomarker candidates for esophageal adenocarcinoma"

MCP Papers in Press. Published on September 24, 2015 as Manuscript M115.050922

Serum glycoprotein biomarker discovery and qualification pipeline reveals novel diagnostic biomarker candidates for esophageal adenocarcinoma

Alok K. Shah[1], Kim-Anh Lê Cao[1], Eunju Choi[1,2†], David Chen[3], Benoît Gautier[1], Derek Nancarrow[4‡], David C. Whiteman[4], Nicholas A. Saunders[1], Andrew P. Barbour[5], Virendra Joshi[6] and Michelle M. Hill[1*]

[1]The University of Queensland Diamantina Institute, The University of Queensland, Translational Research Institute, Brisbane, Queensland, Australia

[2]School of Veterinary Science, The University of Queensland, Gatton, Queensland, Australia

[3]School of Information and Communication Technology, Griffith University, Brisbane, Queensland, Australia

[4]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

[5]School of Medicine, The University of Queensland, Brisbane, Queensland, Australia

[6]Ochsner Health System, Gastroenterology, New Orleans, LA, United States

Current addresses: [†]Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA. [‡]Department of Surgery, University of Michigan, Ann Arbor, Michigan, MI, USA.

***Correspondence:**

Michelle M. Hill

The University of Queensland Diamantina Institute

Level 5, Translational Research Institute, 37 Kent Street, Woolloongabba, QLD AUSTRALIA 4102

Tel: +61 (0)7 3443 7049 Fax: +61 (0)7 3443 5946

E-mail: m.hill2@uq.edu.au

Running title: Serum glycoprotein biomarker discovery pipeline

Abbreviations

A1BG	Alpha-1B-glycoprotein
A2M	Alpha-2-macroglobulin
AAL	*Aleuria aurantia* lectin
APOB	Apolipoprotein B-100
AUROC	Area under receiver operating characteristic curve
BE	Barrett's esophagus
BMI	Body mass index
BPL	*Bauhinia purpurea* lectin
C2	Complement C2
C4B	Complement C4-B
C9	Complement component C9
ConA	Concanavalin A from *Canavalia ensiformis*
CV	Co-efficient of variation
DSA	*Datura stramonium* agglutinin
EAC	Esophageal adenocarcinoma
ECA	*Erythrina cristagalli* agglutinin
EPHA	Erythroagglutinin *Phaseolus vulgaris*
GalNAc	N-acetylgalactosamine
GERD	Gastroesophageal reflux disease
GlcNAc	N-acetylglucosamine
GNL	*Galanthus nivalis* lectin
GSN	Gelsolin
HAA	*Helix aspersa* agglutinin

HGD	High grade dysplasia
HP	Haptoglobin
HPA	*Helix pomatia* agglutinin
JAC	Jacalin from *Artocarpus integrifolia*
LeMBA	Lectin magnetic bead array
LPHA	Leukoagglutinating phytohemagglutinin
MAA	*Maackia amurensis* agglutinin
NPL	*Narcissus pseudonarcissus* lectin
PSA	*Pisum sativum* agglutinin
SBA	Soybean agglutinin
SERPING1	Plasma protease C1 inhibitor
SIS	Stable isotope standard
SNA	*Sambucus nigra* agglutinin
sPLS-DA	Sparse partial least squares-discriminant analysis
STL	*Solanum tuberosum* lectin
UEA	*Ulex europeus* agglutinin-I
WFA	*Wisteria floribunda* agglutinin
WGA	Wheat germ agglutinin

Summary

We report an integrated pipeline for efficient serum glycoprotein biomarker candidate discovery and qualification that may be used to facilitate cancer diagnosis and management. The discovery phase used semi-automated lectin magnetic bead array (LeMBA)-coupled tandem mass spectrometry with a dedicated data-housing and analysis pipeline; GlycoSelector (http://glycoselector.di.uq.edu.au). The qualification phase used LeMBA-multiple reaction monitoring-mass spectrometry incorporating an interactive web-interface, Shiny mixOmics (http://mixomics-projects.di.uq.edu.au/Shiny), for univariate and multivariate statistical analysis. Relative quantitation was performed by referencing to a spiked-in glycoprotein, chicken ovalbumin. We applied this workflow to identify diagnostic biomarkers for esophageal adenocarcinoma (EAC), a life threatening malignancy with poor prognosis in the advanced setting. EAC develops from metaplastic condition Barrett's esophagus (BE). Currently diagnosis and monitoring of at-risk patients is through endoscopy and biopsy which is expensive and requires hospital admission. Hence there is a clinical need for a noninvasive diagnostic biomarker of EAC. In total 89 patient samples from healthy controls, and patients with BE or EAC were screened in discovery and qualification stages. Of the 246 glycoforms measured in the qualification stage, 40 glycoforms (as measured by lectin affinity) qualified as candidate serum markers. The top candidate for distinguishing healthy from BE patients' group was *Narcissus pseudonarcissus* lectin (NPL)-reactive Apolipoprotein B-100 (*P* value=0.0231; AUROC=0.71); BE vs EAC, *Aleuria aurantia* lectin (AAL)-reactive complement component C9 (*P* value=0.0001; AUROC=0.85); healthy vs EAC, Erythroagglutinin *Phaseolus vulgaris* (EPHA)-reactive gelsolin (*P* value=0.0014; AUROC=0.80). A panel of 8 glycoforms showed an improved AUROC of 0.94 to discriminate EAC from BE. Two biomarker candidates were independently verified by LeMBA-immunoblotting, confirming the validity of the relative quantitation approach. Thus,

we have identified candidate biomarkers which, following large-scale clinical evaluation, can be developed into diagnostic blood tests. A key feature of the pipeline is the potential for rapid translation of the candidate biomarkers to lectin-immunoassays.

Introduction

Biomarkers play a central role in health care by enabling accurate diagnosis and prognosis; hence there is extensive research on the identification and development of novel biomarkers. However, despite numerous biomarker publications over the years (1), only a handful of new cancer biomarkers have successfully completed the journey from discovery, qualification, to verification and validation (2-4). One possible way to overcome this challenge is to develop an integrated biomarker pipeline that facilitates the smooth and successful transition from discovery to validation (5-10). The first and foremost consideration in an integrated pipeline is the sample source. In general, most of the proteomics based workflows use tissues or proximal fluids during the discovery phase, with the goal of extending the findings to plasma. Although this approach avoid the high complexity serum/plasma proteome and the associated requisite multi-dimensional sample separation in discovery stages, it often leads to failure when the candidates are not detected in plasma due to the limited sensitivity of the available analytical methods, or the absence of candidates in the plasma (11). To overcome this pitfall, we have developed an integrated glycoprotein biomarker pipeline which can simply and rapidly isolate glycosylated proteins from serum to enable high throughput analysis of differentially glycosylated proteins in discovery and qualification stages.

The workflow utilizes naturally occurring glycan binding proteins, lectins, in a semi-automated high throughput workflow called lectin magnetic bead array-tandem mass spectrometry (LeMBA-MS/MS) (12, 13). Although lectins have been well-utilized in glycobiology and biomarker discovery (14-17), the LeMBA-MS/MS workflow demonstrates several unique features. Firstly, serum glycoproteins are isolated in a single-step using 20 individual lectin-coated magnetic beads in microplate format. Secondly, we have optimized the concentrations of salts and detergents for sample denaturation to avoid co-isolation of

protein complexes without adversely affecting lectin pull-down efficiency. Thirdly, a liquid handler is used for sample processing to facilitate high-throughput screening and increase reproducibility. In addition, we have optimized on-bead trypsin digestion and incorporated lectin-exclusion lists during nano-LC-MS/MS to identify non-glycosylated peptides from the isolated glycoproteins. With these innovations, LeMBA-MS/MS demonstrates nanomolar sensitivity and linearity, and applicability across species (12). Compared to existing single, serial or multi-lectin affinity chromatography (18, 19), LeMBA-MS/MS offers the capability to simultaneously screen 20 lectins in a semi-automated, high throughput manner. On the other hand, since LeMBA-MS/MS identifies the non-glycosylated peptides, it cannot be used for glycan site assignment and glycan structure elucidation (20-23). However, the main advantage of LeMBA, we believe, is as a part of an integrated translational biomarker pipeline leading to lectin immunoassays. The lack of glycan structure details is not critical for clinical translation, as exemplified by the alpha-fetoprotein-L3 (AFP-L3) test, which measures the *Lens culinaris* agglutinin (LCA) binding fraction of serum alpha-fetoprotein (24, 25), and has been approved by the U.S. Food and Drug Administration for detection of hepatocellular carcinoma.

In this study, we report the extension of the glycoprotein biomarker pipeline to the qualification phase with LeMBA-MRM-MS, and introduce statistical analysis pipelines GlycoSelector (http://glycoselector.di.uq.edu.au/) and Shiny mixOmics (http://mixomics-projects.di.uq.edu.au/Shiny) for the discovery and qualification phases, respectively. The utility of this integrated serum glycoprotein biomarker pipeline is demonstrated using esophageal adenocarcinoma (EAC) with unmet clinical need for an *in vitro* diagnostic test. EAC is a lethal malignancy of the lower esophagus with very poor 5-year survival rate of less than 25% (26). EAC is becoming increasingly common and its incidence is associated with the prevalent precursor metaplastic condition Barrett's esophagus (BE), but with a low annual

conversion rate of up to 1% (27). A common set of risk factors are described for BE and EAC, include gastroesophageal reflux disease (GERD), obesity, male gender, and smoking (28, 29). The current endoscopy-biopsy based diagnosis is invasive and costly, leading to an ineffective surveillance program. A blood test employing serum biomarkers that can distinguish patients with EAC from those with either BE or healthy tissue would, potentially, change the paradigm for the way in which BE and EAC are managed in the population (30). Serum glycan profiling studies have shown differential expression of glycan structures between healthy, BE, early dysplastic and EAC patients (31-35). However, diagnostic serum glycoproteins showing differential glycosylation hence differential lectin binding remain to be discovered, making it a suitable disease model for this study.

Experimental Procedures

Study design and sample information

The overall biomarker study design was based on the strategy proposed by Rifai *et al.* (3), with the current work spanning discovery and qualification of the described four-phase paradigm. Serum samples were collected as part of the Australian Cancer Study (ACS) (36) and Study of Digestive Health (SDH) (37). All patients in these studies gave written, informed consent, and the studies were approved by the Human Research Ethics Committees of Queensland Institute for Medical Research, the University of Queensland, and all participating hospitals. Identical SOPs were followed for collecting samples for SDH and ACS, and processed by the same person. All 29 serum samples (Healthy-9, BE-10 and EAC-10) used for biomarker discovery phase and 79 serum samples (Healthy-20, BE-20, EAC-20 and population control-19) used for biomarker qualification study were matched by age; all selected patients were male considering the high male-dominance of EAC (29). The samples were stored at -80°C until use. Healthy controls were individuals with no history of esophageal cancer and no evidence of esophageal histological abnormality at the time of endoscopic sample collection. BE patients had a histologically confirmed diagnosis of Barrett's mucosa. EAC patients had histologically confirmed adenocarcinoma within the distal esophagus or gastro-esophageal junction. EAC patient sera were collected prior to the commencement of cancer treatment. Population controls were volunteers with no self-reported history of EAC or BE. Samples were randomized prior to all experiments. Table 1 and 2 describes patient information used in this study. For categorical and numerical variables related to patient information, *P* values were calculated using Fisher's exact test and Kruskal-Wallis test respectively.

Materials

MyOne™ Tosyl activated Dynabeads were from Life Technologies. Lectins AAL, BPL, DSA, EPHA, GNL, JAC, LPHA, MAA, NPL, SNA, STL and UEA were from Vector Laboratories. Modified sequencing grade trypsin was from Promega. Protein assay Bradford reagent, Triton X-100 and sodium dodecyl sulfate solution were from Bio-rad. Tris base, glycine, sodium chloride and acrylamide/bis-acrylamide solution 40% 29:1 were from Amresco. Glycerol, disodium hydrogen phosphate dihydrate, sodium dihydrogen phosphate dihydrate, calcium chloride dihydrate and Tween-20 were from Ajax Finechem. Magnesium chloride and manganese chloride were from Univar. For quadrupole time of flight runs, acetonitrile, isocratic HPLC grade was from Scharlau and for triple quadrupole runs, acetonitrile CHROMASOLV® gradient grade was from Sigma. Heavy labeled stable isotope-labeled standard (SIS) peptides were from Sigma. All other reagents including lectins not listed above were from Sigma unless otherwise specified.

Serum glycoprotein biomarker discovery and qualification pipeline

Figure 1 represents the integrated glycoprotein biomarker discovery and qualification pipeline developed using LeMBA. The discovery phase aimed to identify changes in the lectin binding of medium to high abundance serum proteins which can distinguish between different phenotypes. To enable economic and high throughput label-free quantitation while controlling for sample processing, including tryptic digestion, we employed a non-labeled spiked-in glycoprotein standard at the very first step of the workflow prior to denaturation (Figure 1). Pilot experiments identified chicken ovalbumin as a suitable internal standard, with low homology to species of interest (human or mouse) that bound to all 20 lectins experimentally. Optimization experiments determined that 10 picomole ovalbumin to be added to each sample (50 µg of serum) per lectin pull-down. Depending upon the individual

lectin, between 3 and 5 ovalbumin peptides (out of 7) were used for normalization (Supplemental Table 1). Details about the data normalization and statistical analyses platforms GlycoSelector and Shiny mixOmics can be found in Supplemental Methods.

Briefly, for discovery data, two different normalization approaches (i) based upon total ovalbumin protein intensity or (ii) using individual ovalbumin peptide intensity were evaluated (Supplemental Figure 1A). There was a strong correlation between the two normalization approaches (Supplemental Figure 1B), and we selected the second normalization method for the pipeline as it gave equal weighting to each peptide. For each patient sample in discovery stage, a two-dimensional dataset was generated, consisting of normalized intensity for proteins identified with each of the 20 lectin pull-down procedures. In general, glycoproteins bound several lectins, reflecting heterogeneity and multiplicity of glycosylation. GlycoSelector is a customized database with an incorporated statistical analysis pipeline coded in the R statistical programming language (38) and integrated in PHP server-side scripting language. The pipeline is based on tools developed in mixOmics (39), an R package dedicated to multivariate statistical analysis of 'omics' data, and includes several steps such as data normalization, sample outlier detection, multivariate statistical analysis and group binding analyses. The sample outlier detection step aims to identify possible errors in sample handling/processing (Supplemental Figure 2). As an example of its utility, sample run ID 63 shown in Supplemental Figure 2A to 2D was considered to be an outlier due to consistent anomalous results detected in all 4 graphical outputs. The error was at the mass spectrometry step, because when the sample was re-run after mass spectrometer re-calibration, it was no longer detected as an outlier (Supplemental Figure 2E to 2H). To determine changes in the lectin binding of individual proteins between the different conditions, GlycoSelector was designed with two parallel approaches. Firstly, Group Binding Difference analyses were performed to identify on-off changes. In addition, multivariate

statistical analysis based on sparse partial least square-discriminant analysis (sPLS-DA) (40) coupled with stability analysis was used to identify qualitative changes, after the exclusion of common contaminant proteins (Supplemental Table 3).

Based on GlycoSelector analysis, a subset of 6 lectins and 41 glycoprotein candidates were selected for independent qualification (Figure 1). The steps included, (i) MRM-MS assay development including confirmation of linearity and reproducibility, (ii) screening an independent cohort of patient samples using customized LeMBA-MRM-MS, (iii) two-step data normalization (Supplemental Figure 1A and Supplemental Methods), and (iv) univariate and multivariate statistical analysis using Shiny mixOmics.

Lectin magnetic bead array (LeMBA)

LeMBA was performed as previously reported (12, 13) with modifications detailed in Supplemental Methods section.

LC-MS/MS and database search for biomarker discovery

The LeMBA pull-down samples were resuspended in 20 µl of 0.1% v/v formic acid for LC-MS/MS (Agilent 6520 quadrupole time of flight [QTOF] coupled with a Chip Cube and 1200 HPLC). Initial experiments were performed to determine the optimal amount of tryptic peptides for LC-MS/MS: 9 µl were loaded for HAA, HPA and UEA, 6 µl for NPL, STL, GNL, 5 µl for BPL, DSA, ECA, MAA, SBA, WFA, and WGA, 4 µl for AAL, SNA, LPHA, PSA and JAC, 1 µl for EPHA and ConA. The nano pump was set at 0.3 µL/min and the capillary pump at 4 µL/min. The HPLC-chip used contains 160 nl C18 trapping column, and 75 µm x 150 mm 300 Å C18 analytical column (G4240-62010 Agilent Technologies). Buffer A was 0.1% v/v formic acid and Buffer B was 90% v/v acetonitrile containing 0.1% v/v formic acid. Peptides were eluted from the column using a gradient from 6% B to 46% B

at 45 min. Nano pump %B was increased to 95 %B at 45.5 min and plateaued till 55.5 min, then decreased to 6% B at 58.5 min. The mass spectrometer was operated in 2 GHz extended dynamic range and programmed to acquire 8 precursor MS1 spectra per second and 4 MS/MS spectra for each MS1 spectra. Dynamic exclusion was applied after 2 MS/MS within 0.25 min. Exclusion for lectin peptides was applied as reported previously (12). The QTOF was tuned and calibrated prior to analysis. One hundred femtomole/µl of pre-digested bovine serum albumin peptides were used as quality control, before and after each plate. Levels of reference ions 299.2945 and 1221.9906 were maintained at minimum 5000 and 1000 counts respectively.

The raw data was extracted and searched using Spectrum Mill MS proteomics workbench (Agilent Technologies, Rev.B.04.00.127) against Swissprot human database containing 20,242 entries (release 3rd Jan 2012). Similar MS/MS spectra acquired on the precursor *m/z* within ± 1.4 m/z and within ± 15 sec were merged. The following parameters were used for the search: Trypsin for digestion of proteins, 2 maximum missed cleavages, minimum matched peak intensity of 50%, precursor mass tolerance of ± 20 ppm, product mass tolerance of ± 50 ppm, calculate reversed database scores enabled and dynamic peak thresholding enabled. Carbamidomethylation was selected as fixed modification and oxidized methionine was selected as a variable modification. Precursor mass shift range from -17.0 Da to 177.0 Da was allowed for variable modification. Results were filtered by protein score > 15, peptide score > 6, and % scored peak intensity (% SPI) > 60. Automatic validation was used to validate proteins and peptides with default settings and false discovery rates (FDRs) were calculated using reversed hits.

BE-EAC biomarker qualification using LeMBA-MRM-MS

Multiple reaction monitoring-mass spectrometry (MRM-MS) assay was performed on Agilent Technologies 6490 triple quadrupole mass spectrometer coupled with 1290 standard-flow infinity UHPLC fitted with a standard-flow ESI (Jet Stream) source to qualify candidate proteins for customized list of six lectin pull-downs (AAL, EPHA, JAC, NPL, PSA and WGA) in an independent patient cohort. LeMBA was performed as described in Supplemental Methods. During MRM method development and validation stages, LeMBA pull-down of multiple lectins was combined and injected.

Protein, peptide and transition selection for MRM method development

Proteins identified using GlycoSelector with either of the above six lectins were selected for qualification. In addition, a few other proteins that were identified as candidate biomarkers with other lectins were also included for qualification.

MRM selector function of Spectrum Mill was used to retrieve the top ten peptides per protein for MRM method development, using several runs from the LeMBA-QTOF discovery data set. The parameters specified included 10 peptides per protein with a score of above 10 and % score peak intensity of 70%. The top four product y-ions for each precursor ion greater than precursor m/z were selected for MRM method development. The formula Collision energy (CE) = 0.036 *m/z* - 4.8 was used to calculate CE for each precursor. Multiple MRM methods consisting of maximum 200 transitions were created as a first step of method development. All methods were transferred across to Skyline software version 2.1.0.4936 (http://skyline.maccosslab.org/) for visualization, subsequent method refinement and analysis (41). Using LeMBA-MS/MS discovery data (.mzxml and .pepxml files), a reference spectral library was built in Skyline. This reference library was used to compare the peptide fragmentation pattern in the MRM method as compared to QTOF data, and also to rank

transitions. LeMBA pull-down from each of the 6 lectins was combined and run for each method to identify best MRM transitions. Each method incorporated transitions for internal standard chicken ovalbumin. Retention time prediction calculator iRT-C18 of Skyline was used to increase confidence of peptide identification (42). iRT scale was calibrated using the known retention time of the peptides and based on this calibration plot, retention time for the peptide of interest was predicted. MRM transitions showing good response at the correct retention time without any interference were selected for the next step. After the first round of method development, three MRM methods were created and each of these methods was run three times to find transitions showing stable responses. Some product y-ions (greater than precursor m/z) showed considerably low response. So to find out transition with higher response, up to five b- and y-ions less than precursor m/z were tried. Only transitions showing stable response during multiple runs were selected. Using retention time information for each peptide, one final dynamic MRM method was created incorporating a total of 145 peptides and 465 transitions with delta retention times of 2.5, 3 or 4 min, to quantify 41 proteins. Supplemental Table 6 contains a detailed list of transitions used in the method.

LC method development

The UHPLC system consisted of a reverse phase chromatographic column AdvanceBio Peptide Mapping (150×2.1 mm i.d., 2.7 µm, part number 653750-902, Agilent Technologies) with a 5 mm long guard column. Mobile phase A consisted of 0.1% formic acid, and mobile phase B consisted of 100% acetonitrile and 0.1% formic acid. The UHPLC system was operated at 60°C, with a flow rate of 0.4 mL/min. The gradient used for peptide separation was as follows: 3% B at 0 min; 30% B at 20 min; 40% B at 24 min; 95% B at 24.5 min; 95% B at 28.5 min; 3% B at 29 min; followed by conditioning of columns for 5 min at 3% B before injecting the next sample.

Mass spectrometer settings

Agilent 6490 triple quadrupole mass spectrometer was operated in positive ion mode and controlled by Agilent's MassHunter Workstation software (version B.06.00 build 6.0.6025.4 SP4). The MRM acquisition parameters were 150 V high pressure RF, 60 V low pressure RF, 4000 V capillary voltage, 300 V nozzle voltage, 11 L/min sheath gas flow at a temperature of 250 °C, 15 L/min drying gas flow at a temperature of 250 °C, 30 psi nebulizer gas flow, unit resolution (0.7 Da full width at half maximum in the first quadrupole (Q1) and the third quadrupole (Q3), and 200 V delta EMV (+).

Loading capacity determination

Loading capacity for individual lectin pull-down was determined by injecting varying amounts of LeMBA pull-down and monitoring peptide responses using MRM-MS assay. Each LeMBA pull-down sample was resuspended in 20 µl 0.1% formic acid. 10 µl of this reconstituted sample was mixed with 6 µl SIS peptide mixture containing 150 femtomole of ISQAVHAAHAEINEAG**R** and VASMASE**K** each, 300 femtomole of VTSIQDWVQ**K**, and 30 femtomole of AVEVLP**K**. Out of the total 16 µl mixture, the optimized loading for AAL, JAC, NPL and PSA lectin was 13 µl, EPHA lectin was 11.5 µl and WGA lectin was 12.5 µl.

Linearity and reproducibility of MRM-MS

Linearity of the MRM-MS method was determined by injecting varying concentrations of aforementioned four SIS peptides spiked-into combined LeMBA pull-down sample of multiple lectins. The amount of SIS peptide spiked-in for each of four peptides was adjusted in such a manner that the response from the 1X labeled peptide mix fell within a 5-fold range of the cognate natural peptide. The concentration of spiked-in SIS peptide varied from 0.008X to 25X covering 3125 fold linear range where 1X concentration indicates

mixture of 150 femtomole of ISQAVHAAHAEINEAG**R** and VASMASE**K** each, 300 femtomole of VTSIQDWVQ**K**, and 30 femtomole of AVEVLP**K**. All dilutions were run in triplicate on each day for three consecutive days (n = 9). The ratio of SIS peptide response/natural peptide response was plotted.

Reproducibility of MRM-MS assay was determined by injecting combined LeMBA pull-down sample from the 6 lectins in quadruplicate on each day for four consecutive days (n = 16). This experiment also determined the stability of the sample resuspended in 0.1% formic acid under the storage condition in the auto sampler. It was anticipated that once samples were resuspended in 96 well plates, they would be run within three days. Hence reproducibility was checked for four consecutive days after reconstituting samples. Percent coefficient of variation (% CV) between runs was calculated using peptide responses normalized with respect to ovalbumin peptide.

Screening samples for LeMBA-MRM-MS qualification

Lectin-beads sufficient for biomarker qualification experiments were made in a single batch to minimize experimental variation. Serum samples were randomized for LeMBA-MRM-MS experiments. % CV of the entire LeMBA-MRM-MS screen was calculated based on response of heavy labeled SIS peptides and internal standard natural chicken ovalbumin peptides. All three SIS peptides, except methionine containing heavy labeled peptides, showed a % CV of less than 20% while % CV for normalized intensity of natural internal standard ovalbumin peptide was around or below 20% (Supplemental Table 8), suggesting robustness (43) of the LeMBA pull-down and mass spectrometric measurement. Interestingly, normalized intensity of natural methionine containing peptide VASMASEK showed less variation as compared to non-normalized intensity, suggesting SIS peptide VASMASE**K** containing methionine was able to correct for batch effects in methionine

oxidation. One hundred femtomole of synthetic peptides mixture (G2455-85001, Agilent technologies) containing seven different human serum albumin peptides was run regularly to check linearity and reproducibility of the mass spectrometer results. When necessary, the mass spectrometer was tuned and calibrated.

Data processing

Raw data from MRM-MS experiment was processed using Skyline. All peaks were manually checked for correct integration, and peak area for each peptide (sum of all transitions) was exported for further analysis. For linearity experiments, the ratio of SIS:Natural peptide was calculated and plotted against SIS peptide spiked-in concentration. Median normalization was performed for each lectin dataset separately (Supplemental Figure 1A). Natural ovalbumin peptide peak intensity was first normalized with respective SIS labeled ovalbumin peptides. Next, using normalized intensity of natural ovalbumin peptide, the intensity of all other peptides was normalized. Methionine and non-methionine containing peptides were dealt with separately during normalization steps, to account for batch effects in methionine oxidation. For reproducibility experiments, the normalized response with respect to ovalbumin peptide was calculated for each run, and the % CV of 16 injections of the same sample run over a period of four days calculated. Detailed statistical analysis was performed using normalized peptide intensity in the computing environment R. Supplemental Table 7 contains normalized intensity of LeMBA-MRM-MS data.

Biomarker qualification at protein level using LeMBA-western immunoblotting

The top two candidates AAL-HP and AAL-GSN, identified using sPLS-DA/stability analysis and the on/off change function of GlycoSelector, respectively, were verified using LeMBA-western immunoblotting in two sets of patient cohorts, firstly using serum samples from the same patients used for the discovery phase, and secondly using an independent set

of 60 serum samples used for qualification. After AAL lectin pull-down, beads were directly boiled in 2X Laemmli sample buffer to elute captured glycoproteins. The denatured samples were separated by SDS-PAGE and proteins were transferred to polyvinylidene difluoride membrane (Millipore) using wet transfer. To compare results across membranes, an AAL pull-down sample from one healthy volunteer serum (unrelated to samples used in screen) was loaded in equal amounts on all gels. Membranes were blocked with 5% bovine serum albumin (BSA) in Tris-buffered saline-0.1% Tween-20 (TBST) for 1 hr, and then carefully cut into two halves. The upper part of the membrane was incubated overnight with anti-gelsolin antibody (Epitomics #EP1940Y; 1: 3000 dilution in 5% BSA/TBST), while the lower portion was incubated with anti-haptoglobin antibody (Gen Way Biotech #GWB-16A7EA; 1: 1000 dilution in 5% BSA/TBST) at 4 ºC overnight. This was followed by three TBST washes and incubation with 1: 3000 dilution of HRP-labeled anti-rabbit secondary antibody for 1 hr at room temperature (Invitrogen #A10547). The blots were developed using SuperSignal West Pico chemiluminescence (Thermo Scientific), and captured on film (Fuji film; Developer and fixer solutions were from Kodak). Densitometric analysis was performed using ImageJ (NIH, USA) (44). Raw densitometric values were normalized using the internal control sample loaded onto each gel.

Analysis for confounders

To check the impact of confounding covariates [reflux frequency, body mass index (BMI), smoking, and alcohol consumption] on biomarker candidates, an additional 19 population control (electoral roll) serum samples were measured using LeMBA-MRM-MS, to achieve sufficient number of disease-free samples for statistical analysis. Healthy and population control sample groups were merged and categorized according to reflux frequency, BMI, cumulative smoking history and alcohol consumption. Kruskal-Wallis test was applied to all the qualified candidates for each confounding factor. Candidates that

showed P < 0.05 for BMI, reflux, cumulative smoking history or alcohol consumption were considered as false positives and removed prior to multivariate analysis.

Functional annotation analysis

A list of candidates that differentiated EAC from BE was determined based on univariate Kruskal-Wallis test ($P < 0.05$) or multivariate analysis using sPLS-DA (stability > 70%) (Supplemental Table 10). The combined list of differential proteins was used in order to assess gene ontology differences between sample groups. We used the plasma proteome gene list (45), converted to DAVID IDs as a background in order to test for ontology differences by over-representation analysis using the DAVID (http://david.abcc.ncifcrf.gov/) website (46, 47) with default feature and algorithm settings. Ontology categories with adjusted FDR *P*-values < 0.05 were recorded. While we report individual ontologies, we applied the built-in functional annotation clustering function to help select representative ontologies for each main cluster (cluster scores over 3).

Results

Esophageal adenocarcinoma biomarker discovery

An overview of the integrated glycoprotein biomarker discovery and qualification pipeline is shown in Figure 1. The discovery phase used age-matched serum samples from 9 healthy, 10 BE and 10 EAC male patients (Table 1). BE and EAC patient groups had a significantly higher proportions of patients with hiatus hernia compared to healthy controls, as has previously been reported (48). We identified a total of 195 unique proteins from the MS/MS data (Supplemental Table 2). There was no difference between total number of proteins identified between healthy, BE and EAC patient groups (Supplemental Figure 3). The discovery LeMBA-MS/MS data were uploaded to GlycoSelector for data housing and statistical analysis. The sPLS-DA sample representation, including the top 100 candidates (lectin-protein pairs) in the model, showed clear separation of the samples according to their phenotype (Figure 2A and Supplemental Figure 4).

To select the most consistent candidates across patients to the qualification stage, we used the stability function built into GlycoSelector, which utilizes a leave-one-out strategy to assess the utility of each candidate biomarker. A relatively non-stringent cut-off of 70% was chosen for this purpose. Out of the top 100 lectin-protein pairs, 57 candidates passed the stability cut-off of 70% between healthy vs BE, 72 candidates passed for BE vs EAC, and 76 candidates, passed for healthy vs EAC (Figure 2B, Supplemental Figure 4, and Supplemental Table 4). A second, parallel approach used the group binding difference tool in GlycoSelector, to select on-off candidates which may not be selected by the statistical approach. Using relatively non-stringent criteria of 60%/40% presence/absence, this approach identified another 14, 20 and 26 candidates respectively for healthy vs BE, BE vs EAC and healthy vs EAC analyses (Supplemental Table 4). Candidates identified using sPLS-DA and

the group binding differences tools were complementary and showed no overlap between lectin-protein candidates, justifying the use of two different approaches for candidate selection. All 20 lectins used in the discovery phase showed differential binding with anywhere between one [e.g. *Helix aspersa* agglutinin (HAA)] to twenty five [e.g. *Narcissus pseudonarcissus* lectin (NPL)] glycoprotein candidates for pair-wise comparison between patient phenotypes (Figure 2C). This suggests widespread changes in the serum glycosylation profile between healthy, BE and EAC samples in agreement with previous studies (31-35). There was considerable overlap between glycoprotein candidates identified between healthy vs BE, BE vs EAC and healthy vs EAC patient groups (Figure 2D).

Immunoblotting was used for orthogonal protein level confirmation of the LeMBA-MS/MS screen. We chose two protein candidates which showed altered binding to *Aleuria aurantia* lectin (AAL) and for which antibodies were commercially available. AAL-haptoglobin (HP; Uniprot entry: P00738) was one of the top ranked candidates in sPLS-DA analysis for healthy vs EAC and BE vs EAC, while AAL-gelsolin (GSN; Uniprot entry: P06396) was identified using the group binding difference function of GlycoSelector as on-off change between BE vs EAC and healthy vs EAC. Using the same set of discovery serum samples, we performed pull-down using AAL and measured haptoglobin and gelsolin binding by immunoblotting. A control serum sample was loaded on every blot as a normalizer between membranes. LeMBA-immunoblotting confirmed the MS/MS results (Figure 2E, 2F and Supplemental Table 5), and showed higher sensitivity as it detected low levels of gelsolin in all patient samples, when some were undetectable by MS/MS [AAL-HP: label-free proteomics P value = 0.0868, western immunoblotting P value = 0.0267; AAL-GSN: label-free proteomics P value = 0.0254, western immunoblotting P value = 0.0019].

Biomarker qualification

Multiple reaction monitoring-mass spectrometry (MRM-MS) assay was optimized for 41 target protein candidates based on the GlycoSelector results with 2-3 peptides per protein and 2-3 transitions per peptide (Figure 3A, list of transitions in Supplemental Table 6). The linearity of this multiplexed assay was evaluated by spiking 4 stable isotope standard (SIS) peptides spanning a 3125 fold dilution range (Figure 3B). The reproducibility of the MRM-MS assay was determined by running the same sample in quadruplicate for four consecutive days. As illustrated in Figure 3C**,** 86% of the peptides measured using MRM-MS assay showed % CV below 10%, while 9% of peptides showed % CV between 10-20%, and only 5% of the peptides were above 20% suggesting overall reproducibility of MRM-MS assay.

For the qualification cohort (20 healthy, 20 BE, and 20 EAC; Table 2), the prevalence of reflux and obesity was consistent with a previous report (49) showing higher frequency in BE/EAC patient groups compared to the healthy group. Age matched electoral roll control and healthy groups were very similar across all measured covariates (Table 2). Based on the GlycoSelector results, we selected 6 lectins (AAL, Erythroagglutinin *Phaseolus vulgaris* [EPHA], jacalin [JAC], NPL, *Pisum sativum* agglutinin [PSA], and wheat germ agglutinin [WGA]) for qualification in this independent cohort of samples using MRM-MS assay for 41 target proteins, hence measuring a total of 246 lectin-protein candidates.

Two sequential steps were used to evaluate and select candidate biomarkers from the qualification data; first, Kruskal-Wallis non-parametric test to assess statistical significance of each individual candidate, then area under receiver operating characteristic (AUROC) curve was used to measure the diagnostic potential of each marker. Pairwise comparisons were made between the three phenotypes: healthy vs BE, BE vs EAC and healthy vs EAC. Out of 246 lectin-protein candidates, 45 candidates were significantly different between any two groups (FDR < 0.05) (Table 3). Amongst them, 26 lectin-protein candidates showed

AUROC of more than 0.7 in at least one of the three phenotype comparisons. Boxplots and ROC curves of the top candidates for healthy vs BE, BE vs EAC and healthy vs EAC are shown in Figure 4A to 4F respectively. Supplemental Figure 5 and 6 contain boxplots and ROC curves for all the candidates. As shown in Figure 4G, 16 candidates overlapped between healthy vs EAC and BE vs EAC analysis and might be of greatest interest as they can differentiate EAC from healthy as well as BE phenotype.

Orthogonal qualification at protein level using LeMBA-immunoblotting (IB) was performed for AAL-HP and AAL-GSN using samples from the qualification cohort. Once again, there was agreement between peptide level quantitation using MRM-MS and protein level quantitation using IB (Supplemental Figure 7 and Supplemental Table 9)**,** validating the LeMBA-MRM-MS workflow [AAL-HP: MRM-MS *P* value = 0.0235, western immunoblotting *P* value = 0.1037, MRM-MS AUROC = 0.69, western immunoblotting AUROC = 0.69; AAL-GSN: MRM-MS *P* value = 0.0120, western immunoblotting *P* value = 0.0203, MRM-MS AUROC = 0.70, western immunoblotting AUROC = 0.73]. For further evaluation, we undertook functional enrichment analysis of the list of candidates that differentiated EAC from BE, which included 17 unique proteins from 59 lectin-protein pairs (Supplemental Table 10). In agreement with the glycoprotein enrichment strategy, the top Annotation Cluster with an Enrichment Score of 10.4 included SP_PIR_KEYWORD glycoprotein (*P* = 1.82E-08) and the UP_SEQ_FEATURE glycosylation site:N-linked (GlcNAc...) (*P* = 2.32E-06). Additional clusters related to acute inflammation, complement cascade pathway, and endopeptidase inhibition, were over-represented within the 17 genes that discriminated BE and EAC. KEGG "Complement and coagulation cascades" pathway (hsa04610) was significantly over-represented (*P* = 4.6E-18), including ten genes compared to the full list of plasma proteins. This result is in agreement with the involvement of

inflammation in BE to EAC pathogenesis (50, 51), and points to alterations in the complement cascade in EAC development.

Identification of candidates affected by confounding covariates

As expected from the known risk factors, healthy, BE and EAC patient groups significantly differ according to BMI and reflux frequency (Table 2). Compared to healthy patients, BE and EAC patient groups had a higher proportion of patients who were obese or experienced frequent GERD. Moreover, functional annotation analysis suggest enrichment of the pathways related to inflammation between BE and EAC. Therefore, it may be possible that some of the candidates identified are due to confounding covariates rather than the actual disease phenotype. To evaluate this hypothesis, firstly the cohort size of healthy phenotype was increased by LeMBA-MRM-MS measurement of an additional 19 control serum samples collected as disease-free, electoral roll samples. These 39 disease-free patient samples were then classified according to potential confounding variables (reflux frequency, BMI, cumulative smoking history and alcohol consumption). The statistical significance of each 45 lectin-protein candidates for each of the four covariates was assessed using a Kruskal-Wallis test. Most of the candidate biomarkers were not significantly correlated with the covariates. As examples, boxplots of the data for the top 3 biomarker candidates of the disease-free cohort classified according to covariates are shown in Figure 5. Out of the four covariates studied, reflux frequency is perhaps the most important factor to be considered in the context of BE/EAC. Notably, none of the candidates were affected by reflux frequency, suggesting specificity of the candidates to diagnose disease phenotype. Five candidates significantly correlated with covariates (Supplemental Table 11 and Supplemental Figure 8). Apolipoprotein B-100 (APOB; Uniprot entry: P04114) showed differential binding with lectins AAL, JAC, and NPL according to BMI classification. This is most likely due to increased levels of total APOB with increase in BMI, suggesting underlying changes in the

lipoprotein metabolism (52). Plasma protease C1 inhibitor (SERPING1; Uniprot entry: P05155) showed significantly reduced binding with JAC lectin in samples classified as overweight and obese as compared to healthy while JAC-alpha-1B-glycpoprotein (A1BG; Uniprot entry: P04217) varied according to alcohol consumption. This covariate analysis led us to eliminate 5 candidates from the qualified biomarker list, leaving 40 biomarker candidates for future studies. Out of the 5 candidates that were eliminated, JAC-APOB was identified in healthy vs BE analysis, AAL-APOB and JAC-A1BG were identified in BE vs EAC analysis, JAC-SERPING1 was identified in healthy vs EAC analysis while NPL-APOB was significantly different in healthy vs BE and BE vs EAC analysis. Notably, none of the 16 lectin-protein candidates that distinguish EAC from BE and healthy phenotype were identified as confounding candidates.

Multimarker panel for EAC

Next we examined the potential of protein glycoforms as complementary biomarkers, focusing on differential diagnosis of EAC and BE, since this is critical for making clinical decisions. After removal of confounding candidates, sPLS-DA was used to derive a multimarker panel that distinguish BE and EAC (Figure 6A). The biomarker panel (BE vs EAC) included four unique proteins namely complement component C9 (C9; Uniprot entry: P02748), alpha-1B-glycoprotein (A1BG; Uniprot entry: P04217), complement C4-B (C4B; Uniprot entry: P0C0L5) and complement C2 (C2; Uniprot entry: P06681) with each of the six lectins appearing at least once in the panel. Using 5-fold cross-validation repeated 1000 times on this multimarker panel, the model showed cross-validation error rate of 37.47% and moderate separation of the BE and EAC sample representations (Figure 6A). The combined signature of the eight candidates gave an AUROC of 0.9425 with 95% specificity and 80% sensitivity (Figure 6B).

Discussion

In this study we present an alternative workflow to identify glycosylation changes in medium to high abundant glycoproteins using serum as the sample source and lectins to interrogate glycan moieties throughout discovery and qualification. Our workflow was designed to enhance the feasibility of glycoprotein biomarker discovery and translation, through scientific rigor while managing the experimental cost. Firstly, serum was used as the sample source throughout discovery and qualification, hence eliminating the risk of switching tissue type during biomarker development. Secondly, single step enrichment using liquid handler assisted LeMBA-system reduced sample processing variability. Thirdly, we utilized the comparatively inexpensive approach of label-free proteomics using relative quantitation with respect to a spiked-in internal standard chicken ovalbumin. This approach achieved the necessary analytical linearity and reproducibility throughout the more than 2000 hr of total mass spectrometer run time performed in the study. This cost-effective strategy can be applied across other existing proteomics platforms to account for variations during sample processing and enable relative quantitation for a large number of candidates without costly SIS labeled peptides. Fourthly, we applied a sequential filtering approach (53) in which many candidates were selected from biomarker discovery proteomics data, and qualified using MRM-MS with increasing sample size in a cost-effective manner. Finally, we introduced software tools for data visualization and statistical analysis in the form of web-interfaces. Both GlycoSelector (http://glycoselector.di.uq.edu.au) and Shiny mixOmics (http://mixomics-projects.di.uq.edu.au/Shiny) platforms are publicly available, user-friendly and require minimal and no background in statistics or computer programming.

The main feature of both these web-interfaces is the use of multivariate sPLS-DA method which enables data dimension reduction, insightful graphical outputs and the

identification of key discriminative features with respect to the biological outcome of interest. In Shiny mixOmics we have added important preliminary data mining steps for 'omics' data visualization such as sample boxplots, coefficient of variation barplots, hierarchical clustering, PCA in order to identify potential outliers prior to statistical analyses. The univariate statistical analysis step includes Krukal-Wallis, ANOVA tests as well as ROC analysis that can be performed efficiently on thousands of variables and results can also be output in a common file format. While similar sorts of analyses can be performed using commercially available software packages such as GraphPad Prism or Origin, these require additional computer programming skills in order to automate the analysis for hundreds of data points. Finally, the multivariate statistical analysis step with sPLS-DA (also separately available in the R package mixOmics) (39) has been shown to identify relevant biological features, with a classification performance similar to other statistical approaches (40). The major advantage of such an approach is graphical representation of the results that univariate approaches cannot provide. Importantly, in this study we have shown that such multivariate methods can be used efficiently and reliably on proteomics data characterized by highly skewed and non normal distributions. Collectively, the two web-interface statistical tools that we propose enable data mining, univariate and multivariate statistical analyses which can be applied to other 'omics' datasets.

The success of cancer screening programs in improving outcomes for many cancer types emphasizes the importance of early diagnosis and the development of screening/surveillance tools (54). The lack of cost-effective screening/surveillance methodology to facilitate early diagnosis of EAC is one of the main reasons for the high mortality. Current endoscopy-based screening is costly, requires specialist appointment, and is not suitable for frequent large scale at risk population monitoring (27, 30). Several innovative screening methods are being evaluated, including advanced imaging (55, 56), non-

endoscopic sampling (57, 58), and blood biomarkers (30). Recently conducted genome profiling studies using next-generation sequencing platforms have concluded that the majority of key mutations are already acquired at the metaplastic stage of BE and only few driver mutations lead to progression of dysplasia and EAC (59, 60). This suggests that genomics-based screening approaches may have limitations as a screening technology. Despite this, the evidence for limited genomic changes between BE and EAC raises the possibility that more functional level changes (e.g. protein expression, protein glycosylation, metabolic changes etc.) may be driving the development of dysplasia/carcinoma from metaplastic condition. In line with this, studies have shown differential expression of glycan structures in tissue and serum samples during metaplasia-dysplasia-carcinoma sequence (23, 31-35, 61-65).

Out of the 246 lectin-protein candidates measured for qualification, 45 candidates (18.3%) were qualified in an independent cohort of patients. Interestingly, only 3 out of 45 candidates were significantly different between healthy vs BE comparison, with the other 42 candidates differentially present in EAC as compared to either healthy or BE samples. This suggests that EAC phenotype is significantly different from BE and healthy in terms of serum glycan expression. The lack of glycosylation changes between healthy and BE was somewhat surprising because genomic studies suggest that BE and EAC share a common mutational profile that differs from healthy samples (59, 60). The top candidate that differentiated between healthy and BE, NPL-APOB, was influenced by BMI. Hence, except EPHA-Alpha-2-macroglobulin (A2M), this study did not find any candidate that can differentiate BE from healthy, suggesting little or no change in glycosylation of serum proteins in the development of BE. However, the critical diagnostic need is to identify patients at early dysplasia or early stages of EAC, or those at high risk of progression. To progress towards this goal, the lectin-protein biomarker candidates should be evaluated in a patient cohort including low grade

dysplasia (LGD) and high grade dysplasia (HGD) phenotypes to precisely determine what disease stage can actually be diagnosed by the glycoprotein biomarker candidates.

Previous serum glycan profiling studies found reductions in total N-linked fucosylation in EAC as compared to healthy patient groups (33, 35). Here we used two fucose specific lectins AAL and PSA, both of them showed differential binding with 6 glycoproteins for BE vs EAC pair-wise comparison. Out of 6 glycoproteins that showed differential binding to fucose specific lectins, 4 showed increased levels in EAC samples for AAL lectin pull-down, while 5 candidates showed increased levels in EAC samples as compared to BE for PSA lectin pull-down. These data suggest that serum glycan changes are specific to the glycoprotein of origin, and this property could be exploited as a specific biomarker compared to overall changes in serum fucosylation.

Apart from the major goal of translating the biomarkers for diagnosis, the verified biomarkers could shed light on the pathogenesis of EAC. To this end, functional annotation analysis of the candidates was able to distinguish between EAC and BE through enrichment of "complement and coagulation cascades" pathway. Very recently Song and colleagues (23) also identified changes in the glycosylation of complement proteins for EAC and high grade dysplasia compared to a healthy phenotype. They used lectin-affinity chromatography (a mix of fucose and sialic acid binding lectin) and hydrazide chemistry-based glycoprotein enrichment methods to identify complement C3 and complement C1r subcomponent as differentially present in HGD and EAC samples respectively, as compared to serum from healthy cohort. The differences between these complement proteins, and those that discriminate BE and EAC, in our results, may be the result of divergent sample processing steps. For example, Song et al. (23) used serum sample after depletion of the seven most abundant proteins as compared to our workflow where as we denatured the serum samples to break protein complexes without depletion of abundant proteins. In our workflow we used an

individual lectin for enrichment of a particular type of glycan while Song and colleagues used a mixture of sialic acid binding *Sambucus nigra* agglutinin (SNA) and fucose specific AAL lectin for glycoprotein enrichment. Nonetheless, changes observed in the glycosylation of complement proteins may suggest a role for inflammation in EAC pathogenesis (50).

While lectins are a useful tool for discovery and translation, a limitation of our pipeline is the lack of identification of the actual glycosylation sites and glycan structures. For rapid translation of the verified biomarkers using a simple lectin-immunoassay format that can be readily achieved, the only information required is the lectin affinity and the protein identity. Hence, we have not incorporated detailed glycosylation site or glycan structure analysis to the current pipeline. Following further clinical evaluation, the final glycoprotein candidates could be subjected to full glycomics characterization to determine the changes in the glycan structure and/or site of glycosylation between different disease states. This may provide additional insight into the pathological basis of the cancer-associated glycosylation changes. We anticipate 3 possible scenarios for a glycoprotein to show differential lectin binding in our LeMBA based workflow. (i) Total glycoprotein level changes would lead to overall increased/decreased binding with multiple lectins. (ii) Changes in the glycan occupancy at a particular glycosylation site will lead to differential binding with multiple lectins. (iii) Differential expression of a specific glycan structure will alter binding of a glycoprotein to a particular lectin or a group of lectins. Further studies following biomarker qualification will be required to identify the exact mechanism of differential lectin binding for each candidate.

In summary, we have developed novel tools for glycoprotein biomarker discovery using serum. The cross-sectional pre-clinical biomarker exploratory study conducted using our workflow has identified a list of serum glycoprotein candidate biomarkers that can distinguish EAC from healthy and BE phenotype. These candidates will need to be further

evaluated in independent cohorts of patient samples that include different disease grades and subtypes, prior to prospective trials. The pipeline developed can be applied to other diseases with software tools GlycoSelector and Shiny mixOmics available online at http://glycoselector.di.uq.edu.au/ and http://mixomics-projects.di.uq.edu.au/Shiny.

The raw mass spectrometry data along with database search results including sequence database used for searches have been deposited to the publicly accessible platform ProteomeXchange Consortium (66) via the PRIDE partner repository with the dataset identifier PXD002442. The peptide identification results can be viewed using MS-Viewer (http://prospector2.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer) (67), using search key jn7qafftux.

Acknowledgements

This project was supported in part by grants from the University of Queensland-Ochsner Seed Fund for Collaborative Research (to M.M.H., A.P.B. and V.J.). Specialized instrumentation for the project was partly supported by the Ramaciotti Foundations and the Australian Cancer Research Foundation (ACRF). Serum samples for the study were acquired through The Australian Cancer Study (ACS) which was funded by Program Grants from the National Health and Medical Research Council (NHMRC) of Australia (199600, 552429), and Study of Digestive Health (SDH) which was supported by grant number 5 RO1 CA 001833-02 from the National Cancer Institute. M.M.H. was supported by the NHMRC of Australia Career Development Fellowship (569512). K.A.L.C. is supported in part by the Diamantina Individualised Oncology Care Centre (ACRF). D.C.W is supported by a Research Fellowship (APP1058522) from the NHMRC of Australia. N.A.S. is supported by a Senior Research Fellowship awarded by the Cancer Council Queensland, and grants from the Australian NHMRC (APP1049182) and the Cancer Council Queensland (APP1025479). A.K.S. is a recipient of International Postgraduate Research Scholarship and The University of Queensland Centennial Scholarship. E. C. was supported by The University of Queensland International Research Tuition Award and The University of Queensland Research Scholarship. We thank Dorothy Loo (The University of Queensland Diamantina Institute, The University of Queensland) and Dr Thomas Hennessey (Agilent Technologies) for technical assistance during mass spectrometric analysis. We would like to thank Peter Baker and Robert Chalkley of University of California, San Francisco and Karl Clauser of Broad Institute of Harvard and MIT for their generous and extended support to enable MS-Viewer to visualize our proteomics dataset.

References

1. Anderson, N. L., Ptolemy, A. S., and Rifai, N. (2013) The riddle of protein diagnostics: future bleak or bright? *Clin. Chem.* 59, 194-197
2. Pavlou, M. P., Diamandis, E. P., and Blasutig, I. M. (2013) The long journey of cancer biomarkers from the bench to the clinic. *Clin. Chem.* 59, 147-157
3. Rifai, N., Gillette, M. A., and Carr, S. A. (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* 24, 971-983
4. Diamandis, E. P. (2010) Cancer biomarkers: can we turn recent failures into success? *J. Natl. Cancer Inst.* 102, 1462-1467
5. Addona, T. A., Shi, X., Keshishian, H., Mani, D. R., Burgess, M., Gillette, M. A., Clauser, K. R., Shen, D., Lewis, G. D., Farrell, L. A., Fifer, M. A., Sabatine, M. S., Gerszten, R. E., and Carr, S. A. (2011) A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. *Nat. Biotechnol.* 29, 635-643
6. Keshishian, H., Burgess, M. W., Gillette, M. A., Mertins, P., Clauser, K. R., Mani, D. R., Kuhn, E. W., Farrell, L. A., Gerszten, R. E., and Carr, S. A. (2015) Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury. *Mol. Cell. Proteomics* 14, 2375-2393
7. Ademowo, O. S., Hernandez, B., Collins, E., Rooney, C., Fearon, U., van Kuijk, A. W., Tak, P. P., Gerlag, D. M., FitzGerald, O., and Pennington, S. R. (2014) Discovery and confirmation of a protein biomarker panel with potential to predict response to biological therapy in psoriatic arthritis. *Ann. Rheum. Dis.* DOI 10.1136/annrheumdis-2014-205417
8. Whiteaker, J. R., Zhang, H., Zhao, L., Wang, P., Kelly-Spratt, K. S., Ivey, R. G., Piening, B. D., Feng, L. C., Kasarda, E., Gurley, K. E., Eng, J. K., Chodosh, L. A.,

Kemp, C. J., McIntosh, M. W., and Paulovich, A. G. (2007) Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. *J. Proteome Res.* 6, 3962-3975

9. Liu, N. Q., Braakman, R. B., Stingl, C., Luider, T. M., Martens, J. W., Foekens, J. A., and Umar, A. (2012) Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. *J. Mammary Gland Biol. Neoplasia* 17, 155-164
10. Whiteaker, J. R., Lin, C., Kennedy, J., Hou, L., Trute, M., Sokal, I., Yan, P., Schoenherr, R. M., Zhao, L., Voytovich, U. J., Kelly-Spratt, K. S., Krasnoselsky, A., Gafken, P. R., Hogan, J. M., Jones, L. A., Wang, P., Amon, L., Chodosh, L. A., Nelson, P. S., McIntosh, M. W., Kemp, C. J., and Paulovich, A. G. (2011) A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat. Biotechnol.* 29, 625-634
11. Paulovich, A. G., Whiteaker, J. R., Hoofnagle, A. N., and Wang, P. (2008) The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. *Proteomics Clin. Appl.* 2, 1386-1402
12. Choi, E., Loo, D., Dennis, J. W., O'Leary, C. A., and Hill, M. M. (2011) High-throughput lectin magnetic bead array-coupled tandem mass spectrometry for glycoprotein biomarker discovery. *Electrophoresis* 32, 3564-3575
13. Loo, D., Jones, A., and Hill, M. M. (2010) Lectin magnetic bead array for biomarker discovery. *J. Proteome Res.* 9, 5496-5500
14. Fanayan, S., Hincapie, M., and Hancock, W. S. (2012) Using lectins to harvest the plasma/serum glycoproteome. *Electrophoresis* 33, 1746-1754
15. Drake, R. R., Schwegler, E. E., Malik, G., Diaz, J., Block, T., Mehta, A., and Semmes, O. J. (2006) Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. *Mol. Cell. Proteomics* 5, 1957-1967

16. Kim, E. H., and Misek, D. E. (2011) Glycoproteomics-based identification of cancer biomarkers. *Int J Proteomics* 2011, 601937

17. Kuzmanov, U., Kosanam, H., and Diamandis, E. P. (2013) The sweet and sour of serological glycoprotein tumor biomarker quantification. *BMC Med.* 11, 31

18. Cummings, R. D., and Kornfeld, S. (1982) Fractionation of asparagine-linked oligosaccharides by serial lectin-Agarose affinity chromatography. A rapid, sensitive, and specific technique. *J. Biol. Chem.* 257, 11235-11240

19. Yang, Z., and Hancock, W. S. (2004) Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J. Chromatogr. A* 1053, 79-88

20. Drake, P. M., Schilling, B., Niles, R. K., Braten, M., Johansen, E., Liu, H., Lerch, M., Sorensen, D. J., Li, B., Allen, S., Hall, S. C., Witkowska, H. E., Regnier, F. E., Gibson, B. W., and Fisher, S. J. (2011) A lectin affinity workflow targeting glycosite-specific, cancer-related carbohydrate structures in trypsin-digested human plasma. *Anal. Biochem.* 408, 71-85

21. Li, Y., Shah, P., De Marzo, A. M., Van Eyk, J. E., Li, Q., Chan, D. W., and Zhang, H. (2015) Identification of glycoproteins containing specific glycans using a lectin-chemical method. *Anal. Chem.* 87, 4683-4687

22. Zhou, H., Froehlich, J. W., Briscoe, A. C., and Lee, R. S. (2013) The GlycoFilter: a simple and comprehensive sample preparation platform for proteomics, N-glycomics and glycosylation site assignment. *Mol. Cell. Proteomics* 12, 2981-2991

23. Song, E., Zhu, R., Hammoud, Z. T., and Mechref, Y. (2014) LC-MS/MS quantitation of esophagus disease blood serum glycoproteins by enrichment with hydrazide chemistry and lectin affinity chromatography. *J. Proteome Res.* 13, 4808-4820

24. Kagebayashi, C., Yamaguchi, I., Akinaga, A., Kitano, H., Yokoyama, K., Satomura, M., Kurosawa, T., Watanabe, M., Kawabata, T., Chang, W., Li, C., Bousse, L., Wada, H. G., and Satomura, S. (2009) Automated immunoassay system for AFP-L3% using on-chip electrokinetic reaction and separation by affinity electrophoresis. *Anal. Biochem.* 388, 306-311

25. Sato, Y., Nakata, K., Kato, Y., Shima, M., Ishii, N., Koji, T., Taketa, K., Endo, Y., and Nagataki, S. (1993) Early recognition of hepatocellular carcinoma based on altered profiles of alpha-fetoprotein. *N. Engl. J. Med.* 328, 1802-1806

26. Hur, C., Miller, M., Kong, C. Y., Dowling, E. C., Nattinger, K. J., Dunn, M., and Feuer, E. J. (2013) Trends in esophageal adenocarcinoma incidence and mortality. *Cancer* 119, 1149-1158

27. Spechler, S. J., and Souza, R. F. (2014) Barrett's esophagus. *N. Engl. J. Med.* 371, 836-845

28. Reid, B. J., Li, X., Galipeau, P. C., and Vaughan, T. L. (2010) Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. *Nat. Rev. Cancer* 10, 87-101

29. Rutegard, M., Lagergren, P., Nordenstedt, H., and Lagergren, J. (2011) Oesophageal adenocarcinoma: the new epidemic in men? *Maturitas* 69, 244-248

30. Shah, A. K., Saunders, N. A., Barbour, A. P., and Hill, M. M. (2013) Early diagnostic biomarkers for esophageal adenocarcinoma--the current state of play. *Cancer Epidemiol. Biomarkers Prev.* 22, 1185-1209

31. Gaye, M. M., Valentine, S. J., Hu, Y., Mirjankar, N., Hammoud, Z. T., Mechref, Y., Lavine, B. K., and Clemmer, D. E. (2012) Ion mobility-mass spectrometry analysis of serum N-linked glycans from esophageal adenocarcinoma phenotypes. *J. Proteome Res.* 11, 6102-6110

32. Hu, Y., Desantos-Garcia, J. L., and Mechref, Y. (2013) Comparative glycomic profiling of isotopically permethylated N-glycans by liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 27, 865-877

33. Mechref, Y., Hussein, A., Bekesova, S., Pungpapong, V., Zhang, M., Dobrolecki, L. E., Hickey, R. J., Hammoud, Z. T., and Novotny, M. V. (2009) Quantitative serum glycomics of esophageal adenocarcinoma and other esophageal disease onsets. *J. Proteome Res.* 8, 2656-2666

34. Mitra, I., Zhuang, Z., Zhang, Y., Yu, C. Y., Hammoud, Z. T., Tang, H., Mechref, Y., and Jacobson, S. C. (2012) N-glycan profiling by microchip electrophoresis to differentiate disease states related to esophageal adenocarcinoma. *Anal. Chem.* 84, 3621-3627

35. Hammoud, Z. T., Mechref, Y., Hussein, A., Bekesova, S., Zhang, M., Kesler, K. A., and Novotny, M. V. (2010) Comparative glycomic profiling in esophageal adenocarcinoma. *J. Thorac. Cardiovasc. Surg.* 139, 1216-1223

36. Whiteman, D. C., Sadeghi, S., Pandeya, N., Smithers, B. M., Gotley, D. C., Bain, C. J., Webb, P. M., Green, A. C., and Australian Cancer, S. (2008) Combined effects of obesity, acid reflux and smoking on the risk of adenocarcinomas of the oesophagus. *Gut* 57, 173-180

37. Smith, K. J., O'Brien, S. M., Smithers, B. M., Gotley, D. C., Webb, P. M., Green, A. C., and Whiteman, D. C. (2005) Interactions among smoking, obesity, and symptoms of acid reflux in Barrett's esophagus. *Cancer Epidemiol. Biomarkers Prev.* 14, 2481-2486

38. R Core Team. (2014) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. http://www.R-project.org/.

39. Le Cao, K.A., Gonzalez, I., Dejean, S., with key contributors Rohart, F., Gautier, B., contributions from Monget, P., Coquery, J., Yao, FZ., and Liquet, B. (2015) mixOmics: Omics data integration project. R package version 5.0-4. http://CRAN.R-project.org/package=mixOmics.

40. Le Cao, K. A., Boitard, S., and Besse, P. (2011) Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* 12, 253

41. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966-968

42. Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., MacCoss, M. J., and Rinner, O. (2012) Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* 12, 1111-1121

43. Percy, A. J., Chambers, A. G., Yang, J., Domanski, D., and Borchers, C. H. (2012) Comparison of standard- and nano-flow liquid chromatography platforms for MRM-based quantitation of putative plasma biomarker proteins. *Anal. Bioanal. Chem.* 404, 1089-1101

44. Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9, 671-675

45. Nanjappa, V., Thomas, J. K., Marimuthu, A., Muthusamy, B., Radhakrishnan, A., Sharma, R., Ahmad Khan, A., Balakrishnan, L., Sahasrabuddhe, N. A., Kumar, S., Jhaveri, B. N., Sheth, K. V., Kumar Khatana, R., Shaw, P. G., Srikanth, S. M., Mathur, P. P., Shankar, S., Nagaraja, D., Christopher, R., Mathivanan, S., Raju, R., Sirdeshmukh, R., Chatterjee, A., Simpson, R. J., Harsha, H. C., Pandey, A., and

Prasad, T. S. (2014) Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res.* 42, D959-965

46. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44-57

47. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1-13

48. Andrici, J., Tio, M., Cox, M. R., and Eslick, G. D. (2013) Hiatal hernia and the risk of Barrett's esophagus. *J. Gastroenterol. Hepatol.* 28, 415-431

49. Thrift, A. P., Shaheen, N. J., Gammon, M. D., Bernstein, L., Reid, B. J., Onstad, L., Risch, H. A., Liu, G., Bird, N. C., Wu, A. H., Corley, D. A., Romero, Y., Chanock, S. J., Chow, W. H., Casson, A. G., Levine, D. M., Zhang, R., Ek, W. E., MacGregor, S., Ye, W., Hardie, L. J., Vaughan, T. L., and Whiteman, D. C. (2014) Obesity and risk of esophageal adenocarcinoma and Barrett's esophagus: a mendelian randomization study. *J. Natl. Cancer Inst.* 106, DOI 10.1093/jnci/dju252

50. Hardikar, S., Onstad, L., Song, X., Wilson, A. M., Montine, T. J., Kratz, M., Anderson, G. L., Blount, P. L., Reid, B. J., White, E., and Vaughan, T. L. (2014) Inflammation and oxidative stress markers and esophageal adenocarcinoma incidence in a Barrett's esophagus cohort. *Cancer Epidemiol. Biomarkers Prev.* 23, 2393-2403

51. Picardo, S. L., Maher, S. G., O'Sullivan, J. N., and Reynolds, J. V. (2012) Barrett's to oesophageal cancer sequence: a model of inflammatory-driven upper gastrointestinal cancer. *Dig. Surg.* 29, 251-260

52. Riches, F. M., Watts, G. F., Naoumova, R. P., Kelly, J. M., Croft, K. D., and Thompson, G. R. (1998) Hepatic secretion of very-low-density lipoprotein

apolipoprotein B-100 studied with a stable isotope technique in men with visceral obesity. *Int. J. Obes. Relat. Metab. Disord.* 22, 414-423

53. Feng, Z., Prentice, R., and Srivastava, S. (2004) Research issues and strategies for genomic and proteomic biomarker discovery and validation: a statistical perspective. *Pharmacogenomics* 5, 709-719

54. Siegel, R., Ma, J., Zou, Z., and Jemal, A. (2014) Cancer statistics, 2014. *CA Cancer J. Clin.* 64, 9-29

55. Suter, M. J., Gora, M. J., Lauwers, G. Y., Arnason, T., Sauk, J., Gallagher, K. A., Kava, L., Tan, K. M., Soomro, A. R., Gallagher, T. P., Gardecki, J. A., Bouma, B. E., Rosenberg, M., Nishioka, N. S., and Tearney, G. J. (2014) Esophageal-guided biopsy with volumetric laser endomicroscopy and laser cautery marking: a pilot clinical study. *Gastrointest. Endosc.* 79, 886-896

56. Leggett, C. L., and Gorospe, E. C. (2014) Application of confocal laser endomicroscopy in the diagnosis and management of Barrett's esophagus. *Ann Gastroenterol* 27, 193-199

57. Lao-Sirieix, P., Boussioutas, A., Kadri, S. R., O'Donovan, M., Debiram, I., Das, M., Harihar, L., and Fitzgerald, R. C. (2009) Non-endoscopic screening biomarkers for Barrett's oesophagus: from microarray analysis to the clinic. *Gut* 58, 1451-1459

58. Kadri, S. R., Lao-Sirieix, P., O'Donovan, M., Debiram, I., Das, M., Blazeby, J. M., Emery, J., Boussioutas, A., Morris, H., Walter, F. M., Pharoah, P., Hardwick, R. H., and Fitzgerald, R. C. (2010) Acceptability and accuracy of a non-endoscopic screening test for Barrett's oesophagus in primary care: cohort study. *BMJ* 341, c4372

59. Agrawal, N., Jiao, Y., Bettegowda, C., Hutfless, S. M., Wang, Y., David, S., Cheng, Y., Twaddell, W. S., Latt, N. L., Shin, E. J., Wang, L. D., Wang, L., Yang, W., Velculescu, V. E., Vogelstein, B., Papadopoulos, N., Kinzler, K. W., and Meltzer, S.

J. (2012) Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discov.* 2, 899-905

60. Weaver, J. M., Ross-Innes, C. S., Shannon, N., Lynch, A. G., Forshew, T., Barbera, M., Murtaza, M., Ong, C. A., Lao-Sirieix, P., Dunning, M. J., Smith, L., Smith, M. L., Anderson, C. L., Carvalho, B., O'Donovan, M., Underwood, T. J., May, A. P., Grehan, N., Hardwick, R., Davies, J., Oloumi, A., Aparicio, S., Caldas, C., Eldridge, M. D., Edwards, P. A., Rosenfeld, N., Tavare, S., Fitzgerald, R. C., and Consortium, O. (2014) Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat. Genet.* 46, 837-843

61. Bird-Lieberman, E. L., Dunn, J. M., Coleman, H. G., Lao-Sirieix, P., Oukrif, D., Moore, C. E., Varghese, S., Johnston, B. T., Arthur, K., McManus, D. T., Novelli, M. R., O'Donovan, M., Cardwell, C. R., Lovat, L. B., Murray, L. J., and Fitzgerald, R. C. (2012) Population-based study reveals new risk-stratification biomarker panel for Barrett's esophagus. *Gastroenterology* 143, 927-935.e3

62. Bird-Lieberman, E. L., Neves, A. A., Lao-Sirieix, P., O'Donovan, M., Novelli, M., Lovat, L. B., Eng, W. S., Mahal, L. K., Brindle, K. M., and Fitzgerald, R. C. (2012) Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Barrett's esophagus. *Nat. Med.* 18, 315-321

63. Iwaya, Y., Hasebe, O., Koide, N., Kitahara, K., Suga, T., Shinji, A., Muraki, T., Yokosawa, S., Yamada, S., Arakura, N., Tanaka, E., and Nakayama, J. (2014) Reduced expression of alphaGlcNAc in Barrett's oesophagus adjacent to Barrett's adenocarcinoma--a possible biomarker to predict the malignant potential of Barrett's oesophagus. *Histopathology* 64, 536-546

64. Neumann, H., Wex, T., Monkemuller, K., Vieth, M., Fry, L. C., and Malfertheiner, P. (2008) Lectin UEA-I-binding proteins are specifically increased in the squamous epithelium of patients with Barrett's esophagus. *Digestion* 78, 201-207

65. Shimamoto, C., Weinstein, W. M., and Boland, C. R. (1987) Glycoconjugate expression in normal, metaplastic, and neoplastic human upper gastrointestinal mucosa. *J. Clin. Invest.* 80, 1670-1678

66. Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., Martinez-Bartolome, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 32, 223-226

67. Baker, P. R., and Chalkley, R. J. (2014) MS-viewer: a web-based spectral viewer for proteomics results. *Mol. Cell. Proteomics* 13, 1392-1396

Figure Legends

Figure 1. Generalized workflow schematic for serum glycoprotein biomarker discovery and qualification. Serum samples from respective patient groups were enriched for sub-glycoproteomes using 20 individual lectin coated magnetic beads, followed by on-bead trypsin digest and tandem mass spectrometry for label-free quantitation referencing to internal standard chicken ovalbumin. In-house database and statistical analysis pipeline "GlycoSelector" (http://glycoselector.di.uq.edu.au/) identified lectin-protein pairs present in one patient group and absent in the other. Sparse partial least squares-discriminant analysis (sPLS-DA) combined with stability analysis was used to generate ranked lists of lectin-protein candidates. For biomarker qualification, selected candidates were measured using multiple reaction monitoring-mass spectrometry (MRM-MS) in an independent patient cohort for a subgroup of lectin pull-downs. Dedicated statistical analysis tool "Shiny mixOmics" (http://mixomics-projects.di.uq.edu.au/Shiny) was developed incorporating tools to plot receiver operating characteristic (ROC) curve and to perform univariate/multivariate statistical analyses. LeMBA-immunoblotting (IB) was used as an orthogonal method to verify peptide level MS data for selected candidates at the protein level.

Figure 2. Biomarker discovery and protein level qualification of two candidates. Serum samples from 29 patients (healthy-9, BE-10 and EAC-10) were screened using the LeMBA-GlycoSelector pipeline. **(A)** The sPLS-DA sample representation based on the top 100 lectin-protein candidates that differentiate EAC from BE. **(B)** Amongst the top 100 sPLS-DA candidates, 72 candidates passed the stability criteria of 70% based on leave-one-out cross-validation. Results of sPLS-DA and stability analysis for healthy vs BE and healthy vs EAC are available in Supplemental Figure 4. **(C)** Number of unique candidate proteins identified for each lectin in LeMBA-GlycoSelector analysis. All 20 lectins used for screening identified

at least one protein candidate. **(D)** Overlap between lectin-protein candidates that differentiate BE from healthy, EAC from BE, and EAC from healthy phenotype. **(E)** AAL-HP and **(F)** AAL-GSN were the top two candidates identified using sPLS-DA and group binding difference tool, respectively. **(E and F, top panel)** Label-free proteomics relative quantitation results for AAL-HP and AAL-GSN respectively. **(E and F, lower panel)** Normalized intensity for AAL-HP and AAL-GSN using immunoblotting. Raw densitometry values are provided in Supplemental Table 5.

Figure 3. Multiple reaction monitoring-mass spectrometry (MRM-MS) assay development outline including determination of assay linearity and reproducibility. (A) Outline of MRM-MS assay development. **(B)** Linearity of MRM-MS assay confirmed using SIS labeled peptide mix of 4 peptides diluted across 3125 fold and spiked-into a constant amount of LeMBA pull-down sample. **(C)** Reproducibility of MRM-MS assay for 16 replicate injections ran over 4 days period.

Figure 4. Qualification of lectin-protein biomarker candidates in an independent patient cohort. (A to F) Boxplots and ROC curves of top biomarker candidate for healthy vs BE, BE vs EAC, and healthy vs EAC comparison, respectively. **(G)** Overlap between lectin-protein candidates that differentiate BE from healthy, EAC from BE, and EAC from healthy phenotype. *P* values were calculated using Kruskal-Wallis test and $P < 0.05$ was considered to be statistically significant.

Figure 5. Assessing effect of confounding covariates on the top 3 biomarker candidates. Levels of NPL-APOB, AAL-C9 and EPHA-GSN were monitored in 39 serum samples (healthy-20 and population control-19) using MRM-MS. Samples were categorized according to **(A)** reflux frequency, **(B)** BMI, **(C)** smoking history and **(D)** alcohol consumption. $P < 0.05$ using Kruskal-Wallis test was considered to be statistically significant. Out of the top 3 candidates, only NPL-APOB was significantly different according BMI categorization.

Figure 6. Multimarker panel to distinguish EAC from BE. (A) sPLS-DA and **(B)** ROC curve analysis of a multimarker panel consists of AAL-C9, EPHA-A1BG, EPHA-C9, JAC-C9, NPL-C2, NPL-C4B, PSA-C9, and WGA-C9.

Tables

Table 1. Clinical characteristics of the patient cohort for biomarker discovery. For categorical and numerical variables, *P* values were calculated using Fisher's exact test and Kruskal-Wallis test respectively.

Variables	Healthy	BE	EAC	*P* value (Healthy vs BE vs EAC)
Sample size	9	10	10	
Age (Median ± SD)	66 ± 10	62 ± 15	66 ± 8	0.9311
Gender	All male	All male	All male	
Protein concentration (μg/μl)	95 ± 19	85 ± 13	82 ± 13	0.3641
Gastritis*	1 (11.1%)	1 (11.1%)	1 (10.0%)	1.0000
Peptic ulcer	3 (33.3%)	2 (20.0%)	3 (30.0%)	0.8792
Hiatus hernia	0 (0.0%)	4 (40.0%)	6 (60.0%)	0.0217
Other malignancy	1 (11.1%)	2 (20.0%)	2 (20.0%)	1.0000

*All the analyses were performed based on available patient information. Gastritis status for one BE patient was missing.

Table 2. Clinical characteristics of the patient cohort for biomarker qualification. For categorical and numerical variables, *P* values were calculated using Fisher's exact test and Kruskal-Wallis test respectively.

Variables	Healthy	BE	EAC	*P* value (Healthy vs BE vs EAC)	Population Control	*P* value (Healthy vs Pop. Control)
Sample size	20	20	20		19	
Gender	All male	All male	All male		All male	
Age in years (Median ± SD)	64 ± 8	60 ± 8	61 ± 7	0.4283	62 ± 7	0.2793
Protein concentration (µg/µl)	83 ± 10	78 ± 12	85 ± 13	0.6486	89 ± 13	0.0785
Reflux frequency* (10 years before diagnosis)				0.0108		0.2155
<Once/month	9 (47.4%)	7 (35.0%)	2 (10.0%)		14 (73.7%)	
Monthly (few times/month)	6 (31.6%)	6 (30.0%)	3 (15.0%)		4 (21.1%)	
Weekly or daily	4 (21.0%)	7 (35.0%)	15 (75.0%)		1 (5.3%)	
Body mass index				0.0076		0.6090
Healthy (<25)	5 (25.0%)	5 (25.0%)	1 (5.0%)		7 (36.8%)	
Overweight (25-30)	8 (40.0%)	12 (60.0%)	5 (25.0%)		8 (42.1%)	
Obese (>=30)	7 (35.0%)	3 (15.0%)	14 (70.0%)		4 (21.1%)	
Smoking history				0.6116		0.7813
Never smoked	8 (40.0%)	8 (40.0%)	4 (20.0%)		7 (36.8%)	
1-29.9 pack per year	8 (40.0%)	9 (45.0%)	10 (50.0%)		6 (31.6%)	
30+ pack per year	4 (20.0%)	3 (15.0%)	6 (30.0%)		6 (31.6%)	
Alcohol consumption				0.6637		0.8379
<1 standard drink/week	3 (15.0%)	3 (15.0%)	1 (5.0%)		2 (10.5%)	
1-6 standard drink/week	3 (15.0%)	4 (20.0%)	6 (30.0%)		5 (26.3%)	
7-20 standard drink/week	8 (40.0%)	4 (20.0%)	6 (30.0%)		6 (31.6%)	
21+ standard drink/week	6 (30.0%)	9 (45.0%)	7 (35.0%)		6 (31.6%)	

*All the analyses were performed based on available patient information. Reflux frequency for one healthy patient was missing.

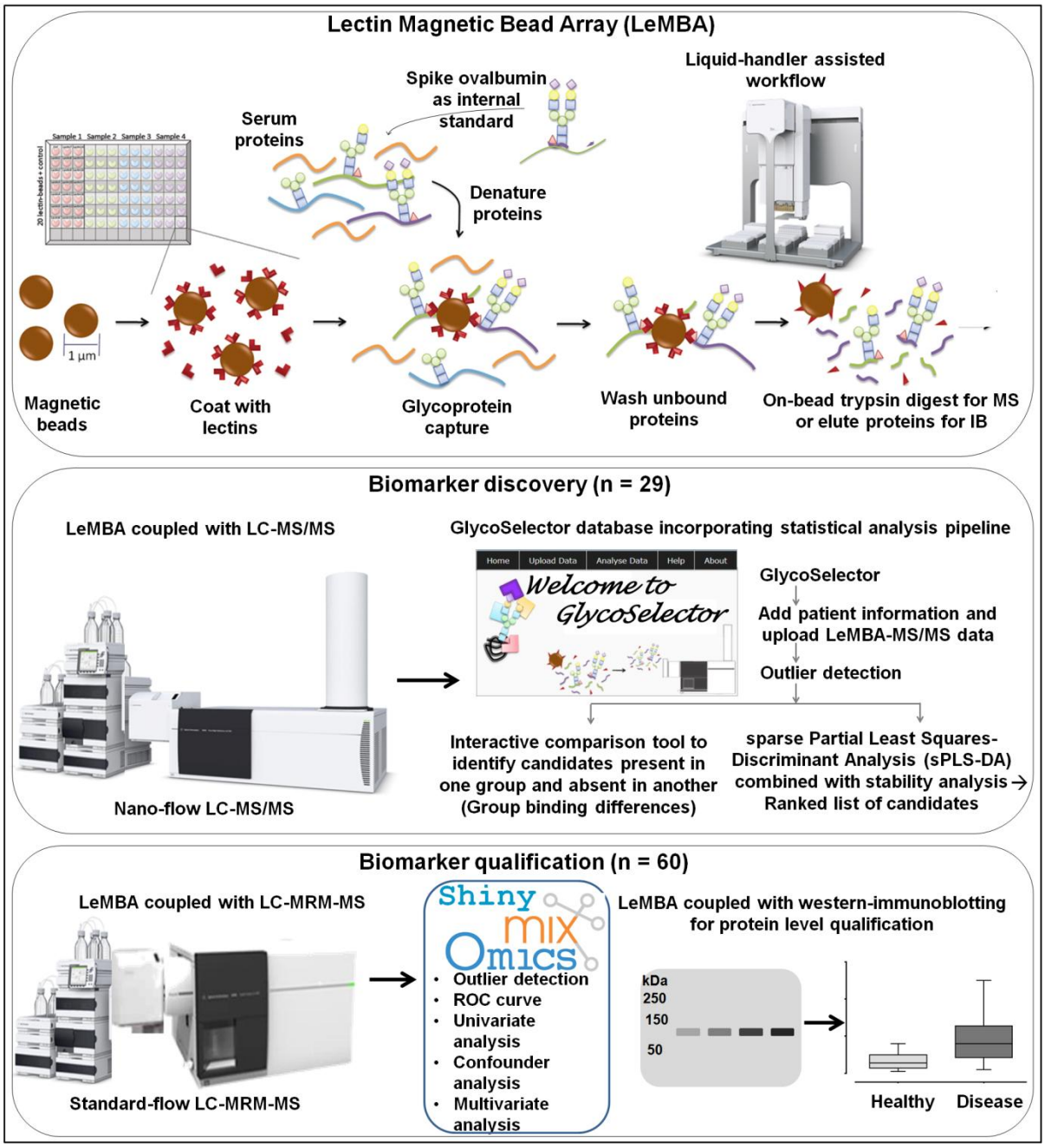
Table 3. Verified list of candidates shown by lectin affinity-protein ID that were significantly different for either healthy vs BE or BE vs EAC or healthy vs EAC analysis. Proteins are denoted using gene symbol; number in the bracket denotes Uniprot accession number. AUROC values of more than 0.7 are highlighted in bold.

Lectin-Protein	Healthy vs BE		BE vs EAC		Healthy vs EAC	
	***P* value**	**AUROC**	***P* value**	**AUROC**	***P* value**	**AUROC**
AAL-APOB (P04114)	0.1368	0.6375	0.0453	0.6850	0.9569	0.4950
AAL-C5 (P01031)	0.6073	0.5475	0.0483	0.6825	0.2340	0.6100
AAL-C7 (P10643)	0.2793	0.6000	**0.0063**	**0.7525**	0.3169	0.5925
AAL-C9 (P02748)	0.2793	0.6000	**0.0001**	**0.8525**	**0.0161**	**0.7225**
AAL-GSN (P06396)	0.7455	0.5300	**0.0087**	**0.7425**	**0.0265**	**0.7050**
AAL-HP (P00738)	0.8711	0.4850	0.0398	0.6900	0.0583	0.6750
EPHA-A2M (P01023)	**0.0248**	**0.7075**	0.9138	0.4900	**0.0186**	**0.7175**
EPHA-AHSG (P02765)	0.5162	0.5600	0.1941	0.6200	0.0483	0.6825
EPHA-C7 (P10643)	0.1368	0.6375	0.0398	0.6900	0.6849	0.5375
EPHA-C9 (P02748)	0.0583	0.6750	**0.0003**	**0.8375**	**0.0265**	**0.7050**
EPHA-GSN (P06396)	0.2036	0.6175	**0.0200**	**0.7150**	**0.0014**	**0.7950**
EPHA-HP (P00738)	0.7455	0.5300	**0.0200**	**0.7150**	0.0305	0.7000
EPHA-SERPINA3 (P01011)	0.4171	0.5750	**0.0265**	**0.7050**	0.0620	0.6725
EPHA-TF (P02787)	0.7455	0.4700	0.0326	0.6975	0.0935	0.6550
JAC-A1BG (P04217)	0.6263	0.5450	0.0483	0.6825	0.1231	0.6425
JAC-APOB (P04114)	0.0305	0.7000	0.0699	0.6675	0.5700	0.5525
JAC-C4BPA (P04003)	0.7251	0.5325	0.0935	0.6550	**0.0128**	**0.7300**
JAC-C5 (P01031)	0.6073	0.4525	0.0425	0.6875	0.0483	0.6825
JAC-C7 (P10643)	0.2914	0.5975	**0.0094**	**0.7400**	0.0834	0.6600
JAC-C9 (P02748)	0.2914	0.5975	**0.0007**	**0.8125**	**0.0029**	**0.7750**
JAC-CFB (P00751)	0.9353	0.5075	0.0373	0.6925	0.0373	0.6925
JAC-GSN (P06396)	0.8498	0.5175	0.0305	0.7000	**0.0215**	**0.7125**
JAC-HP (P00738)	0.9569	0.5050	0.0483	0.6825	0.0583	0.6750
JAC-HPX (P02790)	0.7868	0.5250	0.0742	0.6650	**0.0200**	**0.7150**
JAC-SERPINA1 (P01009)	0.3040	0.5950	0.0453	0.6850	0.2448	0.6075
JAC-SERPINA3 (P01011)	0.9569	0.5050	**0.0102**	**0.7375**	0.0305	0.7000
JAC-SERPIND1 (P05546)	0.1368	0.6375	0.4819	0.5650	0.0483	0.6825
JAC-SERPING1 (P05155)	0.5518	0.5550	0.2559	0.6050	**0.0200**	**0.7150**
NPL-AFM (P43652)	0.5338	0.5575	0.0483	0.6825	0.1762	0.6250

Lectin-Protein	Healthy vs BE		BE vs EAC		Healthy vs EAC	
	***P* value**	**AUROC**	***P* value**	**AUROC**	***P* value**	**AUROC**
NPL-APOB (P04114)	**0.0231**	**0.7100**	**0.0231**	**0.7100**	0.8924	0.5125
NPL-C4BPA (P04003)	0.0989	0.6525	0.6849	0.5375	**0.0231**	**0.7100**
NPL-C9 (P02748)	0.5885	0.5500	**0.0049**	**0.7600**	**0.0074**	**0.7475**
NPL-GSN (P06396)	0.8924	0.5125	**0.0173**	**0.7200**	0.0583	0.6750
NPL-HP (P00738)	0.8077	0.5225	0.0884	0.6575	0.0326	0.6975
NPL-SERPINA3 (P01011)	0.5518	0.5550	0.0989	0.6525	0.0305	0.7000
PSA-C5 (P01031)	0.4017	0.5775	0.0453	0.6850	0.3040	0.5950
PSA-C7 (P10643)	0.2914	0.5975	**0.0019**	**0.7875**	0.0742	0.6650
PSA-C9 (P02748)	0.2036	0.6175	**0.0008**	**0.8100**	**0.0161**	**0.7225**
PSA-GSN (P06396)	0.3577	0.5850	0.0483	0.6825	**0.0110**	**0.7350**
PSA-HP (P00738)	0.8498	0.4825	0.0483	0.6825	0.0425	0.6875
PSA-SERPINA3 (P01011)	0.8077	0.5225	0.0425	0.6875	0.0834	0.6600
WGA-C9 (P02748)	0.4819	0.5650	**0.0032**	**0.7725**	**0.0053**	**0.7575**
WGA-GSN (P06396)	0.7868	0.5250	**0.0119**	**0.7325**	0.0742	0.6650
WGA-HP (P00738)	0.7455	0.5300	0.0483	0.6825	**0.0215**	**0.7125**
WGA-SERPINA3 (P01011)	0.4819	0.5650	0.0989	0.6525	**0.0063**	**0.7525**

Figures

Figure 1



ASBMB
MOLECULAR & CELLULAR PROTEOMICS
MCP

Figure 2

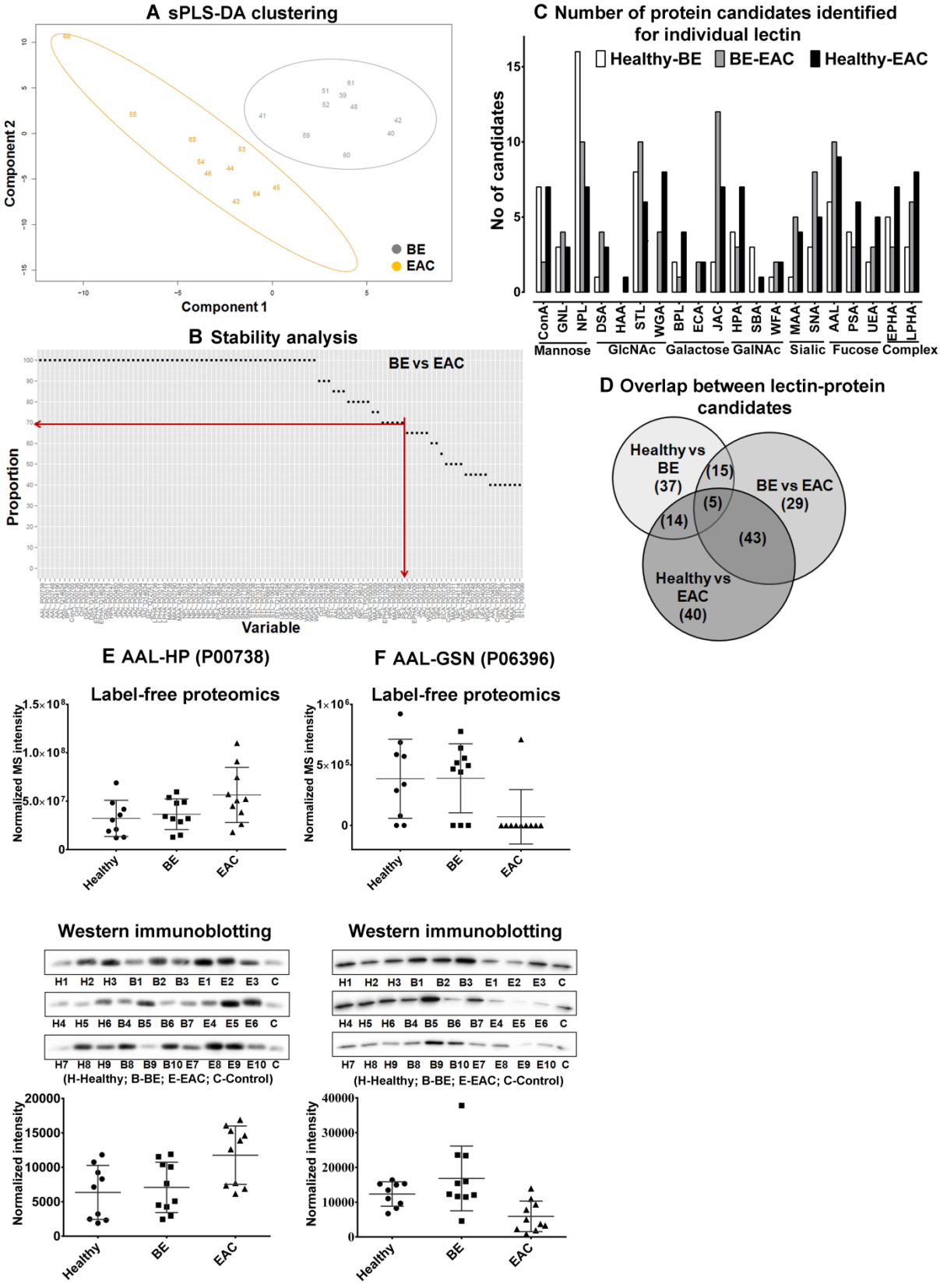


Figure 3

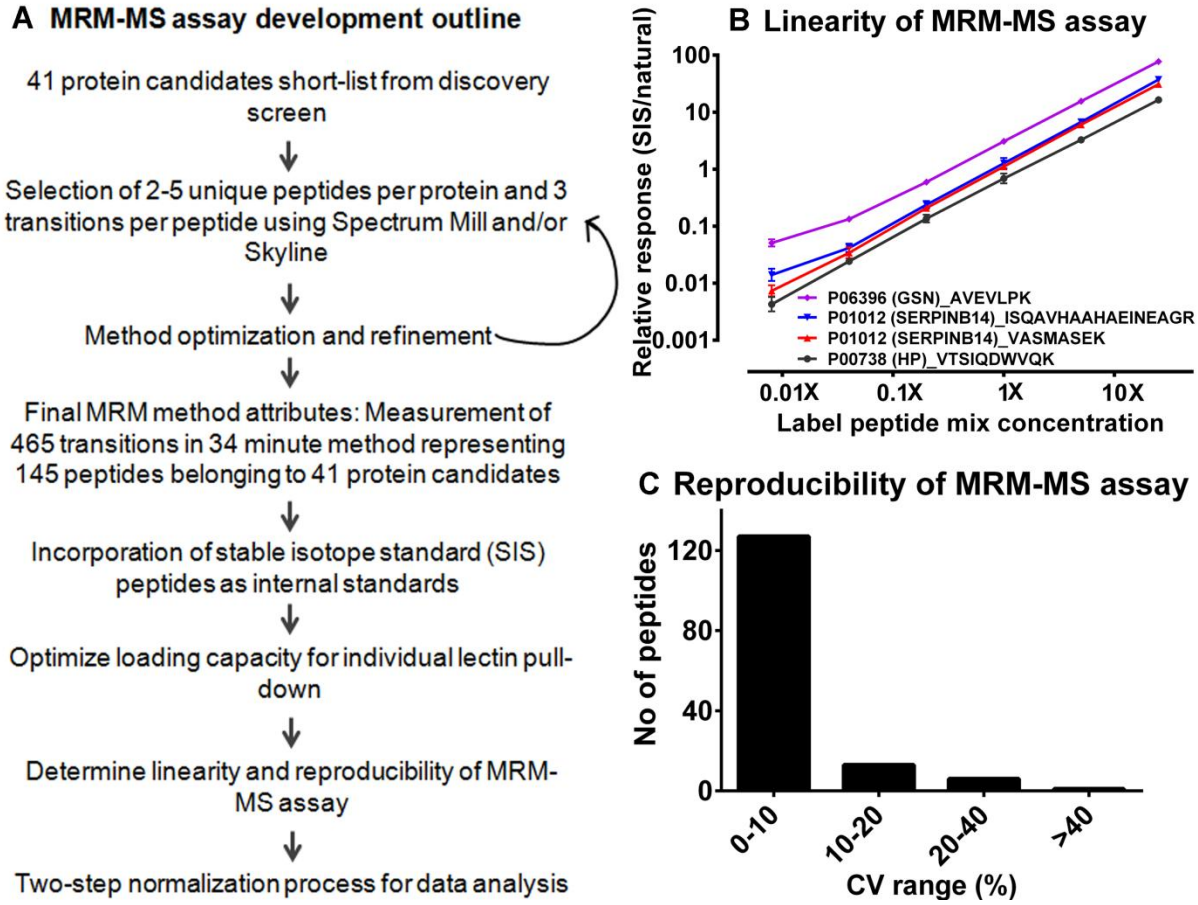


Figure 4

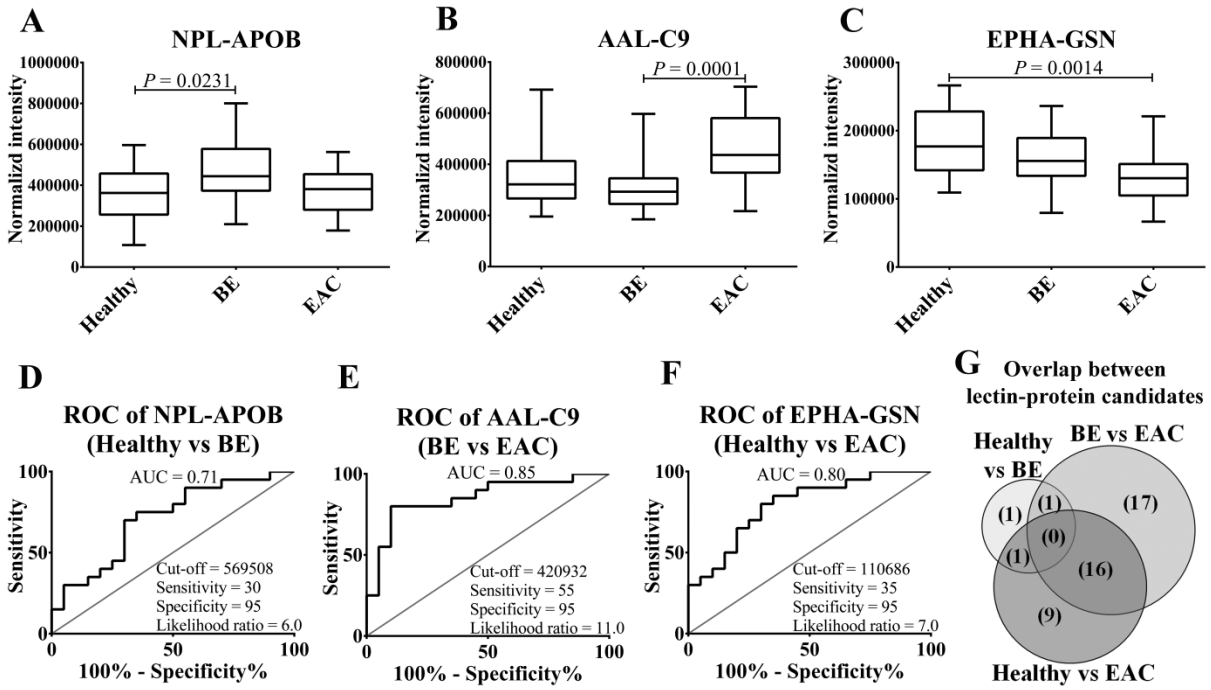


Figure 5

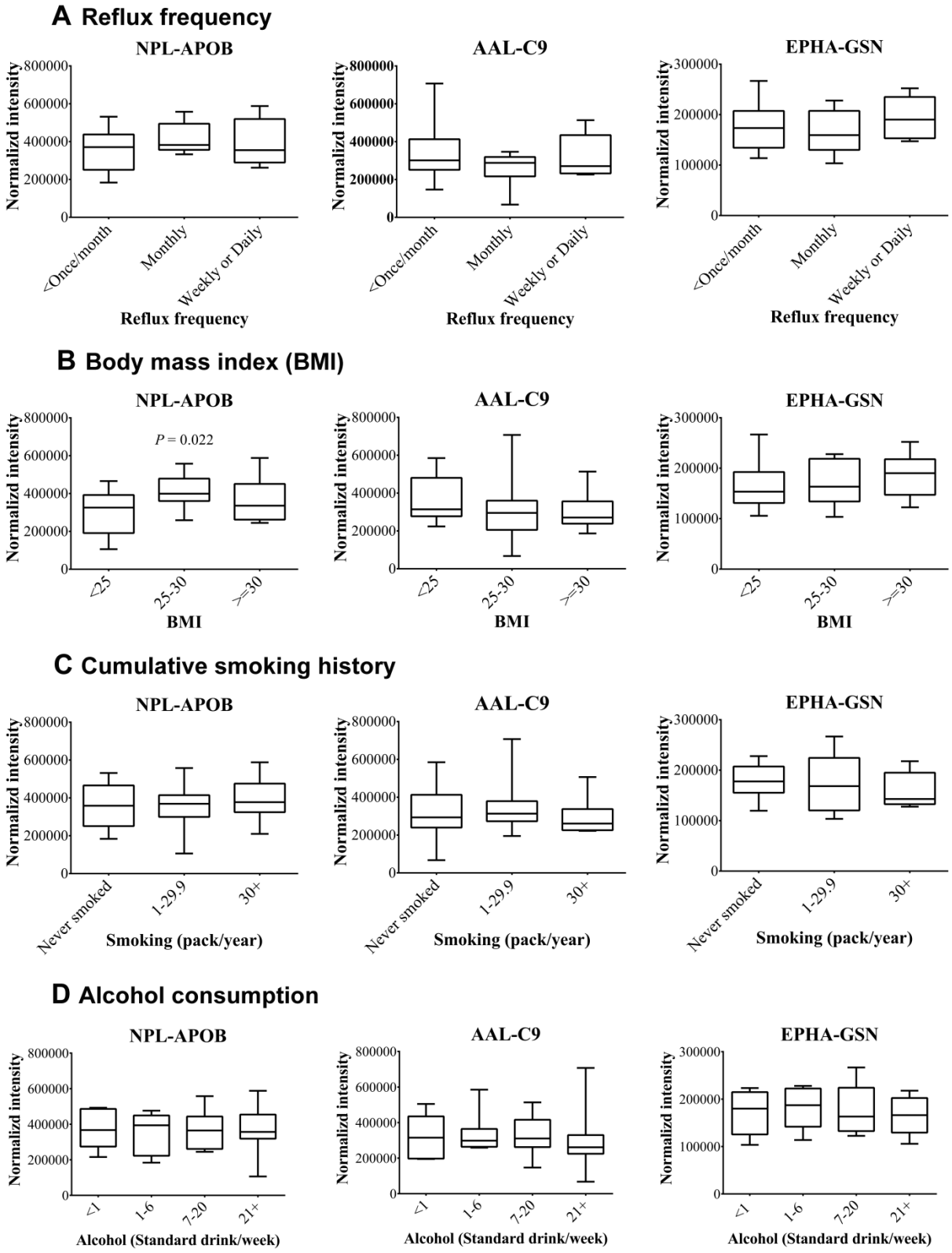
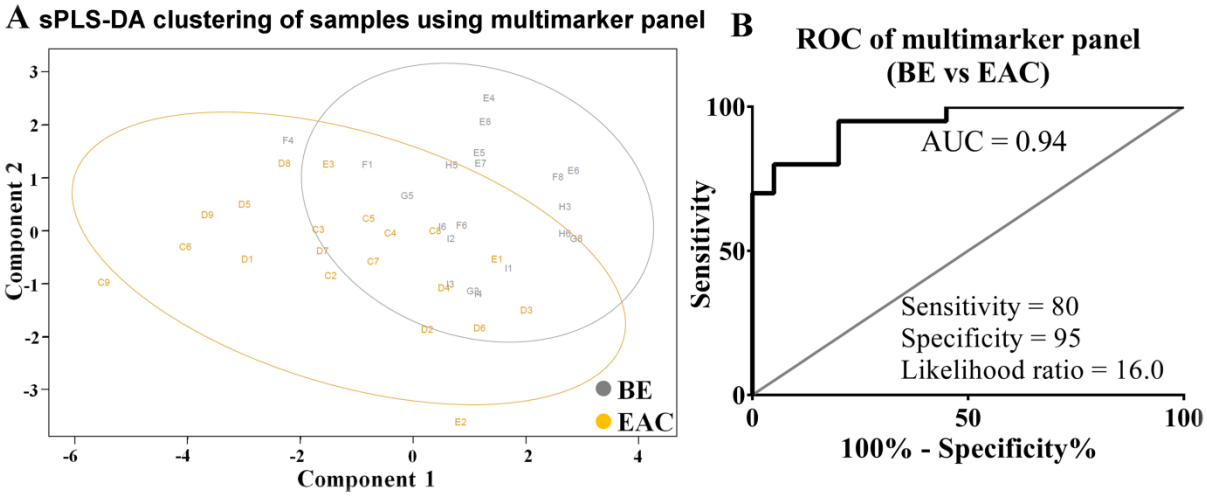
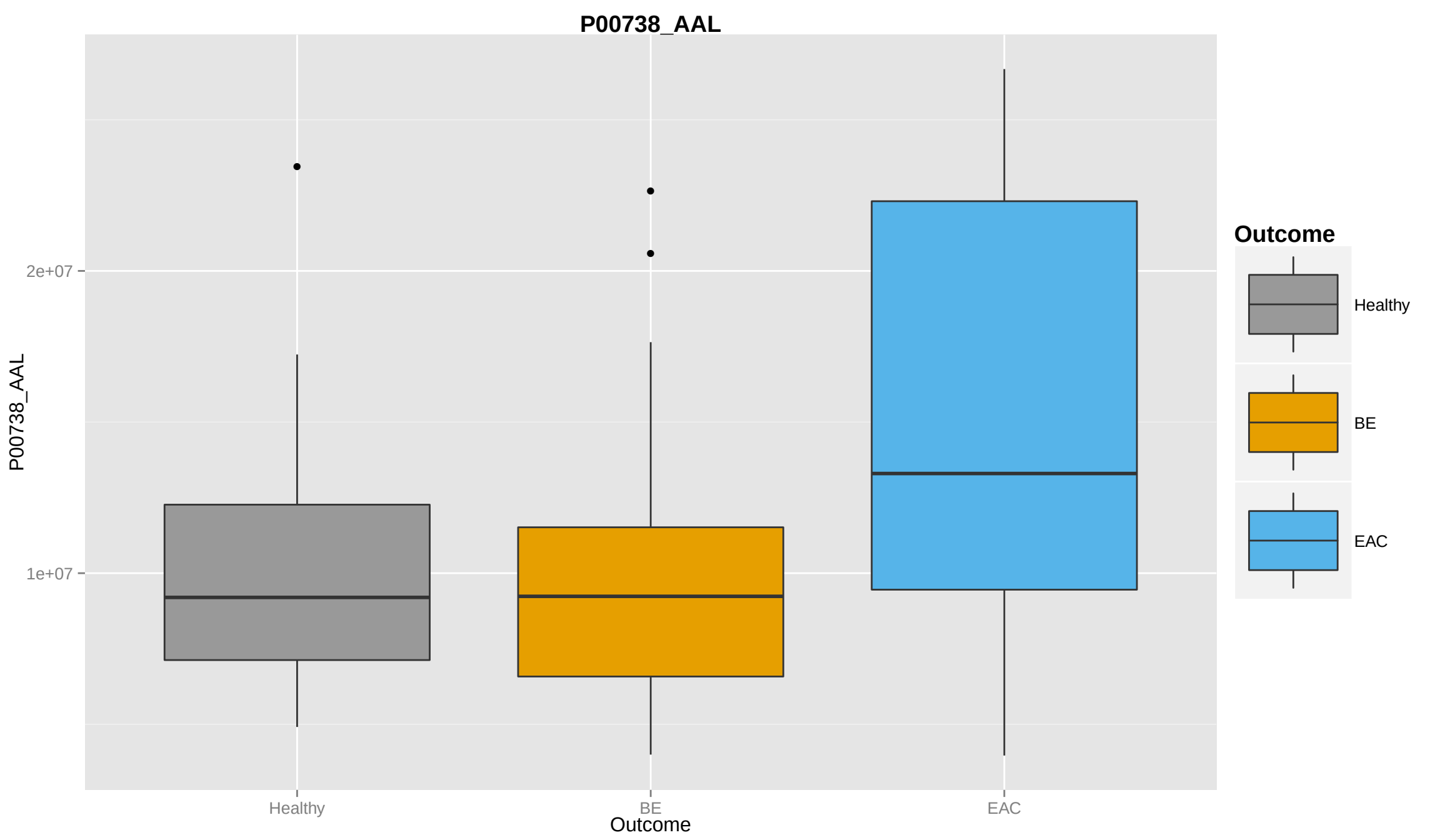
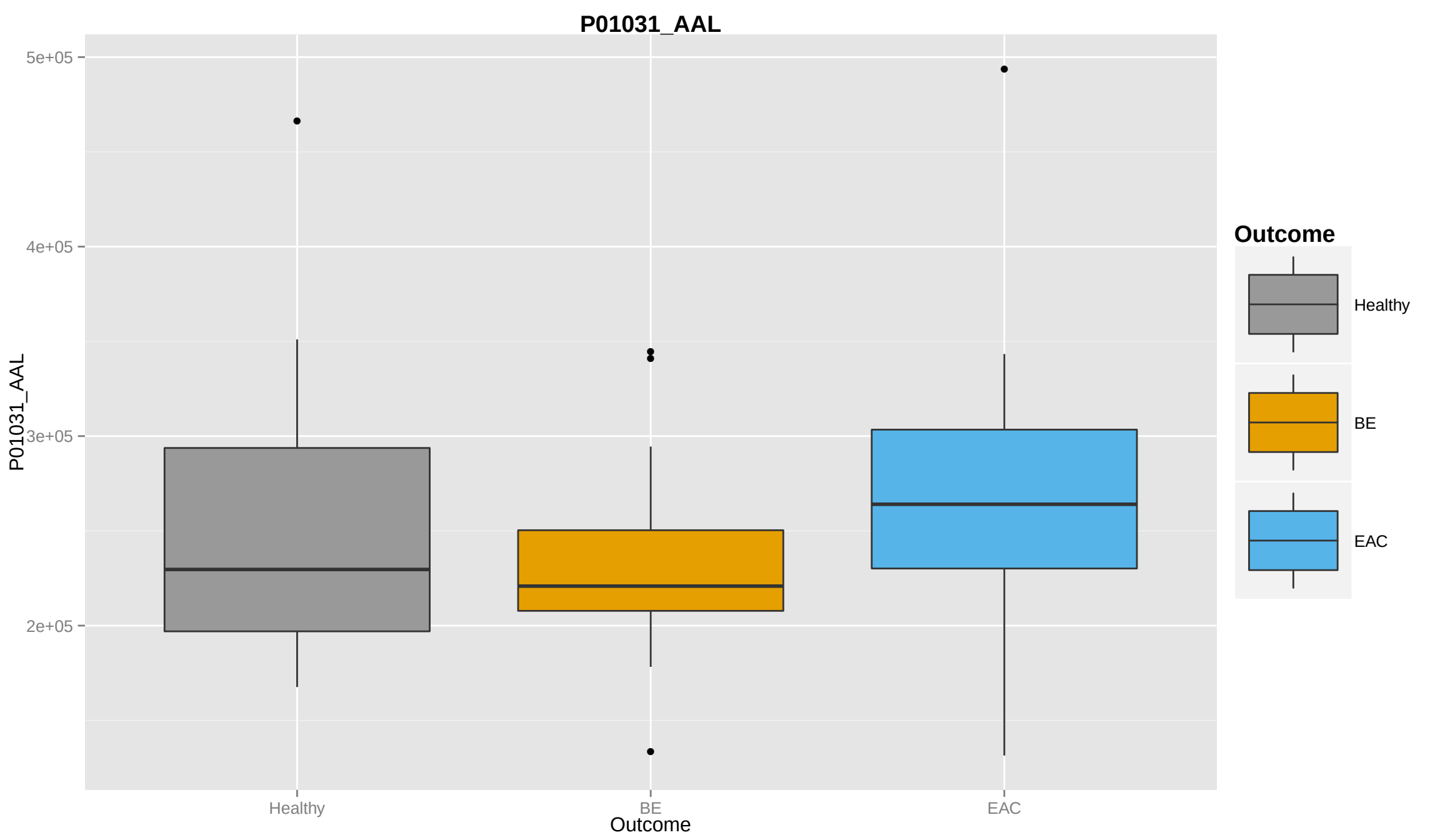

A Reflux frequency
NPL-APOB
AAL-C9
EPHA-GSN
Normalizd intensity
<Once/month
Monthly
Weekly or Daily
Reflux frequency
B Body mass index (BMI)
P = 0.022
<25
25-30
>=30
BMI
C Cumulative smoking history
Never smoked
1-29.9
30+
Smoking (pack/year)
D Alcohol consumption
<1
1-6
7-20
21+
Alcohol (Standard drink/week)

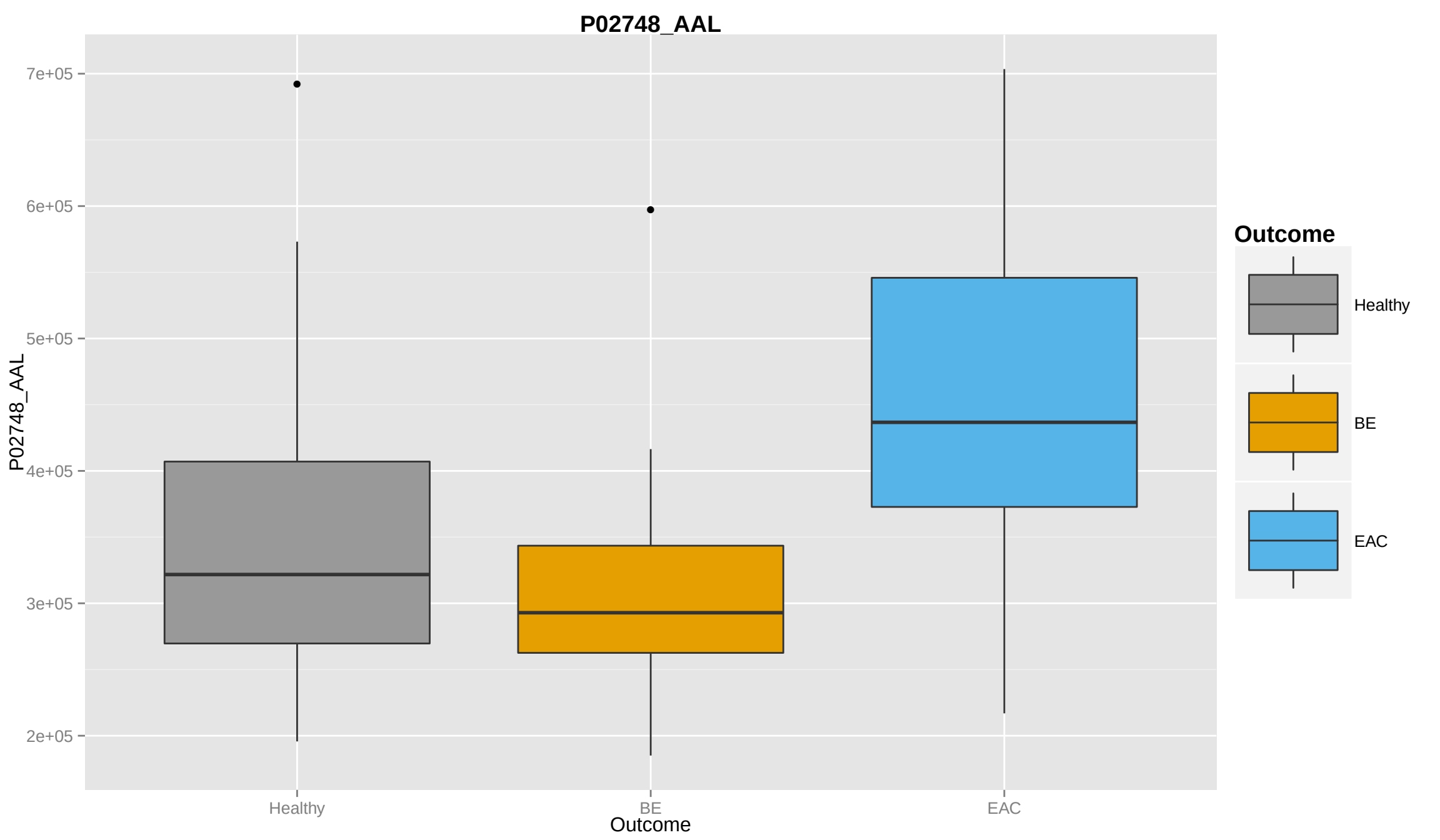
Figure 6

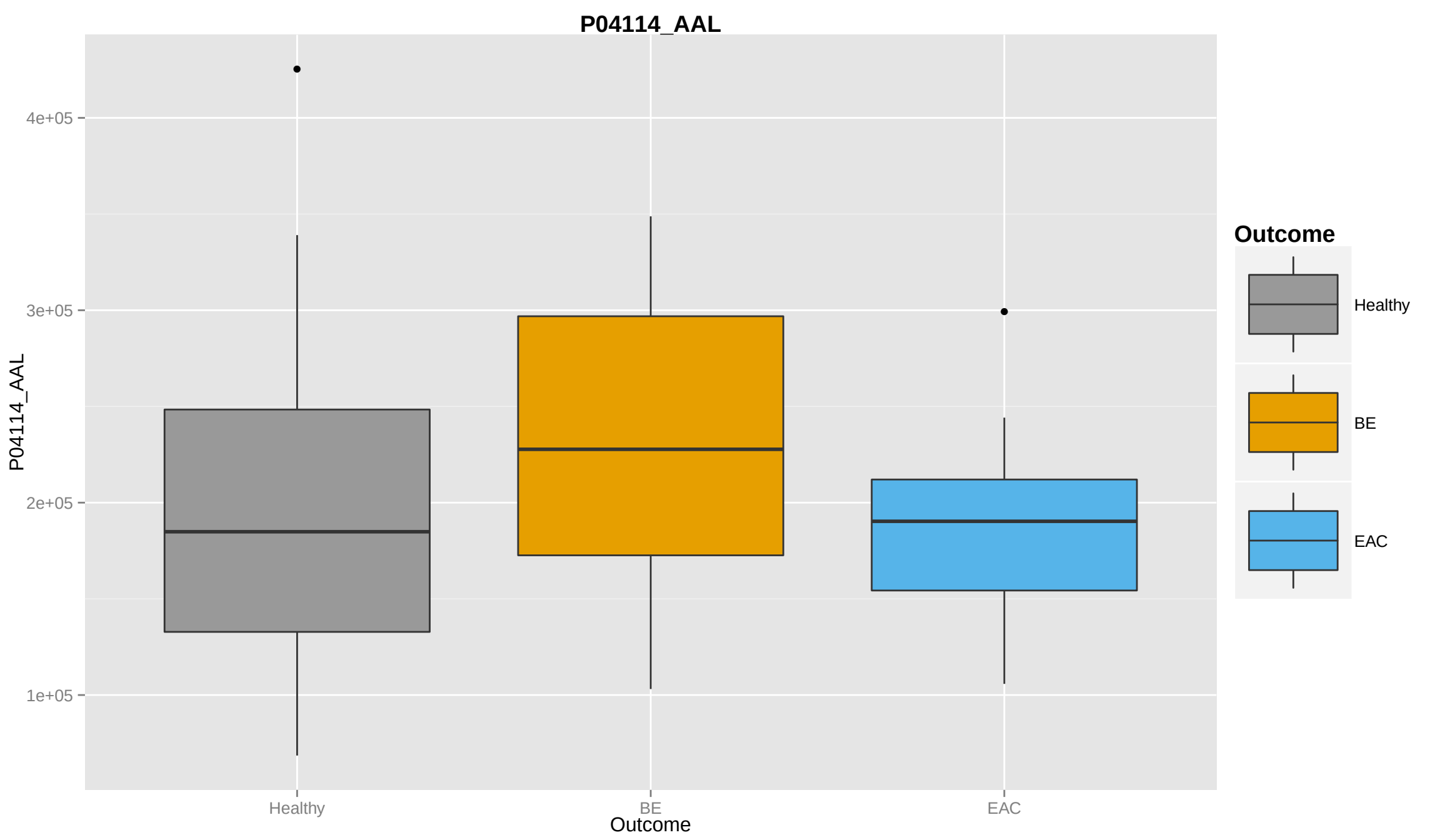


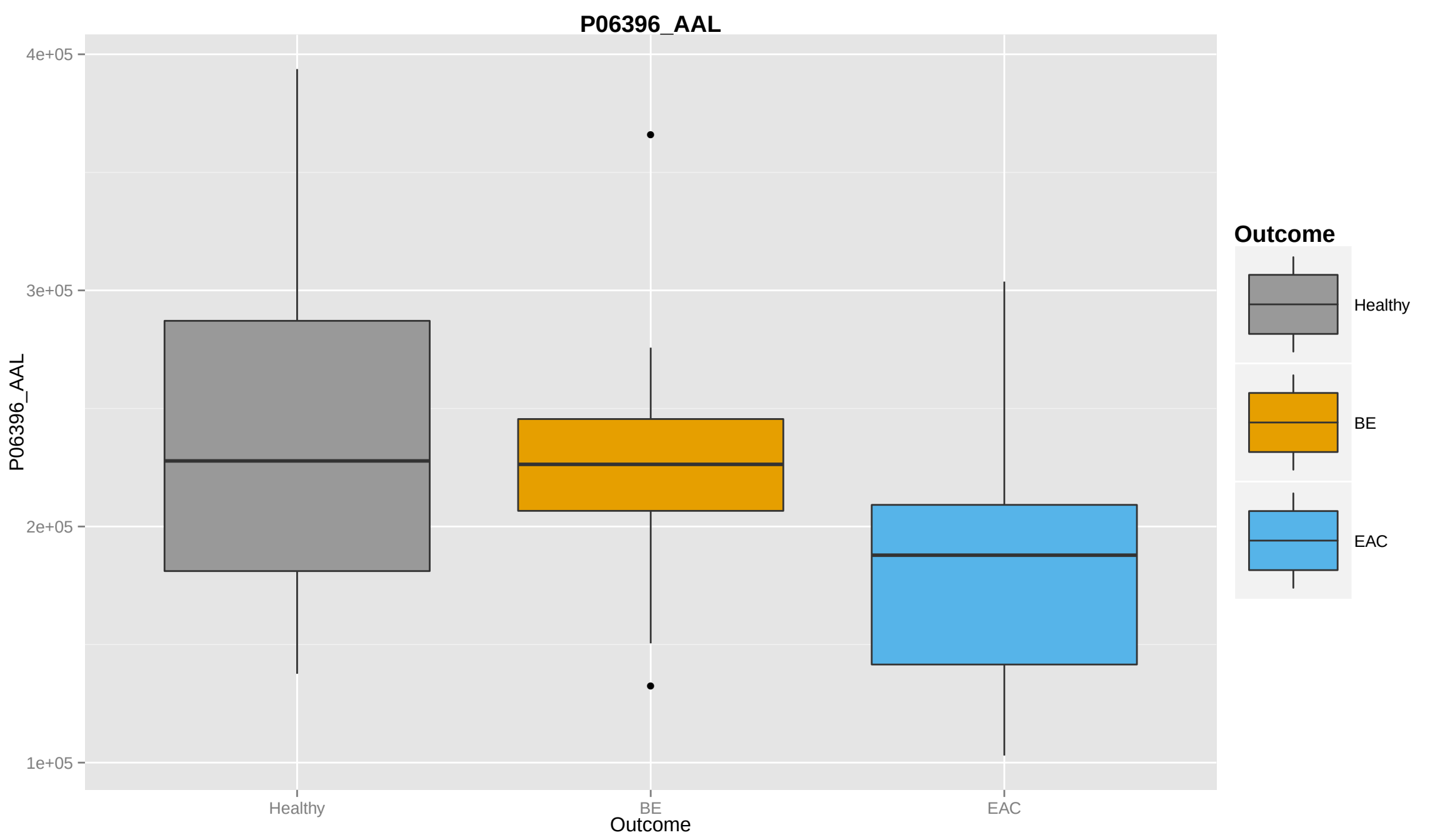
9.3 Appendix III: Boxplots and ROC curves of the verified candidates

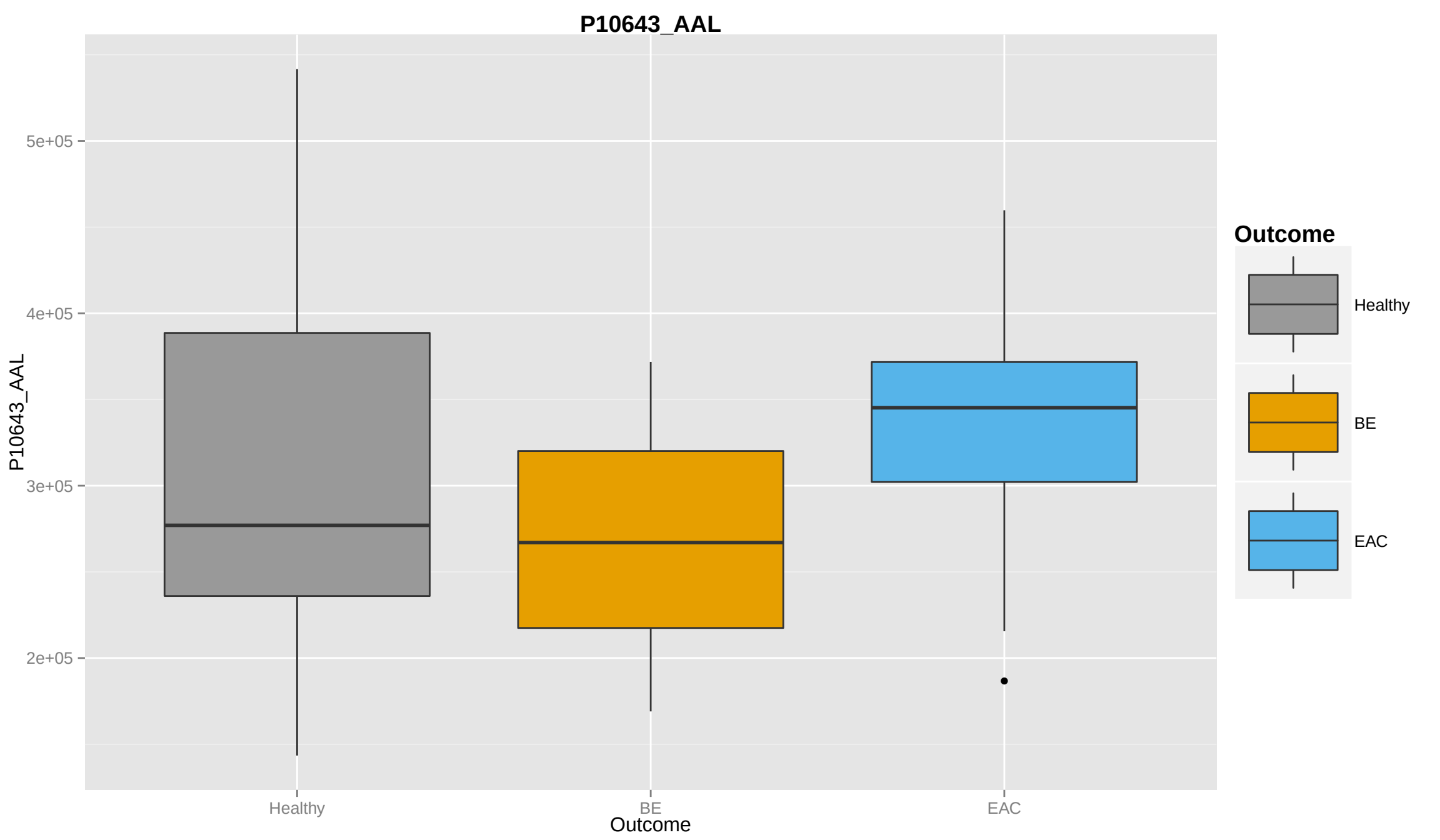

P00738_AAL
2e+07
1e+07
P00738_AAL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

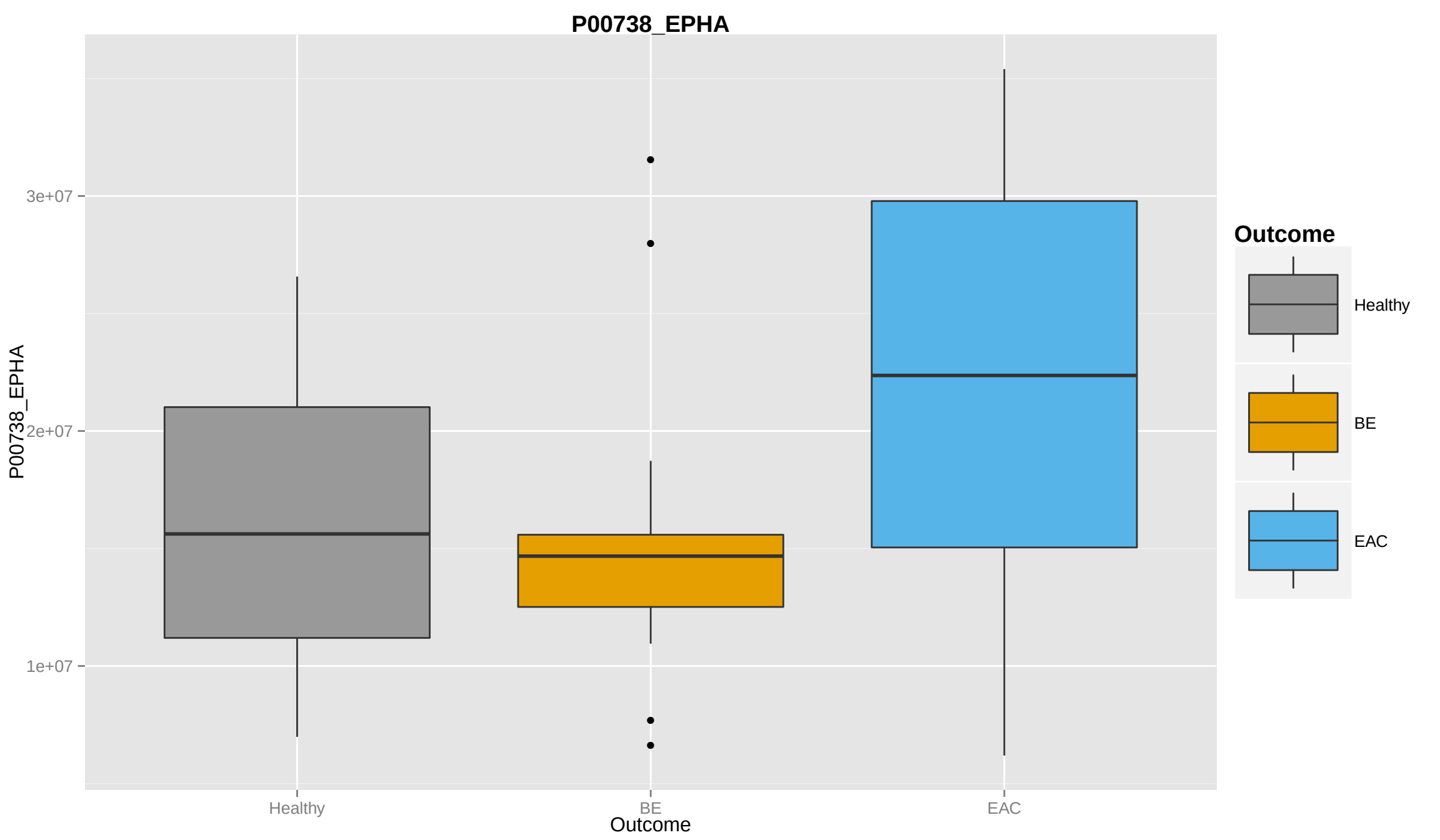

P01031_AAL
5e+05
4e+05
3e+05
2e+05
P01031_AAL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

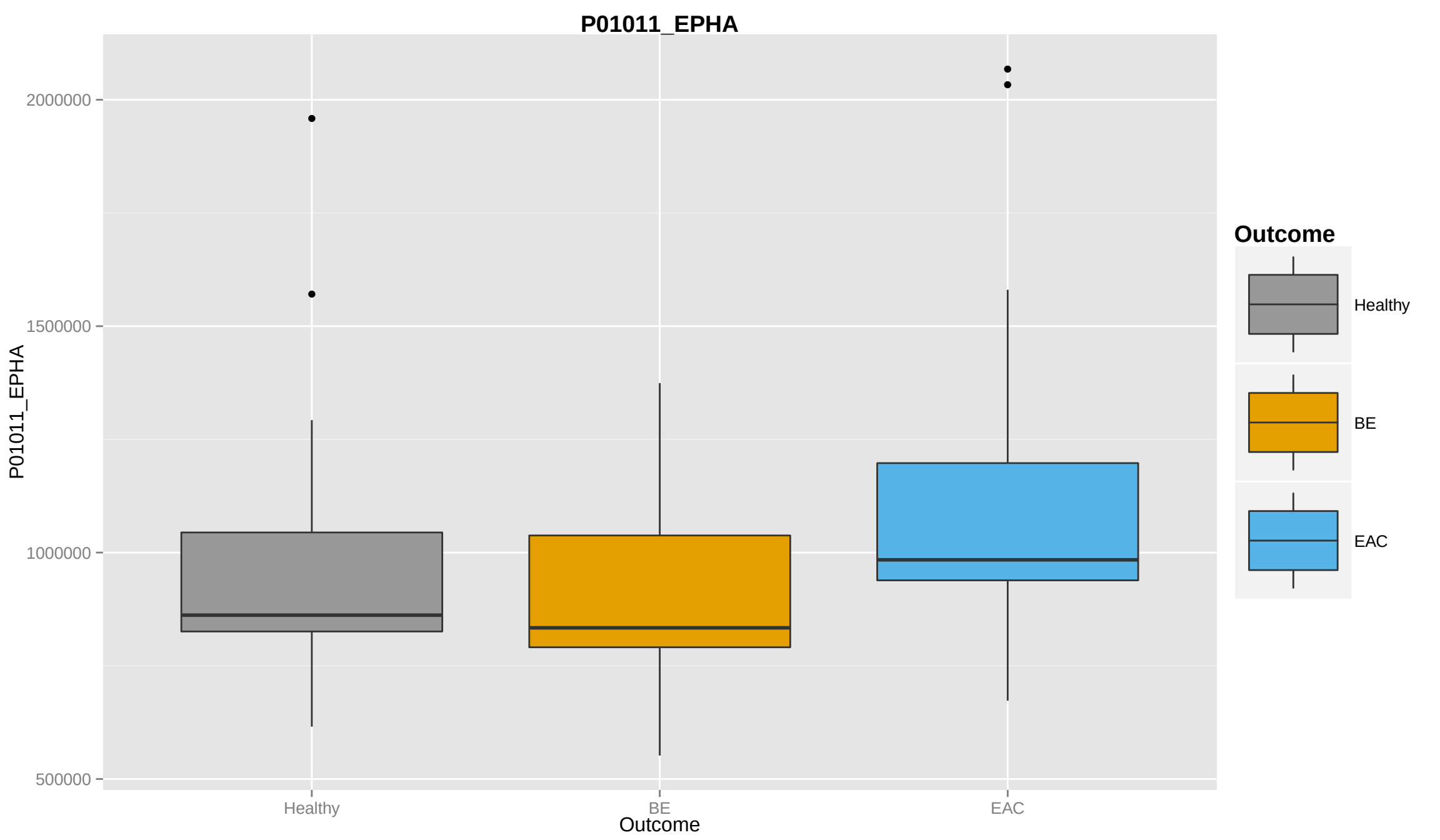

P02748_AAL
7e+05
6e+05
5e+05
4e+05
3e+05
2e+05
P02748_AAL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

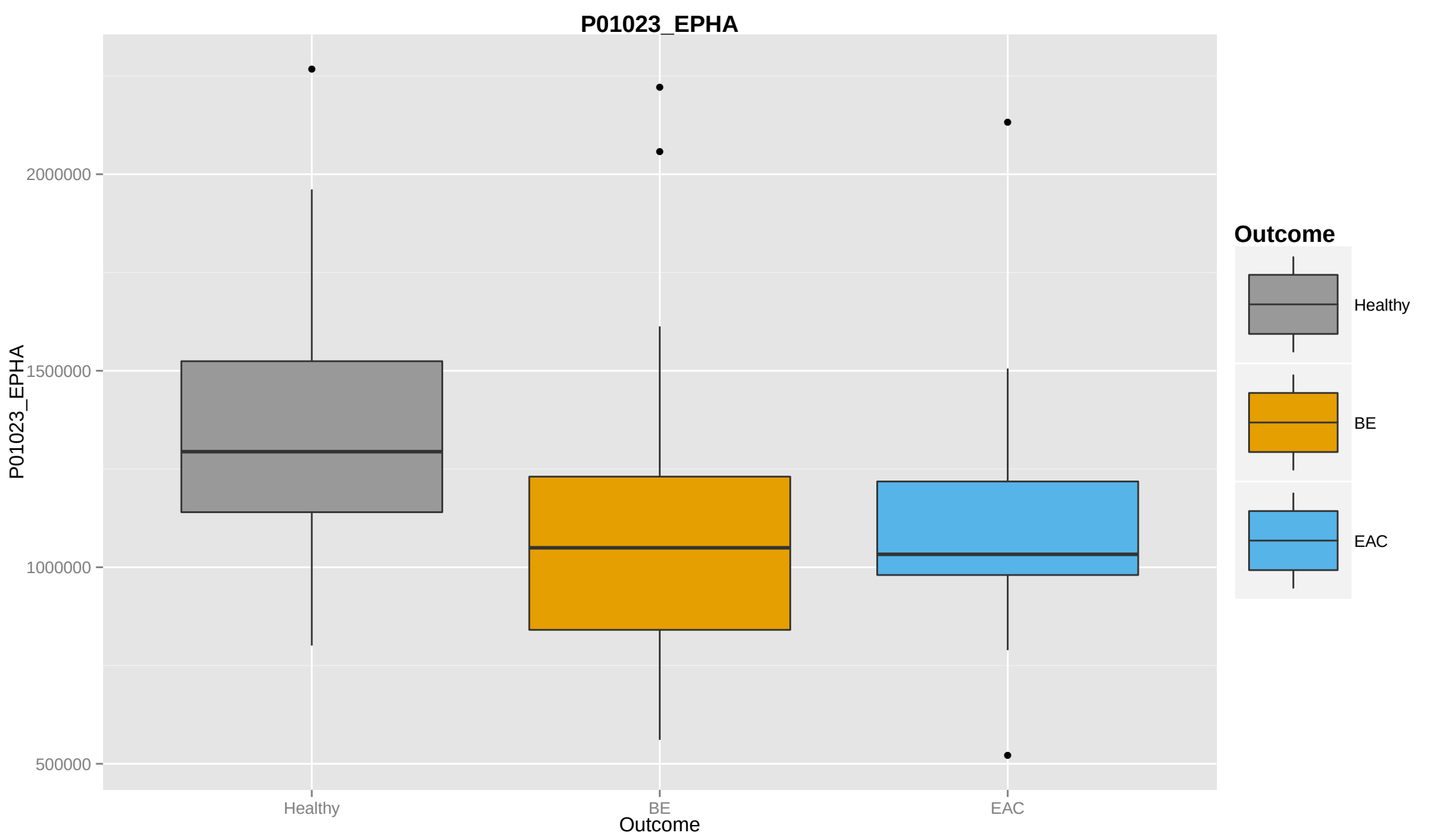

P04114_AAL
4e+05
3e+05
2e+05
1e+05
P04114_AAL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

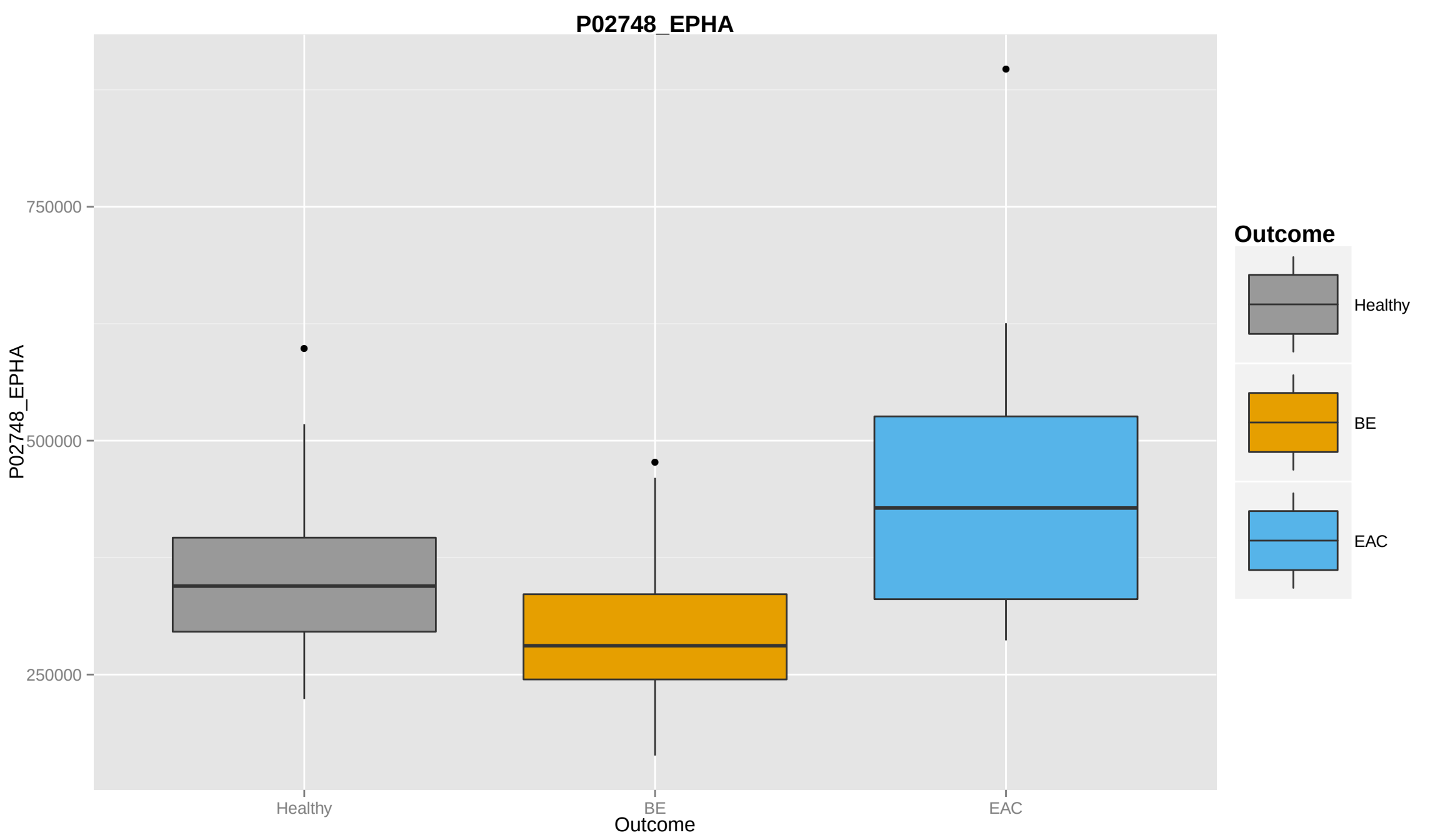

P06396_AAL
4e+05
3e+05
2e+05
1e+05
P06396_AAL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

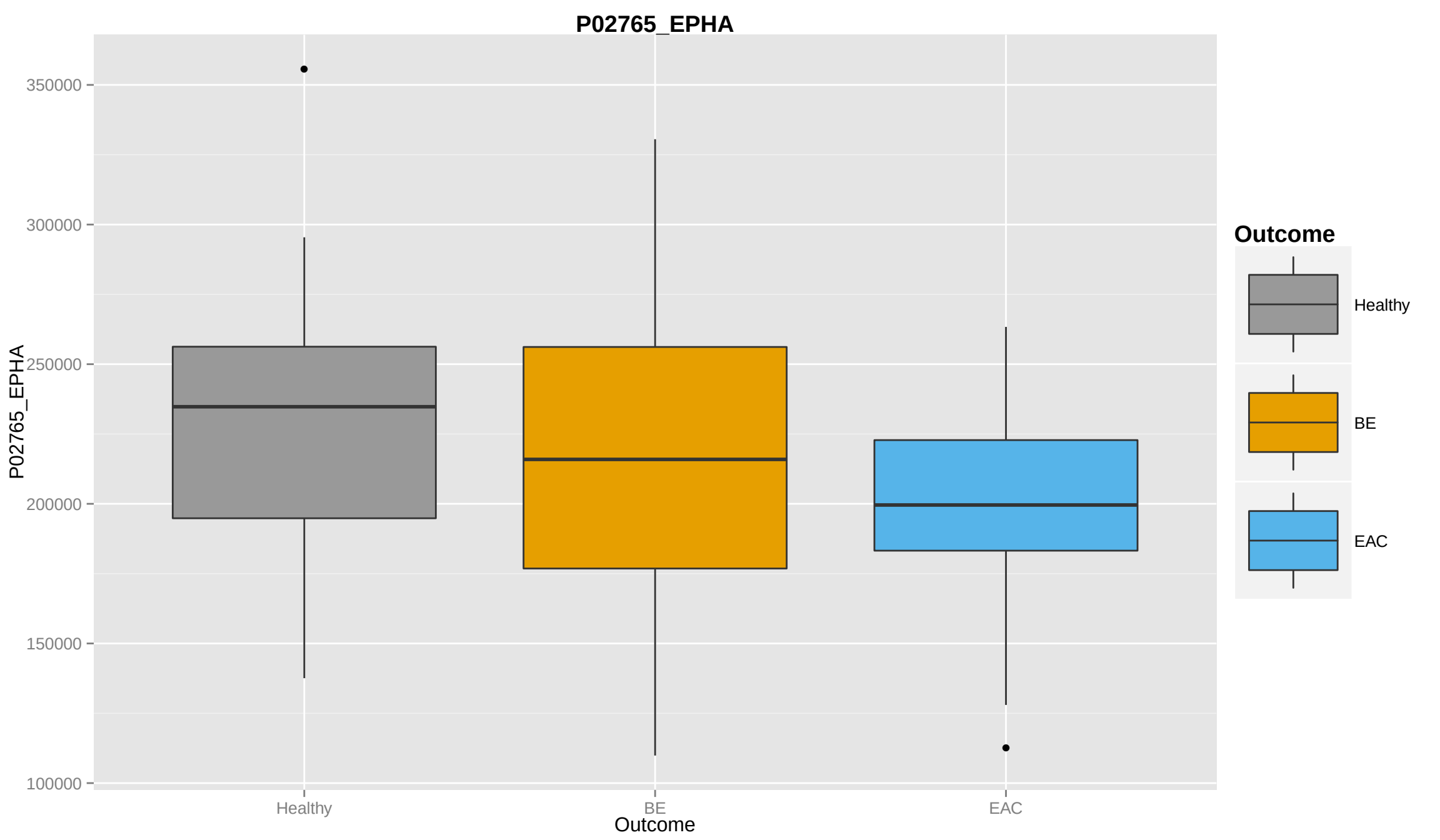

P10643_AAL
5e+05
4e+05
3e+05
2e+05
P10643_AAL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

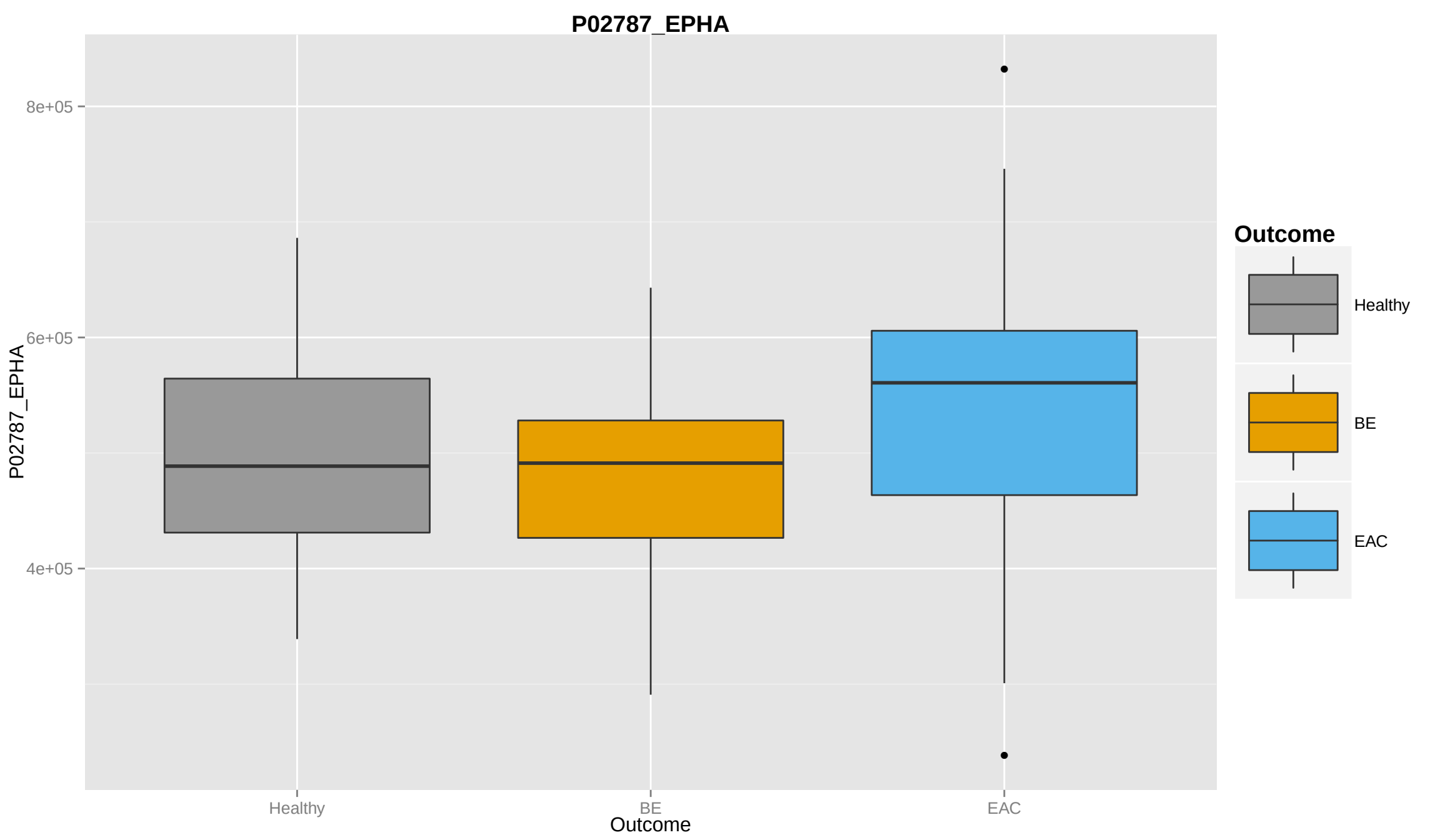

P00738_EPHA
3e+07
2e+07
1e+07
P00738_EPHA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

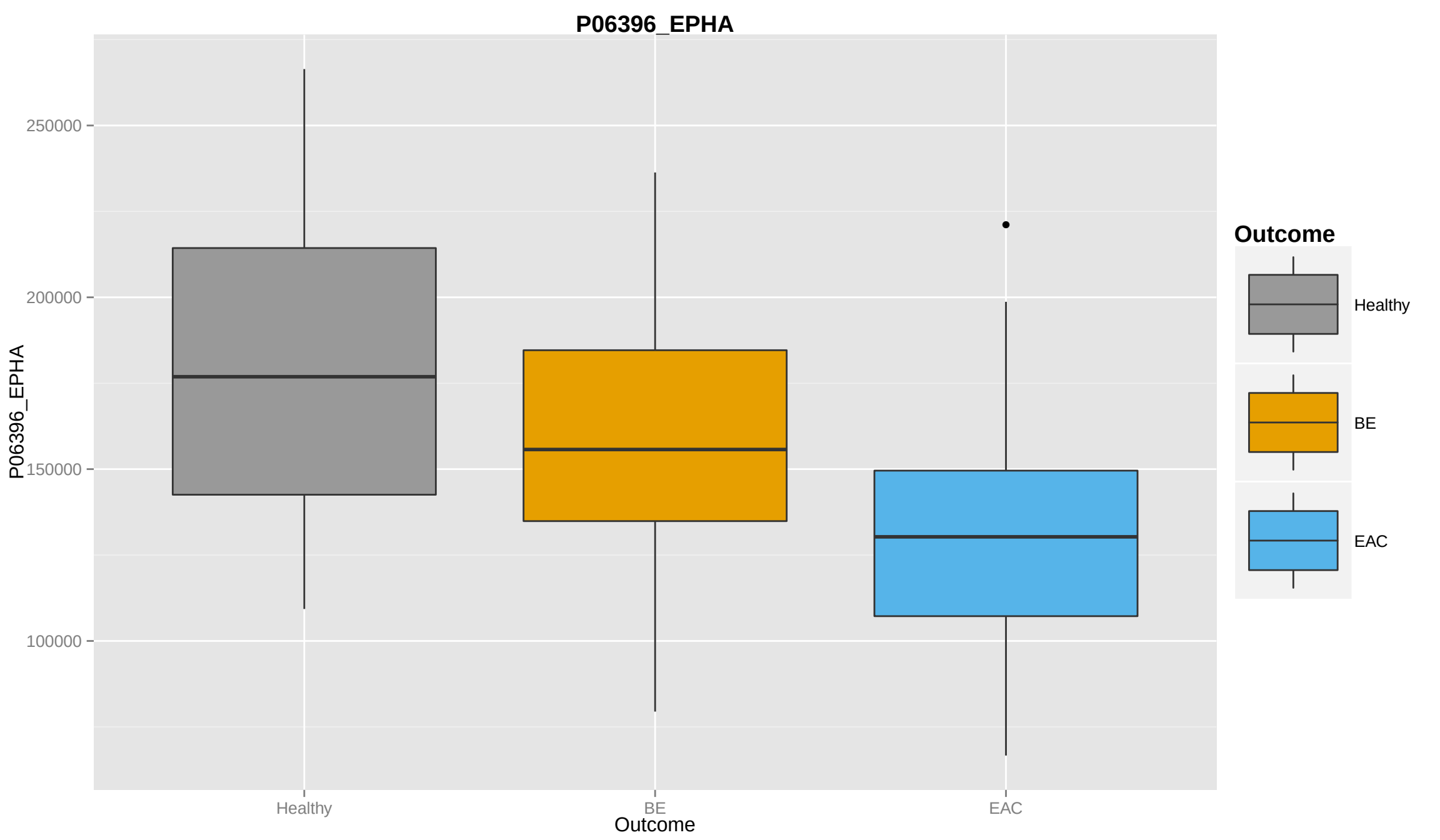

P01011_EPHA
2000000
1500000
1000000
500000
P01011_EPHA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

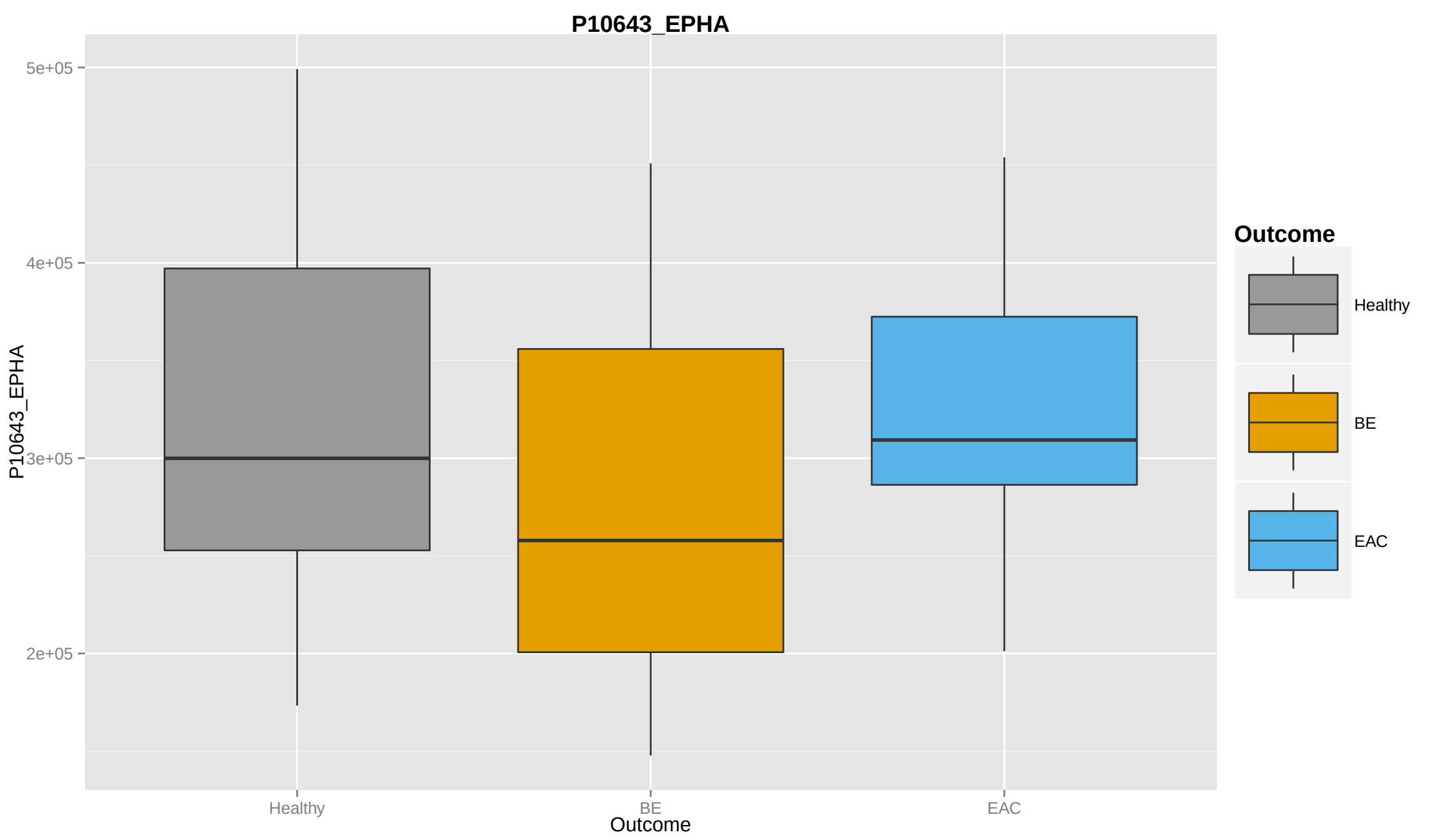

P01023_EPHA
2000000
1500000
1000000
500000
P01023_EPHA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

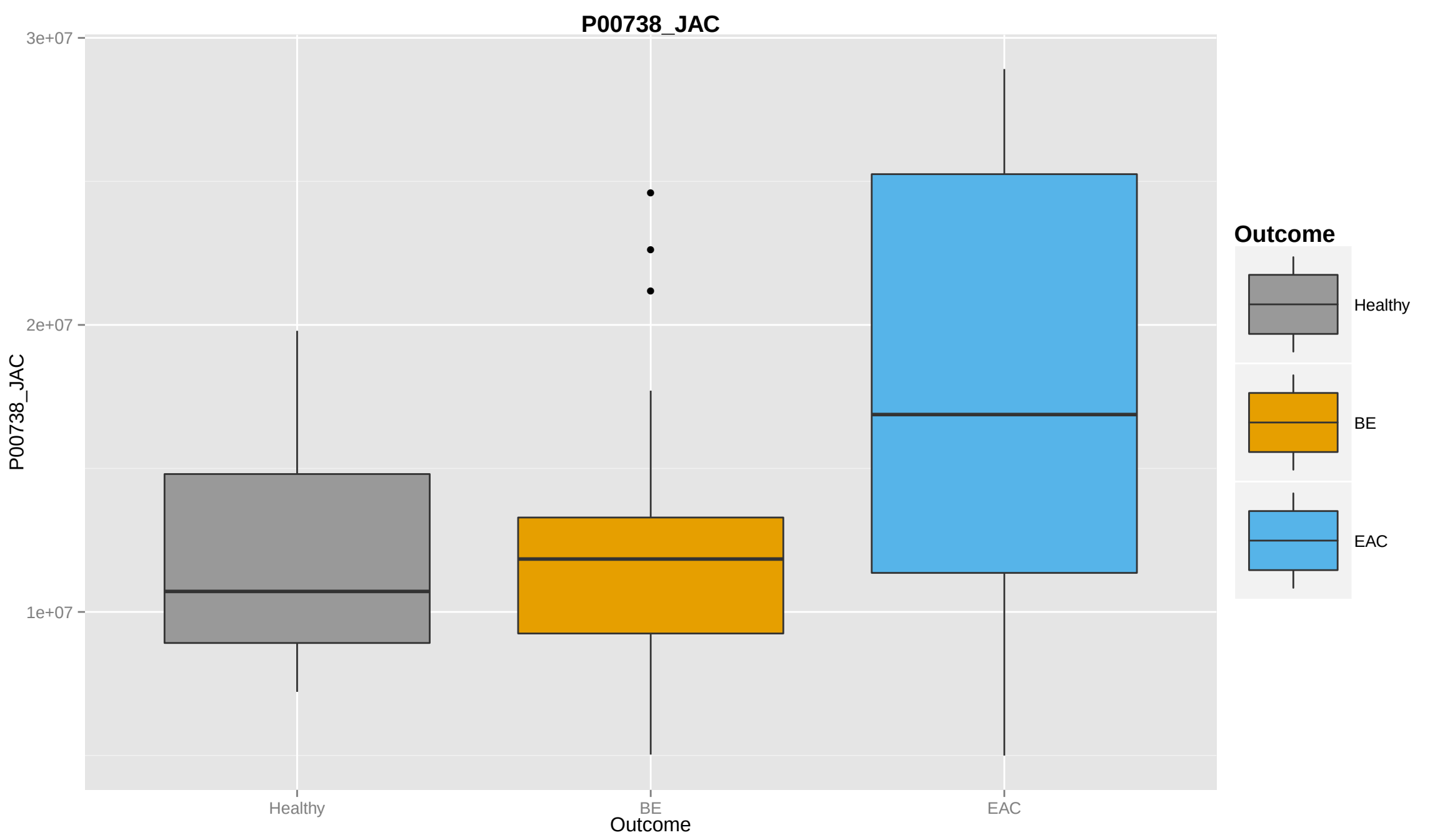

P02748_EPHA
750000
500000
250000
P02748_EPHA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

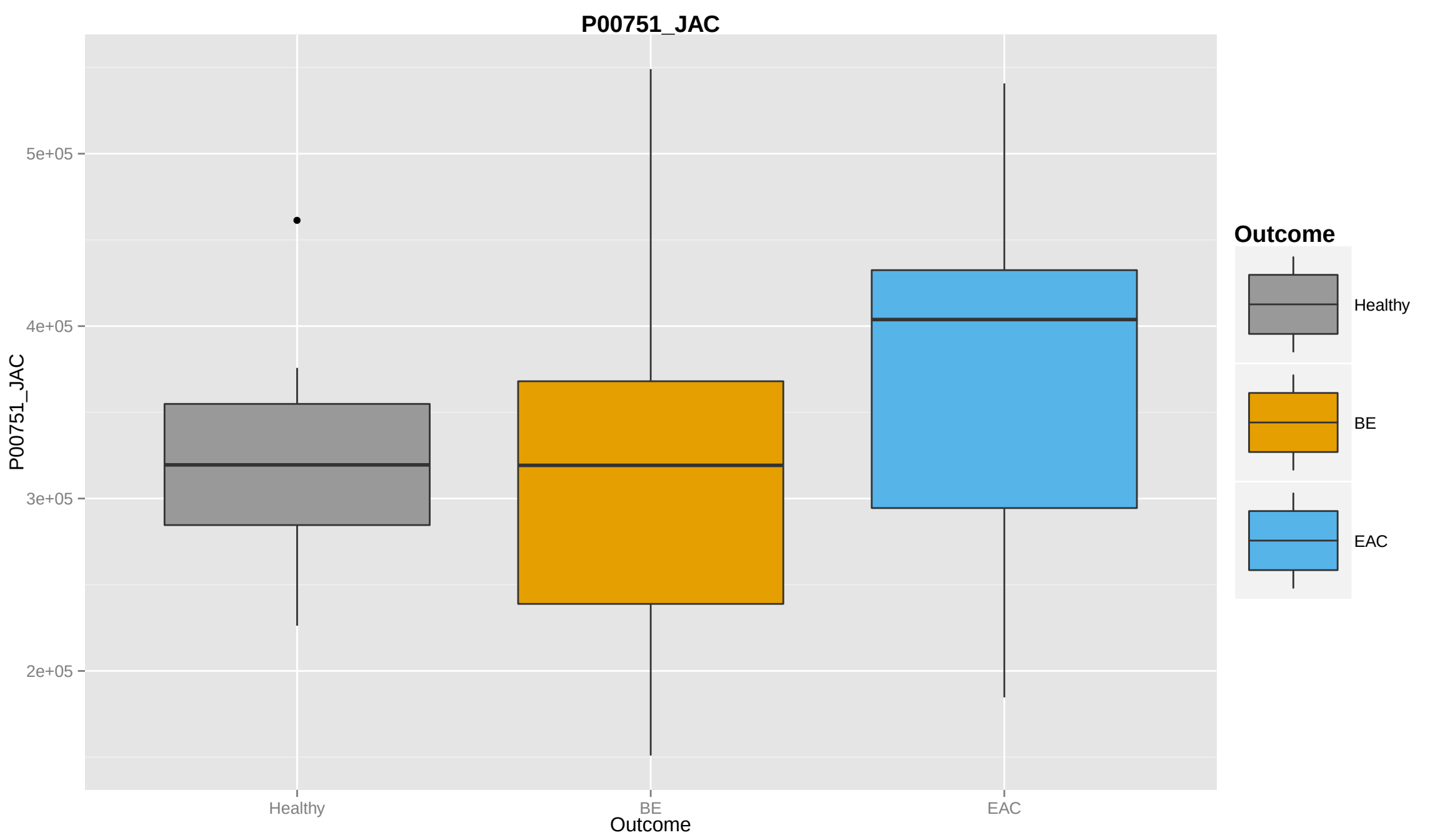

P02765_EPHA
350000
300000
250000
200000
150000
100000
P02765_EPHA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

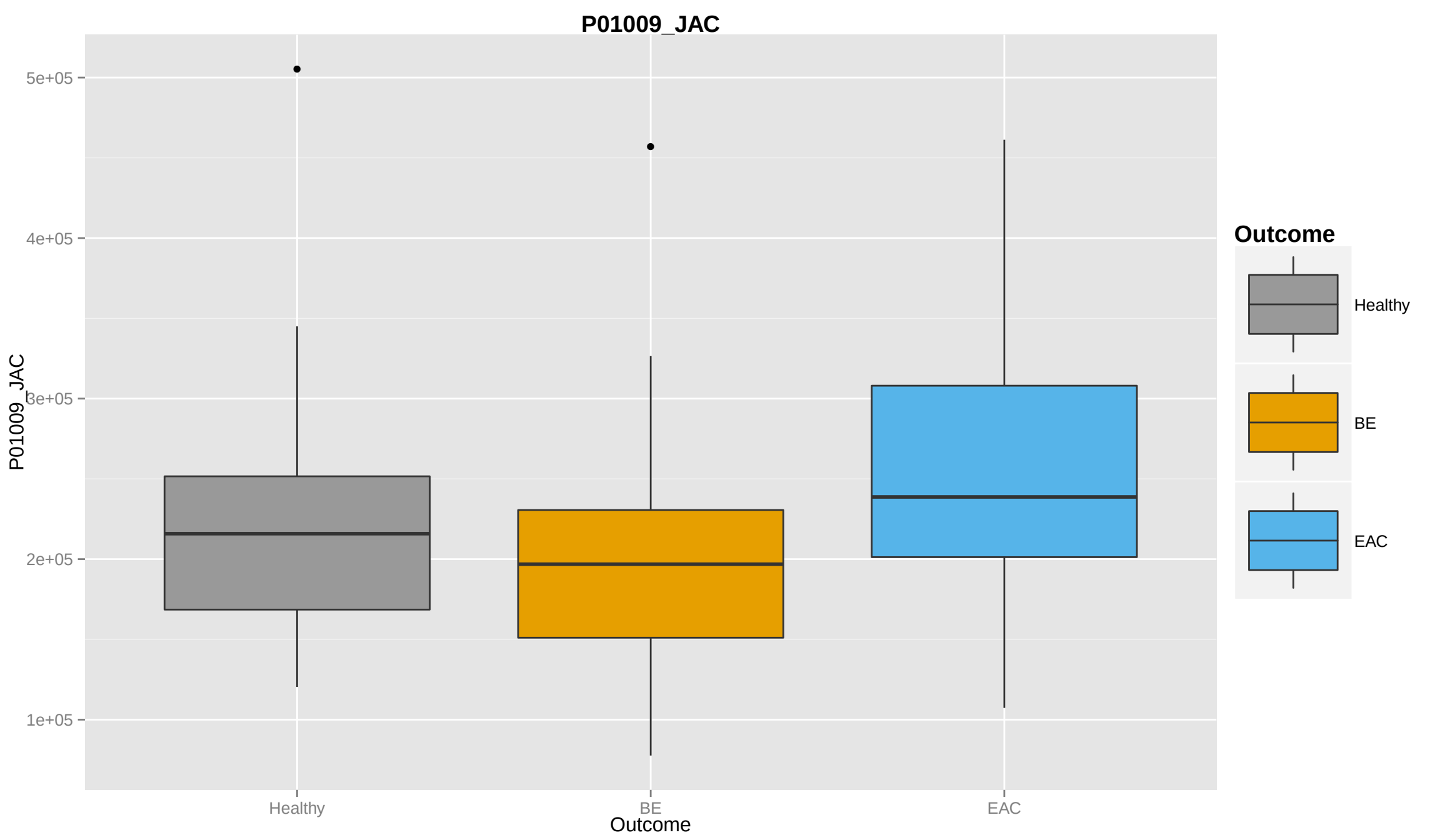

P02787_EPHA
8e+05
6e+05
4e+05
P02787_EPHA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

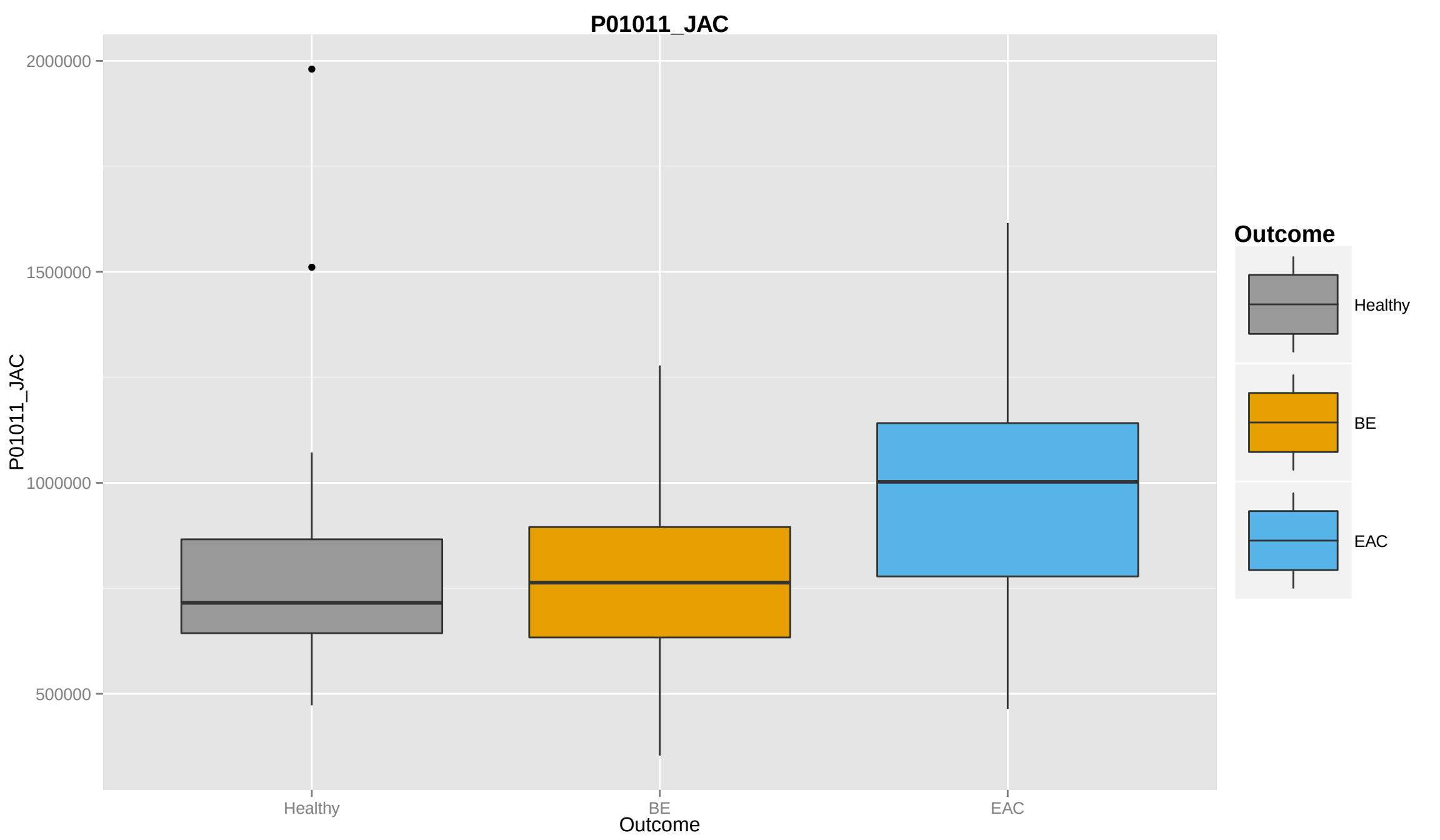

P06396_EPHA
250000
200000
150000
100000
P06396_EPHA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

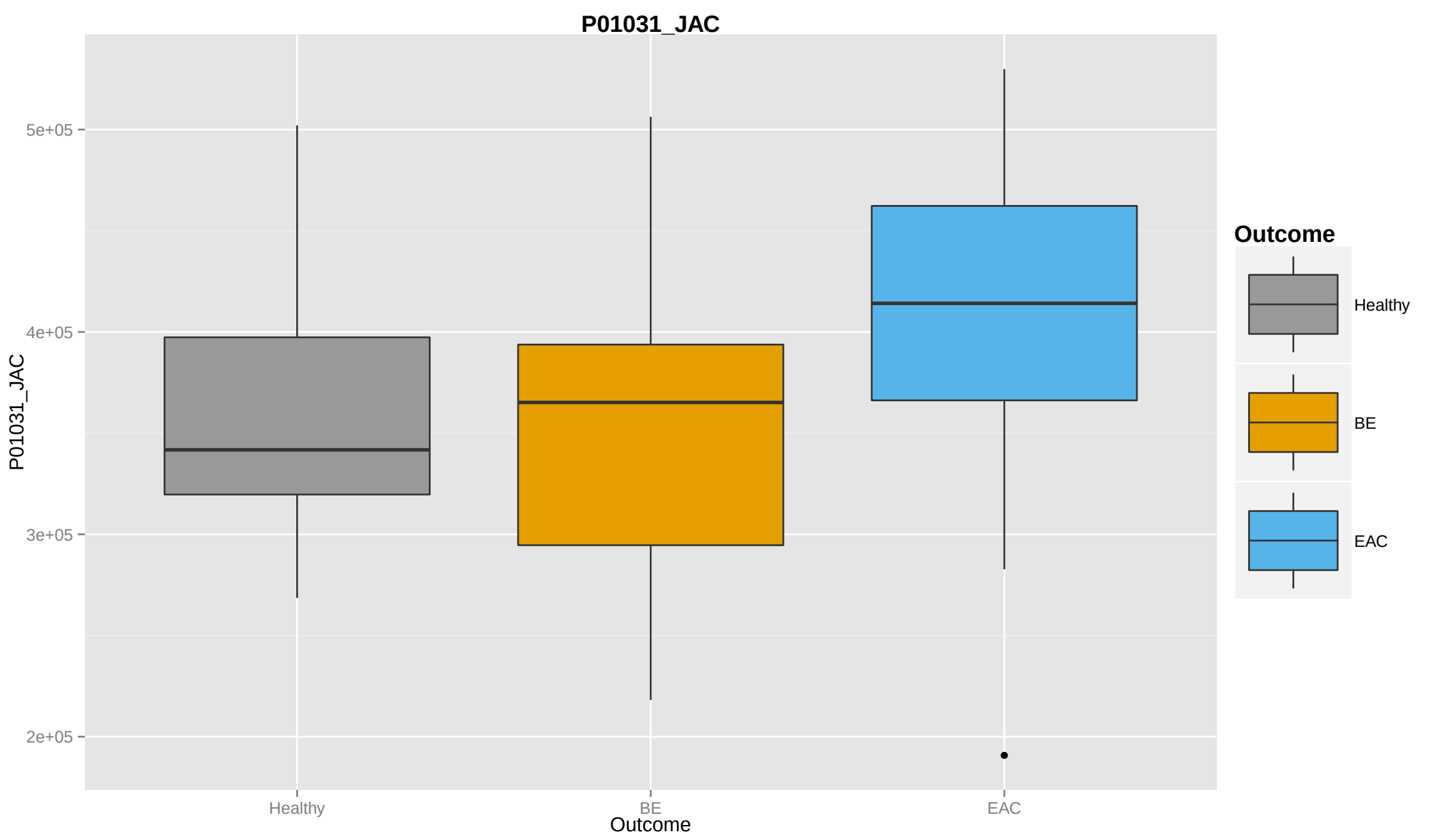

P10643_EPHA
5e+05
4e+05
3e+05
2e+05
P10643_EPHA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

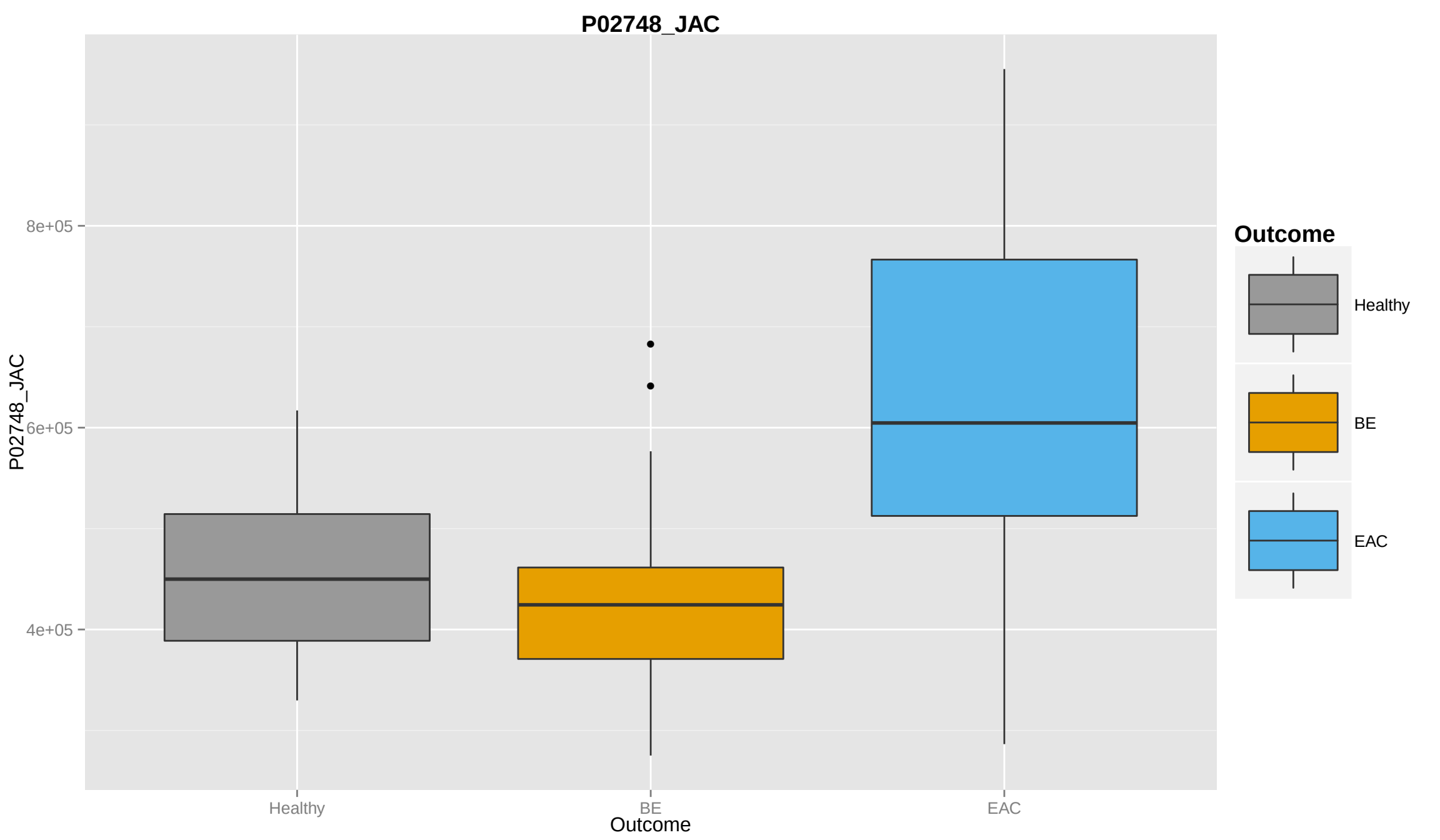

P00738_JAC
3e+07
2e+07
1e+07
P00738_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

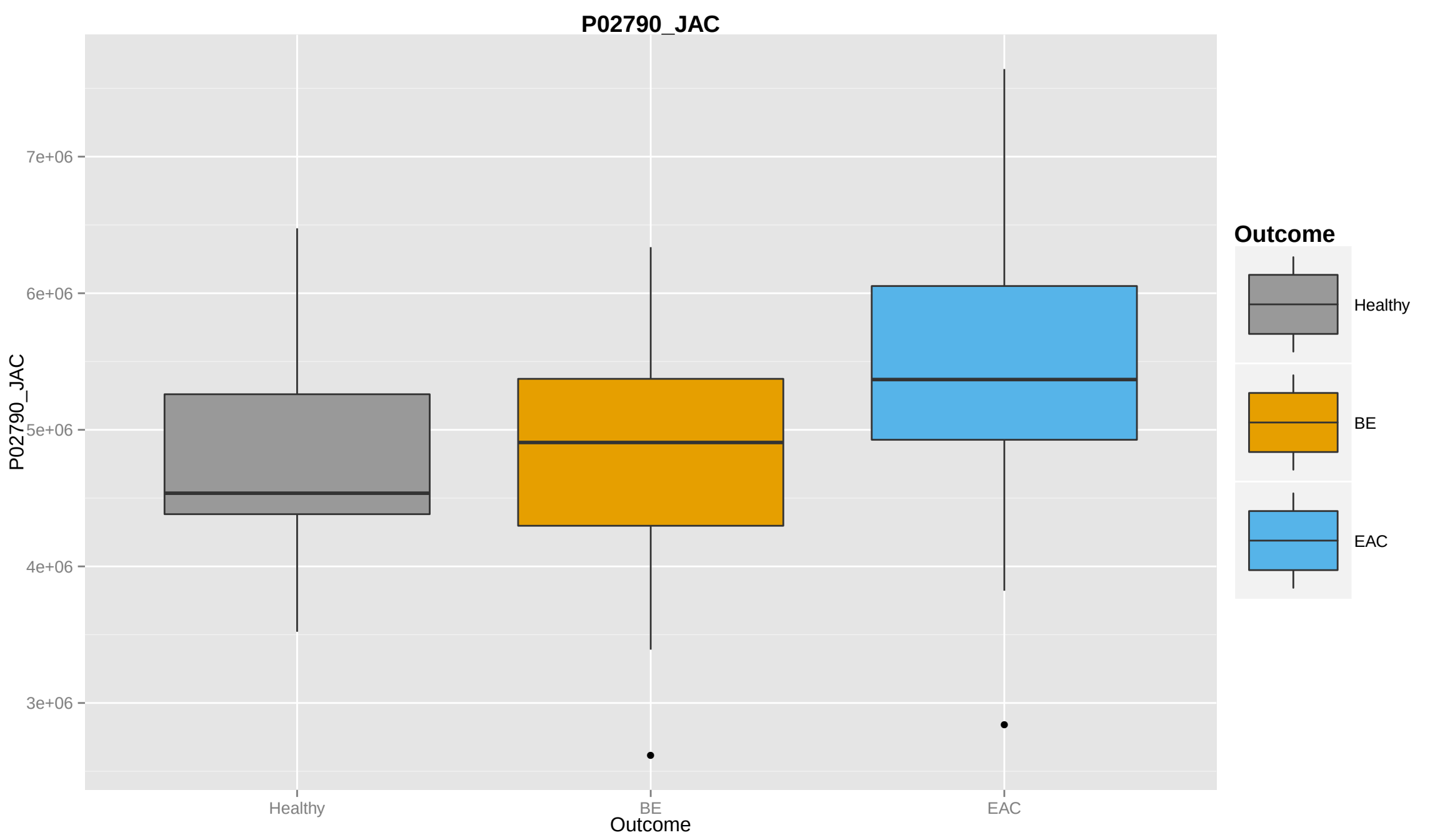

P00751_JAC
5e+05
4e+05
3e+05
2e+05
P00751_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

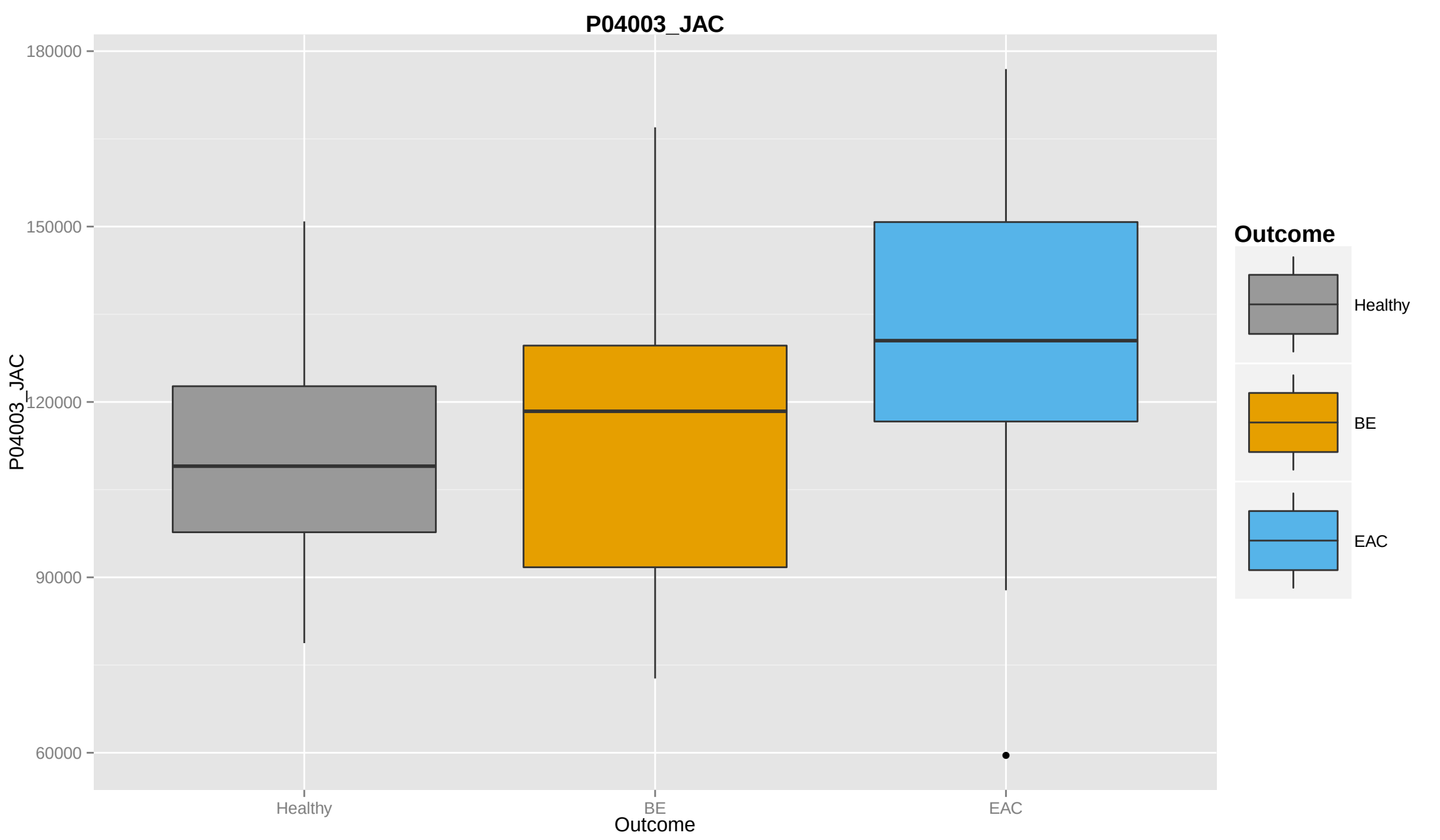

P01009_JAC
5e+05
4e+05
3e+05
2e+05
1e+05
P01009_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

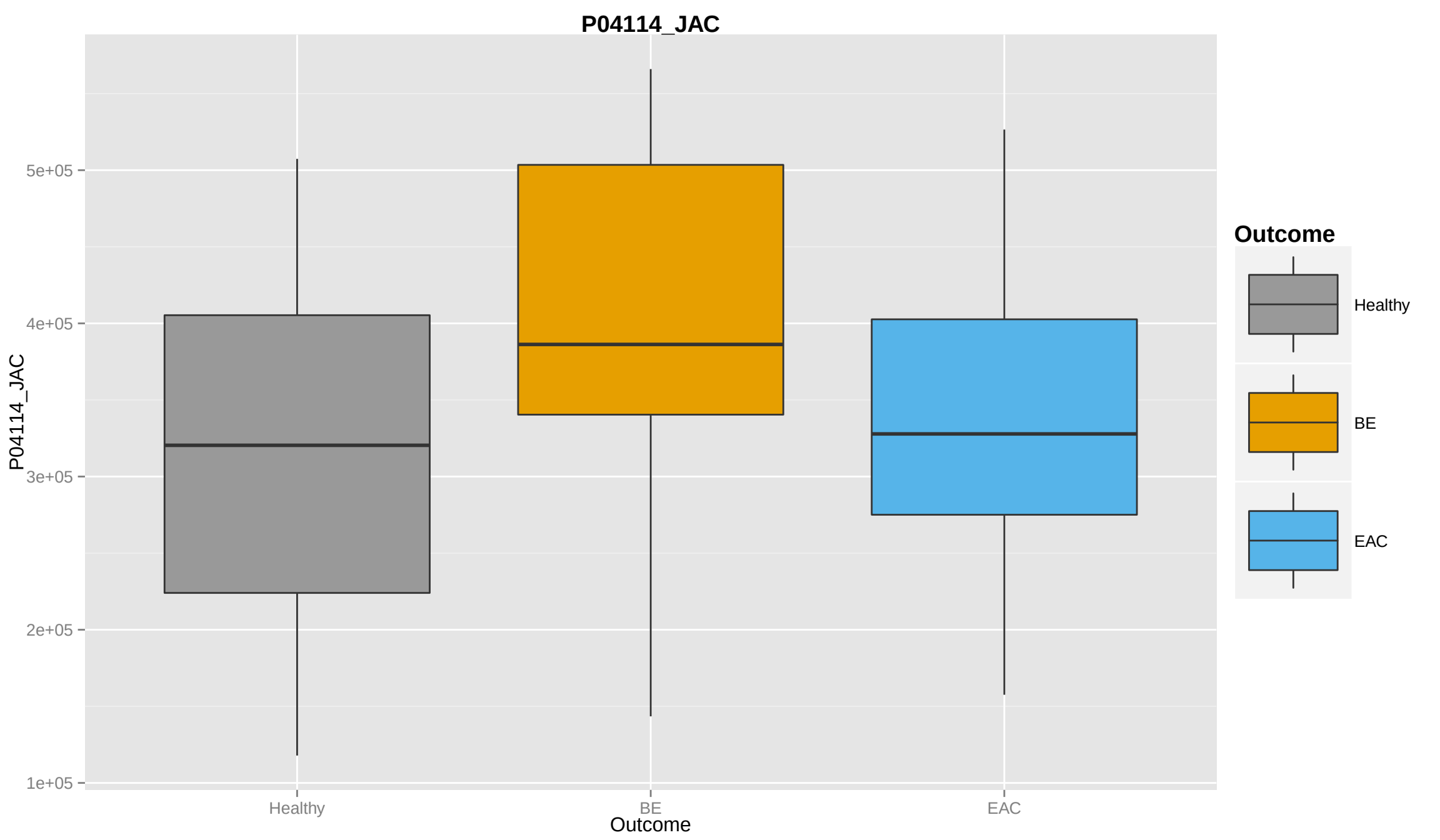

P01011_JAC
2000000
1500000
1000000
500000
P01011_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

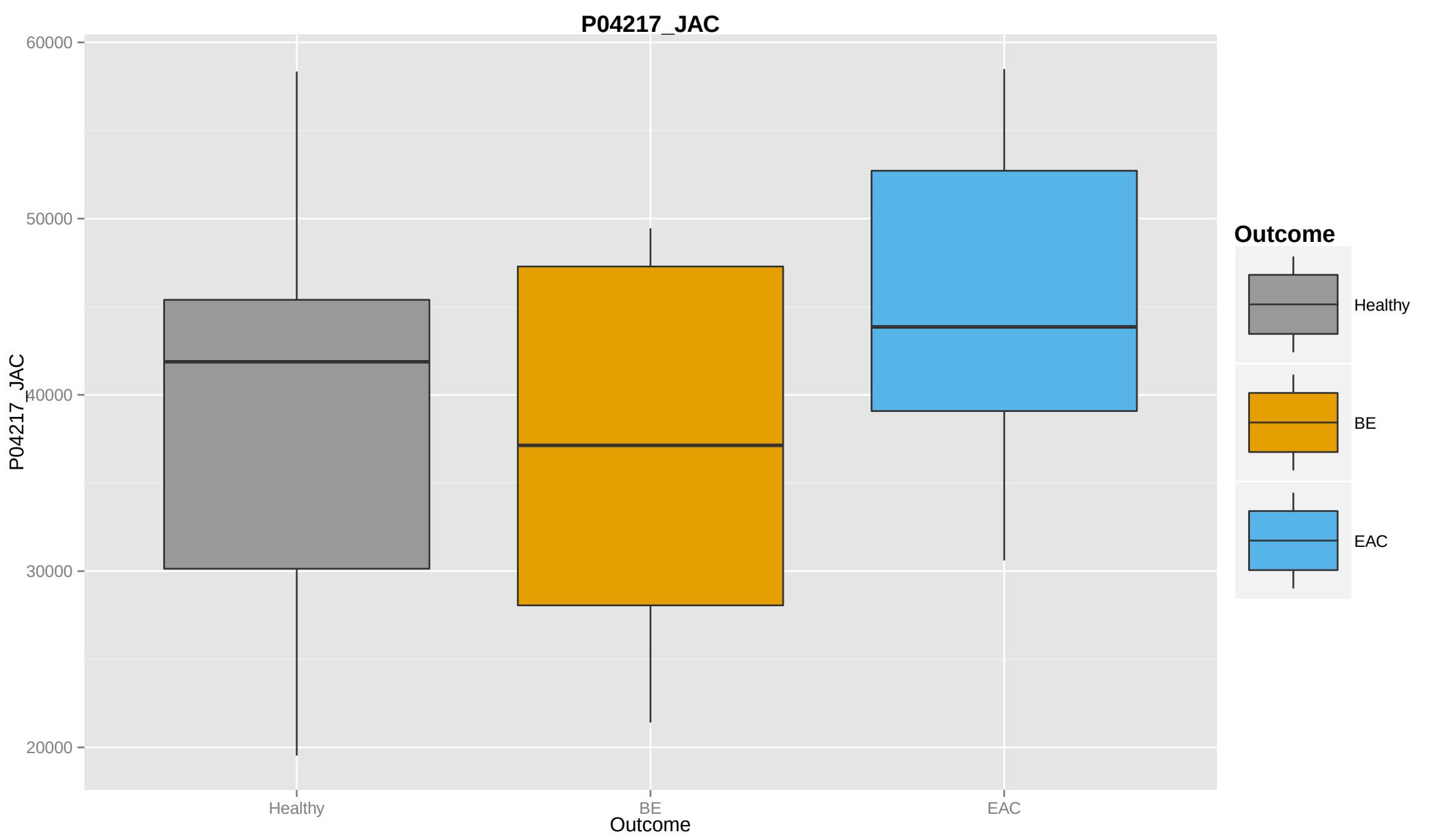

P01031_JAC
5e+05
4e+05
3e+05
2e+05
P01031_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

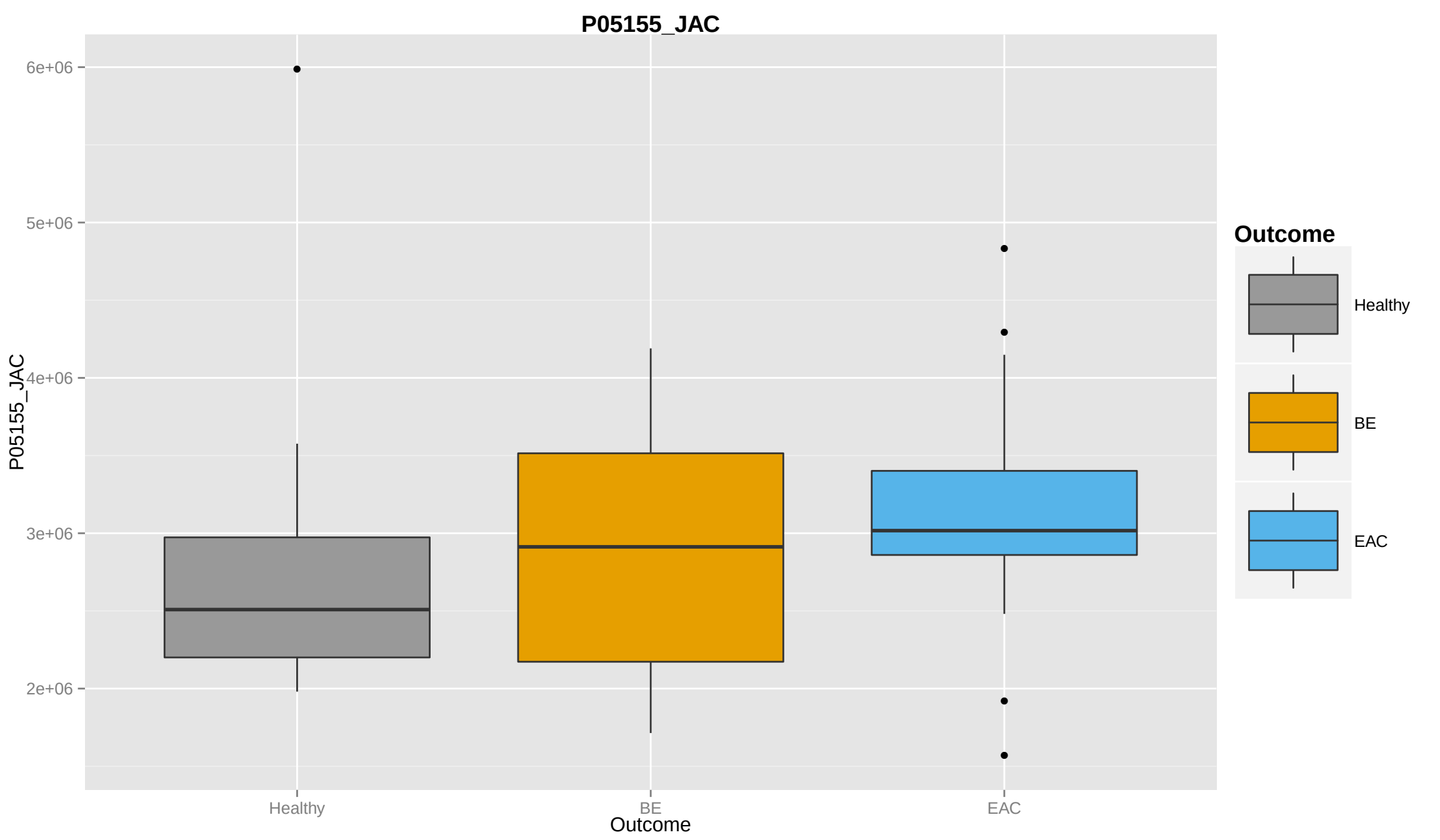

P02748_JAC
8e+05
6e+05
4e+05
P02748_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

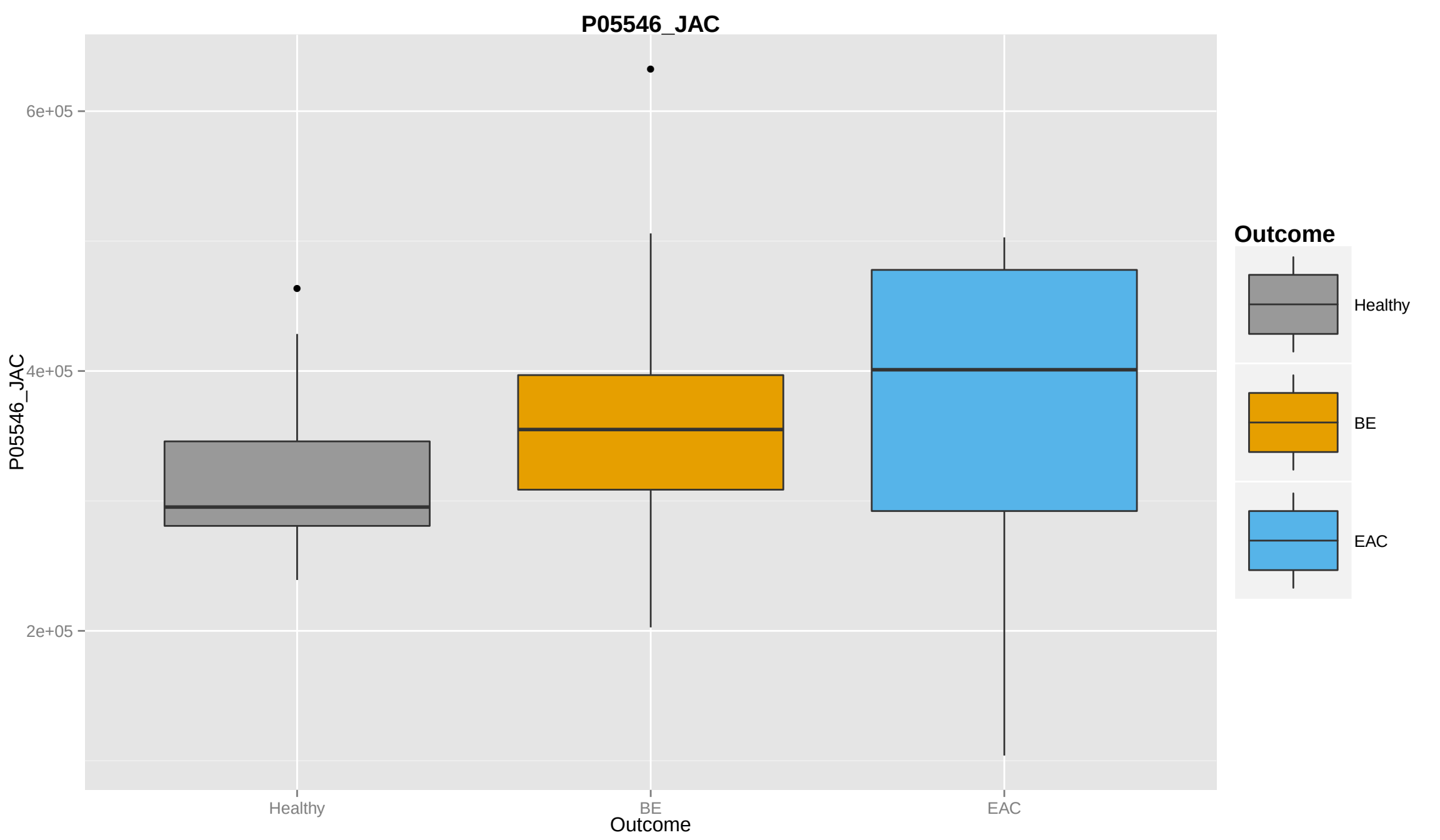

P02790_JAC
7e+06
6e+06
5e+06
4e+06
3e+06
P02790_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

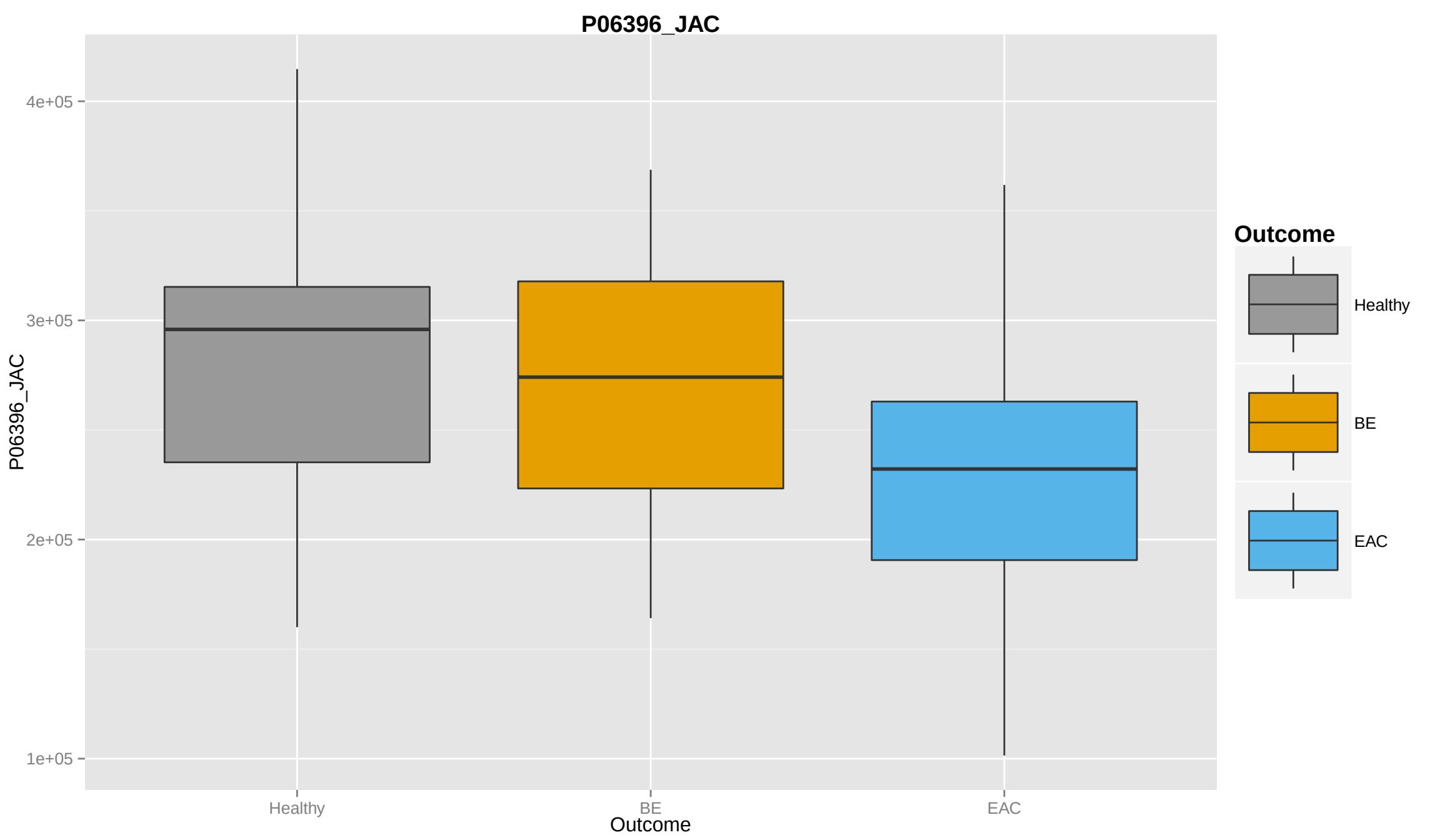

P04003_JAC
180000
150000
120000
90000
60000
P04003_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

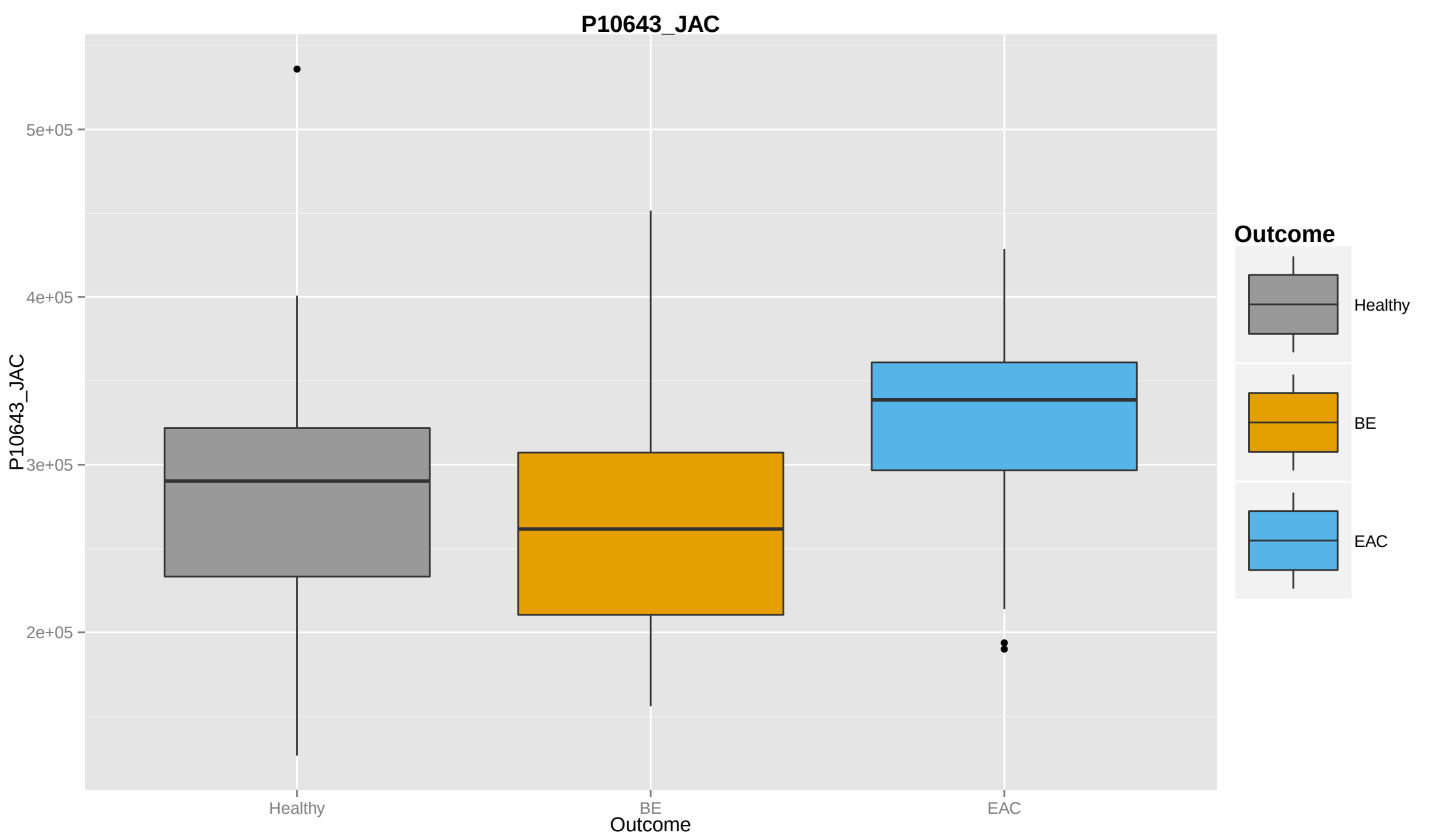

P04114_JAC
5e+05
4e+05
3e+05
2e+05
1e+05
P04114_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

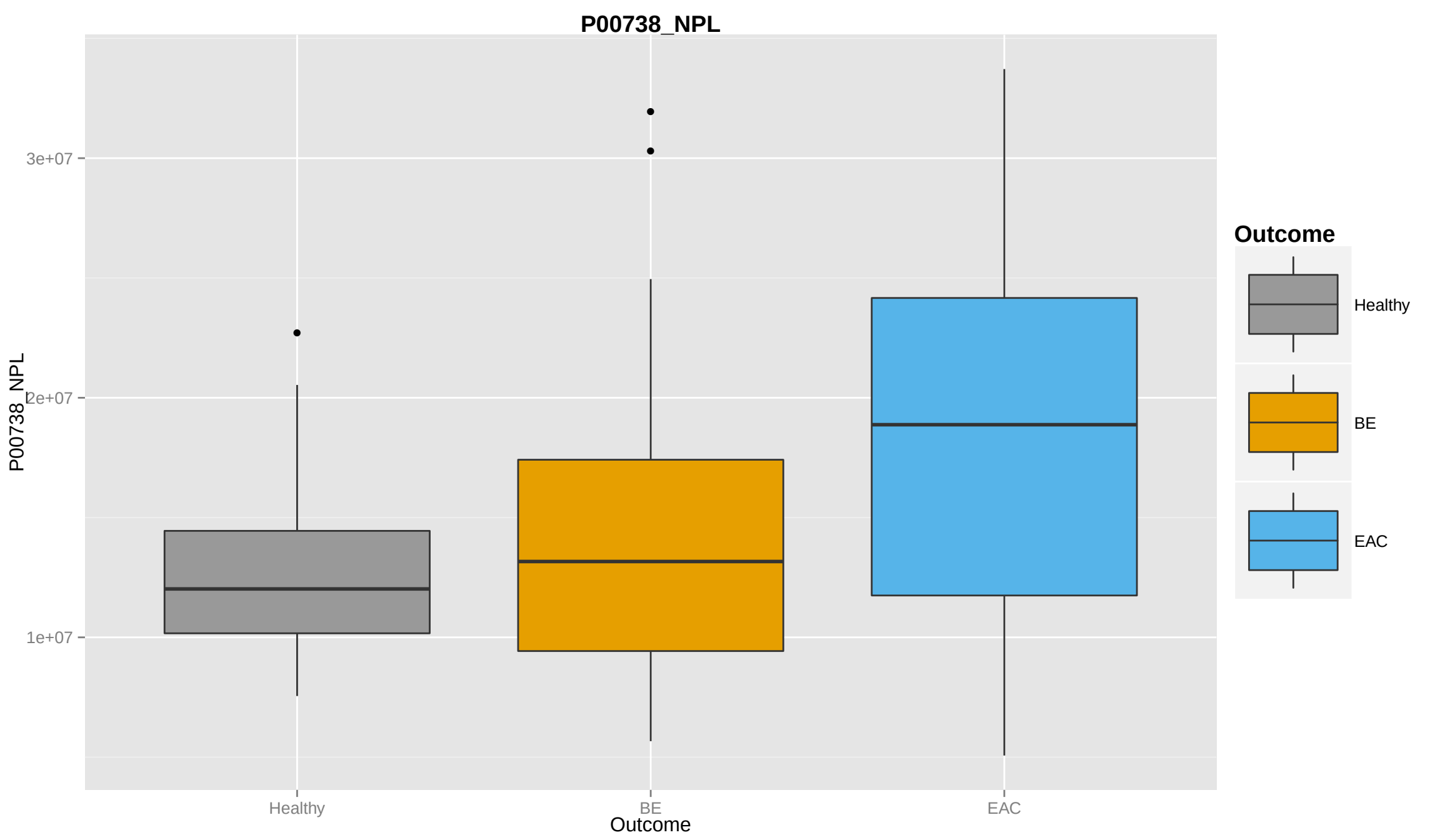

P04217_JAC
60000
50000
40000
30000
20000
P04217_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

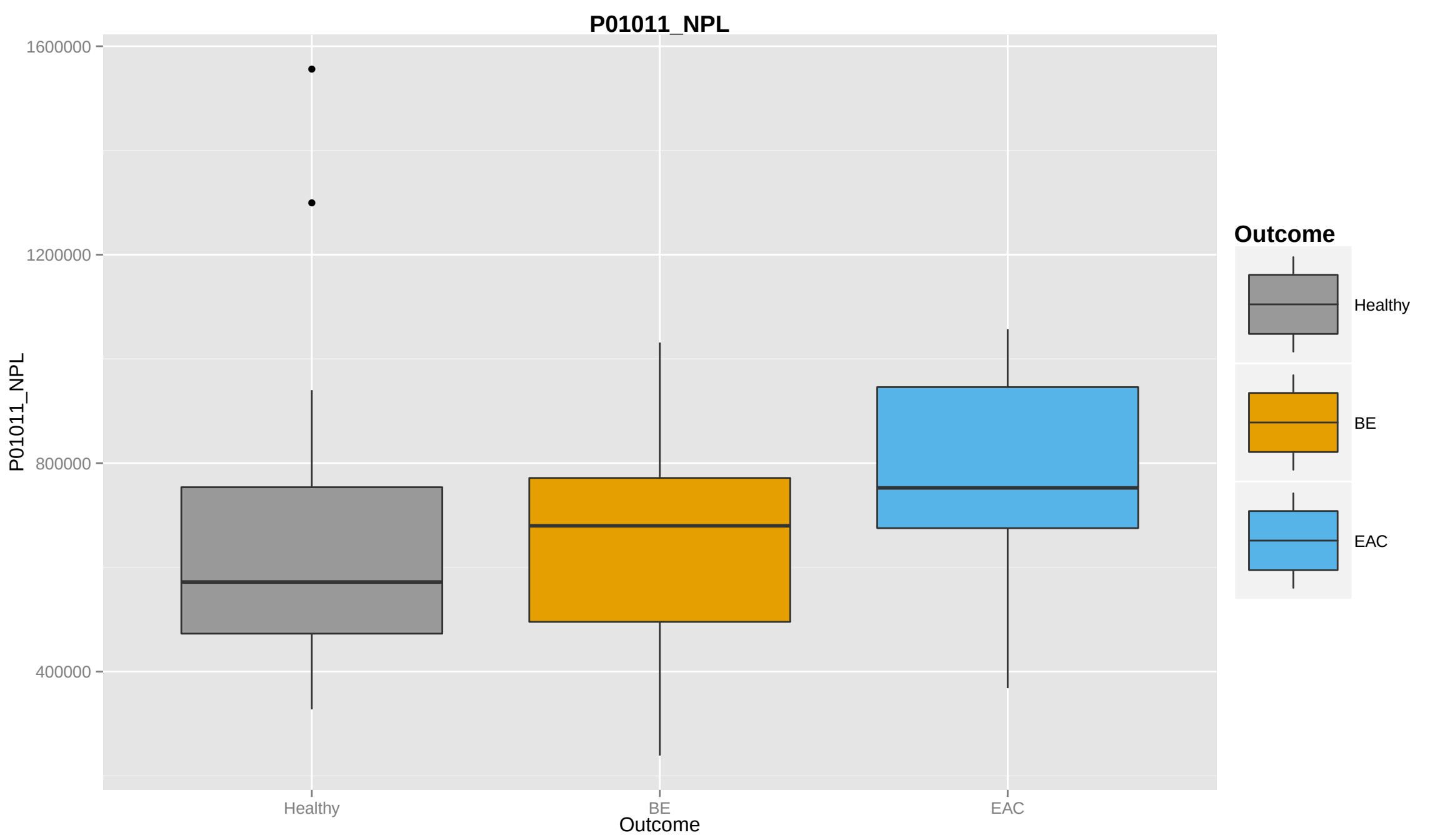

P05155_JAC
6e+06
5e+06
4e+06
3e+06
2e+06
P05155_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

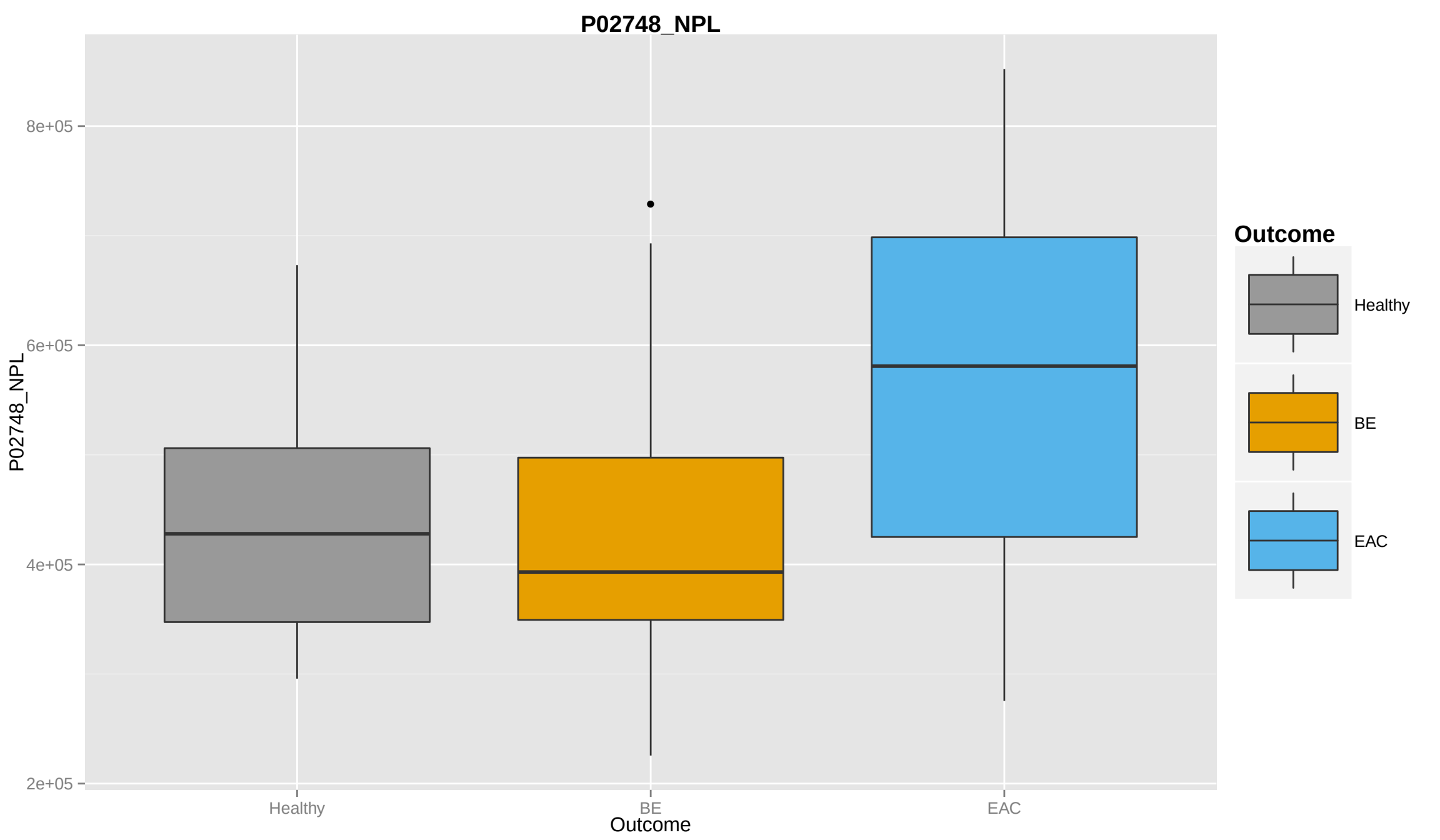

P05546_JAC
6e+05
4e+05
2e+05
P05546_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

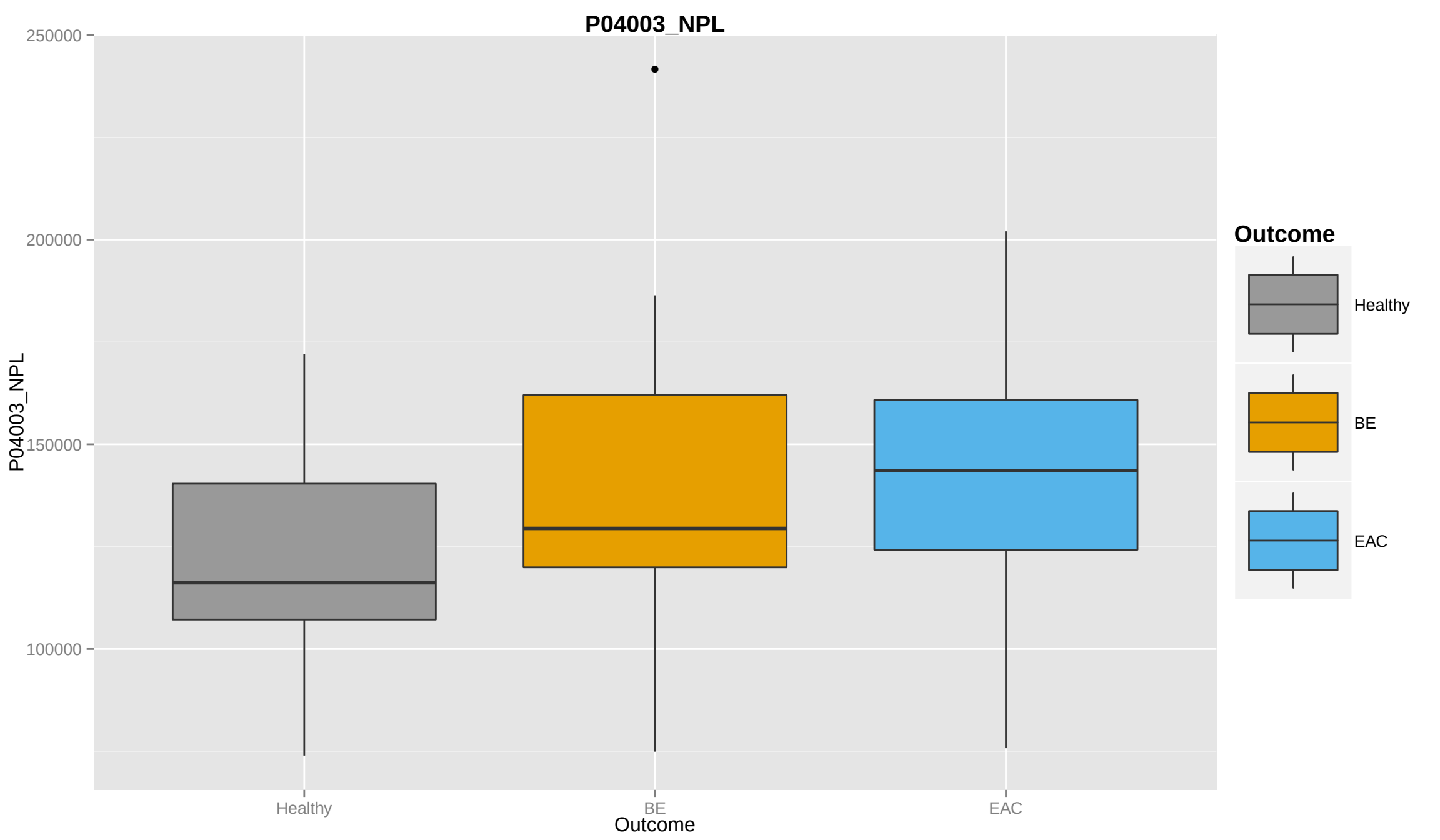

P06396_JAC
4e+05
3e+05
2e+05
1e+05
P06396_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

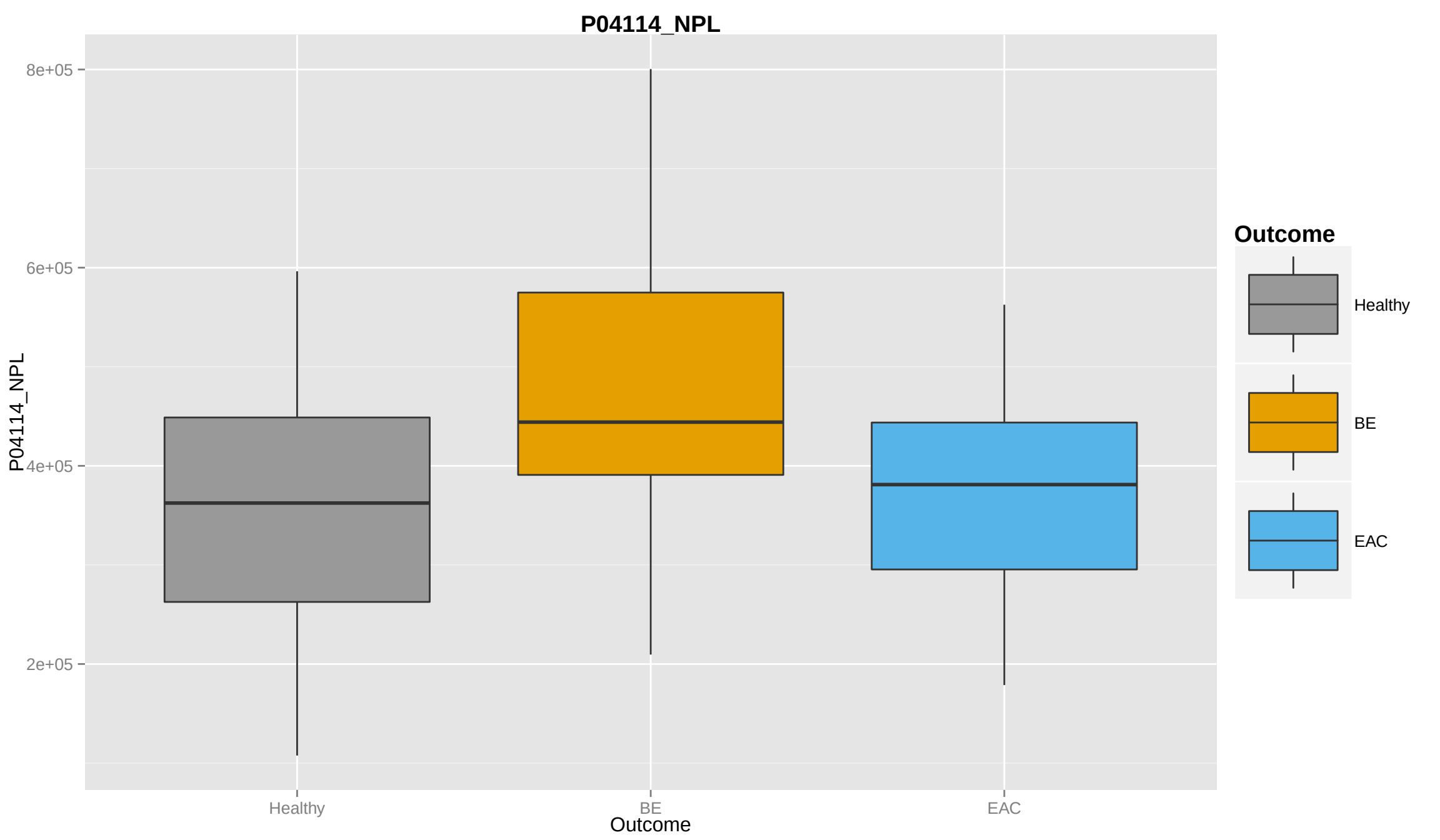

P10643_JAC
5e+05
4e+05
3e+05
2e+05
P10643_JAC
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

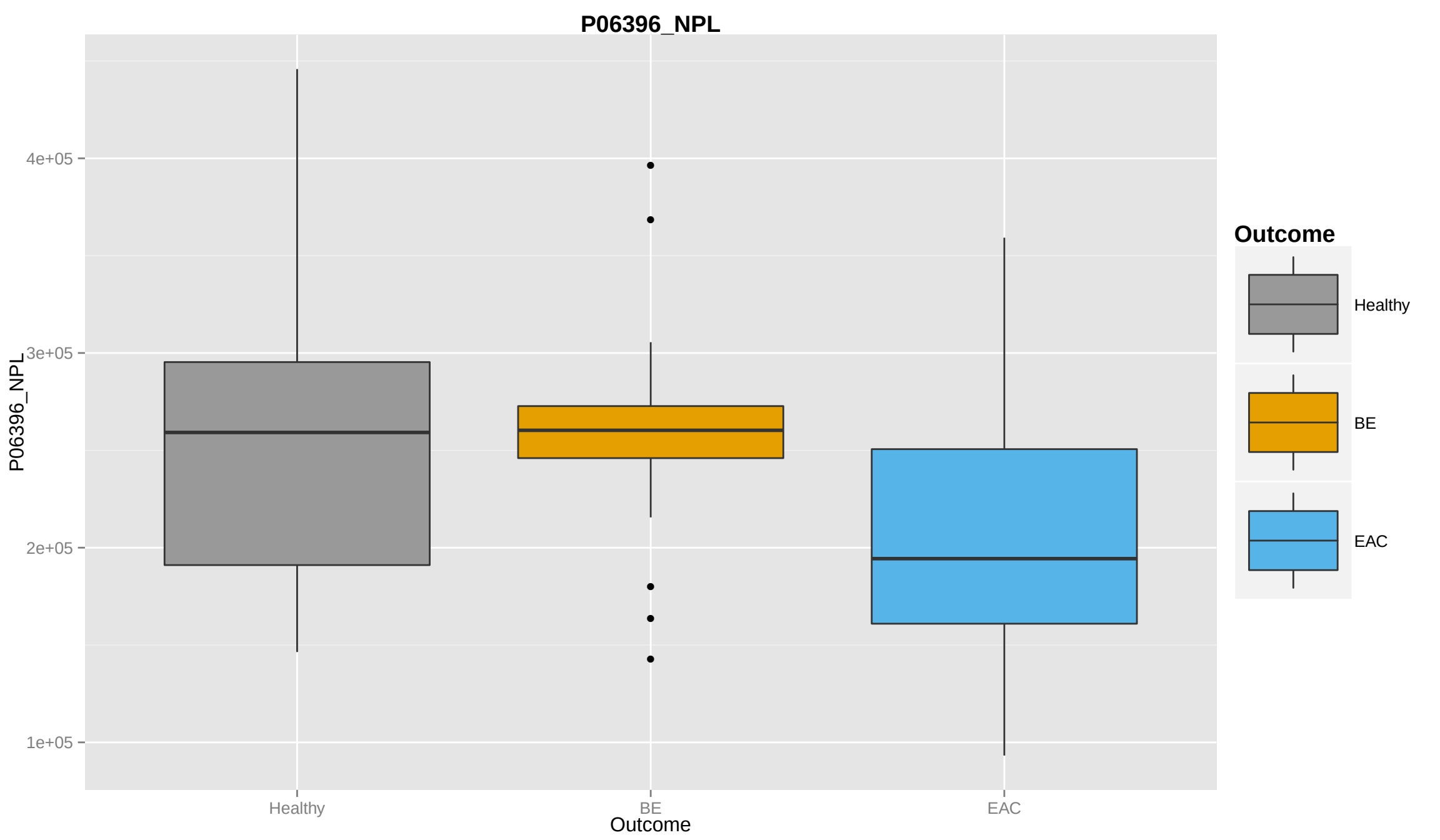

P00738_NPL
3e+07
2e+07
1e+07
P00738_NPL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

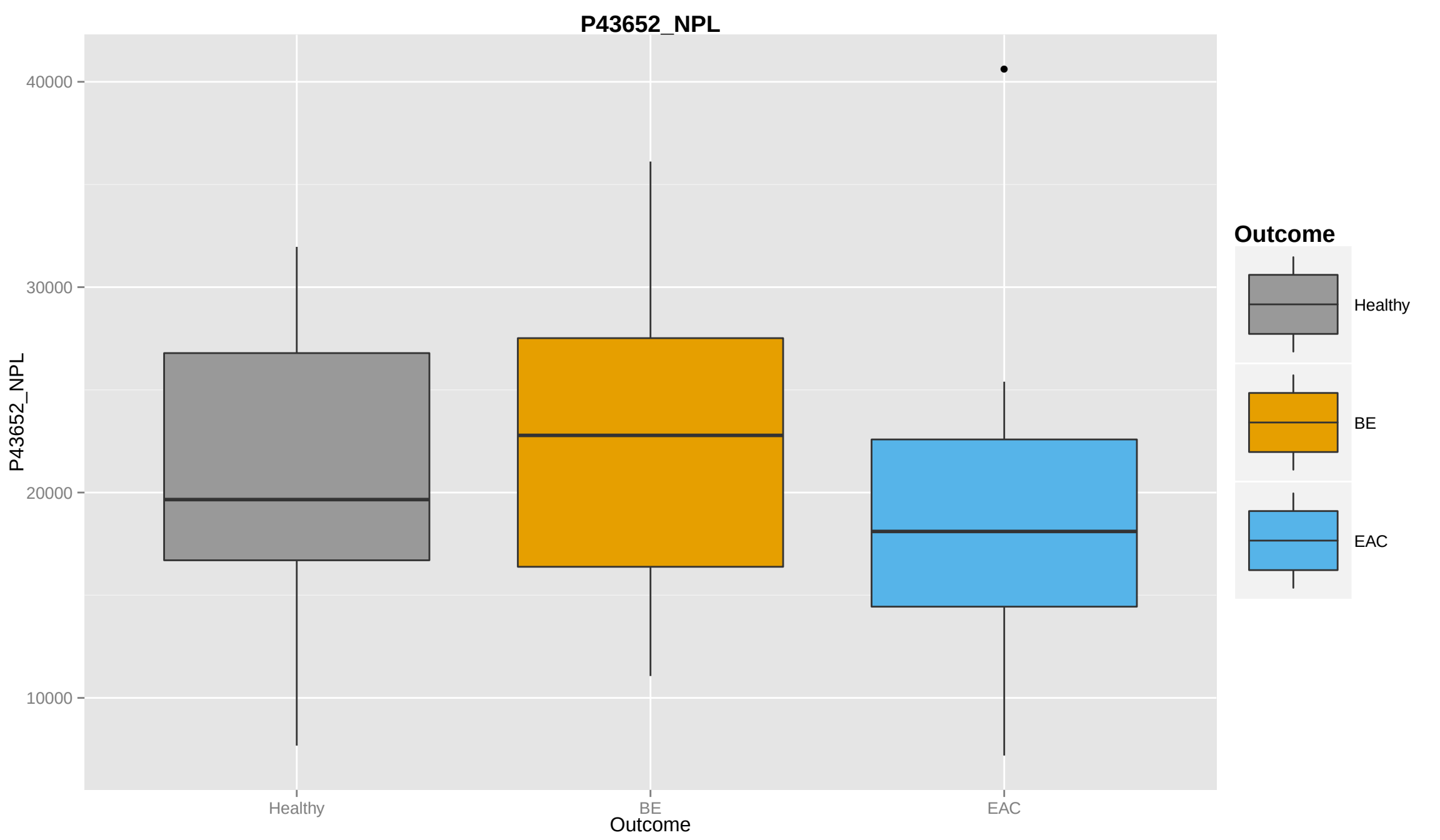

P01011_NPL
1600000
1200000
800000
400000
P01011_NPL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

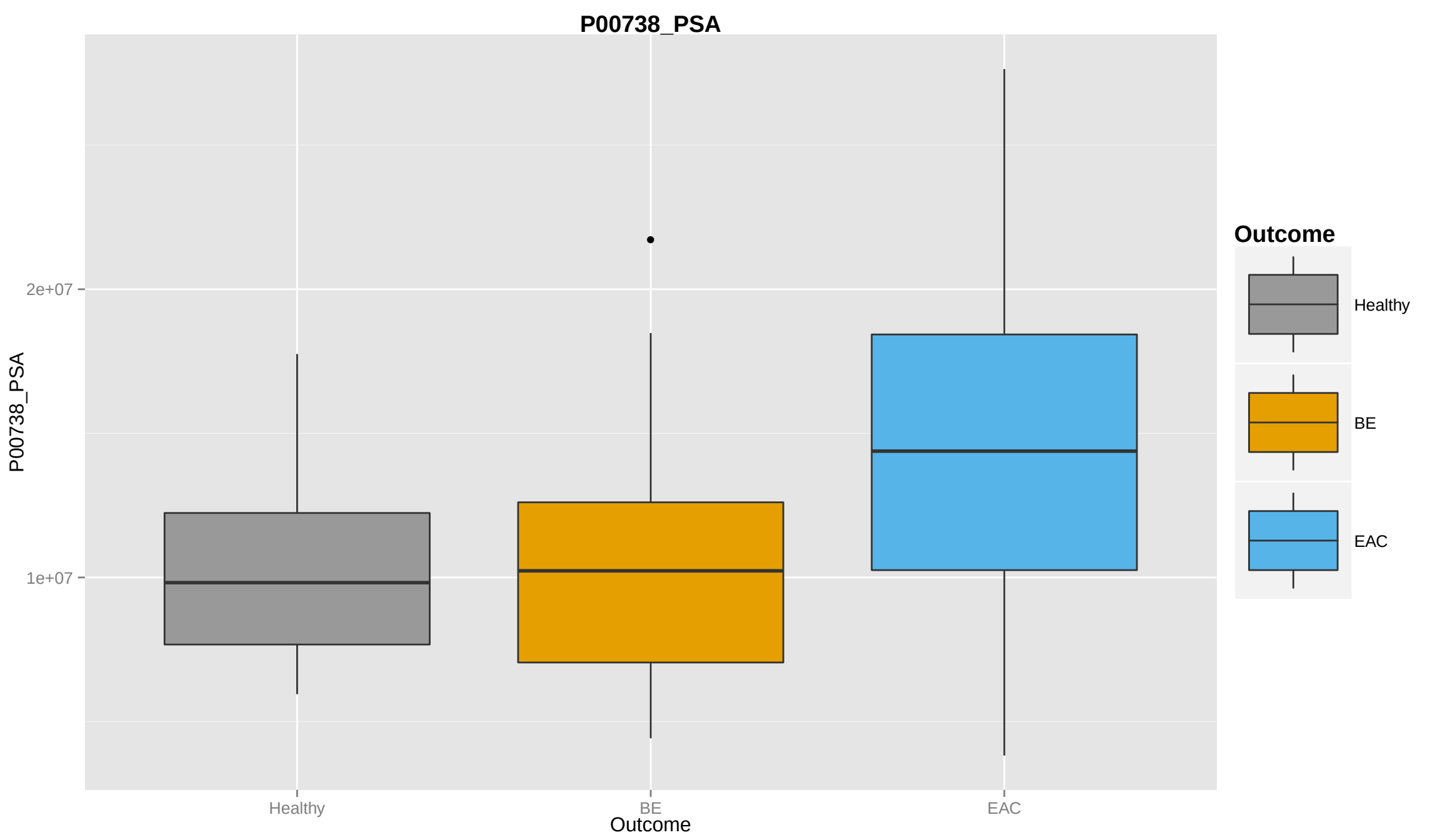

P02748_NPL
8e+05
6e+05
4e+05
2e+05
P02748_NPL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

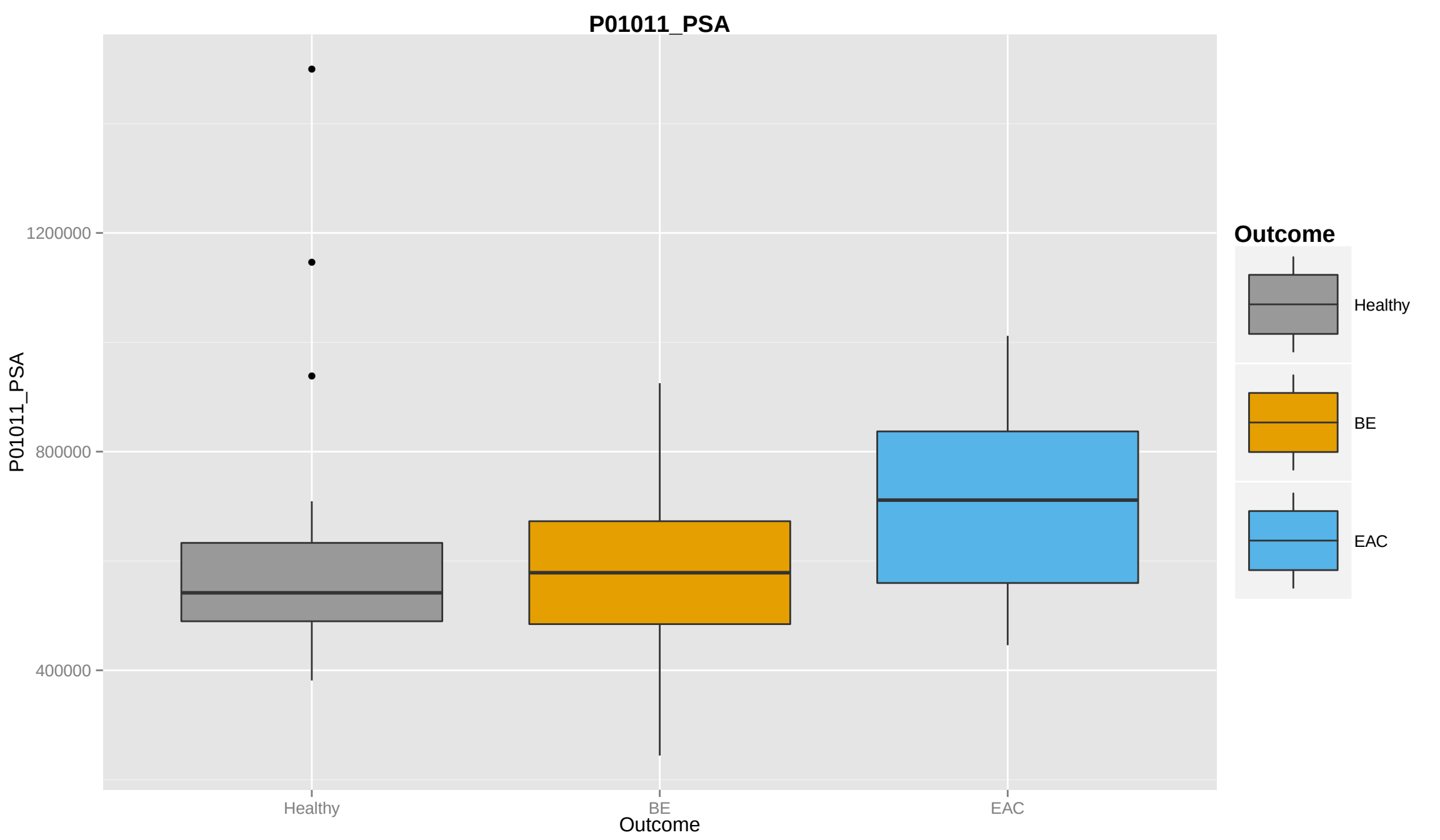

P04003_NPL
250000
200000
150000
100000
P04003_NPL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

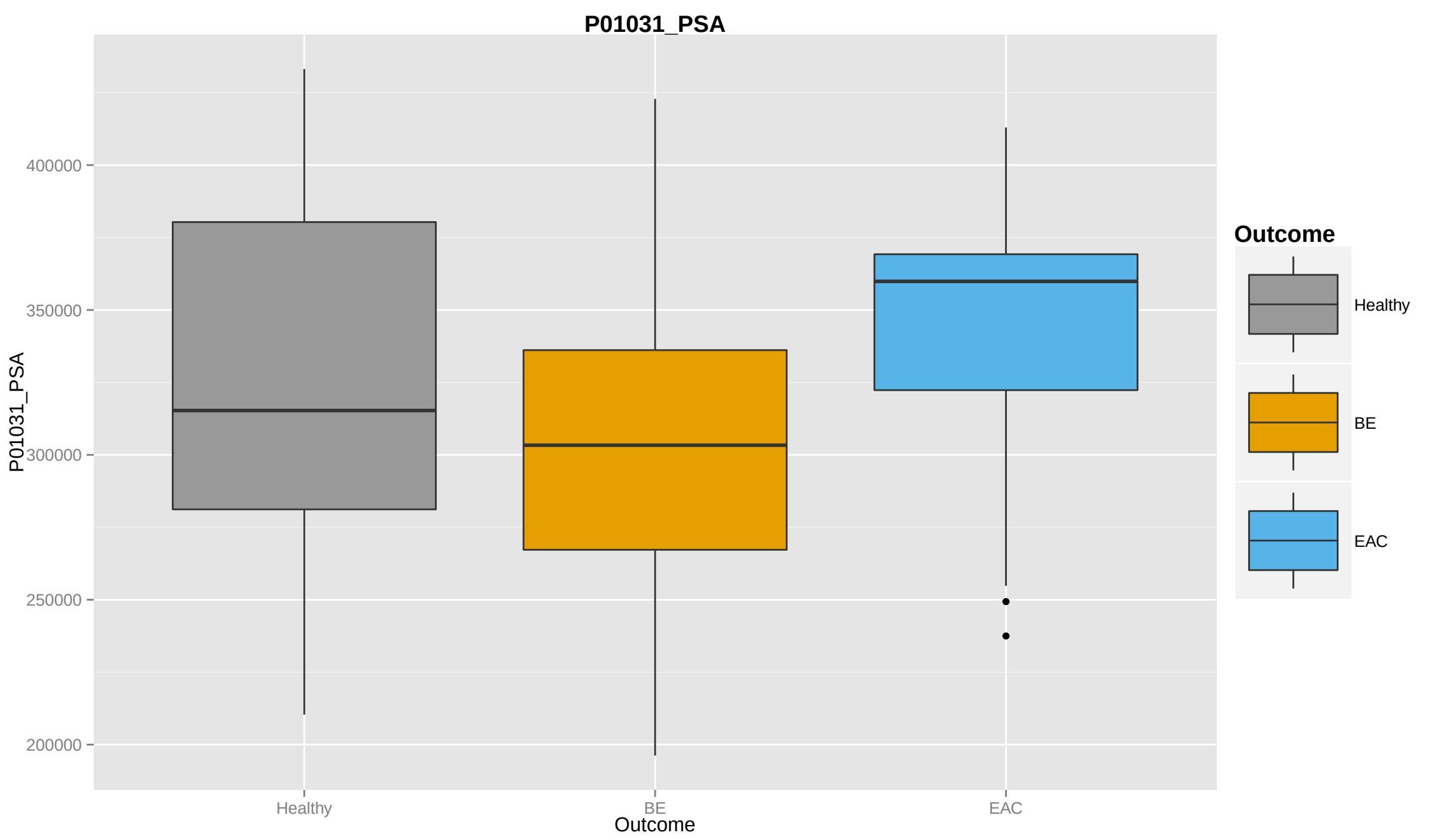

P04114_NPL
8e+05
6e+05
4e+05
2e+05
P04114_NPL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

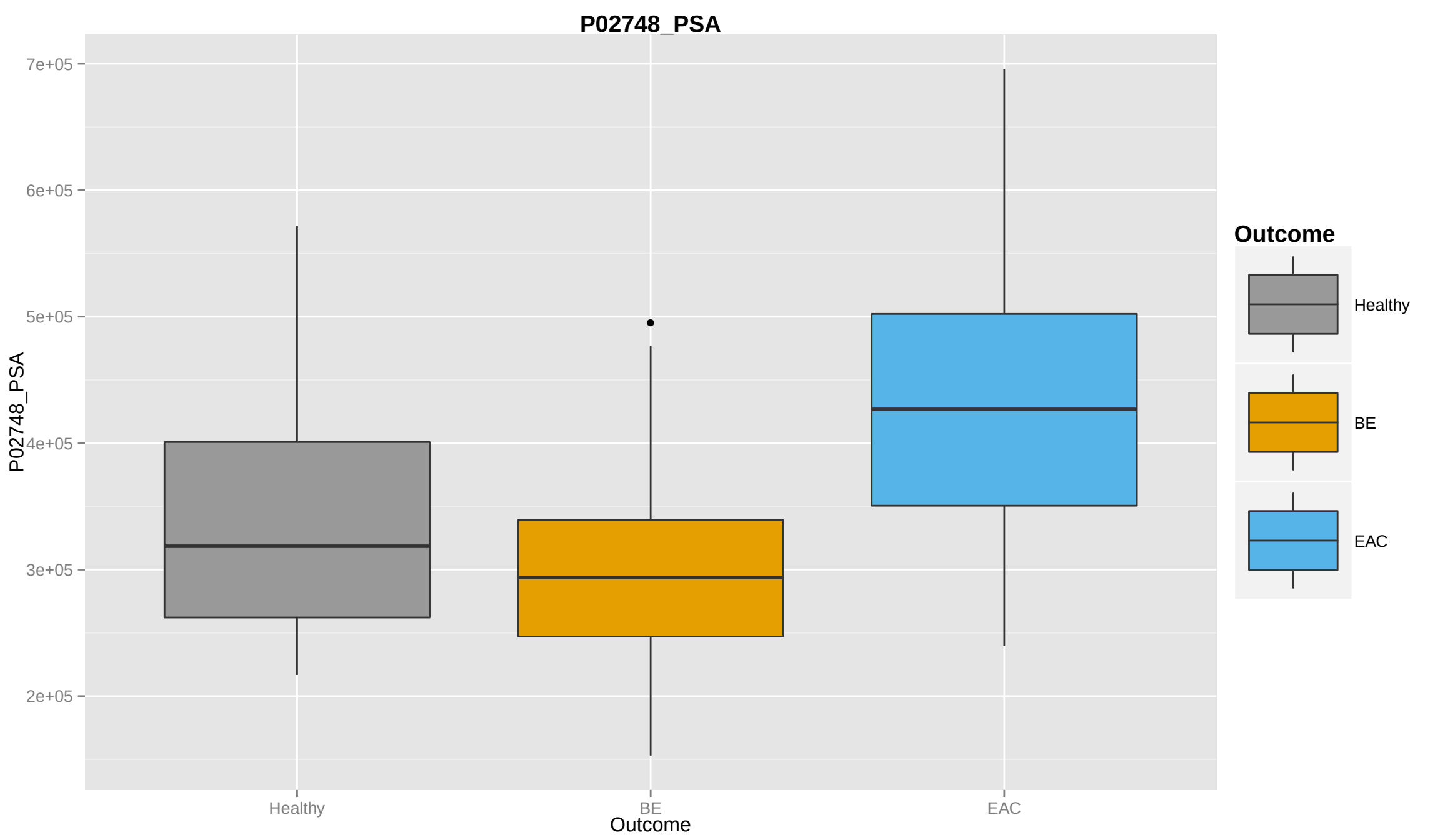

P06396_NPL
4e+05
3e+05
2e+05
1e+05
P06396_NPL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

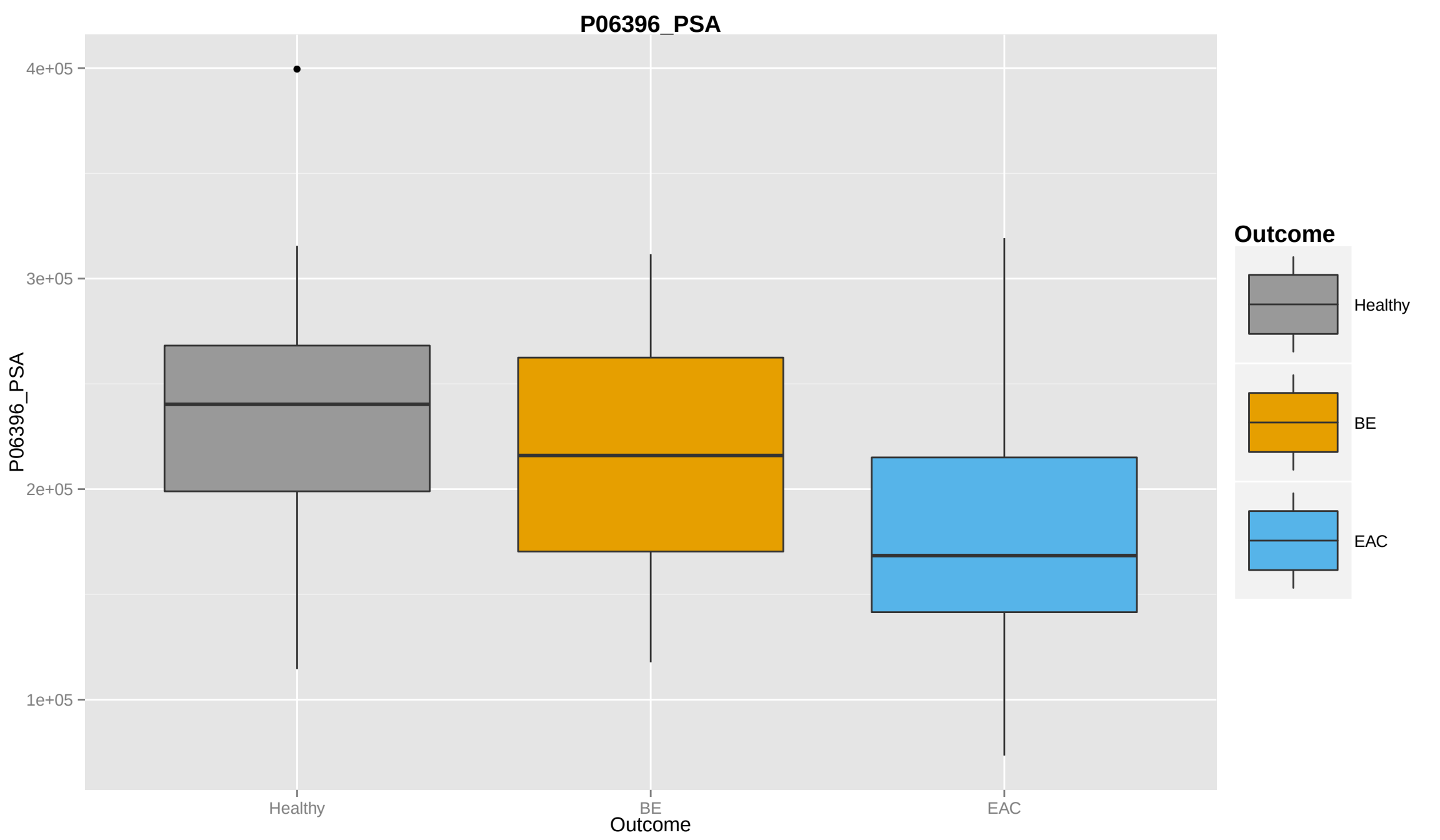

P43652_NPL
40000
30000
20000
10000
P43652_NPL
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

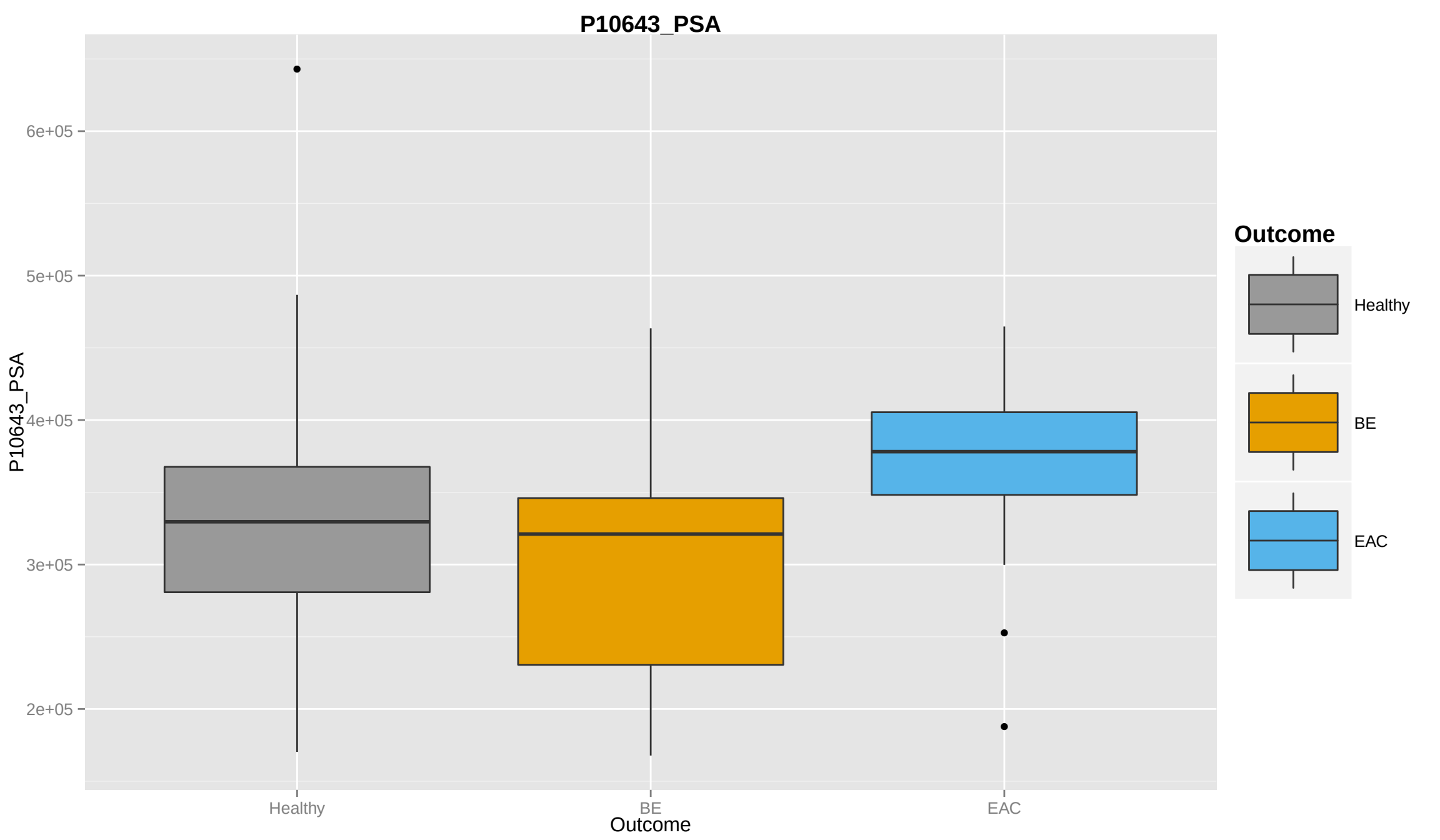

P00738_PSA
2e+07
1e+07
P00738_PSA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

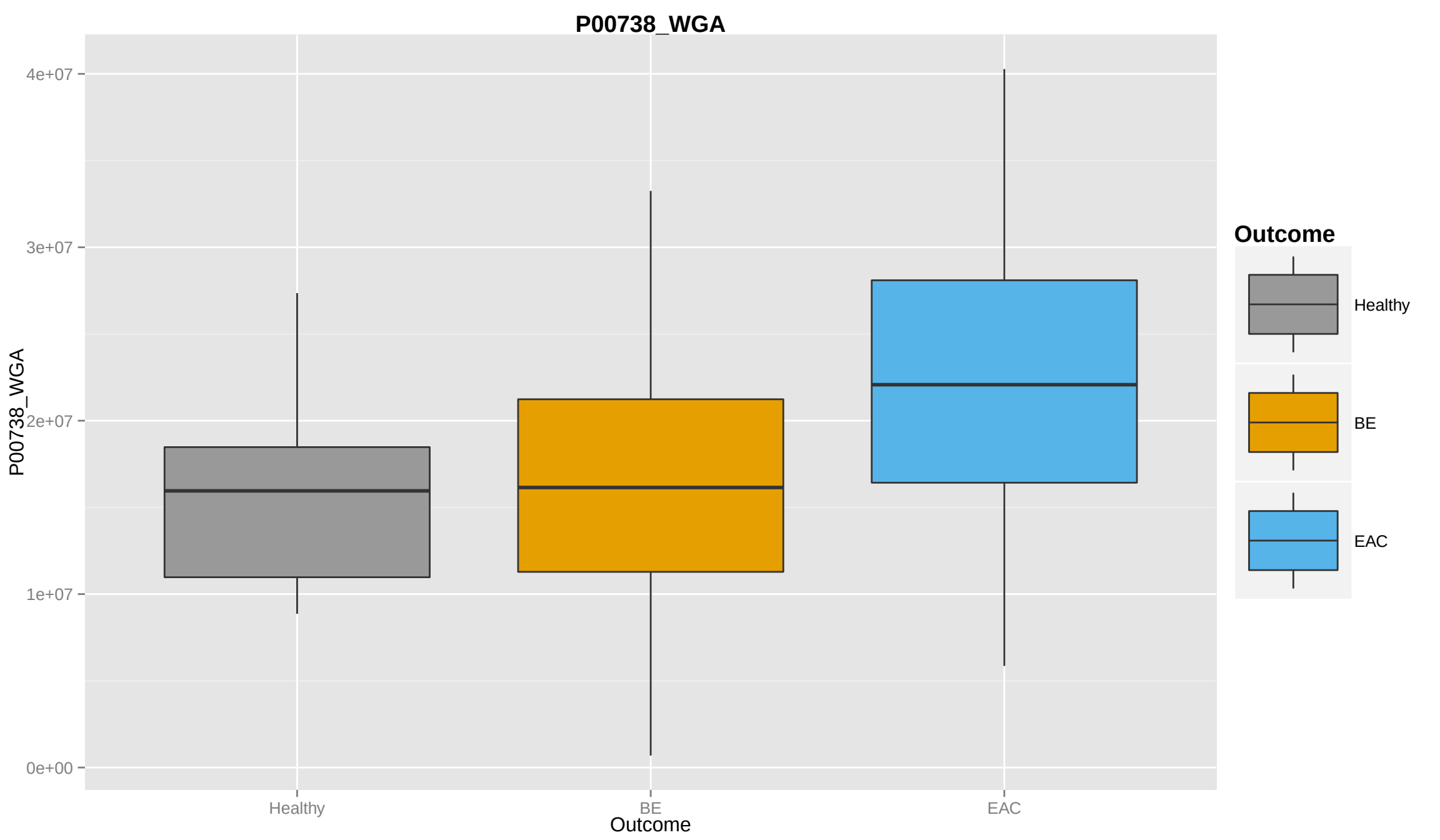

P01011_PSA
1200000
800000
400000
P01011_PSA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

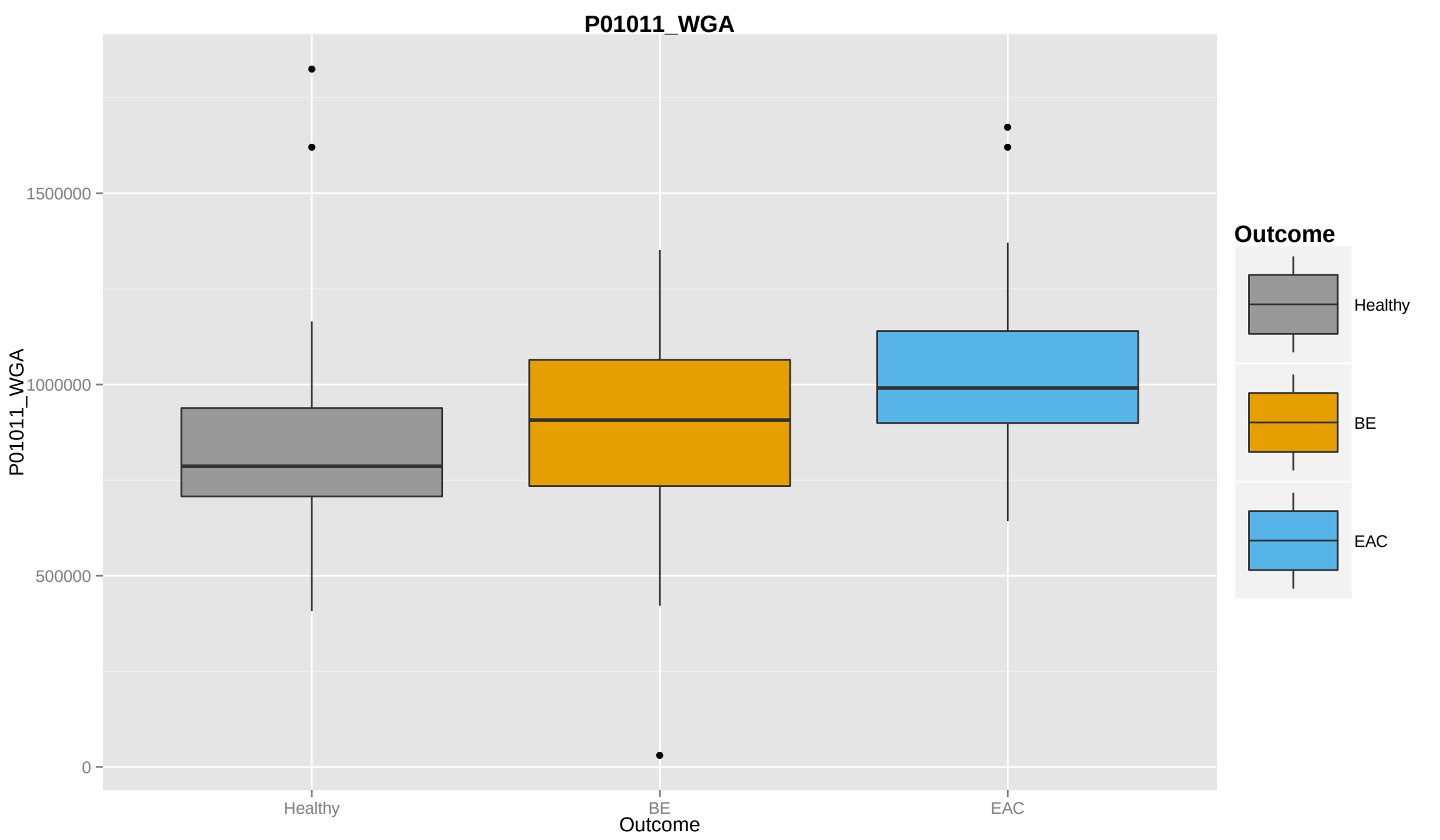

P01031_PSA
400000
350000
300000
250000
200000
P01031_PSA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

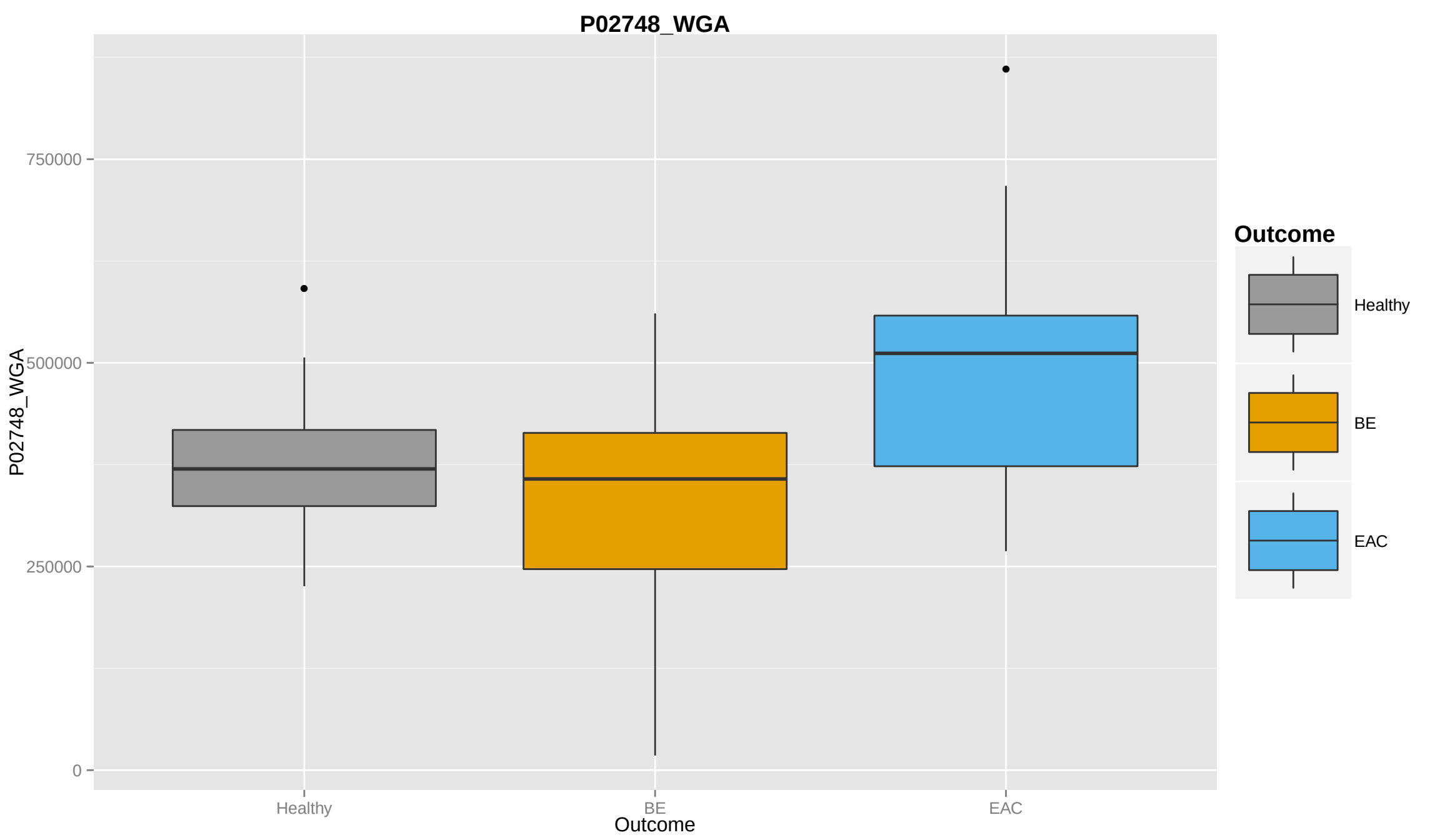

P02748_PSA
7e+05
6e+05
5e+05
4e+05
3e+05
2e+05
P02748_PSA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

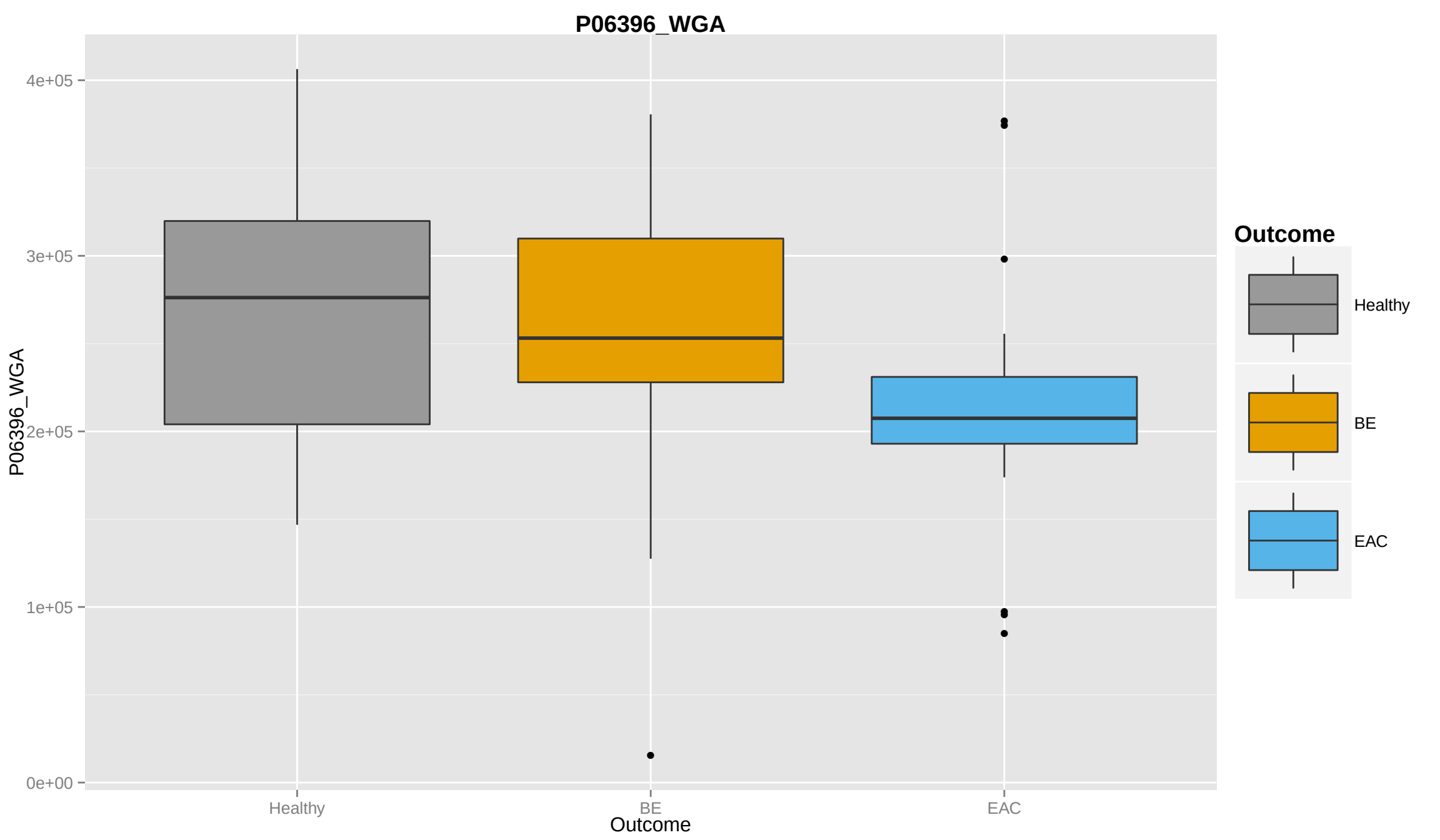

P06396_PSA
4e+05
3e+05
2e+05
1e+05
P06396_PSA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

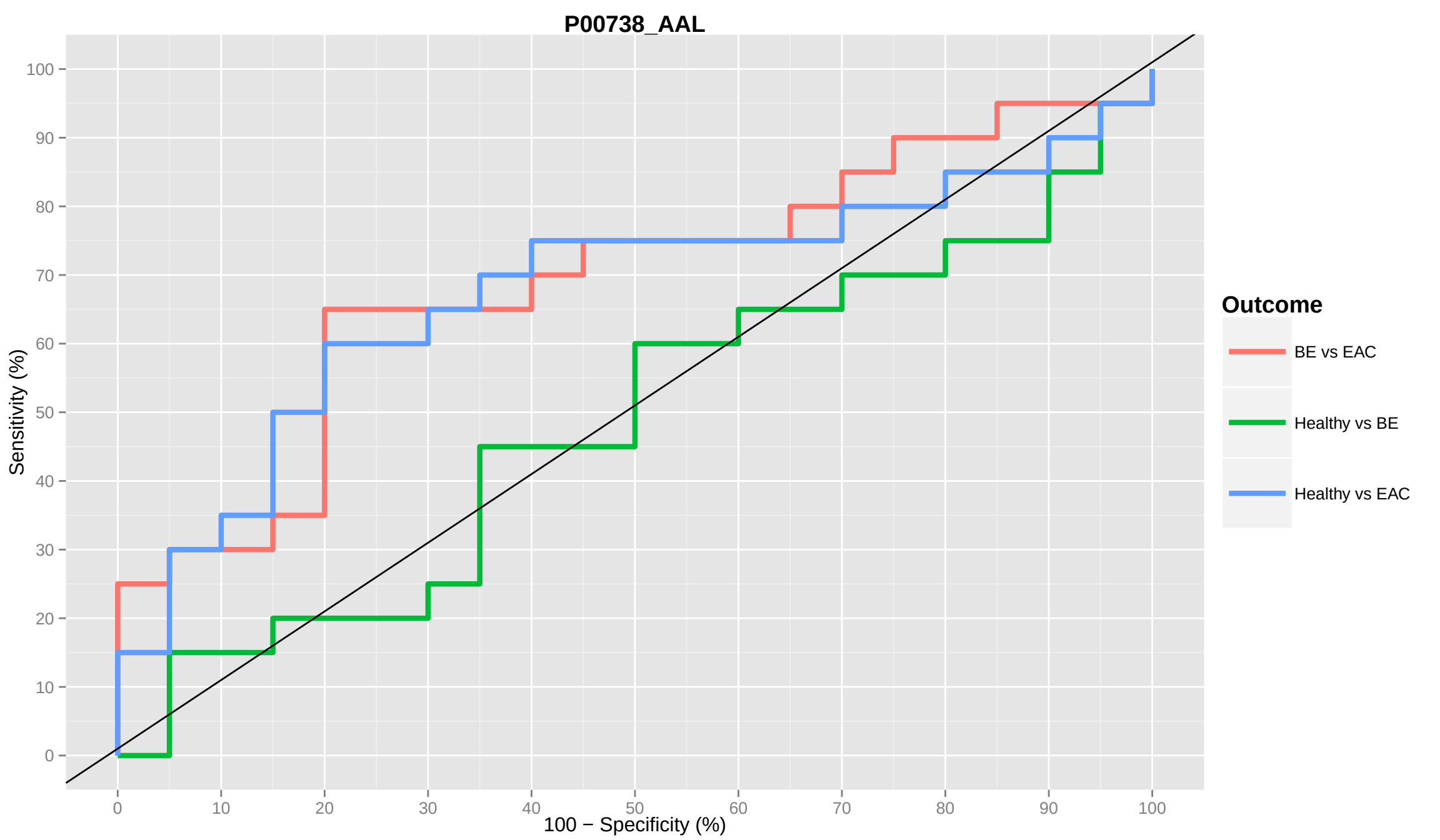

P10643_PSA
6e+05
5e+05
4e+05
3e+05
2e+05
P10643_PSA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

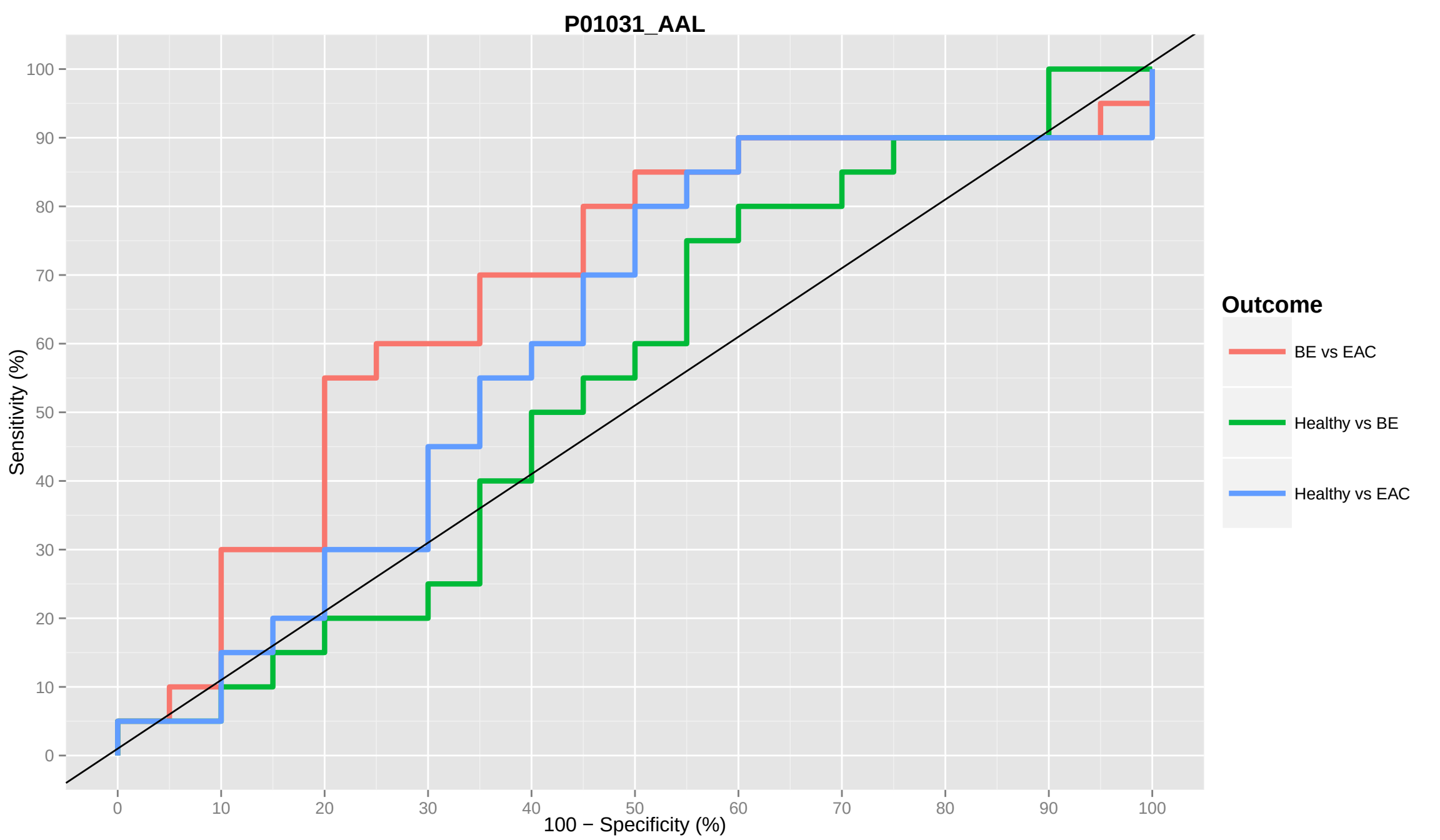

P00738_WGA
4e+07
3e+07
2e+07
1e+07
0e+00
P00738_WGA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

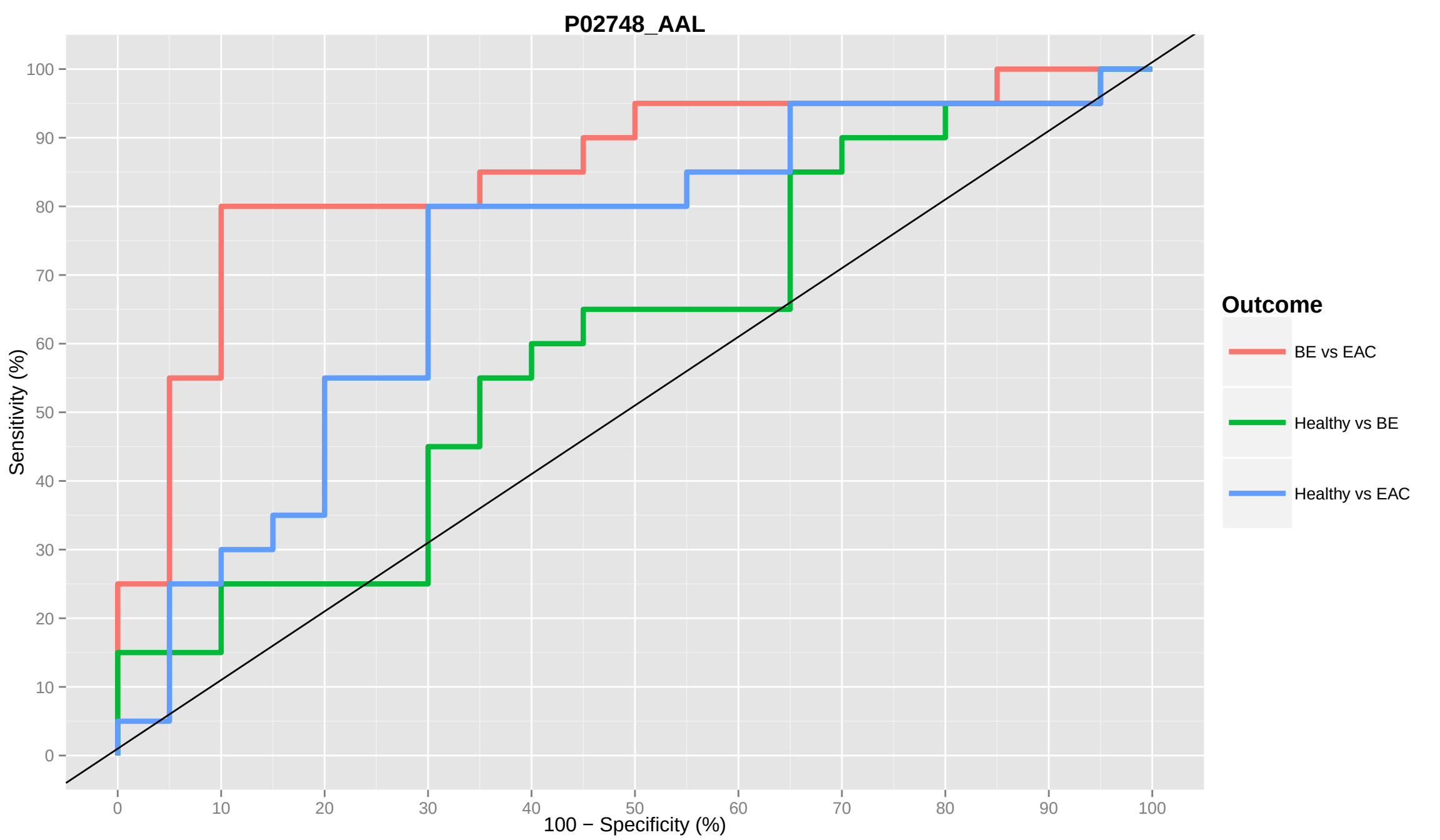

P01011_WGA
1500000
1000000
500000
0
P01011_WGA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

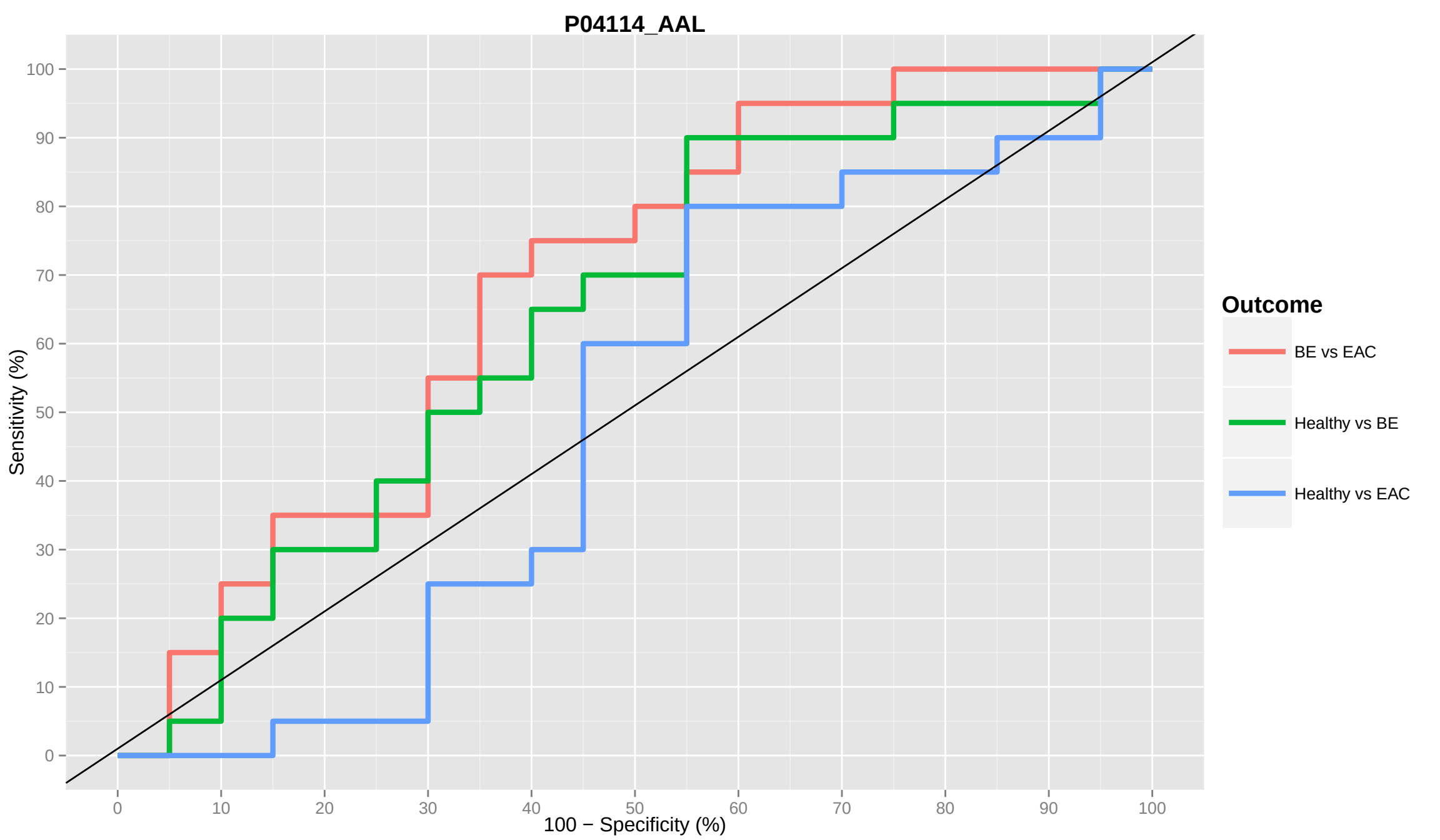

P02748_WGA
750000
500000
250000
0
P02748_WGA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

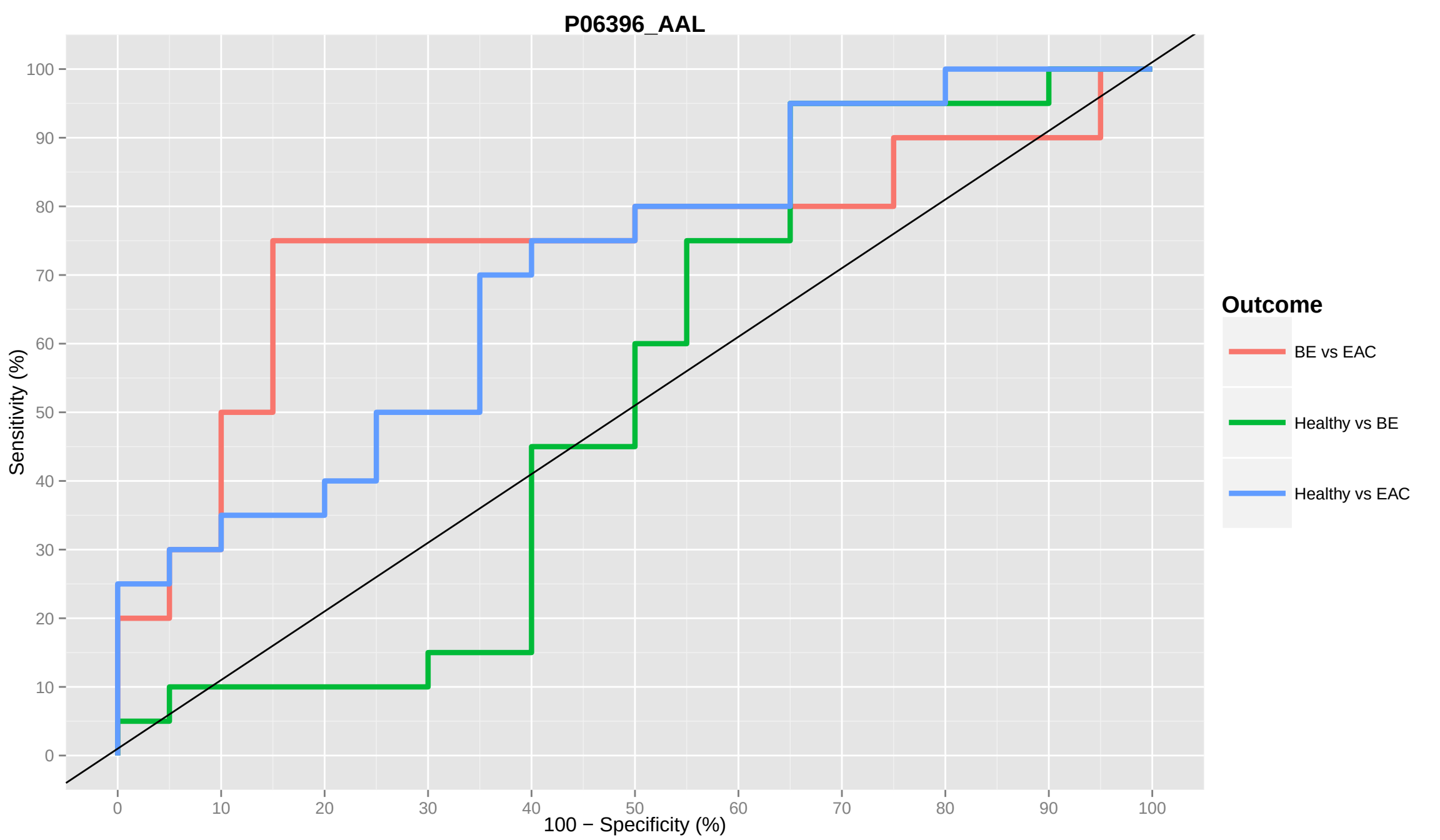

P06396_WGA
4e+05
3e+05
2e+05
1e+05
0e+00
P06396_WGA
Healthy
BE
EAC
Outcome
Outcome
Healthy
BE
EAC

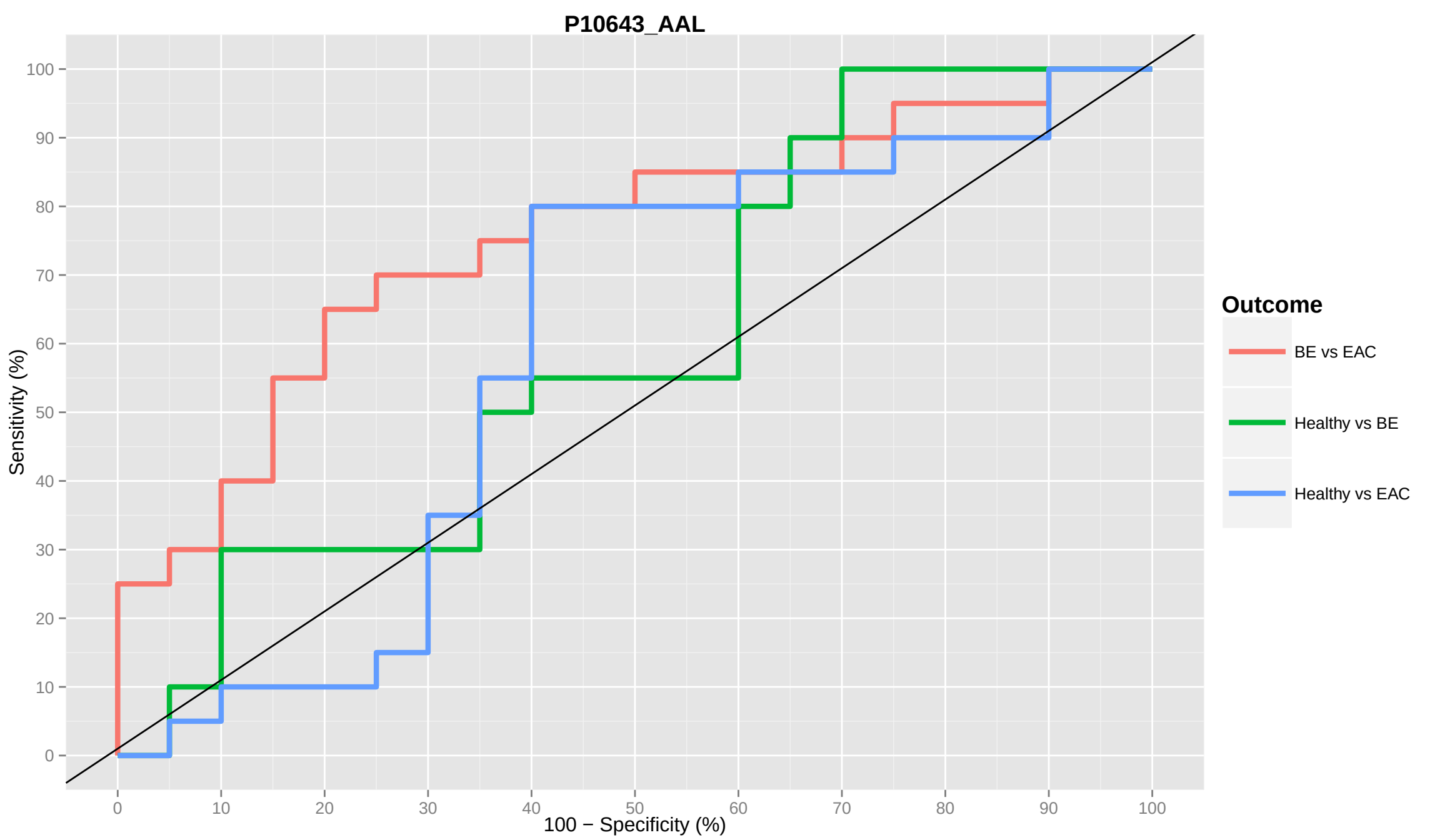

P00738_AAL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

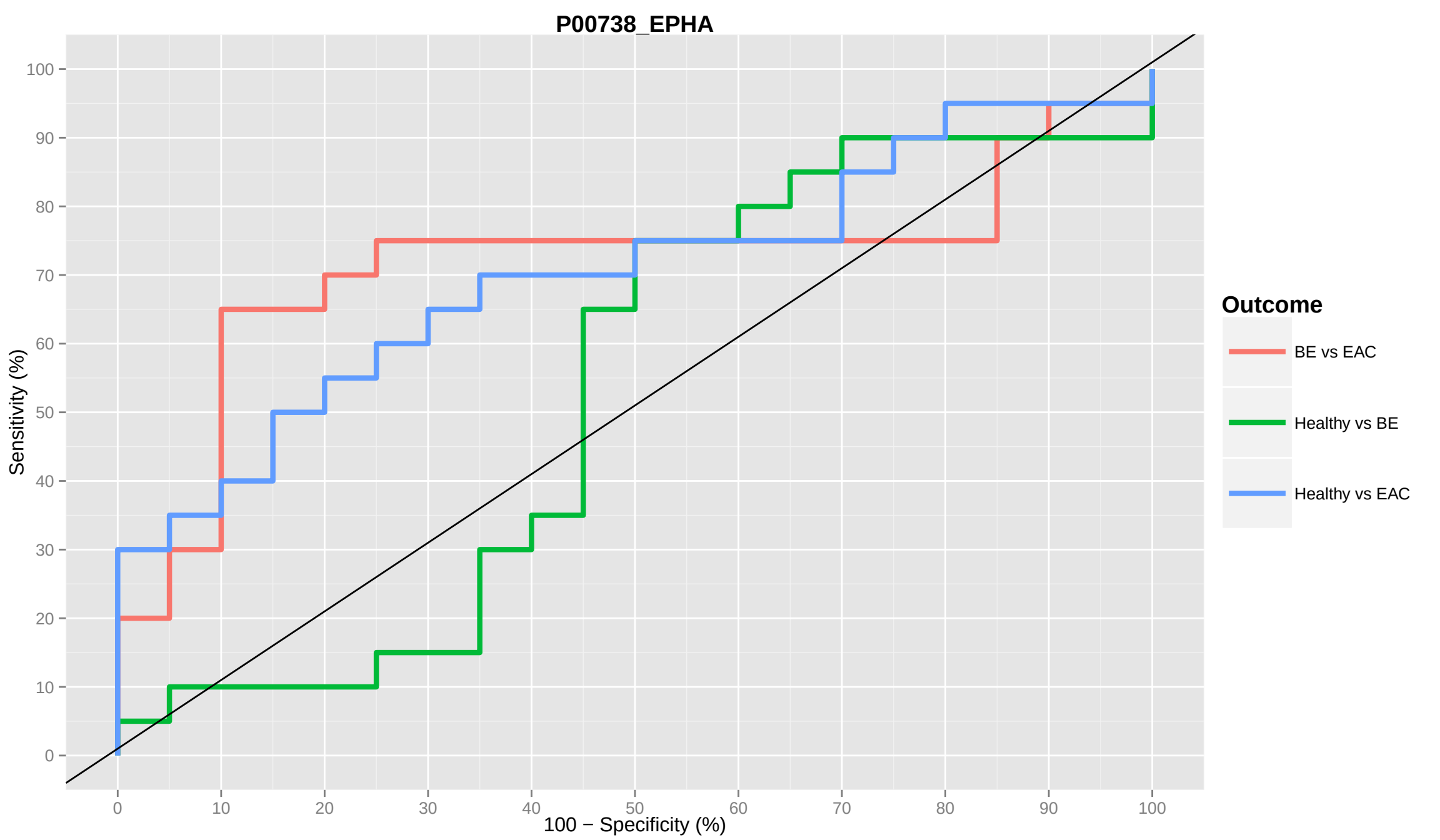

P01031_AAL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

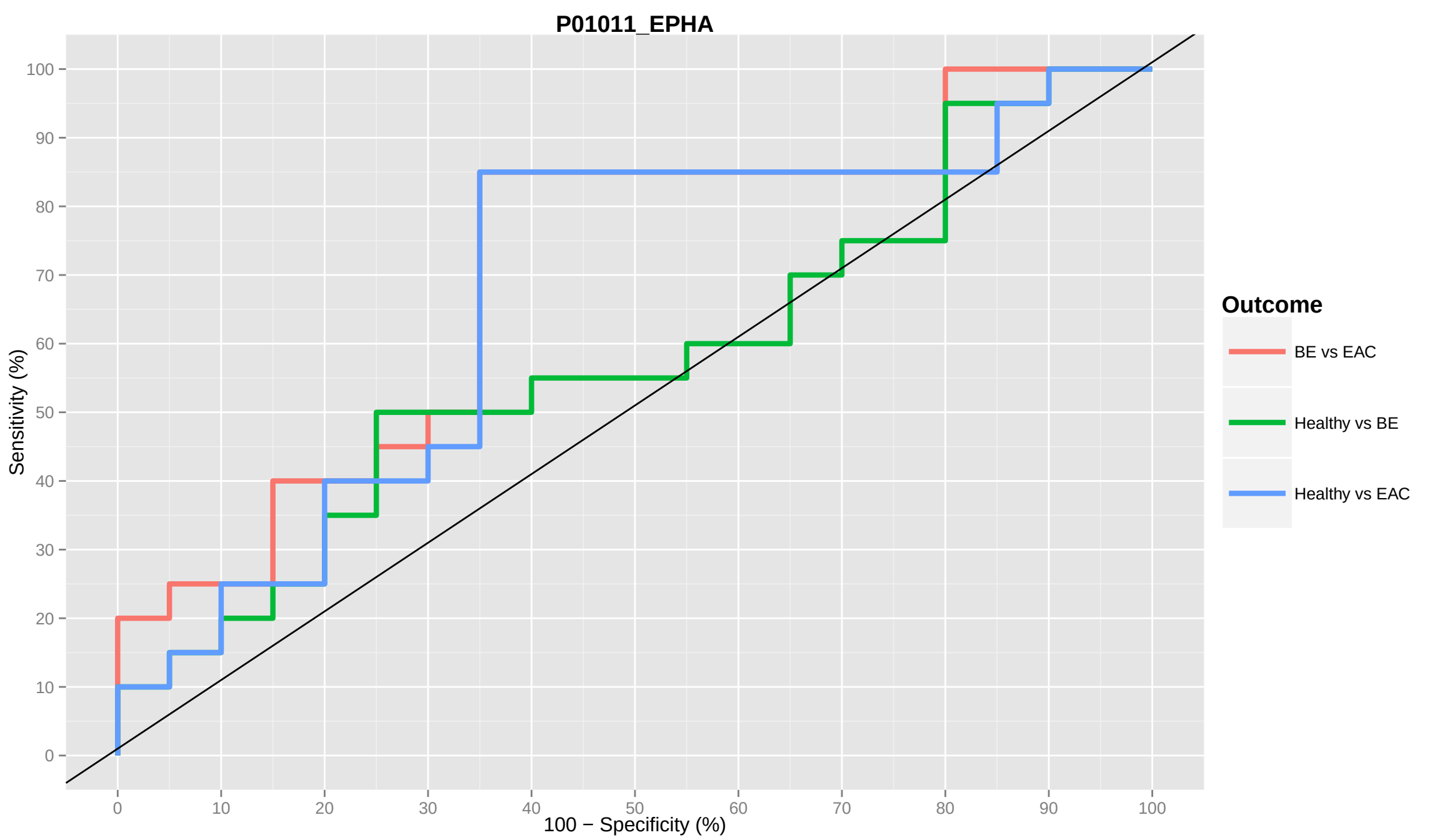

P02748_AAL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

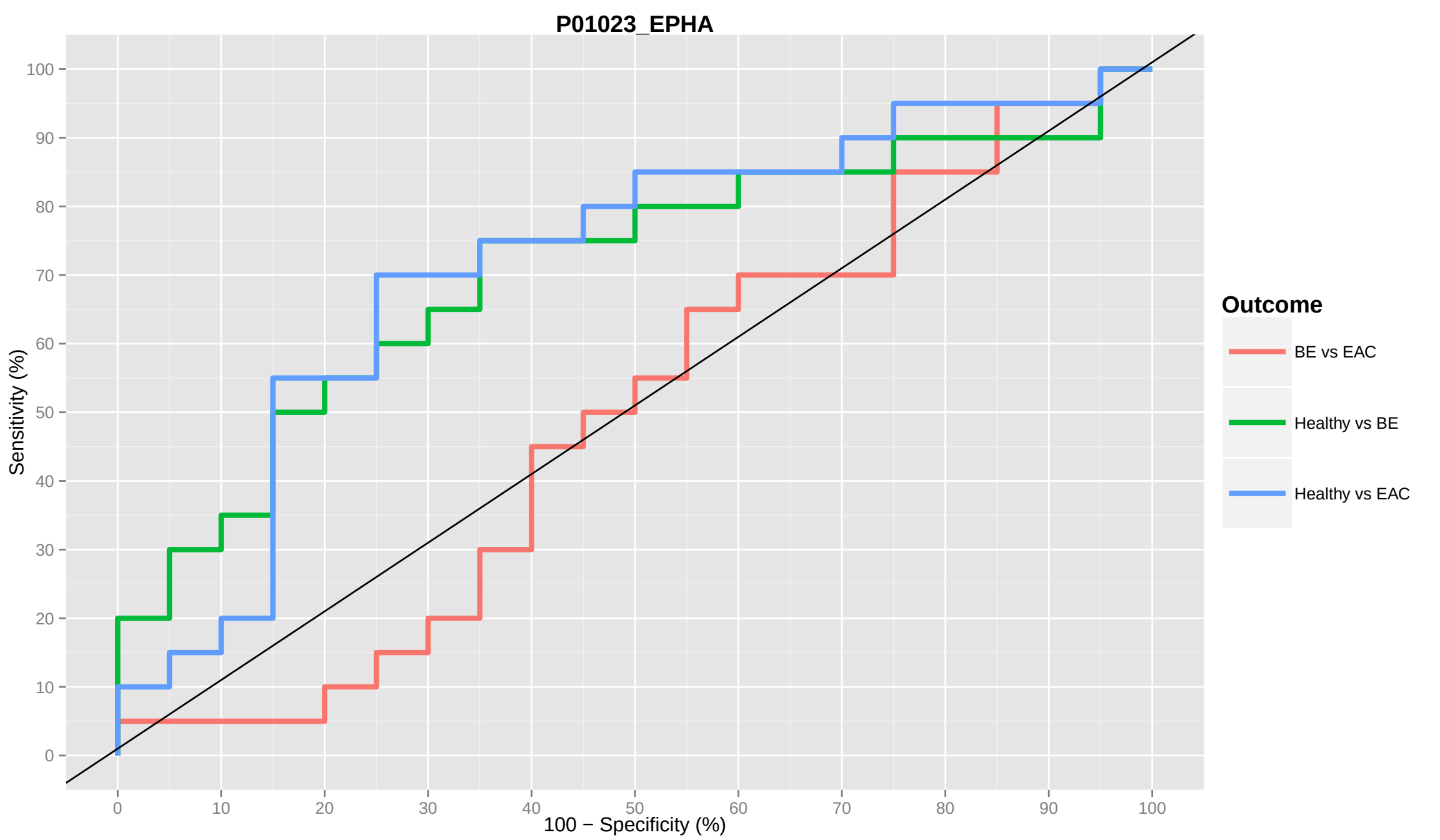

P04114_AAL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

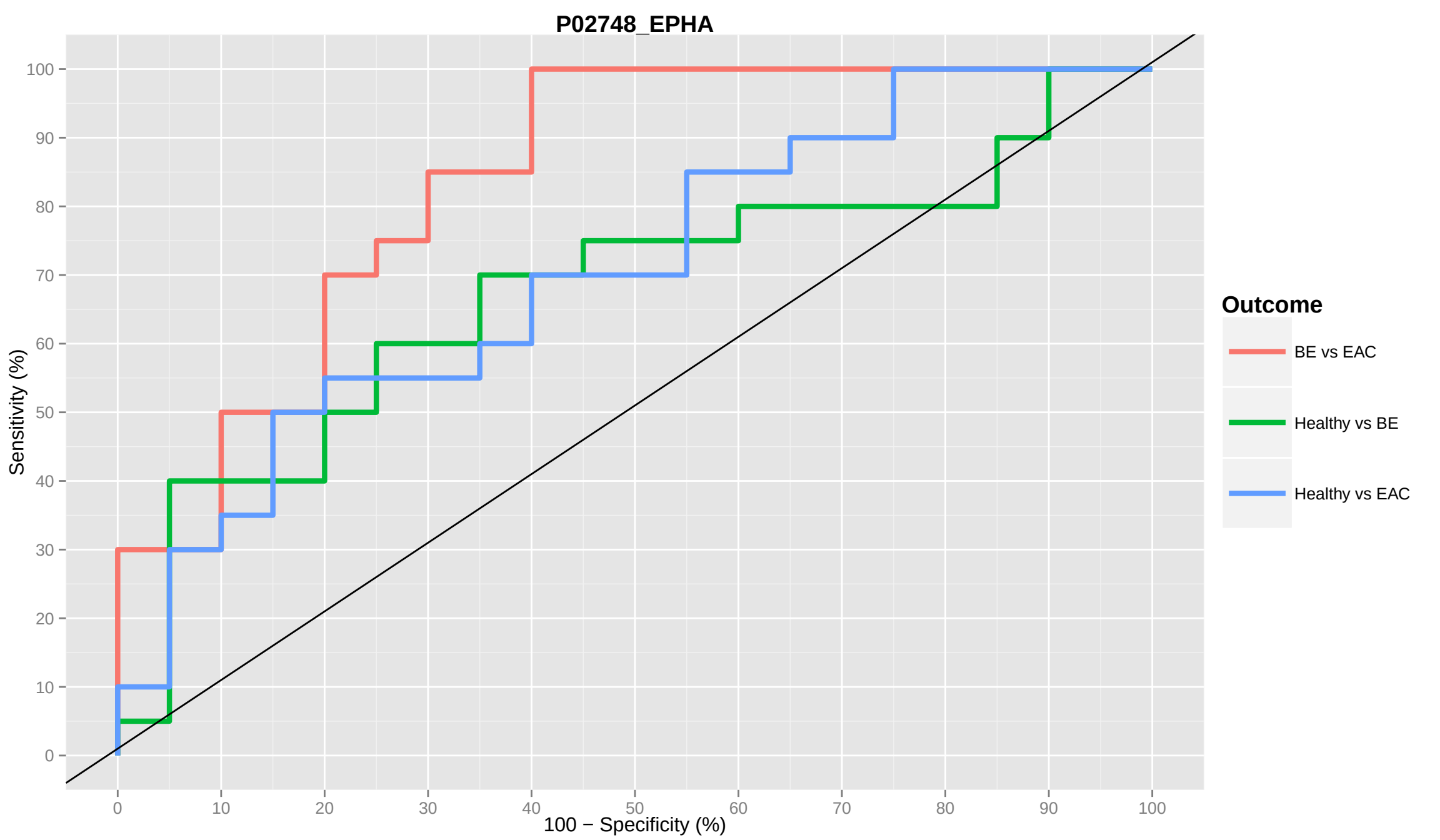

P06396_AAL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

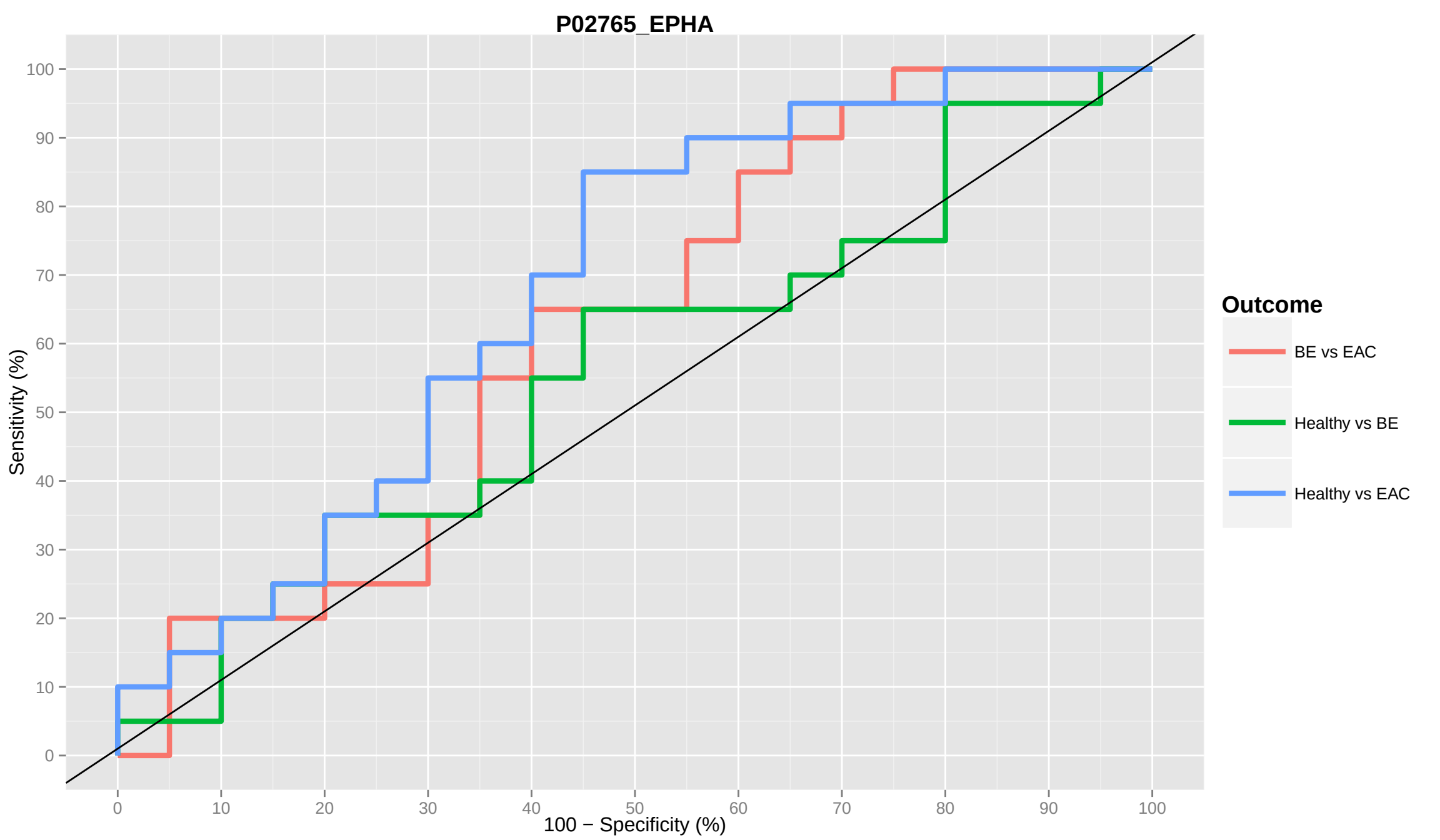

P10643_AAL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

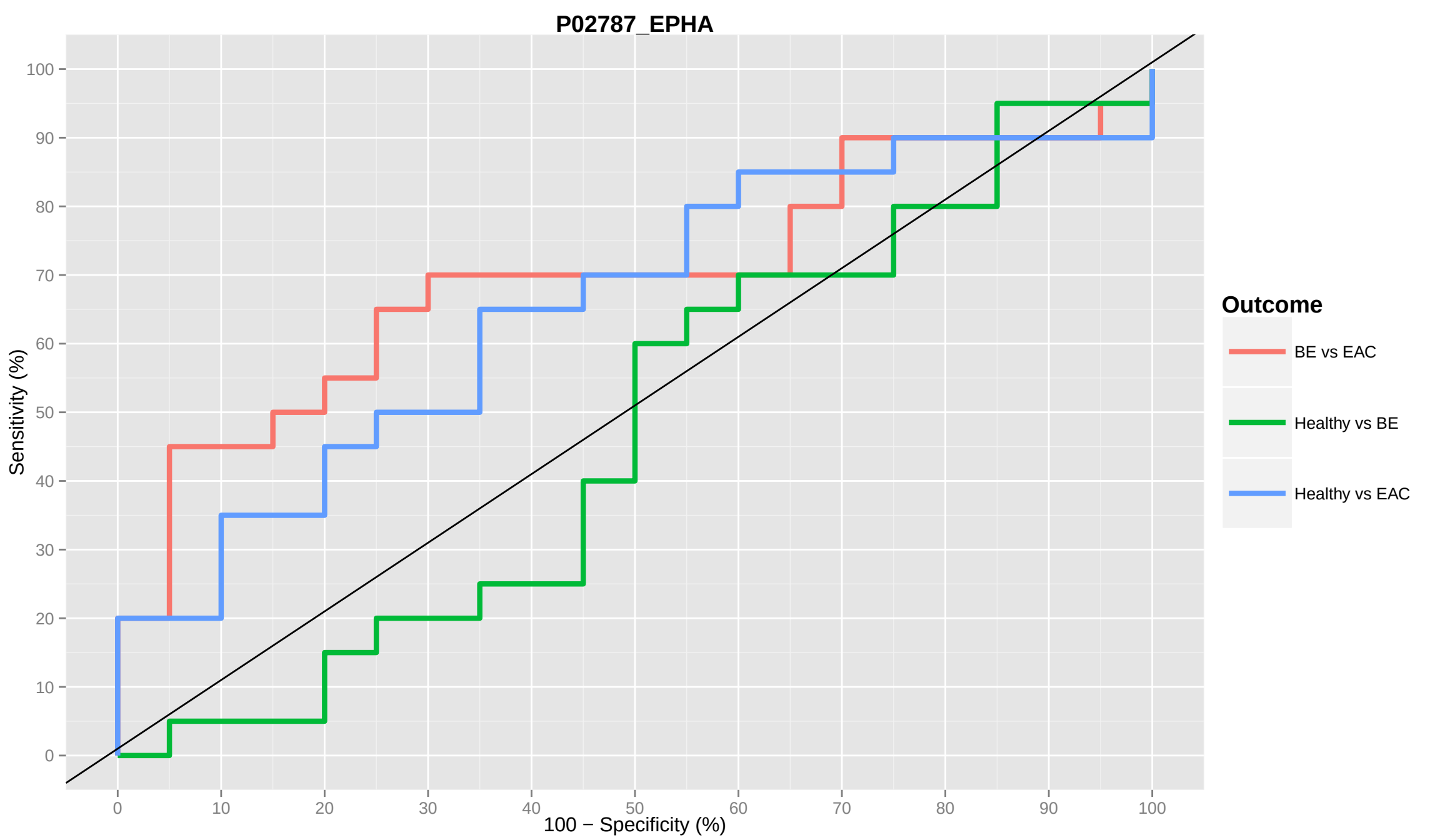

P00738_EPHA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

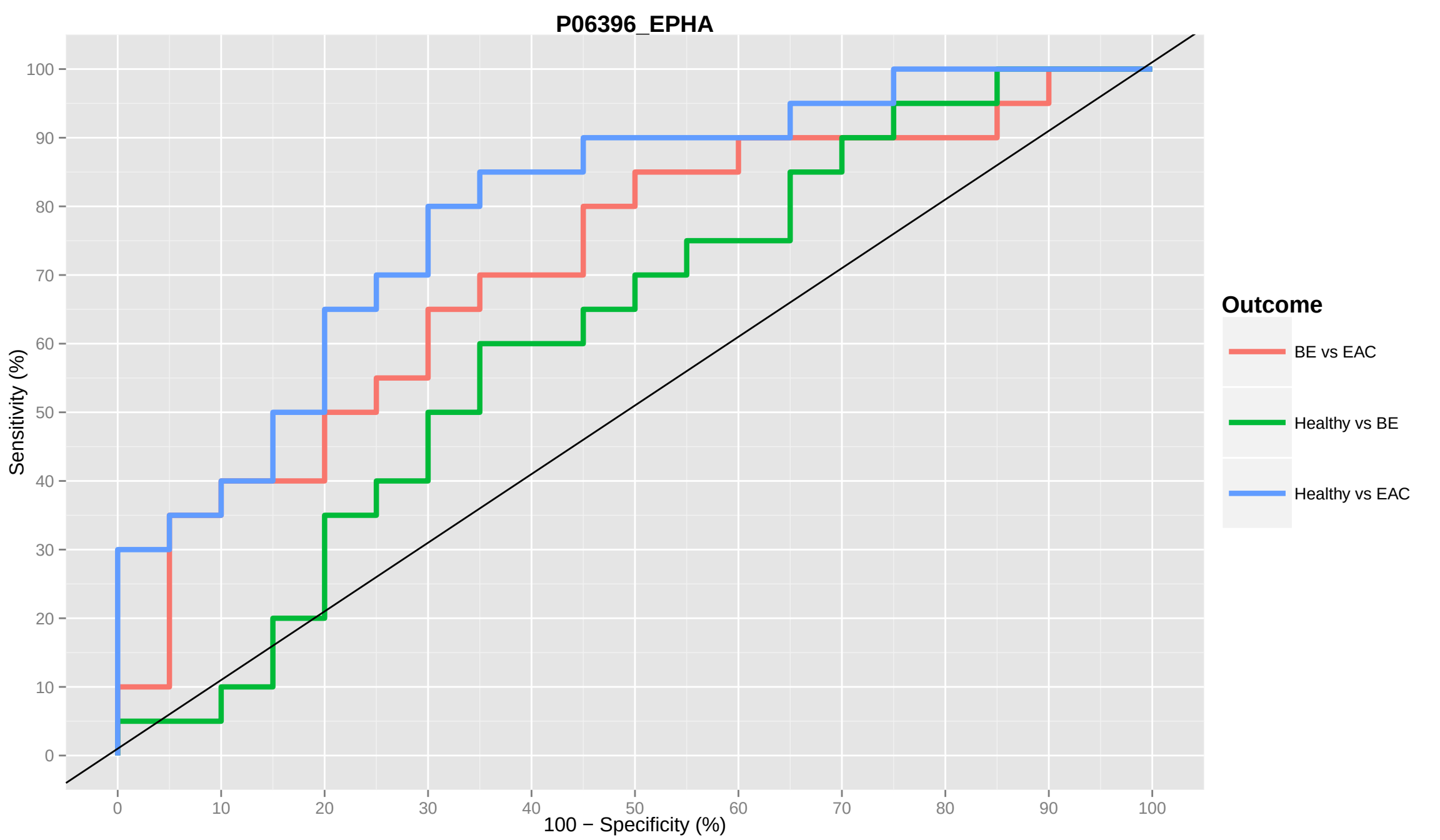

P01011_EPHA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

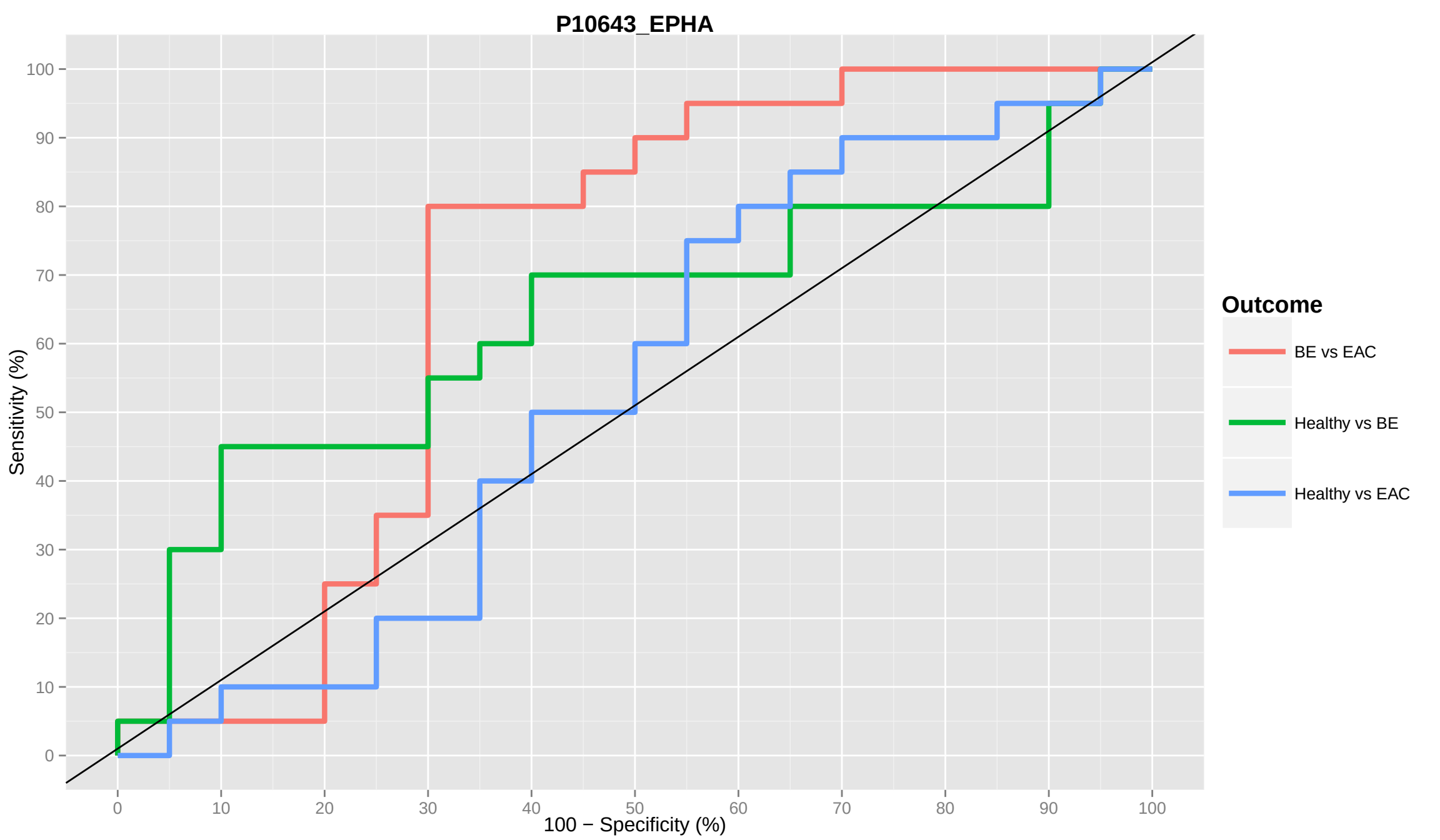

P01023_EPHA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

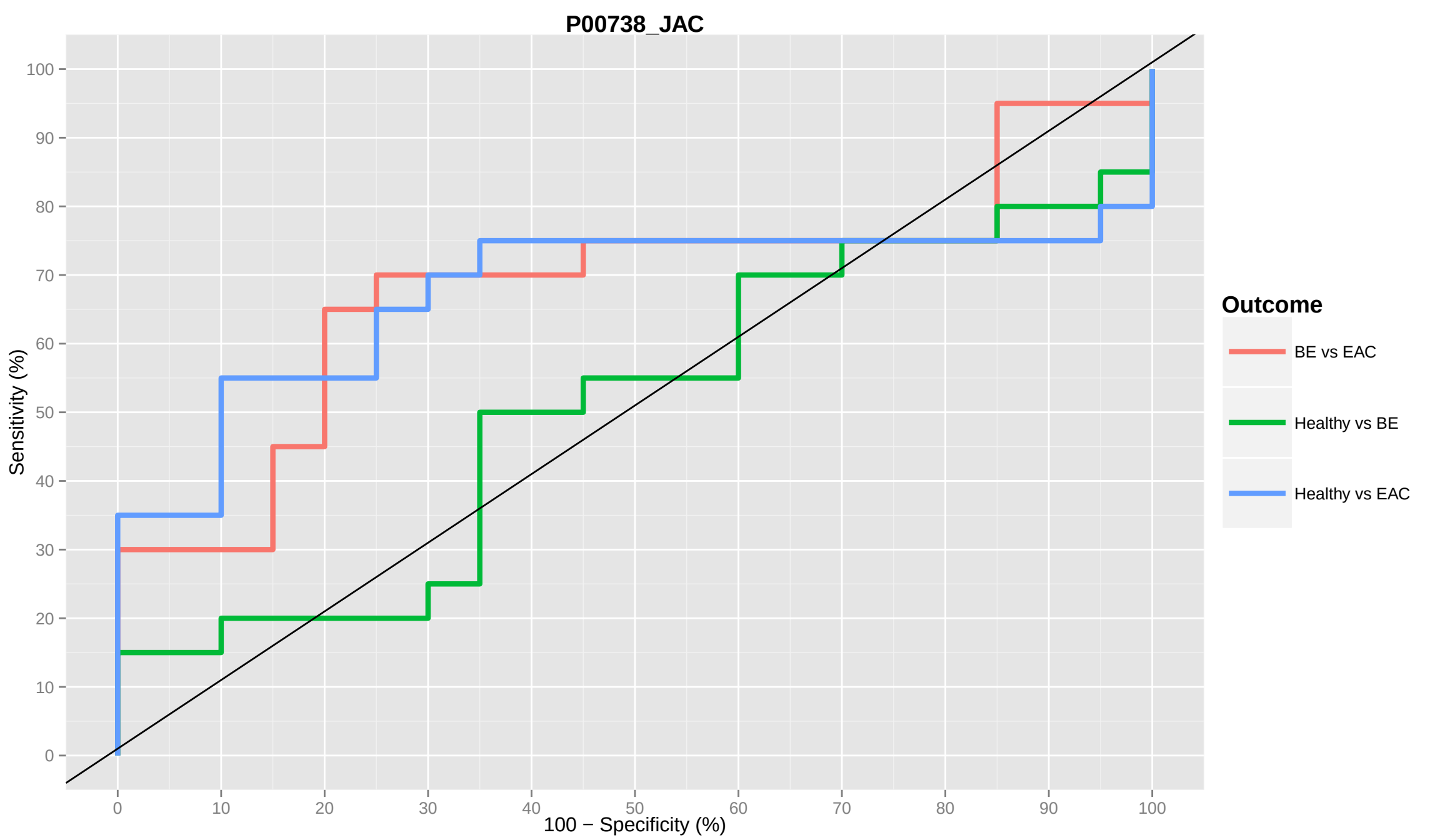

P02748_EPHA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

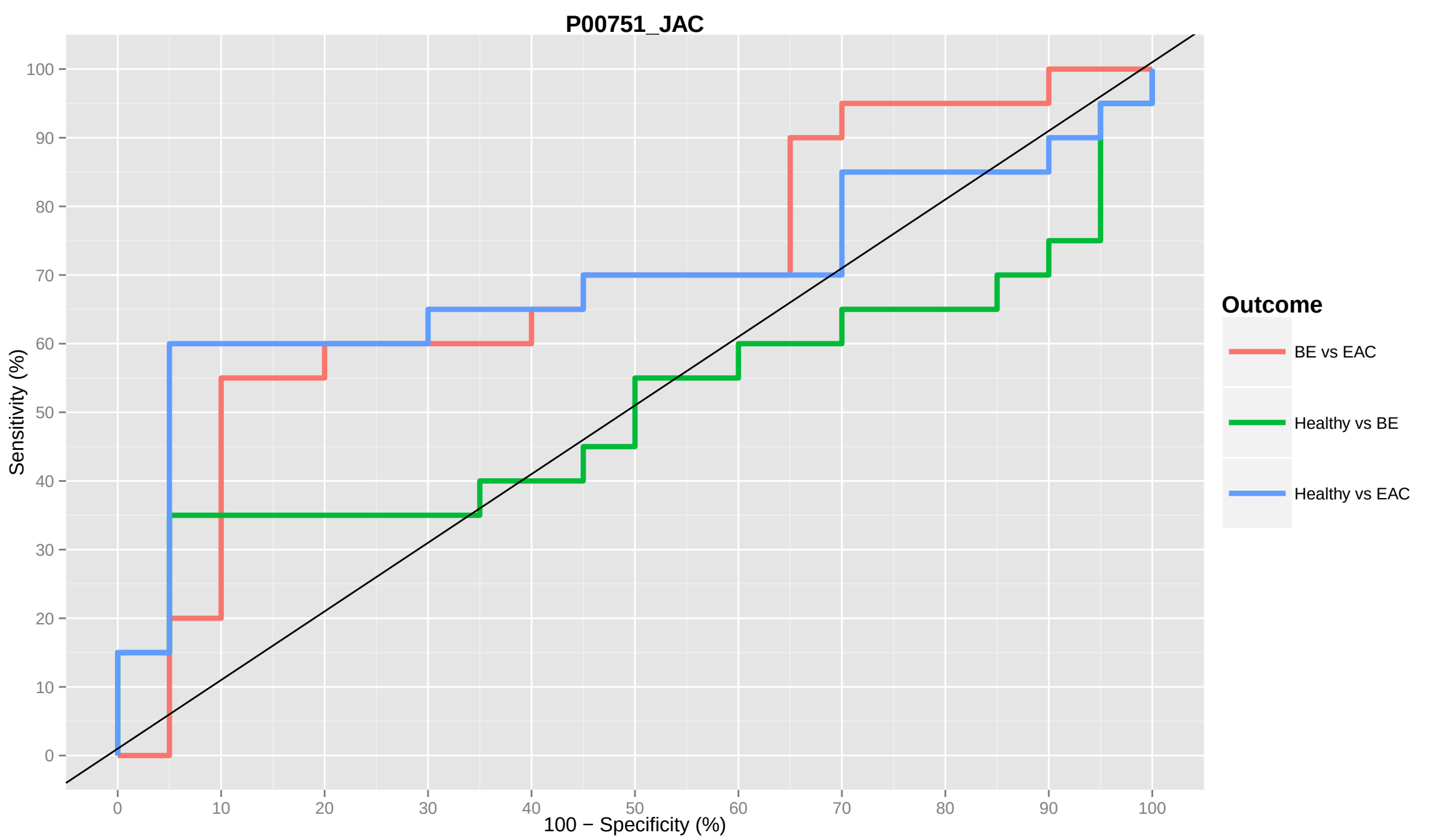

P02765_EPHA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

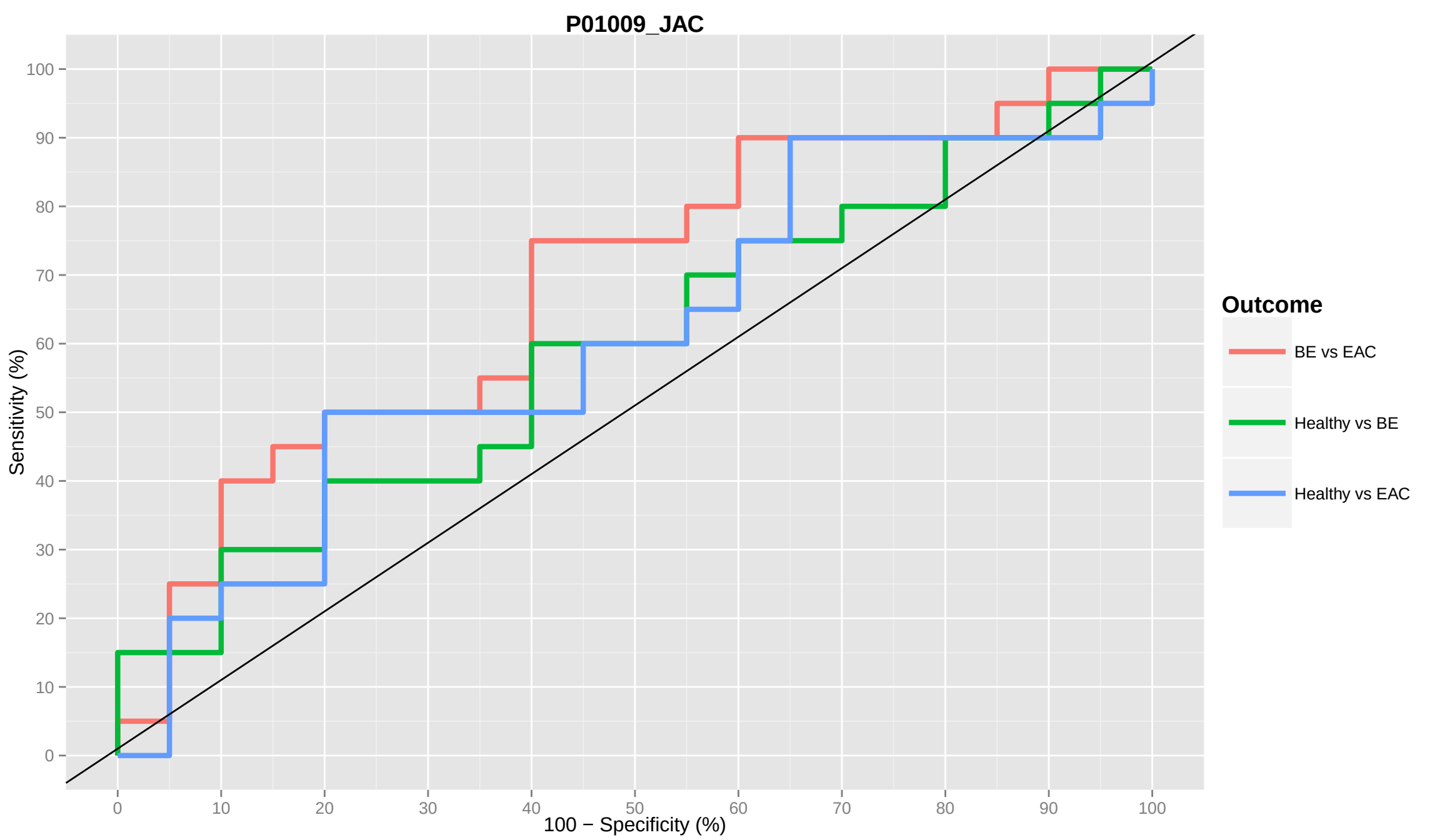

P02787_EPHA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

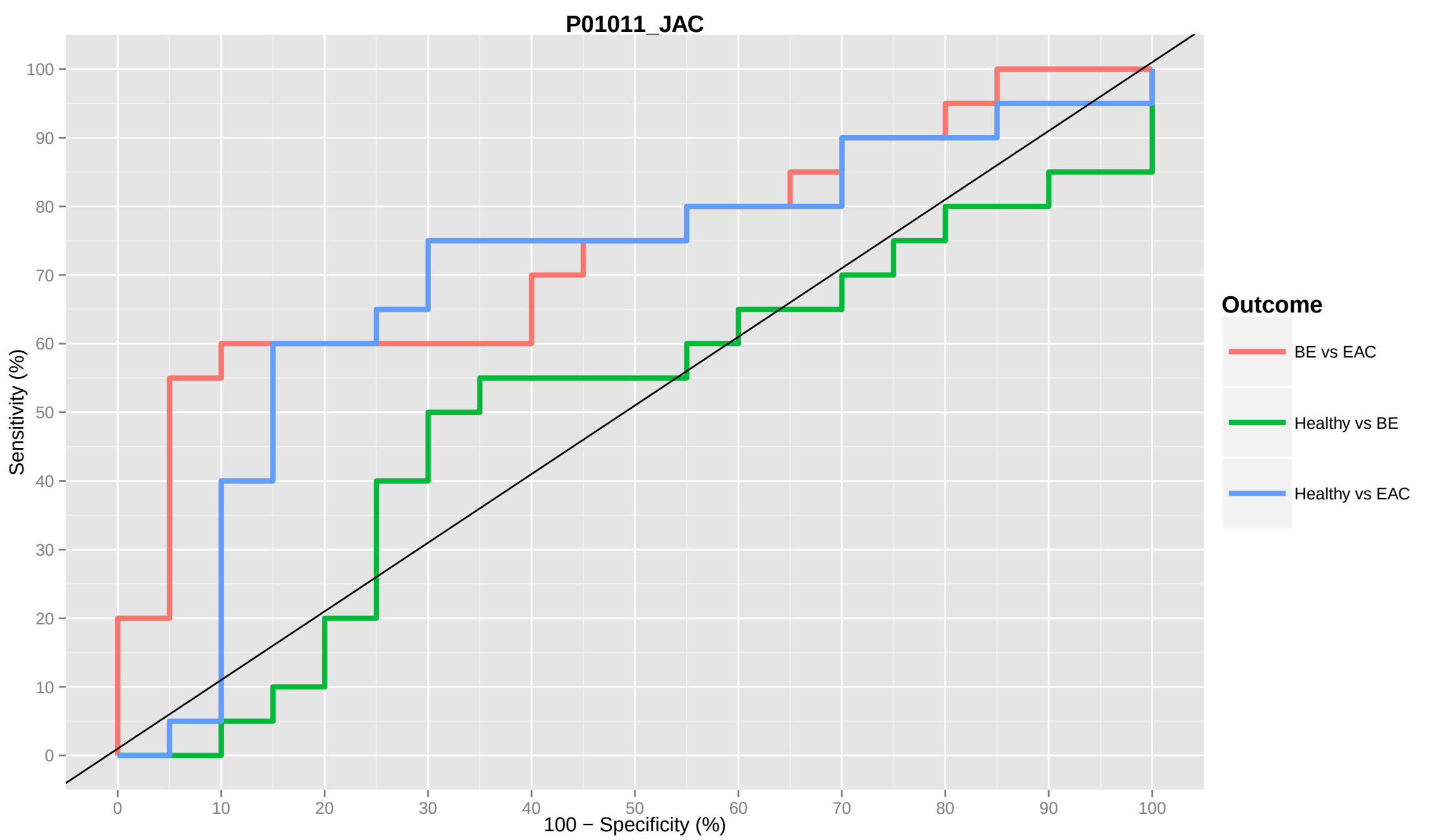

P06396_EPHA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

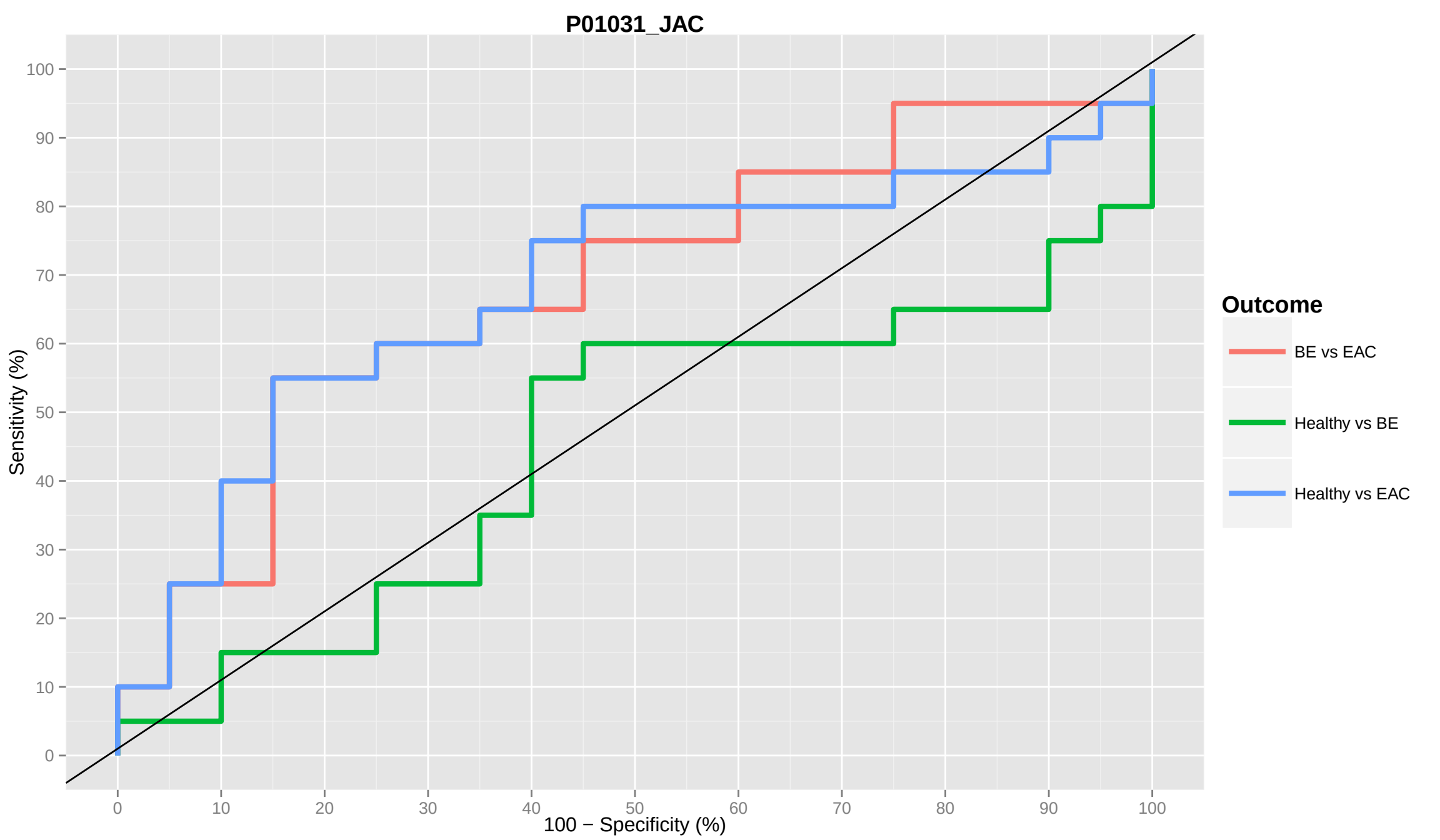

P10643_EPHA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

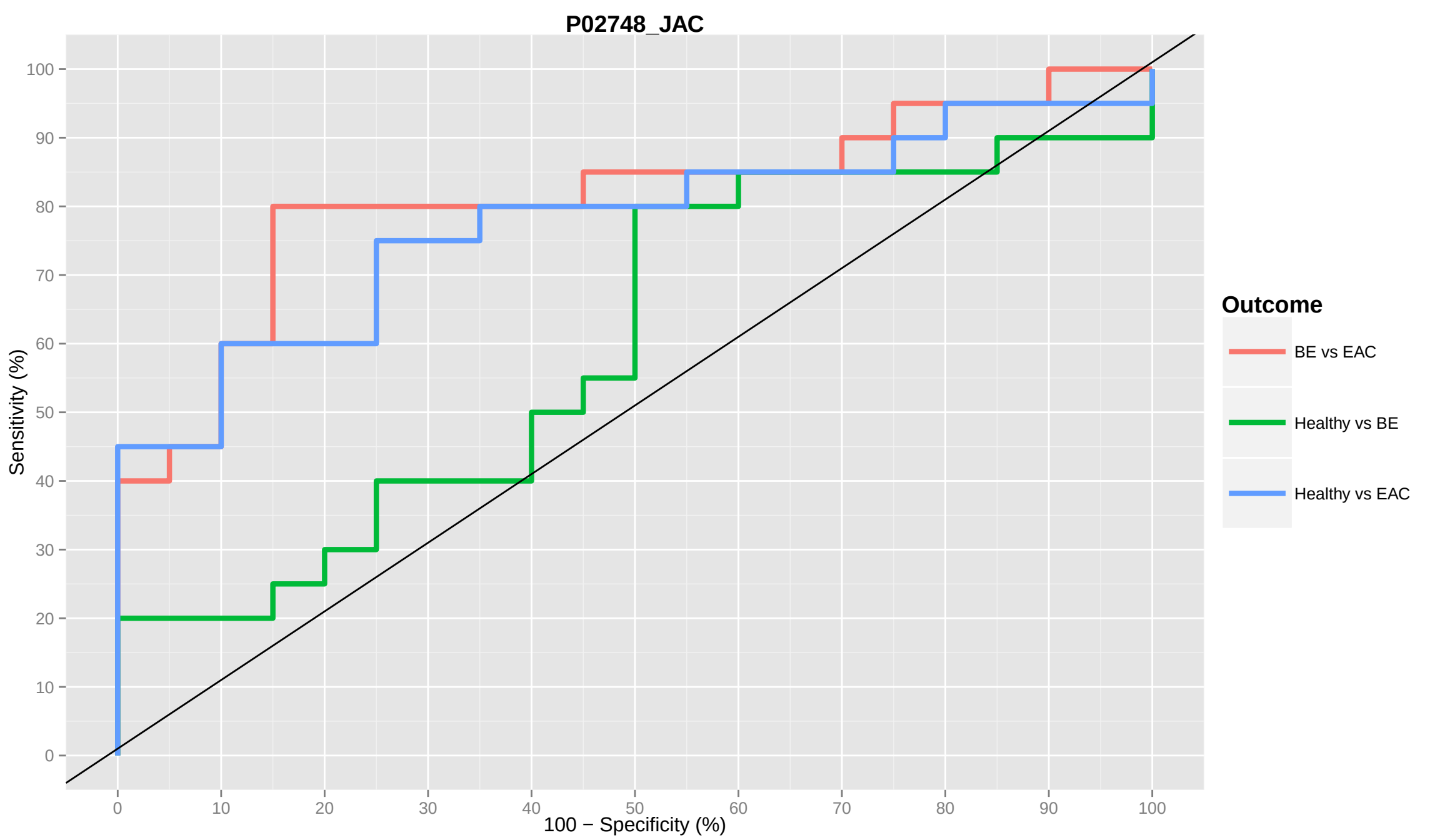

P00738_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

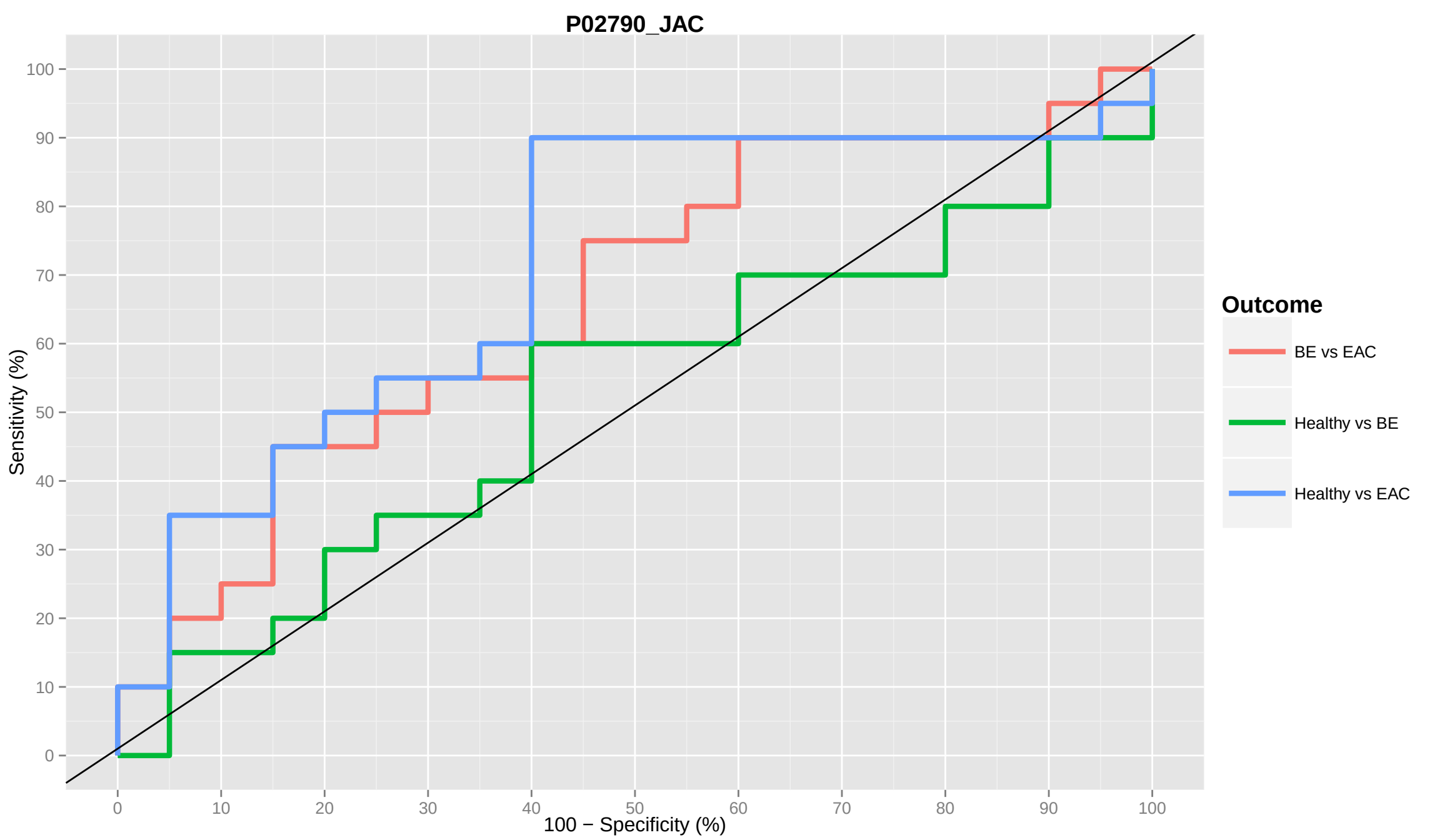

P00751_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

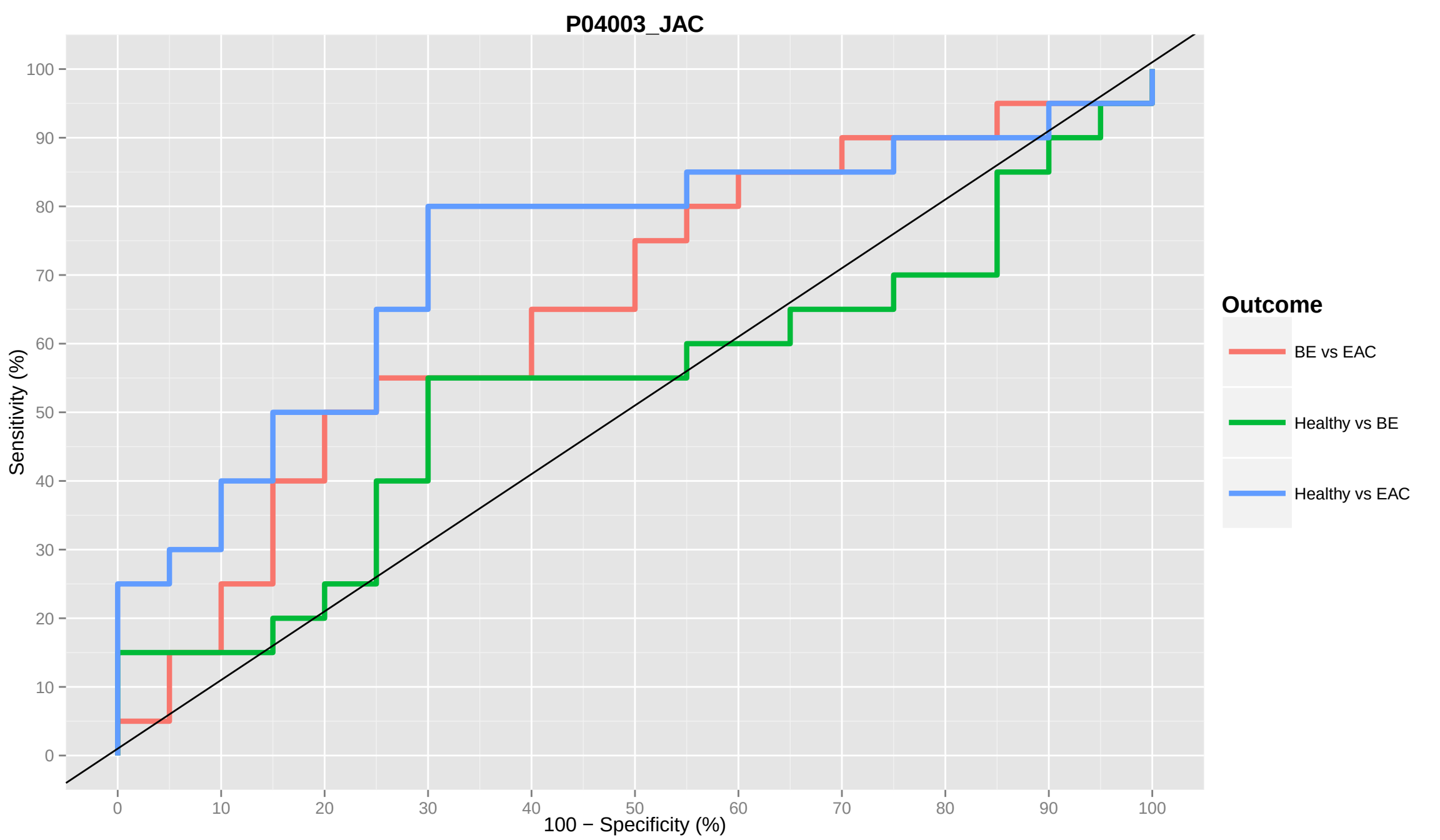

P01009_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

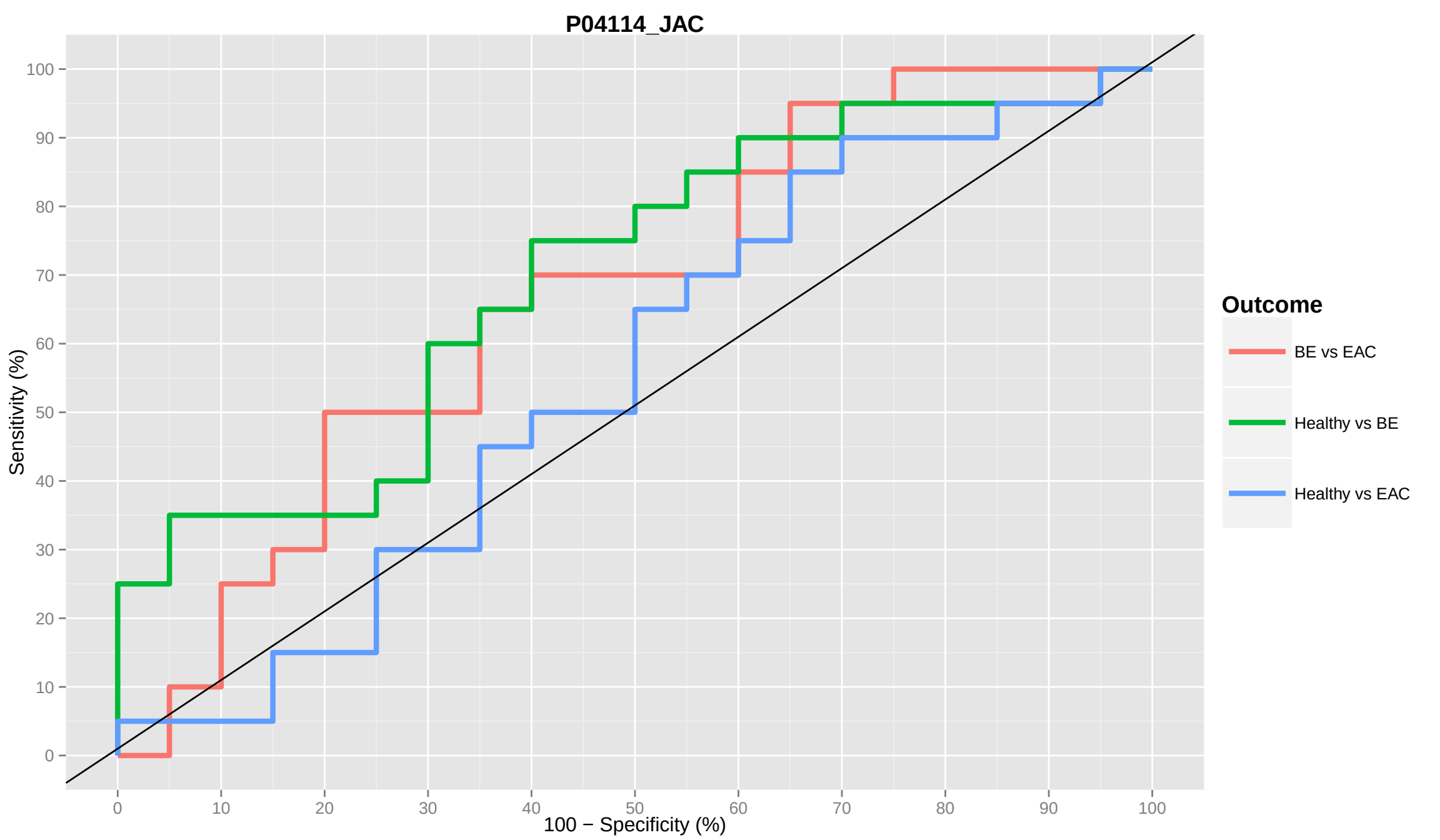

P01011_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

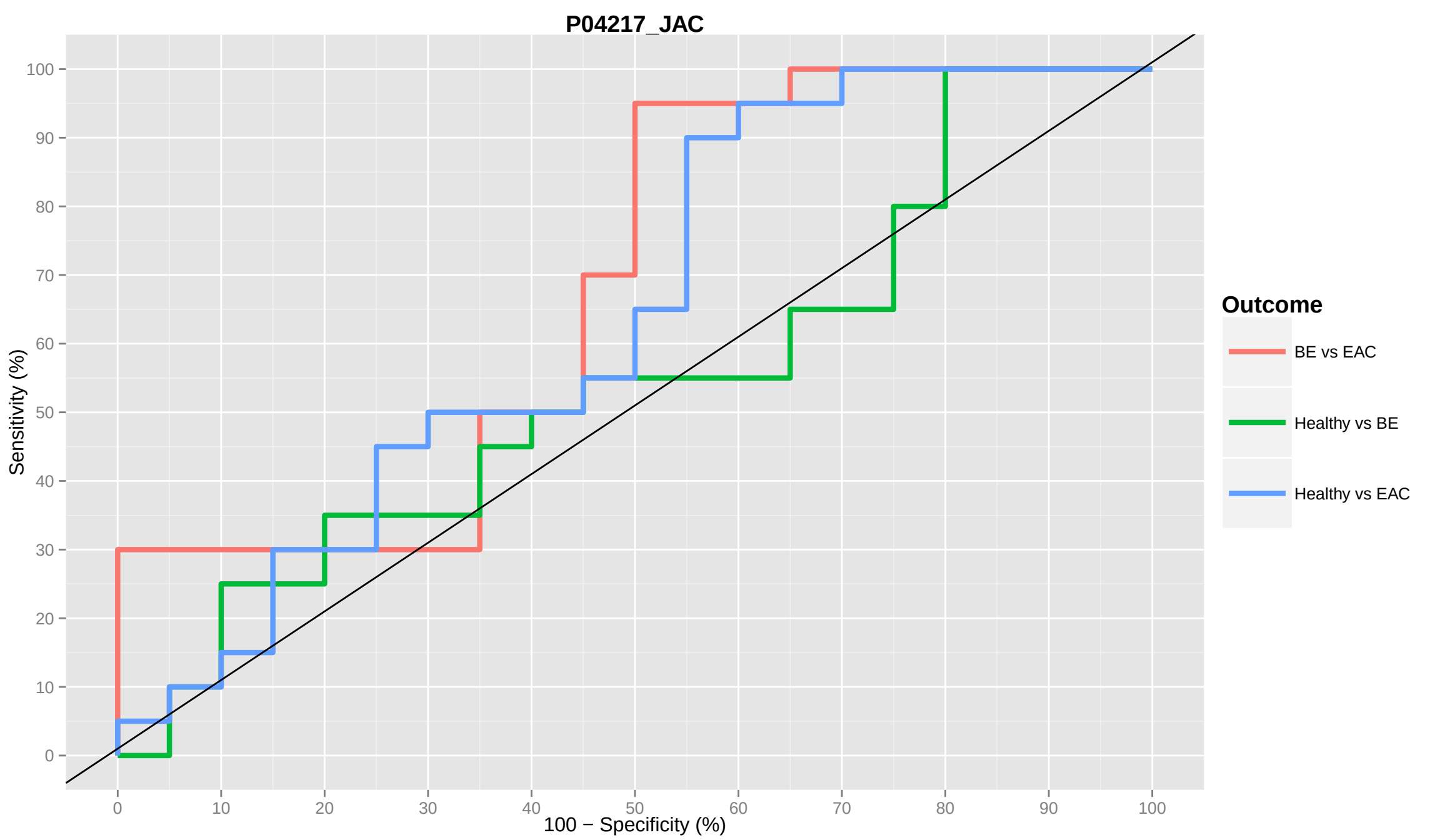

P01031_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

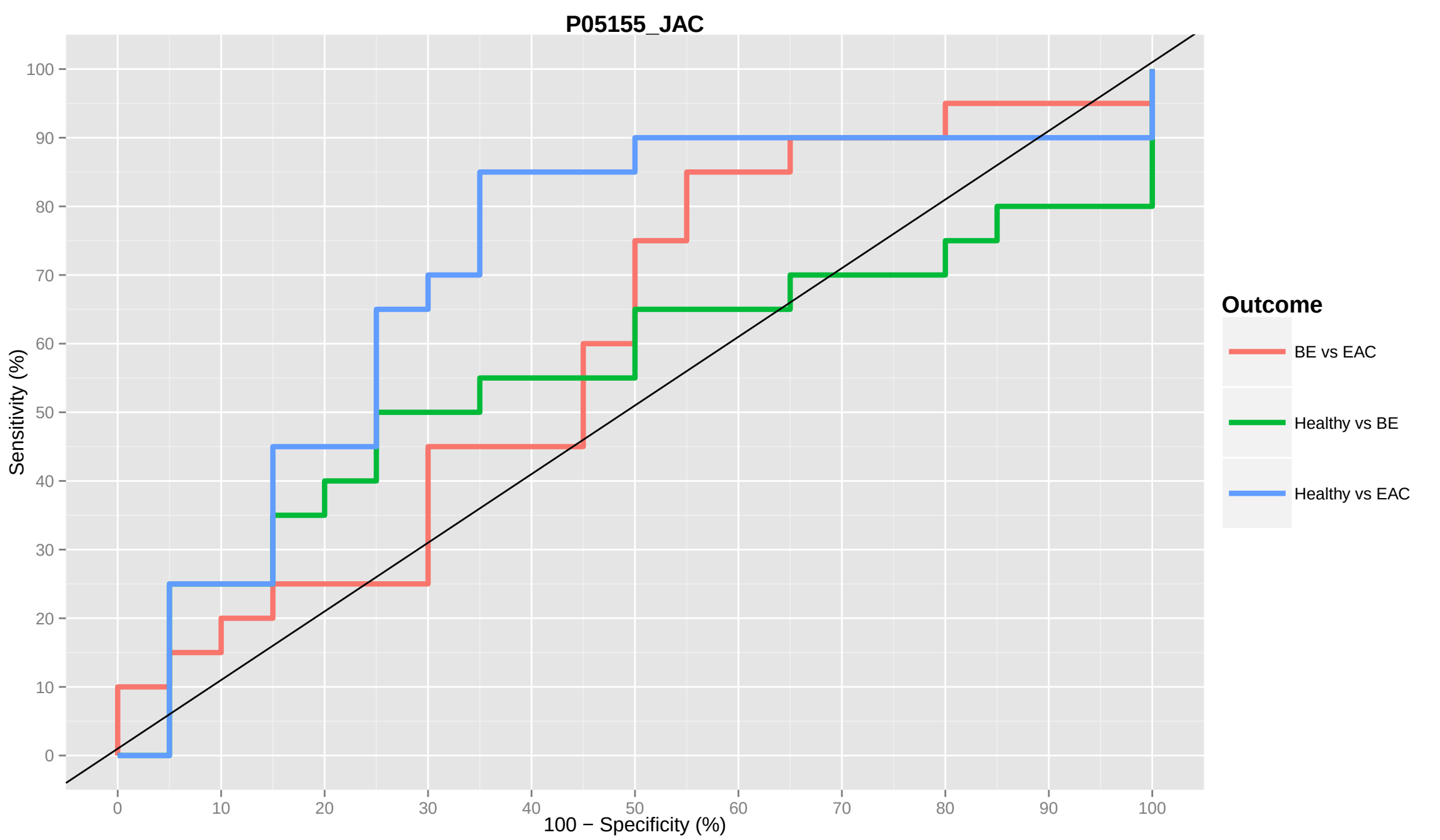

P02748_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

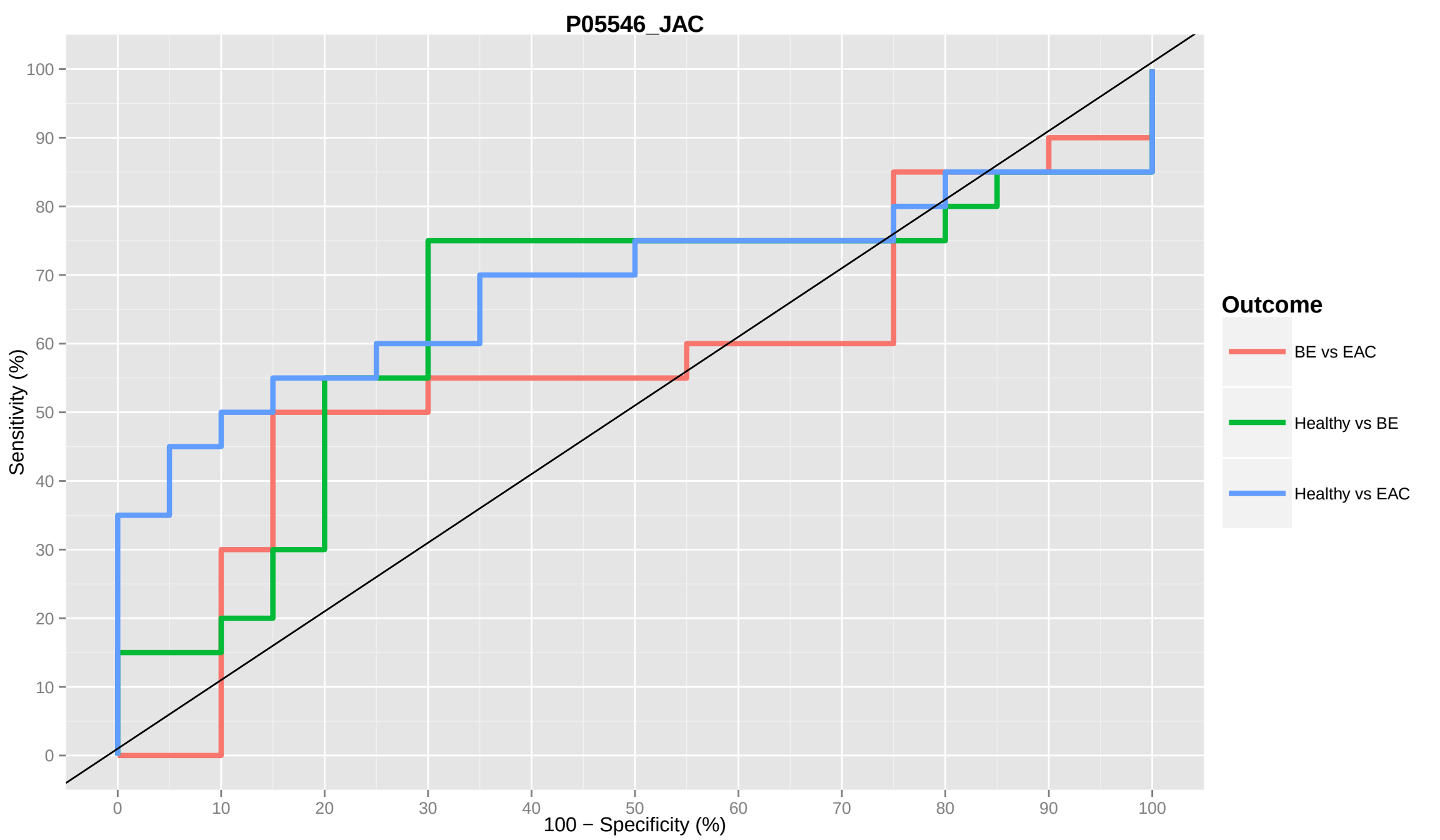

P02790_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

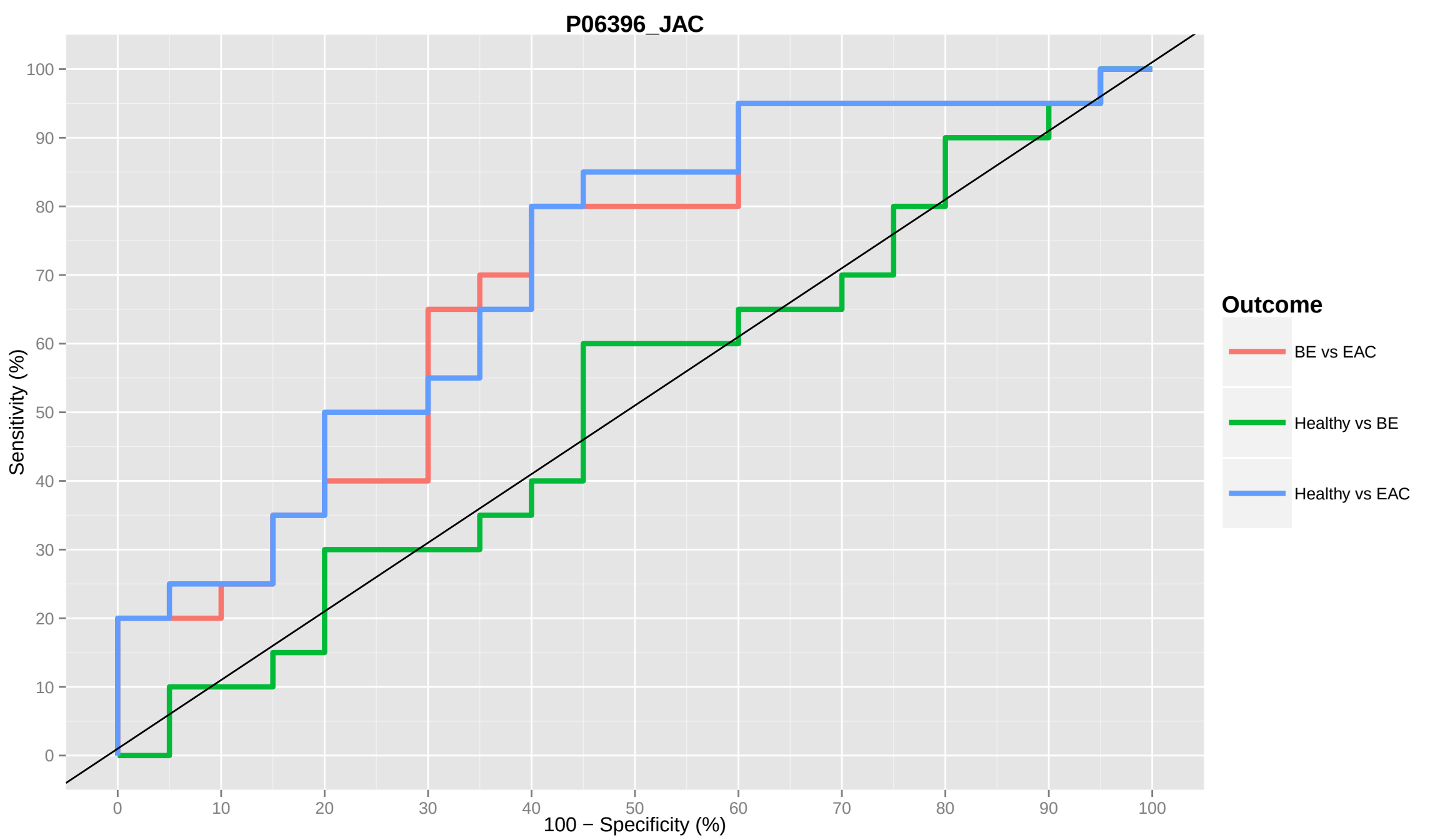

P04003_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

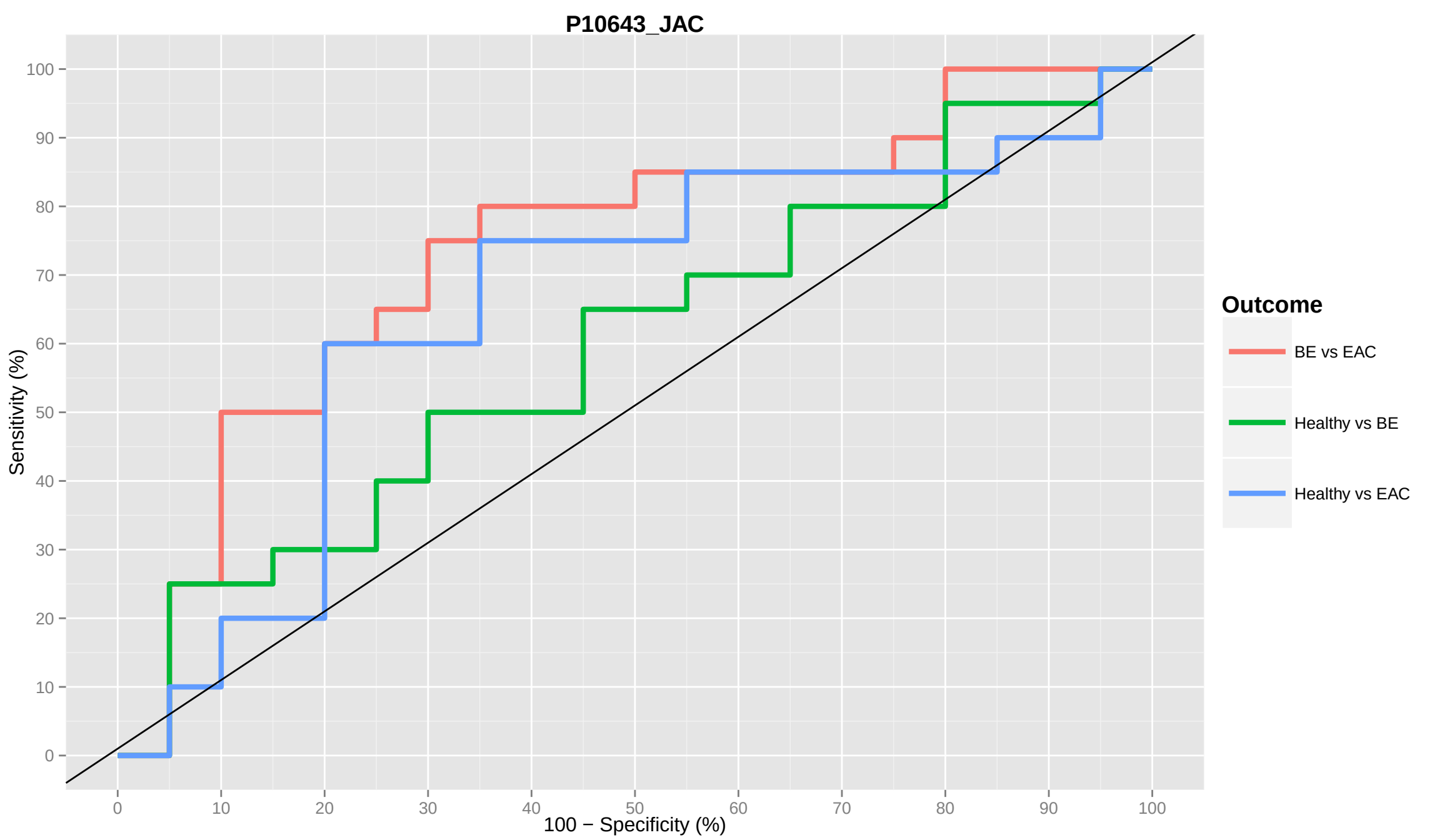

P04114_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

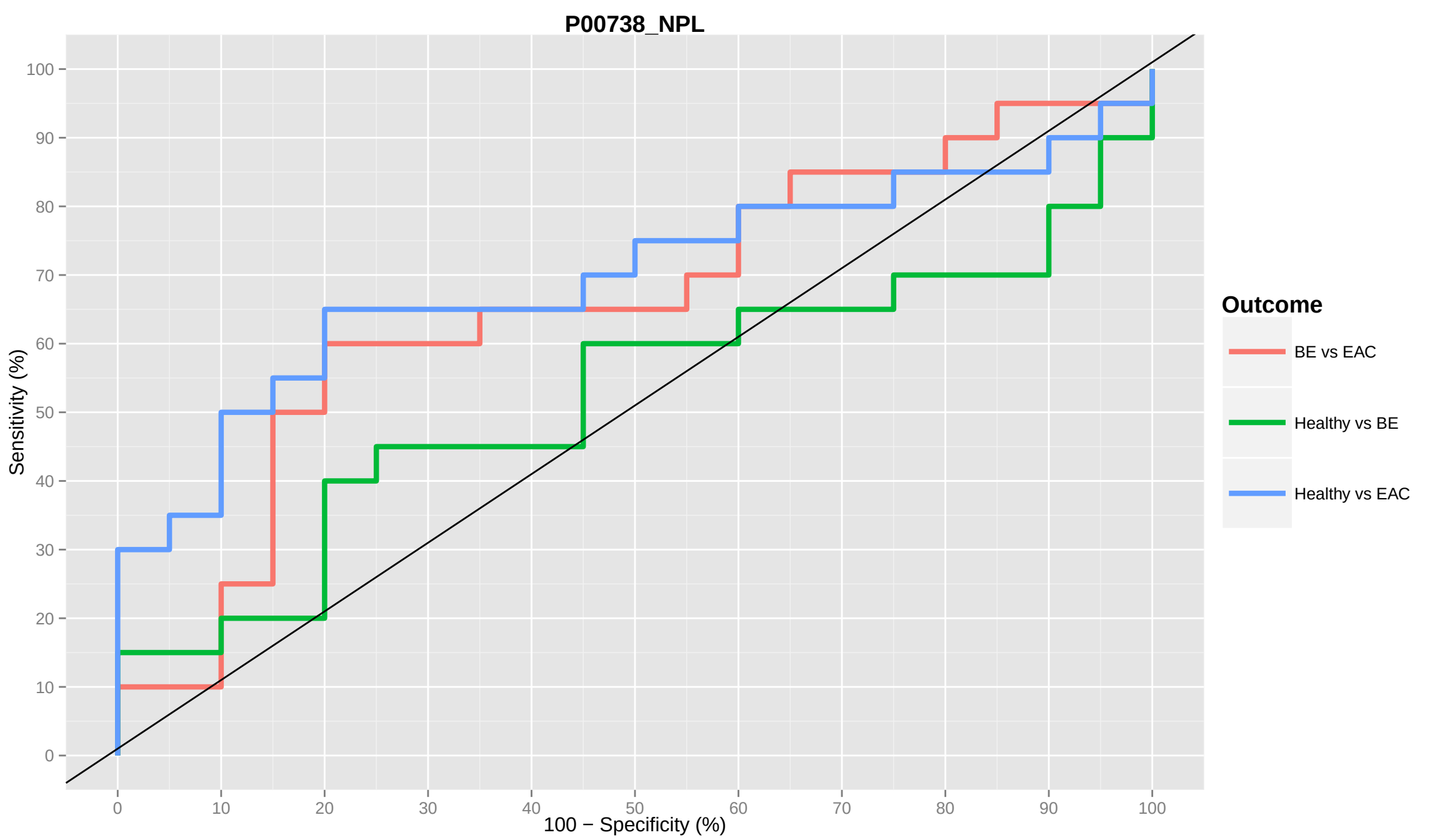

P04217_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

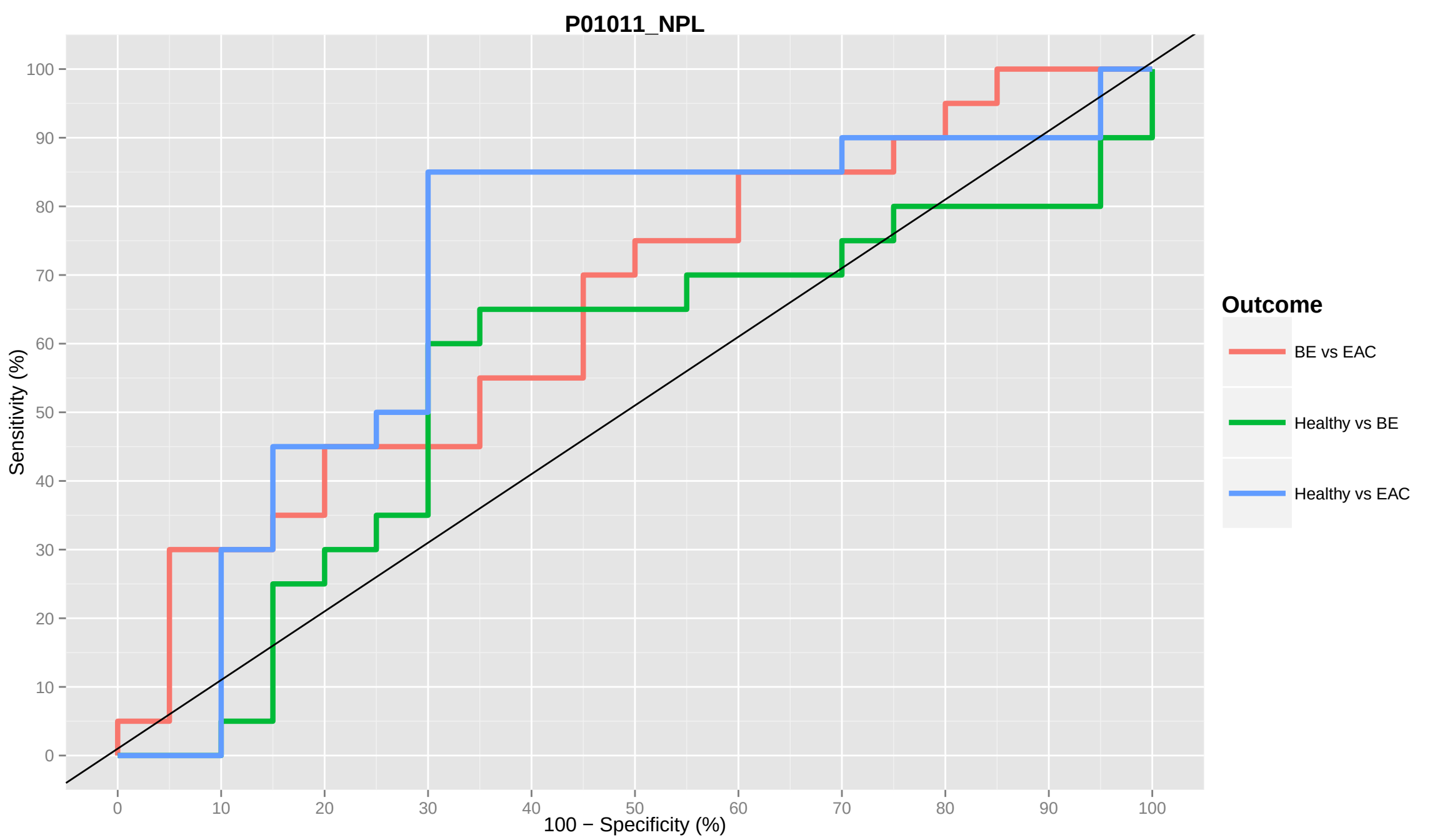

P05155_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

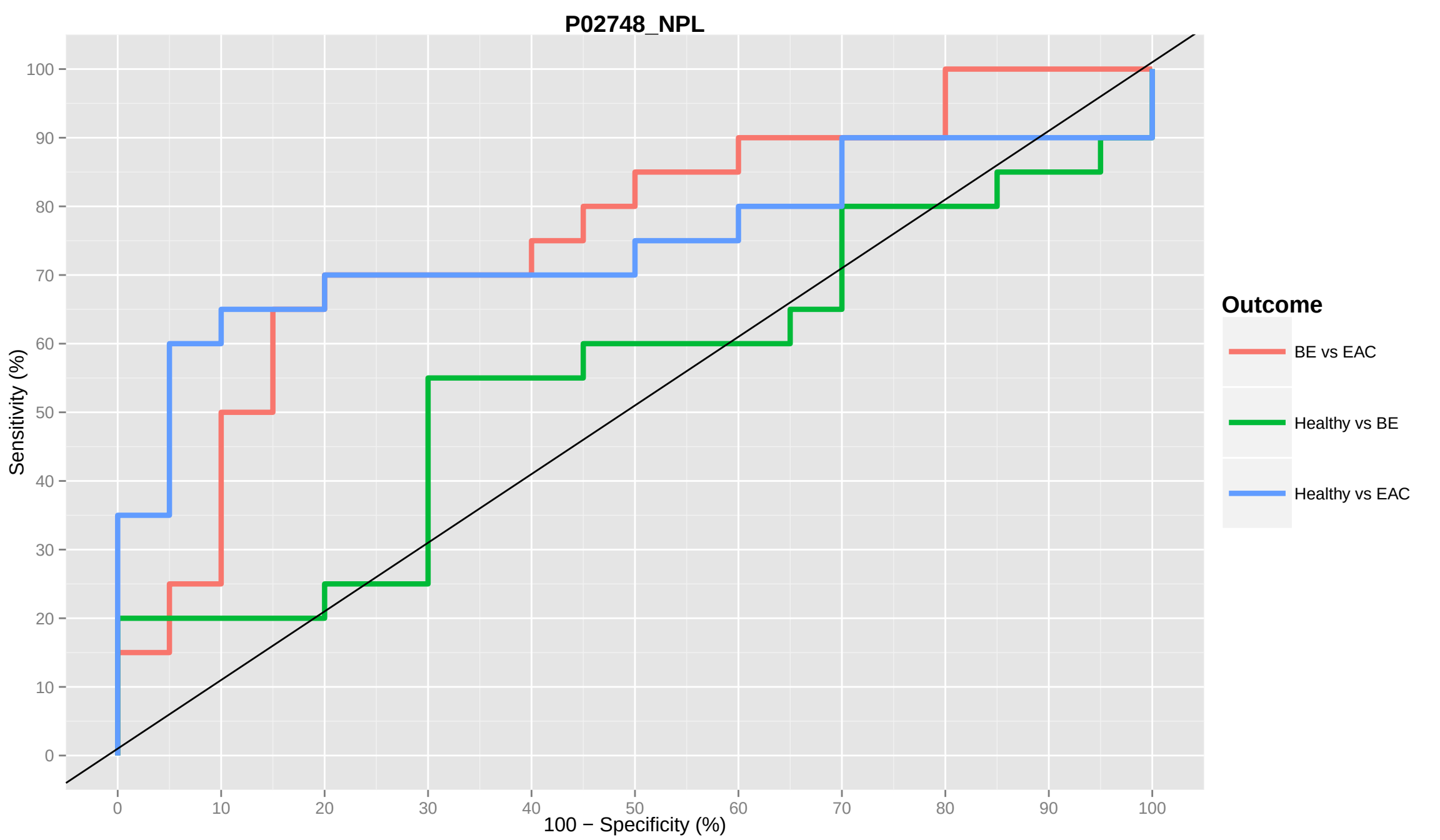

P05546_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

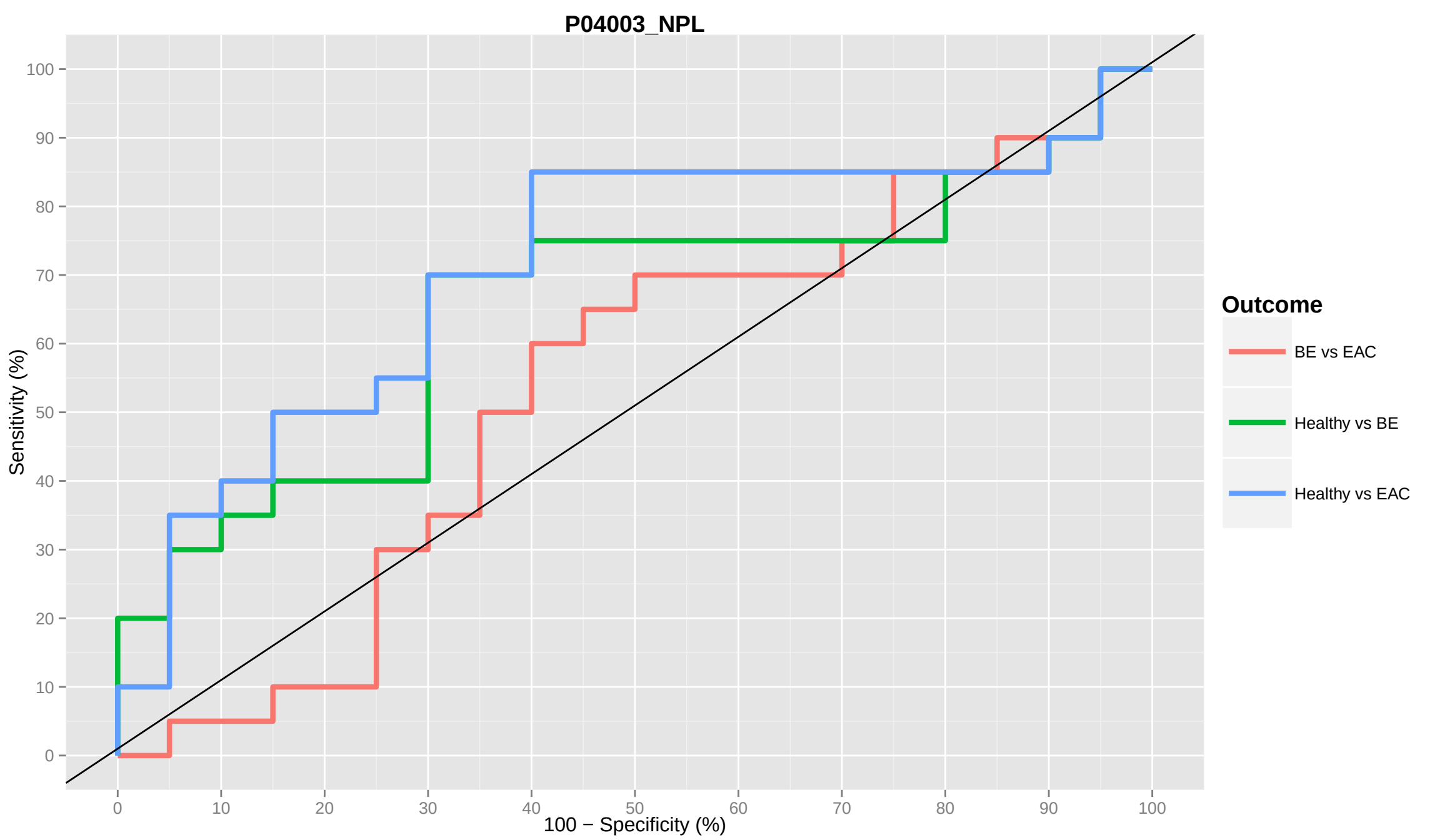

P06396_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

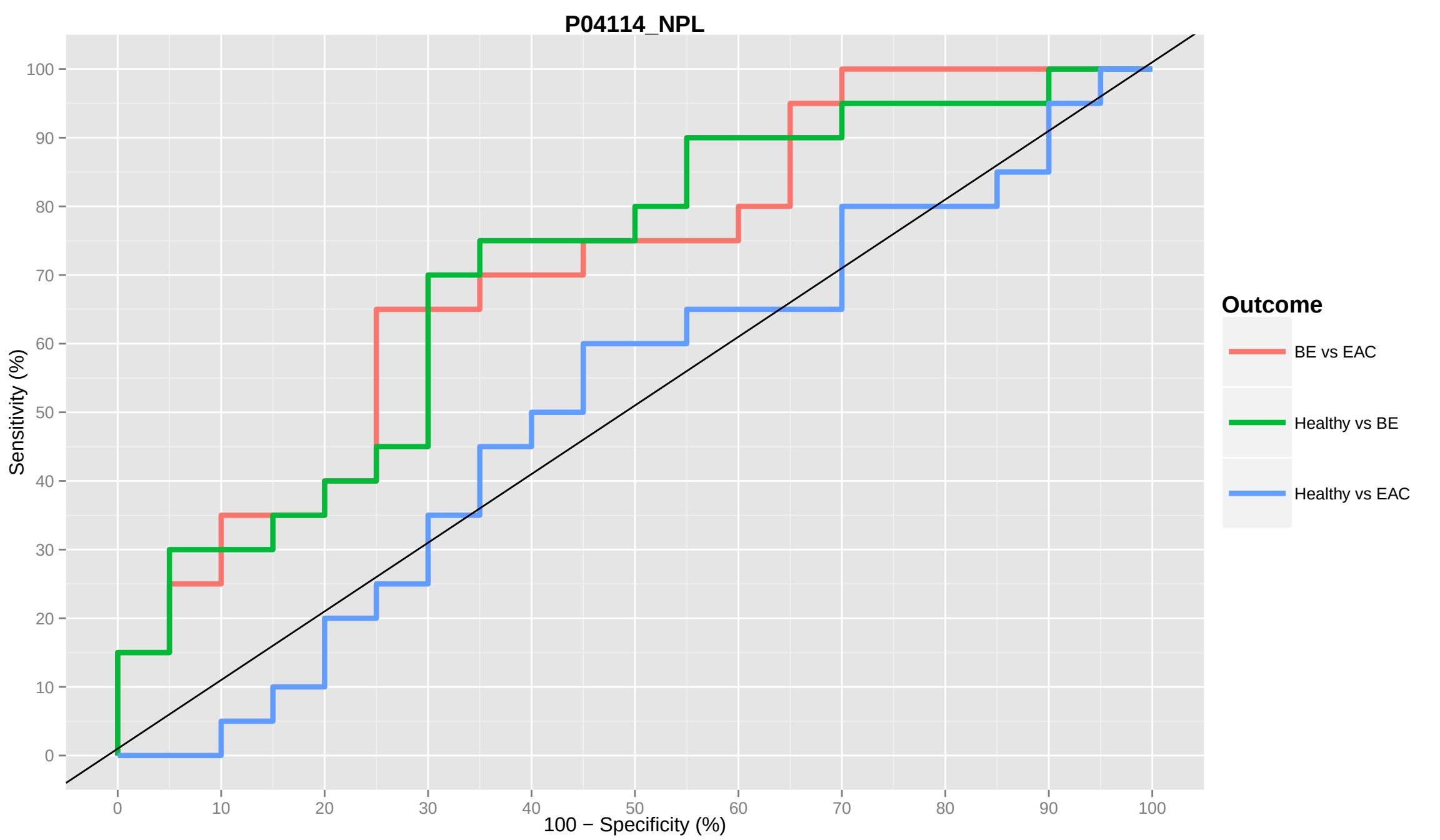

P10643_JAC
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

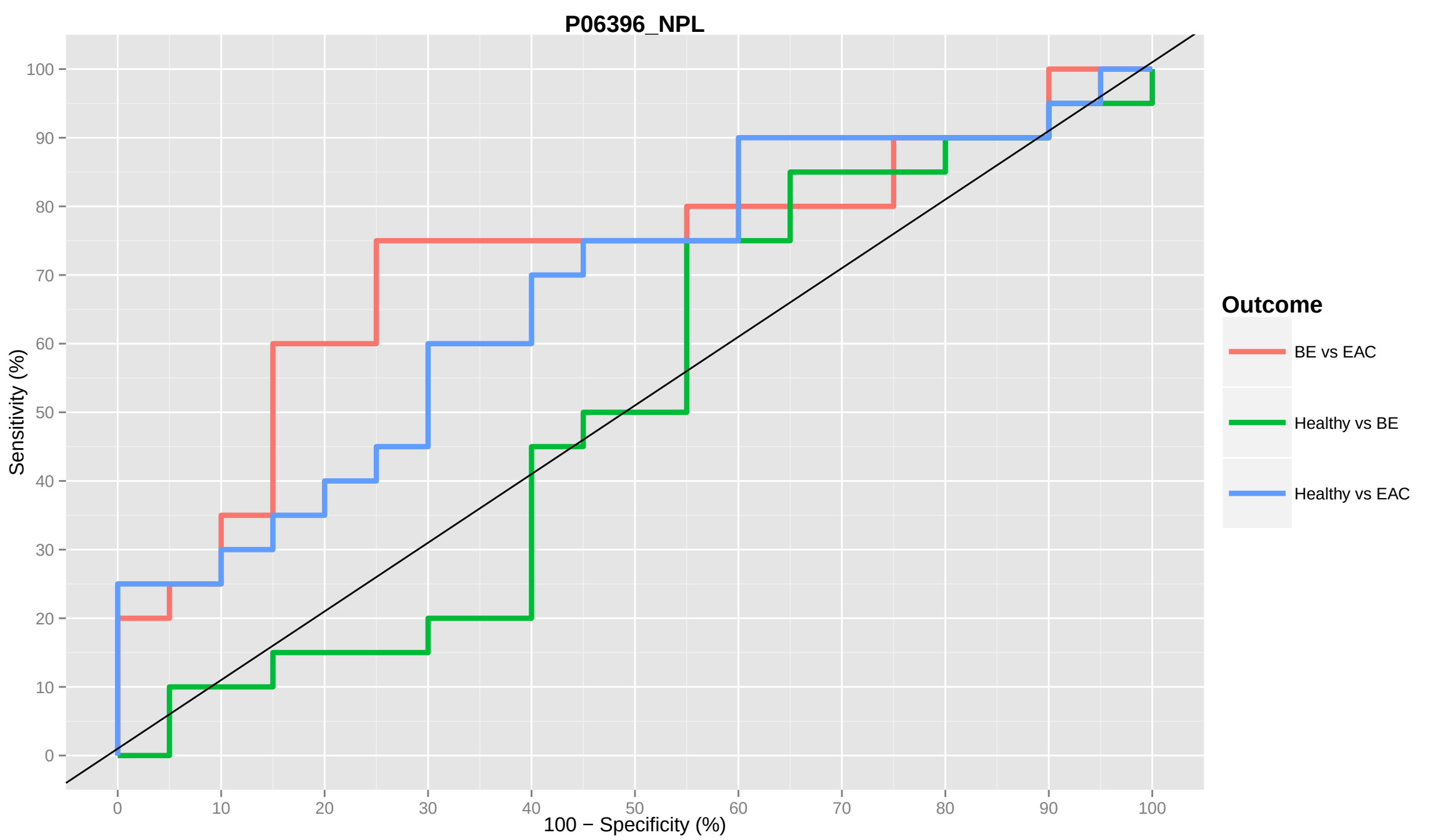

P00738_NPL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

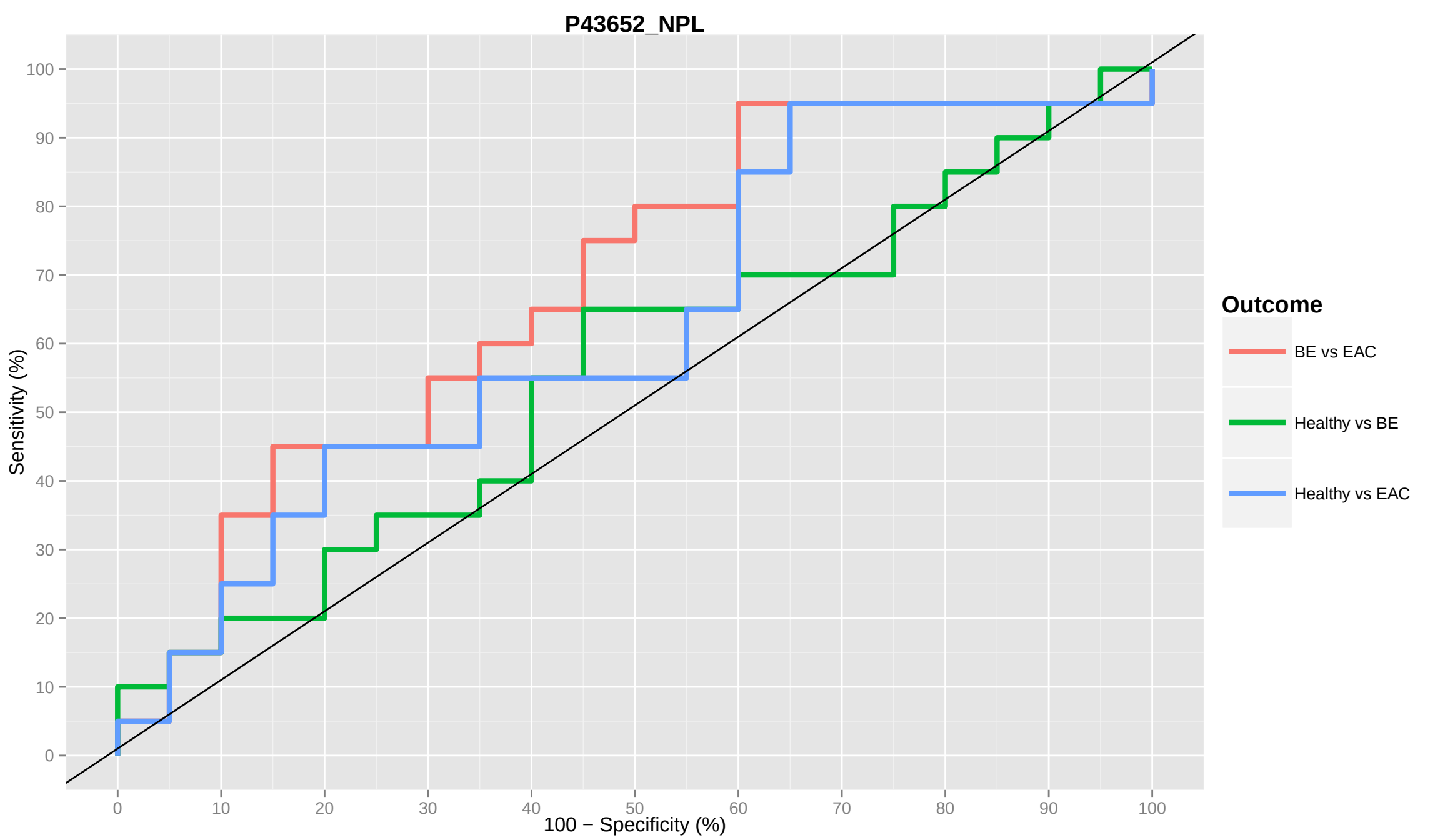

P01011_NPL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

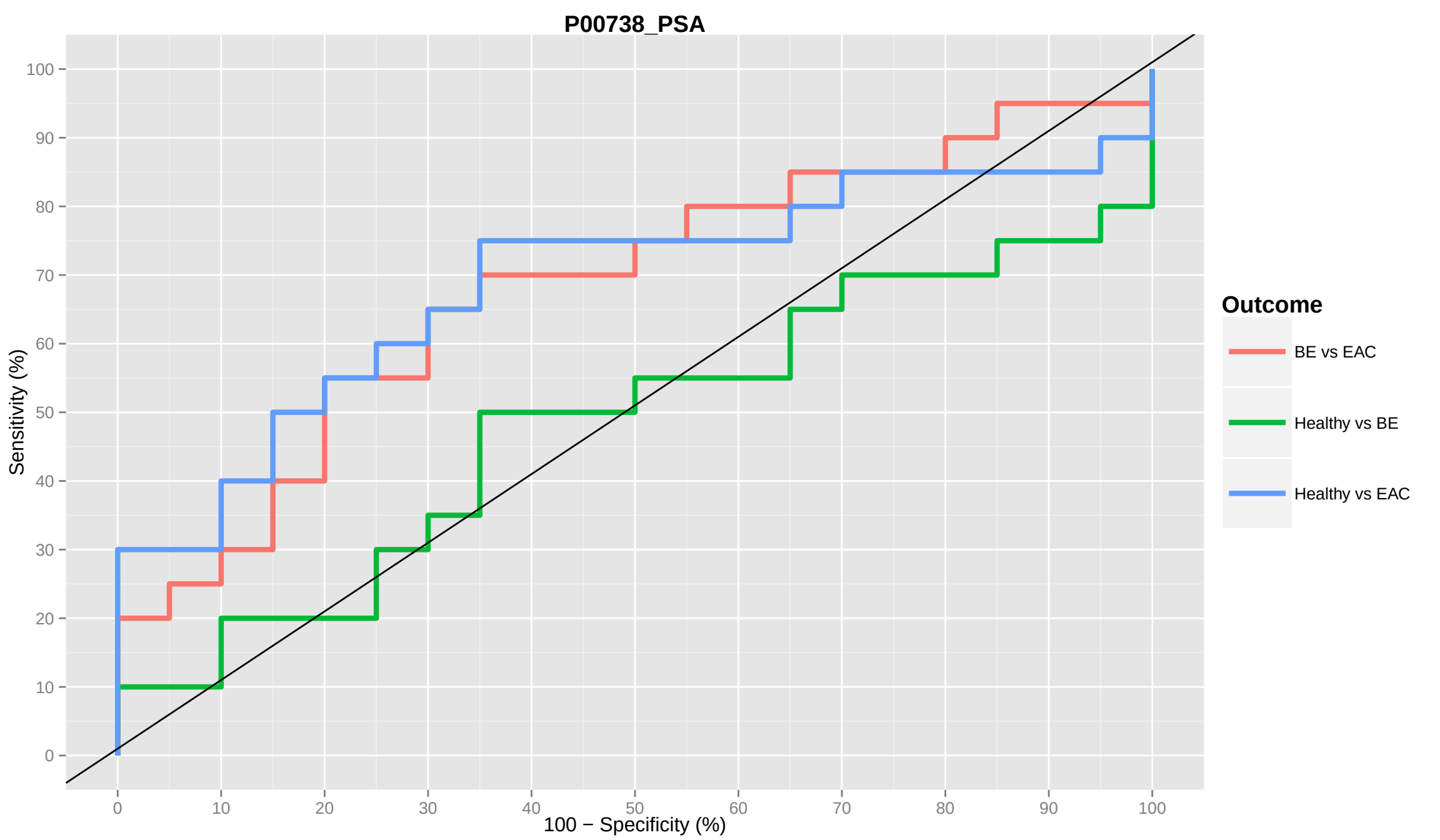

P02748_NPL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

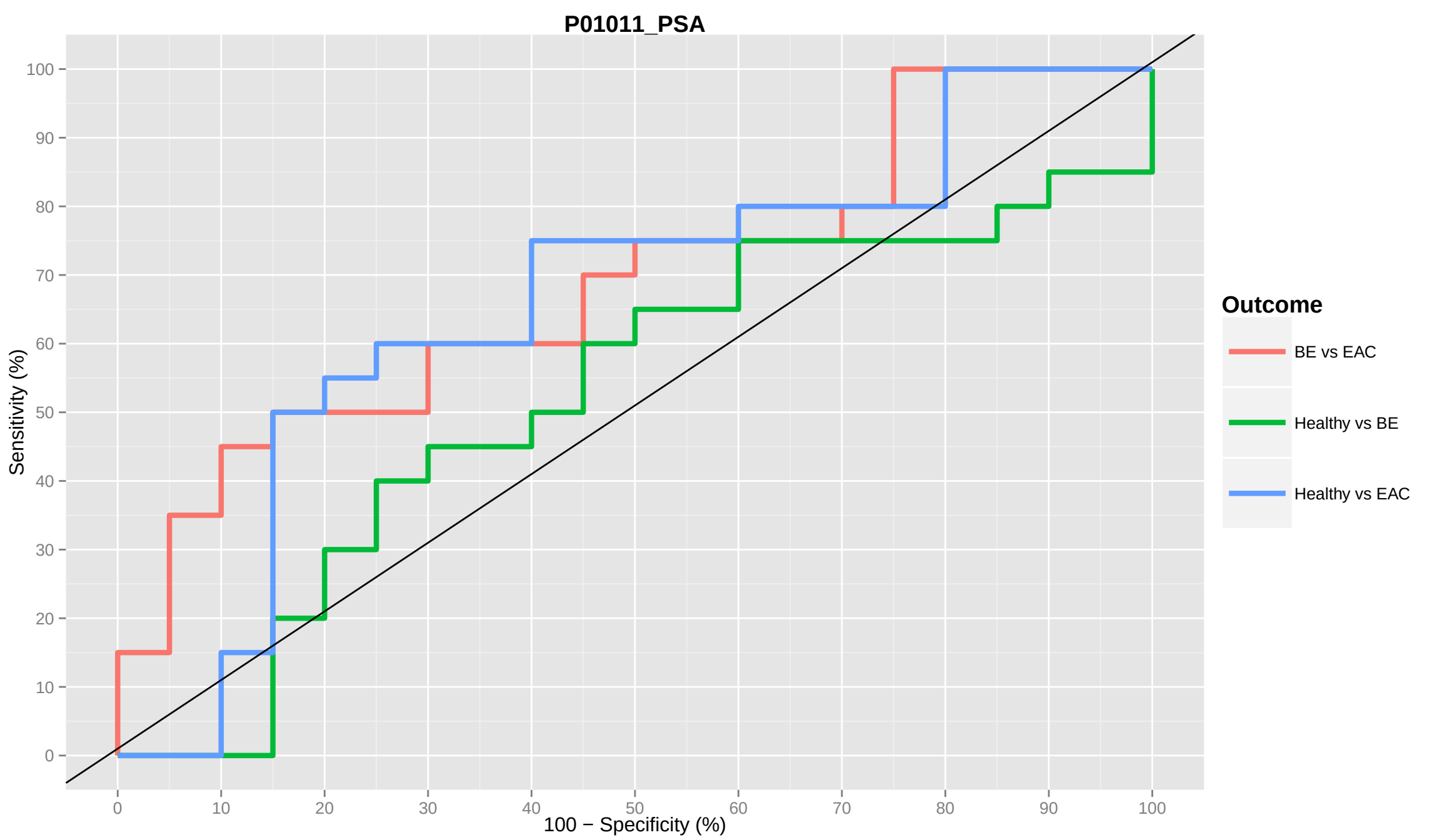

P04003_NPL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

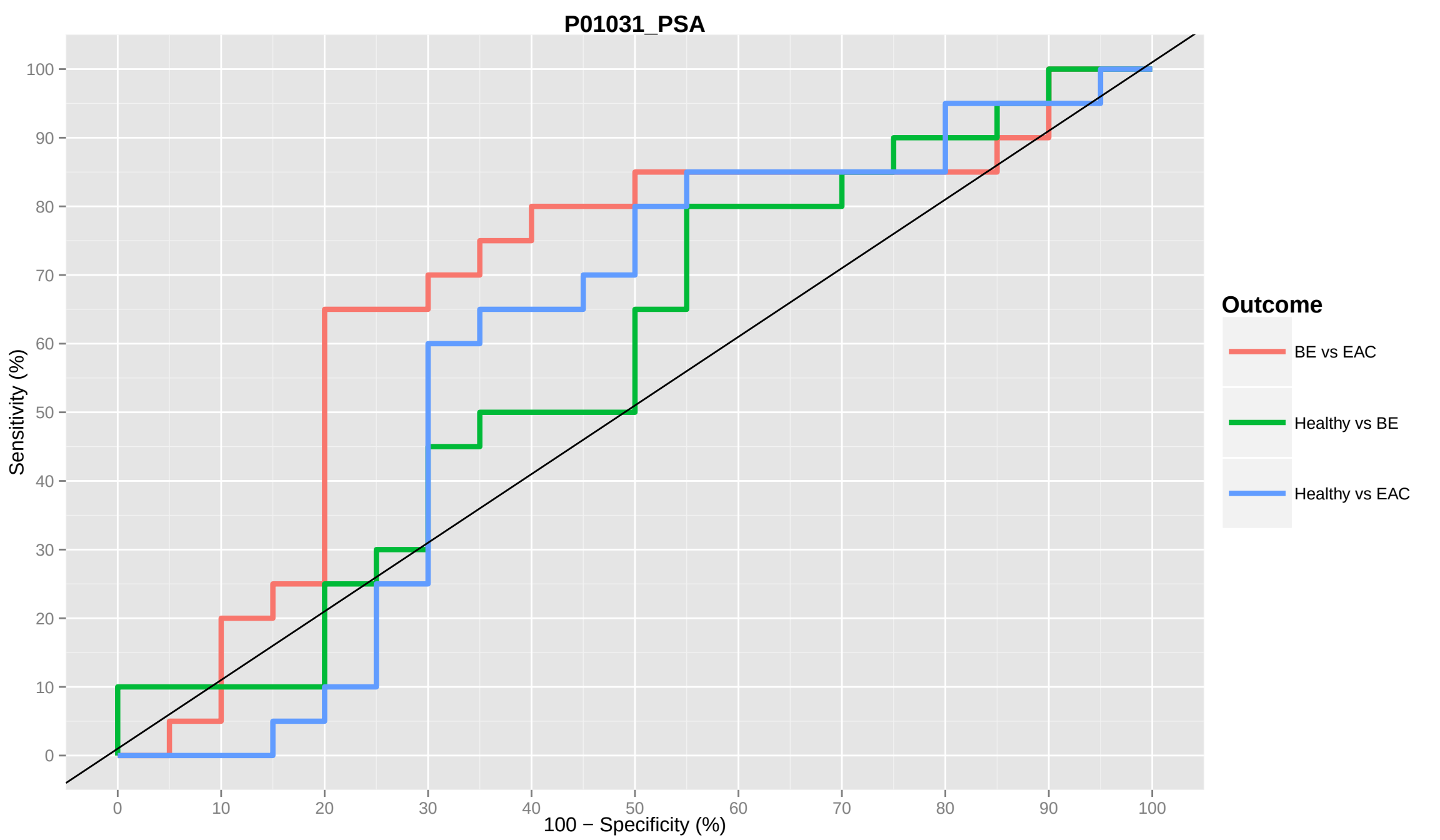

P04114_NPL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

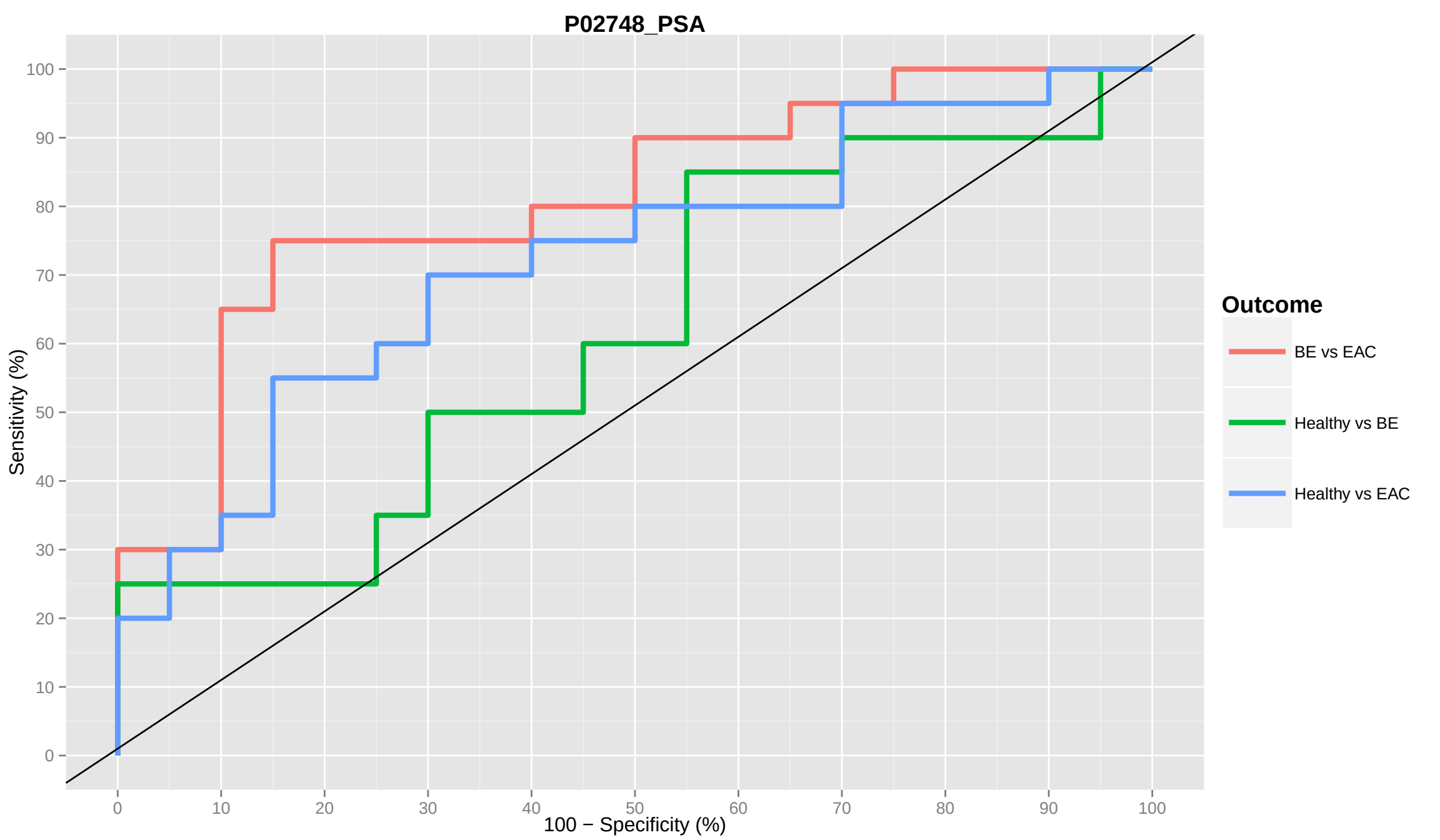

P06396_NPL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

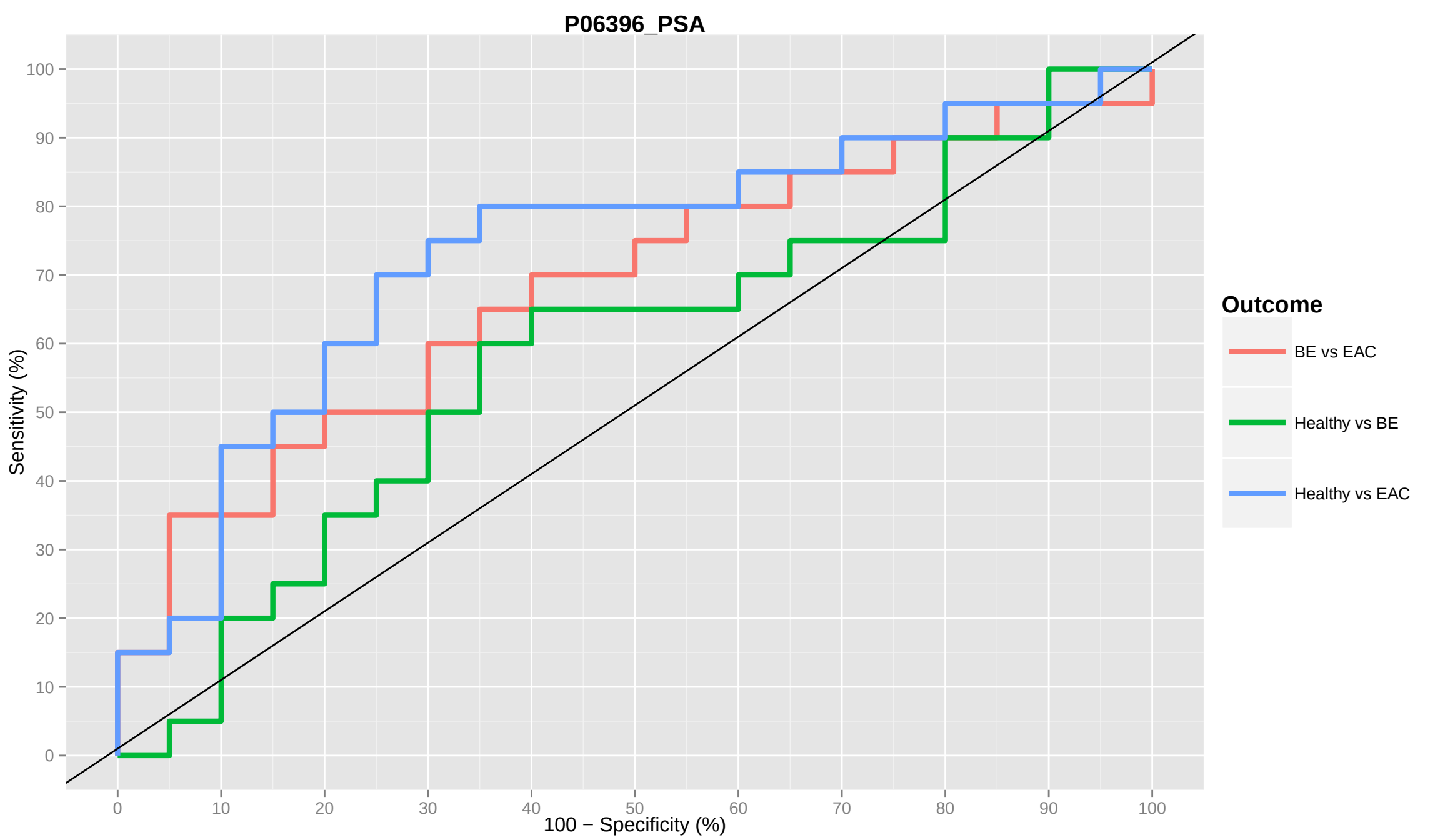

P43652_NPL
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

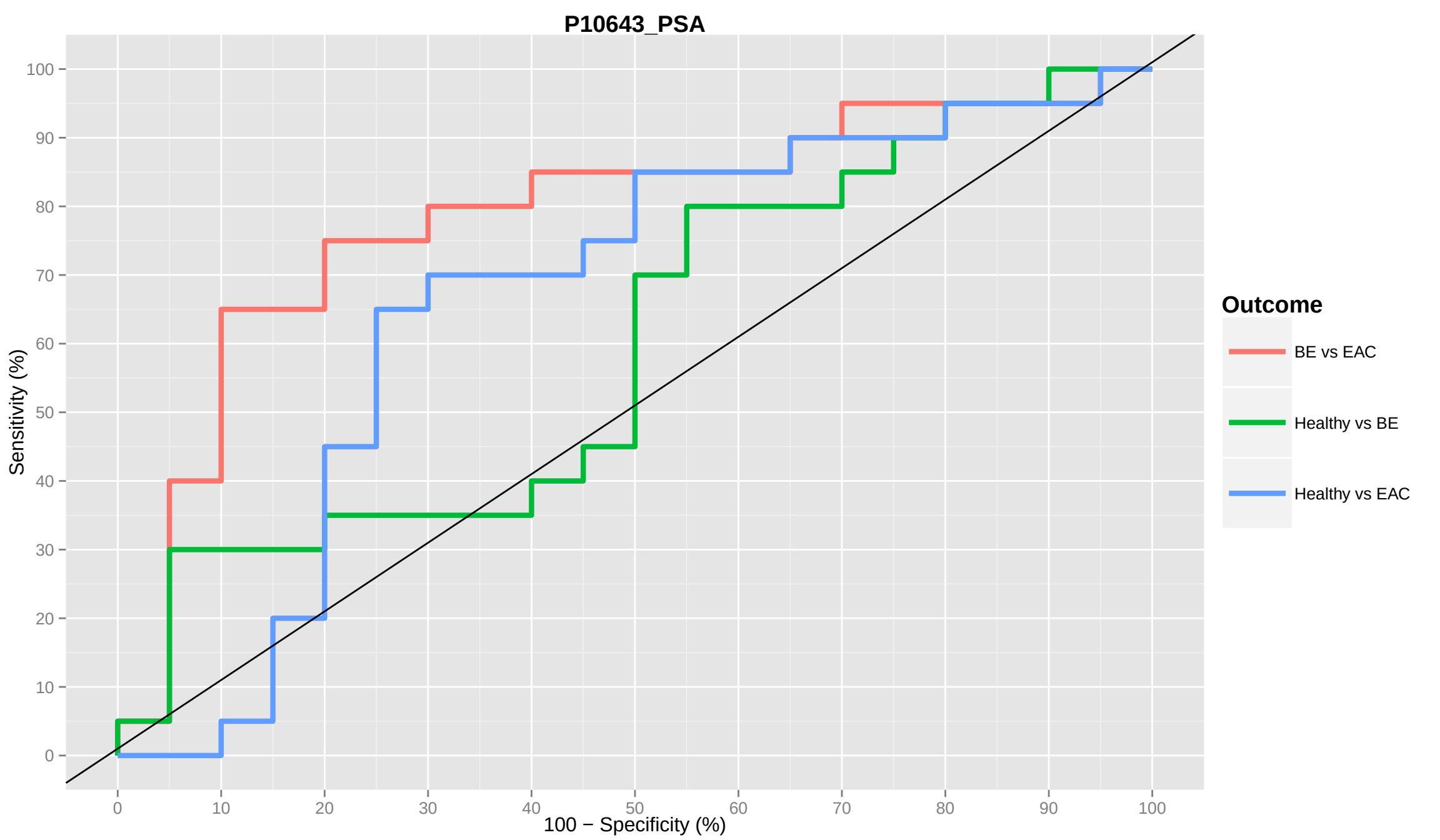

P00738_PSA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

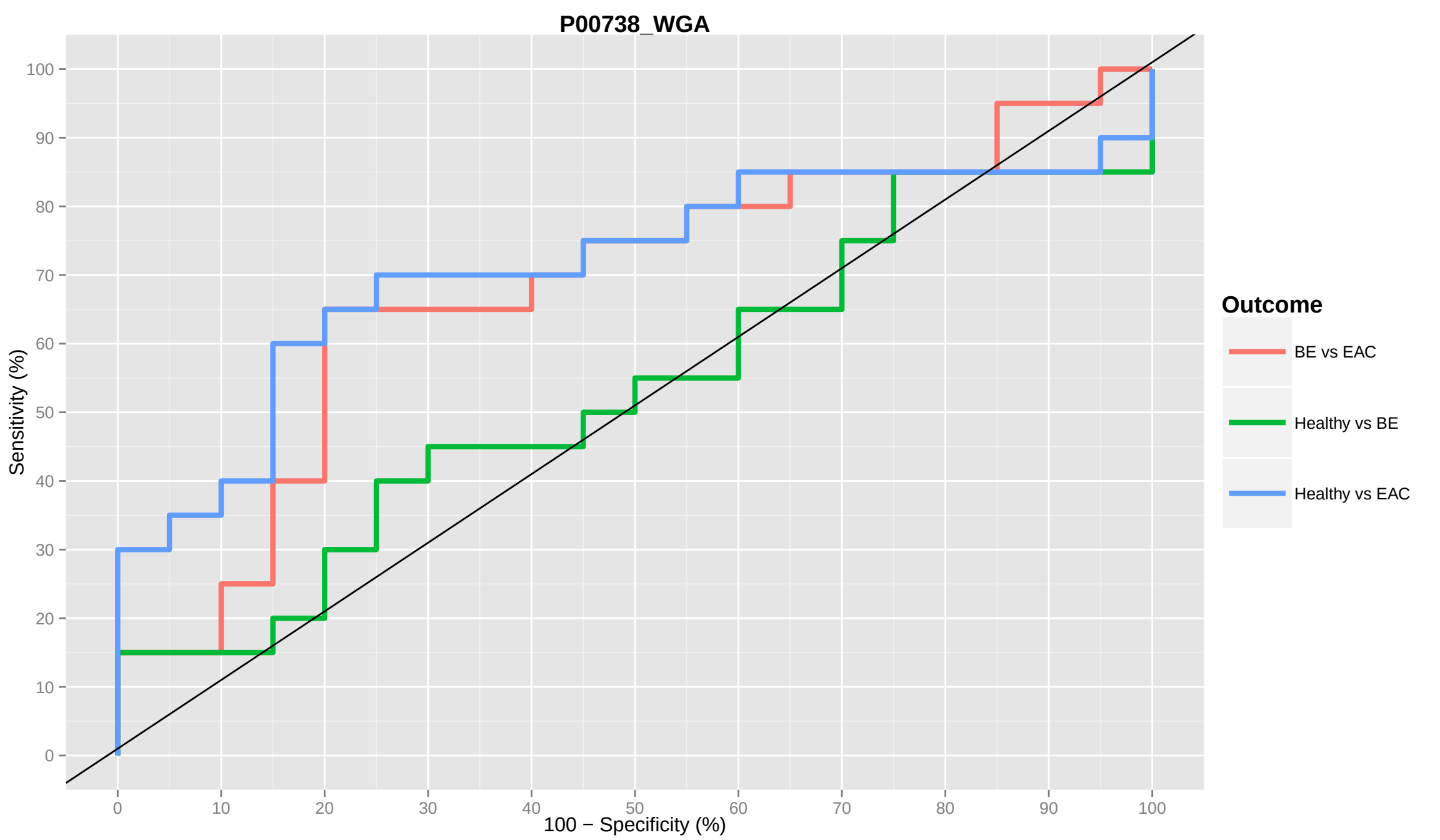

P01011_PSA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

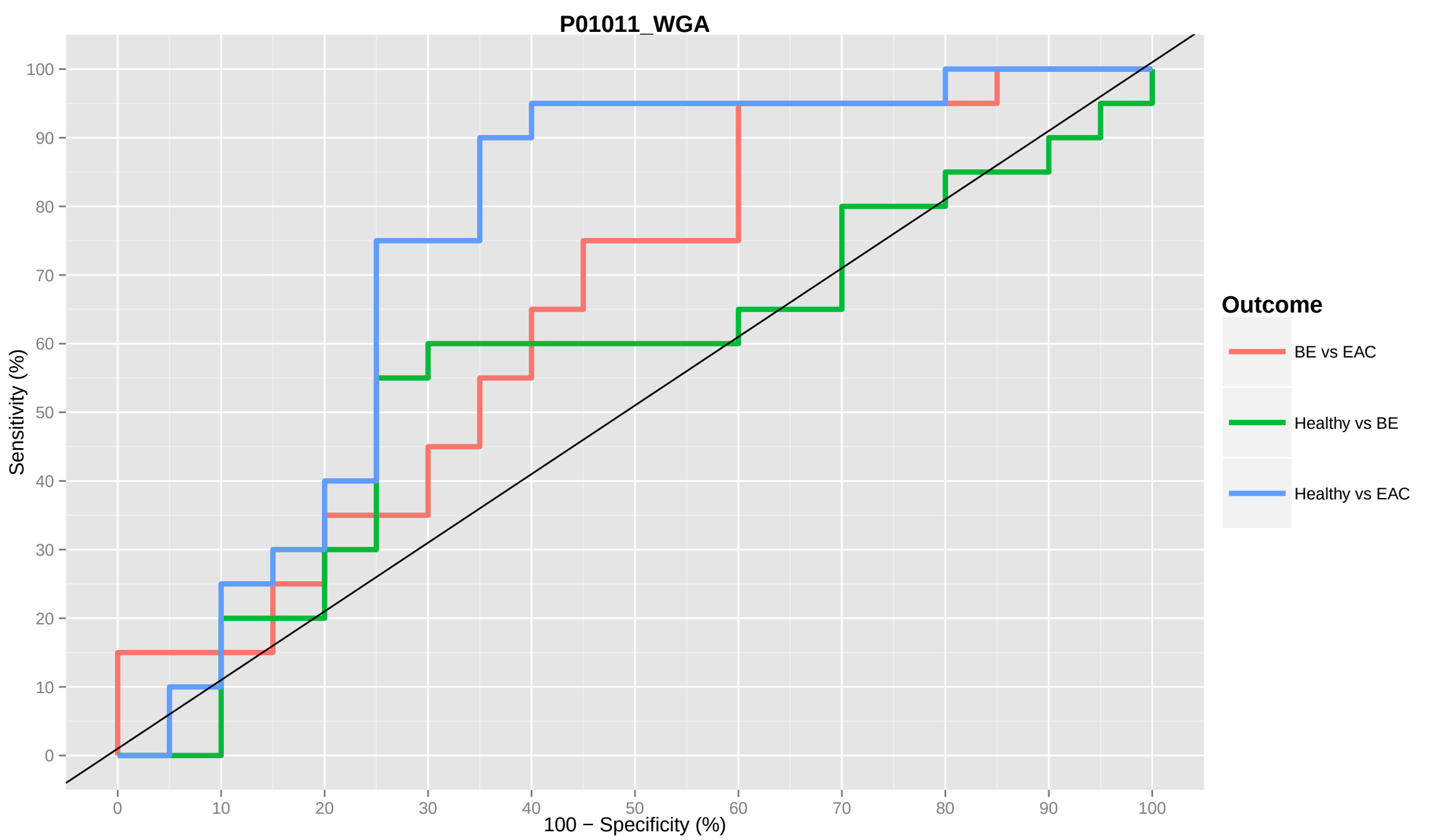

P01031_PSA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

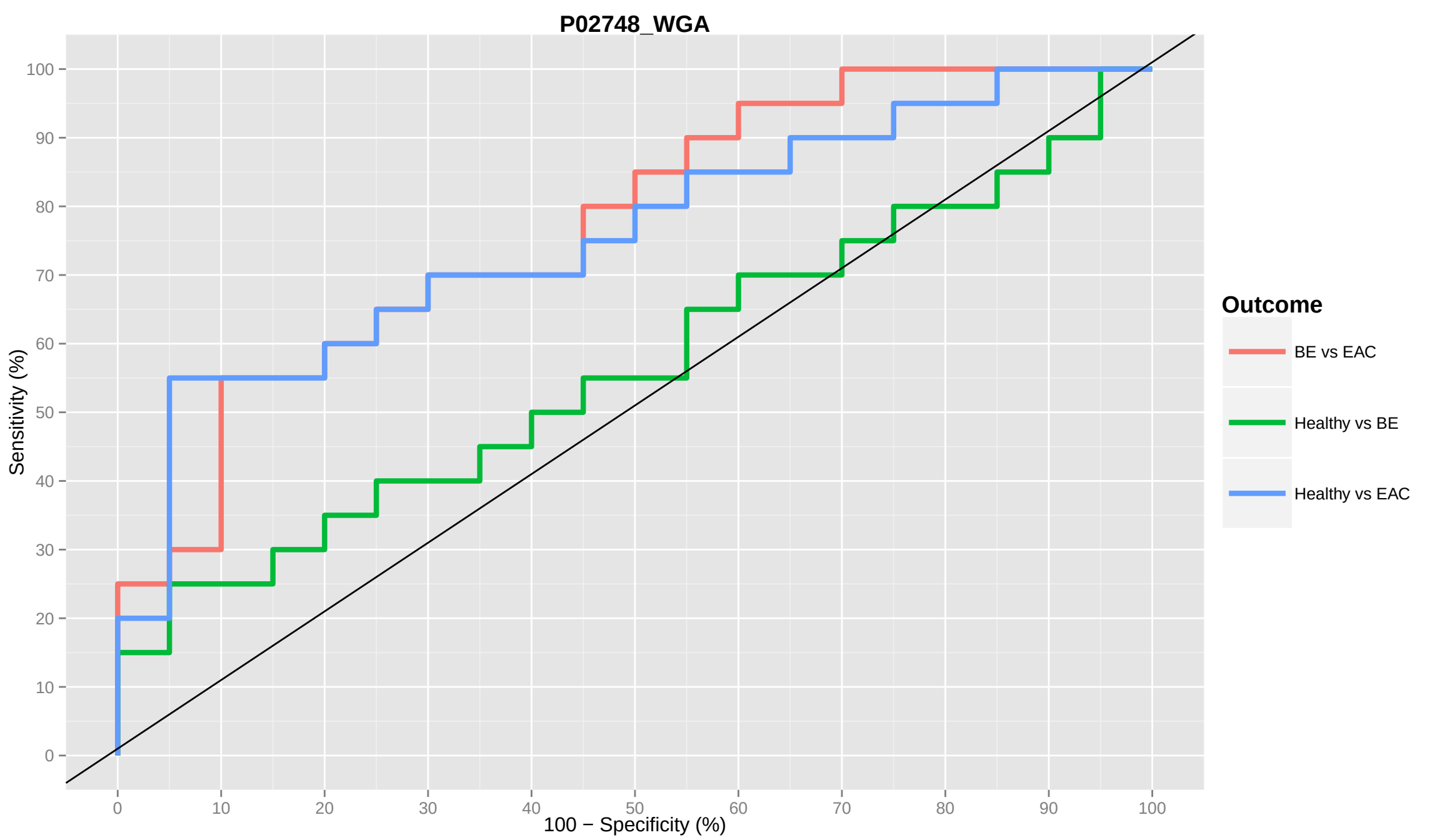

P02748_PSA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC

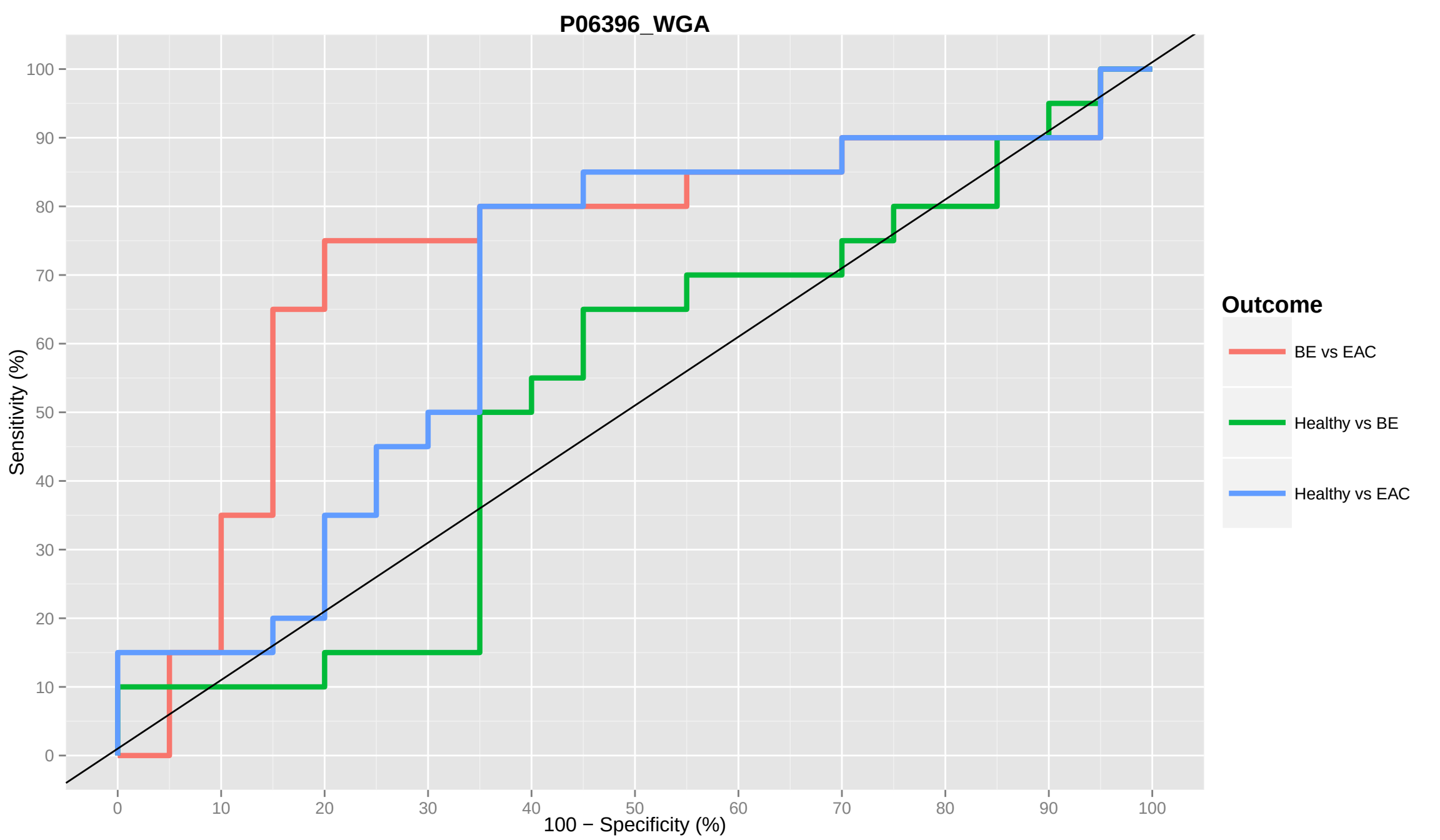

P06396_PSA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC


P10643_PSA
Sensitivity (%)
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100
100 − Specificity (%)
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC


P00738_WGA
Sensitivity (%)
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100
100 − Specificity (%)
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC


P01011_WGA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC


P02748_WGA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC


P06396_WGA
Sensitivity (%)
100 − Specificity (%)
0
10
20
30
40
50
60
70
80
90
100
Outcome
BE vs EAC
Healthy vs BE
Healthy vs EAC