

# Dissecting the genetic heterogeneity of gastric cancer



Timo Hess,<sup>a,b,de</sup> Carlo Maj,<sup>a,c,de</sup> Jan Gehlen,<sup>a</sup> Oleg Borisov,<sup>c</sup> Stephan L. Haas,<sup>d</sup> Ines Gockel,<sup>e</sup> Michael Vieth,<sup>f</sup> Guillaume Piessen,<sup>g</sup> Hakan Alakus,<sup>h</sup> Yogesh Vashist,<sup>ij</sup> Carina Pereira,<sup>k,l</sup> Michael Knapp,<sup>m</sup> Vitalia Schüller,<sup>a</sup> Alexander Quaas,<sup>n</sup> Heike I. Grabsch,<sup>o,p</sup> Jessica Trautmann,<sup>b</sup> Ewa Malecka-Wojcieszko,<sup>q</sup> Anna Mokrowiecka,<sup>q</sup> Jan Speller,<sup>m</sup> Andreas Mayr,<sup>m</sup> Julia Schröder,<sup>b</sup> Axel M. Hillmer,<sup>n</sup> Dominik Heider,<sup>r</sup> Florian Lordick,<sup>s</sup> Angeles Pérez-Aísa,<sup>t</sup> Rafael Campo,<sup>u</sup> Jesús Espinel,<sup>v</sup> Fernando Gejjo,<sup>w</sup> Concha Thomson,<sup>x</sup> Luis Bujanda,<sup>y,z</sup> Federico Sopeña,<sup>y,aa,ab</sup> Ángel Lanás,<sup>y,ab,ac</sup> María Pellisé,<sup>y,ad</sup> Claudia Paulig,<sup>ae,af</sup> Thorsten Oliver Goetze,<sup>ae,af</sup> Carolin Zelck,<sup>a</sup> Julian Reingruber,<sup>a</sup> Emadeldin Hassanin,<sup>c</sup> Peter Elbe,<sup>d</sup> Sandra Alsabeah,<sup>d</sup> Mats Lindblad,<sup>ag,ah</sup> Magnus Nilsson,<sup>ag,ah</sup> Nicole Kreuzer,<sup>e</sup> René Thieme,<sup>e</sup> Francesca Tavano,<sup>ai</sup> Roberta Pastorino,<sup>aj</sup> Dario Arzani,<sup>ak</sup> Roberto Persiani,<sup>al,am</sup> Jin-On Jung,<sup>h,an</sup> Henrik Nienhüser,<sup>an</sup> Katja Ott,<sup>ao</sup> Ralf R. Schumann,<sup>ap</sup> Oliver Kumpf,<sup>aq</sup> Susen Burock,<sup>ar</sup> Volker Arndt,<sup>as</sup> Anna Jakubowska,<sup>at,au</sup> Małgorzata Ławniczak,<sup>av</sup> Victor Moreno,<sup>aw,ax,ay</sup> Vicente Martín,<sup>az,ba,bb</sup> Manolis Kogevinas,<sup>az,bc,bd,be</sup> Marina Pollán,<sup>az,bf</sup> Justyna Dąbrowska,<sup>bg</sup> Antonio Salas,<sup>bh,bi,bj</sup> Olivier Cussenot,<sup>bk,bl</sup> Anne Boland-Auge,<sup>bm</sup> Delphine Daian,<sup>bm</sup> Jean-Francois Deleuze,<sup>bm</sup> Erika Salvi,<sup>bn</sup> Maris Teder-Laving,<sup>bo</sup> Gianluca Tomasello,<sup>bp,bq</sup> Margherita Ratti,<sup>bp</sup> Chiara Senti,<sup>bp,br</sup> Valli De Re,<sup>bs</sup> Agostino Steffan,<sup>bt</sup> Arnulf H. Hölscher,<sup>h</sup> Katharina Messerle,<sup>h</sup> Christiane Josephine Bruns,<sup>h</sup> Armands Siviņš,<sup>bu</sup> Inga Bogdanova,<sup>bu</sup> Jurgita Skieceviciene,<sup>bv</sup> Justina Arstikyte,<sup>bv</sup> Markus Moehler,<sup>bw</sup> Hauke Lang,<sup>bx</sup> Peter P. Grimminger,<sup>bx</sup> Martin Kruschewski,<sup>by</sup> Nikolaos Vassos,<sup>bz</sup> Claus Schildberg,<sup>ca</sup> Philipp Lingohr,<sup>cb</sup> Karsten Ridwelski,<sup>cc</sup> Hans Lippert,<sup>cd</sup> Nadine Fricker,<sup>b,ce</sup> Peter Krawitz,<sup>c</sup> Per Hoffmann,<sup>b,ce</sup> Markus M. Nöthen,<sup>b,ce</sup> Lothar Veits,<sup>f</sup> Jakob R. Izbicki,<sup>i</sup> Adrianna Mostowska,<sup>bg</sup> Federico Martínón-Torres,<sup>bj,cf,cg</sup> Daniele Cusi,<sup>ch,ci</sup> Rolf Adolfsson,<sup>cj</sup> Geraldine Cancel-Tassin,<sup>bk,bl</sup> Aksana Höblinger,<sup>ck</sup> Ernst Rodermann,<sup>d</sup> Monika Ludwig,<sup>cm</sup> Gisela Keller,<sup>cn</sup> Andres Metspalu,<sup>bo</sup> Hermann Brenner,<sup>as,co,cp</sup> Joerg Heller,<sup>cq</sup> Markus Neef,<sup>cr</sup> Michael Schepke,<sup>cr</sup> Franz Ludwig Dumoulin,<sup>cs</sup> Lutz Hamann,<sup>ap</sup> Renato Cannizzaro,<sup>ct</sup> Michele Ghidini,<sup>bp,bq</sup> Dominik Pfafsmann,<sup>cu</sup> Michael Geppert,<sup>cv</sup> Peter Malfertheiner,<sup>cw,cx</sup> Olivier Gehlen,<sup>cy</sup> Tomasz Skoczylas,<sup>cz</sup> Marek Majewski,<sup>cz</sup> Jan Lubiński,<sup>at</sup> Orazio Palmieri,<sup>ai</sup> Stefania Boccia,<sup>aj,ak</sup> Anna Latiano,<sup>ai</sup> Nuria Aragones,<sup>az,da</sup> Thomas Schmidt,<sup>h,an</sup> Mário Dinis-Ribeiro,<sup>l,db</sup> Rui Medeiros,<sup>k,dc</sup> Salah-Eddin Al-Batran,<sup>ae,af</sup> Mărcis Leja,<sup>bu,dd</sup> Juozas Kupcinskas,<sup>bv</sup> María A. García-González,<sup>y,aa,ac</sup> Marino Venerito,<sup>cw,df</sup> and Johannes Schumacher<sup>a,df,\*</sup>

<sup>a</sup>Institute of Human Genetics, University of Marburg, Marburg, Germany

<sup>b</sup>Institute of Human Genetics, University of Bonn, School of Medicine & University Hospital Bonn, Bonn, Germany

<sup>c</sup>Medical Faculty, Institute for Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany

<sup>d</sup>Department of Upper GI Diseases, Karolinska Institutet, Karolinska University Hospital and Unit of Gastroenterology and Rheumatology, Stockholm, Sweden

<sup>e</sup>Department of Visceral, Transplant, Thoracic and Vascular Surgery, University Hospital Leipzig, Leipzig, Germany

<sup>f</sup>Institute for Pathology, Friedrich-Alexander-University Erlangen-Nuernberg, Klinikum Bayreuth, Bayreuth, Germany

<sup>g</sup>Department of Digestive and Oncological Surgery, Claude Huriez Hospital, CHU Lille, Lille, France

<sup>h</sup>Department of General, Visceral, Cancer and Transplant Surgery, University of Cologne, Cologne, Germany

<sup>i</sup>Department of General, Visceral and Thoracic Surgery, University Medical Center Hamburg-Eppendorf, University of Hamburg, Hamburg, Germany

<sup>j</sup>Organ Transplant Center of Excellence, King Faisal Specialist Hospital and Research Center, Riyadh, Kingdom of Saudi Arabia

<sup>k</sup>Molecular Oncology and Viral Pathology Group, IPO Porto Research Center (CI-IPOP), Portuguese Institute of Oncology of Porto (IPO Porto), Porto 4200-072, Portugal

<sup>l</sup>Porto Comprehensive Cancer Center & RISE @ CI-IPO, University of Porto, Porto 4200-450, Portugal

<sup>m</sup>Medical Faculty, Institute of Medical Biometrics, Informatics and Epidemiology (IMBIE), University of Bonn, Bonn, Germany

<sup>n</sup>Medical Faculty, Institute of Pathology, University Hospital Cologne, University of Cologne, Germany

<sup>o</sup>Department of Pathology, GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, the Netherlands

<sup>p</sup>Pathology and Data Analytics, Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, United Kingdom

<sup>q</sup>Department of Digestive Tract Diseases, Medical University of Lodz, Lodz, Poland

<sup>r</sup>Department of Mathematics and Computer Science, University of Marburg, Marburg, Germany

<sup>s</sup>University Cancer Center Leipzig, Leipzig University Medical Center, Leipzig, Germany

<sup>t</sup>Department of Gastroenterology, Hospital Costa del Sol, Marbella, Spain

<sup>u</sup>Department of Gastroenterology, Hospital Parc Tauli, Sabadell, Spain

<sup>v</sup>Department of Gastroenterology, Complejo Hospitalario, León, Spain

<sup>w</sup>Department of Gastroenterology, Hospital Clínico Universitario, Salamanca, Spain

<sup>x</sup>Department of Gastroenterology, Hospital Obispo Polanco, Teruel, Spain

<sup>y</sup>CIBER de Enfermedades Hepáticas y Digestivas (CIBERehd), Spain

<sup>z</sup>Department of Gastroenterology, Hospital Donostia/Instituto Biodonostia, Universidad del País Vasco (UPV/EHU), San Sebastián, Spain

<sup>aa</sup>Instituto de Investigación Sanitaria Aragón (IIS Aragón), Zaragoza, Spain

\*Corresponding author. Center for Human Genetics, University Hospital of Marburg, Baldingerstraße, Marburg 35033, Germany.

E-mail address: johannes.schumacher@uni-marburg.de (J. Schumacher).

<sup>de</sup>Contributed equally.

<sup>df</sup>Jointly directed the study.

- <sup>ab</sup>Department of Gastroenterology, Hospital Clínico Universitario Lozano Blesa, Zaragoza, Spain
- <sup>ac</sup>Instituto Aragonés de Ciencias de la Salud (IACS), Zaragoza, Spain
- <sup>ad</sup>Department of Gastroenterology, Hospital Clinic of Barcelona, Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain
- <sup>ae</sup>Krankenhaus Nordwest, University Cancer Center, Frankfurt, Germany
- <sup>af</sup>Institut für Klinische Krebsforschung IKF GmbH am Krankenhaus Nordwest, Frankfurt, Germany
- <sup>ag</sup>Division of Surgery, Department of Upper GI Diseases, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden
- <sup>ah</sup>Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden
- <sup>ai</sup>Gastroenterology Unit, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy
- <sup>aj</sup>Department of Woman and Child Health and Public Health - Public Health Area, Fondazione Policlinico Universitario A. Gemelli IRCCS, Roma, Italy
- <sup>ak</sup>Section of Hygiene, University Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Roma, Italy
- <sup>al</sup>Dipartimento di Medicina e Chirurgia Traslazionale, Università Cattolica del Sacro Cuore, Roma, Italy
- <sup>am</sup>Dipartimento di Scienze Mediche e Chirurgiche, Fondazione Policlinico Universitario A. Gemelli IRCCS, Roma, Italy
- <sup>an</sup>Department of General, Visceral and Transplantation Surgery, University of Heidelberg, Heidelberg, Germany
- <sup>ao</sup>Department of Surgery, RoMed Klinikum Rosenheim, Rosenheim, Germany
- <sup>ap</sup>Institute of Microbiology, Infectious Diseases and Immunology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
- <sup>aq</sup>Department of Anaesthesiology and Operative Intensive Care Medicine, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, and Humboldt-Universität zu Berlin, Berlin, Germany
- <sup>ar</sup>Charité Comprehensive Cancer Center, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, and Humboldt-Universität zu Berlin, Berlin, Germany
- <sup>as</sup>Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
- <sup>at</sup>Department of Genetics and Pathology, Pomeranian Medical University in Szczecin, Poland
- <sup>au</sup>Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University in Szczecin, Poland
- <sup>av</sup>Department of Gastroenterology, Pomeranian Medical University in Szczecin, Poland
- <sup>aw</sup>Unit of Biomarkers and Susceptibility, Oncology Data Analytics Program, Catalan Institute of Oncology (ICO), Hospital Duran I Reynals, Barcelona, Spain
- <sup>ax</sup>Colorectal Cancer Group, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain
- <sup>ay</sup>Faculty of Medicine, Department of Clinical Sciences, University of Barcelona, Barcelona, Spain
- <sup>az</sup>Consortium for Biomedical Research in Epidemiology & Public Health (CIBER en Epidemiología y Salud Pública e CIBERESP), Spain
- <sup>ba</sup>The Research Group in Gene - Environment and Health Interactions (GIIGAS)/Institute of Biomedicine (IBIOMED), Universidad de Leon, Leon, Spain
- <sup>bb</sup>Faculty of Health Sciences, Department of Biomedical Sciences, Area of Preventive Medicine and Public Health, Universidad de Leon, Leon, Spain
- <sup>bc</sup>ISGlobal, Barcelona, Spain
- <sup>bd</sup>Universitat Pompeu Fabra (UPF), Campus Del Mar, Barcelona, Spain
- <sup>be</sup>IMIM (Hospital Del Mar Medical Research Institute), Barcelona, Spain
- <sup>bf</sup>Cancer and Environmental Epidemiology Unit, Department of Epidemiology of Chronic Diseases, National Center for Epidemiology, Carlos III Institute of Health, Madrid, Spain
- <sup>bg</sup>Department of Biochemistry and Molecular Biology, Poznan University of Medical Sciences, Poland
- <sup>bh</sup>Unidade de Xenética, Instituto de Ciencias Forenses (INCIFOR), Facultade de Medicina, Universidade de Santiago de Compostela, Santiago de Compostela, Spain
- <sup>bi</sup>GenPoB Research Group, Instituto de Investigación Sanitaria (IDIS), Hospital Clínico Universitario de Santiago (SERGAS), Santiago de Compostela, Spain
- <sup>bj</sup>Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBER-ES), Madrid, Spain
- <sup>bk</sup>CeRePP, Paris, France
- <sup>bl</sup>GRC n°5 Predictive Onco-Urology, Tenon Hospital, Sorbonne University, Paris, France
- <sup>bm</sup>Centre National de Recherche en Génomique Humaine, CEA, University Paris-Saclay, Evry, France
- <sup>bn</sup>Neuroalgology Unit Fondazione IRCCS, Istituto Neurologico 'Carlo Besta' Milan, Milan, Italy
- <sup>bo</sup>Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
- <sup>bp</sup>Medical Oncology Unit, ASST of Cremona, Cremona, Italy
- <sup>bq</sup>Medical Oncology Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy
- <sup>br</sup>Department of Medical Oncology, Institut Jules Bordet - Université Libre de Bruxelles (ULB), Brussels, Belgium
- <sup>bs</sup>Unit of Immunopathologia e Biomarcatori Oncologici/Bio-proteomics Facility, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, Aviano, Italy
- <sup>bt</sup>Unit of Immunopathologia e Biomarcatori Oncologici, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, Aviano, Italy
- <sup>bu</sup>Institute of Clinical and Preventive Medicine, University of Latvia, Riga East University Hospital, Riga, Latvia
- <sup>bv</sup>Gastroenterology Department and Institute for Digestive Research, Lithuanian University of Health Sciences, Kaunas, Lithuania

- <sup>bw</sup>Department of Medicine I, University Medical Center of Johannes Gutenberg University Mainz, Mainz, Germany
- <sup>bx</sup>Department of General, Visceral and Transplant Surgery, University Medical Center, University of Mainz, Mainz, Germany
- <sup>by</sup>Department of General and Visceral Surgery, Klinikum Frankfurt (Oder), Germany
- <sup>bz</sup>Division of Surgical Oncology and Thoracic Surgery, Department of Surgery, University Medical Center Mannheim, University of Heidelberg, Mannheim, Germany
- <sup>ca</sup>Department of General Surgery, Brandenburg Medical School Theodor Fontane, University Hospital Brandenburg, Brandenburg, Germany
- <sup>cb</sup>Department of General, Visceral, Thoracic and Vascular Surgery, University Hospital Bonn, Bonn, Germany
- <sup>cc</sup>Department of General and Visceral Surgery, Klinikum Magdeburg GmbH, Magdeburg, Germany
- <sup>cd</sup>Institute of Quality Assurance in Operative Medicine, Otto-von-Guericke University Hospital, Magdeburg, Germany
- <sup>ce</sup>Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany
- <sup>cf</sup>Department of Pediatrics, Translational Pediatrics and Infectious Diseases Section, Hospital Clínico Universitario de Santiago (SERGAS), Santiago de Compostela, Spain
- <sup>cg</sup>Genetics, Vaccines, Infectious Diseases and Pediatrics Research Group GENVIP, Instituto de Investigación Sanitaria de Santiago (IDIS), Universidade de Santiago de Compostela, Santiago de Compostela, Spain
- <sup>ci</sup>Institute of Biomedical Technologies, National Research Council of Italy, Milan, Italy
- <sup>cj</sup>Bio4Dreams-Business, Nursery for Life Sciences, Milan, Italy
- <sup>ck</sup>Department of Clinical Sciences, Umeå University, Umeå, Sweden
- <sup>cl</sup>Department of Internal Medicine I, Community Hospital Mittelrhein, Koblenz, Germany
- <sup>cm</sup>Association of Medical Practices in Hematology and Internal Oncology, Troisdorf, Germany
- <sup>cn</sup>Association for Oncological Studies (Gefos), Dortmund, Germany
- <sup>co</sup>Institute of Pathology, Technical University of Munich, Munich, Germany
- <sup>cp</sup>Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
- <sup>cq</sup>German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
- <sup>cr</sup>Department of Gastroenterology, Marienhaus Hospital Ahrweiler, Ahrweiler, Germany
- <sup>cs</sup>Department of Gastroenterology, Helios Hospital Siegburg, Siegburg, Germany
- <sup>ct</sup>Department of Gastroenterology, St. Elisabeth Hospital Bonn, Bonn, Germany
- <sup>cu</sup>Unit of Oncological Gastroenterology, Centro di Riferimento Oncologico di Aviano (CRO), IRCCS, Aviano, Italy
- <sup>cv</sup>Gastroenterological Centre, Dr. Plafmann, Bonn, Germany
- <sup>cw</sup>Gastroenterological Centre, Dr. Geppert, Bayreuth, Germany
- <sup>cx</sup>Department of Gastroenterology, Hepatology and Infectious Diseases, Otto-von-Guericke University Hospital, Magdeburg, Germany
- <sup>cy</sup>Department of Internal Medicine II, Hospital of the Ludwig Maximilians University of Munich, Munich, Germany
- <sup>cz</sup>Department of Surgical Oncology, Centre Hospitalier Lyon-sud, Lyon, France
- <sup>d1</sup>2nd Department of General Surgery, Medical University of Lublin, Lublin, Poland
- <sup>da</sup>Epidemiology Section, Public Health Division, Department of Health of Madrid, Madrid, Spain
- <sup>db</sup>Gastroenterology Department, Portuguese Institute of Oncology of Porto, Porto 4200-072, Portugal
- <sup>dc</sup>Research Department of the Portuguese League Against Cancer-North (LPCC-NRNorte), Porto 4200-177, Portugal
- <sup>dd</sup>Digestive Diseases Centre GASTRO, Riga, Latvia

## Summary

**Background** Gastric cancer (GC) is clinically heterogenous according to location (cardia/non-cardia) and histopathology (diffuse/intestinal). We aimed to characterize the genetic risk architecture of GC according to its subtypes. Another aim was to examine whether cardia GC and oesophageal adenocarcinoma (OAC) and its precursor lesion Barrett's oesophagus (BO), which are all located at the gastro-oesophageal junction (GOJ), share polygenic risk architecture.

**Methods** We did a meta-analysis of ten European genome-wide association studies (GWAS) of GC and its subtypes. All patients had a histopathologically confirmed diagnosis of gastric adenocarcinoma. For the identification of risk genes among GWAS loci we did a transcriptome-wide association study (TWAS) and expression quantitative trait locus (eQTL) study from gastric corpus and antrum mucosa. To test whether cardia GC and OAC/BO share genetic aetiology we also used a European GWAS sample with OAC/BO.

**Findings** Our GWAS consisting of 5816 patients and 10,999 controls highlights the genetic heterogeneity of GC according to its subtypes. We newly identified two and replicated five GC risk loci, all of them with subtype-specific association. The gastric transcriptome data consisting of 361 corpus and 342 antrum mucosa samples revealed that an upregulated expression of *MUC1*, *ANKRD50*, *PTGER4*, and *PSCA* are plausible GC-pathomechanisms at four GWAS loci. At another risk locus, we found that the blood-group 0 exerts protective effects for non-cardia and diffuse GC,

eBioMedicine  
2023;92: 104616

Published Online xxx  
<https://doi.org/10.1016/j.ebiom.2023.104616>

while blood-group A increases risk for both GC subtypes. Furthermore, our GWAS on cardia GC and OAC/BO (10,279 patients, 16,527 controls) showed that both cancer entities share genetic aetiology at the polygenic level and identified two new risk loci on the single-marker level.

**Interpretation** Our findings show that the pathophysiology of GC is genetically heterogeneous according to location and histopathology. Moreover, our findings point to common molecular mechanisms underlying cardia GC and OAC/BO.

**Funding** German Research Foundation (DFG).

**Copyright** © 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Gastric cancer; Oesophageal adenocarcinoma; Genome-wide association study (GWAS); Transcriptome-wide association study (TWAS)

### Research in context

#### Evidence before this study

We searched PubMed on June 30, 2022, to identify genetic risk variants and corresponding risk genes for gastric cancer that have been identified through genome-wide association studies. We did not apply any publication date restrictions. The search was restricted to papers published in the English language. Search terms were: (“gastric cancer” OR “gastric carcinoma” OR “gastric adenocarcinoma”) AND (“genome wide association study” OR “GWAS”). Nine genome-wide association studies for gastric cancer have been published to date in the East-Asian population and one in the European population. In none of these studies, patients’ tumours were characterised with regard to both subtypes, namely location (cardia/non-cardia) and histopathology (diffuse/intestinal). In addition, only one risk locus could be functionally characterized in these studies. The expression of the gene *PSCA* is regulated by an identified gastric cancer risk variant. Furthermore, it has not been examined so far whether gastric cancer at the cardia and oesophageal adenocarcinoma, which are both localised at the gastro-oesophageal junction, share genetic aetiology.

#### Added value of this study

Within a European consortium, we did a meta-analysis of ten datasets available to date from genome-wide association

studies, including more than 16,000 individuals. We newly identified two and replicated five risk loci, all of them contribute subtype-specific risk to gastric cancer. For the identification of risk genes among the identified loci we used transcriptome-wide expression data from gastric corpus and antrum mucosa. This revealed that an upregulated expression of the genes *MUC1*, *ANKRD50*, *PTGER4*, and *PSCA* are plausible pathomechanisms at four risk loci. At another risk locus, we found that the blood-group O exerts protective effects for non-cardia and diffuse gastric cancer, while blood-group A increases risk for both subtypes. Furthermore, our data on cardia gastric cancer as well as oesophageal adenocarcinoma (26,000 individuals) showed that both cancer entities share genetic aetiology at the polygenic level and identified two new risk loci on the single-marker level.

#### Implications of all the available evidence

Our results highlight the genetic heterogeneity of gastric cancer and provide insights into distinct and shared aetiological processes underlying different gastric cancer subtypes. This should further help to elucidate the pathomechanisms underlying each gastric cancer subtype. The findings may also serve as basis for a more biologically driven stratification of gastric cancer in the clinical setting.

### Introduction

Gastric adenocarcinoma, here called gastric cancer (GC), has a multifactorial aetiology and is clinically heterogeneous. Anatomically, GC is subdivided into a cardia and non-cardia type.<sup>1</sup> Cardia GC is located in the proximal stomach at the gastro-oesophageal junction (GOJ), while non-cardia GC resides in the distal stomach. Various characteristics point to pathomechanisms that are specific for each location-type. Cardia GC is related to gastro-oesophageal reflux (GOR) and its prevalence is rising.<sup>1</sup> In contrast, the prevalence of non-cardia GC has decreased due to a decline in *Helicobacter*

*pylori* infections, the main risk factor.<sup>1</sup> Histopathologically, GC is subdivided into Lauren’s diffuse- and intestinal-type.<sup>1</sup> Different pathophysiological processes are also specific for each of these subtypes. While intestinal GC develops from a metaplasia-neoplasia sequence, diffuse GC shows no preneoplastic condition.<sup>1</sup>

In multifactorial diseases genome-wide association studies (GWAS) have led to the identification of risk variants. Accordingly, nine GC GWAS have been carried out in the East-Asian population (Appendix p 4). Four of them focused on non-cardia GC<sup>2-5</sup> and three also used cardia GC.<sup>6-8</sup> Because the Lauren classification is not

widely used in Asia, no information on histopathological types were available in most studies. Only two GWAS in the Japanese population used Lauren's diffuse- and intestinal-type without information on tumour location.<sup>9,10</sup> In total, 12 GC loci have been identified in the East-Asian GWAS (Appendix p 5). Only one GC GWAS has been carried out in Europeans. This study used an Icelandic sample comprising 400 cases along with 2100 patients' relatives that were counted as cases (Appendix p 4). GC-associations were found on chromosome 11q22 and three East-Asian GC loci were replicated.<sup>11</sup> However, no information on GC location- or Lauren-type was available in this study.

To assess the GC genetic susceptibility due to common germline variants we carried out the largest GWAS in the European population and the first GWAS where information on patients' tumours according to location- and Lauren-type were available. We next used expression data from different gastric regions for a transcriptome-wide association study (TWAS) and expression quantitative trait locus (eQTL) study to identify GC risk genes at GWAS loci. Finally, we examined whether oesophago-gastric adenocarcinomas at the GOJ share genetic aetiology. In addition to our GWAS on cardia GC this involved a GWAS on oesophageal adenocarcinoma (OAC) and its precursor lesion Barrett's oesophagus (BO).<sup>12</sup>

## Methods

### Study design and participants

The GC GWAS consisted of ten European case-control samples (Appendix p 6). All patients had a histopathologically confirmed diagnosis of gastric adenocarcinoma and were recruited between 1990 and 2017 across 24 sites (Appendix p 7). The ethnically-matched controls were recruited between 1988 and 2017 across ten sites (Appendix p 8). For the TWAS/eQTL study we obtained genotype and transcriptome data from healthy European individuals that were recruited between 2016 and 2017 across nine sites (Appendix p 9). Informed consent was obtained from all participants and ethics approval was obtained from ethics boards at each participating institution (Appendix p 40).

### Genome-wide genotyping, quality control and imputation

Except for cases from United Kingdom (UK) and Estonia, for which genotypes were drawn from the UK and Estonian Biobank,<sup>13,14</sup> DNA-samples from all GWAS patients and TWAS/eQTL participants were genotyped within this study. For the control samples, genotypes were partly obtained from previous studies. All genotyping array types that were used are listed in the Appendix (p 10).

The pre-imputation quality control (QC) is described in the Appendix (p 3). Single nucleotide polymorphisms

(SNPs) were then imputed for all GWAS samples – with exception of the UK and Estonian datasets – using the TOPMed Imputation Server and TOPMed Reference panel.<sup>15</sup> For the TWAS/eQTL samples, SNPs were imputed using Impute2<sup>16</sup> and 1000 Genomes Phase 3 as reference.<sup>17</sup>

In the post-imputation QC we excluded variants with  $r^2$  less than 0.3 (GWAS) and an information score less than 0.8 (TWAS/eQTL), p-values for Hardy-Weinberg equilibrium (HWE) less than  $1.0 \times 10^{-6}$  in patients and less than  $1.0 \times 10^{-4}$  in non-patients, a minor allele frequency (MAF) less than 0.01 or a SNP-missing rate less than 0.05 for best-guessed genotypes at posterior probability of more than 0.9.

### Gene-expression analysis

For the TWAS/eQTL analysis, tissue biopsies from five gastric regions (cardia, corpus, fundus, antrum and angulus) were collected from healthy individuals during routine gastroscopies. Pre-experiments revealed that the expression in corpus and antrum covers the transcriptome-variance in almost all gastric regions (Appendix p 20). Thus, both were subjected to 3'-mRNA Sequencing for the TWAS/eQTL analysis using the QuantSeq 3'-mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen, Austria). Details on the QC, gene-alignment and expression-quantification are provided in the Appendix (p 3).

### Statistical analysis

GWAS-association testing was performed considering an additive genetic model adjusting for five principal components (PCs) using PLINK2.<sup>18</sup> After computing SNP-associations at single-sample level we performed a meta-analysis considering the fixed-effects inverse variance-weighting approach implemented in METAL.<sup>19</sup> To test for independence of associated variants we applied a conditional analysis by a stepwise selection procedure implemented in GCTA-COJO.<sup>20</sup>

SNP-based GC heritability was calculated using LD score regression (LDSR).<sup>21</sup> We also applied LDSR to determine the genetic correlation between GC and reported risk factors,<sup>1</sup> for which GWAS data from Europeans were available.<sup>22</sup> In total, we used four obesity-, two reflux-, five smoking-, five alcohol- and four education-/employment-related phenotypes that have been assessed by the UK Biobank.<sup>13</sup> As significance threshold we applied a Bonferroni-correction considering the number of investigated traits ( $p = 0.05/20 = 2.5 \times 10^{-3}$ ).

For the TWAS we created expression prediction models for all genes in the transcriptome with FUSION<sup>23</sup> using local SNPs present in the HapMap3 and GC GWAS data (500 kb up- and downstream of the annotated gene start and stop). For each expressed gene we correlated the predicted gene-expression with our GWAS data using linkage



disequilibrium (LD) data from prediction models to identify significant expression disease-associations after Bonferroni-correction for the number of tested genes ( $p_{\text{corpus}} = 0.05/3269 = 1.5 \times 10^{-5}$ ,  $p_{\text{antrum}} = 0.05/4182 = 1.1 \times 10^{-5}$ ).

For the eQTL analysis expression and genotype data were analysed using QTLtools<sup>24</sup> according to parameters used by The Genotype-Tissue Expression (GTEx) project.<sup>25</sup> Briefly, sex, three genotype-based PCs and a set of PEER-factors derived from the normalized expression data were used for adjustment. A window of 1 Mb around the transcription start site (TSS) of each gene was defined for cis-eQTL detection. SNP-gene pairs were considered significant with a nominal p-value below a genome-wide empirical p-value threshold ( $p_i$ ) determined for each gene by extrapolation from a Beta distribution fitted to adaptive permutations. In contrast to the polygenic TWAS models, eQTL analyses evaluate the effect of single genetic marker on gene expression and are not followed up by testing for association in GWAS datasets.

To examine the genetic architecture of oesophago-gastric adenocarcinomas at the GOJ we computed polygenic risk scores (PRS) by considering as base in-house GWAS data from a European OAC/BO case-control sample that was part of a GWAS published previously<sup>12</sup> and for which individual GWAS data were available (Appendix p 11). PRS were calculated after clumping (250 kb regions, clump-p = 1, clump- $r^2$  = 0.1) by testing different p-value thresholds (from genome-wide significant ( $p = 5.0 \times 10^{-8}$ ) to the full model ( $p = 1$ )) using PRSice tool.<sup>26</sup> As for the GWAS, the analysis was performed independently in each sample considering the overall GC status. Logistic regression models between PRS and the phenotypic status were computed. The single-sample PRS regression-coefficients (beta and standard error) were then combined into a meta-analysis using the restricted maximum-likelihood (REML) estimator as implemented in the R package *metaphor*.

The PRS analysis revealed a shared polygenic risk architecture between cardia GC and OAC/BO. We, thus, performed a cross-trait GWAS meta-analysis using our cardia GC and the in-house OAC/BO samples. In addition, we used GWAS summary statistics from the other OAC/BO datasets that were published previously.<sup>12</sup> All samples that were included are listed in the Appendix (p 11). The cross-trait GWAS meta-analysis was performed considering a fixed-effects inverse variance-weighting approach implemented in METAL.<sup>19</sup>

### Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all data and final responsibility for the decision to submit for publication.

## Results

In total, information on 7,829,683 SNPs in 5815 patients and 10,999 controls from ten European GWAS were included in the GC meta-analysis. Furthermore, 1291 cardia and 3183 non-cardia GC patients were available for the location-specific as well as 1368 diffuse and 1696 intestinal GC patients for the Lauren-specific GWAS meta-analysis. All Q-Q and Manhattan plots from the GWAS are shown in the Appendix (pp 21–25). The genomic inflation factor lambda was 1.11, 1.07, 1.11, 1.06, and 1.07 for the entire, cardia, non-cardia, diffuse, and intestinal GC GWAS. The lambda values suggest the presence of moderate inflation which can be expected given the experimental design of a GWAS meta-analysis with this sample size including cohorts from different European populations and centres.

We identified four genome-wide significant GWAS loci, all with subtype-specific GC-association. Two of them – on chromosome 1q22 and 8q24 – have been reported previously,<sup>3–5,7,8,10,11</sup> but not with respect to subtype-specificity. The other two – on chromosome 2p23 and 17q12 – have not been described before. Table 1 lists all associated loci and Fig. 1 shows the subtype-specificity. The consistency of all associations across samples is shown in the Appendix (p 26).

SNP rs760077 on chromosome 1q22 near *MUC1* showed genome-wide significant GC-association ( $p = 5.23 \times 10^{-21}$ , odds ratio (OR) of 1.27 (95% confidence interval (CI) 1.21–1.34)). The risk variant was strongest associated to non-cardia ( $p = 5.12 \times 10^{-17}$ , OR of 1.31 (95% CI 1.23–1.40)) and diffuse GC ( $p = 7.41 \times 10^{-11}$ , OR of 1.35 (95% CI 1.23–1.48)), while the association was nearly absent in cardia GC ( $p = 0.045$ ) (Table 1 and Fig. 1). Moreover, conditioning on the lead SNP using COJO<sup>20</sup> revealed an independent GC-association. SNP rs67579710 showed genome-wide significant association to entire ( $p = 5.89 \times 10^{-13}$ , OR of 1.21 (95% CI 1.11–1.31)) and non-cardia GC ( $p = 4.62 \times 10^{-12}$ , OR of 1.27 (95% CI 1.15–1.42)) (Table 1).

On chromosome 8q24 near *PSCA* we found the strongest GC-association for rs2920293 in the entire sample ( $p = 2.84 \times 10^{-32}$ , OR of 1.39 (95% CI 1.31–1.47)). Also this association was strongest to non-cardia ( $p = 1.80 \times 10^{-30}$ , OR of 1.46 (95% CI 1.36–1.55)) and diffuse GC ( $p = 8.10 \times 10^{-17}$ , OR 1.46 (95% CI 1.33–1.60)), while cardia GC showed no association ( $p = 0.114$ ) (Table 1 and Fig. 1).

SNP rs11677924 on chromosome 2p23 within intron 4 of *ALK* showed suggestive association in the entire GC sample ( $p = 2.37 \times 10^{-7}$ , OR of 1.19 (95% CI 1.11–1.27)) and genome-wide significant association to intestinal GC ( $p = 2.04 \times 10^{-8}$ , OR of 1.34 (95% CI 1.21–1.49)) (Table 1). In contrast, this variant was only moderately associated to diffuse GC ( $p = 0.002$ ) (Table 1 and Fig. 1).

The association to intestinal GC was even more pronounced for rs17138478 on chromosome 17q12

within intron 4 of *HNF1B*. This variant showed suggestive association in the entire GC sample ( $p = 4.30 \times 10^{-6}$ , OR of 1.19 (95% CI 1.10–1.28)) and genome-wide significant association to intestinal GC ( $p = 1.83 \times 10^{-8}$ , OR of 1.44 (95% CI 1.27–1.64)) (Table 1). In contrast, rs17138478 was not associated to diffuse GC ( $p = 0.883$ ) (Table 1 and Fig. 1).

We next focused on all 10 remaining GC risk loci that have been reported previously (Appendix p 13).<sup>2-11</sup> We could replicate disease-associations at three loci. On chromosome 4q28, we observed the strongest association for rs10029005 near *ANKRD50* in non-cardia GC patients ( $p = 4.37 \times 10^{-4}$ , OR of 1.12 (95% CI 1.05–1.19)) (Table 1). On chromosome 5q13, rs6897169 near *PTGER4* was also strongest associated to non-cardia GC ( $p = 9.40 \times 10^{-5}$ , OR of 1.18 (95% CI 1.09–1.28)) (Table 1). Moreover, rs532436 on chromosome 9q34 within intron 1 of *ABO* encoding for the blood-group system was associated to non-cardia ( $p = 5.82 \times 10^{-6}$ , OR of 1.19 (95% CI 1.10–1.28)) and – even stronger – to diffuse GC ( $p = 8.52 \times 10^{-7}$ , OR of 1.29 (95% CI 1.17–1.44)) (Table 1). We, thus, inferred the *ABO* blood-groups using rs8176719 and rs8176746, which determine the *ABO* system (Appendix p 14). This revealed that blood-group 0 protects against non-cardia ( $p = 2.16 \times 10^{-4}$ , OR of 0.86 (95% CI 0.90–0.82)) and diffuse GC ( $p = 3.04 \times 10^{-5}$ , OR of 0.76 (95% CI 0.70–0.82)), while blood-group A increases risk for both GC subtypes (non-cardia GC:  $p = 9.27 \times 10^{-10}$ , OR of 1.28 (95% CI 1.32–1.24), diffuse GC:  $p = 8.01 \times 10^{-6}$ , OR of 1.31 (95% CI 1.37–1.25)) (Appendix p 14).

In pre-experiments for our TWAS/eQTL study we compared the expression profiles from five gastric regions, which showed that the expression in corpus and antrum covers the transcriptome-variance in almost all gastric regions (Appendix p 20). We, thus, selected the corpus and antrum mucosa from 362 and 342 individuals for transcriptome-wide profiling. This revealed specific expression profiles in both gastric regions (Appendix p 27), which were followed-up in our TWAS/eQTL study to prioritize GC risk genes.

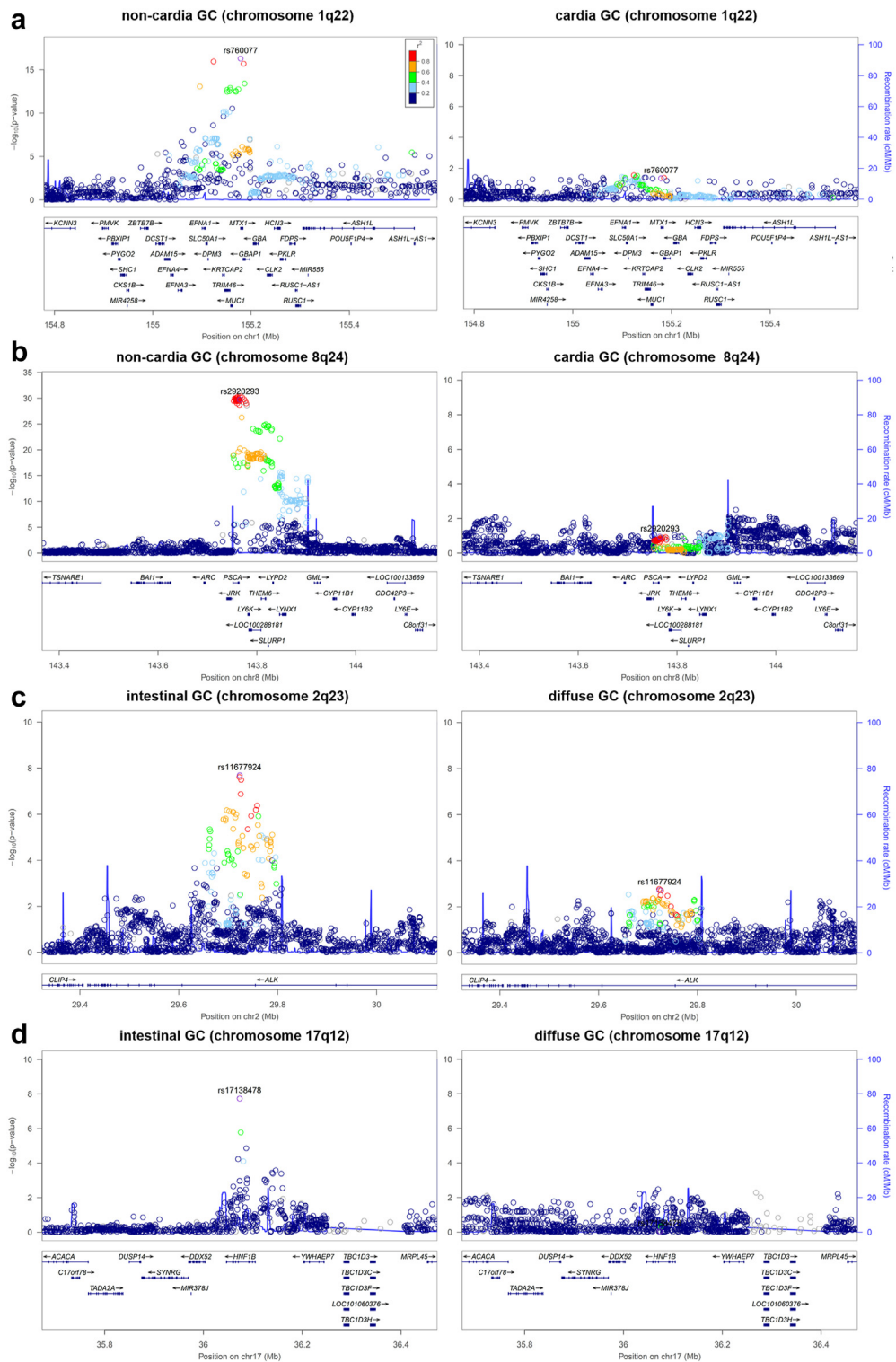
On chromosome 1q22, we found significant TWAS-effects to non-cardia ( $p = 3.83 \times 10^{-13}$ ) and diffuse GC ( $p = 2.40 \times 10^{-10}$ ) with an upregulated *MUC1*-expression in corpus mucosa (Table 2 and Appendix (pp 28–31)). Conditioning on these findings revealed that the *MUC1*-expression explains most of the GWAS signals (Appendix pp 32–33). Although the upregulated *MUC1*-expression was also present in antrum mucosa, this TWAS-effect was considerably less significant (non-cardia GC:  $p = 5.26 \times 10^{-6}$ , diffuse GC:  $p = 1.01 \times 10^{-5}$ ) (Table 2).

On chromosome 8q24, an upregulated expression of four genes – *PSCA*, *LY6K*, *THEM6* and *LYNX1* – showed significant GC-associations in the TWAS (Table 2 and Appendix (pp 28–31)). The upregulated *PSCA*-expression showed the most significant

SNP	Chromosome (position in bp (hg38))	Effect /other allele <sup>a</sup>	Entire GC sample			Cardia GC sample			Non-cardia GC sample			Diffuse GC sample			Intestinal GC sample		
			p value	OR	95% CI	p value	OR	95% CI	p value	OR	95% CI	p value	OR	95% CI	p value	OR	95% CI
rs760077	1q22 (155,208,991)	T/A	5.23E-21	1.27	1.21-1.34	4.45E-02	1.09	1.00-1.19	5.12E-17	1.31	1.23-1.40	7.41E-11	1.35	1.23-1.48	1.78E-07	1.24	1.14-1.35
rs67579710 <sup>b</sup>	1q22 (155,203,736)	G/A	5.89E-13	1.21	1.11-1.31	-	-	-	4.62E-12	1.27	1.15-1.42	-	-	-	-	-	-
rs11677924	2p23 (29,500,326)	G/C	2.37E-07	1.19	1.11-1.27	5.91E-03	1.17	1.04-1.32	1.29E-05	1.20	1.10-1.30	1.90E-03	1.20	1.07-1.35	2.04E-08	1.34	1.21-1.49
rs10029005	4q28 (124,530,209)	A/G	4.69E-04	1.09	1.04-1.15	9.31E-01	1.00	0.91-1.09	4.37E-04	1.12	1.05-1.19	2.24E-02	1.11	1.01-1.21	7.72E-02	1.08	0.99-1.16
rs6897169	5q13 (40,726,036)	C/T	2.96E-04	1.13	1.06-1.21	2.65E-01	1.07	0.95-1.19	9.40E-05	1.18	1.09-1.28	4.20E-02	1.13	1.00-1.23	1.18E-02	1.15	1.03-1.27
rs2920293 <sup>c</sup>	8q24 (142,683,996)	G/C	2.84E-32	1.39	1.31-1.47	1.14E-01	1.08	0.98-1.19	1.80E-30	1.46	1.36-1.55	8.10E-17	1.46	1.33-1.60	4.05E-09	1.27	1.17-1.37
rs532436 <sup>c</sup>	9q34 (133,274,414)	A/G	7.51E-05	1.22	1.07-1.22	4.95E-01	1.04	0.93-1.17	5.82E-06	1.19	1.10-1.28	8.52E-07	1.29	1.17-1.44	3.02E-03	1.15	1.05-1.26
rs17138478	17q12 (37,713,312)	C/A	4.30E-06	1.19	1.10-1.28	3.96E-02	1.15	1.00-1.31	6.98E-04	1.17	1.07-1.29	8.83E-01	1.00	0.88-1.14	1.83E-08	1.44	1.27-1.64

The associations are shown for the risk alleles (effect alleles) in the entire GC sample as well as in the location- (cardia, non-cardia) and Lauren-specific (diffuse, intestinal) GC samples. p-values, odds ratios (ORs) and the corresponding 95% confidence intervals (CIs) are shown. Allele frequencies for the associated SNPs among patients and controls are not given, as the GWAS samples were meta-analysed. Instead the frequency of effect alleles in the European population are shown according to gnomAD.<sup>27</sup> In addition, on chromosome 8q24 and 9q34 the LD between the lead GC SNPs in the present and East Asian GWAS is shown. At all replicated loci, the same associations contribute to GC risk across populations. All genome-wide significant associated GC loci also remain genome-wide significant associated after applying a genomic-control correction (genomic-control corrected p-values are shown in the Appendix (p 19)).<sup>a</sup>Frequency of effect alleles in the European (non-Finnish) population according to gnomAD.<sup>27</sup> rs760077 allele T 59%, rs67579710 allele G 90%, rs11677924 allele G 15%, rs10029005 allele A 42%, rs6897169 allele C 17%, rs2920293 allele G 47%, rs532436 allele G 20%, rs17138478 allele C 86%.<sup>b</sup>rs67579710 was the only variant that showed additional association in the entire study after conditioning on lead SNPs using COJO.<sup>20</sup> The independent association signal on chromosome 1q22 appeared in the entire and non-cardia GC sample and, thus, the associations are not shown for the other location- or Lauren-specific GC samples. <sup>c</sup> $r^2$  of 0.88 between rs2920293 and rs2978977, the lead GC SNP on chromosome 8q24 in the East Asian population.<sup>8</sup> rs532436 corresponds to the lead SNP rs7849280 in the East Asian population, where a high LD between both variants is present ( $r^2 = 0.76$ ).<sup>27</sup> However, no LD between both variants is observed in the European population ( $r^2 = 0.01$ ).<sup>27</sup>

**Table 1:** Lead associations of genome-wide significant and replicated GC risk loci.



**Fig. 1: Regional association plots of GC risk loci.** Disease associations are shown for non-cardia and cardia GC on chromosome 1q22 (a) and chromosome 8q24 (b). In addition, disease associations are shown for intestinal and diffuse GC on chromosome 2q23 (c) and chromosome 17q12 (d). At each locus, associations ( $-\log_{10}(p\text{-value})$ ) are shown for SNPs flanking 400 kb on either side of the lead associated SNP (Position



Chromosomal region	Tissue	GC type	Gene	NWGT SNPs	TWAS Z score	TWAS p value	
1q22	Corpus	non-cardia	MUC1	2	7.26	3.83E-13	
		diffuse	MUC1	2	6.33	2.40E-10	
	Antrum	non-cardia	MUC1	1	4.55	5.26E-06	
		diffuse	MUC1	1	4.41	1.01E-05	
8q24	Corpus	non-cardia	PSCA	30	11.46	2.14E-30	
			LY6K	1	8.73	2.46E-18	
			THEM6	1	9.31	1.23E-20	
			LYNX1	26	-7.23	4.85E-13	
			diffuse	PSCA	30	8.14	4.11E-16
			LY6K	1	6.31	2.84E-10	
		THEM6	1	6.34	2.28E-10		
		intestinal	PSCA	30	-4.65	3.35E-06	
			LY6K	1	5.96	2.60E-09	
			THEM6	1	5.11	3.19E-07	
			LYNX1	1	5.24	1.62E-07	
			Antrum	non-cardia	PSCA	46	11.34
	LY6K				26	6.75	1.46E-11
	THEM6	1		9.20	3.48E-20		
	LYNX1	1		-6.57	5.17E-11		
	diffuse	PSCA	46	8.04	8.84E-16		
		LY6K	26	5.90	3.69E-09		
		THEM6	1	6.26	3.87E-10		
		LYNX1	1	-4.29	1.81E-05		
		intestinal	PSCA	46	5.81	6.32E-09	
THEM6			1	5.17	2.39E-07		

In total, two risk loci (1q22, 8q24) and three GC types (non-cardia, diffuse, intestinal) as well as five genes were implicated. The number of weighted SNPs (NWGT SNPs, maximum of 50 variants) in the best predicting expression model is shown. In addition, TWAS Z scores indicating the effect of GC expression-association (downregulated/upregulated) and corresponding TWAS p-values are shown.

**Table 2: Transcriptome-wide significant GC-associations.**

association (Table 2). Moreover, *PSCA* was the only gene whose expression explained most of the GWAS signals in all conditional analyses (Appendix pp 34–37). Thus, *PSCA* can be prioritized as the most relevant GC gene at this locus. The *PSCA*-association was strongest to non-cardia (corpus:  $p = 2.14 \times 10^{-30}$ , antrum:  $p = 8.42 \times 10^{-30}$ ) and diffuse GC (corpus:  $p = 4.11 \times 10^{-16}$ , antrum:  $p = 8.84 \times 10^{-16}$ ) (Table 2).

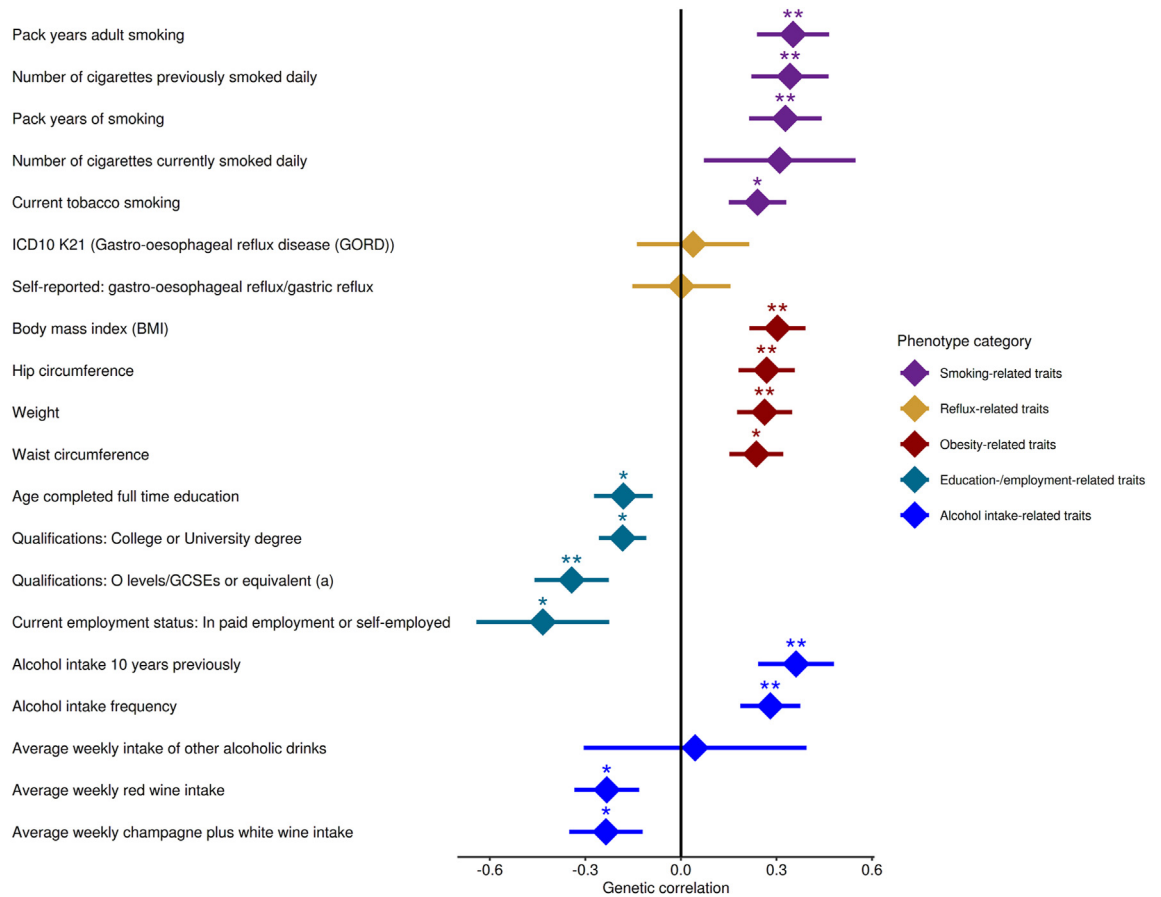
Individual eQTL-effects were comparatively smaller than TWAS-effects at most GWAS loci, highlighting the polygenic component of gene-expression. However, significant single-marker eQTLs at two replicated GC loci were found. On chromosome 4q28, risk allele-carrier at rs10029005 showed an upregulated *ANKRD50*-expression in corpus mucosa ( $p = 2.77 \times 10^{-12}$ ) (Appendix p 38). On chromosome 5q13, an upregulated *PTGER4*-expression in corpus ( $p = 5.66 \times 10^{-21}$ ) and to a lesser extent in antrum mucosa ( $p = 4.67 \times 10^{-6}$ ) was present among risk allele-carrier at rs6897169 (Appendix p 38). Both

eQTL-effects most probably did not reach significance in the TWAS because the significance of disease-association influences the TWAS outcome.

We next applied LDSR and estimated the SNP-based GC heritability to be  $8.48 \pm 3.12\%$  standard deviation (SD). Unfortunately, we were not able to apply LDSR using GC subtypes due to the limited size of our location- and Lauren-specific samples. However, in the genetic correlation analysis using 20 traits belonging to five phenotype-categories representing GC risk factors, a significant genetic correlation with GC was found for three obesity-related traits, one smoking- and one alcohol-related trait (Fig. 2 and Appendix (p 15)).

Cardia GC and OAC are located in close proximity at the GOJ and belong to oesophago-gastric adenocarcinomas. Both cancer types share common characteristics, as they are related to GOR and their prevalence is rising.<sup>1</sup> Thus, it remains controversial whether cardia GC and OAC rather represent one cancer entity.<sup>28</sup>

in hg19). The lead variant is shown in purple. Other markers at each locus are displayed by different colours, which indicates different levels of LD ( $r^2$ ) to the lead SNP. Furthermore, all annotated genes within each region are shown with arrows indicating their transcription direction. The location- and Lauren-specific effects at each locus are also shown in a case-case comparison in the Appendix (p 12).

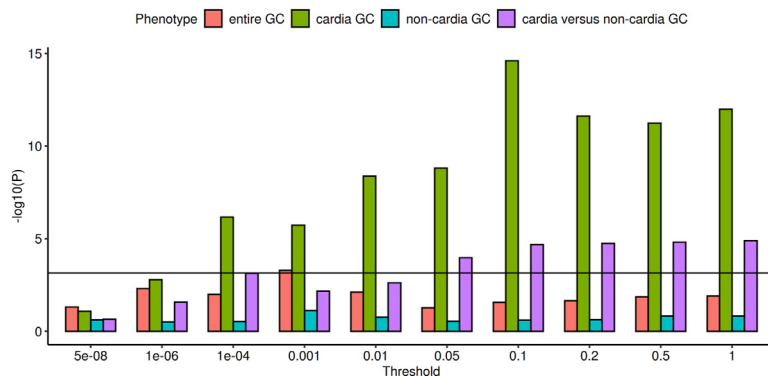


**Fig. 2: Genetic correlation between GC and risk factors.** Genetic correlations determined with LDSR between GC and 20 traits belonging to five phenotype-categories that represent risk factors for GC development are shown. For each trait the genetic correlation (dot) and the standard deviation (line) is given. The significance levels of the genetic correlation are indicated by asterisks (\*p value < 0.05, \*\*p value < 0.0025). Body mass index ( $r_g = 0.303$ ,  $p = 6.0 \times 10^{-4}$ ), hip circumference ( $r_g = 0.269$ ,  $p = 2.3 \times 10^{-3}$ ) and weight ( $r_g = 0.262$ ,  $p = 2.4 \times 10^{-3}$ ) showed Bonferroni-corrected positive GC correlation along with pack years of adult smoking ( $r_g = 0.352$ ,  $p = 2.0 \times 10^{-3}$ ) and alcohol intake 10 years previously ( $r_g = 0.361$ ,  $p = 2.0 \times 10^{-3}$ ).

To examine the genetic relation between cardia GC and OAC we performed PRS analyses. We calculated PRS in two in-house discovery GWAS datasets with OAC as well as OAC and its precursor lesion BO.<sup>12</sup> We then determined the proportion of variance explained by PRS from the discovery sets in four target datasets. We used the cardia GC GWAS as target set to examine the genetic relation to OAC and OAC/BO. In addition, we used the GWAS of the entire GC, non-cardia GC and cardia-versus-non-cardia GC as target sets to determine whether associations between cardia GC and OAC as well as OAC/BO are specific. In the Appendix (pp 16–17) all p-value thresholds, the number of SNPs included in each PRS and the observed associations are shown. We found highly significant associations between cardia GC and PRS derived from OAC ( $p_{\text{threshold}} = 0.001$ ,  $p_{\text{association}} = 2.37 \times 10^{-8}$ ) and OAC/BO ( $p_{\text{threshold}} = 0.2$ ,  $p_{\text{association}} = 2.79 \times 10^{-17}$ ) (Fig. 3). In contrast, no

PRS-associations were present in the other GC case-control datasets (Fig. 3). Accordingly, in the case-case comparison (cardia-versus-non-cardia GC) we found significant associations to PRS derived from OAC ( $p_{\text{threshold}} = 0.5$ ,  $p_{\text{association}} = 2.18 \times 10^{-9}$ ) and OAC/BO ( $p_{\text{threshold}} = 0.2$ ,  $p_{\text{association}} = 4.33 \times 10^{-7}$ ) (Fig. 3). The results imply a shared genetic aetiology of cardia GC and OAC as well as OAC/BO, while OAC- and OAC/BO-PRS enable to discriminate between cardia and non-cardia GC.

Based on the shared polygenic risk of cardia GC and OAC/BO we performed a meta-analysis combining our GWAS datasets. This included 1291 cardia GC and 10,279 OAC/BO cases as well as 27,326 controls (Appendix 11). In total, we identified 17 genome-wide significant associated loci for oesophago-gastric adenocarcinoma that are – along with the corresponding Manhattan plot – shown in the Appendix (p 18 and p



**Fig. 3: Polygenic risk score associations for OAC in the target GC subtypes.** The association-values in dependence of different significance thresholds for PRS variant selection are given. The horizontal black line represents the Bonferroni correction threshold. All PRS associations for OAC/BO in the target GC subtypes are shown in the [Appendix](#) (p 17).

39). Two of the identified loci have not been described before in the OAC/BO-only GWAS.<sup>12</sup> SNP rs1817002 near *HNF4G* on chromosome 8q21 showed disease-association with  $p = 4.10 \times 10^{-8}$  (OR of 1.11 (95% CI 1.06–1.14)) and rs234506 near *NR2F2* on chromosome 15q26 showed disease-association with  $p = 1.56 \times 10^{-9}$  (OR of 1.12 (95% CI 1.07–1.16)).

## Discussion

The strength of this study is that the genetic heterogeneity of GC could be assessed, as patients' tumours were characterised according to location and histopathology. This revealed that the known risk loci on chromosome 1q22 (*MUC1*), 4q28 (*ANKRD50*), 5p13 (*PTGER4*), 8q24 (*PSCA*), and 9q34 (*ABO*)<sup>3,4,7–9,11</sup> contribute almost exclusively to non-cardia GC. Moreover, the risk loci on chromosome 1q22 (*MUC1*), 8q24 (*PSCA*), and 9q34 (*ABO*) confer substantially higher risk to diffuse than to intestinal GC. In contrast, the newly identified risk loci on chromosome 2p23 (*ALK*) and 17q12 (*HNF1B*) contribute almost exclusively to intestinal GC.

Our TWAS/eQTL study revealed that an upregulated expression of *MUC1*, *ANKRD50*, *PTGER4*, and *PSCA* are the most plausible GC-pathomechanism at four GWAS loci. These effects were strongest in corpus mucosa, except for *PSCA* that also showed considerable GC-association in antrum mucosa. Moreover, the upregulated *MUC1*- and *PSCA*-expression contributed strongest to diffuse GC.

While it has been shown previously that *PSCA* represents the most plausible risk gene at chromosome 8q24,<sup>29,30</sup> the expression-effects of the remaining genes on GC-pathophysiology are novel. *MUC1* exerts proto-oncogenic effects via a cytoplasmic domain (CD).<sup>31</sup> This might explain why *MUC1* particularly increases risk for diffuse GC. It has been shown, that over-expression of *MUC1*-CD disrupts E-cadherin functionality via increased binding of  $\beta$ -catenin.<sup>31</sup> The

E-cadherin/ $\beta$ -catenin complex plays an important role in maintaining epithelial integrity and mutations in *CDH1*—encoding *E-cadherin*—lead to hereditary diffuse GC (HDGC), the most common monogenic GC.<sup>1</sup> So far, little is known about the function of *ANKRD50*. However, with regard to cancer *ANKRD50* represents a hotspot for the genomic integration of human papillomaviruses in the context of cervical carcinoma development.<sup>32</sup> *PTGER4* encodes the prostaglandin E2 (PGE2) receptor 4, which mediates cellular responses to PGE2. It has been shown previously that PGE2 is important for the inflammatory microenvironment in tumours.<sup>33</sup> Based on these findings, functional studies are now required to identify the GC-relevant disease mechanisms in detail.

Multiple studies, mainly from the 1950s and 1960s, reported an association between ABO blood-groups and gastrointestinal cancers, most strongly for gastric and pancreatic cancer.<sup>34,35</sup> We confirm these findings on the genetic level and show that the protective effect of blood-group 0 and the risk effect of blood-group A is restricted to non-cardia and diffuse GC in Europeans. As with the risk genes mentioned above, functional studies are now required to show how *ABO* contributes to non-cardia and diffuse GC at the cellular level.

Although we could not functionally characterise the remaining GWAS loci, promising GC candidate genes are located in close proximity to the associated risk variants. The risk SNP on chromosome 2p23 is located in intron 4 of *ALK*, which encodes for a receptor tyrosine kinase that plays a role in cancer development. Germline mutations in *ALK* lead to monogenic neuroblastomas<sup>36</sup> and somatic *ALK* mutations are key driver events in non-small cell lung cancers,<sup>37</sup> but have also been observed in GC.<sup>38</sup> On chromosome 17q12 the risk SNP is located in intron 4 of *HNF1B*, which encodes a transcription factor that plays a role in cancer development. Accordingly, SNPs in *HNF1B* have been repeatedly identified as risk factors in GWAS of prostate,

endometrial, ovarian, and pancreatic cancer.<sup>39–42</sup> However, we could not functionally characterise both GWAS loci and, thus, it remains speculative that *ALK* and *HNF1B* represent the true susceptibility genes for GC at these loci.

This study is the first that analysed GWAS data for GC on the polygenic level. We estimated the SNP-based GC heritability to be  $8.48 \pm 3.12$  SD, which already explains around 30% of the reported GC twin-based heritability of 28%.<sup>43</sup> Furthermore, we found genetic correlations between GC and GC risk factors obesity, smoking and alcohol-intake on the polygenic level.

Most importantly, we found highly significant associations between cardia GC and PRS derived from OAC and OAC/BO. Moreover, OAC- and OAC/BO-PRS enabled to discriminate between cardia and non-cardia GC. The findings imply that cardia GC and OAC or oesophago-gastric adenocarcinomas at the GOJ share polygenic risk architecture, which points to common molecular mechanisms conferring to both cancer types.

Promising candidate genes are also in close vicinity to the newly identified risk SNPs for oesophago-gastric adenocarcinoma, namely *HNF4G* on chromosome 8q21 and *NR2F2* on chromosome 15q26. *NR2F2* is a known coregulator of *HNF4G* and both genes play a prominent role in intestinal-like cell transformations in gastric cell lineages.<sup>44</sup>

Our study has limitations. Because GWAS only allow the identification of common risk variants for multifactorial diseases, we could not identify rare risk variants in our patients. Thus, we are not able to completely capture the genetic risk architecture underlying GC. Future studies are warranted that cover the entire genetic variability leading to GC development. Furthermore, the size of our GWAS sample was not large enough to develop GC PRS that can be used for disease prediction. For other cancer types, this has already become possible due to the availability of larger GWAS samples.

In summary, this European GWAS shows that all known and new risk loci contribute subtype-specifically to GC. Our findings should further help to elucidate the pathomechanisms underlying each GC type. At four GWAS loci we identified that an upregulated *MUC1*, *ANKRD50*-, *PTGER4*-, and *PSCA*-expression represent the most plausible GC-pathomechanisms. In addition, we show that blood-group 0 exerts protective effects against non-cardia and diffuse GC, while blood-group A increases risk for both GC subtypes. Finally, we identified that cardia GC and OAC share genetic aetiology and, thus, support that common molecular pathomechanisms confer risk to oesophago-gastric adenocarcinomas at the GOJ.

#### Contributors

CarM, JG, JohS, MarV, and TH designed the study, verified the underlying data, and wrote the manuscript. AndMa, CarM, EH, JG, JanS, Mick, OB, PK, TH, and VS performed statistical analysis and conducted

experiments. AdrM, AgoS, AH, AndMe, AngL, APA, AJ, AnnL, AnnM, ABA, AntS, ArmS, AHH, CarP, CZ, ChiS, CJB, ClaP, ClaS, CT, DC, DA, DD, DP, ES, ER, EMP, FMT, FS, FG, FL, FT, FLD, GCT, GT, GK, GP, HA, HanL, HL, HenN, HB, IG, IB, JRI, JL, JFD, JT, JE, JOJ, JH, JK, JurS, JA, JD, KR, KM, KO, LV, LB, LH, MagN, Mall, ManK, MarL, MarMa, MR, MAGG, MPo, MPe, MarV, MDR, MTL, MarMo, MarN, MarK, MatL, MicGe, MS, MicGh, MonL, NF, NK, NV, NA, OK, OC, OG, OP, PE, PM, PPG, PL, RafC, RRS, RenC, RusT, RobPa, RobPe, RA, RM, SEAB, SA, SB, SLH, SusB, ThoS, TOG, TomS, VDR, VicMa, VicMO, VA, and YV contributed to sample collection, including handling of data. AQ, AMH, DH, HIG, JulS, JR, MMN, MicV, and PH reviewed data and provided critical comments or suggestions. JohS had primary responsibility for the final content. All authors read and approved the final version of the manuscript.

#### Data sharing statement

Summary statistics of the GC GWAS are available on request from the corresponding author ([Johannes.schumacher@uni-marburg.de](mailto:Johannes.schumacher@uni-marburg.de)).

#### Declaration of interests

MDR reports consulting fees from Roche Diagnostics and Medtronic; leadership or fiduciary role in the European Society of Gastrointestinal Endoscopy (ESGE) and World Endoscopy Organization (WEO). All other authors report no conflict of interest.

#### Acknowledgments

This study was supported by the German Research Foundation (DFG) (SCHU 1596/6–1, KN 378/3–1, LO 1147/1–1, and VE 917/1–1). Additional funding institutions that indirectly supported the study can be found in the Appendix (p 40). The authors thank all study participants, scientists and clinicians who contributed to this study.

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2023.104616>.

#### References

- 1 Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. *Lancet*. 2020;396(10251):635–648.
- 2 Jin G, Ma H, Wu C, et al. Genetic variants at 6p21.1 and 7p15.3 are associated with risk of multiple cancers in Han Chinese. *Am J Hum Genet*. 2012;91(5):928–934.
- 3 Shi Y, Hu Z, Wu C, et al. A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat Genet*. 2011;43(12):1215–1218.
- 4 Wang Z, Dai J, Hu N, et al. Identification of new susceptibility loci for gastric non-cardia adenocarcinoma: pooled results from two Chinese genome-wide association studies. *Gut*. 2017;66(4):581–587.
- 5 Yan C, Zhu M, Ding Y, et al. Meta-analysis of genome-wide association studies and functional assays decipher susceptibility genes for gastric cancer in Chinese populations. *Gut*. 2020;69(4):641–651.
- 6 Abnet CC, Freedman ND, Hu N, et al. A shared susceptibility locus in *PLCE1* at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat Genet*. 2010;42(9):764–767.
- 7 Hu N, Wang Z, Song X, et al. Genome-wide association study of gastric adenocarcinoma in Asia: a comparison of associations between cardia and non-cardia tumours. *Gut*. 2016;65(10):1611–1618.
- 8 Ishigaki K, Akiyama M, Kanai M, et al. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet*. 2020;52(7):669–679.
- 9 Study Group of Millennium Genome Project for Cancer, Sakamoto H, Yoshimura K, Saeki N, et al. Genetic variation in *PSCA* is associated with susceptibility to diffuse-type gastric cancer. *Nat Genet*. 2008;40(6):730–740.
- 10 Tanikawa C, Kamatani Y, Toyoshima O, et al. Genome-wide association study identifies gastric cancer susceptibility loci at 12q24.11–12 and 20q11.21. *Cancer Sci*. 2018;109(12):4015–4024.
- 11 Helgason H, Rafnar T, Olafsdottir HS, et al. Loss-of-function variants in *ATM* confer risk of gastric cancer. *Nat Genet*. 2015;47(8):906–910.

- 12 Gharahkhani P, Fitzgerald RC, Vaughan TL, et al. Genome-wide association studies in oesophageal adenocarcinoma and Barrett's oesophagus: a large-scale meta-analysis. *Lancet Oncol.* 2016;17(10):1363–1373.
- 13 Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779.
- 14 Leitsalu L, Haller T, Esko T, et al. Cohort profile: Estonian biobank of the Estonian genome center, University of Tartu. *Int J Epidemiol.* 2015;44(4):1137–1147.
- 15 Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature.* 2021;590(7845):290–299.
- 16 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5(6):e1000529.
- 17 Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68–74.
- 18 Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–575.
- 19 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190–2191.
- 20 Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet.* 2012;44(4):369–375, S1–S3.
- 21 Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015;47(11):1236–1241.
- 22 Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics.* 2017;33(2):272–279.
- 23 Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet.* 2016;48(3):245–252.
- 24 Fort A, Panousis NI, Garieri M, et al. MBV: a method to solve sample mislabeling and detect technical bias in large combined genotype and sequencing assay datasets. *Bioinformatics.* 2017;33(12):1895–1897.
- 25 GTEx Consortium, Laboratory, Data Analysis and Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204–213.
- 26 Choi SW, O'Reilly PF. PRSice-2: polygenic Risk Score software for biobank-scale data. *Gigascience.* 2019;8(7).
- 27 Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434–443.
- 28 Hayakawa Y, Sethi N, Sepulveda AR, Bass AJ, Wang TC. Oesophageal adenocarcinoma and gastric cancer: should we mind the gap? *Nat Rev Cancer.* 2016;16(5):305–318.
- 29 Sung H, Hu N, Yang HH, et al. Association of high-evidence gastric cancer susceptibility loci and somatic gene expression levels with survival. *Carcinogenesis.* 2017;38(11):1119–1128.
- 30 Heinrichs SKM, Hess T, Becker J, et al. Evidence for PTGER4, PSCA, and MBOAT7 as risk genes for gastric cancer on the genome and transcriptome level. *Cancer Med.* 2018;7(10):5057–5065.
- 31 Farahmand L, Merikhian P, Jalili N, Darvishi B, Majidzadeh AK. Significant role of MUC1 in development of resistance to currently existing anti-cancer therapeutic agents. *Curr Cancer Drug Targets.* 2018;18(8):737–748.
- 32 Yang-Chun F, Sen-Yu W, Yuan Z, Yan-Chun H. Genome-wide profiling of human papillomavirus DNA integration into human genome and its influence on PD-L1 expression in Chinese Uygur cervical cancer women. *J Immunol Res.* 2020;2020:6284960.
- 33 Echizen K, Hirose O, Maeda Y, Oshima M. Inflammation in gastric cancer: interplay of the COX-2/prostaglandin E2 and Toll-like receptor/MyD88 pathways. *Cancer Sci.* 2016;107(4):391–397.
- 34 Aird I, Bentall HH, Roberts JA. A relationship between cancer of stomach and the ABO blood groups. *Br Med J.* 1953;1(4814):799–801.
- 35 Marcus DM. The ABO and Lewis blood-group system. Immunology, genetics and relation to human disease. *N Engl J Med.* 1969;280(18):994–1006.
- 36 Mosse YP, Laudenslager M, Longo L, et al. Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature.* 2008;455(7215):930–935.
- 37 Soda M, Choi YL, Enomoto M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature.* 2007;448(7153):561–566.
- 38 Wen Z, Xiong D, Zhang S, et al. Case report: RAB10-ALK: a novel ALK fusion in a patient with gastric cancer. *Front Oncol.* 2021;11:645370.
- 39 Gudmundsson J, Sulem P, Steinthorsdottir V, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet.* 2007;39(8):977–983.
- 40 Klein AP, Wolpin BM, Risch HA, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat Commun.* 2018;9(1):556.
- 41 Pharoah PD, Tsai YY, Ramus SJ, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet.* 2013;45(4):362–370.
- 42 Spurdle AB, Thompson DJ, Ahmed S, et al. Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat Genet.* 2011;43(5):451–454.
- 43 Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med.* 2000;343(2):78–85.
- 44 Sousa JF, Nam KT, Petersen CP, et al. miR-30-HNF4gamma and miR-194-NR2F2 regulatory networks contribute to the upregulation of metaplasia markers in the stomach. *Gut.* 2016;65(6):914–924.