Molecular Profiling Plasma Extracellular Vesicles From Breast Cancer Patients

By

Zhenyu Zhong

A Dissertation Presented in Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2018 by the
Graduate Supervisory Committee:

David Spetzler, Co-Chair
Hao Yan, Co-Chair
Douglas Lake
Marco Mangone

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Extracellular vesicles (EVs) represent a heterogeneous population of small vesicles, consisting of a phospholipidic bilayer surrounding a soluble interior cargo. These vesicles play an important role in cellular communication by virtue of their protein, RNA, and lipid content, which can be transferred among cells. Peripheral blood is a rich source of circulating EVs. An analysis of EVs in peripheral blood could provide access to unparalleled amounts of biomarkers of great diagnostic, prognostic as well as therapeutic value. In the current study, a plasma EV enrichment method based on pluronic co-polymer was first established and characterized. Plasma EVs from breast cancer patients were then enriched, profiled and compared to non-cancer controls. Proteins signatures that contributed to the prediction of cancer samples from non-cancer controls were created by a random-forest based cross-validation approach. We found that a large portion of these signatures were related to breast cancer aggression. To verify such findings, KIAA0100, one of the features identified, was chosen for *in vitro* molecular and cellular studies in the breast cancer cell line MDA-MB-231. We found that KIAA0100 regulates cancer cell aggression in MDA-MB-231 in an anchorage-independent manner and is particularly associated with anoikis resistance through its interaction with HSPA1A. Lastly, plasma EVs contain not only individual proteins, but also numerous molecular complexes. In order to measure millions of proteins, isoforms, and complexes simultaneously, Adaptive Dynamic Artificial Poly-ligand Targeting (ADAPT) platform was applied. ADAPT employs an enriched library of single-stranded oligodeoxynucleotides to profile complex biological samples, thus achieving a deep coverage of system-wide, native biomolecules. Profiling of EVs from breast cancer

patients was able to obtain a prediction AUC performance of 0.73 when compared biopsy-positive cancer patient to healthy controls and 0.64 compared to biopsy-negative controls and such performance was not associated with the physical breast condition indicated by BIRAD scores. Taken together, current research demonstrated the potential of profiling plasma EVs in searching for therapeutic targets as well as diagnostic signatures.

DEDICATION

To my parents, my wife, my kids, my brother and sister, and the rest of my family. Also to my high school biology teacher who motivated me to a career in biological science, as well as to all my friends who have been a constant source of support that kept me going.

ACKNOWLEDGMENTS

I give my sincere appreciation to all people who have offered supports through my entire graduate study.

Dr. David Spetzler, my PhD advisor, introduced me to the wonderful scientific research and offered tremendous guidance, encouragement and support during my graduated student life.

My graduate committee members, Dr. Hao Yan, Dr. Douglas Lake, Dr. Marco Mangone, kindly supported me in my research through years.

I really enjoyed the graduate life with all fellow students in ASU and research scientists, fellow colleagues in Caris Life Sciences. This includes Kasuen Kotagama in Dr. Mangone's lab and Dr. Valeriy Domenyuk, Dr. Matthew Rosenow, Dr. Vaishali Pannu, Adam Stark, Dr. Xianghua Liu, Melissa Richard, Janet Duncan, Gerri Ortiz, Stephen Logie, Dr. Xixi Wei, Radhika Santhanam, Dr. Xiaowei Liu, Dr. Jelena Zarkovic, Symon Levenberg, Jim Abraham, Dr. Heather O'Neill, Dr. Tassilo Hornung, Dr. Ryan Wang, Dr. Qing Zhang, Dr. Nick Xiao as well as my other colleagues from Caris Life Science.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Significance of Extracellular Vesicles

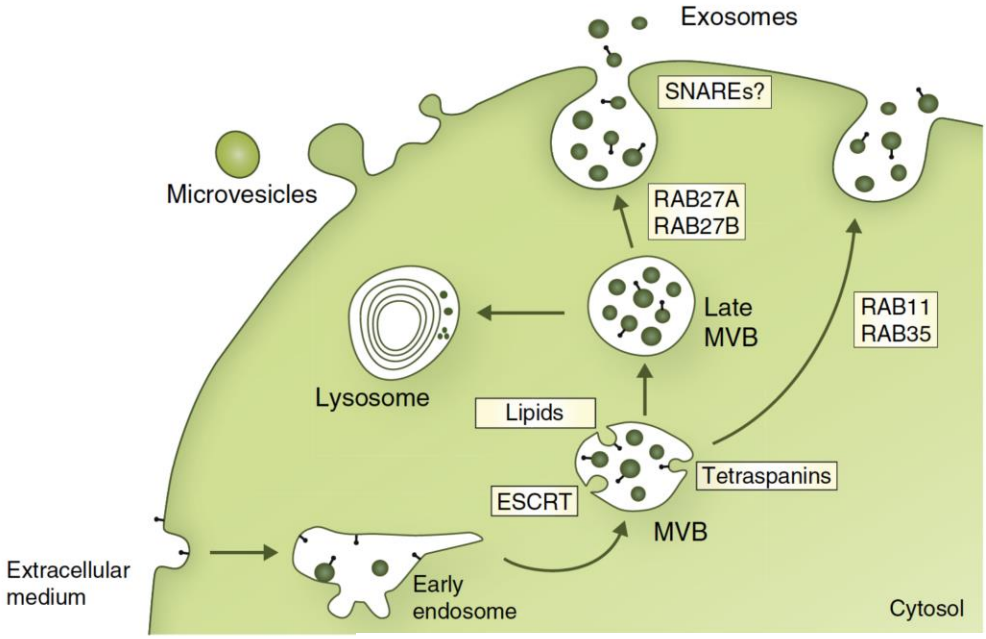
Extracellular Vesicles (EVs)/exosomes are nanosized membrane vesicles released by fusion of an organelle of the endocytic pathway, the multi-vesicular body (MVB), with the plasma membrane. This process was discovered more than 30 years ago. During these years, EVs have gone from being considered cellular waste to a novel mechanism of mediating cell-to-cell communication[1]. The exponential interest in EVs during recent years is due to their important roles in health and disease as well as to their potential clinical applications for therapeutic and diagnostic[1].

EVs were first discovered by Peter Wolf, who referred these vesicles as "platelet dust"[2] and later on were considered as "garbage cans"[3] as well as "cellular waste"[4-5]. Since then, all biological fluids tested have been shown to contain vesicles. In vitro grown cell lines have been shown to release vesicles to different extents[1, 6-7]. Two major types of vesicles have been described based on their mechanisms of release and size: EVs/exosome (less than 150 nm in diameter) and microvesicle particles (MPs). MPs are released directly from the plasma membrane (PM), which include microvesicles (MVs), apoptotic bodies and apoptotic microparticles are typically small particles ranging from 50nm to 3μm in diameter[8]. Secreted MPs derived from MVBs fused with the plasma membrane are termed EVs, while MPs released from the surface of plasma membranes are referred to as MVs, membrane particles, apoptotic vesicles[8]. Lately, EVs are to play a role not only in intercellular communication but also numerous physiological and

pathological functions[1, 6, 9-10]. EVs from cancer cells have been shown to promote angiogenesis, modulate the immune system and remodel the surrounding parenchymal tissue to support tumor progression. In particular, EVs have been shown to participate in the generation of the pre-metastatic niche[11-13], suggesting they might contain critical information related to aggressive.

EV biogenesis starts within the endosomal system. Figure 1.1.1 shows the schematic representation of the origin and release of EVs by eukaryotic cells[14]. EVs are formed as intraluminal vesicles (ILVs) by budding into early endosomes and MVBs. Several molecules are involved in the biogenesis of ILVs, such as the ESCRT machinery, lipids (e.g. ceramide) and tetraspanins proteins[14]. It is still unknown whether these mechanisms act simultaneously on the same MVB or on different MVBs. The fate of MVBs can be either fusion with lysosomes or fusion with the PM, which allows the release of their content into the extracellular milieu. Several RAB proteins (RAB11, RAB27 and RAB35) have been shown to be involved in the transport of MVBs to the PM and in EV secretion[14].

Figure 1.1.1 Biogenesis of EVs[14].



Curr Opin Cell Biol 2014, 29, 116-25

EVs contain proteins, lipids and nucleic acids. Their compositions to some extent are cell-type dependent and can be influenced by different cellular conditions or treatments[1]. Moreover, EVs released by cell lines are quite heterogeneous[1, 15], but less is known about how the cargos are sorted into the vesicles. Proteomic studies revealed that EVs may contain a specific subset of proteins from endosomes, the PM and the cytosol[14]. In contrast, proteins from intracellular organelles such as the nucleus, mitochondria, and Golgi are or under-represented in EVs[14]. Additionally, EVs contain DNA bound by nuclear proteins[16]; mitochondrial-derived vesicles (MDVs) are also able to fused with the MVBs[17]. Further, HSP60, a mitochondrial heat-shock protein, was found in tumor cell plasma membrane, EV membrane as well as in the Golgi apparatus[18]. These all suggest that though other cellular compartments that may not be directly involved in the EV biogenesis, the constant communication between these compartments and MVBs results in these proteins being found in MVBs or EVs.

Observations of protein, RNA and lipid composition in numerous studies have been made accessible through the creation of the EV protein, RNA and lipid database ExoCarta[19]. Although EVs should be enriched in endosomal components, as compared to MPs, the two types of vesicles display a large overlap in composition, and it is not possible today to name universal (i.e. valid for any cell type) protein markers specific for EVs versus MPs[1].

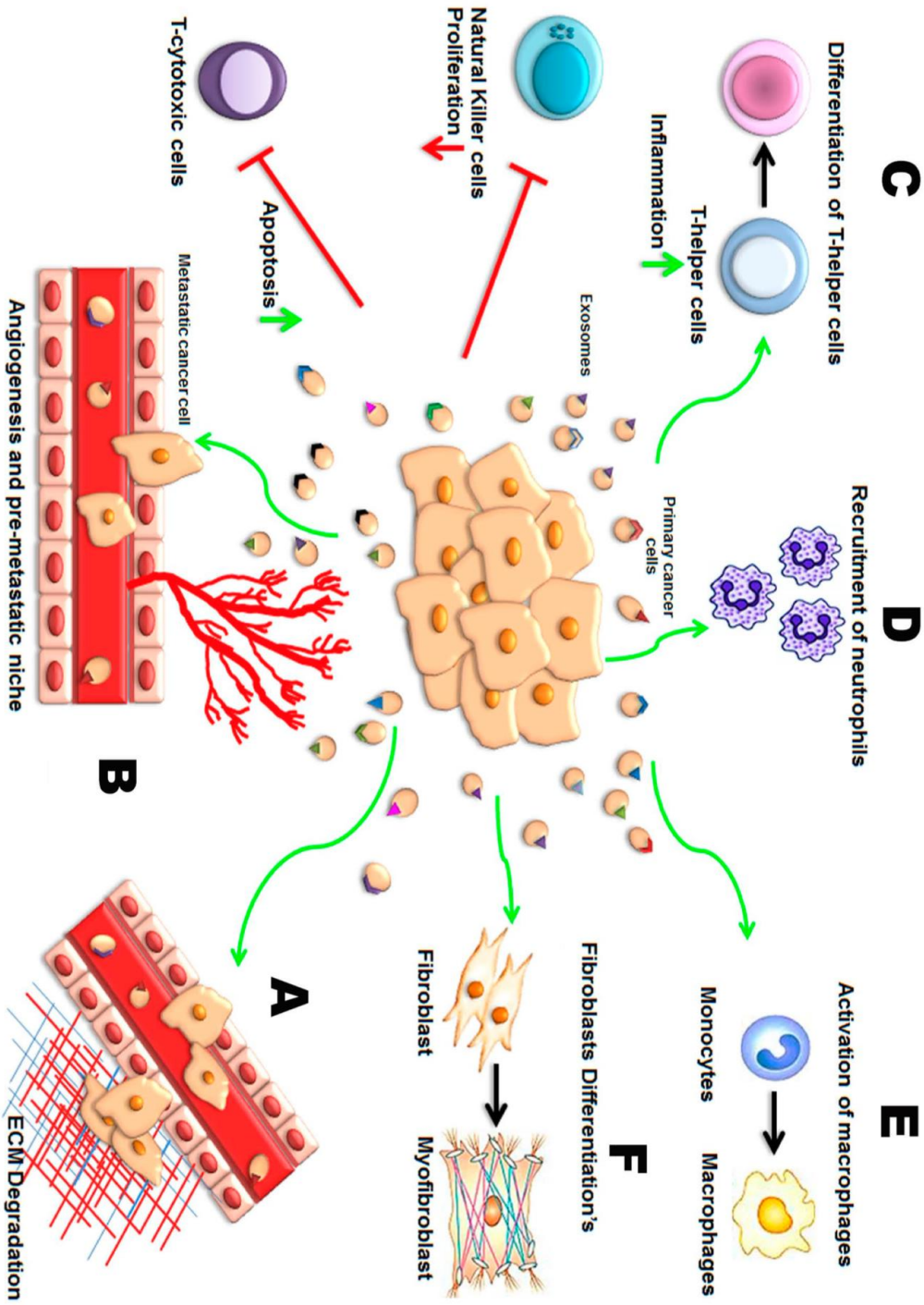
Interestingly, emerging studies are now showing EV release may be a highly regulated process instead of a random process for the removal of unwanted internal cellular debris as originally thought[20]. MVBs can either be directed to lysosomes where their content is degraded, or transported to the plasma membrane for EV release[1]. The

molecular mechanisms and the cellular states that regulate this balance are yet to be elucidated. Transporting MVBs to the plasma membrane depends on their interaction with actin and the microtubule cytoskeleton[1, 21-22]. Knock-down or over-expression of the actin binding protein cortactin has been shown to decrease or increase EV release respectively[23]; live-cell imaging indicated that cortactin is involved in both trafficking and docking of MVBs to the plasma membrane[23]. Treatment with paclitaxel (PTX), a chemotherapy that stabilizes microtubules, of breast cancer cell MDA-MB-231, resulted in elevated amount of EVs that are highly enriched with the cell survival protein and cancer markers[24], implying stabilization of microtubule network may be enhancing EV transportation and release. Moreover, Rab GTPases, the largest family of small GTPases[25], regulate many steps of membrane trafficking, including vesicle budding, transport of vesicles along actin and tubulin, as well as membrane fusion. Interestingly, several Rab GTPases have been shown to play a role in EV secretion, although their precise mechanism of action in this process is not yet known[1].

EVs located in biological fluids attract enormous interest because of their potential use as a source of protein and nucleic acid biomarkers, or as a prospective delivery system for various therapeutic purposes[26]. The study of EVs is opening new horizons for their potential application not only as therapeutic tools but also as clinical biomarkers for monitoring disease progression[27]. Increased levels of EVs have been detected in the body fluid relating to a varieties of diseases, including cardiovascular and inflammatory pathologies, obesity, atherosclerosis, diabetes, metabolic syndrome, neurodegenerative diseases and multiple sclerosis[27]. In term of cancer biology, one of the hallmarks of cancer cells is that they react to their micro-environment and can

communicate and exchange information to the host immune system. This occurs through the secretion of growth factors, cytokines, chemokines, small molecular mediators as well as EVs[28], which may be a package of all the above. As cell-to-cell messenger mediators of communication, EVs could have profound effects on cancer development such as invasion, metastasis, angiogenesis, and responses to hypoxia as well as immunomodulation (Figure 1.1.2). Therefore, profiling the EVs from biological fluids might lead to identification of new target with both therapeutic and diagnostic potential.

Figure 1.1.2 EVs in Cancer Progression[28]. EV released from tumor cells affect the local tumor microenvironment and are critically involved in tumor initiation, growth, progression, and metastasis by transferring oncogenic proteins and nucleic acids. (A) EVs travel to distant sites to promote generation of the pre-metastatic niche; (B) Angiogenesis is increased and endothelial and stromal cell differentiation is induced, leading to a pro-tumor environment; (C) EVs have immunosuppressive effects and assist cancers in immune evasion. Cytotoxic T cells induced apoptosis, while natural killer cell proliferation is impaired, and T-helper cells differentiate toward a T-regulatory cell phenotype; (D) Bone marrow-derived cells are recruited to tumor and pre-tumor tissue where they contribute to cancer development; (E) EVs are also responsible for the recruitment and activation of tumor-associated macrophages (TAMs) by promoting their polarization. TAMs support diverse phenotypes within the primary tumor, including growth, angiogenesis, and invasion, by secreting a plethora of pro-tumorigenic proteases, cytokines, and growth factors; (F) EVs can functionally modify fibroblasts by reprogramming these cells to cancer-associated fibroblasts (CAFs), which exhibit myofibroblastic differentiation. *Red arrows indicate a negative contribution or repression and green arrows indicate an activation or positive function.


C
Differentiation of T-helper cells
T-helper cells
Inflammation
Natural Killer cells
Proliferation
T-cytotoxic cells
Apoptosis
Exosomes
Primary cancer cells
D
Recruitment of neutrophils
E
Activation of macrophages
Monocytes
Macrophages
F
Fibroblasts Differentiation's
Fibroblast
Myofibroblast
Metastatic cancer cell
B
Angiogenesis and pre-metastatic niche
A
ECM Degradation

1.2 Current EV Enrichment Methods

Varieties of EV enrichment methods are currently available, each with advantages and disadvantages. Importantly, it is currently technically challenging to obtain a pure EV fraction free from non-vesicular components for functional studies[29]. Therefore, each methology may only recover a subset of the EV population. The choice of which method to use is highly dependent on the expected abundance of EV, sample amount, scale of the study, EV purity requirments, potential contamination, as well as complexity of the protocol. Restrictions and limitations imposed by these considerations could resulted in different profiles which may be observed in downstream applications.

In cell line EV release studies, EVs have often been isolated by sequential centrifugation. This method includes 2–3 centrifugations at low speeds to remove cells, cell debris and MPs, followed by ultracentrifugation at approximately 100,000g for hours with multiple PBS washes[1]. The principle of this method implies a mixed EVs population will be collected since the size of EVs and MPs are partially overlapped. When combined with density gradient, it does provid the most pure EVs to date[30-32]. However, this comes with a cost of extremely low recovery efficiency and reproducibility[30]. In additon, it is relatively tedious procedures, time consuming and technique sensitive, requiring prior training and the availability of an ultracentrifuge[30]. However, when the purpose of the study is to enrich the EVs generated by cell lines, this may be acceptable because the source is almost unlimited and the purity requirement of such EVs is critical for the follow-up functional studies. In contrast, UC for large scale clinical studies on body fluid EVs, where sample amount and source limited. In this case, reproducibility and of the procedure priority, and minor contamination be acceptable.

The term "EVs" is used to refer to the pellet obtained after 100,000g ultracentrifugation[1]. Although this pellet is enriched in EVs, it also contains small MPs as well as protein aggregates. In fact, it was recently suggested that this pellet should be referred to as small EVs (sEVs) instead of EVs, as it has been shown to contain different vesicle populations[1, 15]. Therefore, this pellet not only reflects changes in EV release, but also in small MPs or other similar density elements. Another concern that is associated with UC is that particle size distribution curves from UC EV preparations shifted slightly to the right (larger) as compared to the other commercial kits. The cause could be the centrifugation at such high speeds potentially led to fusion of the particles with contaminants and other proteins, therefore affecting the physical properties of the EVs and impacting the sensitivity of proteomic analysis[30, 33-35].

Immunoisolation of EVs based on specific proteins at the EV membrane is another method that results a in relatively pure EV sub-populations[1, 15, 36]. However, it requires that the selected membrane protein is present in the EVs and that the immunoisolation protocol is optimized. A list of identified cell line EV surface markers is shown in Table 1.2.1[1] and the protein families presented in EVs from different cell types is shown in Table 1.2.2[37]. The most common proteins that could be found in EVs includes tetraspanin proteins (CD9, CD63, CD81). In the case of tumor, epithelial cell adhesion molecules (EpCAM), IGF-1R α units (α-IGF-1R), CA125, CD41b, and E-cadherin[38] are also found in EVs.

Unfortunately, immunisolation methods have not been very successful due to the the heterogeneity of the EVs popultation. EVs released by a cell line are already quite heterogeneous[15]. Many studies have described the proteins, lipids and RNA cargos of

EVs, but have not shown how the cargos are sorted into the vesicles[1]. The proteins in Table 1.2.1 and Table 1.2.2 are rather a list of potential indicators for particular sub-population than univeral EV markers. Though varieties of proteins markers have been shown associated with EV, none of these proteins are universal maker for any EV populations. Thus the immunoisolation method may only be suitable in the situation where the EVs pupulation is closely related to the souce cells with known surface markers. For samples like plasma, EVs may be derived from diverse sources, suggesting that immunoisolation would only recover a small portion from all the potential EVs, and may not suitable for the study of the whole EV populations.

Table 1.2.1 Cell line EV Surface Markers[1]

Protein/lipid/modification	Cell line	Secretion quantified by
Hrs	DCs, HEK293, SCC61, SCC25-H1047R HNSCC, HeLa	Ubiquitinated proteins, TSG101, VPS4B, Evi, Wnt3A (WB), NTA, CD81/HLA-DR (FC), MHC II, HSC70, CD63 (WB)
STAM1	HeLa	CD81/HLA-DR (FC), MHC II, HSC70, CD63 (WB)
TSG101	HeLa	CD81/HLA-DR (FC), MHC II, HSC70, CD63 (WB)
CHMP4C	HeLa	CD81/HLA-DR (FC)
ALIX	HeLa, DCs, MCF-7	CD81/HLA-DR (FC), CD63, HSP70, syndecan (WB), NTA
VTA1	HeLa	CD81/HLA-DR (FC)
VPS4	HeLa	CD81/HLA-DR (FC), MHC II, HSC70, CD63 (WB)
Syntenin	MCF-7	CD63, HSP70, syndecan (WB), NTA
Syndecan	MCF-7	CD63, HSP70, ALIX, syntenin (WB), NTA
CD9	HEK293, BMDCs	β-Catenin, flotillin-1 (WB)
CD82	HEK293	β-Catenin (WB)
CD63	HEK293	NTA
LMP1	HEK293	NTA, acetylcholinesterase activity, Alix, HSC70, CD63, and TSG101 (WB)
Tspan8	Adenocarcinoma cells	mRNA microarray, RT-qPCR, proteomics, WB
Synaptotagmin-7	SCC61, SCC25-H1047R HNSCC	NTA
VAMP7	K562	Acetylcholinesterase activity
YKT6	HEK293 and A549	TSG101, WNT3A and VPS26/35 (WB)
Syntaxin 1A	*Drosophila* S2	Evi (WB)
PKM2	A549, HeLa	NTA
SNAP-23	A549	NTA
RalA and RalB	4T1	EM, ALIX, CD63, HSC70 and TSG101 (WB)
Rab2b	HeLa	HLA-DR (FACS-assay)
Rab5a	HeLa	HLA-DR (FACS-assay)
Rab9a	HeLa	HLA-DR (FACS-assay)
Rab7	MCF-7, HUVEC	CD63, syntenin and syndecan (WB), miR-143 (qPCR)
Rab11	K562, *Drosophila* S2	Transferrin receptor, Lyn, HSC70 and Evi (WB)
Rab27a	HeLa, 4T1, TS/A, B-16-F-10, SK-Mel-28, SCC61, SCC25-H1047R HNSCC, Du145	HLA-DR (FACS-assay), total protein, HLA-DR, HSC70, TSG101, CD63, ALIX and LAMP2 (WB), NTA, CD9 (ELISA)
Rab27b	HeLa, HUVEC	HLA-DR (FACS-assay), total protein, HLA-DR, HSC70 and TSG101 (WB), miR-143 (qPCR)
Rab35	Oli-neu	PLP (WB)
Citron kinase	HeLa, 293T	HSC70, CD82, Lamp-1 (WB)
Cortactin	SCC61	NTA, TSG101, CD63 and flotillin-1 (WB)
ISGylation	Jurkat T and HEK293	CD63, CD81, TSG101 and flotillin (WB), NTA
SIMPLE	COS	Fluorescence of LactC2-RFP, NTA, CD63, ALIX (WB)
nSMase2	Oli-neu, PC-3, HEK293, J77	PLP, EGFP-CD63 (WB), miR-16, miR-146a (qPCR), total protein, CD81 (WB)
DGKα	J-HM1-2.2	CD63, Lamp-1, FasL, (WB)
PLD2	RBL-2H3, MCF-7	Bodipy-ceramide label (FACS), syntenin, ALIX, CD63 (WB)
ARF6	MCF-7	Syntenin, ALIX, CD63, SDC1CTF (WB)
Cholesterol	Oli-neu, PC-3	Flotillin-2, ALIX, EGFP-CD63, PLP-myc, caveolin-1, Lamp-1 (WB)
Ether lipid (hexadecylglycerol)	PC-3	NTA, total protein
V-ATPase	HeLa	EM, CD63, ALIX, TSG101 (WB)
Tetherin	HeLa	EM, CD63, ALIX, TSG101 (WB)
Hypoxia	MCF-7, SKBR3, MDA-MB 231	NTA, CD63 (WB)
Irradiation	LNCaP, 22Rv1, PBMC	Vybrant DiI (fluorescent staining), B7-H3 (WB), NTA, total protein
Cisplatin	A549	Total protein
PIKfyve	PC-3	NTA, total protein, MS-proteomics
ER stress (tunicamycin)	MEFs	qNano
Autophagy (starvation)	K562	Acetylcholinesterase activity, HSC70 (WB)
ATG12-ATG3	MEFs	Total protein, ALIX, TSG101, GAPDH, HSC70 (WB)
Autophagy (ATG7)	DCs	GAPDH (WB)
Intracellular calcium (monensin, A23187, ionomycin)	K562, oligodendrocytes	HSC70 (WB), acetylcholinesterase activity, PLP (WB)

Cell Mol Life Sci 2018, 75 (2), 193-208

Table 1.2.2 EV Surface Marker from Different Cell Types[37]

Exosome protein	Cell type
Antigen presentation	
MHC class I	B cells Dendritic cells Enterocytes Tumours T cells
MHC class II	B cells Dendritic cells Enterocytes (IFN-γ-treated) Mastocytes T cells
Integrins	
α4β1	Reticulocytes
αMβ2	Dendritic cells
β2	T cells
αLβ2	Mastocytes
Immunoglobulin-family members	
ICAM1/CD54	B cells Dendritic cells Mastocytes
P-selectin	Platelets
A33 antigen	Enterocytes
Cell-surface peptidases	
Dipeptidylpeptidase IV/CD26	Enterocytes
Aminopeptidase n/CD13	Mastocytes
Tetraspanins	
CD63	B cells Dendritic cells Enterocytes Platelets T cells Mastocytes
CD37, CD53, CD81, CD82	B cells
CD9	Dendritic cells
Heat-shock proteins	
HSC70	Reticulocytes Dendritic cells Tumours
HSP84/90	Dendritic cells Enterocytes
Cytoskeletal proteins	
Actin	Dendritic cells Enterocytes Mastocytes
Actin-binding proteins (cofilin)	Dendritic cells
Tubulin	Dendritic cells Enterocytes
Membrane transport and fusion	
Annexins I, II, IV, V, VI	Dendritic cells
Annexin VI	Mastocytes
RAB7/RAP1B/RABGDI	Dendritic cells
Signal transduction	
Gi2α/14-3-3	Dendritic cells
CBL/LCK	T cells
Metabolic enzymes	
Enolase-1	Enterocytes
Thioredoxine peroxidase	Dendritic cells

Nat Rev Immunol 2002, 2 (8), 569-79

Recently, commercial kits have been made available for EV isolation using precipitation via PEG polymer, including Exoquick and Total Exosome Isolation (TEI). PEG has been used in a range of protein applications utilizing volume exclusion mechansim[39]. Once the sample is mixed with the reagent, the solubility of the proteins is lowered and results in the precipitation of EVs. Using these kits, EVs can be separated at a low g force and with higher yield as well as improved reproducibility relative to UC[38]. The precipitation is not dependent on EV sub-populations is a simpler procedure, has better yield and reproducibility, which makes made this method attractive for EV studies of complex samples like plasma.

However, PEG, especially PEG8000, which is included in these commercial kits, is well-known for precipitating IgG, fibrinogens, and components of complement system in the plasma, and thus has limited selectivity[40-41]. Moreover, the use of linear, long PEG polymer results in much high viscosity[41], which compromise some downstream applications, such as electro-microscopy, mass spectrometry and flow cytometry unless tedious procedures are adopted to remove the polymer after the precipitation.

Other than the UC, immunoisolation and precipitation based EV enrichment methods, other techniques, such a membrane filtration, and size exlusion chromatography are available. A comparison of advantages and disadvantages of different methods is summarized in Table 1.2.3 (https://www.labome.com/method/Exosomes-isolation-and-characterization-methods-and-specific-markers.html). Previous studies have already compared the reproducibility and efficiency of enriching plasma-derived EVs between UC and PEG-based methods[30-32]. Though results might vary from study to study, the conclusion in general was that UC may obtain EV with relatively better purity but with

very low efficiency. In contrast, PEG-based method offers higher efficiency and better reproducibility, but at a cost of significant contamination of high abundant plasma proteins. Therefore, though EV enrichment method based on precipitation is more preferred for the body fluid study like plasma, increasing the precipitation specificity and reducing the recovery of high abundant plasma proteins may be necessary to improve its reliability.

Table 1.2.3 Advantages and Disadvantages of Different EV Isolation Methods

Methods	Mechanism	Advantages	Disadvantages
Differential Centrifugation	The method consists of several centrifugation steps aiming to remove cells, large vesicles and debris and precipitate EVs	Differential centrifugation is the standard and very common method used to isolate EVs from biological fluids and media	The efficiency of the method is lower when viscous biological fluids such as plasma and serum are used for analysis
Density gradient centrifugation	This method combines ultracentrifugation with sucrose density gradient.	The method allows separation of the low-density EVs from other vesicles, particles and contaminants	Very high sensitivity to the centrifugation time
Size Exclusion Chromatography	Size-exclusion chromatography separates macromolecules on the base of their size. It applies a column packed with porous polymeric beads.	The method allows precise separation of large and small molecules and application of various solutions. Compared to centrifugation methods, the structure of EVs isolated by chromatography is not affected by shearing force.	The method requires a long running time, which limits applications of chromatographical isolation for processing multiple biological samples.
Filtration	Ultrafiltration membranes are used to separate exosomes from proteins and other macromolecules. The EV population is concentrated on the membrane.	Filtration allows separation of small particles and soluble molecules from EVs. During the process the EV population is concentrated by the filtration membrane	EVs can adhere to the filtration membranes and become lost for the following analysis. Also, since the additional force is applied to pass the analyzed liquid through the membranes, the EVs can potentially be deformed or damaged.
Polymer-based precipitation	The technique includes mixing the biological fluid with polymer-containing precipitation solution, incubation step and centrifugation at low speed	The advantages of precipitation include the mild effect on isolated EVs and usage of neutral pH.	Polymer-based precipitation methods co-isolate non-vesicular contaminants, including lipoproteins. Also, the presence of the polymer material may not be compatible with down-stream analysis.
Immuno-isolation	Various immunological methods are applied. Magnetic beads bound to the specific antibodies are used to isolate EVs. Also, ELISA-based separation method was developed.	The method allows isolation of selective subtypes of exosomes. Also, it may be applied for characterization and quantitation of EVs proteins.	The method is not applicable for large sample volumes. Also, the isolated vesicles may lose the functional activity.
Isolation by sieving	This technique isolates EVs by sieving them via a membrane and performing filtration by pressure or electrophoresis.	Relatively short separation time and gives high purity of isolated EVs.	Low recovery of isolated EVs

1.3 Breast Cancer and Breast Cancer Aggression

Breast cancer is the most common noncutaneous cancer in U.S. women, with an estimated 61,000 cases of in situ disease and 246,660 cases of invasive disease in 2016. One out of six women diagnosed with breast cancer is estimated to die of the disease. Population studies from the United States and the United Kingdom also demonstrated an increase in DCIS and invasive breast cancer incidence since the 1970s, attributable to the widespread adoption of both postmenopausal hormone therapy and screening mammography (https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0032676/).

Breast cancer is a heterogeneous disease and several biologic subtypes have been identified as conventional clinical factors such as tumor grade, size, lymph node involvement, and surgical margins are no longer sufficient as sole prognostic factors[42]. Most (80%) breast cancers are invasive, or infiltrating, which means they have broken through the walls of the glands or ducts where they originated and grown into surrounding breast tissue. Although breast cancer generally has been referred to as a single disease, there are up to 21 distinct histological subtypes and at least four different molecular subtypes that differ in terms of risk factors, presentation, response to treatment, and outcomes[43]. Breast cancers are generally divided into subtypes according to the status of their estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) as well as the origins of cancer epithelail cells, including luminal epithelial (types A and B, non-basal like) and basal epithelial (basal-like)[44]. Luminal A is the most common breast cancer subtype and is characterized by ER+ and/or PR+/HER2- status, low grade tumors, and good prognosis[45-47]. Luminal B subtype accounts for roughly 10% of all breast cancer and is distinguished by ER+ and/or

PR+/HER2+ status[42]. Breast cancer subtypes with negative ER, PR, and HER2 status are typically called "triple negative" breast cancers and approximate the basal-like category[42].

The prognosis of invasive breast cancer is strongly influenced by the stage of the disease – that is, the extent or spread of the cancer when it is first diagnosed[43]. There are two main staging systems for cancer[43]. The TNM classification of tumors uses information on tumor size and how far it has spread within the breast and to adjacent tissues (T), the extent of spread to the nearby lymph nodes (N), and the presence or absence of distant metastases (spread to distant organs) (M). Once the T, N, and M are determined, a stage of 0, I, II, III, or IV is assigned[43, 48], with stage 0 being in situ (abnormal cells have not penetrated the ducts or glands from which they originated), stage I being early-stage cancer, and stage IV being the most advanced disease. The overall 5-year relative survival rate is 99% for localized disease, 85% for regional disease, and 27% for distant-stage disease[43]. Therefore, understanding how breast cancer cells are able to acquire metastasis capability might be able to increase the survival rate for the disease.

Many factors may contribute to the risk of transforming locally contained breast cancer cells to a more invasive form, such as age, tumor types/subtypes as well as genetic disposition. For example, among women under age 50, incidence of having invasive breast cancer rates have slowly increased (0.2% per year) since the mid-1990s[43]; lobular carcinoma in situ (LCIS), whose cancer cells were developed in milk producing glands poses an increased risk for developing invasive cancer compared to the ductal carcinoma in situ (DCIS) that arised in epithelial cells lining the breast ducts[49-50]; women with

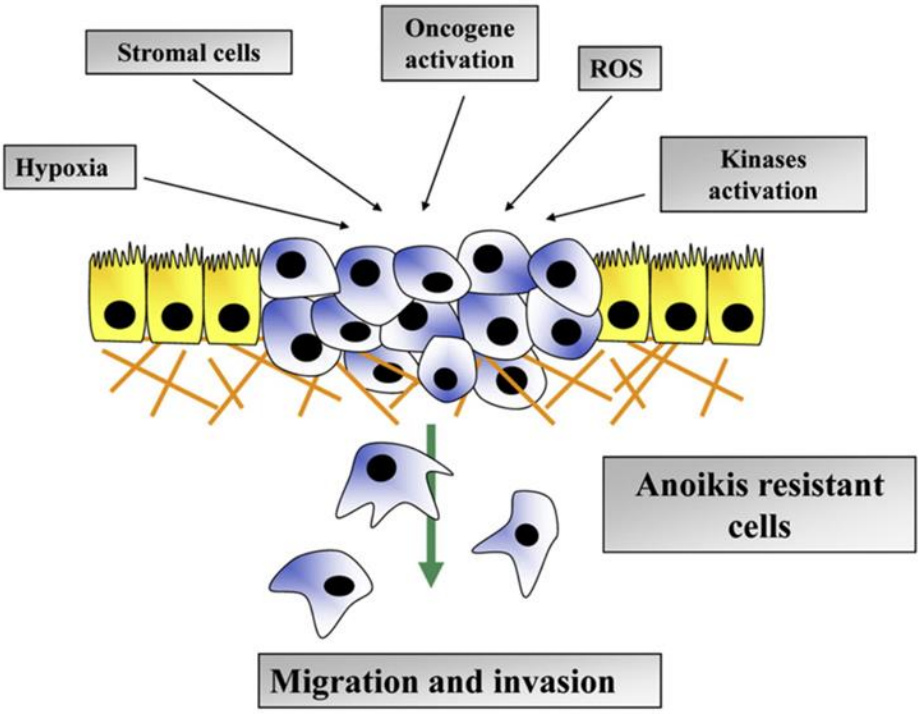
HER2-enriched as well luminal B tumors (HR 1.96, 95% CI: 1.08-3.54) had roughly a two-fold increased adjusted risk of breast cancer mortality[42]. Breast cancers in BRCA1 mutation carriers are associated with more aggressive tumor characteristics compared to BRCA2 and are more difficult to detect using mammography[51].

Escalation of invasive carcinoma begins with the epithelial mesenchymal transition (EMT), a physiological process that allows epithelial cells to remodel the cytoskeleton, release the linkage with vicinal cells and acquire a motile phenotype[52-53]. This phenomenon is usually activated during wound healing, inflammation or embryogenesis. In cancers, EMT has long been associated with the acquisition of malignant cell traits, such as motility, invasiveness and a heightened resistance to various cytotoxic therapies for cancer cells[53], allowing cancer cells to detach from neighboring cells, overcome anoikis and to move from their primary location to invade others tissues[52]. During EMT cancer cells activate epigenetic pathways that lead to the down-regulation of cell–cell adhesion molecules, such as E-cadherins and γ-catenin. At the same time, the expression of mesenchymal markers such as vimentin, fibronectin, α-smooth muscle actin (SMA), N-cadherin, as well as to the activation of MMPs, were increased[52, 54]. It is known that the ability to overcome anoikis is correlated with the acquisition of the mesenchymal phenotype[52]. Otherwise, detaching from the extracellular matrix (ECM) would triggers the apoptotic process for normal epithelial cells[54]. Such ability of cancer cells to survive in the absence of normal matrix components represents an important property for cells undergoing EMT[54]. As shown in Figure 1.3.1**,** during EMT, cancer cells started to break through the local growth barrier. Once they have

acquired the capability to circumvent anoikis, cancer cells are able to migrate outside of the orignal tumor location[52].

Since EVs from tumor cells resies in the micro enviroment around the tumor, there is a increased likelihood of secretion into the plasma. Though invasive behavior of breast cancer may be observed in stage II, it is more frequently found in stage III/IV (https://www.cancer.net/cancer-types/breast-cancer/stages). Thus, the likelihood of detecting EVs that contains information related to the cancer in the plasma, could be higher, profiling of plasma EVs at that stages may lead to discovery of molecules with diagnostic/therapeutic information.

Figure 1.3.1 Cancer Epithelial-Mesenchymal Transition[52]. Hypoxia, stromal cells, oncogene activation, ROS as well as kinase activation drive the cancer cells to a more aggressive, invasive mode. During EMT, under these conditions, cells started to break the local growth barrier, once anoikis resistance is acquired, cancer cells are able to migrate outside of the primary tumor location.



Biochim Biophys Acta 2013, 1833 (12), 3481-3498

1.4 Current Breast Cancer Diagnosis and Limitation

Breast imaging, including mammography, ultrasound and magnetic resonance imaging (MRI) are the most common breast cancer diagnosis tools. Mammography is currently the most common method of early breast cancer detection since it can often identify cancer several years before physical symptoms develop. MRI and ultrasound detection were also reported in conjunction with mammography to improved the detection performance[55-57].

The American Cancer Society (ACS) recommendations for the early detection of breast cancer screening vary depending on a woman's age and include mammography, as well as MRI for women at high risk. In 2015, the American Cancer Society updated its breast cancer screening guideline for average-risk women[58]. Screening mammography in women aged 40 to 69 years was associated with a reduction in breast cancer deaths across a range of study designs, and inferential evidence supported breast cancer screening for women 70 years and older who were in good health[58]. The ACS recommendation[58] for women with an average risk of breast cancer is that regular screening mammography should be started at age 45 years (strong recommendation); Women aged 45 to 54 years should be screened annually (qualified recommendation); Women 55 years and older should transition to biennial screening or have the opportunity to continue screening annually (qualified recommendation); Women should have the opportunity to begin annual screening between the ages of 40 and 44 years (qualified recommendation). Women should continue screening mammography as long as their overall health is good and they have a life expectancy of 10 years or longer (qualified recommendation)[58].

An expert panel convened by the ACS published recommendations for the use of MRI for screening women at increased risk for breast cancer in 2007[56]. The panel recommended annual MRI screening in addition to mammography for women at high lifetime risk (~20%-25% or greater) beginning at 30 years of age[56]. Breast MRI is not recommended as a routine screening tool for all women. However, it is recommended for screening women who are at high risk for breast cancer, usually due to a strong family history and/or a mutation in genes such as BRCA1 or BRCA2.

Ultrasound is an imaging test that sends high-frequency sound waves through the breast and converts them into images on a viewing screen. Ultrasound is an attractive supplement to mammography because it is widely available, relatively inexpensive and well-tolerated by patients[59]. Ultrasound images the tissue to the chest wall and is a cross-sectional technique, displaying tissue without overlap involved[59]. Under age of 30, the doctor may recommend ultrasound before mammography to evaluate a palpable breast lump (a breast lump that can be felt through the skin) because mammograms can be difficult to interpret in young women because their breasts tend to be dense and full of milk glands (http://www.breastcancer.org/symptoms/testing/types/ultrasound). Breast cancer detection doubled from 23 to 46 in 6,425 subjects using automated whole breast ultrasound (AWBU) with mammography, resulting in an increase in diagnostic yield from 3.6 per 1,000 with mammography alone to 7.2 per 1,000 by adding AWBU[59].

Despite the sensitivity of these technologies based on advanced imaging systems, there are fundamental limitations that exist. For mammography, there is a high false-positive rate, leading to over-diagnosis and includes radiation exposure. Mammography often results in false positive leading to biopsies, when there is no cancer[60]. When

screening begins at age 40, the cumulative probability of a woman receiving at least one false-positive result after 10 years was 61.3% (95% CI, 59.4% to 63.1%) with annual and 41.6% (CI, 40.6% to 42.5%) with biennial screening[60]. On average, 1 in 9 women are recalled from each screening examination for further testing, but most (95%) do not have cancer[61]. In addition, over the course of 10 screening examinations, about one-half of women experience a false positive, and about 19% undergo biopsy but do not have cancer[55]. Mammography also likely results in some over-diagnosis, the diagnosis of cancer that would not cause a woman any harm in her lifetime and that would not have progressed or otherwise been detected in the absence of screening[43]. Estimates of the prevalence of over-diagnosis are highly variable, ranging from <5% to more than 30%[62-66]. Additionally, repeated exposure to radiation from mammography could increase the risk of radiation-induced cancer, especially for the BRCA mutation carriers that are recommended to start mammographic screening for breast cancer as early as age 25-30 years[67]. A relative risk model suggested that there would be no net benefit from annual mammographic screening of BRCA mutation carriers at age 25-29 years, the net benefit would be zero or small at age 30-34 years, but there should be some net benefit at age 35 or older[67]. Exposure of the breast to ionizing radiation was predicted to cause breast cancer and cancer related death though the percentage is relatively low (~0.1%)[68]; Moreover, not all early breast cancer will be detected by mammogram, and some cancers that are screen-detected still have poor prognosis. Most women will never be diagnosed with breast cancer, but will undergo regular screening and may experience one or more "false alarms."[43].

On the other hand, breast MRI is also not a perfect tool. Although it is generally considered more sensitive for picking up breast cancer than mammography, it also can miss some cancers that would be detected by mammography. MRI screening is not recommended for women whose lifetime risk of breast cancer is less than 15%. Studies indicated that although MRI is underutilized among high-risk women, it is often used in women who are not at high risk for breast cancer[43, 69], which may be the reason why breast MRI is recommended only in combination with other tests, such as mammogram or ultrasound.

Ultrasound is not used on its own as a screening test for breast cancer. Rather, it is used to complement other screening tests. The performance of detecting breast cancer in ultrasound is comparable with mammography, with a greater proportion of invasive and node-negative cancers[57]. However, similar to mammography, false positive rates are more common using ultrasound detection. If an abnormality is seen on mammography or felt by physical exam, ultrasound is the best way to find out if the abnormality is solid (such as a benign fibroadenoma or cancer) or fluid-filled[57] (such as a benign cyst). It cannot determine whether a solid lump is cancerous, nor can it detect calcifications. Therefore, in addition to the currently available breast imaging tools, a non-invasive test could be desired in the detection of breast cancer and the overall screening performance could be improved. In contrast to the high-cost equipment required by the breast imaging, blood/plasma is routinely collected in most clinical setting. Therefore, a blood-based detecting approach for breast cancer based on the patient's biochemical changes would be able to increase the confidence in clinical decisions and decreases the stress and anxiety that every woman faces.

1.5 Aptamer-Based Profiling of Biological Samples

Aptamers are artificial nucleic acid ligands that can be generated against amino acids, drugs, proteins and other molecules. They are isolated from complex libraries of synthetic nucleic acid by an iterative process of adsorption, recovery and re-amplification. They have potential applications in analytical devices, including biosensors, and as therapeutic agents[70].

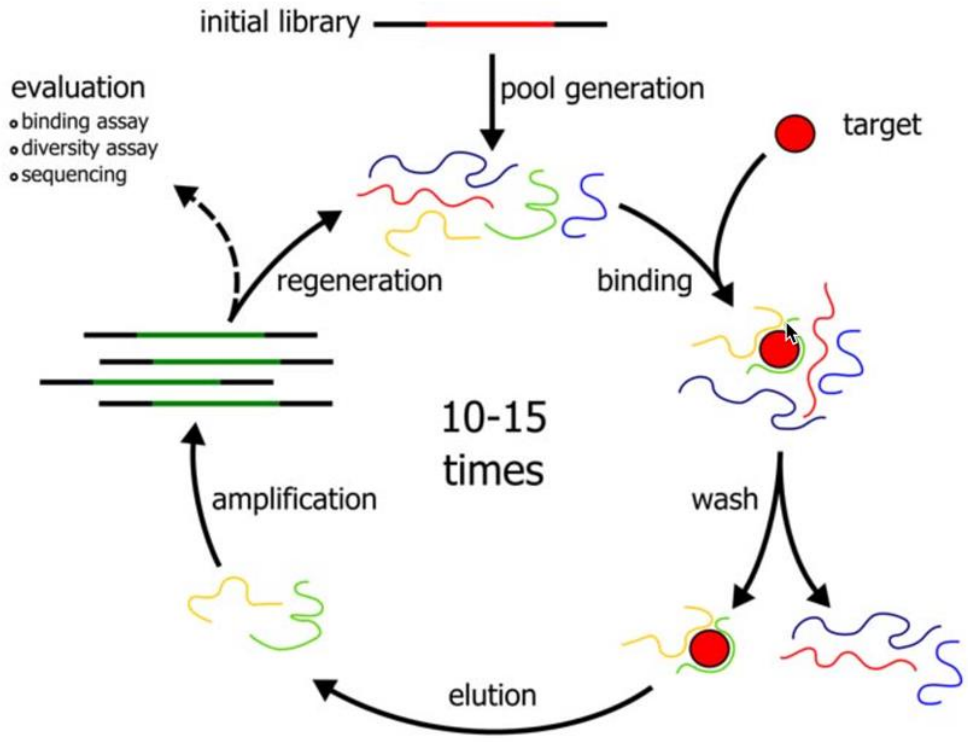
Systematic Evolution of Ligands by EXponential enrichment (SELEX) is an experimental procedure that allows extraction, from an initially random pool of oligonucleotides, of the oligomers with a desired binding affinity for a given molecular target[71]. The procedure can be used to infer the strongest binders for a given DNA or RNA binding protein, and the highest affinity binding sequences isolated through SELEX can have both diagnostic and therapeutic applications. For example, nucleic acid ligands against important targets in infectious, malignant and vascular diseases are continually being discovered by the process of in vitro evolution. Promising methods are being developed for the exploitation of these aptamers in biosensors and other diagnostic devices[72-76].

A large variety of molecules can be targets in SELEX experiments. When the library consists of dsDNA, targets are typically proteins that interact with DNA as a part of their natural function[77]. In the case of single stranded oligonucleotides, a wide range of proteins and small molecules have been used as targets in SELEX experiments[71]. The initial library of oligonucleotides typically consists of a large number (10^{15}-10^{16}) of random sequences. In principle, larger libraries of up to 10^{20} oligonucleotides are technically feasible[78], but are rarely used in practice. Each oligonucleotide consists of a

central region of random sequence, which is flanked by two regions of fixed sequence that enable amplification. The length of the random region is typically between 20 and 30 bp, while each flanking region is typically 15-25 bp long.

Figure 1.5.1 shows the standard SELEX procedure[79]. As indicated, SELEX generally started from a synthetic library, a more diversity-rich pool is first generated by PCR; the library is then mixed and incubated with the desired target; after wash, the aptamers that remain on the target are eluted and new library is regenerated by PCR amplification; the selection is then repeated 6-20 times, during which the enriched library could be characterized and individual aptamer could be chosen for further evaluation.

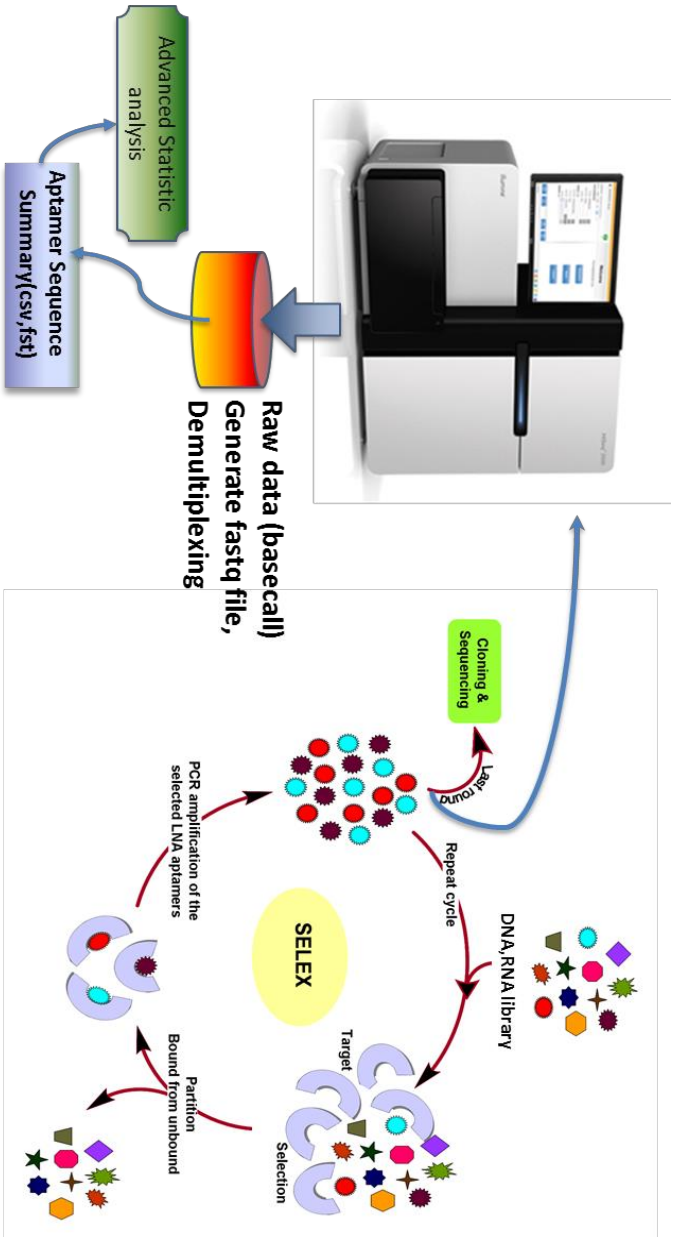
Figure 1.5.1 Standard SELEX Procedure.



With the latest developments and the increasing capacity of Next Generation Sequencing (NGS), instead of ranking the aptamer by random cloning, the enriched library can be sequenced as a whole with output as many as 300 million reads. The relative abundance of the individual aptamers could then be evaluated and ranked as shown in Figure 1.5.2. The advantage of SELEX-NGS combination is that limitation of quantification with traditional SELEX cloning procedure is overcome, which leads to the possibility of more accurate statistical analysis.

Interestingly, recent studies have also shown that cell-based SELEX and in vivo SELEX are both able to obtain specific aptamers which target the tumor cell surface[80-84], and even have identified brain-penetrating aptamers[85], implying that SELEX procedure might not necessarily to be restricted to single, pure targets. Instead, more complex, multi-targets sample could also be used, such as enriched plasma EVs from cancer patients.

Figure 1.5.2 SELEX-NGS Pipeline



1.6 Summary

EV analysis presents several advantages over the traditional analyses of soluble molecules in blood, such as hormones and cytokines[86]. One significant advantage is the inherent protection of the EV cargo of proteins and RNA from degradation. Otherwise, they would be rapidly degraded in blood[87-88]. Moreover, EVs have a relatively long half-life in blood[88]. Therefore, EVs can be transported from any location of the body to the bloodstream, making them easily accessible for analysis compared to biopsies[86]. This also has a significant advantage for the patient because the collection of a blood sample is a minimally invasive procedure associated with much less discomfort than a biopsy. It can often become an issue to detect relevant biomarkers because these diagnostic molecules frequently constitute a small part of the total amount of molecules in a blood sample[86]. With appropriate enrichment method, EV-associated proteins, which represented by only 0.01% of the plasma proteome[86, 89], can be identified. The EV cargos, biomarkers of otherwise undetectable amounts in the plasma, can now be analyzed and interpreted into clinically relevant information. Furthermore, EV cargo content can seemingly change according to the progression of a disease[89-93]. Therefore, EVs from a blood sample may render a sophisticated fingerprint of a disease of diagnostic and prognostic value.

This work is intended to examine whether enriched plasma EVs could be used for diagnostic application. Here we established and compared a nova plasma EV precipitation method based on pluronic block copolymer to the commercially available precipitation methods. The proteomic profiles of enriched plasma EVs from high grade breast cancer patients and the non-cancer patients were then compared and features

selected to generate a classifier between breast cancer patient and non-cancer controls. Cross-validation was used to validate the findings. To further validate our finding that EV associated proteins related to breast cancer was discovered from the cancer vs non-cancer study, one feature that significantly contributed to the model was chosen for further *in vitro* studies. In parallel, we also examined the possibility of differentiating breast cancer patients from controls by profiling the enriched plasma EVs through a SELEX-based profiling platform.

CHAPTER 2

MOLECULAR PROFILING PLASMA EXTRACELLULAR VESICLE UNVEILS FEATURES ASSOCIATED WITH BREAST CANCER AGGRESSION, METASTASIS AND INVASION

Abstract

Extracellular vesicle (EV) based liquid biopsies have been proposed to be a readily obtainable biological substrate recently for both profiling and diagnostics purposes. Development of a fast and reliable preparation protocol to enrich such small particles could accelerate the discovery of informative, disease-related biomarkers. Though multiple EV enrichment protocols are available, in terms of efficiency, reproducibility and simplicity, precipitation-based methods are most amenable to studies with large numbers of subjects. However, the selectivity of the precipitation becomes critical. Here we present a simple plasma EV enrichment protocol based on pluronic block copolymer. The enriched plasma EVs were verified by multiple platforms. Our results showed that the particles enriched from plasma by the copolymer were EV size vesicles with membrane structure. Proteomic profiling showed that EV related proteins were significantly enriched, while high abundant plasma proteins were significantly reduced in comparison to other precipitation-based enrichment methods. Next generation sequencing confirmed the existence of various RNA species that was found in EVs from previous studies. Small RNA sequencing also showed enriched miR species compared to the corresponding plasma. Moreover, plasma EVs enriched from 20 advanced breast cancer patients and 20 age-matched non-cancer controls were profiled by semi-

quantitative mass spectrometry. Protein features were further screened by EV proteomic profiles generated from four breast cancer cell lines, and then selected in random forest-based cross-validation models. In total 60 protein features that highly contributed in model prediction were identified. Interestingly, a large portion of these features were found to be associated with breast cancer aggression, metastasis as well as invasion, consistent with the patients advanced clinical stage status. In summary, we have developed a plasma EV enrichment method with improved precipitation selectivity and it might be suitable for larger scale discovery studies.

Introduction

Extracellular vesicles (EVs) or exosomes, are small, spherical vesicles with a diameter of ~40-100nm that are secreted into the extracellular milieu by many cell types upon fusing the membrane of multi-vesicular bodies (MVBs) with the plasma membrane. In contrast, micro-particles (MPs), which include not only EVs, but also microvesicles (MVs), apoptotic bodies and apoptotic micro-particles are typically small particles ranging from 50nm to 3μm in diameter. Secreted MPs derived from MVBs fused with the plasma membrane are termed EVs, while MPs released from the surface of plasma membranes are referred to as microvesicles, membrane particles, as well as apoptotic vesicles[8]. Recent studies are now showing EV release may be a highly regulated process instead of a random process for the removing unwanted internal cellular debris as originally thought [20]. EVs located in biological fluids attract enormous interest because of their potential uses as a source of protein and nucleic acid biomarkers or as a prospective delivery system for various therapeutic purposes [26]. A simple and efficient method to enrich EVs from complex bio fluids, such as plasma samples, that addresses the many pitfalls of currently used methods would be useful in the research, clinical, and commercial setting.

Blood samples frequently contain large quantities of soluble proteins, aggregates and contaminants from other organelles that restrict the accuracy of EV analysis. Limitations of the current methods of EV enrichment from complex biological fluids include sample volume requirements, carryover of undesired high abundance proteins, and interference of downstream analysis due to leftover of method-based reagents. Differential ultracentrifugation (UC) was the gold standard to separate EVs[94], which aids

in the removal of most of the plasma contaminants. However, time consuming steps and large sample volume requirement leads to poor reproducibility, which restricts its utility for large scale studies. Immunoprecipitation (IP), while providing a faster process, is restricted only to certain EV subpopulation with known surface proteins, which limits its value to identify novel biomarkers. More recently, polymer based enrichment techniques by, for example PEG8000, have been used to enrich EVs[95].

Previous studies have already compared the reproducibility and efficiency in enriching plasma-derived EVs between UC and PEG-based methods[30-32]. Though results might vary from study to study, the conclusion, in general, was that UC-based methods may obtain EV with relatively better purity, but with very low efficiency. In contrast, PEG-based method offers higher efficiency and better reproducibility, but at the cost of huge contamination of high abundant plasma proteins. In addition, high viscosity and carryover of PEG also might affect the compatibility of the enriched sample with downstream applications, such as electro-microscopy (EM), mass spectrometry (MS) and flow cytometry (FL). Procedures that used for the removal of the polymer are often tedious and incomplete.

Pluronic block copolymers consist of ethylene oxide (EO) and propylene oxide (PO) blocks arranged in a triblock structure: EOx–POy–EOx. This arrangement results in an amphiphilic copolymer, in which the hydrophilicity and hydrophobicity could be altered by varying the size of the hydrophilic EO(x) and hydrophobic PO(y) units[96]. Pluronic block copolymers F68 and related F127 have been demonstrated to have broad utilities and biological properties. Copolymers with higher hydrophilic/lipophilic balance (HLB) (e.g., F-68, HLB0.80) were shown capable of being inserted into lipid bilayer

membranes and restoring the integrity of damaged membranes[97-100]; F68 was also shown to be capable of inhibiting protein aggregation[101]. The biological properties of pluronic block copolymers are different from PEG and offer the potential to increase the selectivity of EV precipitation from complex bio fluids, such as plasma.

An ideal method of enriching EVs from plasma might be a balance between high efficiency, low contamination and simplicity. We describe here a plasma EV enrichment method based on the pluronic block copolymer F68, which might be a balance between these three requirements. The enrichment method described here was proven to be more selective for plasma EV fraction with high efficiency, while much less contamination from high abundant plasma proteins, making it ideal for larger scale plasma EV biomarker studies. To demonstrate its potential utility, plasma EVs from 20 high-grade (stage III and IV) breast cancer patients and age-matched non-cancer controls were profiled and compared in the current study to identify features in the EVs that might be related to breast cancer aggression, metastasis and invasion.

Material and Methods

Cell line EV Preparation

Cell line used in this study in includes: Vcap, MCF7, T47D, MDA-MB-231, and MDA-MB-468. Cell culture: FBS (20% in respective medium) was depleted of bovine EV by centrifugation at 100,000g for 16 h at 4°C. Vcap cells were cultured in DMEM, MCF7 cells were culture in EMEM, T47D cells were cultured in RMPI1640, MDA-MB-231/468 were cultured in L15 medium, all supplemented with 10% depleted FBS, 2 mM L-glutamine, 1 U/mL penicillin, and 1 μg/mL streptomycin at 37°C and 5% CO2. Cell line generated EVs were isolated from cell cultures supernatant by sucrose density centrifugation. Briefly, supernatant was cleared of cells and cellular debris by sequential centrifugation at 400g for 10 min and 2000g for 20 min at 4°C. Cleared supernatant was concentrated by centrifugal filtration (Centricon Plus-70, 100 kDa NMWL), layered on a 30% sucrose cushion and centrifuged at 100,000g for 75 min at 4°C. Supernatant was removed and discarded without disrupting the cushion interface, which was then collected, diluted 6-fold with PBS and centrifuged at 100,000g for 70 min at 4°C. The resulting EV pellet was re-suspended in PBS by pipetting and incubation overnight at 4°C, and then stored at -80°C.

Breast Cancer / non-cancer Plasma Samples and Processing

Breast cancer and non-cancer samples were from Caris biorepository, age-match samples (average of 55 years old) were chosen (Figure 2.5). Breast cancer samples were all from advanced clinical stage III or IV. All information regarding to age and stage of the patients were list in Table 2.1. Plasma specimens utilized in this experiment were

obtained under an IRB-approved Biorepository Protocol. All subjects were consented with an IRB approved consent form and per 21 CFR 50.20 guidelines. Blood was collected using standard venipuncture to EDTA tubes and plasma was collected by standardized protocol. Blood were then spun in the Labofuge 200 for 10 minutes at 5300 RPM to remove cell debris. The clear layer of the plasma was aliquoted and transferred into cryovials and store at -80°C until use. Before the EV enrichment described below, the plasma was quickly thawed in water at room temperature (RT) and centrifuged at 4000g for 15min at RT to remove any potential protein aggregation/cell debris. The clear supernatant of the plasma was then used for the EV enrichment. Plasma aliquots for the EV enrichment below were all freeze-thawed only one-time.

EV Enrichment from Plasma

Pluronic block copolymer F68 was obtained from Amresco in a powder format, the block copolymer was dissolved in PBS to make a 15% stock solution. The stock solution is further diluted in PBS in a way that when adding 100ul of plasma to 900ul of such solution, the final F68 concentration for enrichment is 2%; the mixture was then incubated at room temperature for 30min; followed by centrifuge at 20000g for 30min at 4°C. The pellets were then washed with 1% F68 by vortexing following by centrifuge for 10min at 4°C at 20000g; Supernatant was then removed and appropriate amount of PBS were added to re-suspend the fraction. For cell line EV spiked-in sample, 1ug of the Vcap generated EV prepared as shown above, was spiked into 100ul plasma. Plasma EV extraction by Exoquick (System Biosciences, Inc.) and Total Exosome Isolation kit (ThermoFisher, TEI) instructions for 100ul of plasma were followed. Protein

quantification was performed by microBCA kit from Thermoscientific according to the instruction from the product.

DLS, ELISA and Western blot

Particle size distributions were measured with DynaPro Plate Reader II (Wyatt Technology Corporation, Santa Barbara, CA), in three replicates, 5 acquisitions of 5 seconds each at 25°C. PBS buffer and UC-isolated cell line EV were used as control. Direct CD9 ELISA was performed as follow: 1ug of the protein from neat plasma, plasma with cell line EV (Vcap) precipitated by F68 and plasma alone precipitated by F68 were coated on the 96 well ELISA plate overnight at 4°C, the plate were blocked by 1% BSA in PBS for 1 hour; biotinylated anti-CD9 monoclonal antibody (Abcam, ab34161) was added as detection antibody followed by streptavidin-HRP; 100ul of Ultra ELISA substrate (Pierce) were used and OD450 were recorded by the BioTek plate reader. For western blots, protein samples were separated on 4-12% Bis Tris protein gel and transferred to a PDVF membrane. Primary antibodies include mouse anti-CD9 (Abcam), mouse anti-β actin antibody (Abcam), mouse anti-HSP70 antibody (Abcam), rabbit anti-Von Willebrand Factor antibody (Abcam), rabbit anti-human serum albumin antibody (Abcam), mouse anti-alpha2 macroglobulin antibody (Abcam) and rabbit anti human IgG antibody (Abcam). Chemiluminescent signals were captured by imager (PXi, Syngene).

RNA analysis, RNA sequencing, small RNA (smRNA) Sequencing and miR assay

To extract total RNA and smRNA, trizol LS reagent was used (ThermoFisher). RNA, including total RNA and smRNA, was purified through the miRNeasy column (Qiagen) according to the instructions from manufacturer. All samples were eluted in 20ul of RNase-free water. 1ul of these samples were used for bioanalyzer analysis (Agilent). Density analysis of the small RNA bands was performed in ImageJ software package[102]. To prepare total RNA Next Gen Sequencing (NGS) library, SMARTer Stranded Total RNA-Seq Kit (Clontech) was used and were sequenced on the Illumina Miseq instrument according to the instruction from the manufacturer. The data were analyzed by a customized Tophat/cufflinks pipeline[103] against the ensemble-GRCh38 human genome and further categorized according to their biotypes from Ensemble GRCh38.85 RNA annotation. For smRNA sequencing, the library preparation was performed according to the instruction of Clean-Tag smRNA kit from Trilink Biotech. Data were then analyzed by customized smRNA pipeline, reads were normalized to RPM (reads per million mapped read) for comparison; smRNA recovery also examined by let-7a Taqman probe assay (ThermoFisher). The same amount of eluted sample was used. When comparing recovery rate of let-7a from cell line EV-spiked plasma and cell line EV alone, directly Ct comparison was used with normalization to the input sample volume (all samples were processed the same way, eluted by same volume). To comparing enrichment of let-7a from plasma enriched EVs to plasma alone, relative abundance was analyzed by $\Delta\Delta$Ct method with miR-451a as reference miR (one of the most abundant miR found in plasma[104]).

Transmission Electron Microscopy (TEM)

TEM and immuno-TEM analysis of EVs were performed as described previously[94] with modifications. Specifically, the vesicles were re-suspended in 1x PBS and deposited onto a formvar coated slide of copper mesh EM grids (FF300-Cu-50, Electron Microscopy Sciences). The vesicle-coated grids were washed, stained with 1% $UO_2(CH_3COO)_2$ and then viewed by the transmission EM (TEM) using a Philips CM12 with Gatan model 791 camera. For the immuno-gold labelling with antibodies, blocked grids coated with EVs were briefly fixed by ice-cold 1:1 ethanol/methanol for 5min followed by washing with PBS. The grids were then transferred to a drop of the anti-CD9 antibody (Abcam) in PBS with 1% BSA, 0.01% tween20 and incubated for 30 min. The grids were then washed, incubated with gold-labeled secondary antibody (Sigma). The grids were stained with 0.5% $UO_2(CH3COO)_2$ and then imaged by TEM.

Flow Cytometry Analysis

The isolated EVs were labelled with a cocktail of mouse anti-CD9-PE (BD), anti-CD63-PE (BD) and mouse anti-CD81 PE (BD) or mouse IgG1 kappa PE as isotype antibody control (eBioscience). PBS, antibody isotype control, anti-tetraspanin cocktail diluent as well as samples without staining were processed as controls. PE fluorescence intensity associated with particle events was measured by flow cytometry (A50-Micro, Apogee Flow Systems). Data analysis was performed in R3.2.2 environment with customized script.

Mass Spectrometry Analysis

In-solution trypsin digestion was performed according to the manufacturer instruction (ThermoFisher). Samples were analyzed by nanoflow reverse phase liquid chromatography using a Dionex Ultimate 3000 RSLCnano System (ThermoFisher) coupled in-line to a Q Exactive HF mass spectrometer (ThermoFisher). The nano LC system included an Acclaim PepMap 100 C18 5µm 100A 300µm×5mm trap column and an EASY-Spray C18 2µm 100A 50µm×150mm analytical column (ThermoFisher). Peptide samples were eluted with a two-step gradient of 2% to 30% B in 28 min then 30% to 45% B in 5 min, where B consisted of acetonitrile containing 0.1% formic acid. Blank samples consisting of 0.1% formic acid in water were injected between each sample and eluted with the same gradient profile and times as the samples. The LC system was interfaced with the mass spectrometry using an EASY-Spray electrospray ion source (ThermoFisher) and the samples were analyzed using positive ion spray voltage set to 2 kV, S-lens RF level at 65, and heated capillary at 285°C. The Q Exactive HF was operated in the data-dependent acquisition mode for fragmentation. MS1 survey scans (m/z 400–1400) were acquired in the Orbitrap analyzer with a resolution of 120,000 at m/z 200, an accumulation target of 3×10^6, and maximum fill time of 50ms. MS2 scans were collected using a resolution of 30,000 at m/z 200, an accumulation target of 1×10^5, and maximum fill time of 100ms, with an isolation window of 1.5 m/z, normalized collision energy of 28, and charged state recognition between 2 and 7.

ProteoWizard[105] was used for peak-picking, filtering out peaks with intensity less the 100 and converting the file to mzML format. Protein search and identifications were performed using MS-GF+[106] search engine on Homo sapiens (Uniprot TaxID=9606). For

semi-quantification analysis, an algorithm was established according to the previous normalized spectra index (SIN) algorithm[107-108] with modifications: sum of intensity from each peptide MS2 scan were normalized by the parent peptide charge, the normalized intensity for the same peptide from different MS2 scans were then summed together. The sum of the normalized intensity for the same peptide was then further normalized by the experimental molecular weight of the peptide, as a unit of normalized intensity per Dalton (mass). Normalized intensity per mass between different peptides of the same protein detected in the same sample was then averaged to obtain the final spectrum index (SI). For relative abundance, the SI was further normalized by the average total MS2 intensity among all samples.

Protein GO Categorization

To categorize the proteins for their relevance to EVs, individual GO terms were query for GO definition of extracellular exosome (0070062), cytoskeleton (0005856), cell surface receptor (0007166), endosome (0005768), Golgi apparatus (0044431), blood microparticles (0072562), endoplasmic reticulum (0044432), mitochondrion (0044429) and nucleosome (0000788). ExoCarta database version 5 was obtained from www.exocarta.org. For the second categorization, all proteins that could be found with GO terms of extracellular exosome, cytoskeleton, cell surface receptor, endosome as well as the ones that could be found in ExoCarta database were regarded as "Potential EV"; proteins with GO terms of blood microparticles, endoplasmic reticulum, mitochondrion as well as nucleosome were regarded as "Potential Contaminants" and proteins with both potential EV and potential contaminant terms, are regarded as "Undetermined".

Statistical Models

All the analysis was performed on R 3.2.2 environment. Leave-One-Out (LOO) cross-validation model was built based on random forest: Briefly, to predict each test sample (1), the rest of the samples (39) were used as training set to build the model. In order to select the most important features to maximizing the removal of trivial features, a second layer of LOO models was built within the 39 training set samples. In the second layer LOO models, features occur in at least 90% of the individual second layer models were extracted and used to build the final prediction model for that training set, this model (from 39 samples) was then used to predict the test sample. The predicted probabilities for individual samples from the LOO cross-validation model were then used to plot the AUC curve. The final set of selected features was then collected by combining all the features retained in the individual training models and used for the un-supervised cluster heatmap.

Permutation prediction were performed by permuting the sample label (cancer/non-cancer) 10000 times, same models as described above was applied to calculate the AUC for each permutation. Statistical significance p value was given by the percentage of permutation AUC that are equal or higher to the AUC performance from the correct label.

All other statistical analysis, include mean, SD, and t test were calculated either in Excel or in R environment, as were the graphs.

Results

Pluronic Block Copolymer F68 as a Reagent to Enrich EVs from Plasma

In order to examine whether the pluronic copolymer could be used as a reagent to enrich EVs from plasma, we first examined if the enriched fraction from plasma contains EV size particles. Plasma EVs was first enriched by F68 copolymer as described in the methods. Same procedures were also applied to samples spiked UC purified cell line EVs with or without plasma as controls. Dynamic Light Scatter (DLS) and transmission electron microscopy (TEM) were used to determine the size distributions as well as morphology of the plasma particles recovered respectively.

As shown in Figure 2.1Aa, cell line EVs were at size around 100nm and range between 30nm to 300nm, consistent with the EV sizes reported previously[20]. Size distributions in the cell line EVs precipitated by F68 in the absence/presence of plasma showed patterns consistent with the cell line EV alone (Figure 2.1Ab/c). Size distributions for plasma EV enriched by F68 was ranged between 20nm-100nm (Figure 2.1Ad), which matches exactly the size of EV described previously[20]. In contrast, particles in PBS buffer control were at around 10nm (Figure 2.1Ae), which is far below the expected EV size range, indicating the particles in the buffer solution would not interfere the analysis. Besides the major population at around 100nm, a small subpopulation of particles with a size distribution around ~1000nm were found in both cell line EVs from UC and the plasma particles enriched by F68 (Figure 2.1Aa/b/c/d). This small subpopulation may consist of particles made up of microvesicles, apoptotic bodies, or from potential EV aggregates.

Transmission electron microscopy (TEM) was used to capture the morphological detail of the enriched particles. UC purified cell line EVs, alone or recovered by F68 precipitation (in presence or absence of plasma) showed intact round shape vesicles with membrane structure (Figure 2.1B a, b, c), similar to the shape of EV that described in previous study[20]. The native plasma particles recovered by F68 precipitation also show similar shape with membrane structure and size (around 100nm) (Figure 2.1Bd), which is consistent to the DLS results, indicating the particles recovered from the plasma by the copolymer are in EV size range membrane structures.

To find out if the particles could be associated with tetraspanin, a common EV indicator, the enriched plasma EVs were labeled with immune-gold by anti-CD9 antibody[29]. For these scans, a modified protocol with ethanol/methanol fixation procedure was adopted. As a result, shapes of the EVs appeared to be more condensed and with lighter contrast compared to the regular protocol above. TEM scans shows cell line EVs alone or precipitated by F68 were able to be labeled with gold particles against CD9 marker as expected (Figure 2.1Ca/b), indicating the presence of CD9 on their surface. Anti-CD9 gold labeled particles were also found from native plasma EVs enriched by F68 copolymer (Figure 2.1Cc), indicating tetraspanin CD9, as a EV indicator, were also associated with the native plasma EVs enriched by F68.

To confirm whether the CD9 EV indicator was indeed enriched, the levels of CD9 presence in the samples were further examined by a direct CD9 ELISA. As shown in Figure 2.1D, EVs enriched by F68 copolymer from plasma sample with the cell-line EVs spike showed significant higher signal compared to the same amount of plasma protein without enrichment ($p<0.05$). Similarly, plasma EVs enriched by the copolymer also

showed significant higher CD9 signals compared to the plasma control ($p<0.05$), indicating EV marker CD9 was indeed enriched in the F68 precipitated population.

The EV recovery efficiency by the copolymer was also evaluated by the EV small RNA recovery from the precipitation of plasma samples with cell line generated EVs. Three distinct smRNA patterns (~140bp, ~90bp and ~60bp) could be detected from Vcap cell line generated EVs (Figure 2.1Ea). Such smRNA signature could be used to evaluate the recovery rate of EVs by densitometry analysis. We hypothesized that if the precipitate reagent is harmful to the EVs, for example disrupting the membrane structure, internal material including the RNAs, would be degraded or removed during the washing steps and would be reflected on the intensities of the smRNA signature bands. As shown in Figure 2.1Ea patterns and intensity of the cell line EV in PBS precipitated by F68 were very comparable to the direct RNA extraction from same amount of cell line EVs. At the same time, EV enriched by F68 from plasma sample with cell line EVs show same smRNA pattern as the cell line EV alone, the intensities of those signature bands were also not significantly different from the cell line EV controls.

To further evaluate the cell line EV smRNA recovery rate more quantitatively, densitometry analysis was performed for all three individual smRNA bands and summarized in Figure 2.1Eb. As shown, no significant difference was observed from analyzing respective smRNA bands between cell line EV, cell line EV precipitated by F68 and cell line EV in plasma precipitated by F68 samples ($p>0.05$), nor by the combined density of all three signature bands between those three samples ($p>0.05$). The recovery rate of each individual band for cell line EVs in PBS precipitated with F68 were ranged from 93% to 100% with final combined rate of 97% in reference to the cell line

EV control. Recovery rate of each individual band for cell line EVs in plasma enriched by F68 also ranged from 97% to 102% with final combined rate of 98%. The average recovery rate in sample with cell line EVs in plasma was slightly more than the cell line EV in PBS alone, which may be caused by the small amount small RNA recovered from plasma native EVs. High recovery rate of the spiked-in EVs smRNA suggested high recovery efficiency for spiked-in cell line EVs in presence of plasma, in turn implies high recovery rate for the native plasma EVs.

miR let-7a has been reported to reside inside EVs[109-111] and it was highly expressed in our cell line generated EVs (Vcap). In addition to the evaluation of efficiency by the smRNA pattern and densitometry from images, let-7a recovery rate was also examined by qPCR. As expected, no significant difference was observed in let-7a level between cell-line EV spiked plasma precipitated by F68 and the same amount of cell line EV alone (Ct 22.3±0.08 vs Ct 22.8±0.06, $p>0.05$), indicating the level of let-7a recovered by F68 was very comparable, from the same amount of original cell line EVs, consistent to the smRNA analysis that showed high recovery rate of EV smRNA from the F68 enrichment.

Let-7a level from native plasma EVs precipitated by F68 were also compared to the let-7a level from whole plasma extraction. There was no significant difference in recovered let-7a level in the plasma EVs enriched by F68 compared to same amount of plasma RNA extraction (Ct 31.6±0.3 vs Ct 31.4±0.7, $p>0.05$), indicating the precipitation was able to fully recover most of the circulating plasma let-7a microRNAs. Meanwhile, relative enrichment of let-7a in comparison to the original plasma was also evaluated by $\Delta\Delta$Ct analysis by normalizing let-7a levels to miR-451a, one of the most abundant

plasma miR[104]. As show in Figure 2.1F, let-7a level in the plasma EVs fraction precipitated by F68 enriched more than 20 folds from the corresponding neat plasma, mostly because the significant reduction in recovery of miR-451a (Ct 26.4±0.5 vs Ct 22±0.7, $p<0.05$). The results indicate that EV related miR like let-7a was enriched by the F68 precipitation, while plasma related miR was significantly reduced, suggesting the high recovery of native plasma EVs could be obtained by the copolymer precipitation.

Additionally, EV recovery efficiency was analyzed in term of protein recovery rate. Cell line EVs in PBS were precipitated by F68 according to the above protocol. Protein amounts from before and after the precipitation were measured and the protein recovery rate was then calculated. As shown in the Figure 2.1G, about 80% of the cell line EV proteins (before: total protein amount is 9991±1156 ng, after: 7852±852 ng, 4 replicates for each sample) was recovered after the precipitation. The result is largely in agreement with the above result that showed >90% recovery by smRNA density analysis and let-7a qPCR analysis.

To further verify the EV identity of the particles recovered from F68 precipitation and the sample's compatibility to downstream application, flow cytometry (FL) analysis was employed. EVs enriched by F68 precipitation were labeled with a fluorescence anti-tetraspanin (CD9/63/81) antibody cocktail. Figure 2.1H(abc) shows FL analysis of the EVs enriched by F68 from plasma sample with cell line EVs by the light scatter signals: when the enriched particles were not labeled with any antibody, majority of the particles distributed horizontally along the SALS-area axis with very narrow LALS-area range (Figure 2.1H a); similar pattern was also found for the same particles labeled with isotype antibody control (Figure 2.1H b); In contrast, when the particles were labeled with anti-

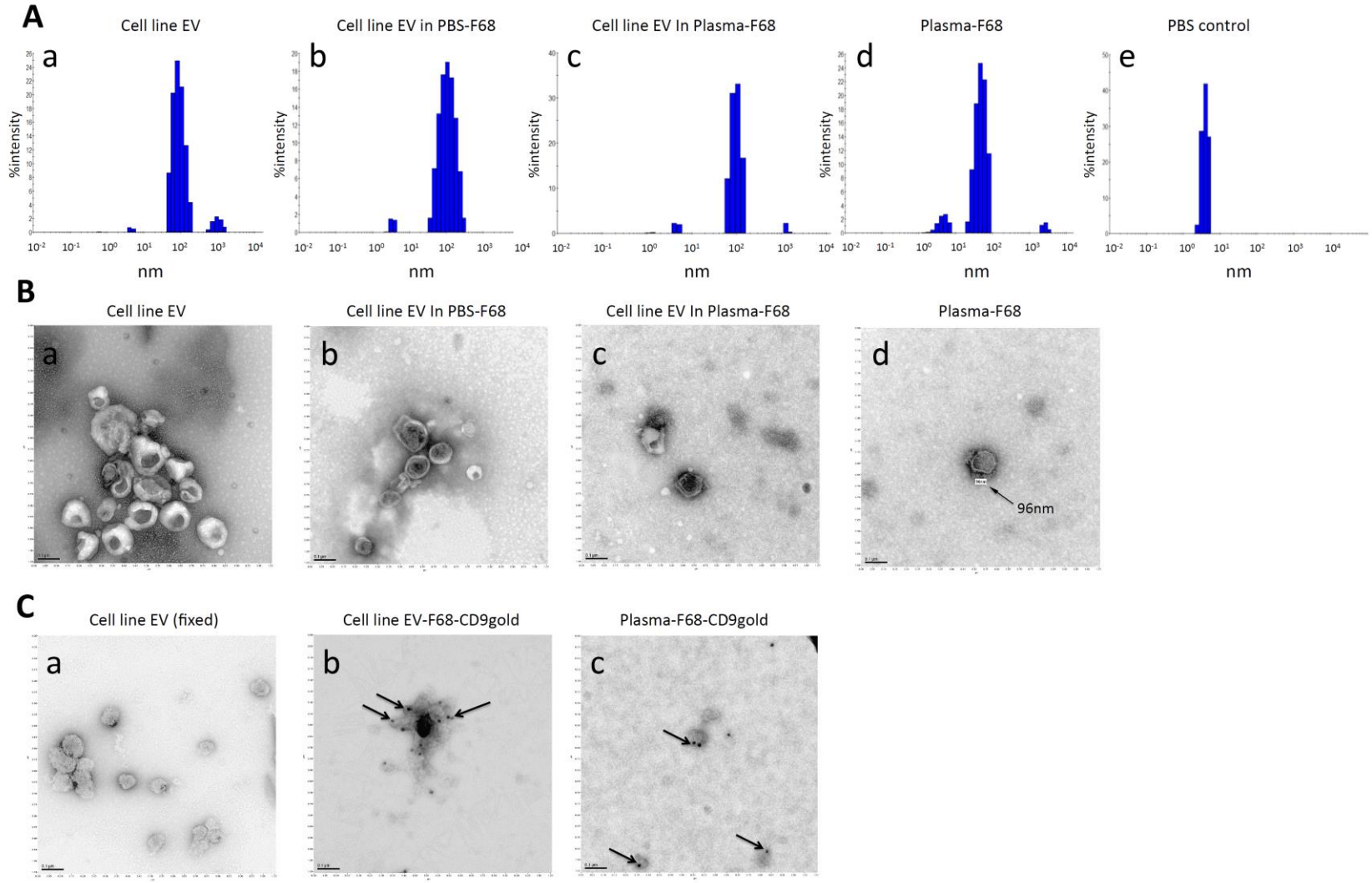
tetraspanin antibody cocktail, a more condensed population, with an correlated LALS-area/SALS-area signal appeared on the plot with an angle against the horizontal minor population(Figure 2.1Hc). This population was clearly distinct from the noise particles found in sample labeled with isotype antibody, as well as non-labeled particles, suggesting a positive binding on the EV size particles by the tetraspanin antibody might have an effect on the integrated light scatter signals, indicating the population of antibody-labeled EVs could be isolated based on such light scatter pattern.

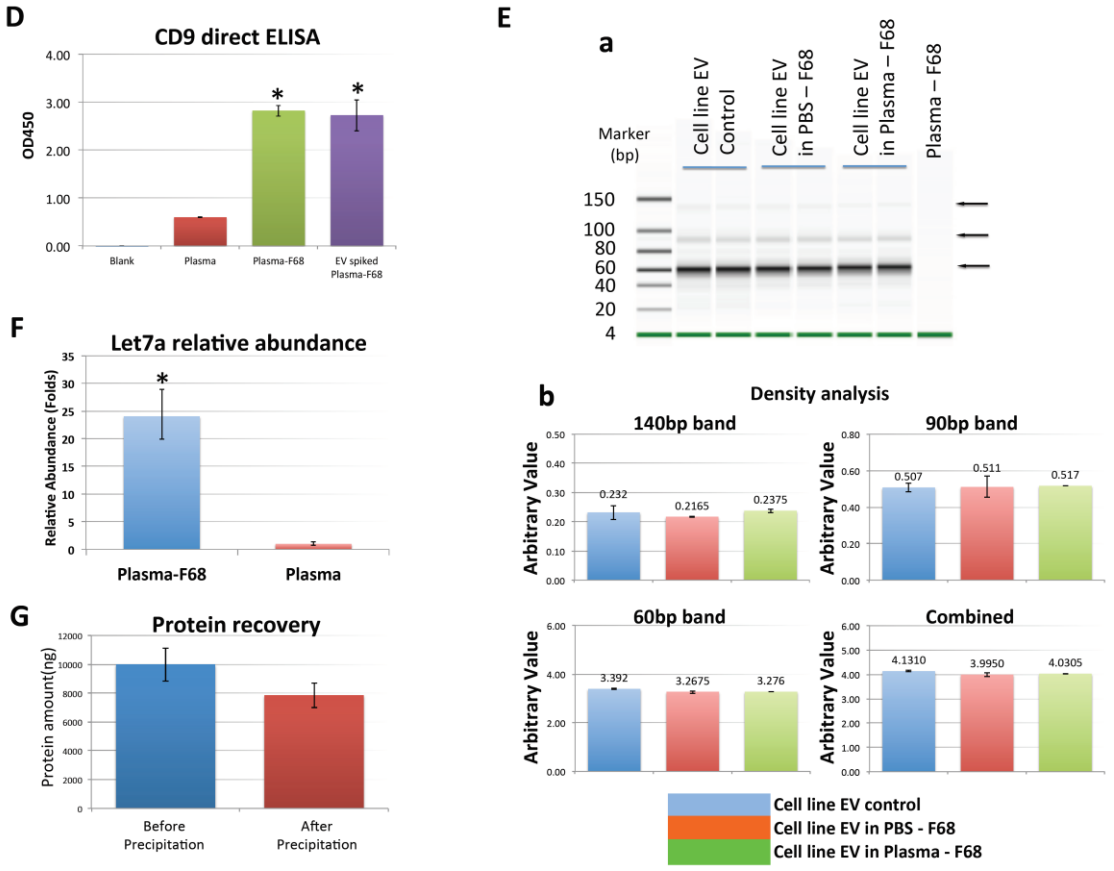
Based on this finding, native plasma EVs enriched by F68 copolymer were also analyzed by the same scheme. A similar pattern was found as shown in Figure 2.1H (def). Native EVs enriched from plasma, when labeled with tetraspanin antibody, show two distinct populations (Figure 2.1H f), one aligned to the distribution from non-staining controls (Figure 2.1H d) and the samples stained with isotype antibody control (Figure 2.1H e), the other population showed signal range similar to the positive control in Figure 2.1H c. Minimal and maximal threshold were then determined manually (Figure 2.1Ha-f, green and red lines, two-step gating see Appendix-A) to select the particles events for fluorescence signal analysis. The selected populations were summarized in a distribution histogram according to their PE fluorescence signals. EVs enriched from a plasma sample spiked with cell-line EVs showed that the majority of the selected populations were tetraspanin positive (green) while the none-specific events that associated with isotype-antibody-PE labeling in the selected population were minimal (blue, almost not visible) (Figure 2.1H g). For the native plasma EVs enriched by F68, a similar pattern was also observed, with more tetraspanin positive events compared to the non-specific signal from isotype-PE labeling (Figure 2.1H h). The PE signal range in the selected

population was very similar to that from cell line EV spiked plasma positive control in Figure 2.1H g. This result further confirms that there is a subpopulation from native plasma EVs enriched by F68 method could be identified with the tetraspanin expression pattern when compared to the cell line generated EVs. (More detail description of the analysis and more controls see Appendix-A)

Figure 2.1 Characterizations of the EVs Enriched by Pluronic Copolymer F68. A) DLS characterization of the particle size: a) cell line EV, b) cell line EV in PBS precipitated by F68, c) cell line EV in plasma enriched by F68, d) native plasma EV enriched by F68, e) PBS buffer control; B) TEM image of EVs from: a) the cell line EVs, b) cell line EV in PBS precipitated by F68, c) plasma with cell line EV spike-in precipitated by F68, d) native plasma EVs enriched by F68 with arrow point to the label of the EV size. All scale in the images are 100nm as shown by the scale bar on lower left corner of the images; C) CD9 Immuno-gold TEM, EVs from a) cell line EVs, b) plasma spiked with cell line EVs and precipitated by F68, and c) native plasma EV enriched by F68, were labeled with anti-CD9 antibody followed by secondary antibody with gold particle; D) Direct CD9 ELISA, 1ug of proteins from each sample were used, two replicate per sample (*significantly higher than corresponding plasma sample, $p<0.05$); E) smRNA recovery analysis: a) gel image from bioanalyzer for comparing smRNA recovery efficiency, all samples were processed in the same way, equal volume of the eluted sample (1ul) were analyzed. Arrow point to the three most concentrated bands for density analysis; b) density analysis of three distinct cell-line EV smRNA bands (140bp, 90bp and 60bp). Plots comparing including in density from individual band and combined density from all three bands are shown; F) Relative abundance ($\Delta\Delta$Ct analysis) of let-7a microRNA from EV fraction enriched by F68 method compared to original plasma, 5 replicate for each sample (plasma as reference sample, miR-451 as reference miR; *significantly higher than corresponding plasma sample, $p<0.05$); G) Protein recovery by F68 precipitation with cell line EVs in PBS. H) Flow Cytometry: a-g) EVs enriched by F68 from plasma spiked with cell line EV; d-h) Native plasma EVs enriched by F68; a, d) samples without

labeling with antibody; b, e) samples labeled with isotype antibody-PE antibody; c, f) samples were labeled with tetraspanin antibody-PE cocktail (CD9/63/81). Green line and red lines are the minimal/maximal thresholds for SALS-area and LALS-area respectively to gate the particle population. The extracted particle events were then projected in the event distribution histogram according to their PE fluorescence signal strength. g) Event distribution histogram for EV enriched from plasma sample spiked with cell line EVs and f) Event distribution histogram for native plasma EVs enriched by F68. (Tet+: tetraspanin CD9/63/81 antibody-PE positive population; isotype+: isotype antibody-PE positive population)


A
Cell line EV
Cell line EV in PBS-F68
Cell line EV In Plasma-F68
Plasma-F68
PBS control
a
b
c
d
e
%intensity
nm
B
Cell line EV
Cell line EV In PBS-F68
Cell line EV In Plasma-F68
Plasma-F68
a
b
c
d
96nm
C
Cell line EV (fixed)
Cell line EV-F68-CD9gold
Plasma-F68-CD9gold
a
b
c


D
CD9 direct ELISA
OD450
4.00
3.00
2.00
1.00
0.00
Blank
Plasma
Plasma-F68
EV spiked Plasma-F68
E
a
Marker (bp)
Cell line EV Control
Cell line EV in PBS – F68
Cell line EV in Plasma – F68
Plasma – F68
150
100
80
60
40
20
4
b
Density analysis
140bp band
90bp band
60bp band
Combined
Arbitrary Value
0.232
0.2165
0.2375
0.507
0.511
0.517
3.392
3.2675
3.276
4.1310
3.9950
4.0305
Cell line EV control
Cell line EV in PBS - F68
Cell line EV in Plasma - F68
F
Let7a relative abundance
Relative Abundance (Folds)
Plasma-F68
Plasma
G
Protein recovery
Protein amount(ng)
Before Precipitation
After Precipitation
H
a No label
b Isotype-PE
c Tet-PE
d No label
e Isotype-PE
f Tet-PE
LALS-area
SALS-area
g
Plasma-F68 with Cell line EV
Tet+
isotype+
Frequency
F2-488(Peak)
h
Plasma-F68
Tet+
isotype+
Frequency
F2-488(Peak)

Proteomic Profiling of Plasma EVs Enriched by Pluronic Copolymer

Our results have shown that the plasma particles enriched by F68 precipitation were EV-size membrane structures, contained RNA/miRNA, part of them contained EV marker tetraspanin proteins. While tetraspanins (CD9/63/81) are frequently used as potential EV markers, they are certainly not universal markers for any EV populations from varieties of cell types[29], especially in plasma. To evaluate what EV-related proteins could be detected and enriched by the copolymer method, shotgun semi-quantitative mass spectrometry (MS) analysis was adopted. In addition, in order to evaluate the purity of the plasma EV recovered by the proposed method, two PEG-based EV enrichment methods commercially available (Exoquick and TEI) were also used as comparison.

The EV fractions from same plasma were enriched from three different protocols: F68, Exoquick and TEI Kit. After precipitation according to respective protocols, protein recoveries were first measured and compared: total protein recovery was 0.46% for F68 protocol, 16% from Exoquick protocol and 4.6% for TEI kit. As >99% of the protein mass in plasma were from high abundant plasma proteins[112], protein mass related to EV was not expected to be high. In contrast, higher yield might imply high abundant plasma protein contamination.

Both high abundant plasma proteins (including ALB, A2M and IgG) and the EV related proteins (including ACTB, HSP70 and VWF) were examined by western blot and semi-quantitative MS analysis. ALB and IgG are the two most abundant plasma proteins, and A2M is also one of the high ranking proteins in the plasma[112]. High amount of ALB not only implies plasma protein contamination, but also might reduce the sensitivity of the MS by narrowing its dynamic range of analysis. On the other hand, CD9, ACTB,

HSP70 and VWF were used as an indicator of plasma EV enrichment. ACTB is a cytoskeleton protein while HSP70 is cytosolic proteins and both were reported to be enriched in EVs[37, 113-114]. VWF has been reported to be one of the protein markers found in platelet derived EVs[113-114], which accounts for 70%-90% of the EVs in the plasma. Protocols that could enrich plasma EV should be most likely to include this protein.

Western blots were first used to examine these general EV protein indicators. As shown in Figure 2.2A, the relative abundance of ALB in the neat plasma was obviously the highest of all, Exoquick and TEI kit was able to reduce that slightly but not very significant based on the density of the images. Level of ALB in plasma EV enriched by F68 dropped dramatically as shown by the blot. In comparison of A2M and IgG levels, EVs enriched by F68 methods consistently showed lowest contamination of these two proteins compared to other two methods. In contrast, Exoquick showed highly enrichment of A2M and IgG, while TEI kit showed slightly reduced in A2M but almost same level of IgG compared to neat plasma.

On the other hand, tetraspanin CD9 and ACTB were detected in plasma EVs enriched by F68 and TEI kit only, but not detected in EVs enriched by Exoquick as well as neat plasma; HSP70 was enriched in all three methods compared to the neat plasma with highest level in EVs enriched by TEI kit, followed by EVs from F68 and Exoquick respectively. VWF was significantly enriched in plasma EVs enriched by F68, which is also higher than those from samples enriched by Exoquick and TEI kit respectively.

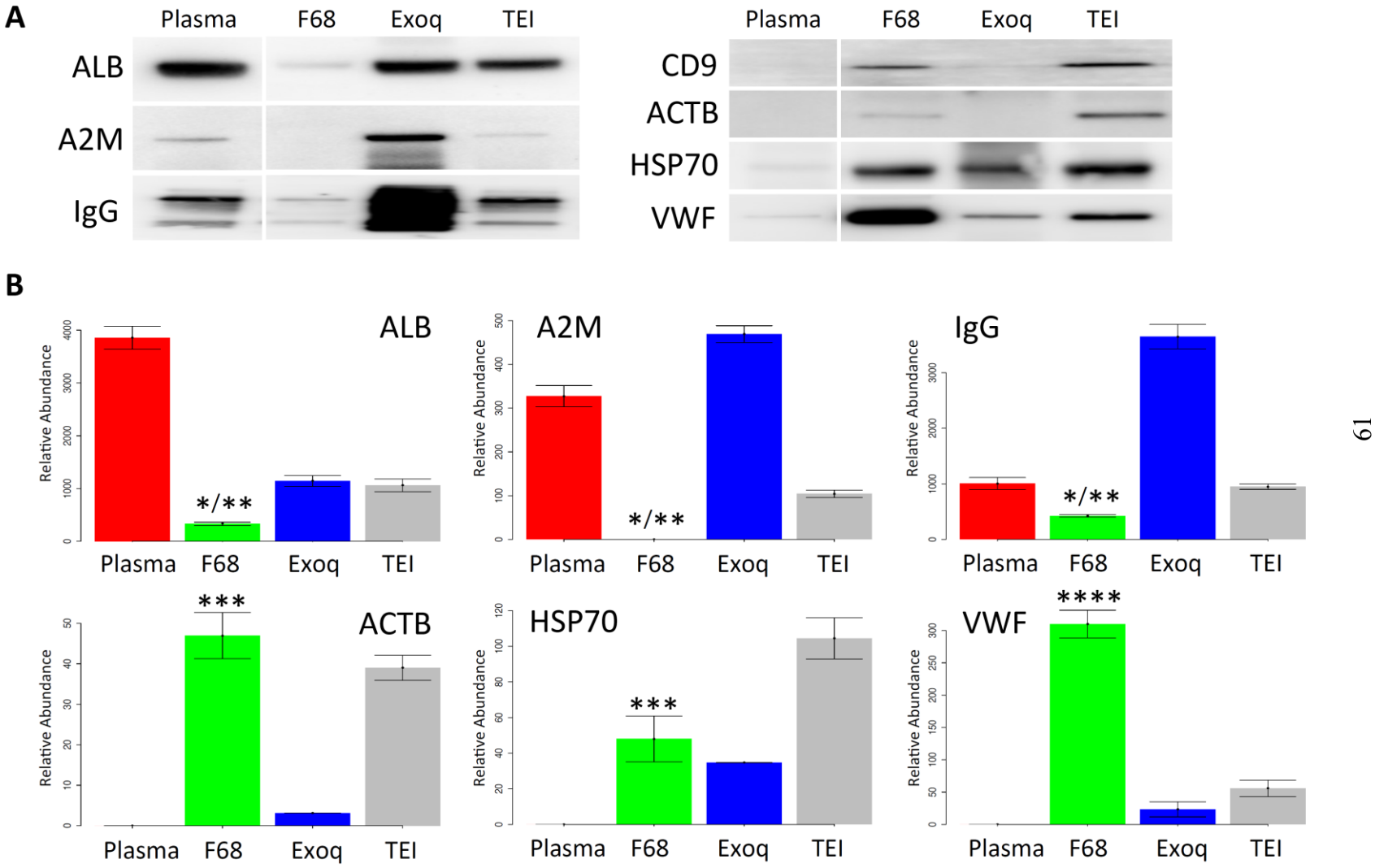
Semi-quantitative MS analyses of the same proteins as above are shown in Figure 2.2B. As expected, the plasma contained highest level of ALB. Plasma EVs enriched by Exoquick and TEI reduced ALB by about 70%, but both were still more than 400%

higher than that from F68 method. A2M protein was significantly reduced by 65% in plasma EVs enriched by TEI kit compared to neat plasma ($p<0.05$), but it was completely removed from the plasma EVs enriched by F68. In contrast, sample enriched by Exoquick showed significant higher level of A2M compared to others ($p<0.05$). Plasma enriched EVs by F68 method significantly reduced the recovery of IgG proteins to about 35% compared to the neat plasma ($p<0.05$), while sample processed by Exoquick again significantly enriched more IgG than the other two methods ($p<0.05$).

Relative abundance of ACTB was enriched in plasma EV from both F68 and TEI protocols compared to the neat plasma ($p<0.05$), while the enrichment of ACTB in Exoquick was significantly lower than the other two methods ($p<0.05$); HSP70 levels was also shown to be enriched by all three methods compared to neat plasma, with highest by TEI and comparable level enrichment between Exoquick and F68. Most significantly, relative abundance of VWF showed to be highly enriched in EV fraction by F68 protocol compared to others ($p<0.05$), which was more than 16 folds higher than that from TEI Kit protocol and over 4 folds higher than that from Exoquick protocol.

The results between western blot and the semi-quantitative MS analysis were highly consistent, which gives us the confidence of using such semi-quantitative method for global comparisons. Though CD9 was readily detected on the western blot, it and other tetraspanins were very difficult to be detected in MS, partially due to high hydrophobicity[115]. However, the western blot of CD9 confirmed the enrichment of the tetraspanin protein in the plasma particles enriched by F68 precipitation, consistent to the result from the above ELISA, TEM and FL analysis.

Figure 2.2 Comparing Different Plasma EV Enrichment Methods. A) Western blot comparison of plasma high abundant protein and EV-related protein among three precipitations reagents used. Plasma high abundant proteins include ALB, A2M, and IgG. Plasma EV-related proteins include CD9, ACTB, HSP70 and VWF; B) Same comparison of the plasma high abundant proteins and EV-related proteins by semi-quantitative shotgun proteomic analysis. (Exoq: Exoquick; TEI: Total Exosome Isolation Kit, 4 replicates per sample; *significantly lower compared to original plasma ($p<0.05$); **significantly lower compared to Exoquick/TEI methods ($p<0.05$); ***significantly enriched compared to original plasma; ****significantly enriched compared to Exoquick/TEI methods.


A
Plasma
F68
Exoq
TEI
ALB
A2M
IgG
CD9
ACTB
HSP70
VWF
B
ALB
A2M
IgG
ACTB
HSP70
VWF
Relative Abundance
*/**
***
****
Plasma
F68
Exoq
TEI

In Figure 2.3A, shotgun proteomics showed that among the >5000 proteins identified from all three methods, over half of those proteins (2648) were shared features for all methods as well as in the corresponding plasma. 93 unique proteins were identified from EVs enriched by F68 compared to others, which is quite similar to that from Exoquick (88) and slightly more than TEI kit (76). Comparing to Exoquick and TEI kit, the unique protein numbers are quite even in either comparison (F68 vs Exoquick: 629 vs 614; F68 vs TEI Kit: 587 vs 595).

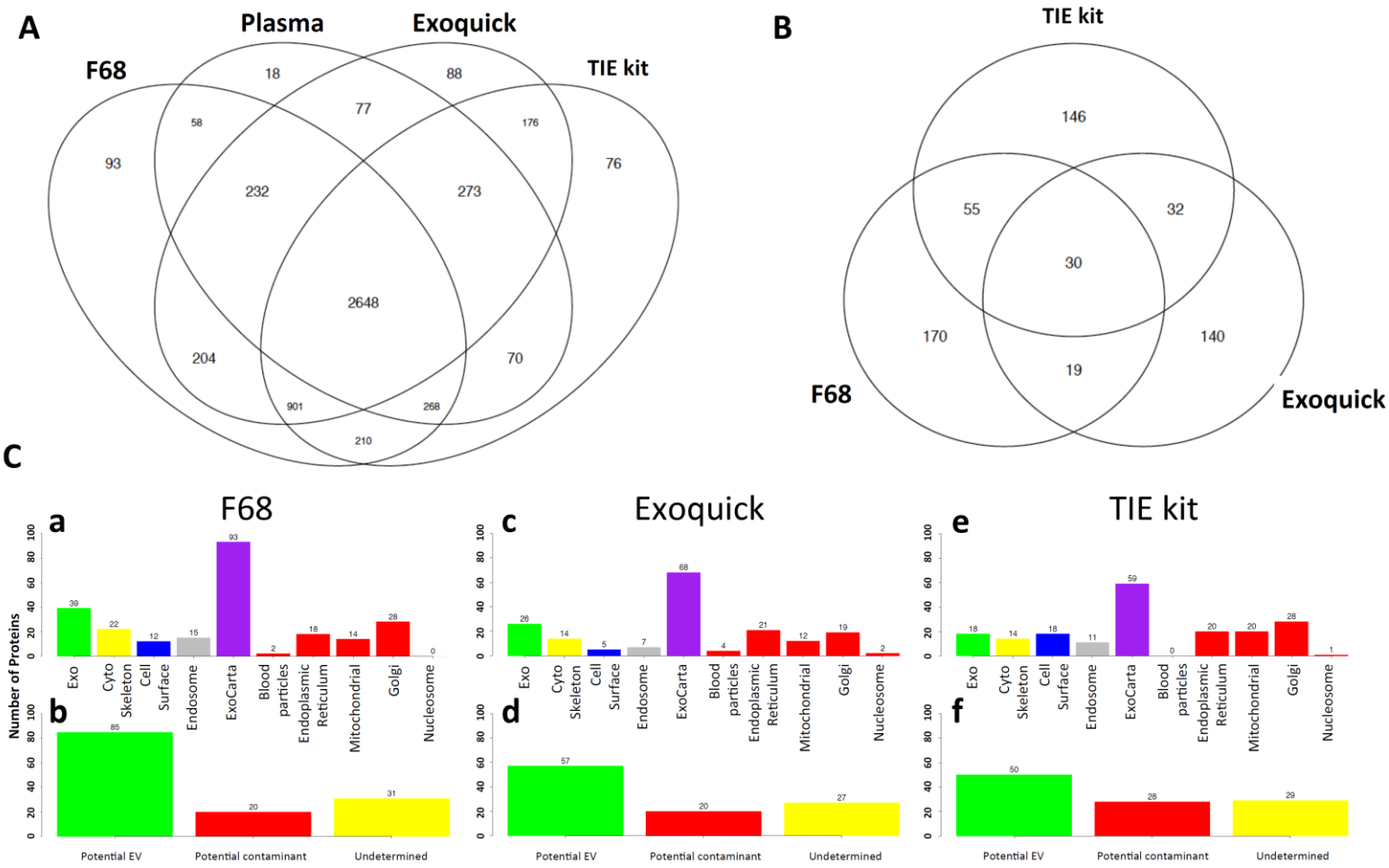
To find out what proteins were enriched in the precipitates recovered by different enrichment methods semi-quantitatively, features were first filtered by none-zero data points from individual methods as well as in the neat plasma, then the average relative abundance (from 4 technical replicates) were compared between individual method to the corresponding neat plasma; fold changes of all valid features in reference to the neat plasma were then obtained, while at the same time t-tests were also performed to gain statistical significance. The highly enriched features were defined as those features where the average relative abundance was 10-fold higher ($p<0.05$ by t-test) than the corresponding neat plasma. Therefore, sets of features with valid quantitative enrichment of equal or above 10 folds compared to that from neat plasma were obtained. These features from different methods were then used in the Venn diagram of Figure 2.3B. Among the 592 shared protein features identified from all three EV enrichment methods with significant enrichment level from plasma by different methods, 170 are unique to F68 protocol, 140 proteins are unique to Exoquick protocol, and 146 proteins from TEI kit.

To explore what kind of cellular functions that these enriched proteins (both unique features in Figure 2.3A and 2.3B) could be involved, the identified proteins were then categorized according to their GO terms. Proteins were first categorized by their association with EV, cytoskeleton, cell surface, endosome, blood macromolecule, endoplasmic reticulum (ER), mitochondrial, Golgi apparatus and nucleosome (Figure 2.3Cace). Cytoskeleton, cell surface and endosome related proteins are ubiquitous, and they are most likely involved in the biogenesis of EV[29]. Proteins were also queried against the EV protein database ExoCarta[19] to match the EV proteins that have been identified from previous studies. These proteins were then categorized as "Potential EV" proteins. In contrast, blood macromolecules, ER, mitochondrial, Golgi and nucleosome related proteins are less unlikely to be present in EVs[29] due to their different biological pathway. They were then all categorized as "Potential Contaminant". Meanwhile, proteins could be identified either as "Potential EV" or "Potential Contaminant" categories were regarded as "Undetermined" (Figure 2.3Cbdf).

In highly enriched proteins from plasma EVs recovered by F68 (including the unique features for individual methods in both Figure 3A and 3B), 40 highly enriched proteins were identified as potential EV proteins, 22 as cytoskeleton proteins, 12 as cell surface proteins, 15 as endosome related proteins and 93 proteins were found matched in ExoCarta EV protein database (Figure 2.3Ca). When categorized only by the three groups ("Potential EV", "Potential Contaminant", and "Undetermined"), there are 85 "Potential EV" proteins, 31 proteins are "Undetermined" proteins, and only 20 are "Potential Contaminant" proteins (Figure 2.3Cb). Same analysis was also applied to the samples from Exoquick and TEI kit preparations (Figure 2.3Ccdef). The EV related proteins from

F68 method showed 50% more than Exoquick (40 to 26) and over 2 folds more than TEI kit (40 to 18). Similarly, the number of highly enriched "Potential EV" proteins in F68 protocol is about 50% more than corresponding Exoquick and TEI kit protocol (85 vs 57 and 60 respectively). As the previous results showed the F68 method significantly reduced the high abundant plasma protein recovery in the enriched EVs fraction, we further showed that the F68 enrichment method was able to enrich more "Potential EV" protein features compared to the other two precipitation-based protocol while dramatically reduce the plasma high abundant protein contaminations.

Figure 2.3 Categorization of Plasma EVs Proteins Enriched from Different Enrichment Method. A) Venn Diagram of unique and shared identified proteins in MS analysis from F68, Exoquick, TEI kit as well as the corresponding plasma; B) Venn Diagram of unique and shared proteins identified from the semi-quantitative MS analysis from different EV enrichment protocols that are significantly higher than original plasma ($p<0.05$ and $>= 10$ folds higher); C) a/b: protein categorization for plasma EV enriched by F68, c/d: protein categorization for plasma EV enriched by Exoquick, e/f: protein categorization for plasma EV enriched by TEI. *unique proteins identified in A as well as the unique proteins that show significantly higher than plasma from individual protocol were categorized to extracellular exosome(Exo), cytoskeleton(Cyto), cell surface, endosome, blood macromolecule, endoplasmic reticulum (ER), mitochondrial, Golgi apparatus and nucleosome. Potential EV includes the proteins from EV, cytoskeleton, cell surface, and endosome as well as proteins that matched in Exocarta. Potential contaminants include blood macromolecule, endoplasmic reticulum (ER), mitochondrial, Golgi apparatus and nucleosome category. Undetermined category includes the proteins that could be identified either as potential EV and potential contaminants.


A
Plasma
Exoquick
F68
TIE kit
18
88
77
58
176
93
76
232
273
2648
204
70
901
268
210
B
TIE kit
146
55
32
30
170
140
19
F68
Exoquick
C
F68
Exoquick
TIE kit
a
b
c
d
e
f
Number of Proteins
Exo
Cyto Skeleton
Cell Surface
Endosome
ExoCarta
Blood particles
Endoplasmic Reticulum
Mitochondrial
Golgi
Nucleosome
Potential EV
Potential contaminant
Undetermined

NGS Profiling of Plasma EVs Enriched by Pluronic Copolymer

The initial let-7a qPCR assay suggested that the RNA inside plasma EVs enriched by F68 could be well preserved. To further identify what RNAs species that exist in the plasma EVs enriched by F68, total RNA sequencing and smRNA sequencing were performed respectively.

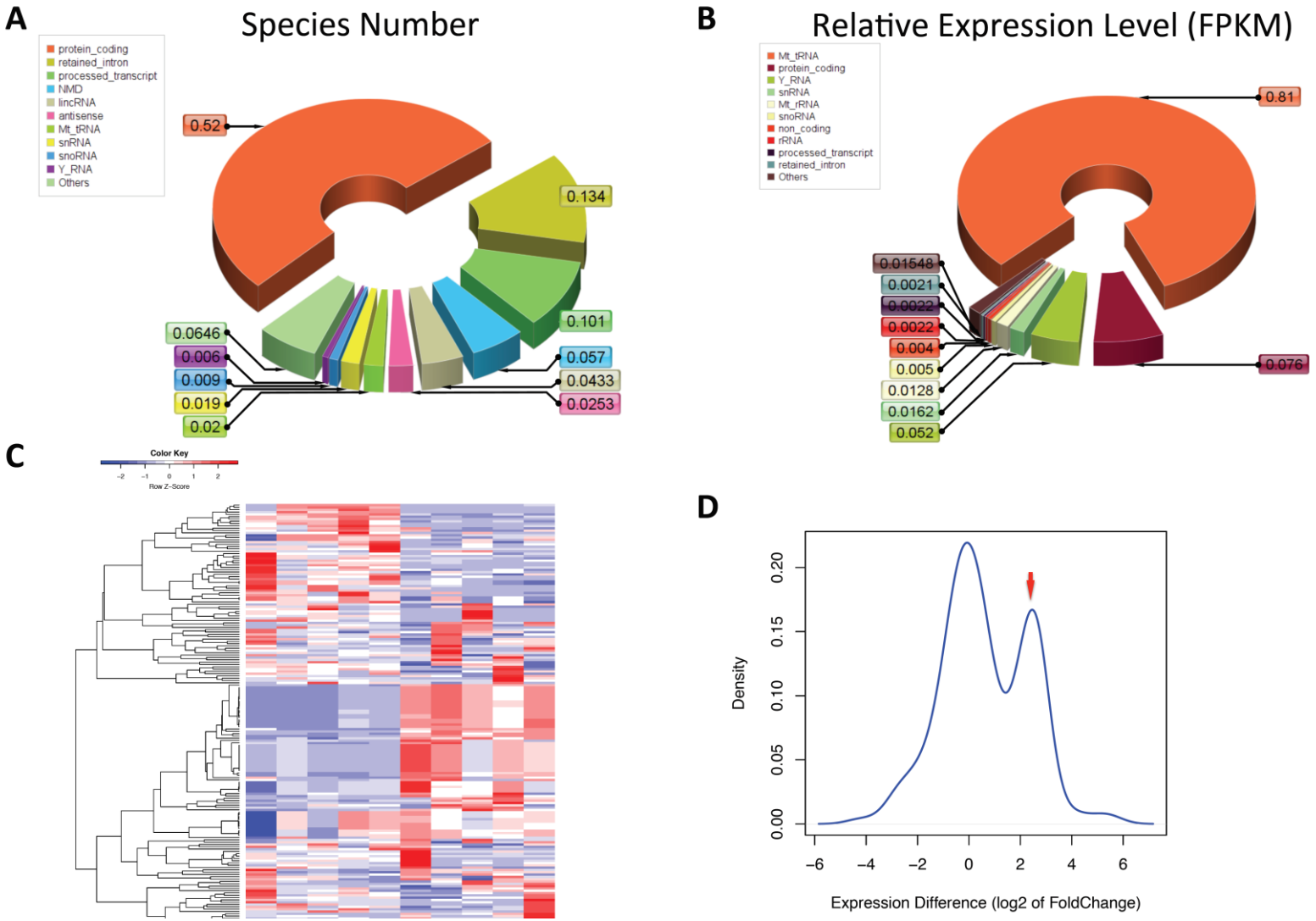
Total RNA profile is summarized in Figure 2.4. The genes detected in the NGS were categorized according to the biotype from Ensemble GRCh38.85 RNA annotation in terms of RNA species count (Figure 2.4A) as well as relative abundance by FPKM (Fragment Per Kilo base of transcript Per Million mapped reads, Figure 2.4B). Among all RNA species detected in the plasma EV enriched by F68: ~58.5% are protein coding RNA species (including protein coding 52%, alternative splicing species(retained intron) ~6.5%); About 5.7% of the transcripts match the non-sense mediated degradation category (NMD), consistent with previous study that showed RNA species subjected to RNA degradation were found in EVs[116]. All other species are different subtypes of non-coding RNAs, including 10% transcripts without valid ORF; 4.3% are Long, intervening noncoding RNA (lincRNA), which also have been reported to be packaged into EVs[6, 117]; 2% mtRNA, including mitochondria-origin tRNA and rRNA, which are also reported in previous researches[116]; Other non-coding RNA species, such as multiple antisense RNA, snRNA, snoRNA, Y_RNA that were reported to be packaged into EVs[6, 116, 118] previously were also found in our study.

In contrast, the relative abundance analysis showed that mt_tRNA was the most abundant RNA species from the plasma EV fraction enriched by F68, account for more than 80%, followed by protein coding transcripts (7.6%). Though quite some species of

NMD RNA were identified above, their levels are relative low in the EV fraction, only accounts for less the 0.08%; Y_RNA has also relative small amount of species number, but it accounts for more than 5% relative abundance of the total RNAs, consistent to the previous report that considerable amount of Y_RNA was found in EVs[116].

Since the total RNA sequencing have limited capability in examining all short RNA fragments, smRNA sequencing was also performed by a ligation-based sequencing method[119], whose optimal templates are 5' phosphorylated RNAs, such as miRNA and those smRNA derived from tRNA (tRFs) and Y_RNA (RNY) but act like miRNA[120-121]. Variety of smRNA species were observed that mapped to miRNA, tRFs, RNY, piRNA as well and lincRNA with length range from 18 to 40 nucleotides. To evaluate if such smRNA profile could distinguish the enriched EVs from the corresponding plasma, a supervised-hierarchical clustering of the smRNA species was performed with the heat map shown in Figure 2.4C. Two smRNA clusters were formed, one show enriched and the other faded in the enriched EVs compared to the corresponding plasma profiles. In addition, relative quantitative measurement showed a bimodal distribution of smRNA in reference to plasma, with one set of smRNA features enriched from the plasma (Figure 2.4D).

Figure 2.4 NGS Profiling of Plasma EVs Enriched by Pluronic Copolymer F68. Total RNA profiling was categorized by either RNA species number or their relative abundance according to their biotype: A) RNA species number according to their biotypes, the frequency in the pie chart indicate the proportion of transcript species identified for each biotype; B) Frequency in the pie chart accounts for the combined relative expression levels in term of FPKM for each biotype. C) Heatmap of supervised hierarchical clustering of smRNA from plasma EV enriched by F68 and their corresponding plasma, from blue to red indicates smRNA expression level (log2 normalized expression). D) Density plot of smRNAs relative abundance in comparison to plasma in terms of fold changes. Fold changes were indicated in log2 scale, the arrow mark the smRNA set that enriched from plasma.


A
Species Number
protein_coding
retained_intron
processed_transcript
NMD
lincRNA
antisense
Mt_tRNA
snRNA
snoRNA
Y_RNA
Others
0.52
0.134
0.101
0.057
0.0433
0.0253
0.0646
0.006
0.009
0.019
0.02
B
Relative Expression Level (FPKM)
Mt_tRNA
protein_coding
Y_RNA
snRNA
Mt_rRNA
snoRNA
non_coding
rRNA
processed_transcript
retained_intron
Others
0.81
0.076
0.01548
0.0021
0.0022
0.0022
0.004
0.005
0.0128
0.0162
0.052
C
Color Key
Row Z-Score
D
Density
0.00
0.05
0.10
0.15
0.20
−6
−4
−2
0
2
4
6
Expression Difference (log2 of FoldChange)

Comparing Plasma EVs Profiles from Advanced Breast Cancer and Non-Cancer Patients

Our results proved that the proposed pluronic copolymer method was able to efficiently enrich plasma EV with less contamination of high abundance plasma protein. To explore its potential utility, plasma EVs from 20 breast cancers with advanced clinical stages (clinical stage III and IV) and 20 age-matched non-cancers were enriched by the F68 method and profiled (Table 2.1). The semi-quantitative MS profiles were then compared to find out whether protein signatures related to breast cancer at such clinical stage could be identified from these enriched plasma EVs.

Table 2.1 Demographic Information of Breast Cancer Patient and Controls

	CaseStatus	stage	stage_M
1	Cancer	III	M0
2	Cancer	III	M0
3	Cancer	IV	M1
4	Cancer	III	MX
5	Cancer	III	M0
6	Cancer	III	M0
7	Cancer	III	M0
8	Cancer	III	M0
9	Cancer	IV	M1
10	Cancer	III	MX
11	Cancer	III	MX
12	Cancer	III	M0
13	Cancer	III	MX
14	Cancer	III	M0
15	Cancer	III	M0
16	Cancer	III	M0
17	Cancer	III	M0
18	Cancer	III	M0
19	Cancer	III	M0
20	Cancer	III	M0
21	Non-Cancer	Unknown	Unknown
22	Non-Cancer	Unknown	Unknown
23	Non-Cancer	Unknown	Unknown
24	Non-Cancer	Unknown	Unknown
25	Non-Cancer	Unknown	Unknown
26	Non-Cancer	Unknown	Unknown
27	Non-Cancer	Unknown	Unknown
28	Non-Cancer	Unknown	Unknown
29	Non-Cancer	Unknown	Unknown
30	Non-Cancer	Unknown	Unknown
31	Non-Cancer	Unknown	Unknown
32	Non-Cancer	Unknown	Unknown
33	Non-Cancer	Unknown	Unknown
34	Non-Cancer	Unknown	Unknown
35	Non-Cancer	Unknown	Unknown
36	Non-Cancer	Unknown	Unknown
37	Non-Cancer	Unknown	Unknown
38	Non-Cancer	Unknown	Unknown
39	Non-Cancer	Unknown	Unknown
40	Non-Cancer	Unknown	Unknown

Total 2108 proteins were detected in both cancer and non-cancer groups. When categorized by their respective GO term as described above, over 300 proteins were identified directly related to EV, 185 proteins related to cytoskeleton, 76 related to cell surface proteins, 74 related to endosome and total 786 could be found in ExoCarta database (Figure 2.5B). Overall, 654 unique proteins that are "Potential EV" proteins features, 282 proteins could be either EV or free contaminant as "Undetermined" proteins while 146 proteins could be "Potential Contaminant" proteins from cellular organelles that are less likely associated with EV. Over four times of the potential EV proteins were identified compared to the potential non-EV contaminants, which was consistent to result above in the methodology characterization.

In order to obtain the features most likely related to breast cancer EVs, proteomic profiling was performed on the four EVs generated from the breast cancer cell-lines MCF7, MDA-MB-468, MDA-MB-231 and T47D (Figure 2.5C). Total 3653, 5550, 3099 and 4994 proteins were identified in respective cell line generated EVs. 872 proteins are shown to be shared features in all four kinds of EVs.

One thing in common for all four breast cancer cell lines is that they have all shown aggressive, metastasis and invasive behavior in varieties of studies[122]. We hypothesized that there might be some common features related to such cancer cell aggression behavior. If such common features exist in cell line EVs, they might also exist in the breast cancer cells from patients. Ideally, they might also be detectable from the enriched plasma EVs. Therefore, features that in cell line EV would be most likely related to the cancer EVs from patient plasma. By overlapping between the features from cell line EVs and the features from patients' enriched plasma EVs, it showed that 440

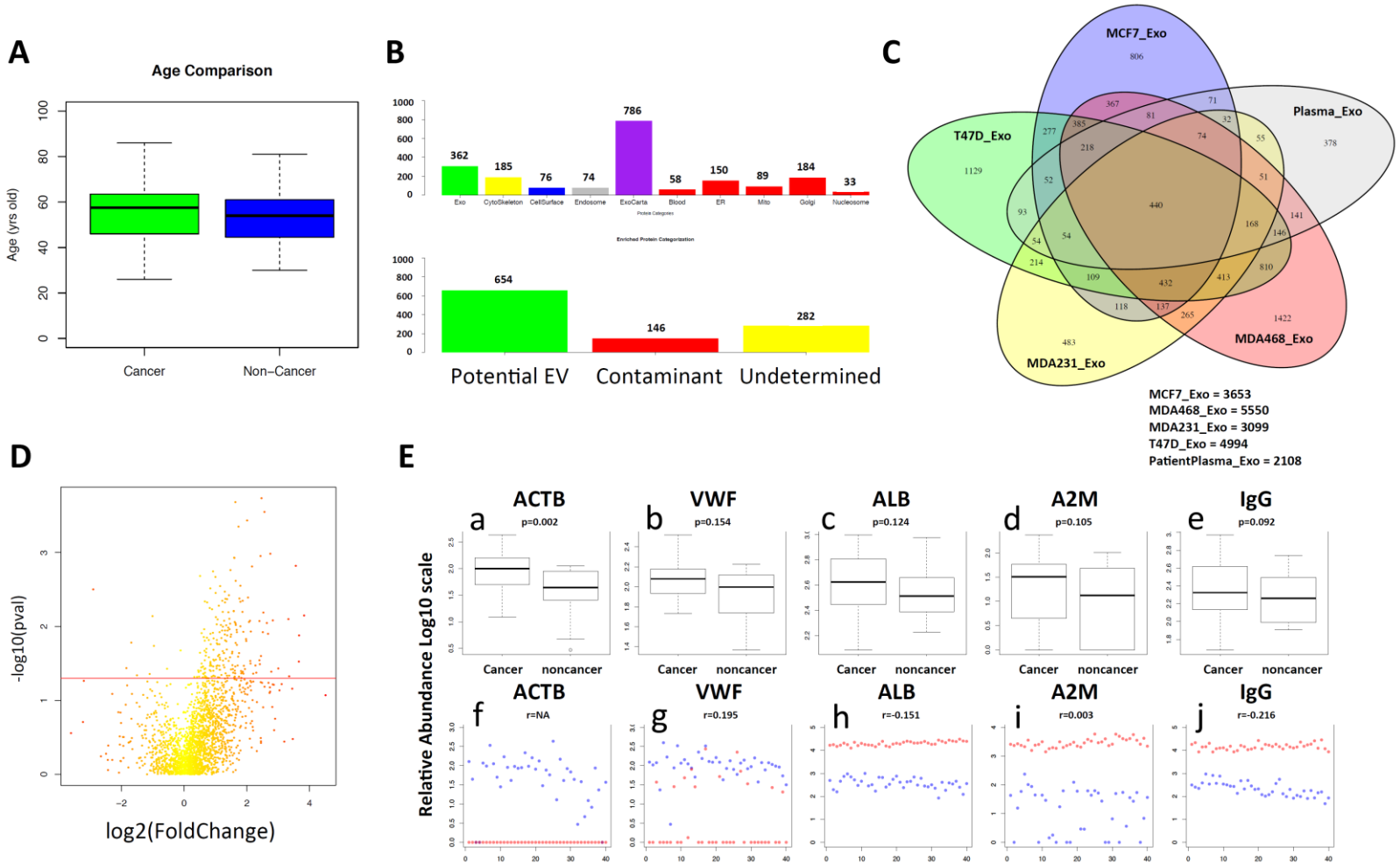
proteins to be shared features between EVs generated. These 440 features were detected in enriched plasma EVs, and then confirmed to be in the four breast cancer cells line EVs.

Figure 2.5D shows the comparison between cancer and non-cancer groups of the 2108 features detected in the patient samples. 217 features show statistically significant difference between cancer and non-cancer groups ($p<0.05$). 197 features show the relative abundance are significant higher in cancer group compared to non-cancer group by equal or more than 1.5 folds ($p<0.05$), while 6 of these features show relative abundance are significantly lower in the cancer group compared to non-cancer group by equal or less than 0.5 folds ($p<0.05$). More features on the side of elevated abundance were consistent to the expectation of EVs enrichment.

General features that found in plasma EVs and high abundant plasma proteins were also compared between cancer and non-cancer groups (Figure 2.5Ea-e), as well as between plasma EV fractions and the corresponding neat plasma samples (Figure 2.5Ea-e). As general EV marker, ACTB showed statistically significant between cancer and non-caner group (Figure 2.5Ea $p<0.05$). In contrast, VWF showed no statistically significant difference between the two groups (Figure 2.5Eb $p>0.05$). High abundant plasma proteins such as ALB, A2M and IgG showed no statistically significant difference between cancer and non-cancer group (Figure 2.5Ec-e, $p>0.05$), indicating the source of high abundant plasma protein contaminations in the enriched EV fraction were quite consistent across groups. At the same time, Figure 2.5Ef-g showed that the general EV marker, the relative abundance of ACTB and VWF were both significantly higher in the enriched EV fraction compared to the corresponding neat plasma. In contrast, all plasma high abundant proteins (ALB, A2M and IgG) showed magnitudes higher in neat plasma

sample compared to the corresponding enriched EV fraction, which was consistent to the previous result that the EVs from the patient plasma were enriched by the F68 copolymer with minimal high abundant plasma protein contaminations. Meanwhile, there was little correlation in the high abundant plasma proteins in the EV fractions with those from the corresponding neat plasma, suggesting trace of plasma leftover should not be the cause of such signal, especially for those proteins that was categorized as "Undetermined" above that showed potential in either as EV or plasma proteins category.

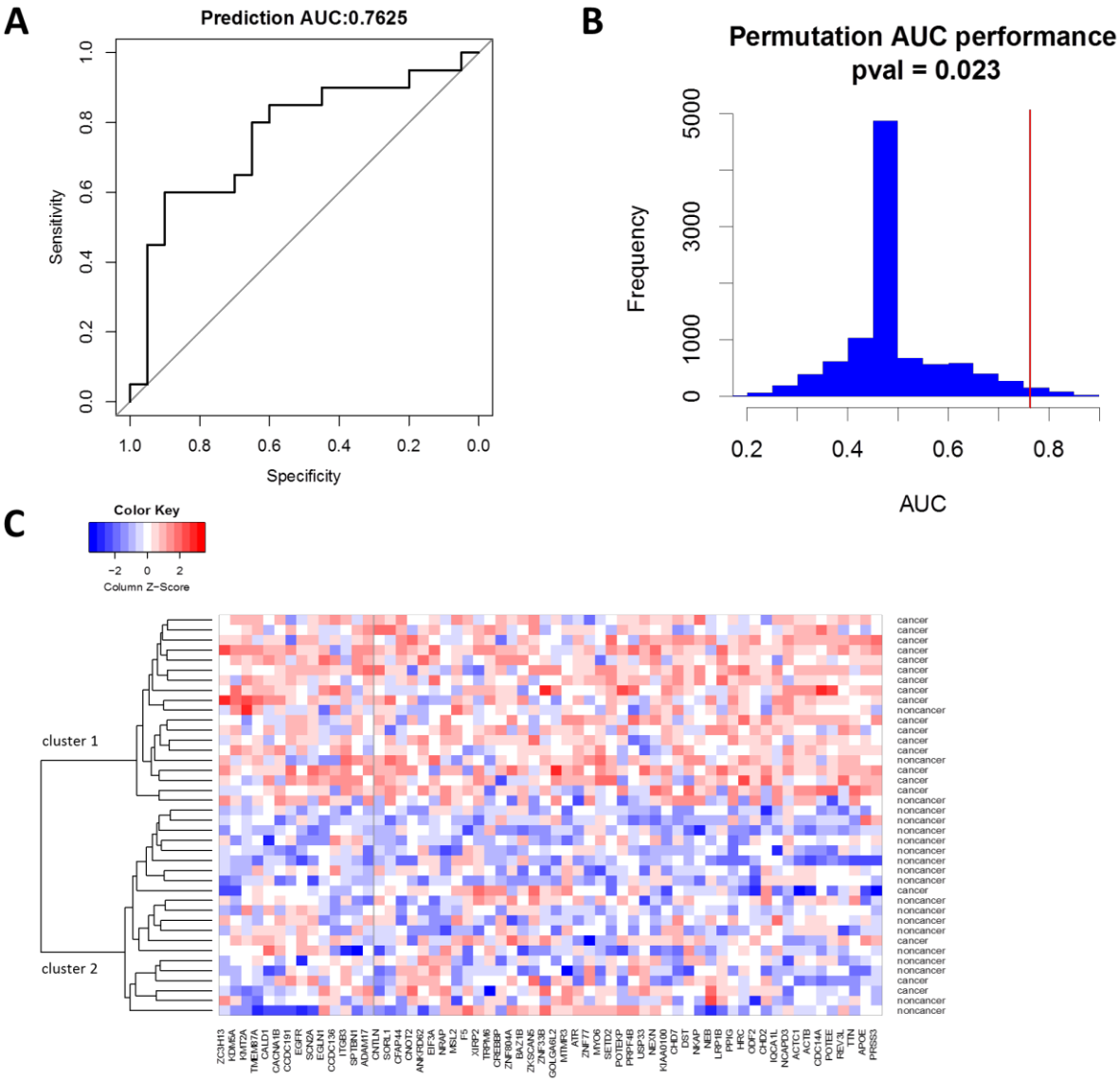
Figure 2.5 Clinical Comparisons between Advanced Breast Cancer and Non-Cancer. A) Age distribution in cancer and non-cancer group in Whiskers box plot; B) The unique proteins identified in the patient sample (protein detected at least in 10 samples in each group) were categorized to extracellular exosome, cytoskeleton, cell surface, endosome, blood macromolecule, endoplasmic reticulum (ER), mitochondrial, Golgi apparatus and nucleosome. “Potential EV”, “Potential Contaminant” and “Undetermined” protein categories were separated as described above by their respective GO term; C) Proteins overlapping between four breast cancer cell lines generated EVs and the clinical patient plasma sample EVs; D) Volcano plot of the cancer vs non-cancer plasma EV samples, the t-test p value (-log10) was plotted against the fold changes of the average relative abundance between cancer and non-cancer plasma EV samples; E) Relative abundance of proteins enriched from plasma EV comparison between cancer and non-cancer group by Whiskers box plot, a:ACTB, b:VWF, c:ALB, d:A2M and e:IgG. Comparison of relative protein abundance from enriched plasma EV (blue) to the neat plasma(red) in dot plot from individual patients, f:ACTB, g:VWF, h:ALB, i:A2M and j:IgG. Pearson correlation between plasma EV and neat plasma were also calculated. Y axis is the log10 transformed relative abundance.


A
Age Comparison
Age (yrs old)
Cancer
Non−Cancer
B
362
185
76
74
786
58
150
89
184
33
Exo
CytoSkeleton
CellSurface
Endosome
ExoCarta
Blood
ER
Mito
Golgi
Nucleosome
Protein Categories
Enriched Protein Categorization
654
146
282
Potential EV
Contaminant
Undetermined
C
MCF7_Exo
806
Plasma_Exo
378
T47D_Exo
1129
MDA468_Exo
1422
MDA231_Exo
483
440
MCF7_Exo = 3653
MDA468_Exo = 5550
MDA231_Exo = 3099
T47D_Exo = 4994
PatientPlasma_Exo = 2108
D
-log10(pval)
log2(FoldChange)
E
Relative Abundance Log10 scale
ACTB
p=0.002
VWF
p=0.154
ALB
p=0.124
A2M
p=0.105
IgG
p=0.092
Cancer
noncancer
ACTB
r=NA
VWF
r=0.195
ALB
r=-0.151
A2M
r=0.003
IgG
r=-0.216

In order to assess the potential classification/prediction power by the proteomic profiles from the enriched plasma EVs, a Leave-One-Out (LOO) cross-validation model was built. The prediction performance in term of the area under curve (AUC) value was obtained. Figure 6A show the overall AUC performance is 0.7625 with optimal specificity of 0.9 and sensitivity about 0.6. To examine whether such performance was due to a random chance, AUC performance was calculated by randomly permuting the sample labels 10000 times. The distribution of AUC performance from the permutations is shown on Figure 2.6B. Compared to the permutation AUC performance, the AUC 0.7625 obtained from the correct sample label is statistically significant (Figure 2.6B, $p<0.05$), indicating the classification performance between the original cancer/non-cancer group was not due to a random chance, suggesting the features from enriched plasma EVs was capable of differentiating the advanced breast cancer patient from non-cancer controls.

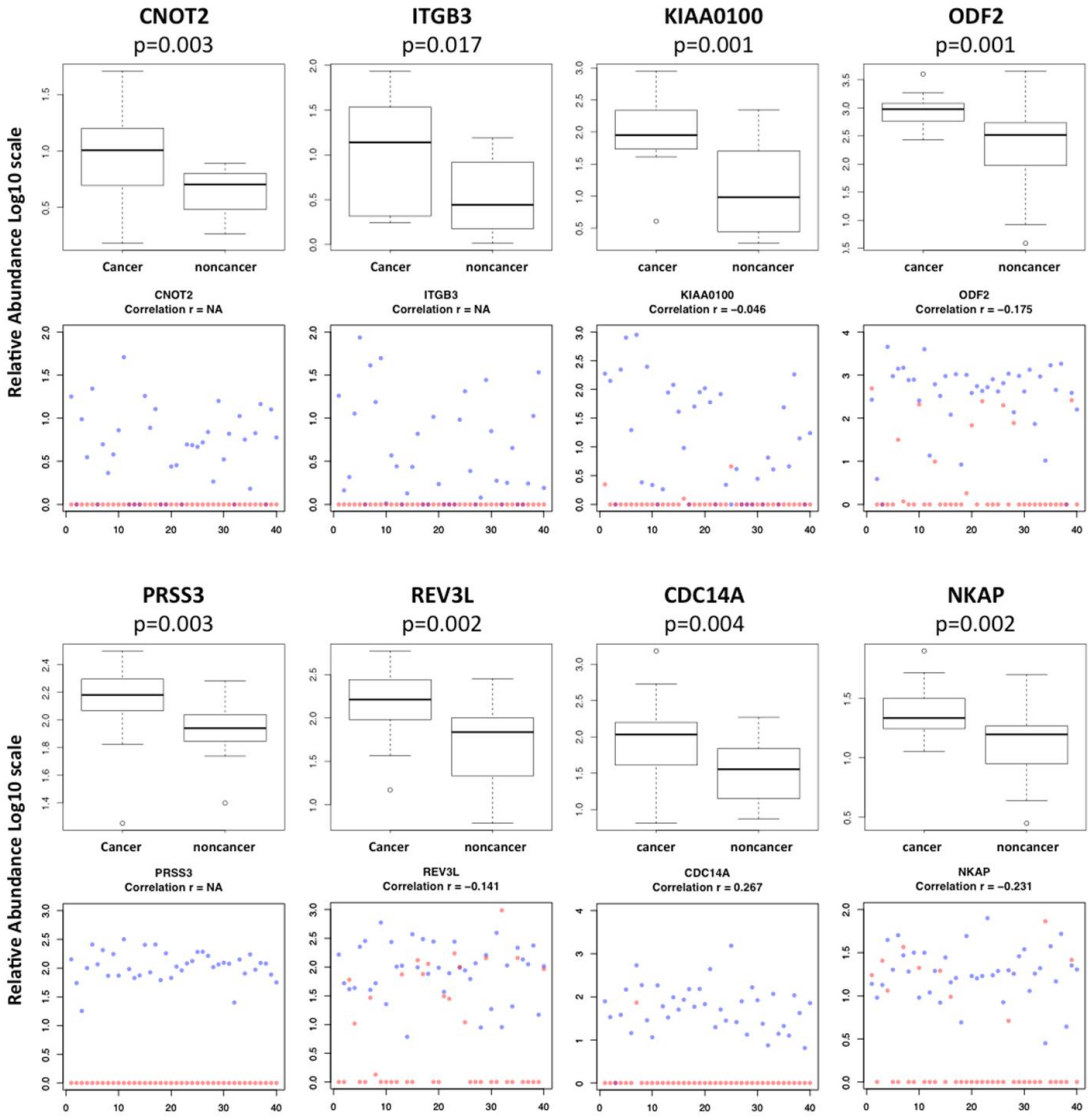
60 features were found consistently contributed to the model in differentiating cancer from non-cancer sample. An unsupervised hierarchical clustering was performed to classify the current patient samples by those features. As shown in Figure 2.6C, two clusters were clearly formed, one with mostly cancer sample (16 cancers vs 3 non-cancers) and the other is mostly non-cancer (17 non-cancers vs 4 cancers), with sensitivity of 80% and specificity of 85%, confirms the above feature selection process was able to screen features that separates advanced cancer patients from non-cancer controls.

Figure 2.6 Classification Performances and Feature Selection. A) Classification performance of the semi-quantitative mass spectrometry data was evaluated by a LOO cross-validation model based on random forest. The overall performance is AUC 0.7625; B) Classification performance from random permutations. Distribution of the AUC performance from all permutations was shown as blue, red vertical line marks the AUC performance of the correct sample label. The p value was estimated using the percentage of permutations with equal or larger AUC performance compared to the original AUC; C) Non-supervised hierarchical clustering of the patient samples by the features selected from the prediction model above. Two clusters were clearly formed in the patient samples, cluster 1 contains 16 cancers and 3 non-cancers (80% sensitivity), cluster 2 contains 4 cancers and 17 non-cancers (85% specificity).


A
Prediction AUC:0.7625
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
B
Permutation AUC performance
pval = 0.023
Frequency
0
1000
3000
5000
0.2
0.4
0.6
0.8
AUC
C
Color Key
−2
0
2
Column Z-Score
cluster 1
cluster 2
cancer
noncancer

To further illustrate that the signal detected in the plasma EVs for the selected features was not from the trace of leftover plasma contaminations, Figure 2.7 shows more of those features with the most significant difference between advanced breast cancer and the non-cancer patients, as well as the comparison of their signals in plasma EVs to the corresponding neat plasma. Not only did these features show significantly higher relative abundance in the advanced breast cancer group compared to the control group ($p<0.05$), but they were also over expressed in the enriched plasma relative to neat plasma. Little correlation between the signals in enriched plasma EVs and the neat plasma, indicating the selected features were most likely to be enriched by the proposed procedure instead from trace of the plasma leftover.

Figure 2.7 Features That Contributed in Breast Cancer Classification. Top features selected from the comparison (cancer vs non-cancer, plasma EVs vs neat plasma). Whisker Boxplot comparison between cancer/non-cancer; dot plots comparison of individual enriched plasma EV and their corresponding neat plasma (blue: plasma EV, red: neat plasma).



Discussion

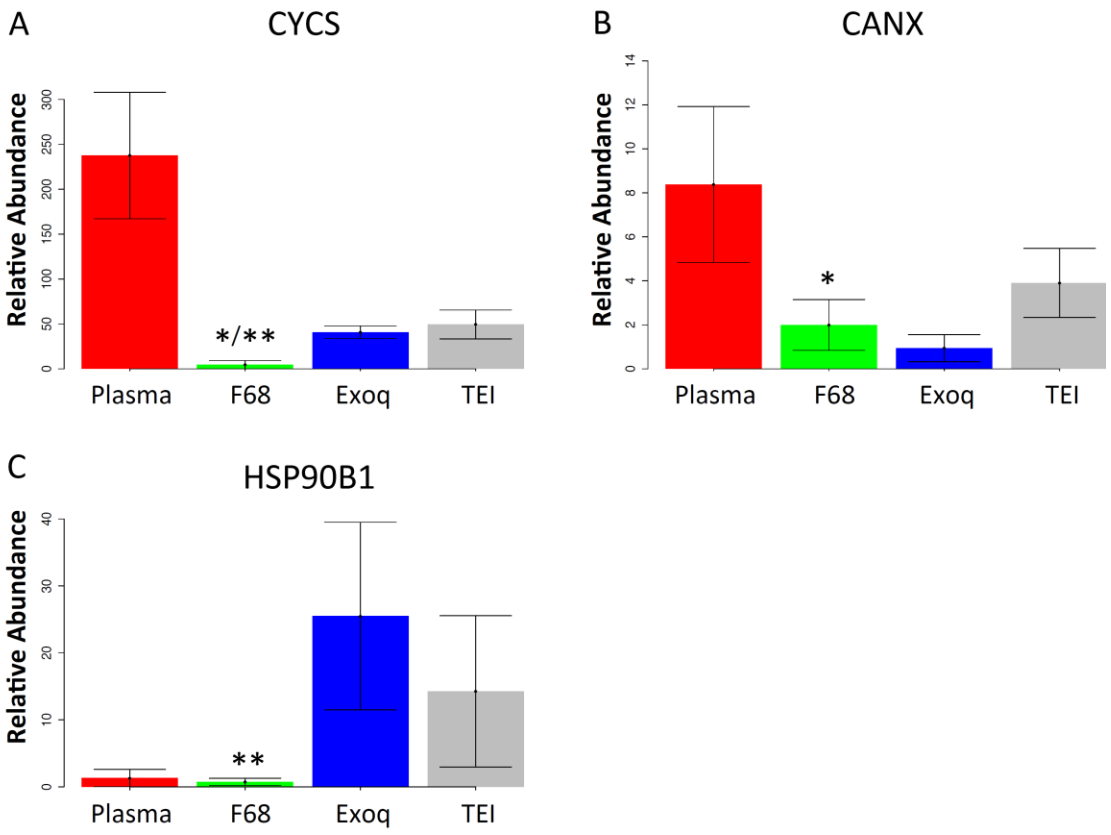
Secreted EV encompasses a very rapidly growing scientific field in biology and medicine. According to International Society for Extracellular Vesicles (ISEV), it is yet technically challenging to obtain a totally pure EV fraction free from non-vesicular components for functional studies. Though EVs are proved to contain proteins, lipid, mRNA and non-coding RNA species, the field of EV has not yet matured to the point that a list of "specific markers" could be given so that the EV population could be clearly separated from each subpopulation as well as the contaminants[29].

An ideal plasma EV enrichment method should be that the EV fraction is enriched with high efficiency while containing as less as possible of the plasma protein contaminations. In terms of the efficiency, reproducibility and simplicity, precipitation-based methods are more competitive compared to UC or IP based methods. That in turn, also depends on whether the selectivity of the precipitation and compatibility the enriched fraction to downstream applications. There is no single platform or characterization yet, by its own, sufficient to indicate the enrichment of EV. Multiple platforms and characterization might be more appropriate in evaluating the quality of the enrichment. In our study, we showed that the Pluronic copolymer F68 was able to enrich the plasma EVs with membrane structure, ranging from 20nm to 100nm in size, contained highly enriched set of EV-related proteins as well as varieties of RNA species. At the same time, much less enrichment in high abundant plasma proteins contaminations was detected in the enriched EV fraction. The enriched EVs were also proven to have broad compatibility for variety of downstream applications including TEM, FL and MS without any additional procedures to remove the copolymer. According to the minimal experimental

requirements for definition of EVs, at least 3 proteins should be reported in at least semi-quantitative manner in any EV preparation[29]. Here we reported not only the higher enrichment of EV related protein, including CD9, ACTB, HSP70 and VWF, but also lower the potential high abundant protein contaminants (ALB, A2M, and IgGs) from plasma by two semi-quantitative methods. Shotgun proteomic further found that the method enriched more potential EV proteins than others. However, since unique features were still found in different enrichment methods, the possibility that the existence of such EV subpopulation that might only be enriched by specific method/reagent, could not be excluded.

In another remark, according to the minimal experimental requirements for definition of EVs, markers that are supposed to be absent or under-represented may also necessary to evaluate the purity of the enriched EVs[29], semi-quantitative comparison for couple of such markers were shown in Figure 2.8A-C, including CYCS (mitochondria), CANX (ER) and HSP90B1 (ER). Compared to other methods, F68 precipitation was the only method that consistently showed significantly lower enrichment of these mitochondria/endoplasmic reticulum protein markers in comparison to the corresponding neat plasma as well as to the other precipitation-based methods. However, it is worth to note that the meaning of interpreting absence or under-represented marker that is already absence or under-represented in plasma might be limited and subjected for debate, partially because lack of a standard on what should be considered as normal for such markers in plasma sample as well as incomplete understanding of the cellular compartmental communications.

Figure 2.8 Markers Considered Not Related to EVs from Different Plasma EV Enrichment Protocols. Semi-quantitative mass spectrometry analysis of proteins expected to be absence or under-presented in EVs, comparison of relative abundance of CYCS, CANX and HSP90B1, * significant lower compared to original plasma ($p<0.05$); ** significant lower compared to Exoquick/TEI methods ($p<0.05$).



Almost half of the proteins that are highly enriched by F68 method (122 out of 263) were not able to match the categories described in Figure 2.3C, further studies might be necessary to illustrate whether these proteins are cargos that are packaged by either specific or non-specific means into EVs other than through the EV biogenesis pathway. Interestingly, most of these proteins belong to three sub-categories, including integral components of membrane, cytoplasm (cytosol), and nucleus proteins. The integral components of membranes include those proteins that might be involved in interactions with the membrane bound receptor signaling cascades, such as CCR8, CYSLTR2, GPR151, ABCG5. A lot of proteins are also identified as cytoplasm/cytosol and nucleus proteins with RNA/DNA binding capability, including YTHDF1, CSRNP1, PHF21A, RBM42, ZFAT, TLE1, HMBOX1, AICDA, KLF17, which are involved in varieties of functions from potential translation regulation, tumor suppressor to transcription factor activity. The findings are consistent to the study that reveals transcriptional regulator proteins were highly abundant in EVs[123], the potential biological implication of such factors being highly enriched in the plasma EVs might deserve more investigations.

Although it has been shown that variety of RNA species exist in EVs[116], profiling of RNA/smRNA from plasma EVs has been challenging. The profile was highly dependent on the method used in preparing the EV fraction. This was further complicated by the fact that smRNAs are well-known to circulate in the bloodstream and other body fluids in a stable, cell-free status[124], resulted in the RNA profiles may varies by different methods and may contains undistinguishable contaminants in various degrees. Despite that, with highly enriched EV related protein and much less enrichment for high abundant

plasma protein in the current method, supporting improvement could be achieved in the current study.

In a broader view, the RNA/smRNA profiles for plasma EVs enriched by F68 are consistent to other PEG-based enrichment methods from other studies (either from Exoquick or from TEI kit preparations)[116, 125]. miRNA, piRNA, tRNA, tRFs, Y_RNA, RNY, MtRNA, protein coding RNA, rRNA, lincRNA as well as snRNA and snoRNA were all detectable. Interestingly, large amount reads were mapped to mtRNA. There might be two explanations: 1) small mitochondrial debris might still be co-precipitated with the copolymer, which is supported by the existence of small portion of particles with larger size, even in the UC cell line EV preparations; 2) alternatively, a recent study showed that mitochondria are also able to generate small vesicular carriers that transport mitochondrial contents to other intracellular organelles. One of the targets of such transportation from mitochondria was the late endosome, or multi-vesicular body, where the EVs were formed[17], suggesting potential communication between mitochondrial and endosome, or maybe MVB as a result. This might be one of the reasons for such high level of mitochondrial RNA in the plasma EV fraction.

To explore the potential application for this enrichment method, proteomic profiling was performed and used to compare plasma EVs enriched from advanced breast cancer and non-cancer controls. The breast cancer plasma sample were all from advanced clinical stage (III and IV), the reasons of using such settings are in two folds: 1) According to the definition of breast cancer stages (https://www.cancer.net/cancer-types/breast-cancer/stages), major character of clinical stage III and IV is that cancer of any size has spread to 4 to 9 axillary lymph nodes or has spread to the chest wall or

caused swelling, or ulceration of the breast, as well as further spread to other organs. Symptoms like these might suggest that cancer cells at these stages are actively spreading to distal tissue and maybe constantly in contact with the blood stream, implying cancer cell generated EVs may be more likely expelled into the blood stream. Therefore, proteins related to cancer aggression behavior at such advanced stages might be more likely to be captured from plasma EVs; 2) by adopting such setting, features selected from the model were expected to be related to the cancer aggression, metastasis, invasion or other characters that are generally associated with such advanced stages. Therefore, it provided a clear direction for potential result interpretation.

Feature selection process in the study also focused on those features that were shared between enriched plasma EVs and breast cancer cell line EVs. This is because the purity of the enriched EVs is yet to be perfect even the enrichment method had been improved. Therefore, without such screening, it might be difficult to ensure the authenticity for such association of the features to EVs. Secondly, EVs in plasma may come from variety of sources but not necessary related to breast cancer cells. By using the features that could be detected from EVs generated by breast cancer cell lines only, the chance of the selected features that are actually from breast cancer related sources could be significant increased. Combined with these two filters, the resulted features may have a better chance of actually associating with breast cancer at that stage.

Characterization of the four different breast cancer cell lines MCF7, MDA-MB-231, MDA-MB-468 and T47D was previously reported[122]. Though genotypes, morphologies as well as growth rates might be different for these cell lines, one character in common for these cell lines is that they were all derived from cancer

metastasis/invasion sites[126-129] and they are all able to show aggression, metastasis and invasion behavior in different models[122, 130], implying certain common features might be involved in regulating such aggression/metastasis/invasion behavior. One of the expectations for the clinical comparison study was to find out whether such common features related to cancer aggression could be found in the plasma EV and whether these features could be used to differentiate the cancer patient from the non-cancer patients.

In the LOO cross-validation model, 60 core features were selected. They are features that are detected in both patient plasma EVs and the breast cancer cell lines generated EVs. These features were able to separate the advanced cancer sample population from non-cancer patient in both the cross-validation model and the un-supervised hierarchical cluster method. Interestingly, according to the information from The Human Protein Atlas database[131-133], as shown in Table 2.2, the expression of over 70% of these features (43 out of 60) have been confirmed by immunohistochemistry on the breast cancer tissues, only 17 proteins are yet to be confirmed, partially because lacking of suitable antibodies, consistent to our expectation by screening the initial plasma EV features with features from EVs generated by breast cancer cell lines.

Also according The Human Protein Atlas database, the prognostic performances of these 60 features were also analyzed by respective RNAseq dataset generated by TCGA[134]. The analysis was based on the FPKM value of each gene, patients were classified into two expression groups and the correlation between expression level and patient survival was examined by Kaplan-Meier survival estimators, and the survival outcomes of the two groups were compared by log-rank tests (https://www.proteinatlas.org/). Almost half of those proteins (28 out of 60) features

show significant prognostic values on varieties of cancers, implying their potential relevance to breast cancer.

Also, through literature searching, 28 out of 60 features show potential association to breast cancer aggression/metastasis/invasion (Table 2.2) from previous studies. For example, CDC14A was shown to be recruited to the cell leading edge to regulate cell migration and adhesion in breast cancer cells[135]; CNOT2 was shown to promote proliferation and angiogenesis via VEGF signaling in breast cancer cells[136]; Downregulation of ITGB3 integrin modulates cell adhesion and invasion by interrupting Erk/Ets1 network in triple-negative breast cancer[137]; Tumor but not stromal expression of ITGB3 is essential and required early bone-metastatic breast cancer[138]; High levels of KIAA0100 expression were associated with poor prognosis in patients with invasive ductal carcinomas of the breast[139]. Such large proportion of the final selected features shown related to cancer aggression, metastasis and invasion was consistent to our initial expectation.

Table 2.2 Summary of 60 Features Selected from the LOO Model. Prognostic potential by their RNA expression level in cancer tissue from TCGA database, analyzed by The Human Protein Atlas database (NA: not available); Protein expression of each protein from breast cancer tissues from The Human Protein Atlas database (NA: not available). The literature references of association of these proteins to breast cancer aggression, metastasis and invasion. The p value from comparing plasma EV samples from advanced breast cancer and non-cancer group.

Gene Name	Prognostic Value in Cancer	IHC expression breast cancer tissue	References to Breast Cancer	p vaule (t-test)
ACTB	Renal, Head and Neck Cancer	Yes	[140]	**0.002**
ACTC1	Head and Neck, Urothelial Cancer	Yes	[141]	**0.001**
ANKRD62	NA	NA		**0.004**
APOE	NA	Yes	[142]	**0.019**
ATR	Pancreatic, Liver Cancer	NA		**0.004**
CDC14A	NA	Yes	[135, 143-144]	**0.004**
CFAP44	Urothelial Cancer	NA		**0.005**
CHD2	NA	Yes	[145]	**0.009**
CHD7	Endometrial Cancer	Yes	[146]	**0.02**
CNOT2	Renal, Liver, Melanoma Cancer	Yes	[136, 147]	**0.003**
EGFR	Urothelial Cancer	NA	[148]	**0.003**
HRC	NA	NA		**0.003**
ITGB3	NA	NA	[137-138, 149-150]	**0.017**
LRP1B	NA	Yes		**0.002**
MTMR3	NA	Yes	[151]	**0.009**
NKAP	Breast Cancer	Yes	[152-153]	**0.002**
ODF2	Liver, Prostate Cancer	Yes		**0.001**
POTEE	NA	Yes	[154]	**0.008**
POTEKP	NA	NA		**0.003**
PPIG	Renal Cancer	Yes		**0.002**
PRSS3	Endometrial Cancer	Yes	[155-157]	**0.003**
REV3L	NA	Yes	[158-159]	**0.002**
SCN2A	NA	NA	[160]	**0.003**
SPTBN1	Renal Cancer	Yes	[161-162]	**0.003**
TMEM87A	NA	Yes		**0.019**
XIRP2	NA	Yes		**0.013**
ZNF33B	NA	Yes		**0.006**
ZNF804A	NA	Yes		**0.002**
PRPF4B	Urothelial, Liver Cancer	Yes		**0.006**
BAZ1B	NA	NA		**0.004**
CCDC191	Urothelial Cancer	Yes		**0.013**
CNTLN	NA	Yes		**0.011**

KIAA0100	Liver Cancer	Yes	139, 163	**0.001**
NEB	NA	NA		**0.024**
TRPM6	NA	NA		**0.017**
ZKSCAN5	Liver, Thyroid Cancer	Yes		**0.024**
ZC3H13	renal Cancer	Yes		**0.063**
DST	NA	Yes	164-166	**0.008**
EGLN1/PHD2	Cervical Cancer	Yes	167-169	**0.026**
KDM5A	NA	NA	170-172	**0.02**
NCAPD3	NA	NA		**0.011**
NEXN	Colorectal Cancer	NA		**0.011**
SETD2	Renal, Melanoma Cancer	Yes	173-174	**0.022**
ADAM17	Urothelial Cancer	Yes	175-177	**0.043**
CACNA1B	NA	Yes	178	**0.035**
GOLGA6L2	NA	NA		**0.018**
IQCA1L	NA	NA		**0.024**
TTN	NA	Yes		**0.02**
F5	Stomach Cancer	Yes		**0.011**
EIF3A	NA	Yes	179	**0.019**
MYO6	Renal Cancer	Yes	180-181	**0.05**
CALD1	Renal, Melanoma Cancer	Yes	182	**0.023**
NRAP	NA	Yes		**0.058**
CREBBP	Renal Cancer	Yes		**0.023**
CCDC136	NA	NA		**0.017**
MSL2	Cervical Cancer	Yes		**0.015**
KMT2A	NA	Yes	183	**0.056**
ZNF77	Lung, Endometrial, Cervical Cancer	Yes		**0.053**
USP33	NA	Yes	184	**0.044**
SORL1	Renal Cancer	Yes		**0.024**

It is worth emphasizing that even though the EV fraction precipitated by the copolymer method was much cleaner, the possibility of contamination from certain amount plasma or other organelle debris could not be completely excluded, as they might be indistinguishable from EV in term of size as well as membrane properties. Indeed, our result showed the presence of a small percentage of larger vesicles and enrichment of some non-EV related organelle proteins. On the other hand, lack of universal EV unique marker might raise some questions regarding to the categorization accuracy, as proteins might locate in multiple organelles. For example, EVs may even contain DNA that bound by nucleus proteins[16]; mitochondrial-derived vesicles (MDVs) was also able to fused with the MVBs[17]; HSP60, a mitochondrial heat-shock protein, was found in tumor cell plasma membrane, EV membrane as well as in the Golgi apparatus[18]. Therefore, as more studies come out, more explicit definition would be expected. Despite of these concerns, the simple procedure, high enrichment, and much less high abundant plasma protein contamination in the current method might offer more advantages.

CHAPTER 3

KIAA0100 MODULATES BREAST CANCER CELL AGGRESSION THROUGH INTERACTION WITH MICROTUBULE AND HEAT SHOCK PROTEINS

Abstract

KIAA0100 gene was isolated from human immature myeloid cell line cDNA library twenty year ago. Recent studies have shown that its expression is elevated in breast cancer and associated with more aggressive cancer types as well as poor outcomes. However, its cellular and molecular function is yet to be understood. Here we showed that silencing KIAA0100 by siRNA in breast cancer cell line MDA-MB-231 significantly reduced the cancer cells aggression behavior, including cell aggregation, reattachment, cell metastasis and invasion. Most importantly, silencing the expression of KIAA0100 sensitized the quiescent, suspended cancer cells to anoikis. Immunoprecipitation, mass spectrometry and immunofluorescence analysis revealed that KIAA0100 may play multiple roles, including stabilizing microtubule structure as a microtubule binding protein, and contributing to cancer cell anoikis resistance by its interaction with stress protein HSPA1A. Our study also implies that the interaction between KIAA0100 and HSPA1A may be targeted for new drug development to specifically induce anoikis cell death in cancer cell.

Introduction

Human KIAA0100 protein belongs to the Human Unidentified Gene-Encoded (HUGE) database, which consists of over 2400 novel large human protein genes that have multiple domains potentially capable of binding many kinds of partners[185]. It was identified from a cDNA library in the Kazusa cDNA sequencing project in 1995[186].

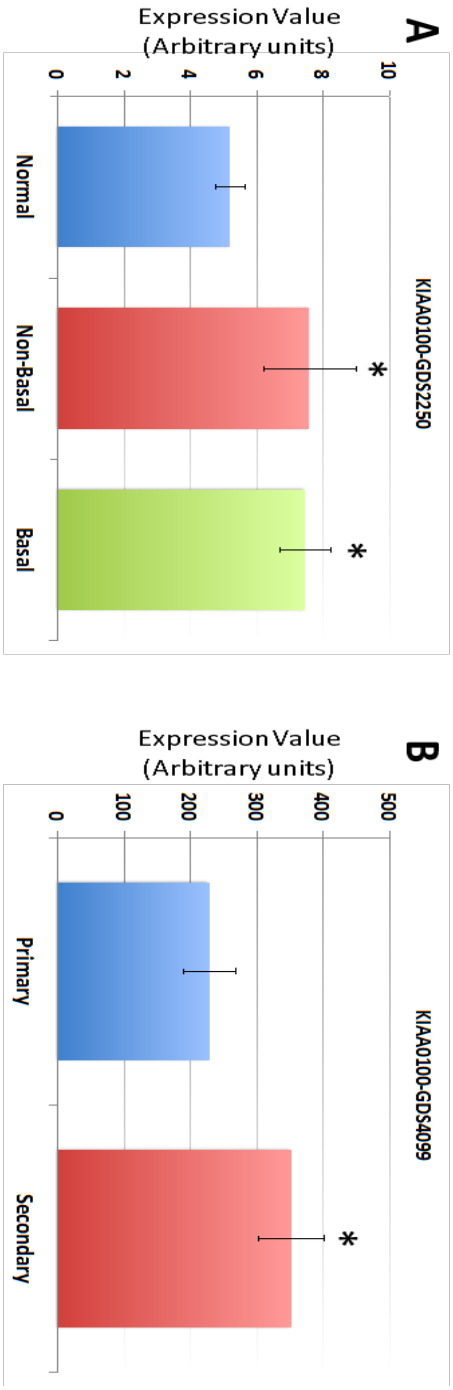
In previous studies, it has been shown that among 71 cancer type-specific tags, 28 were specific tags for breast carcinomas[187]; one tag of "GGTCCCCTAC" has been demonstrated to be highly expressed in the SAGE libraries of breast carcinomas and breast cancer cell lines compared to the libraries from normal mammary epithelium. The reliable UniGene cluster that matched to the tag of GGTCCCCTAC was identified to be Hs.151761, which was corresponding to the gene of KIAA0100[163]. In the most recent EST report, KIAA0100 was reported to have a total 448 EST sequences from a variety of human tissues, with 26 of the ESTs confirmed to be in mammal grand tissue (NCBI Unique Gene:KIAA0100).

Interestingly, data from multiple studies have shown that KIAA0100 might be related to breast cancer: in the non-cancerous human mammary breast epithelial cell line MCF10A transformed by v-Src, KIAA0100 protein was significantly up-regulated in response to the malignant transformation by proteomic profiling[188]. Genomic location of KIAA0100 (17q11) was also found to be within a close proximity to 17q12 chromosomal region. Amplification of this region was found in approximately 25% of breast tumors, which was associated with poor prognosis[189], implying the expression of KIAA0100 may be affected if such event occurs; Also both ERRα and ER-α, were found to be recruited to the promoter region of KIAA0100 in the mouse model of ERBB2-initiated mammary

tumorigenesis[190], implying the expression of KIAA0100 may be potentially up-regulated through these factor in breast cancer. High level of KIAA0100 expression was also shown to be associated with poor prognosis in patients with invasive ductal breast carcinomas[139].

Our recent data-mining from NCBI Gene Expression Omnibus (GEO) database also revealed a compelling expression pattern of KIAA0100 in breast cancer patients as well as in *in-vivo* tumor models: the expression level of KIAA0100 was significantly elevated in both basal-like and non-basal like breast cancer compared to normal controls [191-192] (Figure 3.1A, GDS2250), suggesting its involvement might be necessary in both breast cancer types. In a mouse HER2 positive breast cancer model, the secondary tumor showed significantly higher expression of KIAA0100 compared to the primary tumor[193] (Figure 3.1B GDS4099), indicating its expression may be associated with the increasing cancer cells aggressive. Meanwhile, multiple bioinformatics tools have been employed to predict potential functions of KIAA0100 and show that it might be an anti-apoptotic factor related to carcinogenesis or progression[186, 194]. However, the molecular and cellular functions that KIAA0100 plays and how it contributes to cancer development and progression, especially in breast cancer cells, remain elusive.

Figure 3.1 Expression of KIAA0100 in GEO Datasets



In the previous chapter, plasma EV profiles comparison between advanced breast cancer and non-cancer patients showed that large portion of selected features were related to cancer aggression, metastasis and invasion. Considering that all current EV preparation methods are not perfect, the question becomes whether these features are really attributed to breast cancer development and aggressive behavior, and whether there is a link between these features and EVs. KIAA0100 was one of those features that showed significant higher level in the advanced cancer patient compared to the non-cancer controls. While most of those selected features have been extensively studied, the cellular and molecular functions of KIAA0100 have not yet been investigated, though multiple studies have shown its potential association to the breast cancer aggression. Therefore, functional study of KIAA0100 not only would be able to strengthen the conclusion in chapter 2, but also potentially may lead to the discovery of new drug targets to treat aggressive breast cancer.

Cancer cell aggression can be exhibited in a variety of ways, including cell proliferation/growth[195]. However, other aggressive behavior, such as cell anchorage/re-attachment[196], cell adhesion/aggregation[197-198], anoikis resistance[52], a form of apoptosis after the cells' detachment from the extracellular matrix (ECM), and metastasis/invasion[199], all contribute in demonstrating the aggressive nature of the breast cancer cells. In the current study, we adopted siRNA technology to knock down the expression of KIAA0100 in MDA-MB-231 cells, a highly aggressive triple negative breast cancer cell line[122, 200], as a model to study its potential molecular and cellular roles associated with aggressive behavior of breast cancer cells. HEK293 over-expressing KIAA0100 recombinant protein was also employed as an additional cell-line model to

investigate its potential protein-protein interactions accounts for the underlying molecular mechanisms.

Materials and methods

Cell Cultures

MDA-MB-231 was maintained in L15 medium, supplemented with 10% FBS, 2 mM L-glutamine, 1U/ml penicillin, and 1μg/ml streptomycin at 37°C and 0% CO2. HEK293 cells were maintained in DMEM with 2 mM L-glutamine, 1U/ml penicillin, and 1μg/ml streptomycin and 10% of FBS with 5% CO2. pKIAA0100 (pKIA) expression vector with FLAG and Myc tag on its C-terminal was from Origene. HEK293 were transfected with pKIA by Transfectamine 3000 reagent (ThermoFisher) according to the manufactory's instruction. Stable cell line HEK293/pKIA was established by screening the transfected cells with 300ug/ml Geneticin (ThermoFisher). For cell culture in suspension, the culture plates were coated with poly-HEMA as described before[201].

siRNA Transfection, Demecolcine Treatment

Gene specific KIAA0100 siRNA (siRNA id: 219429) targeted exon 25, HSPA1A siRNA (siRNA id: 145248) targeted to exon 1, as well as the negative siRNA control that target no gene product in human sequences (AM4635) were from ThermoFisher. Transfection of the siRNA was performed by the RNAiMAX transfection reagent from ThermoFisher according to the manufactory's instruction. Transfections were performed in either forward or reverse style. For forward transfection, cells were seeded on the culture plate and incubated overnight to let cells settle down and attached to the culture plate surface before transfection; for reverse transfection, transfection mixture was prepared and mixed with cells during seeding to the culture plate coated with poly-HEMA to prevent cell from re-attaching the surface.

Demecolcine (Sigma) was dissolved in DMSO. HEK293 and HEK293/pKIA cells were seeded onto the 96 well tissue culture plates with 1, 10, 100, 1000ng/ml (2.7, 27, 270, 2700nM) final concentration of Demecolcine. Control for 0ng/ml concentration was equal amount of DMSO instead. Cell viability were performed as described below at 0, 24, 48 and 72 hours as described below and normalized to the 0ng/ml control at each time point as 100% viability. For MDA-MB-231 cells, treatments of Demecolcine were at 0, 0.5, 5 and 50uM concentrations and viability were assessed at 24 and 48 hours after treatment.

Cell Growth/Viability, Attachment/Anchorage and Invasion/Metastasis assay

Cell proliferation/growth and viability were tested either by ViaCell instrument (Viacord) or multiTox-Fluor cell viability assay (Promega) according to the manufactory's instruction; Cell attachment/anchorage assays were performed by seeding the floating cell culture onto the culture plate, 12 hours after the seeding, cells not able to attach/anchor to the plate surface were removed by gently washed with warm complete L15 medium, cell growth and viability were then tested as described; Cell invasion/Metastasis was examined by the CultreCoat 96 Well BME-Coated Cell Invasion Optimization Assay (R&D systems) according to the manufactory's instruction.

RT-qPCR, Western blot and Immunoprecipitation (IP)

For RNA expression analysis, cells were lysis by Trizol-LS reagent (ThermoFisher). RNAs were then purified by the RNA minElute kit from Qiagen. Reverse transcription was performed according to the Superscript III RT kit manual from

ThermoFisher with poly-T primer. Taqman probe assay for KIAA0100 and β-actin were from ThermoFisher performed accordingly, relative expression abundances were analyzed by ΔΔCt method. For protein analysis, cells were also lysis by Cell Extraction buffer with protease inhibitor mixture (ThermoFisher), protein concentration was measured by microBCA assay (ThermoFisher). Equal amount of proteins was then separated on either Tris-Acetate or Bis-Tris SDS-PAGE and transfer to the PDVF membrane. The blot was then detected by anti-KIAA0100 antibody (ThermoFisher) and anti-β-Actin antibody (Abcam).

For protein crosslinking, HEK293/pKIA cells were washed by PBS followed by incubating with 6mM of DTBP (ThermoFisher) in PBS for 5min. Cells were then lysed in NP40 cell lysis buffer (ThermoFisher). Anti-FLAG monoclonal antibody (IgG1, Sigma) was used to capture the recombinant KIAA0100 protein; anti-HSPA1A monoclonal antibody (IgG1, Abcam) was used to capture native HSPA1A; anti-Digoxigenin (DIG) monoclonal antibodies (IgG1, Abcam) were used as an isotype antibody control and bare beads were used as negative control. Equal amount of cell lysate was immune-precipitated by the respective antibodies by Dynabead protein-G immunoprecipitation kit (ThermoFisher) according to the manufacturer's instruction. At the end of the immunoprecipitation, the beads were mixed with LDS buffer with reduced reagent and heated at 70°C for 5min. Protein samples were then separated on a 4-12% Bis-Tris SDS-PAGE PAGE gel and transfer to PDVF membrane. The blot was then probed by anti-Myc tag antibody (Abcam), anti-α-Tubulin antibody (Abcam), anti-HSPA1A (Abcam), HSP90AB1 antibody (Abcam) respectively. The mouse monoclonal

antibody used in the immunoprecipitation was detected by anti-mouse-HRP antibody (Santa Cruz).

Mass Spectrometry Analysis

Protein samples were separated on 4-12% Bis-Tris PAGE gel and stained with Brilliant Blue (ThermoFisher); slices containing the protein bands of interested were excised. In-gel digestions by trypsin (ThermoFisher) were then performed. Samples were analyzed by nanoflow reverse phase liquid chromatography using a Dionex Ultimate 3000 RSLCnano System (ThermoFisher) coupled in-line to a Q Exactive HF mass spectrometer (ThermoFisher). The nano LC system included an Acclaim PepMap 100 C18 5μm 100A 300μm×5mm trap column and an EASY-Spray C18 2μm 100A 50μm×150mm analytical column (ThermoFisher). Peptide samples were eluted with a two-step gradient of 2% to 30% B in 28 min then 30% to 45% B in 5 min, where B consisted of acetonitrile containing 0.1% formic acid. Blank samples consisting of 0.1% formic acid in water were injected between each sample and eluted with the same gradient profile and times as the samples. The LC system was interfaced with the mass spectrometry using an EASY-Spray electrospray ion source (ThermoFisher) and the samples were analyzed using positive ion spray voltage set to 2 KV, S-lens RF level at 65, and heated capillary at 285°C. The Q Exactive HF was operated in the data-dependent acquisition mode for fragmentation. MS1 survey scans (m/z 400–1400) were acquired in the Orbitrap analyzer with a resolution of 120,000 at m/z 200, an accumulation target of 3×10^6, and maximum fill time of 50ms. MS2 scans were collected using a resolution of 30,000 at m/z 200, an accumulation target of 1×10^5, and maximum fill time of 100ms,

with an isolation window of 1.5 m/z, normalized collision energy of 28, and charged state recognition between 2 and 7.

ProteoWizard was used for peak-picking, filtering out peaks with intensity less the 100 and converting the file to mzML format. Protein search and identifications were performed using MS-GF+ search engine on homo sapiens (Uniprot TaxID=9606). Protein identification was set with parameter thresholds of peptide E-val of 0.01 and protein FDR of 0.01.

Annexin V staining, Caspase 8, 3/7 Activity Assays

Anoikis was detected by annexin V-FITC Apoptosis Detection Kit (Abcam) according to the manufactory's instruction. The images were taken by a Zeiss Axio Imager 2 Fluorescence Microscope. For caspase activity measurement, cells were lysis by the caspase buffer (Promega), protein concentration were then measured by microBCA assay. Equal amount of protein (5ug) were used for measuring caspase 8, 3/7 activities according to the instructions.

Immunofluorescence

Cells were seeded on 8 well cell culture slides and incubated for designated time points and fixed with 4% formaldehyde for 10 minutes. The slides were then blocked by PBS buffer contains 5% of BSA. Anti-KIAA0100 c-terminal rabbit antibody (ThermoFisher) was co-labeled with α-anti-tubulin monoclonal antibody (ThermoFisher) or anti-HSPA1A monoclonal antibody (Abcam) respectively, following by anti-rabbit-alexa488 and anti-mouse-alexa555 as secondary antibodies, or phalloidin-568

(ThermoFisher) for F-actin detection respectively. Slides were mounted by Prolong Gold anti-fade with DAPI (ThermoFisher). Images were acquired by Zeiss Axio Imager 2 Fluorescence Microscope or Leica TCS SP5 AOBS Spectral Confocal System.

Statistical Analysis

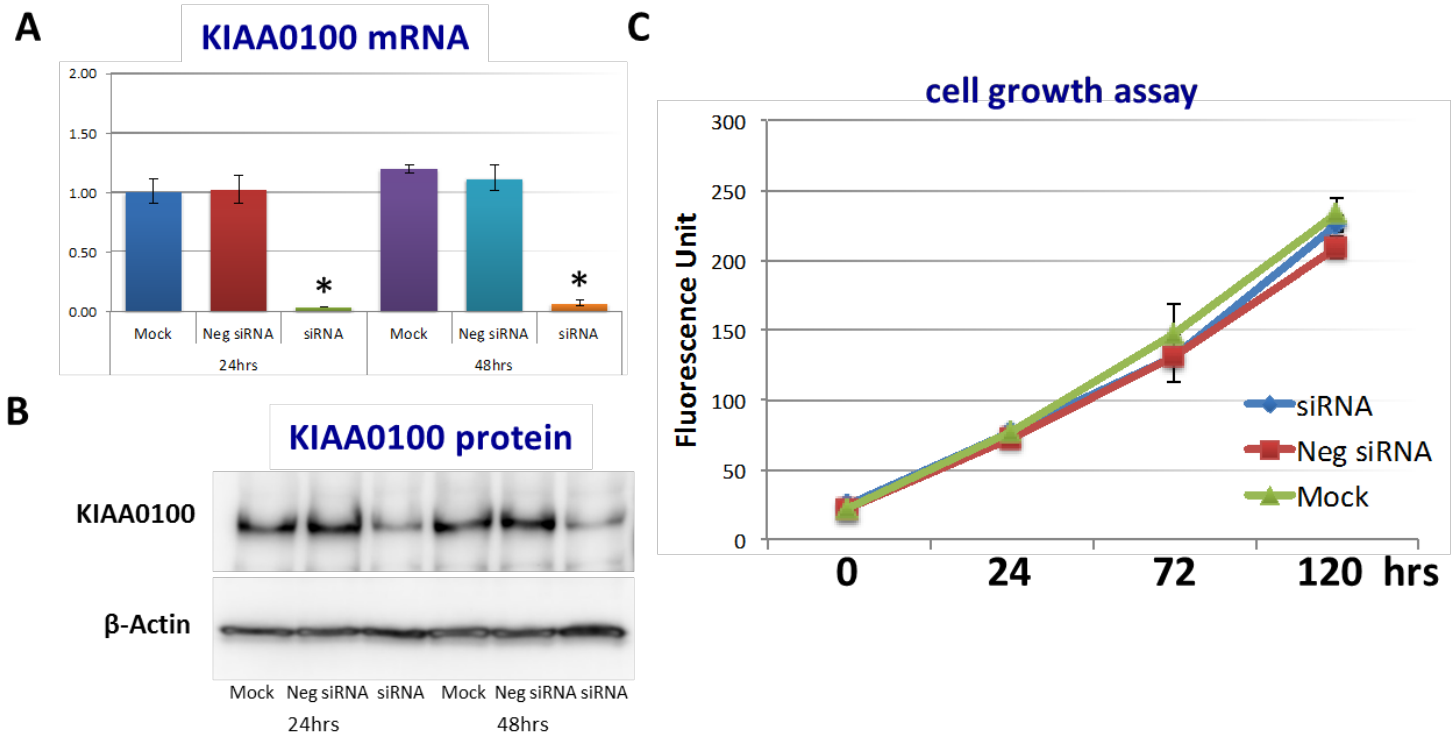
All the experiments were performed at least three times. Mean, SD, and t test were calculated either using Excel software (*, $P < 0.05$). Graphs were generated in Excel.

Results

Silencing KIAA0100 Expression Does Not Affect the Anchorage-Dependent Cancer Cell Growth/Proliferation

To assess the roles of KIAA0100 in breast cancer, its expression in breast cancer cell line MDA-MB-231 was first knocked-down by the siRNA in a forward transfection manner as indicated in Methods. Anchorage-dependent cell proliferation and growth were then examined as shown in Figure 3.2. The expression of KIAA0100 mRNA was effectively reduced more than 90% compared to mock control and the negative siRNA treated control cells within 24 hours and 48 hours after KIAA0100 siRNA treatment (Figure 3.2A). Consistently, the protein levels of KIAA0100 were also significantly decreased within 24 hours after the treatment, indicating the selected siRNA was very efficient in reducing the expression of both KIAA0100 mRNA and protein levels (Figure 3.2B). However, there was no significant difference in the anchorage-dependent cell growth/proliferation between KIAA0100-silenced cells and the control cells for up to 5-day of observation (Figure 3.2C), indicating that silencing the expression of KIAA0100 had no effect on anchorage-dependent cancer cell proliferation/growth. This suggests KIAA0100 might not be directly involved in signal path for the cancer cell aggressive behavior in term of anchorage-dependent cell proliferation and growth.

Figure 3.2 Silencing KIAA0100 Expression Does Not Affect Cell Anchorage-dependent Proliferation/Growth. MDA-MB-231 cells were first seeded on the culture plate and allowed to settle (attachment); cells were then transfected with KIAA0100 siRNA in a forward transfection manner. Efficiency of the siRNA was examined by qPCR and western blot at 24 and 48 hours after transfection. A) Relative KIAA0100 mRNA expression was analyzed by ΔΔCt analysis; mock control at 24hrs was used as reference group and the β-actin was used as reference gene. Y-axis is relative expression level indicated as fold changes in reference to the mock sample at 24hrs. Cells transfected with negative siRNA showed little reduction in KIAA0100 mRNA expression level compared to mock controls ($p>0.05$); KIAA0100 siRNA treatment significantly reduced the KIAA0100 mRNA expression level by over 90% in the first 24hrs and 48hrs ($*p<0.05$); B) KIAA0100 protein levels detected by Western blot. Consistent with the reduction in the mRNA level, protein level of KIAA0100 significantly decreased within 24 and 48 hours, compared to the mock controls and cells treated with negative siRNA; C) Cell growth/proliferation were examined by multiTox-Fluor cell viability assay from day 0 to day 5, assay was performed every two days. No significant difference was observed in the cell proliferation/growth of the cells treated with KIAA0100 siRNA and the negative siRNA as well as the mock controls ($p>0.05$).


A
KIAA0100 mRNA
2.00
1.50
1.00
0.50
0.00
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA
24hrs
48hrs
B
KIAA0100 protein
KIAA0100
β-Actin
Mock Neg siRNA siRNA Mock Neg siRNA siRNA
24hrs
48hrs
C
cell growth assay
Fluorescence Unit
300
250
200
150
100
50
0
0
24
72
120 hrs
siRNA
Neg siRNA
Mock

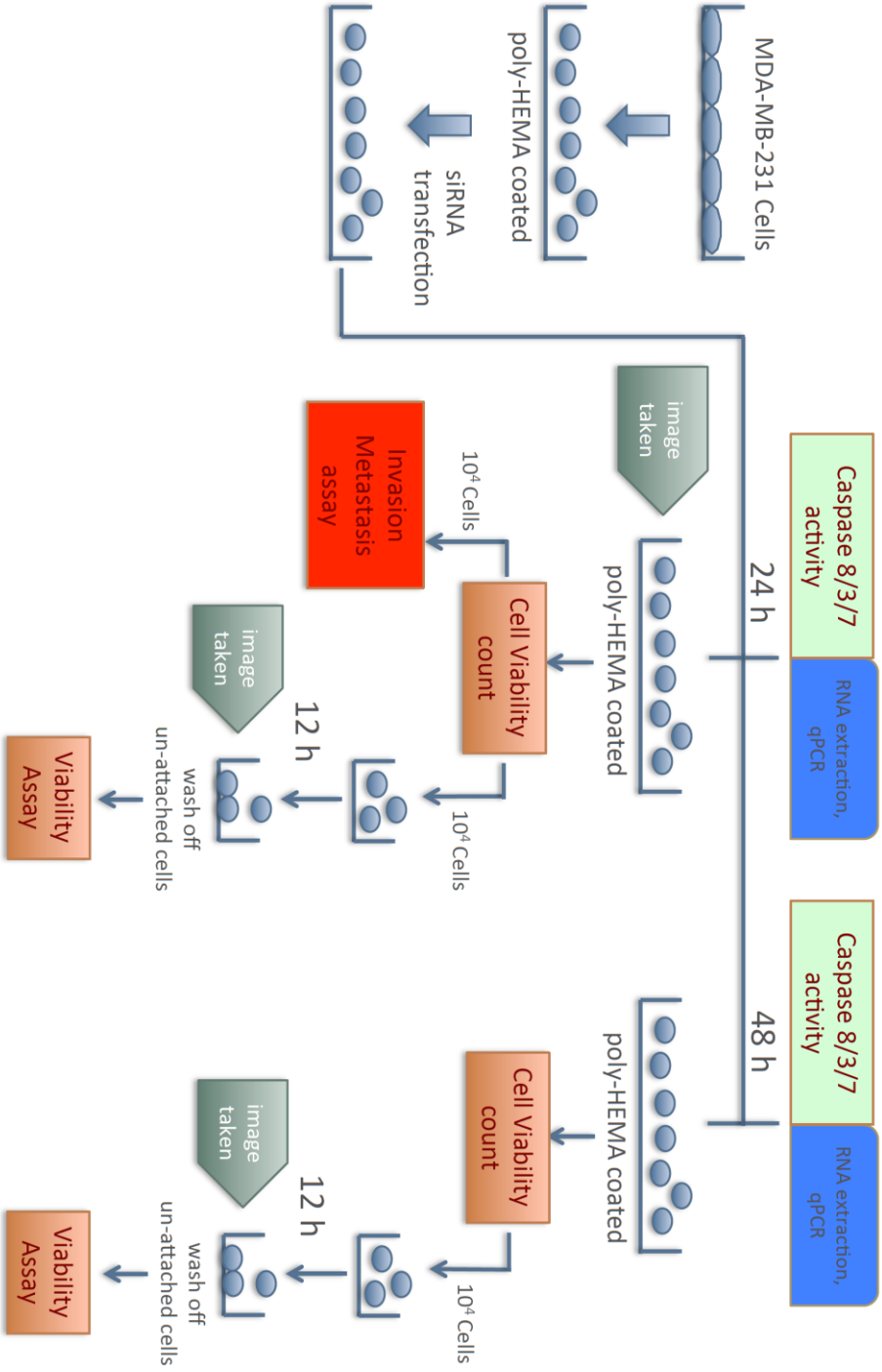
Scheme of Studying Cancer Cell Aggressive Behavior Other Than Anchorage-dependent Cell Growth and Proliferation

Since silencing the expression of KIAA0100 failed to show any significant effect in cancer cells in terms of anchorage-dependent proliferation/growth, effects on anchorage-independent behavior of the breast cancer cells was then needed to be investigated. To this end, cancer cells aggregation in suspension, ability to re-attachment/re-anchorage, anoikis resistance and invasion/metastasis upon KIAA0100 knockdown were studied in the MDA-MB-231 cells. Figure 3.3 illustrates the experiment scheme used to evaluate these potential effects.

MDA-MB-231 cells were maintained in poly-HEMA coated plates, which allowed them to mimic the behavior of cancer cells detaching from ECM and initiating metastasis. Expression of KIAA0100 in the suspended cells was then silenced by the siRNA in a reverse transfection manner. Morphology of the cells (aggregation) in suspension was observed using light-contrast microscopy; total cell number and viability in suspension was then determined by trypan blue based staining on a ViaCell counter (viability). Cells were then further studied for their ability to attach and migrate as follows. First, cells in suspension culture were seeded to the culture plate without poly-HEMA at 24 and 48 hours after siRNA treatment; cells unable to re-attach back to the plate surface were then washed away while cells able to anchor back to the plate were assessed by the viability assay (attachment/anchorage). Second, 24 hours after siRNA treatment, the cells ability to migrate across the BME barrier was examined (invasion/metastasis). Additionally, cell death in the suspension culture due to the silencing of KIAA0100 was evaluated by caspase 8, 3/7 assay as well as annexin V

staining (anoikis). Cells samples were also taken for KIAA0100 protein and mRNA expression analysis. Therefore, cells aggregation/viability in suspension, anoikis resistance, ability of re-attaching, and capability of invasion was then assessed after KIAA0100 silencing without interference from the cell active growth phase.

Figure 3.3 Scheme of Assessing Breast Cancer Cell's Aggressive Behavior After Silencing of KIAA0100



Knocking-Down KIAA0100 Reduced Cancer Cell Aggregation

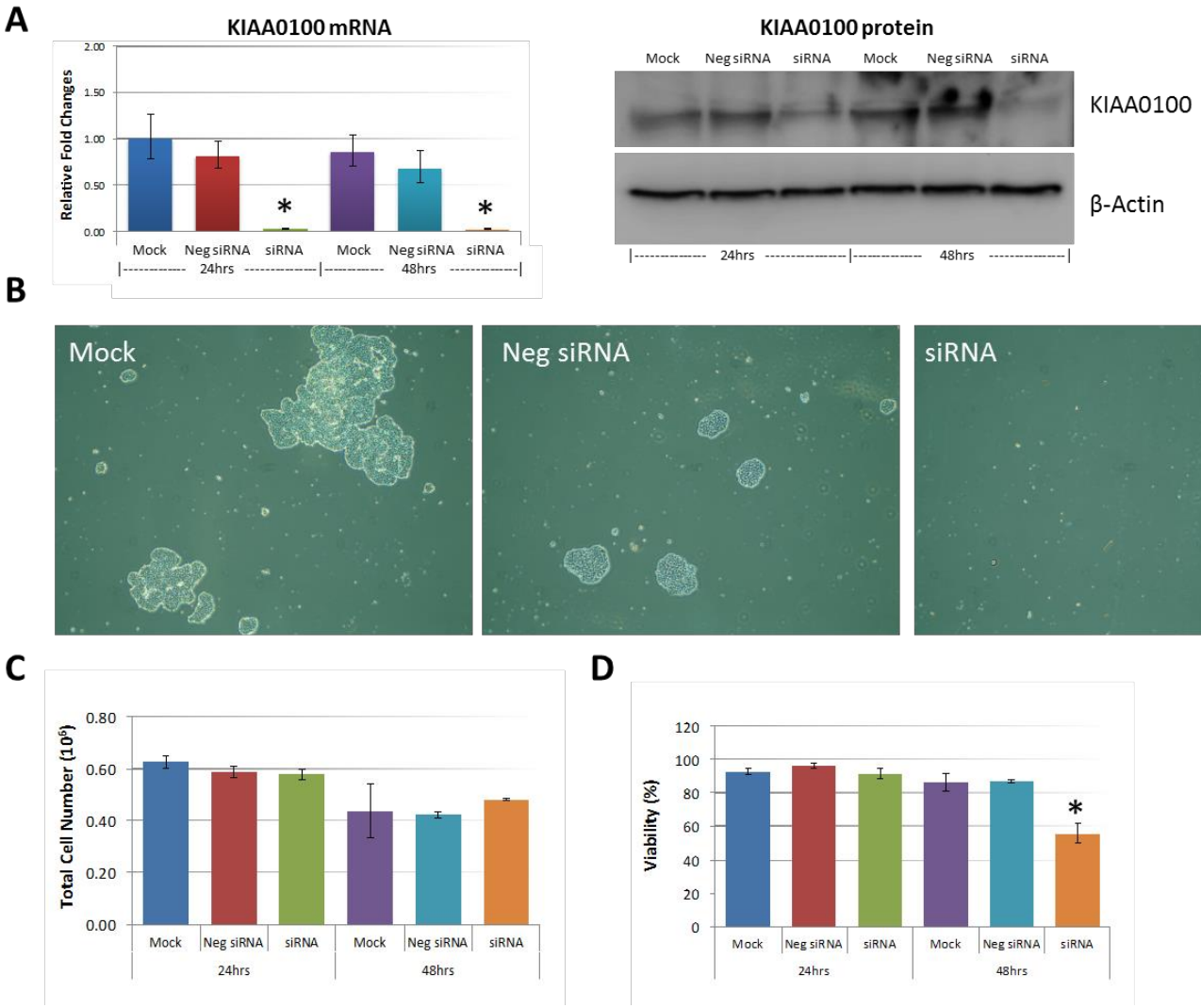
Figure 3.4A shows the efficiency of silencing KIAA0100 in suspended cell culture in a reverse transfection manner. Relative expression levels of KIAA0100 as quantitated by RT-qPCR reduced by more than 90% compared to the mock cell control as well as the cells treated with the negative siRNA. Western blot showed protein levels of KIAA0100 were significantly reduced at 24 and 48 hours after the treatment compared to the mock and the negative siRNA treated control cells, confirming that there was no significant difference in the efficiency of the siRNA treatment between forward transfection scheme and the reverse transfection scheme.

MDA-MB-231 cells are prone to forming loosely connected aggregates when maintained in suspension cultures[202]. Such cell aggregation has been shown to play an important role in cancer cell survival during ECM detachment and loss of anchorage[203-204]. Consistent with this notion, control cells including mock and the negative siRNA treated MDA-MB-231 cells formed cell aggregates that were visible under the microscope as shown in Figure 3.4B. Though the cell aggregates appeared to be different sizes, in general cells were clumped in small spherical units and larger aggregates were formed by multiple units combined. Interestingly, silencing the expression of KIAA0100 eliminated such cell aggregations within 24 hours of treatment and most of the cancer cells appeared as individual suspended cells. The disappearance of cell aggregation upon KIAA0100 siRNA treatment indicates KIAA0100 may play a role in maintaining cell-cell contact and adhesion under the suspension culture condition.

Knocking-Down KIAA0100 Reduced Cancer Cell Viability

Breast cancer cell lines, such as MCF7 and T47D are prone to forming very tight mammosphere in suspension. In contrast, cell aggregates formed by MDA-MB-231 in suspension culture were relatively loose. This suggests that the cell-cell adhesion in MDA-MB-231 cell aggregates may not be very strong. It has been shown that by gentle pipetting, the cell aggregates formed by MDA-MB-231 in suspension could be separated without any deleterious impact on the cells[202]. To take advantage of this, in 24 and 48 hours after the KIAA0100 siRNA treatment, cells were separated by gentle pipetting and the total cell count and viability was examined through trypan blue staining on a Viacell device. Total cell number regardless of viability showed no significant difference at 24 and 48 hours after the siRNA treatment in all samples ($p>0.05$, Figure 3.4 C). Though a slight decrease in total cell number was observed in all three samples at 48 hours, irrespective of the siRNA treatment indicating this effect was not specific to KIAA0100 silenced cells. Cell viability percentage was relatively stable at around 90% at 24 hours after treatment for all samples. However, at 48 hours after treatment, cell viability in KIAA0100 silenced cells was significantly impacted as it decreased to less than 60%. In contrast, no significant difference was observed between the mock and negative siRNA treated control cells, as their viability remained relatively the same as that from the respective 24-hour time point at about 90%. Total cell numbers were at similar levels while only KIAA0100 silenced cells showed reduction in cell viability suggesting a direct effect of silencing the expression of KIAA0100 on the suspension cancer cells.

Figure 3.4 Knocking-down KIAA0100 Reduces Cell Aggregation and Viability in Suspension. MDA-MB-231 cells were reverse transfected with KIAA0100 siRNA and maintained in suspension culture plate coated with poly-HEMA. A) 24 hours and 48 hours after transfection, both mRNA and protein levels of KIAA0100 in cell treated with KIAA0100 siRNA were examined by qPCR and western blot. Both mRNA and protein levels were significantly reduced compared to the cells treated with negative siRNA as well as the mock controls (*$p<0.05$); B) cell aggregation in suspension upon KIAA0100 silencing was significantly reduced compared to the cells treated with negative siRNA and mock control; C) No significant difference in total cell number counts between cells treated with KIAA0100 siRNA and controls at 24 and 48 hours after the treatment by ViaCell ($p>0.05$); D) Cell Viability (in percentage) at 24 and 48 hours after treatment. No significant cell viability difference at 24 hours ($p>0.05$). However, silencing the expression of KIAA0100 resulted in significant drop in cell viability at 48 hours after treatment (*$p<0.05$).


A
KIAA0100 mRNA
Relative Fold Changes
2.00
1.50
1.00
0.50
0.00
*
*
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA
24hrs
48hrs
KIAA0100 protein
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA
KIAA0100
β-Actin
24hrs
48hrs
B
Mock
Neg siRNA
siRNA
C
Total Cell Number (10⁵)
0.80
0.60
0.40
0.20
0.00
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA
24hrs
48hrs
D
Viability (%)
120
100
80
60
40
20
0
*
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA
24hrs
48hrs

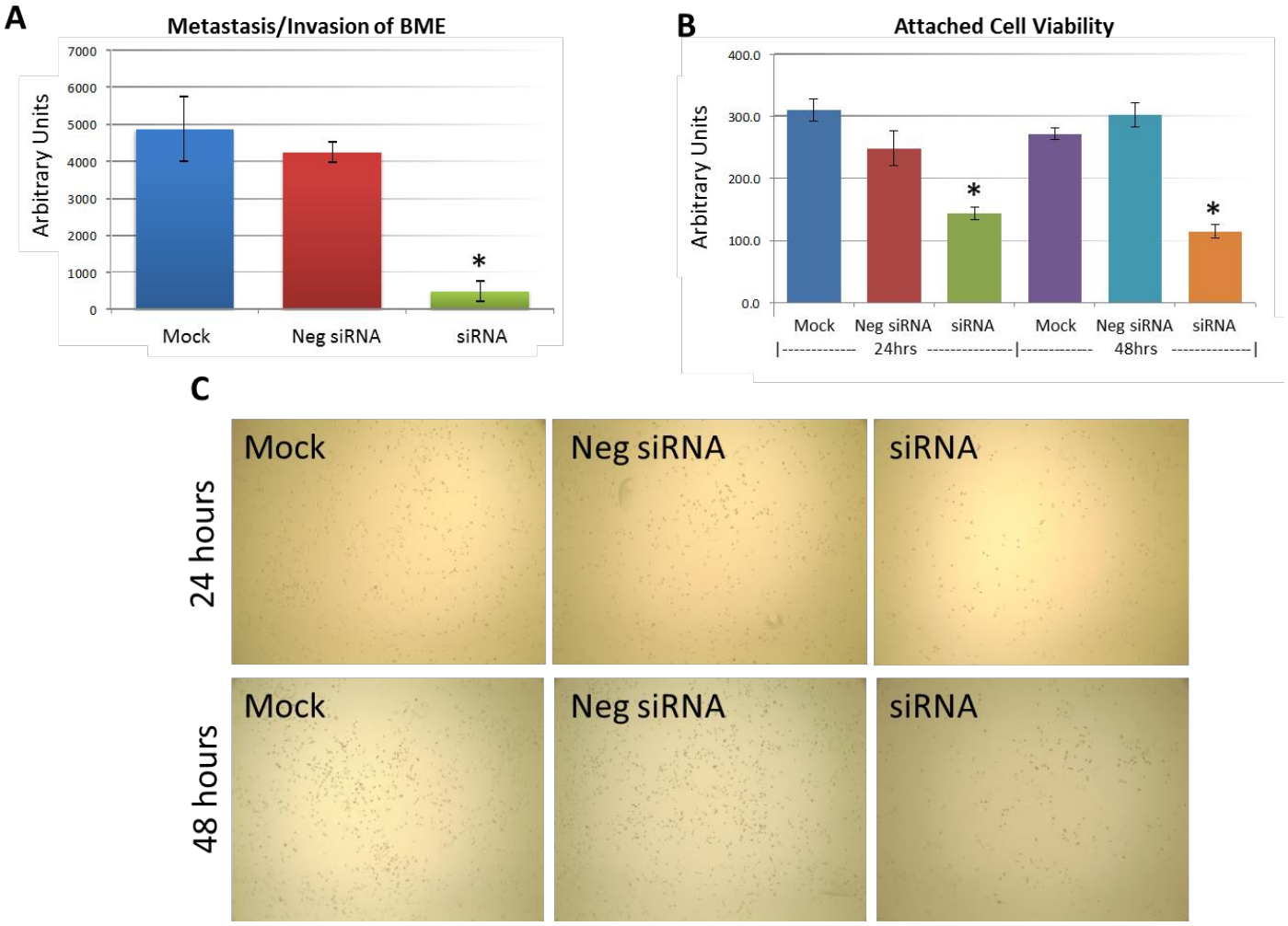
Knocking-Down KIAA0100 Reduced Cancer Cell Invasion/Metastasis

Cell invasion is fundamental to angiogenesis[205], embryonic development[206], immune responses[207], and tumor cell metastasis[208]. MDA-MB-231 cells not only were derived from a metastatic site but have also been shown to be a highly invasive breast cancer cell line in multiple studies[153, 209-210]. To metastasis, cancer cells must detach from the original site, migrate, re-anchor and invade into distal tissue[52]. Figure 3.3 shows how cancer cells maintained in suspension could have the ability to re-anchor and invade/migrate across ECM and then be assessed by a BME assay. The effect of silencing KIAA0100 expression on the cancer cells can then be evaluated. To do so, 24 hours after the KIAA0100 siRNA treatment, aliquots of cells were collected and re-suspended in serum-free medium. Cell invasion/metastasis potential was examined using the cells ability to penetrate BME barrier in response to complete medium with 10% FBS.

As previously described, the expression of KIAA0100 was significantly reduced at 24 hours after the siRNA treatment. However, total cell number and viability at this time point were not affected. Equal numbers of cells from each sample were then used for the BME assay. Figure 3.5A shows cells that were able to penetrate the BME barrier in mock control and negative siRNA transfected cells were at similar levels ($p>0.05$). In contrast, silencing the expression of KIAA0100 significantly reduced the number of cells that were able to migrate across the BME barrier, by more than 80%. In comparison to the control cells ($p<0.05$), this indicates that the capability of re-anchoring and invading/migrating across the ECM matrix in the KIAA0100 silenced MDA-MB-231 cancer cells in suspension was significantly reduced, suggesting KIAA0100 might play a

role in the epithelial to mesenchymal transition of breast cancer cells as well as their capability to colonize distant metastatic sites.

Figure 3.5 Knocking-Down KIAA0100 Reduces Cell Attachment, Metastasis and Invasion. A) Metastasis/Invasion potential was evaluated by the cells' ability to penetrate the BME barrier. Cells able to penetrate the BME barrier were significantly reduced in KIAA0100 knock-down cells. There was approximately a 90% reduction compared to mock control and 80% compared to cells transfected with negative siRNA (*$p<0.05$); B) The number of cells able to attach onto the culture surface was examined by the cell viability assay. Significantly less cells were found to re-attach to the surface from KIAA0100 silenced sample compared to control cells (*$p<0.05$); C) Phase-contrast microscope image shows a reduced density of the KIAA0100 silenced cells compared to controls cells at 24 and 48 hours after treatment.


A
Metastasis/Invasion of BME
Arbitrary Units
7000
6000
5000
4000
3000
2000
1000
0
Mock
Neg siRNA
siRNA
*
B
Attached Cell Viability
Arbitrary Units
400.0
300.0
200.0
100.0
0.0
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA
24hrs
48hrs
*
*
C
24 hours
48 hours
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA

Knocking-Down KIAA0100 Reduced Cancer Cell Re-Attachment

Anchorage dependence of cellular growth and survival prevents inappropriate cell growth or survival in ectopic environments and serves as a potential barrier to metastasis of cancer cells[211]. Cancer aggression manifests as the ability to detach from the original site, and then cancer cell's ability to re-attach and start proliferation at a distant site. Cancer cells ability to re-attachment and invasion was investigated by silencing KIAA0100. The cancer cells ability to re-attach back to the culture surface was assessed. Equal numbers of cells in suspension culture were seeded to the culture plate without poly-HEMA 24 hours after treatment. Cells that were unable to re-attach back to the surface were washed away while cells that able to re-attach back to the plate were examined by a cell viability assay.

As shown in Figure 3.5B, viable cells that were able to re-attach back to the culture plate decreased by more than 40% compared to control cells at 24 hours after treatment ($p<0.05$). At 48 hours after the treatment, the number of viable cells that were able to re-attach back on the culture plate reduced by more than 60% compared to control cells ($p<0.05$). Light-contrast microscope images showed a reduction of cell density upon KIAA0100 knock-down as compared to the controls under the microscope as shown in Figure 3.5C, indicating loss of KIAA0100 expression was able to reduce the cells' ability to re-attach themselves back to the surface, suggesting that KIAA0100 may play a role in regulating the transition of cancer cells from suspension to anchorage.

Knocking-Down KIAA0100 Sensitized Cancer Cells to Anoikis

Cells invasion and reattachment assays indicated that silencing the expression of KIAA0100 in breast cancer cell MDA-MB-231 in suspension hampered the ability of cells to re-attach back to the culture surface. At the same time, reduced cell viability in KIAA0100 silenced cells in suspension also implies KIAA0100 expression knock-down may invoke a cell death mechanism. Anoikis, a cell-detachment induced apoptosis, is one of the primary features required for cancer cells to become a malignant phenotype which supports tumor invasion and metastasis[52]. The untransformed mammary epithelial cell line MCF-10A underwent anoikis when detached and cultured in suspension condition[211]. In contrast, MDA-MB-231 has been shown to have significantly higher anoikis resistance compared to other less aggressive breast cancer cell lines upon detachment from the ECM, consistent with its aggressive invasion and metastasis[212].

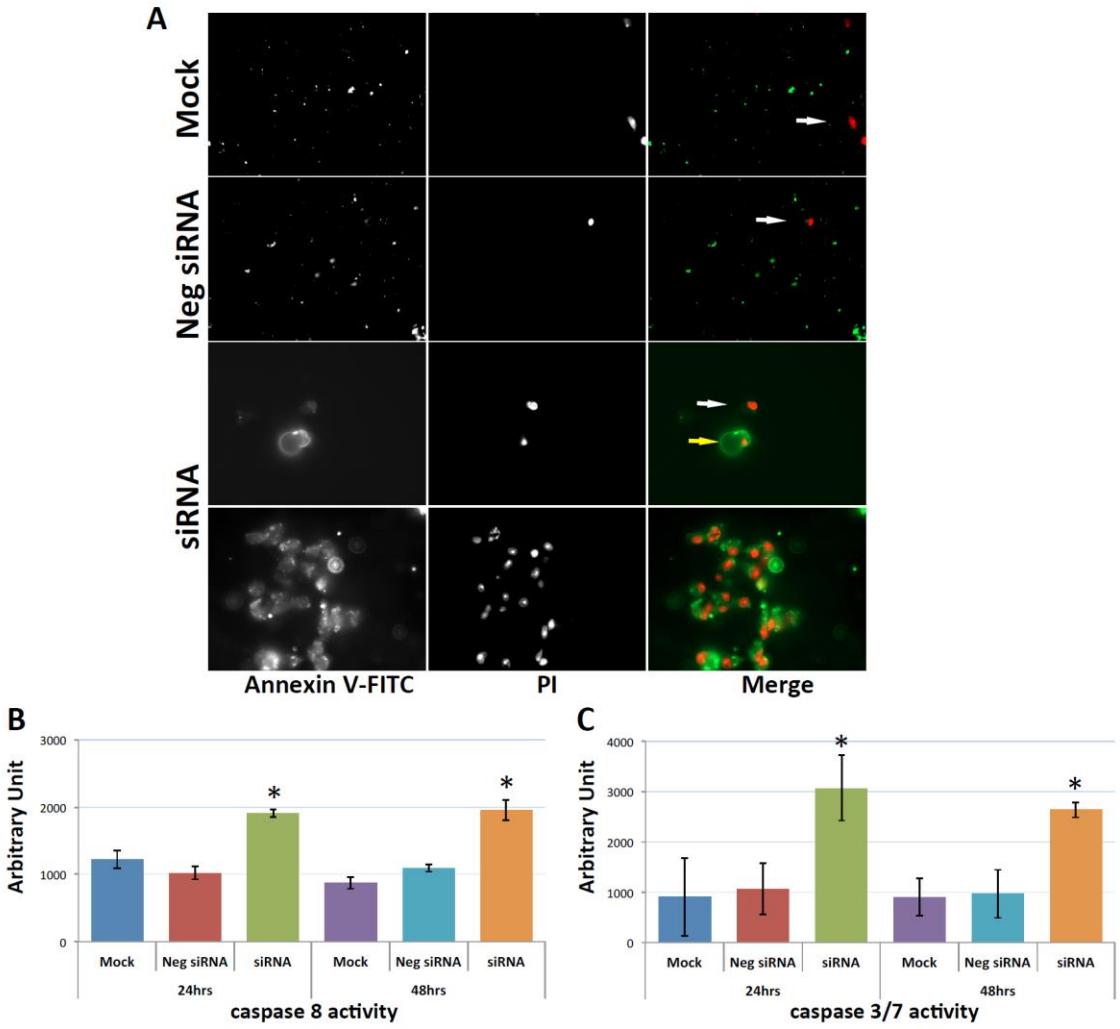
To find out whether depletion of KIAA0100 could compromise MDA-MB-231 cancer cells, anoikis resistance potential and cell death in suspension was evaluated. Anoikis was detected by Annexin V/PI staining 48 hours after the treatment of KIAA0100 siRNA and resulted in maximal protein depletion. As shown in Figure 3.6A, MDA-MB-231 cells in suspension showed no Annexin V staining, with only occasional dead cells shown with staining of PI in controls cells (white arrow pointed). This indicates negligible anoikis induction in control cells, which is consistent with the previous finding that MDA-MB-231 cells were highly resistant to anoikis[212].

In the initial stage of apoptosis/anoikis, cells translocate the membrane phosphatidylserine (PS) from the inner face of the plasma membrane to the cell surface. This was detected by staining with a fluorescent conjugate of annexin V and was showed

a ring shape. In the late stage of apoptosis/anoikis, when the cell membrane structure is disrupted, apoptotic cells may appear double positive for Annexin V and PI staining. Moreover, apoptotic cells may be stained rapidly and strongly with PI and thus may not exhibit Annexin V staining at all[213]. Accordingly, 48 hours after cells treated with KIAA0100 siRNA, MDA-MB-231 cells in suspension showed indications of both early and late stage of anoikis, particularly the late stage anoikis phenotype with halo Annexin V and PI being more dominant(Figure 3.6A).

To confirm the anoikis pathway was activated, activities for caspase 8 and 3/7 were examined. Caspase 8 was considered activated at the initial stage of anoikis and at the beginning of the caspase cascade, while caspase 3/7 was more as the effector caspase in later stage[52] in the cascade. Figure 3.6B shows that both caspase 8 and caspase 3/7 activity were significantly increased compared to control cells ($p<0.05$) at both 24 and 48 hours after the treatment of KIAA0100 siRNA. Therefore, depletion of KIAA0100 in MDA-MB-231 cells in suspension initiate anoikis process was initiated. Specifically, silencing KIAA0100 induced cell anoikis but not general apoptosis since silencing expression of KIAA0100 on cells that firmly attached to the culture plate did not show a similar detrimental effect (Figure 3.2).

Figure 3.6 Knocking-Down KIAA0100 Induces Anoikis. A) Annexin V staining: anoikis of the cell in suspension were assessed by annexin V-FITC apoptosis assay. Most of the mock control and negative siRNA transfected cells showed no visible annexin V-FITC other than the expected random cell death observed with PI (white arrow). Two types of staining pattern for anoikis were shown for cells transfected with KIAA0100 siRNA: early stage of anoikis - cells have lost membrane integrity will show red propidium iodide (PI) staining throughout the nuclei and a ring-like green staining annexin V-FITC on the plasma membrane; late stage of anoikis – cells stained with PI for the nuclei without annexin V staining with or without halo-green like annexin V-FITC staining. Apoptosis/Anoikis activation was examined by caspase 8, 3/7 activity assay: B) Caspase 8 activity significantly increased in cells treated with KIAA0100 siRNA at both 24 and 48 hours after transfection compared to mock control cells and cells treated with negative siRNA (*$p<0.05$); C) Caspase 3/7 activities in cells treated with KIAA0100 siRNA were significantly higher compared to mock control cells and cells treated with negative siRNA (*$p<0.05$).


A
Mock
Neg siRNA
siRNA
Annexin V-FITC
PI
Merge
B
Arbitrary Unit
3000
2000
1000
0
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA
24hrs
48hrs
caspase 8 activity
C
Arbitrary Unit
4000
3000
2000
1000
0
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA
24hrs
48hrs
caspase 3/7 activity

Over-Expressed KIAA0100 Associates with Proteins from Cytoskeleton and Heat Shock Protein (HSP) Family

Our results have shown that silencing the expression of KIAA0100 in MDA-MB-231 breast cancer cells has a profound inhibitory effect on the breast cancer cells aggressive behavior, implying KIAA0100 may be involved in supporting the cell-cell adhesion, enhancing anoikis resistance and invasion. However, how it actually accomplishes these diverse roles remains unknown. In order to get deeper insight into the molecular mechanism of how KIAA0100 affects cancer cells, the open reading frame (ORF) of KIAA0100 was cloned with FLAG and Myc peptide tags added to the C-terminal of the recombinant protein. The recombinant plasmid was then stably transfected into cell line HEK293 (as HEK293/pKIA). Over-expressing recombinant KIAA0100 protein in HEK293 cells instead of breast cancer related cell lines is not ideal, but it is necessary due to the technical difficulty in transfecting large plasmid to multiple breast cancer cell lines, including MDA-MB-231.

Cells over-expressing KIAA0100 (HEK293/pKIA) were viable with minimal morphologically differences compared to the original HEK293 (Figure 3.7A). Expression of recombinant KIAA0100 protein was confirmed by western blot probed by antibody against the Myc tag added to its C-terminal (Figure 3.7B). Immunofluorescence with KIAA0100 was performed by mixing HEK293/pKIA cells with the original HEK293 cells (as KIAA0100 staining negative control). As shown in Figure 3.7C, the over-expressed KIAA0100 recombinant proteins showed a distinguishable localization in the cytoplasm as thread-like networks in HEK293/pKIA cells. Such localization pattern was highly consistent with the microtubule network stained by the anti-TUBA antibody,

implying the recombinant KIAA0100 proteins might be associated with the microtubule network.

Proteins associated with KIAA0100 might provide important clues on its potential molecular functions. To find out what proteins might be associated with KIAA0100, protein-protein interactions in HEK293/pKIA cells were first cross-linked by DTBP[214], the recombinant KIAA0100 protein was then immune-captured by anti-FLAG antibody from the cell lysate. Captured proteins were then separated by SDS-PAGE and stained with coomassie blue. Compared to the control IPs, there are six distinguishable protein bands (Figure 6D, S1-S6) with molecular weight of >260KD, ~260KD, ~90-100 KD, ~70KD, ~55KD and ~45KD. Protein bands of S1, S4 and S6 are unique to the sample that captured by the anti-FLAG antibody; Protein bands of S2 and S3 appeared from IP samples by both anti-FLAG and anti-DIG isotype antibody; however, the intensity of the bands was much stronger in IP by anti-FLAG antibody than the corresponding band from the IP sample captured by the anti-DIG antibody control. Protein band S5 (Figure 3.7D, red arrow) in anti-FLAG IP matched the protein band in control IP sample captured by anti-DIG antibody, whose molecular weight matched the heavy chain of the capture antibody. At the same time, a very faint band (black arrow, S5) underneath that was unique to the anti-FLAG IP was also observed.

Protein slices of S1 to S6 were then excised from the PAGE gel as shown in Figure 3.7D according to the area marked by the rectangles respectively, and then subjected to mass spectrometry analysis. The identified protein list from the mass spectrometry was then screened by the expected molecular weight shown by the corresponding SDS-PAGE slices. Top proteins that matched respective molecular weight

from individual protein slices and the number of peptides identified for those proteins are listed in Figure 3.7E: Slice S1 and S2 were shown to be KIAA0100 as expected with 40 and 75 peptides identified; slice S3 was identified to be HSP90AB1 (HSP90) with 1 peptide identified; slice S4 was identified to be HSPA1A (HSP70) with 19 peptides; band S5 and S6 was identified to be TUBA with 2 peptides and ACTB with 23 peptides respectively. Therefore, the identified proteins belong to two major family, cytoskeleton proteins and heat shock proteins.

To verify the finding from the mass spectrometry analysis, western blot was then performed and probed by specific antibodies against the identified targets respectively. As shown in Figure 3.7F (left), KIAA0100 protein was clearly detected by the anti-Myc antibody from the IP sample captured by the anti-FLAG antibody at the positions of S1 and S2, while no recombinant KIAA0100 was detected from IP captured by the antibody isotype and bare beads controls, confirming the IP captured with anti-FLAG antibody was very specific to the recombinant KIAA0100 proteins; The KIAA0100 recombinant protein was detected at the size around ~260KD, consistent to its calculated molecular weight. The signal from the recombinant protein was also detected at the S1 position (>260KD), which might be a result of larger protein complex formed by KIAA0100 with other proteins, or due to potential post-translational modification; Cytoskeleton protein TUBA and ACTB were detected at the corresponding position of band S5(black arrow) and S6 in sample captured by anti-FLAG IP while none of them was detected in the control IPs; protein corresponding to gel slice S4 was confirmed to be HSPA1A by the western blot. TUBA, ACTB and HSPA1A were confirmed without any non-specific capture from control IPs, suggesting that these proteins were most likely co-precipitated

with the recombinant KIAA0100 proteins. In contrast, the protein corresponding to slice S3 at ~90KD was confirmed to be HSP90AB1 by the western blot, but moderate non-specific capture was also observed by IP with the isotype antibody. The IP input capture monoclonal antibodies were at similar level, suggesting the much higher HSP90AB1 signal detected in comparison to the control may be partially due to protein co-precipitation with KIAA0100.

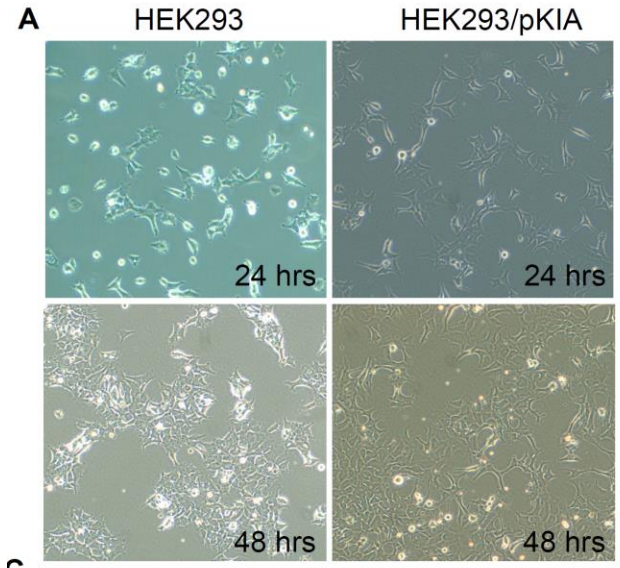
The potential interaction of KIAA0100 with HSPA1A was particularly notable due to its broad involvement in important cellular processes. Interestingly, HSPA1A was also detected by MS in the slice S1, which may be a protein complex formed with KIAA0100, further indicating its potential interaction with KIAA0100. To confirm its association with KIAA0100, IP western blots were also performed with an anti-HSPA1A antibody as capture antibody and probed with same set of antibodies as shown in the recombinant KIAA0100 IP. As shown in Figure 3.7F (right), recombinant KIAA0100 and TUBA proteins co-precipitated with HSPA1A without non-specific capture in control IPs. In contrast, the co-precipitated HSP90AB1 signal was similar to the isotype control IP, therefore most likely HSP90AB1 was not involved in the interaction among KIAA0100, TUBA and HSPA1A. Though ACTB was also co-precipitated by HSPA1A antibody, the signal was relatively low compared to that from recombinant KIAA0100 IP, suggesting the interaction that mainly consists of KIAA0100, HSPA1A and TUBA and may be independent from the relatively minor association of KIAA0100 to HSP90AB1 and ACTB.

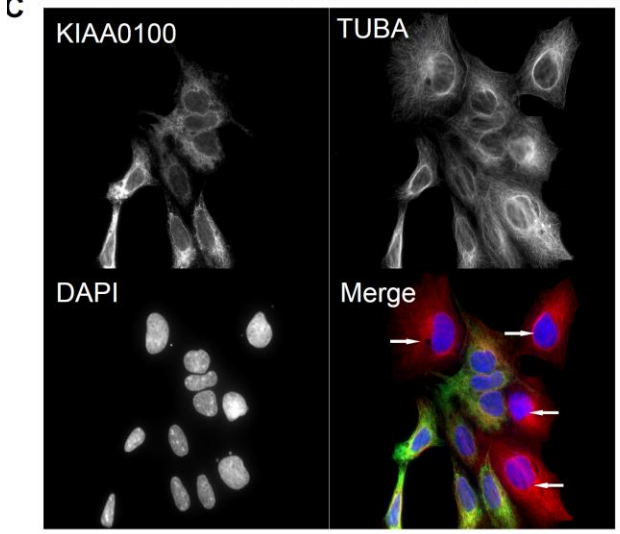
The initial KIAA0100 immunofluorescence in HEK293/pKIA cells in Figure 3.7C shows hints that KIAA0100 might be connected to the microtubule network.

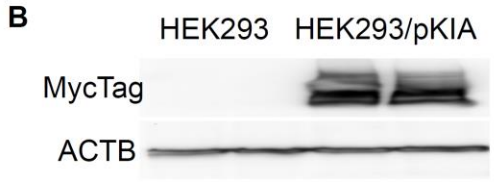
Immunoprecipitation was then used to confirm the association between KIAA0100 and the microtubule protein TUBA. To further verify this result, the individual HEK293/pKIA cells stained with KIAA0100 and TUBA were enlarged in Figure 3.7G. KIAA0100 proteins showed the thread-like network structures. Some but not all of the recombinant proteins were aligned to the same structures stained with the anti-TUBA antibody, which is consistent with the IP and earlier immunofluorescence results. At the same time, co-immunofluorescence was also performed with anti-KIAA0100 and phalloidin, a dye that specifically recognized the cell F-actin structure. As shown in Figure 3.7H, F-actin structures shown by phalloidin staining were mainly shown as fibers on the edge of the cell body with minimal pattern similarity and overlapping with the pattern of KIAA0100.

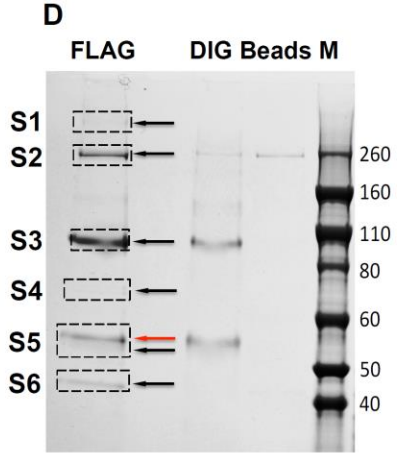
Because the IP mass spectrometry and western blot results also strongly suggested potential the interaction between the recombinant KIAA0100 protein and the stress protein HSPA1A, their cellular localization was determined by IF, where HEK293/pKIA cells were co-labeled with KIAA0100 and HSPA1A antibodies. Consistent with the IP result, Figure 3.7I shows that KIAA0100 also partially co-localized with HSPA1A, indicating such closed contact between KIAA0100 and HSPA1A also exist.

Figure 3.7 Over-Expressing Recombinant KIAA0100 Proteins in HEK293 Cells. KIAA0100 over-expressing cell line HEK293/pKIA was established by transfecting KIAA0100 recombinant plasmid to HEK293 cells: A) Cell morphology comparison between HEK293 and HEK293/pKIA. 24 and 48 hours after seeding same number of cells to the culture plate; B) Western blot confirmed over-expression of the KIAA0100 recombinant protein in HEK293/pKIA cells compared to the parent HEK293 cells; C) Immunofluorescence detection of the recombinant KIAA0100 and microtubule network (TUBA) in HEK293/pKIA cells*. Cells with white arrow did not expressing recombinant KIAA0100 and served as control for the recombinant KIAA0100 staining; D): Recombinant KIAA0100 was captured by anti-FLAG antibody from cell lysate; the captured proteins were separated on 4-12% SDS-PAGE and stained with coomassie blue. Six gel slices were excised as labeled S1 to S6 for mass spectrometry analysis; E) Top protein calls from mass spectrometry analysis that match the molecular weight according to the gel slice positions; F) Western blot was used to confirm the protein identified in the mass spectrometry analysis, IP with anti-FLAG tag antibody and anti-HSPA1A antibody compared to the corresponding controls with anti-DIG antibody as well as empty beads are presented; G) Immunofluorescence of recombinant KIAA0100 in HEK293/pKIA cells in a higher magnification, arrow point to the thread-like structure that both appears in the KIAA0100 staining and the TUBA staining*; H) Immunofluorescence detection recombinant KIAA0100 and F-actin filaments detected by phalloidin in HEK293/pKIA cells; I) Immunofluorescence of HEK293/pKIA cells stained with KIAA0100 and HSPA1A (arrow point to the potential co-localization spot). *In all immunofluorescence images above, KIAA0100 showed in green; other co-stained targets showed in red color.



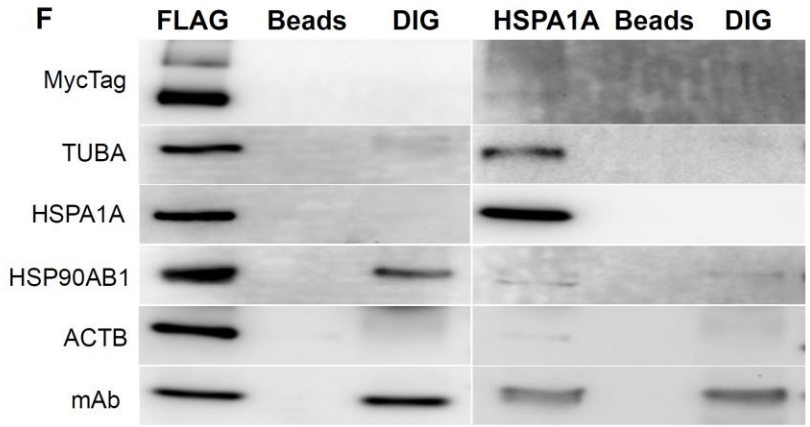


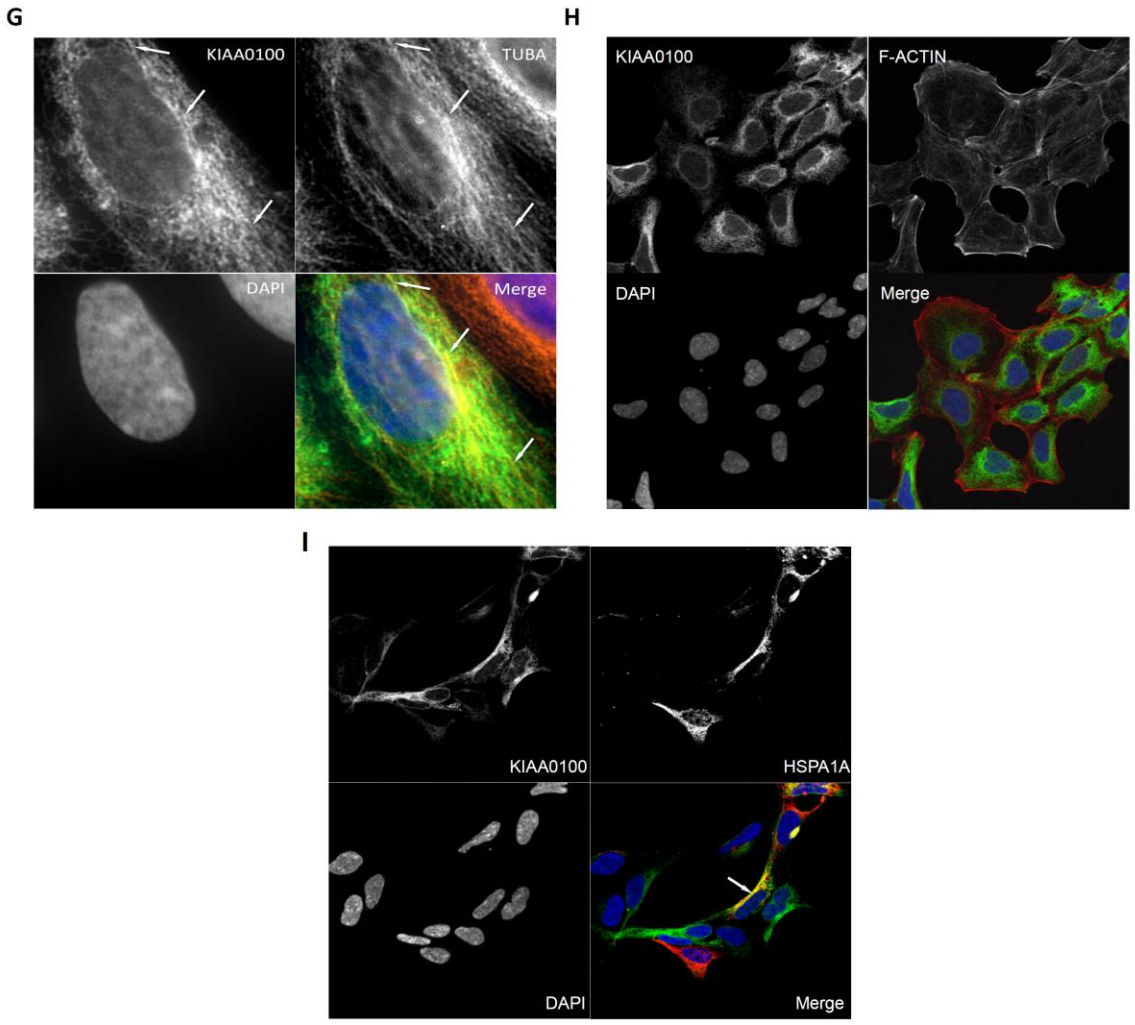




E

Slice	protein identified	peptides number detected
S1	KIAA0100	40
S2	KIAA0100	75
S3	HSP90AB1	1
S4	HSPA1A	19
S5	TUBA	2
S6	ACTB	23



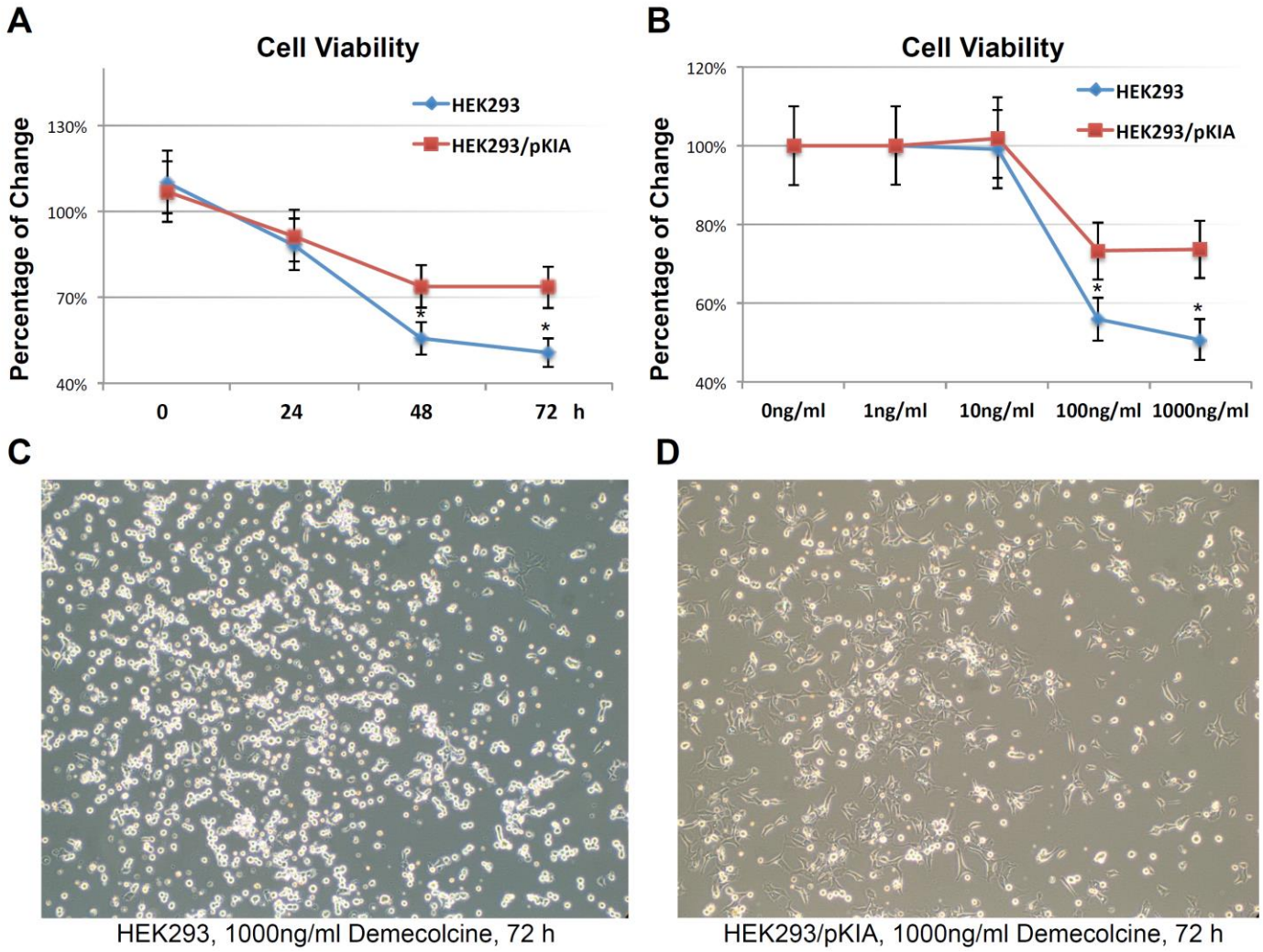

G
KIAA0100
TUBA
DAPI
Merge
H
KIAA0100
F-ACTIN
DAPI
Merge
I
KIAA0100
HSPA1A
DAPI
Merge

Over-expressed recombinant KIAA0100 proteins showed microtubule-like network structures and interacted with TUBA. One explanation could be that KIAA0100 may act as a microtubule binding protein MBP. If so, over-expressing recombinant KIAA0100 protein in HEK293/pKIA cells might be able to prevent microtubule targeting drugs, such as demecolcine from accessing the microtubule structure, in turn rendering the cells more tolerant to the microtubule-binding drugs.

To test this hypothesis, HEK293/pKIA and the parent HEK293 cells were treated with microtubule depolymerizing drug demecolcine[215] ranging from 1ng/ml to 1000ng/ml concentration, endpoint cell viabilities were compared at 72 hours (Figure 7A). Both cells responded to the demecolcine treatment in a dose-dependent manner, but the cytotoxicity of demecolcine on HEK293/pKIA cells was significantly lower than the corresponding parent HEK293 cells: At 100ng/ml of demecolcine, the cell viability for HEK293 was at around 56% while viability for HEK293/pKIA above 70% ($p<0.05$); At 1000ng/ml of demecolcine, the cell viability for HEK293 was at around 50% while viability for HEK293/pKIA still stayed at above 70% ($p<0.05$). Cytotoxicity for both cell lines in response to treatment was also shown in a time-dependent manner. For concentration of 1000ng/ml (Figure 3.8B), cells over-expressing KIAA0100 had a slower declining curve over a 72-hour period compared to the cells without over-expressing the protein. Cytotoxicity was also observed under the light contrast microscope as shown in Figure 3.8CD. For the HEK293 cells, 72 hours after demecolcine treatment, most of the cells become round and lost cell anchorage from the surface while most of the HEK293/pKIA cells remained on the culture plate surface. The results indicate over-expressing KIAA0100 was able to decrease the demecolcine induced cytotoxicity in HEK293 cells,

suggesting the recombinant KIAA0100 proteins prevent the microtubule depolymerizing drug from accessing the microtubule structure.

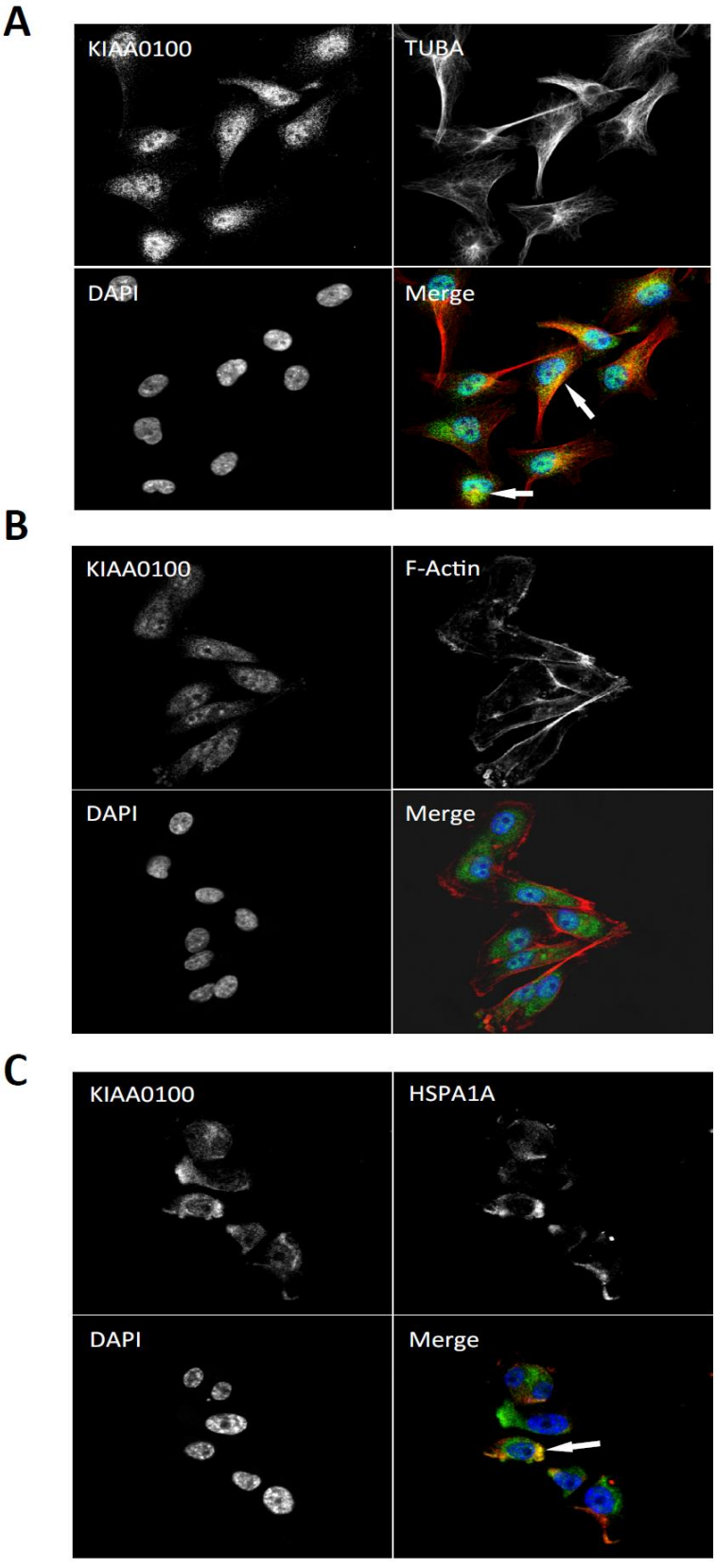
Figure 3.8 Over-Expressing KIAA0100 Increases Tolerance to Microtubule Targeted Drug. A) HEK293 and HEK293/pKIA cells were seeded in 96 well plate with demecolcine concentration ranging from 1ng/ml to 1000ng/ml, cell viability was examined from 0 (DMSO only) to 72 hours. Cell viability for HEK293/pKIA at higher dose of demecolcine (100 and 1000ng/ml) were significantly higher than the corresponding HEK293 cells (*$p<0.05$); B) At 1000ng/ml demecolcine treatment, viability of both cells start to decrease at 48 hours after treatment, but HEK293/pKIA show significantly higher viability compared to HEK293 (*$p<0.05$); Phase contrast microscope image of C) HEK293 and D) HEK293/pKIA cells treated with 1000ng/ml demecolcine at 72 hours after treatment. Most HEK293 cells detached from the plate surface and appeared as round shape. In contrast, though some HEK293/pKIA cells also become round shape, most of the cell were still well attached to the plate surface and well spread.


A
Cell Viability
Percentage of Change
130%
100%
70%
40%
HEK293
HEK293/pKIA
0
24
48
72 h
B
Cell Viability
Percentage of Change
120%
100%
80%
60%
40%
HEK293
HEK293/pKIA
0ng/ml
1ng/ml
10ng/ml
100ng/ml
1000ng/ml
C
HEK293, 1000ng/ml Demecolcine, 72 h
D
HEK293/pKIA, 1000ng/ml Demecolcine, 72 h

KIAA0100 Co-localized with Microtubule Network and HSPA1A in MDA-MB-231

By using the KIAA0100 over-expressing HEK293 model cells, we were able to identify recombinant KIAA0100 proteins mainly associated with microtubule networks and HSPA1A. In order to verify whether the same interactions could exist in the breast cancer cell line MDA-MB-231, co-immunofluorescence was performed on MDA-MB-231 cells and observed under the con-focal microscope. As shown in Figure 3.9A, unlike the recombinant KIAA0100 in HEK293/pKIA cells, the native KIAA0100 proteins in MDA-MB-231 cells showed speckle-like staining. However, the shape of the distribution is roughly consistent with the tubulin staining and is partially co-localized (yellow). In contrast, neither staining pattern similarity nor co-localization with F-actin staining could be observed (Figure 3.9B), consistent with the result from KIAA0100 over-expressing HEK293 cells. The co-localization of the KIAA0100 with HSPA1A in MDA-MB-231 cells was also confirmed in Figure 3.9C, consistent with the above results from KIAA0100 over-expressing HEK293 cells. Our result confirmed that KIAA0100 proteins are mainly associated with microtubule networks and HSPA1A proteins in the breast cancer cells MDA-MB-231.

Figure 3.9 Immunofluorescence of KIAA0100 in MDA-MB-231 Cells. A) MDA-MB-231 cells stained with of KIAA0100 and microtubules TUBA; B) MDA-MB-231 cells stained with KIAA0100 and Phalloidin (F-actin); C) MDA-MB-231 cells stained with KIAA0100 and HSPA1A. *in all figures, KIAA0100 showed as green, other targets showed in red color, arrow point to the yellow spot.


A
KIAA0100
TUBA
DAPI
Merge
B
KIAA0100
F-Actin
DAPI
Merge
C
KIAA0100
HSPA1A
DAPI
Merge

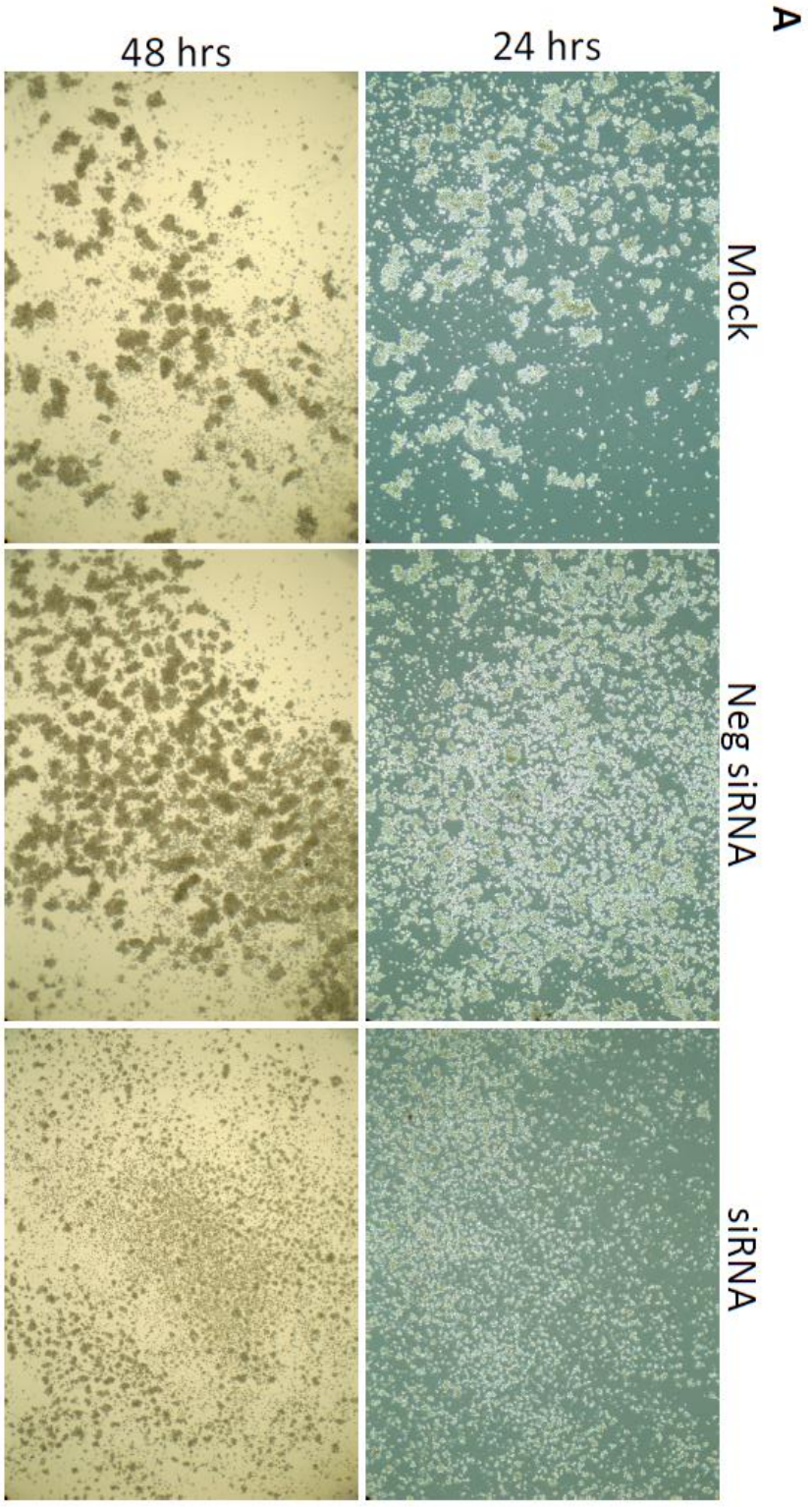
Knocking-down HSPA1A Sensitizing MDA-MB-231 to Anoikis

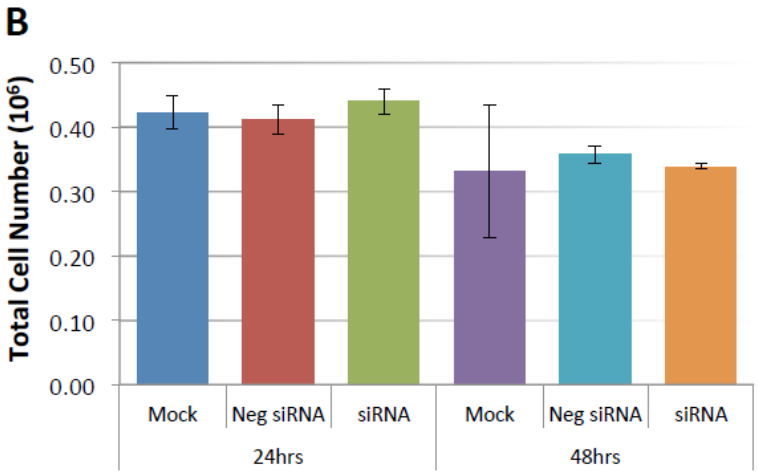
We have shown a strong association between KIAA0100 protein to microtubule networks and HSPA1A in KIAA0100 over-expressed HEK293 and breast cancer MDA-MB-231 cells. Because microtubule networks and HSPA1A have both been reported to be involved in anoikis resistance, we then wondered which protein might be more responsible for the anoikis effect. MDA-MB-231 cells in suspension were treated with either siRNA to silence the expression of HSPA1A, or demecolcine to de-polymerize the microtubule networks. The respective effects were then observed and compared to the KIAA0100 silenced cells in above results.

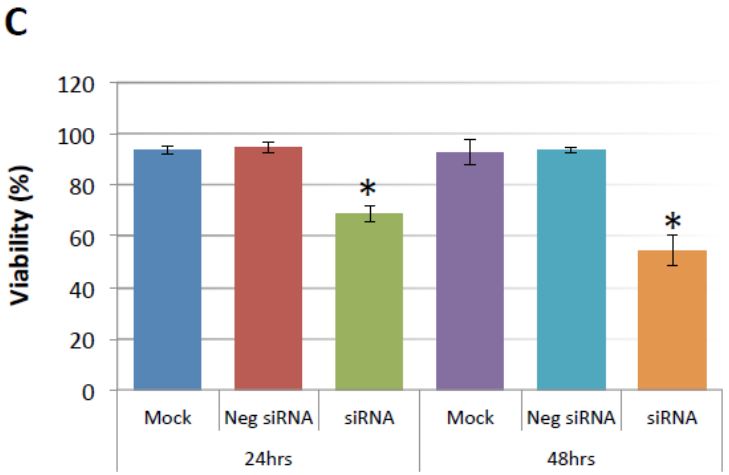
As shown in Figure 3.10A, mock and negative control siRNA treated control cells formed visible cell aggregates in suspension, but the aggregation dramatically dropped in cells treated with HSPA1A siRNA within 24 hours after the treatment and the loss of aggregation became more obvious at 48 hours after treatment. This phenomenon closely resembled the result was observed where diminished cell aggregation in KIAA0100 silenced MDA-MB-231 cells. Consistent with the morphology changes, cell viabilities of HSPA1A silenced cells also dramatically dropped to about 70% at 24 hours and below 60% at 48 hours after treatment while no significant changes in total cell number compared to controls was observed (Figure 3.10B and C). Similarly, depleting the expression of HSPA1A induced anoikis, which was detected by annexin V staining (Figure 3.10E). The decrease in cell viability coupled with an increase in anoikis occurred with KIAA0100 siRNA treatment of MDA-MB-231 cells. This suggests that KIAA0100 and HSPA1A might be involved in the same pathway of anoikis.

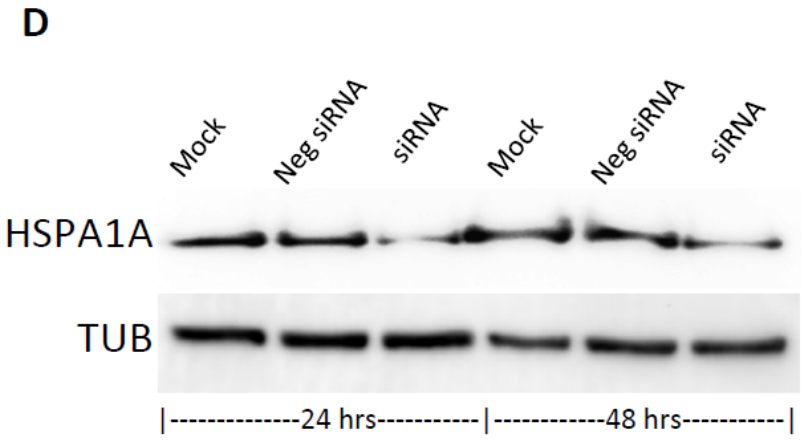
Treating cells with varying concentration of demecolcine (0.5, 5, 50μM) did not show a similar effect of losing cell aggregation and viability (Figure 3.11A and B) on the cells, indicating depolymerization of microtubules alone in the suspended MDA-MB-231 cells was not enough to introduce an anoikis. This suggests that the association between KIAA0100 and HSPA1A drive anoikis, and not microtubule inhibitors. Silencing KIAA0100 resulted in loss of interaction partner to HSPA1A and may be the cause of activating the anoikis pathway. This further implies the interaction between KIAA0100 and HSPA1A may account for the anoikis resistance for MDA-MB-231 cells.

Figure 3.10 Knocking-Down HSPA1A Resembles the Effect of Silencing the Expression of KIAA0100 in MDA-MB-231 Cells. A) MDA-MB-231 Cells were transfected with HSPA1A siRNA and cell morphologies were observed under microscopy at 24 and 48 hours after the treatment. Mock and negative siRNA treated cells show significant aggregation in suspension culture at 24 and 48 hours, while HSPA1A siRNA treated cells lost cell aggregation from 24 hours after the treatment; B) Total Cell number in the suspension culture were similar at 24 and 48 hours between mock, negative siRNA treated cells and the HSPA1A siRNA treated cells; C) Viability of the HSPA1A treated cell dropped to about 70% at 24 hours and then below 60% at 48 hours; D) HSPA1A Silencing by the siRNA was confirmed by western blot analysis; E) annexin V/PI staining of MDA-MB-231 cells treated with HSPA1A siRNA, 24 hours after the treatment, arrows(white) point to PI nucleus staining with halo annexin V staining, arrows(yellow) point to PI nucleus staining with ring-shape annexin V staining.


A
48 hrs
24 hrs
Mock
Neg siRNA
siRNA


B
Total Cell Number (10^6)
0.50
0.40
0.30
0.20
0.10
0.00
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA
24hrs
48hrs


C
Viability (%)
120
100
80
60
40
20
0
*
*
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA
24hrs
48hrs


D
Mock
Neg siRNA
siRNA
Mock
Neg siRNA
siRNA
HSPA1A
TUB
|----------24 hrs----------|----------48 hrs----------|

E

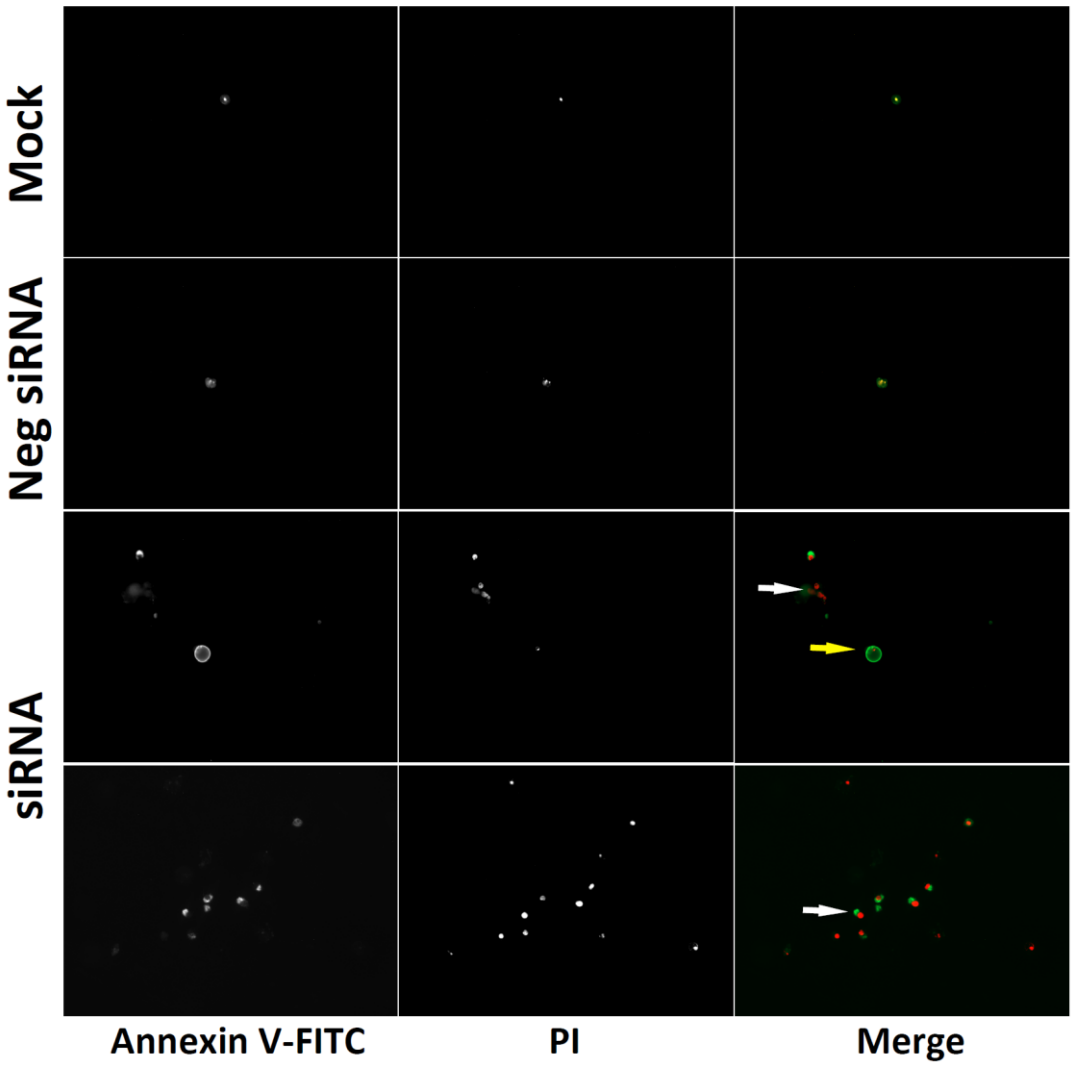
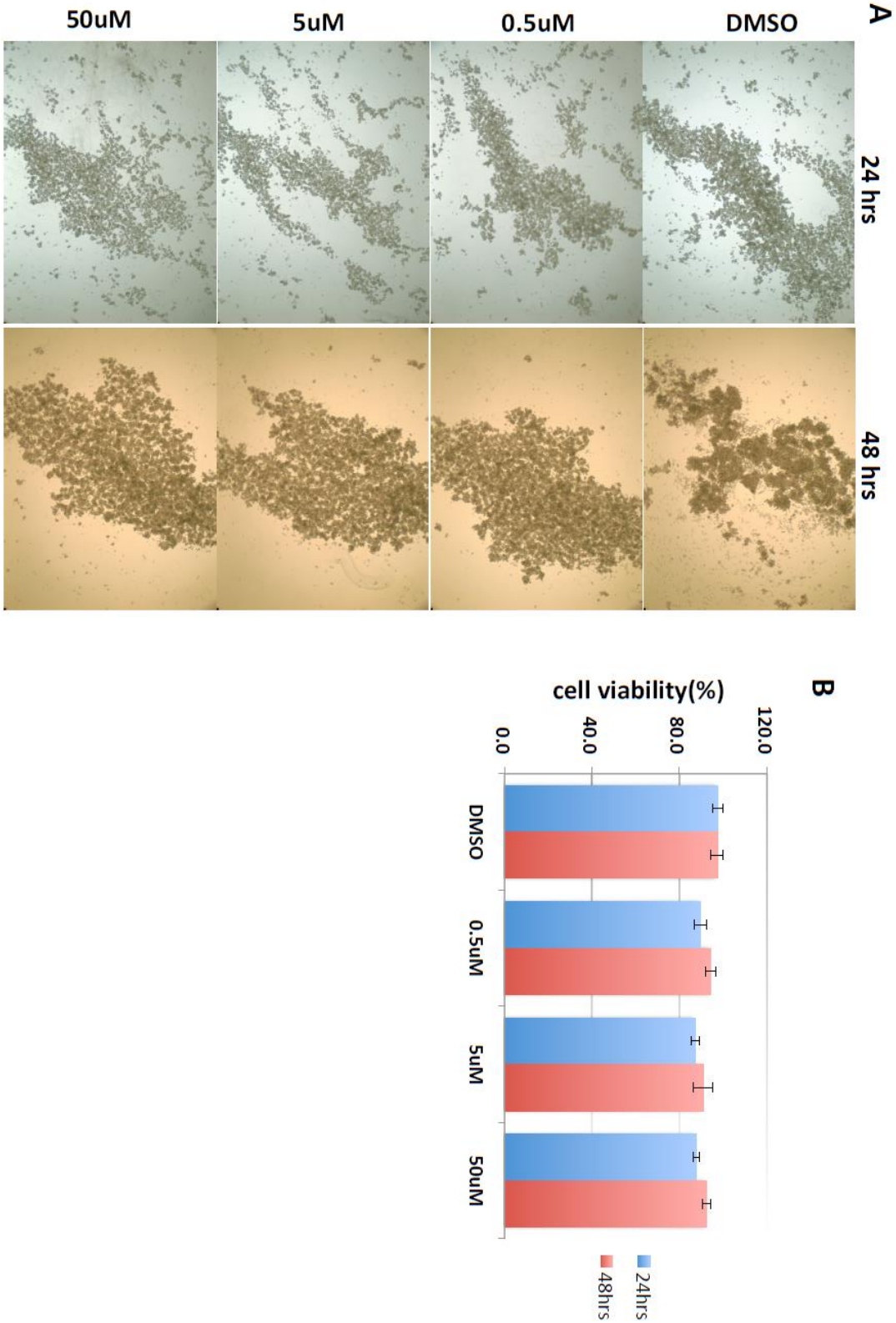


Figure 3.11 Demecolcine Treatments Does Not Induced Anoikis in MDA-MB-231 Cells. A) Morphologies for MDA-MB-231 treated with 0.5, 5, 50uM of demecolcine at 24 and 48 hours. There was no much morphologic difference in terms of cell aggregation between control DMSO treated cells and the cell treated with different concentration of demecolcine; B) Cell viability, both cells treated with DMSO and the cells treated with different concentrations of demecolcine show high cells viability at 24 and 48 hours after the treatment.


A
DMSO
0.5uM
5uM
50uM
24 hrs
48 hrs
B
cell viability(%)
120.0
80.0
40.0
0.0
DMSO
0.5uM
5uM
50uM
24hrs
48hrs

Discussion

Breast tumorigenesis is a multi-step process starting from benign and atypical hyper proliferation, progressing into in situ carcinoma, invasive carcinomas, and culminating in metastatic disease[216]. The development and progression of breast cancer is a complex process including interaction of many genetic, epigenetic and environmental factors at different stages of cancer cells. These include proliferation/growth, anchorage/re-attachment, adhesion/aggregation, anoikis resistance, as well as metastasis/invasion. KIAA0100 has been found to be elevated in breast cancer tissues from multiple studies[139, 163, 191-192] and in malignant transformed benign breast cancer cells[188]. Our study also showed that KIAA0100 protein was found in breast cancer cell lines MCF7, T47D and MDA-MB-231 and their respective EVs. The relative abundance of KIAA0100 in the enriched plasma EV fraction was shown to be elevated in advanced breast cancer patient compared to the non-cancer controls (Chapter 2), suggesting it might be related to breast cancer aggression.

In the current study, the role of KIAA0100 in cancer aggression was investigated by adopting MDA-MB-231 cells - a highly aggressive breast cancer cell line[122, 200] as a model. We found that while depleting the expression of KIAA0100 did not affect cell anchorage-dependent proliferation/growth, it did show a significant detrimental effect on the aggressive behavior in the anchorage-independent cells. The aggressive behavior of breast cancer cells in suspension, such as cell aggregation, re-attachment, migration and invasion were all affected to different extent upon silencing the expression of KIAA0100, suggesting KIAA0100 could be involved in a broad regulation network. Specifically, we found that silencing the expression of KIAA0100 render the MDA-MB-231 cells in

suspension sensitive to anoikis. We also provided evidence that KIAA0100 may be involved in anoikis resistance through its interaction with HSPA1A. At the same time KIAA0100 might also regulate other cancer aggression behavior through its interaction with microtubule networks.

Silencing the expression of KIAA0100 reduced multiple cancer cell aggressive behaviors in MDA-MB-231 cells. However, which was the first one to be triggered, or whether they were all activated at the same time remains unknown, partially due to the fact that these seemingly separated processes are actually intertwined together and they may have mutual influence, which made it more difficult to isolate individual event experimentally. For instance, it has been shown that cell aggregation is important for anoikis resistance and promotes survival[203-204], which in turn would increase the chance of viable cells to regain anchorage; on the other hand, if cells were under anoikis, it would be reasonable to assume that cells would not be able to aggregate as well as less likely to regain the anchorage. However, because MDA-MB-231 cells were reported to be anoikis resistant[212] and its cell aggregates in suspension are relatively loose compared to other breast cancer cell lines[202], suggesting that the aggregation dependent survival may not be the key contributor for its capability of anoikis resistance. The initiation of anoikis process but not the cell aggregation upon the silencing KIAA0100 might be the main driver for the loss of cell viability and resulted in the following diminished cell aggregation and cell reattachment.

It is interesting that silencing the expression of KIAA0100 specifically induce anoikis, a programmed cell death process that is invoked upon cell detachment from extracellular matrix but not the common apoptosis for well anchored cells[52]. MDA-MB-

231 cells that are firmly attached on the culture plate did not respond to the KIAA0100 siRNA treatment for up to 5 days, while the same treatment on the cells in suspension showed a response within 24-48 hours post treatment. One potential explanation is that KIAA0100 may be particularly important to promote cell survival under stress condition caused by the losing of cell anchorage and it may contribute to the acquisition of anoikis resistance, an important step during cancer cell malignant transformation[52]. When a cell is not under stress, such function may not be that significant or may even be redundant. Our results showed that the interaction between KIAA0100 and HSPA1A, a stress protein that are well-known to be involved in promoting cell survival under stress condition, which is consistent with this interpretation[217-219].

Native KIAA0100 in breast cancer cell line MDA-MB-231 was localized in cytoplasm with a speckled pattern (Figure 3.9A), which is consistent with the KIAA0100 staining patterns from Human Protein Atlas database, where KIAA0100 were detected in U-251MG (Glioblastoma), A431 (Epidermoid Carcinoma) and U2-OS (Osteosarcoma) cell lines[220]. Staining suggests the KIAA0100 is part of the microtubules network structure but not the action network. Interestingly, the over-expressed recombinant KIAA0100 proteins in HEK293 cells formed a clear thread-like network structure and showed partial co-localization with the microtubule network, implying the speckle-like distribution of native KIAA0100 proteins might be just protein spots that were distributed sporadically along the microtubule network. This was further supported by the immunoprecipitation which showed tubulin protein, the building block of the microtubule network, was co-immunoprecipitated with recombinant KIAA0100 proteins. Additionally, over-expressing KIAA0100 recombinant proteins in HEK293 cells

significantly increased the cells' resistance to the microtubule depolymerizing drug demecolcine. The existence of similar interaction between native KIAA0100 and microtubule network was confirmed by immunofluorescence in the breast cancer cell line MDA-MB-231. This is consistent with the result from the over-expressing cell line model HEK293 even though it was not a breast cancer cell line, suggesting the interaction between KIAA0100 and microtubule network might be common across difference cell types.

Microtubules are proven critical for anoikis resistance, aggregation, migration and invasion[196-198, 215, 221-222]. Our results suggested KIAA0100 might be functioning as a MBP, which have been shown to be a group of structurally and functionally diverse regulators of microtubule-mediated cellular processes[223]. MBPs have been reported to either stabilize/destabilize microtubules or move transport various cargoes along the microtubule networks[223]. MBPs have also been demonstrated to regulate the binding of microtubule-targeting drugs to microtubules; thereby modulating cancer cell sensitivity to these drugs[223-225]. Our results showed potentially close contact between KIAA0100 with microtubules was consistent to such description. On the other hand, whether KIAA0100 is just bound to microtubules subunits or move along the network with cargoes is still left to be investigated.

Interaction with microtubule network implies KIAA0100 may also have a role in regulating the cell spreading, mobility, metastasis and invasion[198, 215, 226], which are the primary functions of microtubule network. According to the trypan blue cell viability assay, cell viabilities were similar between controls cells and the KIAA0100 silenced cells at 24 hours after treatment. Seeding cells at this time showed that about 50% of the

cells were able to re-attach back to the culture plate, which was roughly consistent to the cell viability at 48 hours after treatment due to the ~12 hours' time gap needed for the re-attachment assay. This suggests that at least part of the reason for the cells' inability to re-attach to the culture surface was due to the loss of cell viability, possibly through activating the anoikis process. Because some cells were also seeded to evaluate their capability of invasion/metastasis potential, an interesting comparison was then demonstrated between the cell re-attachment rate and the migration/invasion rate upon depletion of KIAA0100: theoretically if sensitizing cells to Anoikis is the only effect from silencing the expression of KIAA0100, a roughly similar percentage of signal reduction in the BME metastasis-invasion assay would be expected. In contrast, the BME invasion result showed more than 80% reduction in signal from the KIAA0100-depleted cells compared to controls, much higher than the expected 50% of reduction, indicating that besides its role in anoikis, KIAA0100 may also have other functions that related to the capability of regaining anchorage as well as migration/invasion. The result implies that KIAA0100 might contribute to cell anchorage/migration/invasion by its interaction with the microtubule network.

Multiple members in the HSP protein family were found potentially interacting with KIAA0100, in particular HSPA1A was confirmed by that fact that it was co-precipitated and co-localized with KIAA0100 in HEK293/pKIA and MDA-MB-231 cells. HSP protein family has long been shown to be involved in protein-chaperoning functions such as protein folding and refolding[227-228]. In addition, HSP proteins have also been widely reported to have stress-induced anti-apoptotic properties[217] and they are actively involved in various processes as tumor cell proliferation, invasion, metastases

and death[228]. Interestingly, it has been shown that HSPA1A is required for cancer cell survival under stress condition but not for cell proliferation and growth[218]. Due to such dual functionality, the question becomes whether the interaction between KIAA0100 and HSP proteins found in the current study was a result of a regular protein folding process through HSP's chaperon function, or the possibility that KIAA0100 actually modulates Anoikis through of HSP proteins. Silencing the expression of HSPA1A showed a similar effect on the suspended MDA-MB-231 cells compared to the KIAA0100 silenced cells, in terms of losing cell aggregation and cell death caused by anoikis, suggesting that KIAA0100 and HSPA1A could be involved in the same signal pathway.

KIAA0100 is a large protein with about 250KD in size. Naturally large proteins may contain multiple domains and serves as docking platform for other molecules[185-186]. Therefore, it may not be surprising that KIAA0100 is able to interact with multiple proteins that belong to a wide array of functional categories. Our immunoprecipitation experiments showed that KIAA0100 interacts with microtubules and HSPs, however, whether they interact with all at the same time as a complex or independently remains unknown. Potentially there are four interaction models that could be suggested from the IP results: 1) HSP and KIAA0100 are bound on same microtubule filament but in separated location; 2) HSP, KIAA0100 and microtubule bind to each other as a complex by different domains; 3) HSP interacts with KIAA0100, and KIAA0100 then bind to microtubules; 4) HSP binds to KIAA0100 and KIAA0100 binds to microtubules independently. There are two major domains on HSPA1A protein: The N-terminal nucleotide binding domain (NBD) is responsible for binding and hydrolyzing ATP; the C-terminal substrate-binding domain (SBD) binds to the client/substrate proteins[229].

Interaction network of HSPA1A also has been explored extensively by both experimental and bioinformatics approaches (CORUM: P0DMV8; IntAct: HSPA1A; STRING: HSPA1A)[230-232], but so far no study has demonstrated a direct interactions between HSPA1A and microtubules. Meanwhile, according to the data from The Human Protein Atlas, HSPA1A proteins are mainly localized to vesicles in the cytoplasm but not distribute along the microtubule networks (The Human Protein Atlas: HSPA1A). Silencing HSPA1A led to sensitization of the cells to anoikis, which is similar to that of silencing KIAA0100 while demecolcine treatment failed to do so. This evidence suggests that the interactions between KIAA0100 and microtubule and HSPs may be independent events.

It has also been shown that HSPA1A is capable of binding to different phospholipids[233] or even embedding itself within the lipid bilayer[234]. One of the phospholipids that has high affinity for HSPA1A is bis-(monoacylglycerol)-phosphate (BMP)[235]. The concentration of BMP in the late endosome is 15 mol % of the total lipid content of the organelle and can comprise as much as 70 mol % of the lipid composition of the intra-endosomal vesicles[236-237], which is closely related to multi-vesicular bodies (MVB) and EVs generation[20, 238]. Indeed, HSPA1A was one of the most common proteins that found in EVs[6, 29, 37, 239] as well as in the breast cancer cell line EVs used in Chapter 2. Therefore, it raised the possibility that KIAA0100 might be mediating the intra-vesicle transportation of HSPA1A proteins, MVBs or even EVs might be one of its potential destinations, which is consistent with the finding in the previous chapter that KIAA0100 was found in EVs generated from breast cancer cell lines as well as EVs enriched from plasma. Therefore, the interaction/co-localization between KIAA0100 and

HSPA1A in cell line studies suggests that there is a clear molecular path of associating KIAA0100 to EVs through its interaction with HSPA1A. This further supported the conclusions in Chapter 2 that the F68 co-polymer based plasma EV enrichment enriched plasma EVs and the features discovered from the clinical plasma EVs study (advance breast cancer vs non-cancer control) are closely associated with EVs.

Lastly, one interesting implication is that most of the cancer chemotherapy drugs are targeted to cancer cells with active cell cycles because most enzymes making DNA and RNA will stall at DNA damage. Non-cycling cells can take the time to repair the damage. DNA-damaging agents will primarily hit cells in the S-phase of the cell cycle and spindle poisons cells in mitosis. Hence, cells in the G0/G1 phase of the cell cycle have long been known to be relatively resistant to classical cytotoxic therapy [240]. MDA-MB-231 controls cells in suspension culture were not in an active proliferation status as indicated by negligible increase in total cell number from 24 to 48 hours after treatment and the cell numbers were still very close to the initial seeding cell number of $5x10^5$ (Figure 3.4C and Figure 3.10B). Therefore, it may not be surprising that treatment with demecolcine did not show any toxicity to the cells, partially because the demecolcine targets mitosis cells and not quiescent cells. This not only suggested that under the suspension condition, KIAA0100 contributes to the anoikis resistance through its interaction with HSPA1A, but also that KIAA0100 might be an attractive target aimed at the quiescent cancer cells because anoikis was initiated by silencing its expression alone. Though HSPA1A related small molecules drugs have been developed[241], such chemical may be toxic not only to cancer cells, but also to normal cells due to their inhibitory effects on normal cellular function like protein folding, autophagy, lysosomal function.

Because lack of toxicity on anchored cells by silencing KIAA0100 and the association between KIAA0100 and HSPA1A might be responsible for the anoikis resistance in suspended cells, specifically targeting to KIAA0100 or its interaction with HSPA1A might be able to sensitize the cancer cells that are in the process of metastasis to Anoikis with less cytotoxicity effect to normal cells.

Unfortunately, most of the basic biochemical characterizations for KIAA0100 are not yet available, such as structure, domains, post-translational process/modification, protein-protein interactions as well as potential activity if there is any. Indeed, this protein is so unique in the human protein database that none has been found to have significant homology to it (NCBI BLAST); neither any significant conserved domain could be found from mammalian protein database. Though bioinformatics analysis suggests that KIAA0100 might contain domains like FMP27 and Apt1 (UniqProt: Q14667, analyzed by InterPro, Pfam and SMART), none of these domains originated from mammalian species, in contrast, those domains were rather from yeast and plant database, which might limit their functional interpretation in mammalian cells. Our finding suggested that there might be a microtubule or HSP protein binding domains on KIAA0100, which need to be further explored.

In summary, we investigated the cellular and molecular functions of KIAA0100, a gene that is strongly related to breast cancer aggression. We showed that KIAA0100 might modulate aggressive behaviors of breast cancer cell MDA-MB-231 in suspension, including cell aggregation, adhesion/attachment, and metastasis/invasion as well as anoikis resistance. Interaction between KIAA0100 and HSPA1A may be directly or indirectly responsible for imparting anoikis resistance to the breast cancer cell MDA-

MB-231, thus making KIAA0100 a potential target for chemotherapeutic intervention.

CHAPTER 4

PLASMA EXOSOME PROFILING OF CANCER PATIENTS BY A NEXT GENERATION SYSTEMS BIOLOGY APPROACH

Abstract

Technologies capable of characterizing the full breadth of cellular systems need to be able to measure millions of proteins, isoforms, and complexes simultaneously. We describe an approach that fulfils this criterion: Adaptive Dynamic Artificial Poly-ligand Targeting (ADAPT). ADAPT employs an enriched library of single-stranded oligodeoxynucleotides (ssODNs) to profile complex biological samples, therefore achieving an unprecedented coverage of system-wide, native biomolecules. In order to show its clinical utility, ssODNs libraries were enriched from plasma extracellular vesicles (EVs) from breast cancer patients and controls respectively. ssODNs that showed differentiation between breast cancer and controls were selected and then a ADAPT library with 2000 resynthesized ssODNs were used to profile a large scale clinical samples of total 500 subjects, including 206 breast cancer patients, 177 non-cancer patients, and 117 self-declared healthy controls. Feature selection, patient classification and prediction performance were then analyzed by a random forest-based model. Overall prediction performance of 0.73 in AUC was obtained when comparing biopsy-positive cancer patients to healthy donors, 0.64 when comparing to biopsy-negative patients. In addition, the cancer prediction performance was not associated with Bi-RAD breast composition categories. Our results suggested that ADAPT profiling was able to extract information from plasma EVs that associated with patients' cancer status.

Introduction

Extracellular vesicles (EV), which are secreted into circulation by many cell types, can provide a snapshot of cellular processes active in disease and healthy cells, allowing the EVs in circulation to serve as sentinels of the health of an individual. In cancer, EVs from neoplastic cells are involved in intercellular communication essential for several fundamental aspects of malignancy, including immune-evasion[242], angiogenesis[243], and metastasis[13, 244]. The molecular composition of EVs correlates with the cell-of-origin[6], and alterations in membrane components, luminal contents, and abundance[91] of EVs have been described in a variety of cancers[245-248]. Thus, EVs may be an informative biological substrate, reflecting the dynamic alterations that can occur during tumor progression.

Libraries consisting of several trillion ssODNs encompass nearly infinite numbers of three-dimensional structures due to the vast complexity of DNA sequence space[249-251]. Selection/amplification schemes can be devised to scan this huge structural space for ssODNs that bind to simple or complex targets[252-253]. These qualifications enable parallel profiling of differences in molecular content across a wide range of biological sources without prior knowledge of binding partners[254-255], but this potential has not been fully exploited to date.

Here we described how libraries of ssODNs can be used to profile plasma EVs from women with and without breast cancer. We introduce "Adaptive Dynamic Artificial Poly-ligand Targeting (ADAPT)", a novel approach for monitoring differences in the molecular content of plasma EVs in a massive parallel fashion without prior knowledge of the targets.

Material and Methods

Cell Culture and EVs Isolation

Cell culture: FBS was depleted of bovine EVs by centrifugation at 100,000g for 16 h at 4°C. Vcap cells were cultured in DMEM supplemented with 10% depleted FBS, 2 mM L-glutamine, 1 U/mL penicillin, and 1 µg/mL streptomycin at 37°C and 5% CO2. EVs were isolated from Vcap cell culture supernatant by sucrose density centrifugation. Briefly, supernatant was cleared of cells and cellular debris by sequential centrifugation at 400g for 10 min and 2000g for 20 min at 4°C. Cleared supernatant was concentrated by centrifugal filtration (Centricon Plus-70, 100 kDa NMWL), layered on a 30% sucrose cushion and centrifuged at 100,000g for 75 min at 4°C. Supernatant was removed and discarded without disrupting the cushion interface, which was then collected, diluted 6-fold with PBS and centrifuged at 100,000g for 70 min at 4°C. The resulting EV pellet was re-suspended in PBS by pipetting and incubation overnight at 4°C, and then stored at -80°C.

Ultracentrifugation: EVs were isolated from human plasma by direct ultracentrifugation. Pooled plasma from normal female donors was diluted 1:1 with PBS then centrifuged at 2000g for 30 min at 4 °C to eliminate contaminating cells, followed by 12,000g for 45 min at 4 °C to clear any remaining platelets and cellular debris. EVs were pelleted by centrifugation at 120,000g for 2 h at 4 °C. Pellets were washed twice by resuspension in PBS and subsequent centrifugation at 120,000g for 70 min at 4 °C. The final EV pellet was re-suspended and stored as described above.

Patients

Bio-specimens utilized in this experiment were obtained under an IRB-approved Bio-repository Protocol (see the list of IRB committees below). All subjects were consented with an IRB approved consent form and per 21 CFR 50.20 guidelines. Informed consent was obtained from all subjects. All experimental procedures in this publication were reviewed by the Western Institutional Review Board (WIRB) and deemed to fall under IRB Exemption per 45 CFR 46.101(b)(4). All experiments were approved by WIRB and performed in accordance with the Code of Federal guidelines and regulations. List of committees: Western IRB (Sponsor's Central IRB). Additional Local IRBs consist of: CAMC/WVU Charleston IRB; Phoebe Putney Memorial Hospital (PPMH) IRB; MedCentral IRB; LVHN IRB; Margaret R. Pardee Memorial Hospital IRB; Community Healthcare System Central IRB; Hartford Healthcare IRB; Renown Regional Medical Center IRB; John Muir Health IRB; Jupiter Medical Center IRB; Middlesex Hospital IRB; Orlando Health/Orlando; Regional Healthcare System IRB; MPM-SAH IRB; St. Luke's Hospital and Health Network IRB; CCHC IRB; MCW/FH IRB; Bon Secours Richmond Health System IRB; Roper-St. Francis Healthcare IRB; CPHS; Greater Glasgow & Clyde Health Board.

Demographics and Clinical Characteristics

Patients' information referred to the Appendix-C, Appendix-E, and Appendix-F.

Specimen Characteristics

Blood plasma was collected at multiple collection sites using standardized protocol and the centrifuges provided by Caris Life Sciences. Blood was collected using standard venipuncture technique into four purple-top EDTA tubes, the tubes were spun in the Labofuge 200 for 10 min at 5300 RPM with proper balance. Plasma was then collected and aliquots were transferred into the cryovials, all cryovials were immediately placed into the -80°C freezer until use.

Samples Combinations and Volume for the Following Assays

SELEX: 1.2 ml 59/60 Breast Cancer Biopsy Positive/Negative samples (Negative sample includes 30 biopsy negative patient and 30 self-declared normal) were mixed as respective cancer/non-cancer pool. Each round of enrichment requires 300 µl of pooled plasma. For probing, 200 µl of plasma from individual 500 samples in 3 technical replicates were used.

Library Design and Amplification

ssODN library design as shown in Figure 4.1A, ssODNs library with 35bp random variable region flanked by primer-binding sites. The library was synthesized by Integrated DNA Technologies (IDT, USA) as “ultramer” (high-fidelity version of the synthesis) and purified by standard desalting. In order to avoid the low diversity problem that might happen when reading the sample index (i7) in the Illumina Hiseq platform, the library was synthesized in 4 separated parts with adding 0-3 bases right before the i7 index sequence. 5’-phosphorylated forward primer 5’-TCG TCG GCA GCG TCA-3’

(Tm: 62.6°C), and 5'-biotinylated reverse primer 5'-CTA GCA TGA CTG CAG TAC GT-3' (Tm: 61.1°C) were used for library amplification by high fidelity Q5 DNA polymerase (NEB) with cycling condition: 98°C 30 s, followed by 10 cycles of 98°C 30 s, 60°C for 30 s and 72°C 1 min, and final elongation at 72°C for 5 min. PCR product was purified by the NucleoSpin PCR clean-up kit (Macherey-Nagel) with NTI buffer according to the manufactory's instruction. Purified dsDNA was quantified with a QuBit ds BR/HS assay kit (ThermoFisher). The dsDNA library was then mixed with Lambda Exonuclease (NEB) to degrade the strand with 5'-phosphorylation label at 37°C for 2 hrs. The enzyme was then inactivated at 80°C for 10 min. ssDNA was again purified by NucleoSpin columns with NTC buffer as instructed. Eluted ssDNA was quantified with the QuBit ssDNA assay kit (ThermoFisher).

Library Enrichment

Scheme of the library enrichment was shown in Figure 4.1B. Enrichments were performed for both cancer and non-cancer sample in separated routes. In both case, the initial amplified library (L0) were mixed with the cancer/non-cancer pool respectively; after incubation for 60min at ambient temperature, 20% PEG (Sigma-Aldrich) were added and mixed to final concentration of 6%. The samples were then inverted two times to mix, incubated for 15 min at 4°C, followed by centrifugation at 10,000g for 5 min at ambient temperature; the supernatant was then removed, the pellet was washed by 1.0 ml of 1X PBS with 3.0 mM MgCl2 and re-suspended in 100 μl H2O; the recovered libraries were then re-amplified, ssODN were then prepared and purified as described above.

In the rest of the enrichment rounds, a scheme of positive selection, counter selection followed by another positive selection was followed for each round. In detail, the ssODNs amplified from the initial round of selection from cancer/non-cancer patient above were incubated with the same sample type plasma and repeat the selection as described above. By the end of the selection, ssODNs were recovered and put into the plasma sample with counter sample type as counter selection with same procedure. In counter selection, instead of collect the ssODNs associated with the EV pellet, ssODNs in the supernatant were collected and directly used to repeat a positive selection as input. The ssODNs recovered from this positive selection step were then amplified. The library enriched from cancer sample pool was denoted as L1 and library enriched from non-cancer sample pool was denoted as L2. L1 and L2 were then mixed together and denoted as L3 and used for the next round of enrichment. Total three rounds of enrichment were performed as indicated in Figure 4.1B, the resulted final library gained approximately 10^6 different unique sequences.

Figure 4.1 Design of the ssODN Library and General Enrichment Scheme. A) Design of the ssODN library. The primer highlighted in blue contains a sequence, complementary to the Illumina sequencing primer. The ssODN library was synthesized in a four parts with a staggered design, which differ from each other in length by +0, +1, +2, and +3 nucleotides (green); B) Patterns of the forward strands of the index PCR product; C) ssODNs library enrichment scheme: a high-diversity molecule library (~10^{11}) is mixed with blood plasma from biopsy-positive cancer sample (denoted by C) as well as with plasma from biopsy negative samples (denoted as non-Cancer, nC). In the 2nd step non-bound ssODNs are removed with supernatant and bound molecules are collected in the EV pellet. In the 3rd step, ssODNs recovered from C are incubated with nC (and vice versa) and counter selection; non-binders ssODNs in the pellets are removed. The recovered ssODNs were used as input in in step 4, another positive selection same as step1 and the recovered ssODNs were amplified. Two separated libraries were obtained, L1 from cancer enrichment, L2 from non-cancer enrichment. L1 and L2 were then mixed to obtain the library of L3 and used for the next round. The enrichment process was repeated three times and the final L3 with diversity was estimated to be at ~10^{6}.

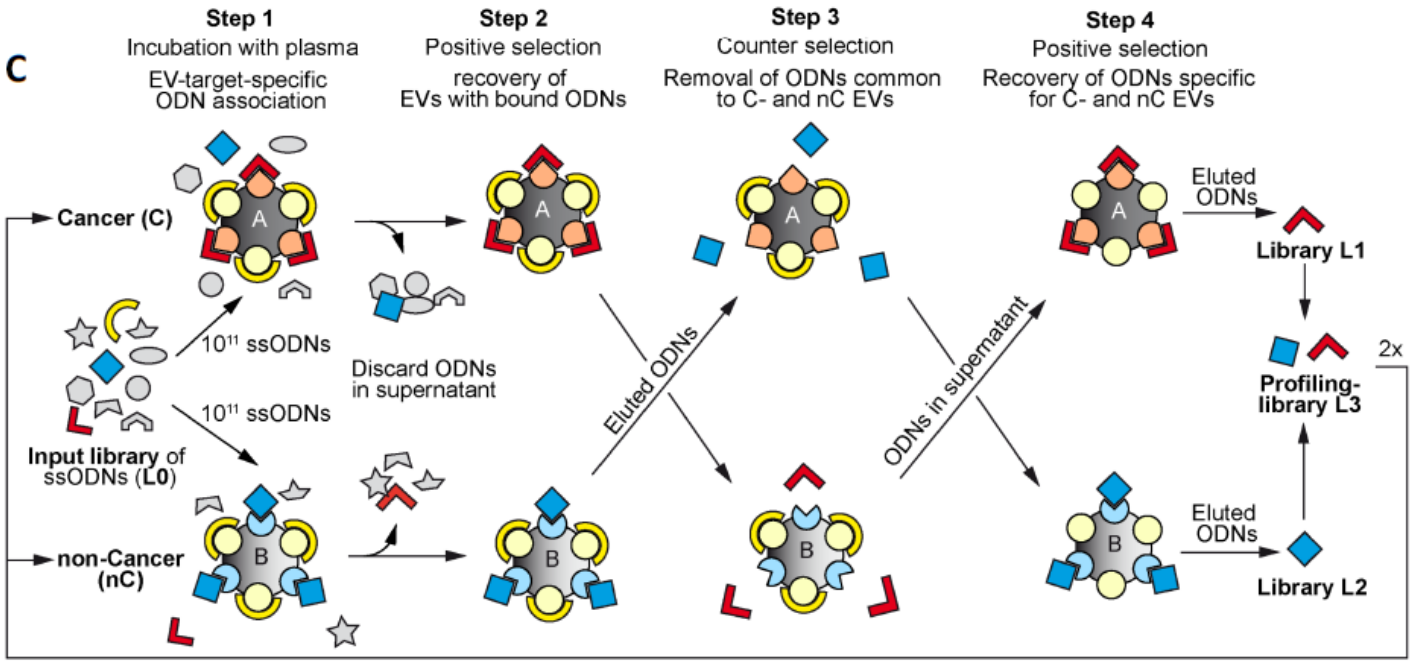
A

Synthesis	Sequence	Length	%GC in variable region
IDT Ultramer	5' CTAGCATGACTGCAGTACGT-35N-CTGTCTCTTATACACATCTGACGCTGCCGACGA 3'	88	50%
	5' CTAGCATGACTGCAGTACGT-35N-ACTGTCTCTTATACACATCTGACGCTGCCGACGA 3'	89	50%
	5' CTAGCATGACTGCAGTACGT-35N-GACTGTCTCTTATACACATCTGACGCTGCCGACGA 3'	90	50%
	5' CTAGCATGACTGCAGTACGT-35N-TGACTGTCTCTTATACACATCTGACGCTGCCGACGA 3'	91	50%

B

5' AATGATACGGCGACCACCGAGATCTACAC-[i5]-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-35n-ACGTACTGCAGTCATGCTAG-[i7]-ATCTCGTATGCCGTCTTCTGCTTG 3'
5' AATGATACGGCGACCACCGAGATCTACAC-[i5]-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGT-35n-ACGTACTGCAGTCATGCTAG-[i7]-ATCTCGTATGCCGTCTTCTGCTTG 3'
5' AATGATACGGCGACCACCGAGATCTACAC [i5] TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCT 35n ACGTACTGCAGTCATGCTAG [i7] ATCTCGTATGCCGTCTTCTGCTTG 3'
5' AATGATACGGCGACCACCGAGATCTACAC-[i5]-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACT-35n-ACGTACTGCAGTCATGCTAG-[i7]-ATCTCGTATGCCGTCTTCTGCTTG 3'

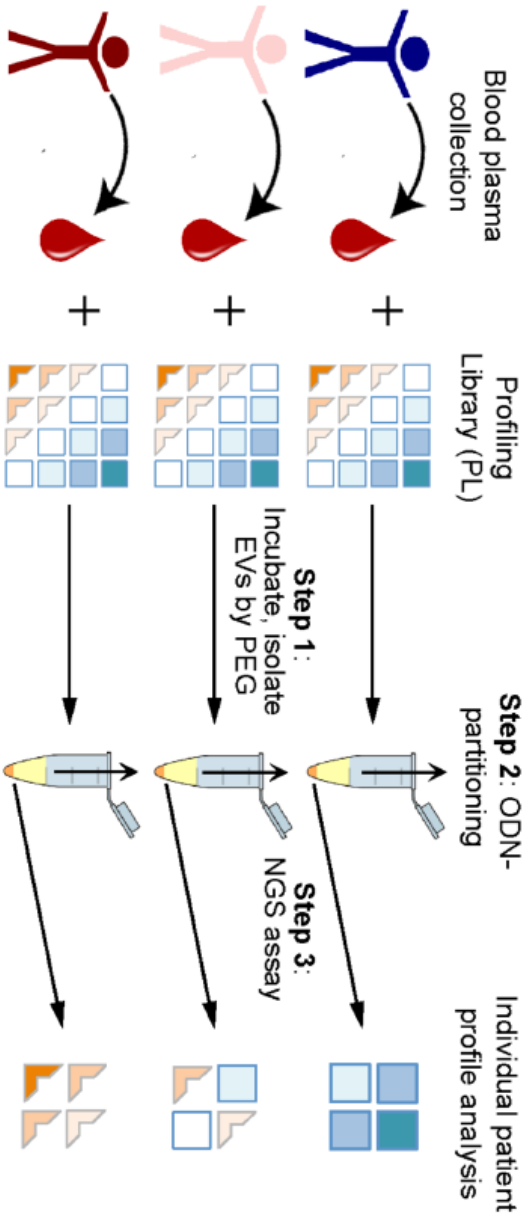
C



Protocol for Profiling Clinical Samples with Enriched or Synthetic Libraries

Scheme profiling clinical sample is shown in Figure 4.2. Frozen plasma (1.0 ml) was thaw and treated with protease inhibitor (20X stock of protease inhibitor (Roche). Cells/debris were removed by centrifugation (10000g, 20 min, 4°C), and the supernatant was collected. 1.0 ml of 2X PBS (SH3025601, GE Healthcare) containing 6.0 mM $MgCl_2$ was added and thoroughly mixed. Dextran sulfate (APT-B006, Aptamer Sciences) were added to a final concentration of 0.01 mM, salmon sperm DNA (ThermoFisher) and tRNA (ThermoFisher) were added to a final 0.8ng/µl and mixed thoroughly. ssODNs aptamer probing library was added to a final concentration of 2.5pg/µl sample in 1X PBS, 3.0 mM MgCl2. The mixtures were then incubated for 1hr at ambient temperature with rotation followed by PEG EV precipitation at final concentration of 6% as described above. After re-suspended in H_2O, the re-suspended sample was then directly amplified and sequenced by NGS.

Figure 4.2 ADAPT Profile Scheme. In the 1st step plasma samples of breast cancer patients are incubated with the profiling library L3. After incubation EVs are precipitated, supernatant is then removed (step 2). EV-associated oligonucleotides are subjected to next generation sequencing (step 3, NGS) to obtain a patient individual profile.



Next Generation Sequencing Library Preparation and HiSeq Sequencing Protocol

One-step NGS prep protocol: libraries were amplified with 1 μM of both index primers as N7xx: CAA GCA GAA GAC GGC ATA CGG AT [i7] CTA GCA TGA CTG-CAG TAC GT and S5xx: AAT GAT ACG GCG ACC ACC GAG ATC TA AC [i5] TCG TCG GCA GCG TCA. PCRs were performed as the protocol described above. Index PCR products were then purified with AMPure beads (Beckman Coulter). The purified product was quantified with QuBit dsDNA kit (ThermoFisher), samples concentrations were normalized and then diluted to 14 nM before multiplexing.

HiSeq2500 sequencing set up for Rapid run mode (single read): 5 μl of each sample (2 nM) was mixed to make particular multiplex, then 5 μl of the pool was mixed with 5 μl of 0.2 M NaOH and incubated for 5 min at RT. This solution was further diluted to 7 pM with chilled HT1 buffer, denatured at 96°C for 2 min, and snap-cooled in 3:1 ice : water bath for 5 min. Sequencing was performed on the HiSeq 2500 (Illumina) TruSeq Rapid SR Cluster Kit - HS and sequencing kit TruSeq Rapid SBS Kit (50 cycle, enough for 74 cycles).

NGS data processing

After completion of sequencing, HiSeq Q score quality metrics were inspected to ensure the high accuracy of the base calling. Only runs with at least 85% of bases above Q30 (probability of correct base call is 99.9%) were considered. Runs with lower percentage were repeated.

HiSeq sequencing data were converted to the ".fastq" file according to the specification of bcl2fastq package from Illumina and then de-multiplexed by a

customized pipeline. Read counts for individual unique sequences were then summarized for each individual sample. Individual sequence counts were then normalized to the average total read output from all samples as the representation of relative abundance (normalized count), which is further used for the statistical analysis. Total output reads refer to all NGS reads for particular sample while valid reads refer to sequences reads with only "ATCG" bases and without undetermined bases such as "N".

Statistical Analysis

ANOVA Permutation test: Number of ANOVA F test significant aptamers was tested by 1000 permutations of the sample labels. A one-way ANOVA model was built for each aptamer using normalized count. P-value of significance was calculated as the probability of Aptamer sequence number with F-test less than 0.05 more than the correct label.

Random Forest Models: Random Forest model analysis was performed in R environment. The probability for individual sample to be cancer was calculated by out-of-bag (OOB) predictions. ROC curve was then plotted from the individual sample OOB predication. Prediction performances denoted by Area under the curve (AUC) were calculated. AUCs from sample label permutation for 1000 times were also calculated. The p-value for statistical significance was defined by the number of permutation AUC that is equal or more than the AUC from original label.

Results

Library Enrichment Performance

Aptamer libraries were enriched against EV fraction from cancer and non-cancer plasma pool, after three rounds enrichments of positive-counter-positive selections, the efficiencies of final library enrichment for L1 and L2 was assessed by average count per unique sequence as shown in Figure 4.3A. In the original library L0, count per sequence is below 5 counts per sequence, this is because the diversity in the original library was so high that is basically saturated the sequencing capacity, which reflected in a low count per sequence number. As the enrichments went on, diversity of the sequences was significantly reduced due to the selection pressure; ssODNs that bound to the EV fraction from plasma samples were able to be maintained and gained more advantages in PCR amplification while ssODNs that not bound to the EV fraction were most likely to be removed and became less likely to be amplified during PCR steps. As ssODNs diversity reduced but NGS outputs were at similar levels, numbers of count per sequence would be expected to increase. As expected, the count per ssODNs in L1 and L2 libraries were more than 10 folds higher than the corresponding L0 library, indicating the enrichment efficiency after the three rounds of enrichment, the ssODNs potentially bound to the EV fraction in the plasma have been enriched more than 10 folds on average. At the same time, as shown in Figure 4.3B, the average output of the NGS output for L0, L1 and L2 library were at similar levels, demonstrated by both total output reads (T) and valid output reads (V).

After obtaining enriched library L1 and L2 in each round of enrichment were then mixed together to gain library L3. The rationale behind this is that assuming L1 library

was enriched against specific targets in plasma EVs from cancer patient and L2 library was enriched specific targets against plasma EVs in non-cancer samples. When mixed, it would be expected that the ssODNs specifically associated with cancer plasma EVs would be more likely to be higher in cancer sample compared to non-cancer samples and *vice versa*. The ssODN diversity of the final L3 library was close to 10^6 and the average count per sequence was greater than 50 (Appendix B).

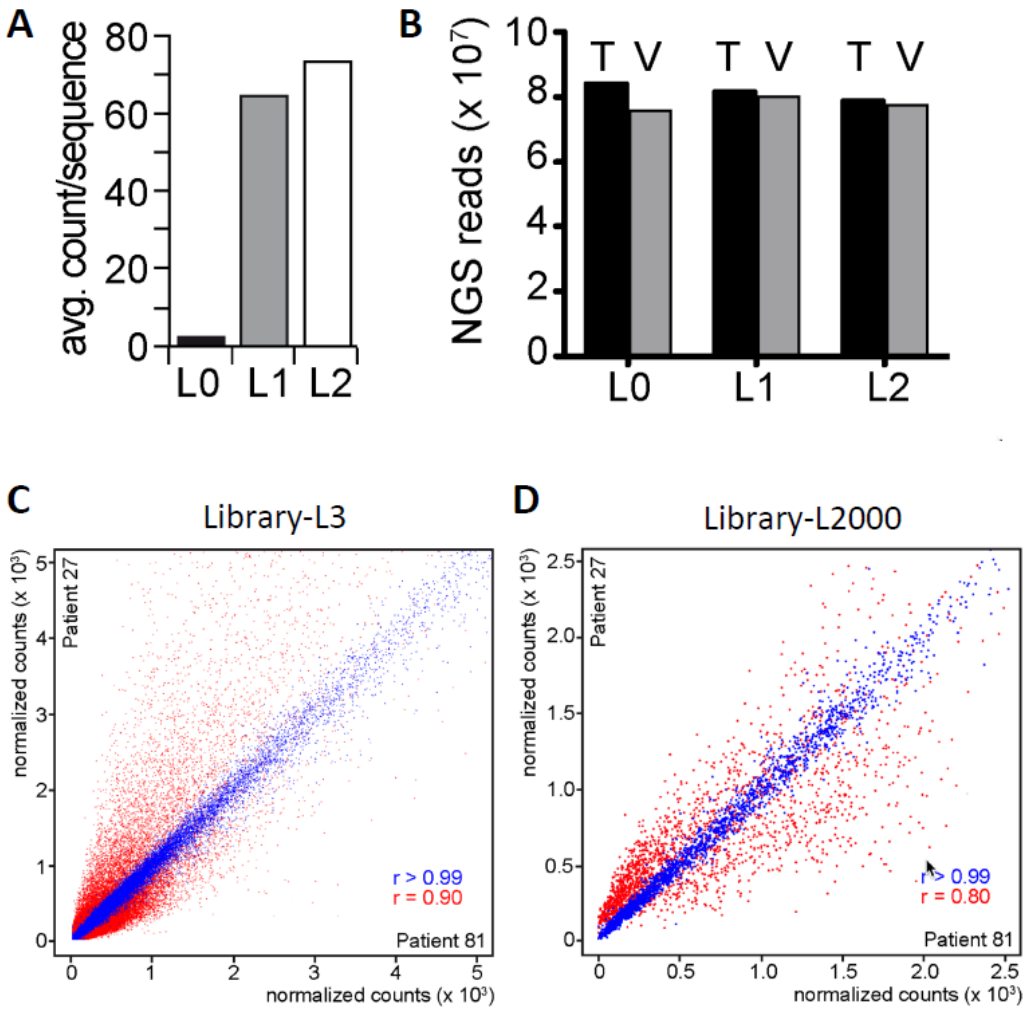
Profiles of relative abundance of the individual aptamer sequences could potentially contains signature pattern to that particular sample as well as pattern that might be specific to the sample status groups (cancer and non-cancer). After the enrichment, the ssODN sequences in the library were ranked according to their relative abundance. However, the quality of such library may not easy to be controlled due to the huge gaps in relative abundance for different sequences. In order to select appropriate ssODNs sequences and reconstruct a more stable library that could be used for the purpose of profiling different samples in large scale study, multiple small-scale profiling studies were performed by the enriched L3 library as probing library (Appendix-C for patient cohort; Appendix-D for individual sequences selected), including cancer vs non-cancer as well as between different cohorts inside cancer group. ssODNs towards either side of the comparisons were selected respectively based on multiple criterions, including t test, fold changes as well as neutral sequences. Total 2000 sequences were selected and chemically synthesized. The new profiling library L2000 was then reconstituted with equal starting concentration of individual ssODNs.

To evaluate whether the re-synthesized L2000 library retained comparable information from L3 library, same plasma samples were profiled by both L3 and L2000

library as shown in Figure 4.3C/D. In general, when comparing different technical replicates for the same sample, the technical variation of the individual profiles is expected to be low, therefore most of the ssODNs are expected to show similar relative abundance and the distribution of the relative abundance of the ssODNs is concentrated on the line of y=x (45° line) graphically. As expected, for the ssODNs profiles between technical replicates were very consistent. The variation of relative abundance of individual ssODNs from the profiles by either L3 or L2000 libraries were very low and display high correlation to the 45° line. Correlations between the technical replicates were more than 99% for both L3 and L2000 libraries. This indicates within the same sample, the ssODNs profiles were very consistent. In contrast, when comparing different sample against each other, the sample variation would be expected much higher in assumption that specific of ssODN pattern in different samples. As showed in Figure 4.3C/D (red), when comparing ssODN profiles from different samples by either L3 or L2000 library, much higher inter-sample variance was observed. Correlation between the different samples reduced from 99% in technical repeats to 90% in inter-sample comparison for L3 library, 99% to 80% for L2000 library respectively. Also, as shown in the figure, the relative abundance of ssODNs were also distributed more scattered (red) inter-sample comparison compared to that from technical replicates(blue), suggesting the potential information represented by the profiles of ssODNs in the original L3 library was retained in the new synthesized, reconstituted L2000 from the selected ssODNs.

Figure 4.3 Performance of Library Enrichment. A) Count per sequence comparison between the initial library(L0) and enriched library (L1 and L2); B) NGS output for the initial library (L0) and enriched library(L1 and L2) in term of total output (T) and valid

output(V); C) Correlation between technical replicates and biological samples for L3. The scatter plot shows the distribution of normalized counts of aptamers recovered from ADAPT with the plasma aliquots of the same patient (blue dots, $r > 0.99$; intra sample: patient 27) or plasma samples from different patients (red dots, $r = 0.90$). Every individual dot represents a unique sequence with the count of that sequence corresponding to representation in different samples or technical replicates; D) Correlation between technical replicates and biological samples for L2000. Scatter plot shows distribution of normalized counts of aptamers recovered from ADAPT with the plasma aliquots of the same patient (blue dots, $r > 0.99$; intra sample: patient 27) or plasma samples from different patients (red dots, $r = 0.80$). Every individual dot represents a unique sequence with the count of that sequence corresponding to representation in different samples or technical replicates.


A
avg. count/sequence
80
60
40
20
0
L0
L1
L2
B
NGS reads (x 10^7)
10
8
6
4
2
0
T V
T V
T V
L0
L1
L2
C
Library-L3
normalized counts (x 10^3)
Patient 27
r > 0.99
r = 0.90
Patient 81
normalized counts (x 10^3)
D
Library-L2000
normalized counts (x 10^3)
Patient 27
r > 0.99
r = 0.80
Patient 81
normalized counts (x 10^3)

Clinical Utility of ADAPT Profiling System

To evaluate the potential clinical utilities of L2000 library, whose sequence was selected from small amount of training samples, and whether the library were able to obtain profiles that could differentiate cancer patients from non-cancer controls other than the training samples, a large scale study with 500 patients, including 206 breast cancer patients (Appendix-E), 177 non-cancer patients (Appendix-F), and 117 self-declared healthy controls were performed and the profiles from respective groups were compared.

The 500 patient samples were profiled by the L2000 library in triplicates, relative abundance was calculated by normalizing individual sequence counts to the average output from the 2000 ssODN sequences total output among all samples instead of total counts indicated above (Appendix-G).

To find out if there were any different patterns detected between the groups, one-way ANOVA F-statistic comparison for each individual aptamer sequence between cancer, non-cancer and healthy controls was calculated. As shown in Figure 4.4A, there are 390 aptamer sequences showed statistically significant ($p<0.05$, red line), indicating for these 390 individual aptamer sequences, at least one group among cancer, non-cancer and health controls was found to be significantly different from the other two. If this is a true performance that might implicate that a specific pattern could exist in differentiating the different groups, otherwise, it might implicate the performance is a random effect. Therefore, in order to assess whether such specific number of ssODNs that reached statistical significance was a result of random effect, a permutation analysis was performed as shown in Figure 4.4A. Specifically, the sample labels (cancer, non-cancer

and healthy controls) were randomly assigned to sample for 1000 times, in each permutation, the one-way ANOVA F-statistic were re-calculated, the number of aptamer sequences that pass the significant level ($p<0.05$) were collected. If the F-statistic was from a random effect, it would expect similar or higher number of ssODN sequences that are able to pass the significant level in majority of the random permutations. In contrast, as shown in Figure 4.4A, most of the random permutation tests only showed less than 100 sequences that pass the F-statistic significant level, much less than the 390 from the correct sample label performance. Number of permutation that have equal of higher Aptamer sequence that pass the F-statistic significant level is less than 50, or less than 5% of the total 1000 random permutations, indicating the number of ssODNs that could pass significant level in the F-statistic with correct sample label is statistically significant ($p<0.05$), suggesting such number of ssODNs passing the statistical significant level for the F-statistics was not a random effect, further implying specific patterns that represented by the ssODN relative abundance from different sample groups may exist.

To further evaluate the ADAPT profile could differentiate cancer sample from all non-cancer samples (including non-caner and healthy controls) as well as its potential prediction performance, prediction models based on random forest (RF) were built. For each model, feature selection was carried out based on the significance of a particular ssODN sequence comparison between cancer and control samples from the training data, including t-test ($p < 0.05$), Wilcoxon rank-sum test ($p < 0.05$), Kolmogorov-Smirnov test ($p < 0.05$) or fold change (>1.2); features that passed these filters in the training set were then used to build the RF model; Out-Of-Box (OOB) was the calculated as the predicted

probability for particular sample to be cancer. All individual sample OOB prediction probabilities were then collected and used to calculate the overall AUC performance.

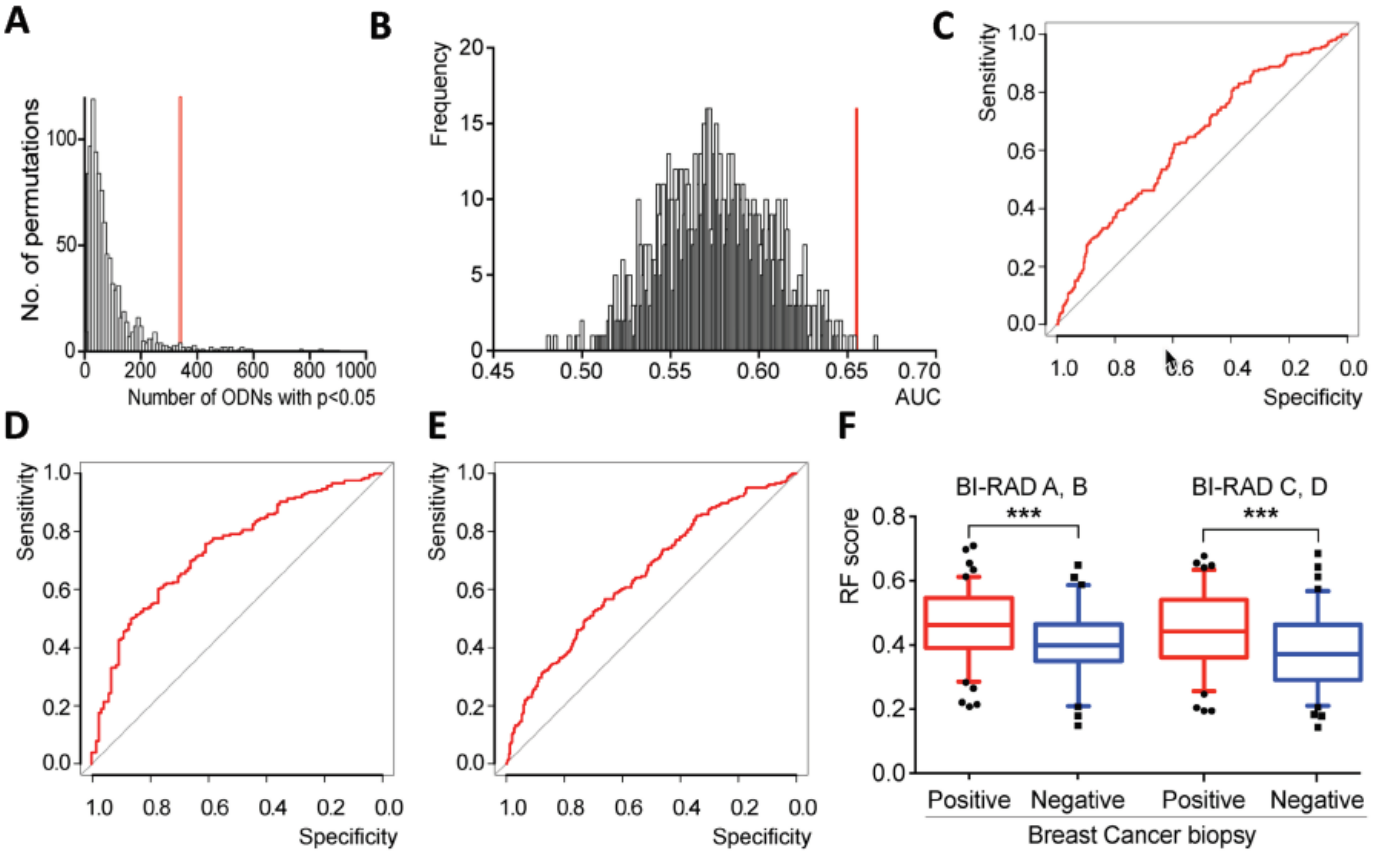
As shown in Figure 4.4B, the overall performance AUC of cancer vs non-cancer samples was 0.66 (red line). Again, to exclude the probability that such AUC performance was a result of random effect, sample labels were then permutated 1000 times, AUC performances from each permutation were re-calculated by the RF models as described above. If the prediction performance AUC of 0.66 was from a random effect, the chance of having AUC performance from randomly permutated datasets to have equal of even better AUC performance than the original label order should not be statistically significant ($p>0.05$). In contrast, most of the AUC performance from the 1000 permutations were less than 0.66 (Figure 4.4B, grey histogram). The number of AUC with equal or better than AUC 0.66 was less than 2 out of 1000 ($p=0.002 < 0.05$). Therefore, there was less than 5% of chance that a AUC performance of 0.66 could be obtained by a random chance, which was statistically significant to reject the null hypothesis that the current AUC of 0.66 was from a random effect, indicating the current prediction performance for the current dataset was unique and suggesting a specific information in the dataset was captured by the RF model, which differentiated the cancer sample from all other controls.

RF model were also applied to differentiate cancer vs non-cancer and cancer vs. healthy controls. Predictive AUC values for each comparison were 0.64, 0.73 respectively with permutation p-values of 0.024, 0.001 (Figure 4.4C, D). The above AUC for cancer vs all non-cancer controls was shown in Figure 4.4E. These analyses indicated that the specific prediction performance found between these groups in comparison were

not by a random effect, confirming that certain specific pattern or features were captured by the RF model so that cancer samples could be separated from non-cancer and healthy controls to some degree.

Interestingly, there was a significant increase in the overall prediction performance for differentiating cancer samples from healthy donors in comparison to the performance in differentiating cancer sample from non-cancer controls (0.73 vs 0.64), implying that the L2000 library may be measuring molecules pattern that related to the abnormalities within the breast other than cancer status itself. To exclude such possibility, OOB predicted cancer probability for biopsy-positive cancer samples and biopsy-negative non-cancer samples were then grouped according to their Bi-RAD composition categories[256], an indicator of breast tissue density (low grade: Bi-RAD category A/B, high grade: Bi-RAD category C/D). As shown in Figure 4.4F, there is no significant difference in the cancer predicted probability ($p>0.05$) between Bi-RAD low grade and high-grade groups within the biopsy-positive samples; similarly, no significant different in the cancer predicted probability between Bi-RAD low grade and high-grade groups within the biopsy-negative samples ($p>0.05$). In contrast, significant different was found between the biopsy-positive cancer and the biopsy-negative controls within the same Bi-RAD grade groups, indicating the L2000 ssODN profiles were able to differentiating the cancer from non-cancer sample regardless of their breast density status, confirming that the L2000 library was able to detect some intrinsic difference from the cancer status rather than some irrelevant physical condition of the breast.

Figure 4.4 Statistical Analysis of the Large-Scale Study. A) Number of sequence that pass of F-test significance ($p<0.05$) among cancer, non-cancer and healthy controls (red line), frequency distribution of number of sequence passes F-test significance from the 1000 permutation. The performance of obtained that number of sequence that could pass the F-test significance is statistically significant ($p<0.05$). *P-val is the number of permutation with number of sequence passes the F-test significance equal of larger than the original correct label; B) Overall prediction performance AUC is 0.66 (red line, cancer vs others) and frequency distribution of the prediction performance AUC from 1000 permutation tests ($p=0.002 < 0.05$). *p value is the number of permutation with equal of higher prediction performance AUC than the original correct label; C) ROC curve for cancer versus control samples with AUC = 0.66, p=0.002; D) ROC curve for cancer versus healthy donors with AUC = 0.73; p=0.001; E) ROC curve for cancer versus non-cancer with AUC = 0.64; p=0.024; F) Prediction probability comparison between cancer (biopsy positive) and non-cancer (biopsy negative) grouped by Bi-RAD component categories (low category: A&B); high category: C&D), ***$p<0.05$.


A
No. of permutations
Number of ODNs with p<0.05
B
Frequency
AUC
C
Sensitivity
Specificity
D
Sensitivity
Specificity
E
Sensitivity
Specificity
F
BI-RAD A, B
BI-RAD C, D
***
***
RF score
Positive
Negative
Positive
Negative
Breast Cancer biopsy

Discussion

Mapping the protein universe is a daunting task. Charting the interactions of proteins in the human proteome, studying their function and how their expression changes in health and disease, may take decades. Compared with about 40,000 genes in the genome, the human proteome is predicted to top 1 million[257]. Traditional proteomics uses techniques such as two-dimensional gel electrophoresis or chromatography, combined with mass spectrometry, to separate and identify proteins on a large scale[258]. Tracking protein interactions and activity and determining functions at this scale is an even greater task, but is beginning to be tackled by protein arrays and multiplexed assays[257]. However, large scale protein array/chip technology also faces immense challenges. For example, one bottleneck that array/chip manufacturers face is that not enough antibodies generated against every human proteins as well as protein isoforms, especially antibodies with high affinity and specificity were even harder to obtain, which would put heavy limitation on the scales; another technical hurdle to producing a reliable protein arrays/chips is the protein stability and the corresponding shelf-life. Most protein arrays use antibodies as the immobilized capture probes. Such process of depositing antibodies onto a solid surface could potentially denature the antibody or affect its recognition properties. In addition, the shelf-life of the antibody products are also limited[257, 259].

Competing with antibodies will be various nucleic acid ligand-like molecules that have been termed aptamers, either RNA or DNA[260]. Nucleic acid aptamers can be selected in vitro based on repeated cycles of binding and PCR. A major advantage of this class of molecules is that it is very straightforward to carry out affinity maturation,

termed SELEX, in the process of the selection, allowing very high affinity ligands to be obtained[261]. Given that they are highly negatively charged molecules, it is likely that nucleic acid aptamers will work well with targets containing positive charge patches on their surface[262]. Traditional SELEX adopted a bottom-up approach as affinity maturation was in general focused on individual purified target first and then start to interpret the potential complex functions. Bottom-up SELEX approach faces multiple problems and limitations. Similar to the problem that are faced by protein/antibody-array, Bottom-up SELEX would not meet the demand of the increasing number of newly discovered novel proteins as well as isoforms in the coming future[257] because the scale of isolation of such huge number of individual proteins is expected to be even more enormous and complicated than the human genome project. It is unlikely the development of aptamer for individual target would be able to catch up. Moreover, on the top of the enormous number of individual proteins, protein interactions that form protein complex and machinery further made the potential functional characterization even more complicated.

Interestingly, recent studies showed that cell-based SELEX as well as in vivo SELEX were both able to obtain specific aptamers that targeting to tumor cell surface[80-84], and even able to identify brain-penetrating aptamers[85], suggesting aptamer-based SELEX affinity maturation was not necessary restricted by the single pure target. With the increasing NGS sequencing capability, a top-down SELEX approach might be possible, which is more focus on generating an overall signature profile from complex samples as well as developing its potential utility first before targets are identified.

Here we have demonstrated a top-down SELEX strategy, where the SELEX affinity maturation was performed using a complex sample and generated affinity

matured libraries. Individual ssODN aptamers were then further selected according to the profiles generated by these affinity-matured libraries and combined as a new multiplexing detection system, termed "ADAPT". ADAPT platform offers high multiplexing capability to generate an unbiased profiling of molecular signatures from complex samples. In the current study, DNA aptamer SELEX affinity maturation were performed on the plasma EV fraction, clinical application was explored by comparing ADAPT profiles from breast cancer and non-cancer patients as well as heathy control samples.

ADAPT platform offers many advantages that offered by other profiling platforms, such as protein, antibody array as well as mass spectrometry and it is compatible to liquid biopsy, multiplexing with high-throughputs. At the same time ADAPT was able to avoid certain disadvantages from other systems, for example ADAPT requires no prior knowledge on the nature of intended targets, therefore sample complexity was not a major hurdle; DNA Aptamers were more stable and easier for storage than proteins like antibodies, therefore the shelf-life of the reagent may be improved significantly.

Few top-down style aptamer-based molecular profiling approaches have been published to date. The closest one is SOMAscan, a Bottom-up platform that utilized more than 1000 aptamers, which was selected on individual known specific targets[263], has been adopted for the proteomic analysis of EVs from the prostate cancer cell line DU145[264]. In comparison, the potential multiplexing and information could be much higher in ADAPT platform. While ADAPT in its current form still demands for further improvement to bring it to a clinical standard, with more extensive libraries training and

quality improvement, information content and better diagnostic value would become available, which will provide better tool for the in-parallel massive analysis of complex samples and clinical utilities.

REFERENCES

1. Hessvik, N. P.; Llorente, A., Current knowledge on exosome biogenesis and release. *Cell Mol Life Sci* **2018,** *75* (2), 193-208.

2. Wolf, P., The nature and significance of platelet products in human plasma. *Br J Haematol* **1967,** *13* (3), 269-88.

3. Melo, S. A.; Sugimoto, H.; O'Connell, J. T.; Kato, N.; Villanueva, A.; Vidal, A.; Qiu, L.; Vitkin, E.; Perelman, L. T.; Melo, C. A.; Lucci, A.; Ivan, C.; Calin, G. A.; Kalluri, R., Cancer exosomes perform cell-independent microRNA biogenesis and promote tumorigenesis. *Cancer Cell* **2014,** *26* (5), 707-21.

4. Johnstone, R. M.; Adam, M.; Hammond, J. R.; Orr, L.; Turbide, C., Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). *J Biol Chem* **1987,** *262* (19), 9412-20.

5. Baixauli, F.; Lopez-Otin, C.; Mittelbrunn, M., Exosomes and autophagy: coordinated mechanisms for the maintenance of cellular fitness. *Front Immunol* **2014,** *5*, 403.

6. Raposo, G.; Stoorvogel, W., Extracellular vesicles: exosomes, microvesicles, and friends. *J Cell Biol* **2013,** *200* (4), 373-83.

7. Yanez-Mo, M.; Siljander, P. R.; Andreu, Z.; Zavec, A. B.; Borras, F. E.; Buzas, E. I.; Buzas, K.; Casal, E.; Cappello, F.; Carvalho, J.; Colas, E.; Cordeiro-da Silva, A.; Fais, S.; Falcon-Perez, J. M.; Ghobrial, I. M.; Giebel, B.; Gimona, M.; Graner, M.; Gursel, I.; Gursel, M.; Heegaard, N. H.; Hendrix, A.; Kierulf, P.; Kokubun, K.; Kosanovic, M.; Kralj-Iglic, V.; Kramer-Albers, E. M.; Laitinen, S.; Lasser, C.; Lener, T.; Ligeti, E.; Line, A.; Lipps, G.; Llorente, A.; Lotvall, J.; Mancek-Keber, M.; Marcilla, A.; Mittelbrunn, M.; Nazarenko, I.; Nolte-'t Hoen, E. N.; Nyman, T. A.; O'Driscoll, L.; Olivan, M.; Oliveira, C.; Pallinger, E.; Del Portillo, H. A.; Reventos, J.; Rigau, M.; Rohde, E.; Sammar, M.; Sanchez-Madrid, F.; Santarem, N.; Schallmoser, K.; Ostenfeld, M. S.; Stoorvogel, W.; Stukelj, R.; Van der Grein, S. G.; Vasconcelos, M. H.; Wauben, M. H.; De Wever, O., Biological properties of extracellular vesicles and their physiological functions. *J Extracell Vesicles* **2015,** *4*, 27066.

8. Orozco, A. F.; Lewis, D. E., Flow cytometric analysis of circulating microparticles in plasma. *Cytometry A* **2010,** *77* (6), 502-14.

9. Mathivanan, S.; Ji, H.; Simpson, R. J., Exosomes: extracellular organelles important in intercellular communication. *J Proteomics* **2010,** *73* (10), 1907-20.

10. Record, M.; Carayon, K.; Poirot, M.; Silvente-Poirot, S., Exosomes as new vesicular lipid transporters involved in cell-cell communication and various pathophysiologies. *Biochim Biophys Acta* **2014,** *1841* (1), 108-20.

11. Peinado, H.; Aleckovic, M.; Lavotshkin, S.; Matei, I.; Costa-Silva, B.; Moreno-Bueno, G.; Hergueta-Redondo, M.; Williams, C.; Garcia-Santos, G.; Ghajar, C.; Nitadori-Hoshino, A.; Hoffman, C.; Badal, K.; Garcia, B. A.; Callahan, M. K.; Yuan, J.; Martins, V. R.; Skog, J.; Kaplan, R. N.; Brady, M. S.; Wolchok, J. D.; Chapman, P. B.; Kang, Y.;

Bromberg, J.; Lyden, D., Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat Med* **2012,** *18* (6), 883-91.

12. Costa-Silva, B.; Aiello, N. M.; Ocean, A. J.; Singh, S.; Zhang, H.; Thakur, B. K.; Becker, A.; Hoshino, A.; Mark, M. T.; Molina, H.; Xiang, J.; Zhang, T.; Theilen, T. M.; Garcia-Santos, G.; Williams, C.; Ararso, Y.; Huang, Y.; Rodrigues, G.; Shen, T. L.; Labori, K. J.; Lothe, I. M.; Kure, E. H.; Hernandez, J.; Doussot, A.; Ebbesen, S. H.; Grandgenett, P. M.; Hollingsworth, M. A.; Jain, M.; Mallya, K.; Batra, S. K.; Jarnagin, W. R.; Schwartz, R. E.; Matei, I.; Peinado, H.; Stanger, B. Z.; Bromberg, J.; Lyden, D., Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat Cell Biol* **2015,** *17* (6), 816-26.

13. Hoshino, A.; Costa-Silva, B.; Shen, T. L.; Rodrigues, G.; Hashimoto, A.; Tesic Mark, M.; Molina, H.; Kohsaka, S.; Di Giannatale, A.; Ceder, S.; Singh, S.; Williams, C.; Soplop, N.; Uryu, K.; Pharmer, L.; King, T.; Bojmar, L.; Davies, A. E.; Ararso, Y.; Zhang, T.; Zhang, H.; Hernandez, J.; Weiss, J. M.; Dumont-Cole, V. D.; Kramer, K.; Wexler, L. H.; Narendran, A.; Schwartz, G. K.; Healey, J. H.; Sandstrom, P.; Labori, K. J.; Kure, E. H.; Grandgenett, P. M.; Hollingsworth, M. A.; de Sousa, M.; Kaur, S.; Jain, M.; Mallya, K.; Batra, S. K.; Jarnagin, W. R.; Brady, M. S.; Fodstad, O.; Muller, V.; Pantel, K.; Minn, A. J.; Bissell, M. J.; Garcia, B. A.; Kang, Y.; Rajasekhar, V. K.; Ghajar, C. M.; Matei, I.; Peinado, H.; Bromberg, J.; Lyden, D., Tumour exosome integrins determine organotropic metastasis. *Nature* **2015,** *527* (7578), 329-35.

14. Kowal, J.; Tkach, M.; Thery, C., Biogenesis and secretion of exosomes. *Curr Opin Cell Biol* **2014,** *29*, 116-25.

15. Kowal, J.; Arras, G.; Colombo, M.; Jouve, M.; Morath, J. P.; Primdal-Bengtson, B.; Dingli, F.; Loew, D.; Tkach, M.; Thery, C., Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc Natl Acad Sci U S A* **2016,** *113* (8), E968-77.

16. Takahashi, A.; Okada, R.; Nagao, K.; Kawamata, Y.; Hanyu, A.; Yoshimoto, S.; Takasugi, M.; Watanabe, S.; Kanemaki, M. T.; Obuse, C.; Hara, E., Exosomes maintain cellular homeostasis by excreting harmful DNA from cells. *Nat Commun* **2017,** *8*, 15287.

17. Sugiura, A.; McLelland, G. L.; Fon, E. A.; McBride, H. M., A new pathway for mitochondrial quality control: mitochondrial-derived vesicles. *EMBO J* **2014,** *33* (19), 2142-56.

18. Campanella, C.; Bucchieri, F.; Merendino, A. M.; Fucarino, A.; Burgio, G.; Corona, D. F.; Barbieri, G.; David, S.; Farina, F.; Zummo, G.; de Macario, E. C.; Macario, A. J.; Cappello, F., The odyssey of Hsp60 from tumor cells to other destinations includes plasma membrane-associated stages and Golgi and exosomal protein-trafficking modalities. *PLoS One* **2012,** *7* (7), e42008.

19. Keerthikumar, S.; Chisanga, D.; Ariyaratne, D.; Al Saffar, H.; Anand, S.; Zhao, K.; Samuel, M.; Pathan, M.; Jois, M.; Chilamkurti, N.; Gangoda, L.; Mathivanan, S., ExoCarta: A Web-Based Compendium of Exosomal Cargo. *J Mol Biol* **2016,** *428* (4), 688-692.

20. Stoorvogel, W.; Kleijmeer, M. J.; Geuze, H. J.; Raposo, G., The biogenesis and functions of exosomes. *Traffic* **2002,** *3* (5), 321-30.

21. Hoshino, D.; Kirkbride, K. C.; Costello, K.; Clark, E. S.; Sinha, S.; Grega-Larson, N.; Tyska, M. J.; Weaver, A. M., Exosome secretion is enhanced by invadopodia and drives invasive behavior. *Cell Rep* **2013,** *5* (5), 1159-68.

22. Mittelbrunn, M.; Gutierrez-Vazquez, C.; Villarroya-Beltri, C.; Gonzalez, S.; Sanchez-Cabo, F.; Gonzalez, M. A.; Bernad, A.; Sanchez-Madrid, F., Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells. *Nat Commun* **2011,** *2*, 282.

23. Sinha, S.; Hoshino, D.; Hong, N. H.; Kirkbride, K. C.; Grega-Larson, N. E.; Seiki, M.; Tyska, M. J.; Weaver, A. M., Cortactin promotes exosome secretion by controlling branched actin dynamics. *J Cell Biol* **2016,** *214* (2), 197-213.

24. Kreger, B. T.; Johansen, E. R.; Cerione, R. A.; Antonyak, M. A., The Enrichment of Survivin in Exosomes from Breast Cancer Cells Treated with Paclitaxel Promotes Cell Survival and Chemoresistance. *Cancers (Basel)* **2016,** *8* (12).

25. Stenmark, H., Rab GTPases as coordinators of vesicle traffic. *Nat Rev Mol Cell Biol* **2009,** *10* (8), 513-25.

26. Livshits, M. A.; Khomyakova, E.; Evtushenko, E. G.; Lazarev, V. N.; Kulemin, N. A.; Semina, S. E.; Generozov, E. V.; Govorun, V. M., Isolation of exosomes by differential centrifugation: Theoretical analysis of a commonly used protocol. *Sci Rep* **2015,** *5*, 17319.

27. La Marca, V.; Fierabracci, A., Insights into the Diagnostic Potential of Extracellular Vesicles and Their miRNA Signature from Liquid Biopsy as Early Biomarkers of Diabetic Micro/Macrovascular Complications. *Int J Mol Sci* **2017,** *18* (9).

28. M, H. R.; Bayraktar, E.; G, K. H.; Abd-Ellah, M. F.; Amero, P.; Chavez-Reyes, A.; Rodriguez-Aguayo, C., Exosomes: From Garbage Bins to Promising Therapeutic Targets. *Int J Mol Sci* **2017,** *18* (3).

29. Lotvall, J.; Hill, A. F.; Hochberg, F.; Buzas, E. I.; Di Vizio, D.; Gardiner, C.; Gho, Y. S.; Kurochkin, I. V.; Mathivanan, S.; Quesenberry, P.; Sahoo, S.; Tahara, H.; Wauben, M. H.; Witwer, K. W.; Thery, C., Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. *J Extracell Vesicles* **2014,** *3*, 26913.

30. Helwa, I.; Cai, J.; Drewry, M. D.; Zimmerman, A.; Dinkins, M. B.; Khaled, M. L.; Seremwe, M.; Dismuke, W. M.; Bieberich, E.; Stamer, W. D.; Hamrick, M. W.; Liu, Y., A Comparative Study of Serum Exosome Isolation Using Differential Ultracentrifugation and Three Commercial Reagents. *PLoS One* **2017,** *12* (1), e0170628.

31. Caradec, J.; Kharmate, G.; Hosseini-Beheshti, E.; Adomat, H.; Gleave, M.; Guns, E., Reproducibility and efficiency of serum-derived exosome extraction methods. *Clin Biochem* **2014,** *47* (13-14), 1286-92.

32. Van Deun, J.; Mestdagh, P.; Sormunen, R.; Cocquyt, V.; Vermaelen, K.; Vandesompele, J.; Bracke, M.; De Wever, O.; Hendrix, A., The impact of disparate isolation methods for extracellular vesicles on downstream RNA profiling. *J Extracell Vesicles* **2014,** *3*.

33. van der Pol, E.; Boing, A. N.; Harrison, P.; Sturk, A.; Nieuwland, R., Classification, functions, and clinical relevance of extracellular vesicles. *Pharmacol Rev* **2012,** *64* (3), 676-705.

34. Rood, I. M.; Deegens, J. K.; Merchant, M. L.; Tamboer, W. P.; Wilkey, D. W.; Wetzels, J. F.; Klein, J. B., Comparison of three methods for isolation of urinary microvesicles to identify biomarkers of nephrotic syndrome. *Kidney Int* **2010,** *78* (8), 810-6.

35. Linares, R.; Tan, S.; Gounou, C.; Arraud, N.; Brisson, A. R., High-speed centrifugation induces aggregation of extracellular vesicles. *J Extracell Vesicles* **2015,** *4*, 29509.

36. Mathivanan, S.; Lim, J. W.; Tauro, B. J.; Ji, H.; Moritz, R. L.; Simpson, R. J., Proteomics analysis of A33 immunoaffinity-purified exosomes released from the human colon tumor cell line LIM1215 reveals a tissue-specific protein signature. *Mol Cell Proteomics* **2010,** *9* (2), 197-208.

37. Thery, C.; Zitvogel, L.; Amigorena, S., Exosomes: composition, biogenesis and function. *Nat Rev Immunol* **2002,** *2* (8), 569-79.

38. Sunkara, V.; Woo, H. K.; Cho, Y. K., Emerging techniques in the isolation and characterization of extracellular vesicles and their roles in cancer diagnostics and prognostics. *Analyst* **2016,** *141* (2), 371-81.

39. Ingham, K. C., Protein precipitation with polyethylene glycol. *Methods Enzymol* **1984,** *104*, 351-6.

40. Jin, Y.; Manabe, T., Analysis of PEG-fractionated high-molecular-mass proteins in human plasma by non-denaturing micro 2-DE and MALDI-MS PMF. *Electrophoresis* **2009,** *30* (20), 3613-21.

41. Sim, S. L.; He, T.; Tscheliessnig, A.; Mueller, M.; Tan, R. B.; Jungbauer, A., Branched polyethylene glycol for protein precipitation. *Biotechnol Bioeng* **2012,** *109* (3), 736-46.

42. Haque, R.; Ahmed, S. A.; Inzhakova, G.; Shi, J.; Avila, C.; Polikoff, J.; Bernstein, L.; Enger, S. M.; Press, M. F., Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades. *Cancer Epidemiol Biomarkers Prev* **2012,** *21* (10), 1848-55.

43. breast-cancer-facts-and-figures-2017-2018. *American Cancer Society*. *Breast Cancer Facts & Figures 2017-2018*. *Atlanta: American Cancer Society, Inc*. *2017*.

44. Perou, C. M.; Sorlie, T.; Eisen, M. B.; van de Rijn, M.; Jeffrey, S. S.; Rees, C. A.; Pollack, J. R.; Ross, D. T.; Johnsen, H.; Akslen, L. A.; Fluge, O.; Pergamenschikov, A.;

Williams, C.; Zhu, S. X.; Lonning, P. E.; Borresen-Dale, A. L.; Brown, P. O.; Botstein, D., Molecular portraits of human breast tumours. *Nature* **2000,** *406* (6797), 747-52.

45. Perou, C. M.; Borresen-Dale, A. L., Systems biology and genomics of breast cancer. *Cold Spring Harb Perspect Biol* **2011,** *3* (2).

46. Blows, F. M.; Driver, K. E.; Schmidt, M. K.; Broeks, A.; van Leeuwen, F. E.; Wesseling, J.; Cheang, M. C.; Gelmon, K.; Nielsen, T. O.; Blomqvist, C.; Heikkila, P.; Heikkinen, T.; Nevanlinna, H.; Akslen, L. A.; Begin, L. R.; Foulkes, W. D.; Couch, F. J.; Wang, X.; Cafourek, V.; Olson, J. E.; Baglietto, L.; Giles, G. G.; Severi, G.; McLean, C. A.; Southey, M. C.; Rakha, E.; Green, A. R.; Ellis, I. O.; Sherman, M. E.; Lissowska, J.; Anderson, W. F.; Cox, A.; Cross, S. S.; Reed, M. W.; Provenzano, E.; Dawson, S. J.; Dunning, A. M.; Humphreys, M.; Easton, D. F.; Garcia-Closas, M.; Caldas, C.; Pharoah, P. D.; Huntsman, D., Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med* **2010,** *7* (5), e1000279.

47. Collins, L. C.; Tamimi, R. M.; Baer, H. J.; Connolly, J. L.; Colditz, G. A.; Schnitt, S. J., Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. *Cancer* **2005,** *103* (9), 1778-84.

48. Singletary, S. E.; Allred, C.; Ashley, P.; Bassett, L. W.; Berry, D.; Bland, K. I.; Borgen, P. I.; Clark, G.; Edge, S. B.; Hayes, D. F.; Hughes, L. L.; Hutter, R. V.; Morrow, M.; Page, D. L.; Recht, A.; Theriault, R. L.; Thor, A.; Weaver, D. L.; Wieand, H. S.; Greene, F. L., Revision of the American Joint Committee on Cancer staging system for breast cancer. *J Clin Oncol* **2002,** *20* (17), 3628-36.

49. Godet, I.; Gilkes, D. M., BRCA1 and BRCA2 mutations and treatment strategies for breast cancer. *Integr Cancer Sci Ther* **2017,** *4* (1).

50. Allred, D. C., Ductal carcinoma in situ: terminology, classification, and natural history. *J Natl Cancer Inst Monogr* **2010,** *2010* (41), 134-8.

51. Krammer, J.; Pinker-Domenig, K.; Robson, M. E.; Gonen, M.; Bernard-Davila, B.; Morris, E. A.; Mangino, D. A.; Jochelson, M. S., Breast cancer detection and tumor characteristics in BRCA1 and BRCA2 mutation carriers. *Breast Cancer Res Treat* **2017,** *163* (3), 565-571.

52. Paoli, P.; Giannoni, E.; Chiarugi, P., Anoikis molecular pathways and its role in cancer progression. *Biochim Biophys Acta* **2013,** *1833* (12), 3481-3498.

53. Brabletz, T.; Kalluri, R.; Nieto, M. A.; Weinberg, R. A., EMT in cancer. *Nat Rev Cancer* **2018,** *18* (2), 128-134.

54. Wu, Y.; Zhou, B. P., New insights of epithelial-mesenchymal transition in cancer metastasis. *Acta Biochim Biophys Sin (Shanghai)* **2008,** *40* (7), 643-50.

55. Elmore, J. G.; Barton, M. B.; Moceri, V. M.; Polk, S.; Arena, P. J.; Fletcher, S. W., Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* **1998,** *338* (16), 1089-96.

56. Saslow, D.; Boetes, C.; Burke, W.; Harms, S.; Leach, M. O.; Lehman, C. D.; Morris, E.; Pisano, E.; Schnall, M.; Sener, S.; Smith, R. A.; Warner, E.; Yaffe, M.; Andrews, K. S.; Russell, C. A.; American Cancer Society Breast Cancer Advisory, G., American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* **2007,** *57* (2), 75-89.

57. Berg, W. A.; Bandos, A. I.; Mendelson, E. B.; Lehrer, D.; Jong, R. A.; Pisano, E. D., Ultrasound as the Primary Screening Test for Breast Cancer: Analysis From ACRIN 6666. *J Natl Cancer Inst* **2016,** *108* (4).

58. Oeffinger, K. C.; Fontham, E. T.; Etzioni, R.; Herzig, A.; Michaelson, J. S.; Shih, Y. C.; Walter, L. C.; Church, T. R.; Flowers, C. R.; LaMonte, S. J.; Wolf, A. M.; DeSantis, C.; Lortet-Tieulent, J.; Andrews, K.; Manassaram-Baptiste, D.; Saslow, D.; Smith, R. A.; Brawley, O. W.; Wender, R.; American Cancer, S., Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. *JAMA* **2015,** *314* (15), 1599-614.

59. Kelly, K. M.; Dean, J.; Comulada, W. S.; Lee, S. J., Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *Eur Radiol* **2010,** *20* (3), 734-42.

60. Hubbard, R. A.; Kerlikowske, K.; Flowers, C. I.; Yankaskas, B. C.; Zhu, W.; Miglioretti, D. L., Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. *Ann Intern Med* **2011,** *155* (8), 481-92.

61. Lehman, C. D.; Arao, R. F.; Sprague, B. L.; Lee, J. M.; Buist, D. S.; Kerlikowske, K.; Henderson, L. M.; Onega, T.; Tosteson, A. N.; Rauscher, G. H.; Miglioretti, D. L., National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology* **2017,** *283* (1), 49-58.

62. Gotzsche, P. C.; Jorgensen, K. J., Screening for breast cancer with mammography. *Cochrane Database Syst Rev* **2013,** (6), CD001877.

63. Marmot, M. G.; Altman, D. G.; Cameron, D. A.; Dewar, J. A.; Thompson, S. G.; Wilcox, M., The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* **2013,** *108* (11), 2205-40.

64. Nelson, H. D.; Tyne, K.; Naik, A.; Bougatsos, C.; Chan, B.; Nygren, P.; Humphrey, L., In *Screening for Breast Cancer: Systematic Evidence Review Update for the US Preventive Services Task Force*, Rockville (MD), 2009.

65. Duffy, S. W.; Tabar, L.; Olsen, A. H.; Vitak, B.; Allgood, P. C.; Chen, T. H.; Yen, A. M.; Smith, R. A., Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. *J Med Screen* **2010,** *17* (1), 25-30.

66. Puliti, D.; Zappa, M.; Miccinesi, G.; Falini, P.; Crocetti, E.; Paci, E., An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. *Eur J Cancer* **2009,** *45* (18), 3166-71.

67. Berrington de Gonzalez, A.; Berg, C. D.; Visvanathan, K.; Robson, M., Estimated risk of radiation-induced breast cancer from mammographic screening for young BRCA mutation carriers. *J Natl Cancer Inst* **2009,** *101* (3), 205-9.

68. Yaffe, M. J.; Mainprize, J. G., Risk of radiation-induced breast cancer from mammographic screening. *Radiology* **2011,** *258* (1), 98-105.

69. Haas, J. S.; Hill, D. A.; Wellman, R. D.; Hubbard, R. A.; Lee, C. I.; Wernli, K. J.; Stout, N. K.; Tosteson, A. N.; Henderson, L. M.; Alford-Teaster, J. A.; Onega, T. L., Disparities in the use of screening magnetic resonance imaging of the breast in community practice by race, ethnicity, and socioeconomic status. *Cancer* **2016,** *122* (4), 611-7.

70. James, W., aptamers_analytical_chemistry.pdf. *Encyclopedia of Analytical Chemistry* **2000**, 4848–4871.

71. Djordjevic, M., SELEX experiments: new prospects, applications and data analysis in inferring regulatory pathways. *Biomol Eng* **2007,** *24* (2), 179-89.

72. James, W., Nucleic acid and polypeptide aptamers: a powerful approach to ligand discovery. *Curr Opin Pharmacol* **2001,** *1* (5), 540-6.

73. Moore, M. D.; Cookson, J.; Coventry, V. K.; Sproat, B.; Rabe, L.; Cranston, R. D.; McGowan, I.; James, W., Protection of HIV neutralizing aptamers against rectal and vaginal nucleases: implications for RNA-based therapeutics. *J Biol Chem* **2011,** *286* (4), 2526-35.

74. Moore, M. D.; Bunka, D. H.; Forzan, M.; Spear, P. G.; Stockley, P. G.; McGowan, I.; James, W., Generation of neutralizing aptamers against herpes simplex virus type 2: potential components of multivalent microbicides. *J Gen Virol* **2011,** *92* (Pt 7), 1493-9.

75. Khati, M.; Schuman, M.; Ibrahim, J.; Sattentau, Q.; Gordon, S.; James, W., Neutralization of infectivity of diverse R5 clinical isolates of human immunodeficiency virus type 1 by gp120-binding 2'F-RNA aptamers. *J Virol* **2003,** *77* (23), 12692-8.

76. James, W., Aptamers in the virologists' toolkit. *J Gen Virol* **2007,** *88* (Pt 2), 351-64.

77. Gu, G.; Wang, T.; Yang, Y.; Xu, X.; Wang, J., An improved SELEX-Seq strategy for characterizing DNA-binding specificity of transcription factor: NF-kappaB as an example. *PLoS One* **2013,** *8* (10), e76109.

78. Gold, L., Oligonucleotides as research, diagnostic, and therapeutic agents. *J Biol Chem* **1995,** *270* (23), 13581-4.

79. Schutze, T.; Wilhelm, B.; Greiner, N.; Braun, H.; Peter, F.; Morl, M.; Erdmann, V. A.; Lehrach, H.; Konthur, Z.; Menger, M.; Arndt, P. F.; Glokler, J., Probing the SELEX process with next-generation sequencing. *PLoS One* **2011,** *6* (12), e29604.

80. Mercier, M. C.; Dontenwill, M.; Choulier, L., Selection of Nucleic Acid Aptamers Targeting Tumor Cell-Surface Protein Biomarkers. *Cancers (Basel)* **2017,** *9* (6).

81. Wang, H.; Zhang, Y.; Yang, H.; Qin, M.; Ding, X.; Liu, R.; Jiang, Y., In Vivo SELEX of an Inhibitory NSCLC-Specific RNA Aptamer from PEGylated RNA Library. *Molecular Therapy - Nucleic Acids* **2018,** *10*, 187-198.

82. Rong, Y.; Chen, H.; Zhou, X. F.; Yin, C. Q.; Wang, B. C.; Peng, C. W.; Liu, S. P.; Wang, F. B., Identification of an aptamer through whole cell-SELEX for targeting high metastatic liver cancers. *Oncotarget* **2016,** *7* (7), 8282-94.

83. Shi, H.; Cui, W.; He, X.; Guo, Q.; Wang, K.; Ye, X.; Tang, J., Whole cell-SELEX aptamers for highly specific fluorescence molecular imaging of carcinomas in vivo. *PLoS One* **2013,** *8* (8), e70476.

84. Zhu, G.; Zhang, H.; Jacobson, O.; Wang, Z.; Chen, H.; Yang, X.; Niu, G.; Chen, X., Combinatorial Screening of DNA Aptamers for Molecular Imaging of HER2 in Cancer. *Bioconjug Chem* **2017,** *28* (4), 1068-1075.

85. Cheng, C.; Chen, Y. H.; Lennox, K. A.; Behlke, M. A.; Davidson, B. L., In vivo SELEX for Identification of Brain-penetrating Aptamers. *Mol Ther Nucleic Acids* **2013,** *2*, e67.

86. Revenfeld, A. L.; Baek, R.; Nielsen, M. H.; Stensballe, A.; Varming, K.; Jorgensen, M., Diagnostic and prognostic potential of extracellular vesicles in peripheral blood. *Clin Ther* **2014,** *36* (6), 830-46.

87. Skog, J.; Wurdinger, T.; van Rijn, S.; Meijer, D. H.; Gainche, L.; Sena-Esteves, M.; Curry, W. T., Jr.; Carter, B. S.; Krichevsky, A. M.; Breakefield, X. O., Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol* **2008,** *10* (12), 1470-6.

88. Vlassov, A. V.; Magdaleno, S.; Setterquist, R.; Conrad, R., Exosomes: current knowledge of their composition, biological functions, and diagnostic and therapeutic potentials. *Biochim Biophys Acta* **2012,** *1820* (7), 940-8.

89. Pant, S.; Hilton, H.; Burczynski, M. E., The multifaceted exosome: biogenesis, role in normal and aberrant cellular function, and frontiers for pharmacological and biomarker opportunities. *Biochem Pharmacol* **2012,** *83* (11), 1484-94.

90. Szajnik, M.; Derbis, M.; Lach, M.; Patalas, P.; Michalak, M.; Drzewiecka, H.; Szpurek, D.; Nowakowski, A.; Spaczynski, M.; Baranowski, W.; Whiteside, T. L., Exosomes in Plasma of Patients with Ovarian Carcinoma: Potential Biomarkers of Tumor Progression and Response to Therapy. *Gynecol Obstet (Sunnyvale)* **2013,** *Suppl 4*, 3.

91. Baran, J.; Baj-Krzyworzeka, M.; Weglarczyk, K.; Szatanek, R.; Zembala, M.; Barbasz, J.; Czupryna, A.; Szczepanik, A.; Zembala, M., Circulating tumour-derived microvesicles in plasma of gastric cancer patients. *Cancer Immunol Immunother* **2010,** *59* (6), 841-50.

92. Kim, H. K.; Song, K. S.; Park, Y. S.; Kang, Y. H.; Lee, Y. J.; Lee, K. R.; Kim, H. K.; Ryu, K. W.; Bae, J. M.; Kim, S., Elevated levels of circulating platelet microparticles, VEGF, IL-6 and RANTES in patients with gastric cancer: possible role of a metastasis predictor. *Eur J Cancer* **2003,** *39* (2), 184-91.

93. Li, J.; Sherman-Baust, C. A.; Tsai-Turton, M.; Bristow, R. E.; Roden, R. B.; Morin, P. J., Claudin-containing exosomes in the peripheral circulation of women with ovarian cancer. *BMC Cancer* **2009,** *9*, 244.

94. Thery, C.; Amigorena, S.; Raposo, G.; Clayton, A., Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr Protoc Cell Biol* **2006,** *Chapter 3*, Unit 3 22.

95. ANTES, T. J.; KWEI, K., Methods for microvesicle isolation and selective removal. Google Patents: 2013.

96. Kabanov, A. V.; Lemieux, P.; Vinogradov, S.; Alakhov, V., Pluronic block copolymers: novel functional molecules for gene therapy. *Adv Drug Deliv Rev* **2002,** *54* (2), 223-33.

97. Shelat, P. B.; Plant, L. D.; Wang, J. C.; Lee, E.; Marks, J. D., The membrane-active tri-block copolymer pluronic F-68 profoundly rescues rat hippocampal neurons from oxygen-glucose deprivation-induced death through early inhibition of apoptosis. *J Neurosci* **2013,** *33* (30), 12287-99.

98. Firestone, M. A.; Wolf, A. C.; Seifert, S., Small-angle X-ray scattering study of the interaction of poly(ethylene oxide)-b-poly(propylene oxide)-b-poly(ethylene oxide) triblock copolymers with lipid bilayers. *Biomacromolecules* **2003,** *4* (6), 1539-49.

99. Lee, R. C.; River, L. P.; Pan, F. S.; Ji, L.; Wollmann, R. L., Surfactant-induced sealing of electropermeabilized skeletal muscle membranes in vivo. *Proc Natl Acad Sci U S A* **1992,** *89* (10), 4524-8.

100. Marks, J. D.; Pan, C. Y.; Bushell, T.; Cromie, W.; Lee, R. C., Amphiphilic, tri-block copolymers provide potent membrane-targeted neuroprotection. *FASEB J* **2001,** *15* (6), 1107-9.

101. Melnichenko, A. A.; Aksenov, D. V.; Myasoedova, V. A.; Panasenko, O. M.; Yaroslavov, A. A.; Sobenin, I. A.; Bobryshev, Y. V.; Orekhov, A. N., Pluronic block copolymers inhibit low density lipoprotein self-association. *Lipids* **2012,** *47* (10), 995-1000.

102. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **2012,** *9* (7), 671-5.

103. Trapnell, C.; Williams, B. A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M. J.; Salzberg, S. L.; Wold, B. J.; Pachter, L., Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **2010,** *28* (5), 511-5.

104. Rekker, K.; Saare, M.; Roost, A. M.; Salumets, A.; Peters, M., Circulating microRNA Profile throughout the menstrual cycle. *PLoS One* **2013,** *8* (11), e81166.

105. Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.;

Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P., A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **2012,** *30* (10), 918-20.

106. Kim, S.; Pevzner, P. A., MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* **2014,** *5*, 5277.

107. Krey, J. F.; Wilmarth, P. A.; Shin, J. B.; Klimek, J.; Sherman, N. E.; Jeffery, E. D.; Choi, D.; David, L. L.; Barr-Gillespie, P. G., Accurate label-free protein quantitation with high- and low-resolution mass spectrometers. *J Proteome Res* **2014,** *13* (2), 1034-44.

108. Sardiu, M. E.; Washburn, M. P., Enriching quantitative proteomics with SI(N). *Nat Biotechnol* **2010,** *28* (1), 40-2.

109. Ohshima, K.; Inoue, K.; Fujiwara, A.; Hatakeyama, K.; Kanto, K.; Watanabe, Y.; Muramatsu, K.; Fukuda, Y.; Ogura, S.; Yamaguchi, K.; Mochizuki, T., Let-7 microRNA family is selectively secreted into the extracellular environment via exosomes in a metastatic gastric cancer cell line. *PLoS One* **2010,** *5* (10), e13247.

110. Hannafon, B. N.; Carpenter, K. J.; Berry, W. L.; Janknecht, R.; Dooley, W. C.; Ding, W. Q., Exosome-mediated microRNA signaling from breast cancer cells is altered by the anti-angiogenesis agent docosahexaenoic acid (DHA). *Mol Cancer* **2015,** *14*, 133.

111. Zhang, J.; Li, S.; Li, L.; Li, M.; Guo, C.; Yao, J.; Mi, S., Exosome and exosomal microRNA: trafficking, sorting, and function. *Genomics Proteomics Bioinformatics* **2015,** *13* (1), 17-24.

112. Hortin, G. L.; Sviridov, D.; Anderson, N. L., High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. *Clin Chem* **2008,** *54* (10), 1608-16.

113. Badimon, L.; Suades, R.; Fuentes, E.; Palomo, I.; Padro, T., Role of Platelet-Derived Microvesicles As Crosstalk Mediators in Atherothrombosis and Future Pharmacology Targets: A Link between Inflammation, Atherosclerosis, and Thrombosis. *Front Pharmacol* **2016,** *7*, 293.

114. Ge, R.; Tan, E.; Sharghi-Namini, S.; Asada, H. H., Exosomes in Cancer Microenvironment and Beyond: have we Overlooked these Extracellular Messengers? *Cancer Microenviron* **2012,** *5* (3), 323-32.

115. Buschow, S. I.; van Balkom, B. W.; Aalberts, M.; Heck, A. J.; Wauben, M.; Stoorvogel, W., MHC class II-associated proteins in B-cell exosomes and potential functional implications for exosome biogenesis. *Immunol Cell Biol* **2010,** *88* (8), 851-6.

116. van Balkom, B. W.; Eisele, A. S.; Pegtel, D. M.; Bervoets, S.; Verhaar, M. C., Quantitative and qualitative analysis of small RNAs in human endothelial cells and exosomes provides insights into localized RNA processing, degradation and sorting. *J Extracell Vesicles* **2015,** *4*, 26760.

117. Hewson, C.; Capraro, D.; Burdach, J.; Whitaker, N.; Morris, K. V., Extracellular vesicle associated long non-coding RNAs functionally enhance cell viability. *Noncoding RNA Res* **2016,** *1* (1), 3-11.

118. Khurana, R.; Ranches, G.; Schafferer, S.; Lukasser, M.; Rudnicki, M.; Mayer, G.; Huttenhofer, A., Identification of urinary exosomal noncoding RNAs as novel biomarkers in chronic kidney disease. *RNA* **2017,** *23* (2), 142-152.

119. Shore, S.; Henderson, J. M.; Lebedev, A.; Salcedo, M. P.; Zon, G.; McCaffrey, A. P.; Paul, N.; Hogrefe, R. I., Small RNA Library Preparation Method for Next-Generation Sequencing Using Chemical Modifications to Prevent Adapter Dimer Formation. *PLoS One* **2016,** *11* (11), e0167009.

120. Venkatesh, T.; Suresh, P. S.; Tsutsumi, R., tRFs: miRNAs in disguise. *Gene* **2016,** *579* (2), 133-8.

121. Hizir, Z.; Bottini, S.; Grandjean, V.; Trabucchi, M.; Repetto, E., RNY (YRNA)-derived small RNAs regulate cell death and inflammation in monocytes/macrophages. *Cell Death Dis* **2017,** *8* (1), e2530.

122. Holliday, D. L.; Speirs, V., Choosing the right cell line for breast cancer research. *Breast Cancer Res* **2011,** *13* (4), 215.

123. Ung, T. H.; Madsen, H. J.; Hellwinkel, J. E.; Lencioni, A. M.; Graner, M. W., Exosome proteomics reveals transcriptional regulator proteins with potential to mediate downstream pathways. *Cancer Sci* **2014,** *105* (11), 1384-92.

124. Liang, H.; Gong, F.; Zhang, S.; Zhang, C. Y.; Zen, K.; Chen, X., The origin, function, and diagnostic potential of extracellular microRNAs in human body fluids. *Wiley Interdiscip Rev RNA* **2014,** *5* (2), 285-300.

125. Huang, X.; Yuan, T.; Tschannen, M.; Sun, Z.; Jacob, H.; Du, M.; Liang, M.; Dittmar, R. L.; Liu, Y.; Liang, M.; Kohli, M.; Thibodeau, S. N.; Boardman, L.; Wang, L., Characterization of human plasma-derived exosomal RNAs by deep sequencing. *BMC Genomics* **2013,** *14*, 319.

126. Soule, H. D.; Vazguez, J.; Long, A.; Albert, S.; Brennan, M., A human cell line from a pleural effusion derived from a breast carcinoma. *J Natl Cancer Inst* **1973,** *51* (5), 1409-16.

127. Brinkley, B. R.; Beall, P. T.; Wible, L. J.; Mace, M. L.; Turner, D. S.; Cailleau, R. M., Variations in cell form and cytoskeleton in human breast carcinoma cells in vitro. *Cancer Res* **1980,** *40* (9), 3118-29.

128. Cailleau, R.; Olive, M.; Cruciger, Q. V., Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. *In Vitro* **1978,** *14* (11), 911-5.

129. Keydar, I.; Chen, L.; Karby, S.; Weiss, F. R.; Delarea, J.; Radu, M.; Chaitcik, S.; Brenner, H. J., Establishment and characterization of a cell line of human breast carcinoma origin. *Eur J Cancer* **1979,** *15* (5), 659-70.

130. Fantozzi, A.; Christofori, G., Mouse models of breast cancer metastasis. *Breast Cancer Res* **2006,** *8* (4), 212.

131. Uhlen, M.; Fagerberg, L.; Hallstrom, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigyarto, C. A.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P. H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; von Heijne, G.; Nielsen, J.; Ponten, F., Proteomics. Tissue-based map of the human proteome. *Science* **2015,** *347* (6220), 1260419.

132. Uhlen, M.; Oksvold, P.; Fagerberg, L.; Lundberg, E.; Jonasson, K.; Forsberg, M.; Zwahlen, M.; Kampf, C.; Wester, K.; Hober, S.; Wernerus, H.; Bjorling, L.; Ponten, F., Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* **2010,** *28* (12), 1248-50.

133. Uhlen, M.; Bjorling, E.; Agaton, C.; Szigyarto, C. A.; Amini, B.; Andersen, E.; Andersson, A. C.; Angelidou, P.; Asplund, A.; Asplund, C.; Berglund, L.; Bergstrom, K.; Brumer, H.; Cerjan, D.; Ekstrom, M.; Elobeid, A.; Eriksson, C.; Fagerberg, L.; Falk, R.; Fall, J.; Forsberg, M.; Bjorklund, M. G.; Gumbel, K.; Halimi, A.; Hallin, I.; Hamsten, C.; Hansson, M.; Hedhammar, M.; Hercules, G.; Kampf, C.; Larsson, K.; Lindskog, M.; Lodewyckx, W.; Lund, J.; Lundeberg, J.; Magnusson, K.; Malm, E.; Nilsson, P.; Odling, J.; Oksvold, P.; Olsson, I.; Oster, E.; Ottosson, J.; Paavilainen, L.; Persson, A.; Rimini, R.; Rockberg, J.; Runeson, M.; Sivertsson, A.; Skollermo, A.; Steen, J.; Stenvall, M.; Sterky, F.; Stromberg, S.; Sundberg, M.; Tegel, H.; Tourle, S.; Wahlund, E.; Walden, A.; Wan, J.; Wernerus, H.; Westberg, J.; Wester, K.; Wrethagen, U.; Xu, L. L.; Hober, S.; Ponten, F., A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* **2005,** *4* (12), 1920-32.

134. Tomczak, K.; Czerwinska, P.; Wiznerowicz, M., The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* **2015,** *19* (1A), A68-77.

135. Chen, N. P.; Uddin, B.; Voit, R.; Schiebel, E., Human phosphatase CDC14A is recruited to the cell leading edge to regulate cell migration and adhesion. *Proc Natl Acad Sci U S A* **2016,** *113* (4), 990-5.

136. Sohn, E. J.; Jung, D. B.; Lee, H.; Han, I.; Lee, J.; Lee, H.; Kim, S. H., CNOT2 promotes proliferation and angiogenesis via VEGF signaling in MDA-MB-231 breast cancer cells. *Cancer Lett* **2017,** *412*, 88-98.

137. Li, W.; Liu, C.; Zhao, C.; Zhai, L.; Lv, S., Downregulation of beta3 integrin by miR-30a-5p modulates cell adhesion and invasion by interrupting Erk/Ets1 network in triple-negative breast cancer. *Int J Oncol* **2016,** *48* (3), 1155-64.

138. Carter, R. Z.; Micocci, K. C.; Natoli, A.; Redvers, R. P.; Paquet-Fifield, S.; Martin, A. C.; Denoyer, D.; Ling, X.; Kim, S. H.; Tomasin, R.; Selistre-de-Araujo, H.; Anderson, R. L.; Pouliot, N., Tumour but not stromal expression of beta3 integrin is essential, and is required early, for spontaneous dissemination of bone-metastatic breast cancer. *J Pathol* **2015,** *235* (5), 760-72.

139. Liu, T.; Zhang, X. Y.; He, X. H.; Geng, J. S.; Liu, Y.; Kong, D. J.; Shi, Q. Y.; Liu, F.; Wei, W.; Pang, D., High levels of BCOX1 expression are associated with poor

prognosis in patients with invasive ductal carcinomas of the breast. *PLoS One* **2014,** *9* (1), e86952.

140. Oliveira, N. C.; Gomig, T. H.; Milioli, H. H.; Cordeiro, F.; Costa, G. G.; Urban, C. A.; Lima, R. S.; Cavalli, I. J.; Ribeiro, E. M., Comparative proteomic analysis of ductal and lobular invasive breast carcinoma. *Genet Mol Res* **2016,** *15* (2).

141. Li, Y.; Rong, G.; Kang, H., Taxotere-induced elevated expression of IL8 in carcinoma-associated fibroblasts of breast invasive ductal cancer. *Oncol Lett* **2017,** *13* (3), 1856-1860.

142. Liu, Y. L.; Zhang, H. M.; Pan, H. M.; Bao, Y. H.; Xue, J.; Wang, T. C.; Dong, X. C.; Li, X. L.; Bao, H. G., The relationship between apolipoprotein E gene epsilon2/epsilon3/epsilon4 polymorphism and breast cancer risk: a systematic review and meta-analysis. *Onco Targets Ther* **2016,** *9*, 1241-9.

143. Stevens, K. N.; Wang, X.; Fredericksen, Z.; Pankratz, V. S.; Cerhan, J.; Vachon, C. M.; Olson, J. E.; Couch, F. J., Evaluation of associations between common variation in mitotic regulatory pathways and risk of overall and high grade breast cancer. *Breast Cancer Res Treat* **2011,** *129* (2), 617-22.

144. Lucci, M. A.; Orlandi, R.; Triulzi, T.; Tagliabue, E.; Balsari, A.; Villa-Moruzzi, E., Expression profile of tyrosine phosphatases in HER2 breast cancer cells and tumors. *Cell Oncol* **2010,** *32* (5-6), 361-72.

145. Russo, J.; Russo, I. H., Molecular basis of pregnancy-induced breast cancer prevention. *Horm Mol Biol Clin Investig* **2012,** *9* (1), 3-10.

146. Chu, X.; Guo, X.; Jiang, Y.; Yu, H.; Liu, L.; Shan, W.; Yang, Z. Q., Genotranscriptomic meta-analysis of the CHD family chromatin remodelers in human cancers - initial evidence of an oncogenic role for CHD7. *Mol Oncol* **2017,** *11* (10), 1348-1360.

147. Faraji, F.; Hu, Y.; Wu, G.; Goldberger, N. E.; Walker, R. C.; Zhang, J.; Hunter, K. W., An integrated systems genetics screen reveals the transcriptional structure of inherited predisposition to metastatic disease. *Genome Res* **2014,** *24* (2), 227-40.

148. Gonzalez-Conchas, G. A.; Rodriguez-Romo, L.; Hernandez-Barajas, D.; Gonzalez-Guerrero, J. F.; Rodriguez-Fernandez, I. A.; Verdines-Perez, A.; Templeton, A. J.; Ocana, A.; Seruga, B.; Tannock, I. F.; Amir, E.; Vera-Badillo, F. E., Epidermal growth factor receptor overexpression and outcomes in early breast cancer: A systematic review and a meta-analysis. *Cancer Treat Rev* **2017,** *62*, 1-8.

149. Qian, W.; Tao, L.; Wang, Y.; Zhang, F.; Li, M.; Huang, S.; Wang, A.; Chen, W.; Yue, Z.; Chen, L.; Liu, Y.; Huang, C.; Zhang, L.; Li, Y.; Lu, Y., Downregulation of Integrins in Cancer Cells and Anti-Platelet Properties Are Involved in Holothurian Glycosaminoglycan-Mediated Disruption of the Interaction of Cancer Cells and Platelets in Hematogenous Metastasis. *J Vasc Res* **2015,** *52* (3), 197-209.

150. Chi, Y.; Huang, S.; Wang, L.; Zhou, R.; Wang, L.; Xiao, X.; Li, D.; Cai, Y.; Zhou, X.; Wu, J., CDK11p58 inhibits ERalpha-positive breast cancer invasion by

targeting integrin beta3 via the repression of ERalpha signaling. *BMC Cancer* **2014,** *14*, 577.

151. Gong, Y.; He, T.; Yang, L.; Yang, G.; Chen, Y.; Zhang, X., The role of miR-100 in regulating apoptosis of breast cancer cells. *Sci Rep* **2015,** *5*, 11650.

152. Marcotte, R.; Sayad, A.; Brown, K. R.; Sanchez-Garcia, F.; Reimand, J.; Haider, M.; Virtanen, C.; Bradner, J. E.; Bader, G. D.; Mills, G. B.; Pe'er, D.; Moffat, J.; Neel, B. G., Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. *Cell* **2016,** *164* (1-2), 293-309.

153. Abdelkarim, M.; Vintonenko, N.; Starzec, A.; Robles, A.; Aubert, J.; Martin, M. L.; Mourah, S.; Podgorniak, M. P.; Rodrigues-Ferreira, S.; Nahmias, C.; Couraud, P. O.; Doliger, C.; Sainte-Catherine, O.; Peyri, N.; Chen, L.; Mariau, J.; Etienne, M.; Perret, G. Y.; Crepin, M.; Poyet, J. L.; Khatib, A. M.; Di Benedetto, M., Invading basement membrane matrix is sufficient for MDA-MB-231 breast cancer cells to develop a stable in vivo metastatic phenotype. *PLoS One* **2011,** *6* (8), e23334.

154. Cine, N.; Baykal, A. T.; Sunnetci, D.; Canturk, Z.; Serhatli, M.; Savli, H., Identification of ApoA1, HPX and POTEE genes by omic analysis in breast cancer. *Oncol Rep* **2014,** *32* (3), 1078-86.

155. Qian, L.; Gao, X.; Huang, H.; Lu, S.; Cai, Y.; Hua, Y.; Liu, Y.; Zhang, J., PRSS3 is a prognostic marker in invasive ductal carcinoma of the breast. *Oncotarget* **2017,** *8* (13), 21444-21453.

156. Ma, R.; Ye, X.; Cheng, H.; Ma, Y.; Cui, H.; Chang, X., PRSS3 expression is associated with tumor progression and poor prognosis in epithelial ovarian cancer. *Gynecol Oncol* **2015,** *137* (3), 546-52.

157. Hockla, A.; Radisky, D. C.; Radisky, E. S., Mesotrypsin promotes malignant growth of breast cancer cells through shedding of CD109. *Breast Cancer Res Treat* **2010,** *124* (1), 27-38.

158. Yang, L.; Shi, T.; Liu, F.; Ren, C.; Wang, Z.; Li, Y.; Tu, X.; Yang, G.; Cheng, X., REV3L, a promising target in regulating the chemosensitivity of cervical cancer cells. *PLoS One* **2015,** *10* (3), e0120334.

159. Varadi, V.; Bevier, M.; Grzybowska, E.; Johansson, R.; Enquist, K.; Henriksson, R.; Butkiewicz, D.; Pamula-Pilat, J.; Tecza, K.; Hemminki, K.; Lenner, P.; Forsti, A., Genetic variation in genes encoding for polymerase zeta subunits associates with breast cancer risk, tumour characteristics and survival. *Breast Cancer Res Treat* **2011,** *129* (1), 235-45.

160. Roger, S.; Gillet, L.; Le Guennec, J. Y.; Besson, P., Voltage-gated sodium channels and cancer: is excitability their primary role? *Front Pharmacol* **2015,** *6*, 152.

161. Gau, D. M.; Lesnock, J. L.; Hood, B. L.; Bhargava, R.; Sun, M.; Darcy, K.; Luthra, S.; Chandran, U.; Conrads, T. P.; Edwards, R. P.; Kelley, J. L.; Krivak, T. C.; Roy, P., BRCA1 deficiency in ovarian cancer is associated with alteration in expression of several key regulators of cell motility - A proteomics study. *Cell Cycle* **2015,** *14* (12), 1884-92.

162. Gur-Dedeoglu, B.; Konu, O.; Kir, S.; Ozturk, A. R.; Bozkurt, B.; Ergul, G.; Yulug, I. G., A resampling-based meta-analysis for detection of differential gene expression in breast cancer. *BMC Cancer* **2008,** *8*, 396.

163. Song, J.; Yang, W.; Shih Ie, M.; Zhang, Z.; Bai, J., Identification of BCOX1, a novel gene overexpressed in breast cancer. *Biochim Biophys Acta* **2006,** *1760* (1), 62-9.

164. Schuetz, C. S.; Bonin, M.; Clare, S. E.; Nieselt, K.; Sotlar, K.; Walter, M.; Fehm, T.; Solomayer, E.; Riess, O.; Wallwiener, D.; Kurek, R.; Neubauer, H. J., Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. *Cancer Res* **2006,** *66* (10), 5278-86.

165. Lee, S.; Stewart, S.; Nagtegaal, I.; Luo, J.; Wu, Y.; Colditz, G.; Medina, D.; Allred, D. C., Differentially expressed genes regulating the progression of ductal carcinoma in situ to invasive breast cancer. *Cancer Res* **2012,** *72* (17), 4574-86.

166. Michael, M.; Begum, R.; Fong, K.; Pourreyrone, C.; South, A. P.; McGrath, J. A.; Parsons, M., BPAG1-e restricts keratinocyte migration through control of adhesion stability. *J Invest Dermatol* **2014,** *134* (3), 773-782.

167. Kuchnio, A.; Dewerchin, M.; Carmeliet, P., The PHD2 oxygen sensor paves the way to metastasis. *Oncotarget* **2015,** *6* (34), 35149-50.

168. Bordoli, M. R.; Stiehl, D. P.; Borsig, L.; Kristiansen, G.; Hausladen, S.; Schraml, P.; Wenger, R. H.; Camenisch, G., Prolyl-4-hydroxylase PHD2- and hypoxia-inducible factor 2-dependent regulation of amphiregulin contributes to breast tumorigenesis. *Oncogene* **2011,** *30* (5), 548-60.

169. Singh, M.; Devi, U.; Roy, S.; Gupta, P. S.; Saraf, S. A.; Kaithwas, G., Prolyl hydroxylase mediated inhibition of fatty acid synthase to combat tumor growth in mammary gland carcinoma. *Breast Cancer* **2016,** *23* (6), 820-829.

170. Paolicchi, E.; Crea, F.; Farrar, W. L.; Green, J. E.; Danesi, R., Histone lysine demethylases in breast cancer. *Crit Rev Oncol Hematol* **2013,** *86* (2), 97-103.

171. Hou, J.; Wu, J.; Dombkowski, A.; Zhang, K.; Holowatyj, A.; Boerner, J. L.; Yang, Z. Q., Genomic amplification and a role in drug-resistance for the KDM5A histone demethylase in breast cancer. *Am J Transl Res* **2012,** *4* (3), 247-56.

172. Gale, M.; Sayegh, J.; Cao, J.; Norcia, M.; Gareiss, P.; Hoyer, D.; Merkel, J. S.; Yan, Q., Screen-identified selective inhibitor of lysine demethylase 5A blocks cancer cell growth and drug resistance. *Oncotarget* **2016,** *7* (26), 39931-39944.

173. Newbold, R. F.; Mokbel, K., Evidence for a tumour suppressor function of SETD2 in human breast cancer: a new hypothesis. *Anticancer Res* **2010,** *30* (9), 3309-11.

174. Al Sarakbi, W.; Sasi, W.; Jiang, W. G.; Roberts, T.; Newbold, R. F.; Mokbel, K., The mRNA expression of SETD2 in human breast cancer: correlation with clinico-pathological parameters. *BMC Cancer* **2009,** *9*, 290.

175. Shen, H.; Li, L.; Zhou, S.; Yu, D.; Yang, S.; Chen, X.; Wang, D.; Zhong, S.; Zhao, J.; Tang, J., The role of ADAM17 in tumorigenesis and progression of breast cancer. *Tumour Biol* **2016**.

176. Narita, D.; Seclaman, E.; Ilina, R.; Cireap, N.; Ursoniu, S.; Anghel, A., ADAM12 and ADAM17 gene expression in laser-capture microdissected and non-microdissected breast tumors. *Pathol Oncol Res* **2011,** *17* (2), 375-85.

177. Zheng, X.; Jiang, F.; Katakowski, M.; Zhang, Z. G.; Lu, Q. E.; Chopp, M., ADAM17 promotes breast cancer cell malignant phenotype through EGFR-PI3K-AKT activation. *Cancer Biol Ther* **2009,** *8* (11), 1045-54.

178. Wang, C. Y.; Lai, M. D.; Phan, N. N.; Sun, Z.; Lin, Y. C., Meta-Analysis of Public Microarray Datasets Reveals Voltage-Gated Calcium Gene Signatures in Clinical Cancer Patients. *PLoS One* **2015,** *10* (7), e0125766.

179. Yamaga, R.; Ikeda, K.; Horie-Inoue, K.; Ouchi, Y.; Suzuki, Y.; Inoue, S., RNA sequencing of MCF-7 breast cancer cells identifies novel estrogen-responsive genes with functional estrogen receptor-binding sites in the vicinity of their transcription start sites. *Horm Cancer* **2013,** *4* (4), 222-32.

180. Bjorklund, S. S.; Panda, A.; Kumar, S.; Seiler, M.; Robinson, D.; Gheeya, J.; Yao, M.; Alnaes, G. I. G.; Toppmeyer, D.; Riis, M.; Naume, B.; Borresen-Dale, A. L.; Kristensen, V. N.; Ganesan, S.; Bhanot, G., Widespread alternative exon usage in clinically distinct subtypes of Invasive Ductal Carcinoma. *Sci Rep* **2017,** *7* (1), 5568.

181. Wang, H.; Wang, B.; Zhu, W.; Yang, Z., Lentivirus-Mediated Knockdown of Myosin VI Inhibits Cell Proliferation of Breast Cancer Cell. *Cancer Biother Radiopharm* **2015,** *30* (8), 330-5.

182. De Marchi, T.; Timmermans, A. M.; Smid, M.; Look, M. P.; Stingl, C.; Opdam, M.; Linn, S. C.; Sweep, F. C.; Span, P. N.; Kliffen, M.; van Deurzen, C. H.; Luider, T. M.; Foekens, J. A.; Martens, J. W.; Umar, A., Annexin-A1 and caldesmon are associated with resistance to tamoxifen in estrogen receptor positive recurrent breast cancer. *Oncotarget* **2016,** *7* (3), 3098-110.

183. Zou, J. X.; Duan, Z.; Wang, J.; Sokolov, A.; Xu, J.; Chen, C. Z.; Li, J. J.; Chen, H. W., Kinesin family deregulation coordinated by bromodomain protein ANCCA and histone methyltransferase MLL for breast cancer cell growth, survival, and tamoxifen resistance. *Mol Cancer Res* **2014,** *12* (4), 539-49.

184. Yuasa-Kawada, J.; Kinoshita-Kawada, M.; Rao, Y.; Wu, J. Y., Deubiquitinating enzyme USP33/VDU1 is required for Slit signaling in inhibiting breast cancer cell migration. *Proc Natl Acad Sci U S A* **2009,** *106* (34), 14530-5.

185. Suyama, M.; Nagase, T.; Ohara, O., HUGE: a database for human large proteins identified by Kazusa cDNA sequencing project. *Nucleic Acids Res* **1999,** *27* (1), 338-9.

186. Nagase, T.; Miyajima, N.; Tanaka, A.; Sazuka, T.; Seki, N.; Sato, S.; Tabata, S.; Ishikawa, K.; Kawarabayasi, Y.; Kotani, H.; et al., Prediction of the coding sequences of unidentified human genes. III. The coding sequences of 40 new genes (KIAA0081-

KIAA0120) deduced by analysis of cDNA clones from human cell line KG-1 (supplement). *DNA Res* **1995,** *2* (1), 51-9.

187. Buckhaults, P.; Zhang, Z.; Chen, Y. C.; Wang, T. L.; St Croix, B.; Saha, S.; Bardelli, A.; Morin, P. J.; Polyak, K.; Hruban, R. H.; Velculescu, V. E.; Shih Ie, M., Identifying tumor origin using a gene expression-based classification map. *Cancer Res* **2003,** *63* (14), 4144-9.

188. Endo, A.; Ly, T.; Pippa, R.; Bensaddek, D.; Nicolas, A.; Lamond, A. I., The Chromatin Assembly Factor Complex 1 (CAF1) and 5-Azacytidine (5-AzaC) Affect Cell Motility in Src-transformed Human Epithelial Cells. *J Biol Chem* **2017,** *292* (1), 172-184.

189. Slamon, D. J.; Clark, G. M.; Wong, S. G.; Levin, W. J.; Ullrich, A.; McGuire, W. L., Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **1987,** *235* (4785), 177-82.

190. Deblois, G.; Chahrour, G.; Perry, M. C.; Sylvain-Drolet, G.; Muller, W. J.; Giguere, V., Transcriptional control of the ERBB2 amplicon by ERRalpha and PGC-1beta promotes mammary gland tumorigenesis. *Cancer Res* **2010,** *70* (24), 10277-87.

191. Richardson, A. L.; Wang, Z. C.; De Nicolo, A.; Lu, X.; Brown, M.; Miron, A.; Liao, X.; Iglehart, J. D.; Livingston, D. M.; Ganesan, S., X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell* **2006,** *9* (2), 121-32.

192. Alimonti, A.; Carracedo, A.; Clohessy, J. G.; Trotman, L. C.; Nardella, C.; Egia, A.; Salmena, L.; Sampieri, K.; Haveman, W. J.; Brogi, E.; Richardson, A. L.; Zhang, J.; Pandolfi, P. P., Subtle variations in Pten dose determine cancer susceptibility. *Nat Genet* **2010,** *42* (5), 454-8.

193. Liu, J. C.; Voisin, V.; Bader, G. D.; Deng, T.; Pusztai, L.; Symmans, W. F.; Esteva, F. J.; Egan, S. E.; Zacksenhaus, E., Seventeen-gene signature from enriched Her2/Neu mammary tumor-initiating cells predicts clinical outcome for human HER2+:ERalpha- breast cancer. *Proc Natl Acad Sci U S A* **2012,** *109* (15), 5832-7.

194. Cui, H.; Lan, X.; Lu, S.; Zhang, F.; Zhang, W., Bioinformatic prediction and functional characterization of human KIAA0100 gene. *Journal of Pharmaceutical Analysis* **2017,** *7* (1), 10-18.

195. Hart, I. R., The selection and characterization of an invasive variant of the B16 melanoma. *Am J Pathol* **1979,** *97* (3), 587-600.

196. Boggs, A. E.; Vitolo, M. I.; Whipple, R. A.; Charpentier, M. S.; Goloubeva, O. G.; Ioffe, O. B.; Tuttle, K. C.; Slovic, J.; Lu, Y.; Mills, G. B.; Martin, S. S., alpha-Tubulin acetylation elevated in metastatic and basal-like breast cancer cells promotes microtentacle formation, adhesion, and invasive migration. *Cancer Res* **2015,** *75* (1), 203-15.

197. Kaverina, I.; Straube, A., Regulation of cell migration by dynamic microtubules. *Semin Cell Dev Biol* **2011,** *22* (9), 968-74.

198. Kadi, A.; Pichard, V.; Lehmann, M.; Briand, C.; Braguer, D.; Marvaldi, J.; Rognoni, J. B.; Luis, J., Effect of microtubule disruption on cell adhesion and spreading. *Biochem Biophys Res Commun* **1998,** *246* (3), 690-5.

199. Appert-Collin, A.; Hubert, P.; Cremel, G.; Bennasroune, A., Role of ErbB Receptors in Cancer Cell Migration and Invasion. *Front Pharmacol* **2015,** *6*, 283.

200. Gordon, L. A.; Mulligan, K. T.; Maxwell-Jones, H.; Adams, M.; Walker, R. A.; Jones, J. L., Breast cell invasive potential relates to the myoepithelial phenotype. *Int J Cancer* **2003,** *106* (1), 8-16.

201. Klein, P. A.; Xiang, J. H.; Kimura, A. K., Melanoma cells growing in aggregates on a non-adhesive poly(HEMA) substrate exhibit polykaryocytosis but do not develop an increased metastatic capability. *Clin Exp Metastasis* **1984,** *2* (4), 287-95.

202. Manuel Iglesias, J.; Beloqui, I.; Garcia-Garcia, F.; Leis, O.; Vazquez-Martin, A.; Eguiara, A.; Cufi, S.; Pavon, A.; Menendez, J. A.; Dopazo, J.; Martin, A. G., Mammosphere formation in breast carcinoma cell lines depends upon expression of E-cadherin. *PLoS One* **2013,** *8* (10), e77281.

203. Rayavarapu, R. R.; Heiden, B.; Pagani, N.; Shaw, M. M.; Shuff, S.; Zhang, S.; Schafer, Z. T., The role of multicellular aggregation in the survival of ErbB2-positive breast cancer cells during extracellular matrix detachment. *J Biol Chem* **2015,** *290* (14), 8722-33.

204. Li, G.; Satyamoorthy, K.; Herlyn, M., N-cadherin-mediated intercellular interactions promote survival and migration of melanoma cells. *Cancer Res* **2001,** *61* (9), 3819-25.

205. Tamilarasan, K. P.; Kolluru, G. K.; Rajaram, M.; Indhumathy, M.; Saranya, R.; Chatterjee, S., Thalidomide attenuates nitric oxide mediated angiogenesis by blocking migration of endothelial cells. *BMC Cell Biol* **2006,** *7*, 17.

206. Borghesani, P. R.; Peyrin, J. M.; Klein, R.; Rubin, J.; Carter, A. R.; Schwartz, P. M.; Luster, A.; Corfas, G.; Segal, R. A., BDNF stimulates migration of cerebellar granule cells. *Development* **2002,** *129* (6), 1435-42.

207. Mohan, K.; Ding, Z.; Hanly, J.; Issekutz, T. B., IFN-gamma-inducible T cell alpha chemoattractant is a potent stimulator of normal human blood T lymphocyte transendothelial migration: differential regulation by IFN-gamma and TNF-alpha. *J Immunol* **2002,** *168* (12), 6420-8.

208. Li, G.; Chen, Y. F.; Greene, G. L.; Oparil, S.; Thompson, J. A., Estrogen inhibits vascular smooth muscle cell-dependent adventitial fibroblast migration in vitro. *Circulation* **1999,** *100* (15), 1639-45.

209. Hsieh, C. Y.; Tsai, P. C.; Chu, C. L.; Chang, F. R.; Chang, L. S.; Wu, Y. C.; Lin, S. R., Brazilein suppresses migration and invasion of MDA-MB-231 breast cancer cells. *Chem Biol Interact* **2013,** *204* (2), 105-15.

210. Rizwan, A.; Cheng, M.; Bhujwalla, Z. M.; Krishnamachary, B.; Jiang, L.; Glunde, K., Breast cancer cell adhesome and degradome interact to drive metastasis. *NPJ Breast Cancer* **2015,** *1*, 15017.

211. Ishikawa, F.; Ushida, K.; Mori, K.; Shibanuma, M., Loss of anchorage primarily induces non-apoptotic cell death in a human mammary epithelial cell line under atypical focal adhesion kinase signaling. *Cell Death Dis* **2015,** *6*, e1619.

212. Akekawatchai, C.; Roytrakul, S.; Kittisenachai, S.; Isarankura-Na-Ayudhya, P.; Jitrapakdee, S., Protein Profiles Associated with Anoikis Resistance of Metastatic MDA-MB-231 Breast Cancer Cells. *Asian Pac J Cancer Prev* **2016,** *17* (2), 581-90.

213. Wlodkowic, D.; Telford, W.; Skommer, J.; Darzynkiewicz, Z., Apoptosis and beyond: cytometry in studies of programmed cell death. *Methods Cell Biol* **2011,** *103*, 55-98.

214. Charulatha, V.; Rajaram, A., Dimethyl 3,3'-dithiobispropionimidate: a novel crosslinking reagent for collagen. *J Biomed Mater Res* **2001,** *54* (1), 122-8.

215. Yang, H.; Ganguly, A.; Cabral, F., Inhibition of cell migration and cell division correlates with distinct effects of microtubule inhibiting drugs. *J Biol Chem* **2010,** *285* (42), 32242-50.

216. Beckmann, M. W.; Niederacher, D.; Schnurch, H. G.; Gusterson, B. A.; Bender, H. G., Multistep carcinogenesis of breast cancer and tumour heterogeneity. *J Mol Med (Berl)* **1997,** *75* (6), 429-39.

217. Portt, L.; Norman, G.; Clapp, C.; Greenwood, M.; Greenwood, M. T., Anti-apoptosis and cell survival: a review. *Biochim Biophys Acta* **2011,** *1813* (1), 238-59.

218. Daugaard, M.; Jaattela, M.; Rohde, M., Hsp70-2 is required for tumor cell growth and survival. *Cell Cycle* **2005,** *4* (7), 877-80.

219. Takayama, S.; Reed, J. C.; Homma, S., Heat-shock proteins as regulators of apoptosis. *Oncogene* **2003,** *22* (56), 9041-7.

220. Thul, P. J.; Akesson, L.; Wiking, M.; Mahdessian, D.; Geladaki, A.; Ait Blal, H.; Alm, T.; Asplund, A.; Bjork, L.; Breckels, L. M.; Backstrom, A.; Danielsson, F.; Fagerberg, L.; Fall, J.; Gatto, L.; Gnann, C.; Hober, S.; Hjelmare, M.; Johansson, F.; Lee, S.; Lindskog, C.; Mulder, J.; Mulvey, C. M.; Nilsson, P.; Oksvold, P.; Rockberg, J.; Schutten, R.; Schwenk, J. M.; Sivertsson, A.; Sjostedt, E.; Skogs, M.; Stadler, C.; Sullivan, D. P.; Tegel, H.; Winsnes, C.; Zhang, C.; Zwahlen, M.; Mardinoglu, A.; Ponten, F.; von Feilitzen, K.; Lilley, K. S.; Uhlen, M.; Lundberg, E., A subcellular map of the human proteome. *Science* **2017,** *356* (6340).

221. Deschesnes, R. G.; Patenaude, A.; Rousseau, J. L.; Fortin, J. S.; Ricard, C.; Cote, M. F.; Huot, J.; R, C. G.; Petitclerc, E., Microtubule-destabilizing agents induce focal adhesion structure disorganization and anoikis in cancer cells. *J Pharmacol Exp Ther* **2007,** *320* (2), 853-64.

222. Srivastava, S.; Mishra, S.; Surolia, A.; Panda, D., C1, a highly potent novel curcumin derivative, binds to tubulin, disrupts microtubule network and induces apoptosis. *Biosci Rep* **2016,** *36* (2).

223. He, X.; Liu, Z.; He, Q.; Qin, J.; Liu, N.; Zhang, L.; Li, D.; Zhou, J.; Shui, W.; Liu, M., Identification of novel microtubule-binding proteins by taxol-mediated microtubule stabilization and mass spectrometry analysis. *Thorac Cancer* **2015,** *6* (5), 649-54.

224. Wang, H.; Liu, B.; Zhang, C.; Peng, G.; Liu, M.; Li, D.; Gu, F.; Chen, Q.; Dong, J. T.; Fu, L.; Zhou, J., Parkin regulates paclitaxel sensitivity in breast cancer via a microtubule-dependent mechanism. *J Pathol* **2009,** *218* (1), 76-85.

225. Sun, X.; Li, D.; Yang, Y.; Ren, Y.; Li, J.; Wang, Z.; Dong, B.; Liu, M.; Zhou, J., Microtubule-binding protein CLIP-170 is a mediator of paclitaxel sensitivity. *J Pathol* **2012,** *226* (4), 666-73.

226. Etienne-Manneville, S., Actin and microtubules in cell motility: which one is in control? *Traffic* **2004,** *5* (7), 470-7.

227. Radons, J., The human HSP70 family of chaperones: where do we stand? *Cell Stress Chaperones* **2016,** *21* (3), 379-404.

228. Lianos, G. D.; Alexiou, G. A.; Mangano, A.; Mangano, A.; Rausei, S.; Boni, L.; Dionigi, G.; Roukos, D. H., The role of heat shock proteins in cancer. *Cancer Lett* **2015,** *360* (2), 114-8.

229. Tutar, Y., Heat shock proteins, substrate specificity and modulation of function. *Protein Pept Lett* **2006,** *13* (7), 699-705.

230. Ruepp, A.; Waegele, B.; Lechner, M.; Brauner, B.; Dunger-Kaltenbach, I.; Fobo, G.; Frishman, G.; Montrone, C.; Mewes, H. W., CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic Acids Res* **2010,** *38* (Database issue), D497-501.

231. Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N. H.; Chavali, G.; Chen, C.; del-Toro, N.; Duesbury, M.; Dumousseau, M.; Galeota, E.; Hinz, U.; Iannuccelli, M.; Jagannathan, S.; Jimenez, R.; Khadake, J.; Lagreid, A.; Licata, L.; Lovering, R. C.; Meldal, B.; Melidoni, A. N.; Milagros, M.; Peluso, D.; Perfetto, L.; Porras, P.; Raghunath, A.; Ricard-Blum, S.; Roechert, B.; Stutz, A.; Tognolli, M.; van Roey, K.; Cesareni, G.; Hermjakob, H., The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* **2014,** *42* (Database issue), D358-63.

232. Szklarczyk, D.; Morris, J. H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N. T.; Roth, A.; Bork, P.; Jensen, L. J.; von Mering, C., The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* **2017,** *45* (D1), D362-D368.

233. McCallister, C.; Kdeiss, B.; Nikolaidis, N., Biochemical characterization of the interaction between HspA1A and phospholipids. *Cell Stress Chaperones* **2016,** *21* (1), 41-53.

234. McCallister, C.; Kdeiss, B.; Nikolaidis, N., HspA1A, a 70-kDa heat shock protein, differentially interacts with anionic lipids. *Biochem Biophys Res Commun* **2015,** *467* (4), 835-40.

235. McCallister, C.; Siracusa, M. C.; Shirazi, F.; Chalkia, D.; Nikolaidis, N., Functional diversification and specialization of cytosolic 70-kDa heat shock proteins. *Sci Rep* **2015,** *5*, 9363.

236. Hullin-Matsuda, F.; Taguchi, T.; Greimel, P.; Kobayashi, T., Lipid compartmentalization in the endosome system. *Semin Cell Dev Biol* **2014,** *31*, 48-56.

237. Kobayashi, T.; Beuchat, M. H.; Chevallier, J.; Makino, A.; Mayran, N.; Escola, J. M.; Lebrand, C.; Cosson, P.; Kobayashi, T.; Gruenberg, J., Separation and characterization of late endosomal membrane domains. *J Biol Chem* **2002,** *277* (35), 32157-64.

238. Scott, C. C.; Vacca, F.; Gruenberg, J., Endosome maturation, transport and functions. *Semin Cell Dev Biol* **2014,** *31*, 2-10.

239. de Jong, O. G.; Verhaar, M. C.; Chen, Y.; Vader, P.; Gremmels, H.; Posthuma, G.; Schiffelers, R. M.; Gucek, M.; van Balkom, B. W., Cellular stress conditions are reflected in the protein and RNA content of endothelial cell-derived exosomes. *J Extracell Vesicles* **2012,** *1*.

240. Stewart, D. J.; Chiritescu, G.; Dahrouge, S.; Banerjee, S.; Tomiak, E. M., Chemotherapy dose--response relationships in non-small cell lung cancer and implied resistance mechanisms. *Cancer Treat Rev* **2007,** *33* (2), 101-37.

241. Leu, J. I.; Pimkina, J.; Frank, A.; Murphy, M. E.; George, D. L., A small molecule inhibitor of inducible heat shock protein 70. *Mol Cell* **2009,** *36* (1), 15-27.

242. Yu, S.; Liu, C.; Su, K.; Wang, J.; Liu, Y.; Zhang, L.; Li, C.; Cong, Y.; Kimberly, R.; Grizzle, W. E.; Falkson, C.; Zhang, H. G., Tumor exosomes inhibit differentiation of bone marrow dendritic cells. *J Immunol* **2007,** *178* (11), 6867-75.

243. Martinez, M. C.; Andriantsitohaina, R., Microparticles in angiogenesis: therapeutic potential. *Circ Res* **2011,** *109* (1), 110-9.

244. Josson, S.; Gururajan, M.; Sung, S. Y.; Hu, P.; Shao, C.; Zhau, H. E.; Liu, C.; Lichterman, J.; Duan, P.; Li, Q.; Rogatko, A.; Posadas, E. M.; Haga, C. L.; Chung, L. W., Stromal fibroblast-derived miR-409 promotes epithelial-to-mesenchymal transition and prostate tumorigenesis. *Oncogene* **2015,** *34* (21), 2690-9.

245. Brett, S. I.; Kim, Y.; Biggs, C. N.; Chin, J. L.; Leong, H. S., Extracellular vesicles such as prostate cancer cell fragments as a fluid biopsy for prostate cancer. *Prostate Cancer Prostatic Dis* **2015,** *18* (3), 213-20.

246. Minciacchi, V. R.; Freeman, M. R.; Di Vizio, D., Extracellular vesicles in cancer: exosomes, microvesicles and the emerging role of large oncosomes. *Semin Cell Dev Biol* **2015,** *40*, 41-51.

247. Rak, J.; Guha, A., Extracellular vesicles--vehicles that spread cancer genes. *Bioessays* **2012,** *34* (6), 489-97.

248. Verma, M.; Lam, T. K.; Hebert, E.; Divi, R. L., Extracellular vesicles: potential applications in cancer diagnosis, prognosis, and epidemiology. *BMC Clin Pathol* **2015,** *15*, 6.

249. Brody, E. N.; Gold, L., Aptamers as therapeutic and diagnostic agents. *J Biotechnol* **2000,** *74* (1), 5-13.

250. Famulok, M.; Hartig, J. S.; Mayer, G., Functional aptamers and aptazymes in biotechnology, diagnostics, and therapy. *Chem Rev* **2007,** *107* (9), 3715-43.

251. Wilson, D. S.; Szostak, J. W., In vitro selection of functional nucleic acids. *Annu Rev Biochem* **1999,** *68*, 611-47.

252. Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J., Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **1992,** *355* (6360), 564-6.

253. Ellington, A. D.; Szostak, J. W., Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature* **1992,** *355* (6363), 850-2.

254. Chu, T.; Ebright, J.; Ellington, A. D., Using aptamers to identify and enter cells. *Curr Opin Mol Ther* **2007,** *9* (2), 137-44.

255. Mi, J.; Liu, Y.; Rabbani, Z. N.; Yang, Z.; Urban, J. H.; Sullenger, B. A.; Clary, B. M., In vivo selection of tumor-targeting RNA motifs. *Nat Chem Biol* **2010,** *6* (1), 22-4.

256. D'Orsi CJ, S. E., Mendelson EB, Morris EA, et al. , ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. *Reston, VA, American College of Radiology* **2013**.

257. Melton, L., Protein arrays: proteomics in multiplex. *Nature* **2004,** *429* (6987), 101-7.

258. Gershon, D., Mass spectrometry goes mainstream. *Nature* **2003,** *424* (6948), 581.

259. Talapatra, A.; Rouse, R.; Hardiman, G., Protein microarrays: challenges and promises. *Pharmacogenomics* **2002,** *3* (4), 527-36.

260. Hesselberth, J.; Robertson, M. P.; Jhaveri, S.; Ellington, A. D., In vitro selection of nucleic acids for diagnostic applications. *J Biotechnol* **2000,** *74* (1), 15-25.

261. Tuerk, C.; Gold, L., Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990,** *249* (4968), 505-10.

262. Kodadek, T., Protein microarrays: prospects and problems. *Chem Biol* **2001,** *8* (2), 105-15.

263. Gold, L.; Walker, J. J.; Wilcox, S. K.; Williams, S., Advances in human proteomics at high scale with the SOMAscan proteomics platform. *N Biotechnol* **2012,** *29* (5), 543-9.

264. Webber, J.; Stone, T. C.; Katilius, E.; Smith, B. C.; Gordon, B.; Mason, M. D.; Tabi, Z.; Brewis, I. A.; Clayton, A., Proteomics analysis of cancer exosomes using a

novel modified aptamer-based array (SOMAscan) platform. *Mol Cell Proteomics* **2014,** *13* (4), 1050-64.

APPENDIX A

DETAIL ANALYSIS OF EV BY FLOW CELL CYTOMETRY

Traditionally, FL analysis (single fluorescence label) involves gating a sample on the light scatter plot (scatter plot) followed by projection of the gated population to a histogram according to the fluorescence signal intensity. The detection of EV by FL has been very challenging, mainly because of the small sizes of EV particles, which is equal to or below the detection limit of the most of the machines currently available (>100nm for Apogee). At the same time, sizes of noise particles from reagents, such as PBS, antibody solution etc. also fall in the range under the machine's detection limit and mixed with the EV particles, resulted in an inability to isolate EV particles from the noise particles based on their light scatter patterns.

In order to solve such problem, we performed FL analysis with 4 different samples, including three positive controls: UC purified cell line EVs in PBS (Figure A1-A), UC purified cell line EV in PBS and precipitated by F68 (Figure A1-B), UC purified cell line EV in plasma and enriched by F68 (Figure A1-C) and one sample of the native plasma EVs enriched by F68 (Figure A1-D). Each sample was analyzed with five negative controls including PBS, isotype-antibody solution, tetraspanin antibody solution, no antibody labeled sample control and isotype antibody-PE stained control.

Figure A1-A shows a typical pure EV sample on the FL analysis as well as the other five negative controls. First the signal from large angle scatter (LALS, equivalent to side scatter in other equipment) and the small angle scatter (SALS, equivalent to forward scatter) are plotted on Figure A1-A row 1 (ignore the red/green lines and the yellow colors at this moment). The signal distributions from noise particles in PBS, isotype-antibody solution, tetraspanin antibody solution were basically the same as samples with EV particles, including unstained EVs, isotype-PE stained EVs and tet antibody-PE

labeled EVs with SALS ranged from ~1.5 to 4 and LALS ranged from 0 to 4 in a log10 scale. Consistent with difficulties of isolating the EV population by traditional light scatter plotting. In the PE labeled fluorescence event distribution histogram (Figure A1-A row 2), most of the control signals are very low compared to EV particles labeled with tet-PE antibody, or may even be undetectable on the graph (blue color: for all controls; green color: the EV sample stained with tet PE antibody, green histogram from EV sample stained with tet-PE antibody were applied in all figures from the negative controls samples in the back at the same scale to compare signal to noise level). This is mainly because the EV particle number is much more abundant than the noise particles. Therefore, lack of distinguishable pattern from light scatter plot prevents the analysis from isolating the desired EV particles from noise. This might be acceptable when the expected EVs particle amount is very large but create significant limitations for the sample with unknown or very low numbers of EVs, such as EVs enriched from plasma.

To exam whether there is any other alternative non-fluorescence light scatter pattern that potentially could be used in assisting the separation between noise and true EVs particles, we found that other than the regular LALS/SALS, the integrated version of LALS and SALS were also collected (LALS-area and SALS-area) by the apogee device during sample analysis. LALS-area against SALS-area plots were then drawn as shown in Figure A1-A row 3, while SALS-area signal showed range between 0 to 200 for all the samples and controls without differentiation, we found that most of the noise particles from PBS, isotype antibody-PE solution, tet antibody-PE solution as well as non-labeled EV sample concentrated in an area with very narrow range of LALS-area values and distributed horizontally (<100, horizontal green line). In contrast, in the tetraspanin

antibody labeled EV sample, most of the particles showed increased signal of LALS-area, correlated with corresponding SALS-area signal and showed saturation of signal at about 4000. Graphically it showed a concentrated bold line with an angle against horizontal noise population (Figure A1-A row 3, sample). The results showed that while traditional light scatter plot was not able to separate the EV particles from noise, regardless the antibody-bound status, once the EV particles were labeled with an antibody targeted to the surface marker, a distinguishable pattern was formed and separated this population from others on the LALS-area/SALS-area scatter plot. This enabled us to separate the antibody-bound EVs from other noise particles as well as none-labeled EVs. Though this still does not directly distinguish the un-labeled EV population from noise, it might be very helpful for initial screening the particles according to the light scatter signals before the fluorescence signal analysis.

Based on this finding, as shown in (Figure A1-A row 3), minimal (green line) and maximal (red line) values for SALS-area (Vertical green and red lines) and LALS-area (Vertical green and red lines) were set and applied to gate the samples as well as the negative controls. The selected particle populations from individual sample and controls were then projected back to the LALS-SALS light scatter plot in yellow color in Figure A1-A row 1. Such high density of yellow-marked population was shown only in the tet-antibody-PE labeled EV sample but no other corresponding negative controls, indicating the population gated by LALS-area/SALS-area does help to filter out the noise and select the correct Tet-antibody bound population, which regular LALS-SALS light scatter plot was not able to achieve.

In order to further refine the population and make it more specific for the Tet-antibody-labeled population, minimal and maximal SALS and LALS thresholds again were set manually according to range of the yellow area shown in Figure A1-A row 1 sample (minimal: green lines; maximal red lines) to filter the population with additional SALS and LALS ranges (referred to as two-step gating below).

After two-step gating, the selected population was then projected to the fluorescence events distribution histogram in Figure A1-A row 4 according to their fluorescence signal intensities. The event distribution histogram from EV samples labeled with Tet-antibody-PE was shown in green and appears in all other histograms from negative controls for comparison and the fluorescence event distribution histograms for individual negative controls were shown in blue in front of the green. While noise particle events were still visible in regular analysis histograms (Figure A1-A row 2), all noise fluorescence events after two-step gating were almost completely disappeared compared to the corresponding tet-antibody-PE labeled events (Figure A1-A row 4), indicating after the two-step gating, an EV population labeled with Tet-antibody-PE could be specifically selected.

We found out by LALS-area/SALS-area light scatter signal plotting could isolate the EV population labeled with surface tetraspanin marker antibody, through two-step gating by LALS-area/SALS-area and LALS/SALS sequentially, a clean EV fluorescence labeled population could be selected. We then examined the sample precipitated by F68 reagent, including cell line EVs in PBS precipitated by F68, plasma sample spiked with cell line EVs enriched by F68 and native plasma EVs enriched by F68 to see if a EV population bound by the tet-antibody could be isolated by the same analysis.

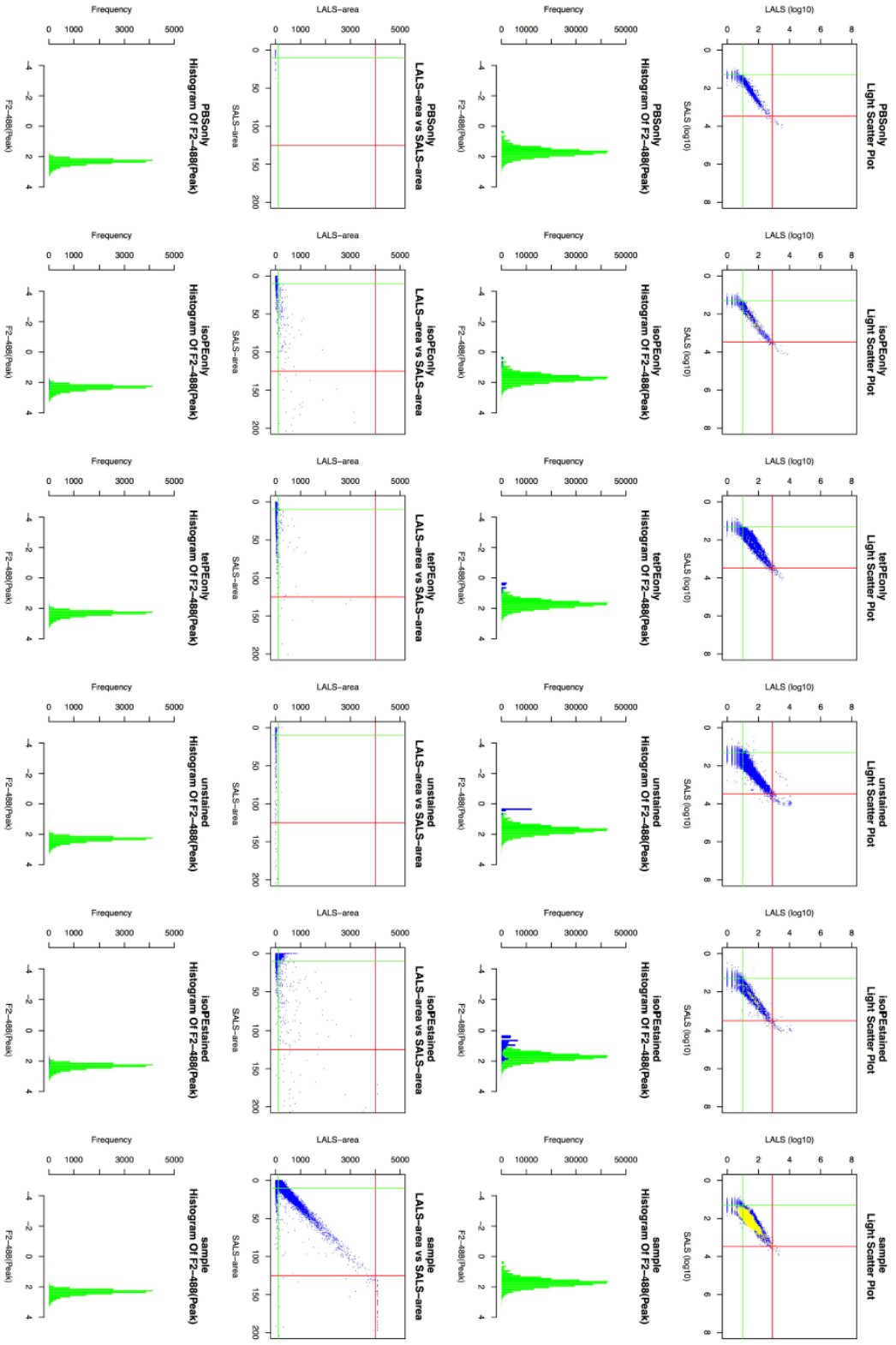
As shown in Figure A1-B, the same analysis layout was applied to the cell line EVs in PBS precipitated by F68. The result patterns are exactly the same as result from cell line EV control sample in Figure A1-A. This not only confirms the patterns observed from cell line EV control above, but also showed that the copolymer F68 was able to precipitate the cell line EVs without interfering its FL analysis.

Figure A1-C further examines the sample that is one more step closer to the real situation with spiked cell line EVs diluted in plasma sample followed by and F68 precipitation. Compared to the cell line EV control in Supplement Figure 1A and 1B, EVs enriched in the presence of plasma sample showed much wider LALS/SALS values in the regular light scatter plot (Figure A1-C row 1) and much higher noise in the fluorescence event distribution histogram, especially in the isotype antibody-PE labeled sample (Figure A1-C row 2). After two-step gating as shown above, a clean population without any noise contamination was isolated as shown in the fluorescence event distribution histogram for the gated population in Figure A1-C row 4. There are almost no noise events visible in blue, compared to the fluorescence event distribution histogram from Tet-antibody-PE labeled sample in green. The result further confirms that the EVs in plasma could be enriched by the F68 method; FL analysis based on the two-step gating could isolate such antibody-bound EV population.

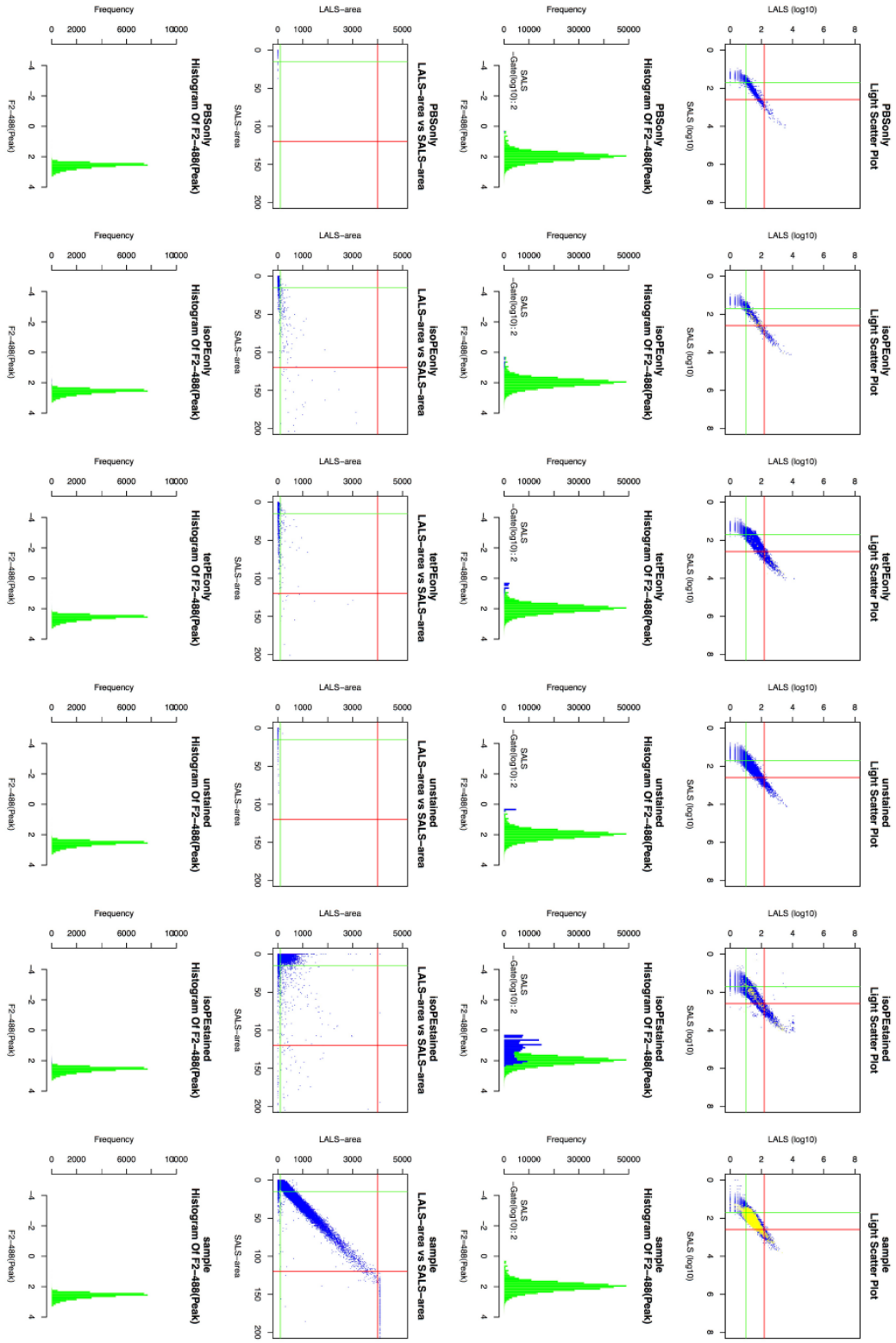
All three positive controls with spiked in cell line EVs show similar results, namely that EV particles bound by the specific antibody could be isolated by the two-step gating strategy compared to native plasma EV particles enriched by F68 were then analyzed to find out if similar population exists. As shown in Figure A1-D row 1 and 2, unlike the samples with cell line EV spiked, the native plasma EVs (regardless antibody-

bound status) was not able to separate itself either from the noise in tradition light scatter plot or in the fluorescence event distribution histogram. The histogram was particularly noisy especially from isotype antibody-PE labeled samples, which was largely overlapped between noise and the tet-antibody-PE labeled sample. In contrast, on the LALS-area/SALS-area plot, two distinguishable populations as described above were again detected (Figure A1-D row 3). Though the population was much smaller compared to the cell line EVs, the expected pattern of two different populations was clearly appeared. Consistently, after two-step gating analysis, the fluorescence event distribution for selected population was shown in Figure A1-D row 4. A tet-antibody bound EV population was clear isolated compared the noise background from isotype-antibody bound events. The result further confirms that an EV population with tetraspanin expression from plasma could be enriched by copolymer F68 precipitation.

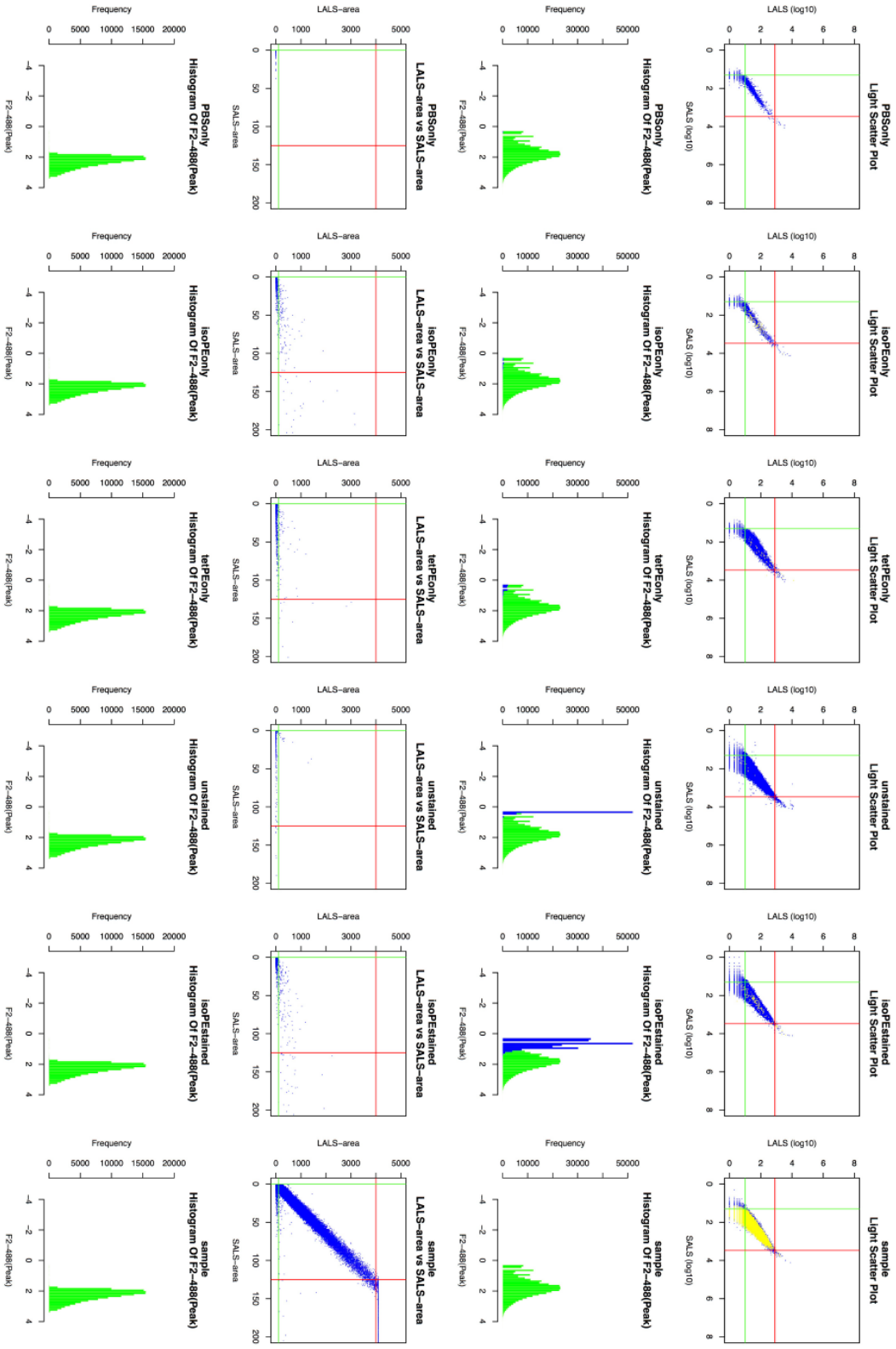
Appendix-A Figure 1A:



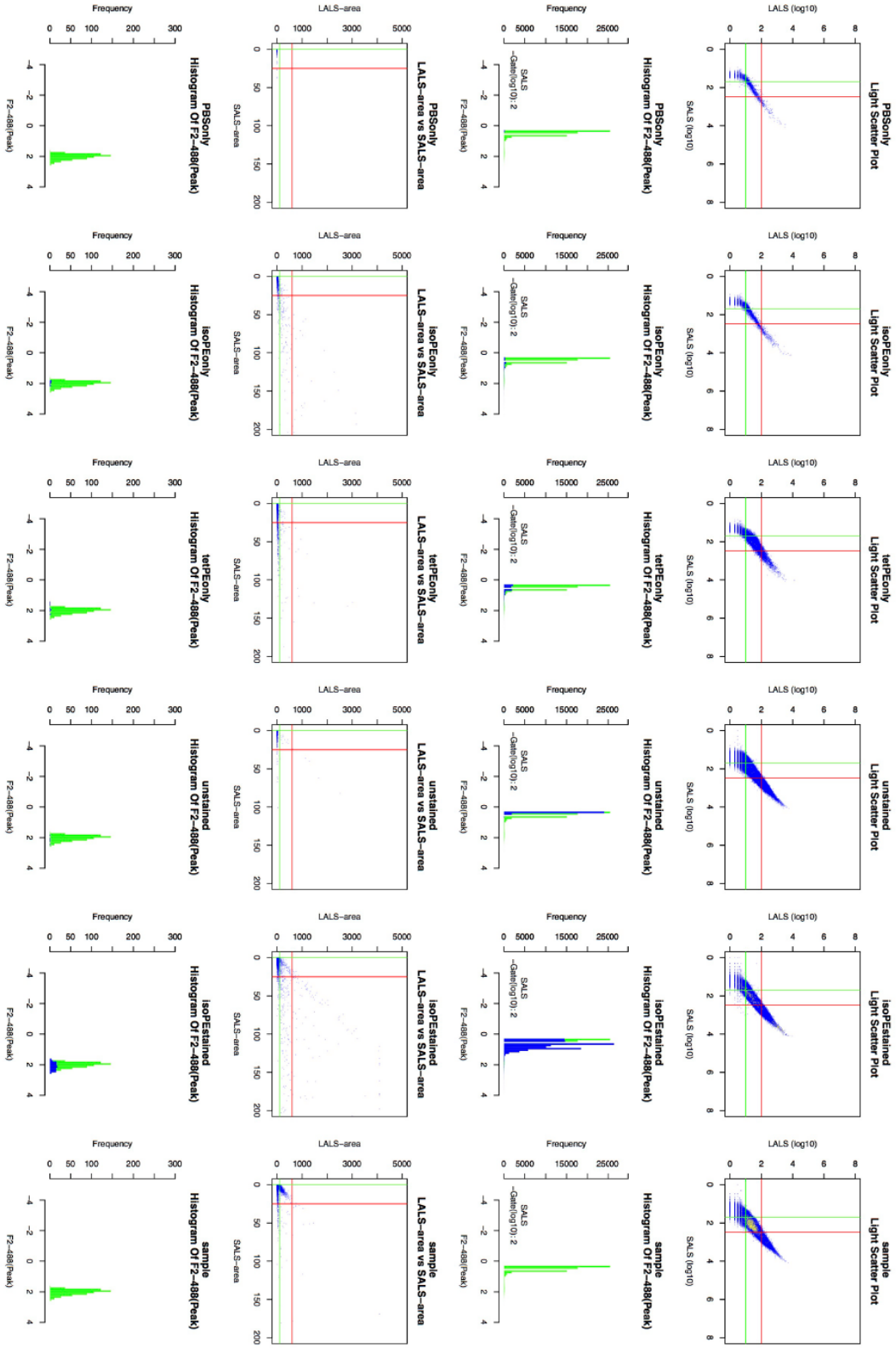
Appendix-A Figure 1B:



Appendix-A Figure 1C:



Appendix-A Figure 1D:



APPENDIX B

NGS SUMMARY FOR LIBRARY ENRICHMENT

Summary of the NGS sequencing reactions, quality matrix and etc. see respective excel worksheets.

APPENDIX C

DEMOGRAPHICS FOR PATIENTS USED IN SMALL SCALE STUDIES

List all the patient demographical information for all the small scale ADAPT studies. See the respective Excel sheet.

APPENDIX D

INFO FOR THE SELECTED 2000 APTAMER SEQUENCES FOR ADAPT

List all the ADAPT ssODNs from small scale study according to the listed criteria for final synthesis.

APPENDIX E

CANCER PATIENT DEMOGRAPHICS IN LARGE SCALE STUDY

Demographical information of cancer patient (biopsy-positive) in the large scale ADAPT study.

APPENDIX F

NON-CANCER PATIENT DEMOGRAPHICS IN LARGE SCALE STUDY

Demographical information of cancer patient (biopsy-negative) in the large scale ADAPT study.

APPENDIX G

DATA TABLE OF 500 LARGE SCALE ADAPT CLINICAL STUDY

Normalized count for the large scale ADAPT study for individual samples.