Druggable Cancer Secretome: Neoplasm-associated Traits

RAMASWAMY NARAYANAN

Department of Biological Sciences, Charles E. Schmidt College of Science, Florida Atlantic University, Boca Raton, FL, U.S.A.

Abstract. *Background: The genome association databases provide valuable clues to identify novel targets for cancer diagnosis and therapy. Genes harboring phenotype-associated polymorphisms for neoplasm traits can be identified using diverse bioinformatics tools. The recent availability of various protein expression datasets from normal human tissues, including the body fluids, enables for baseline expression profiling of the cancer secretome. Chemoinformatics approaches can help identify drug-like compounds from the protein 3D structures. Materials and Methods: The National Center for Biotechnology Information (NCBI) Phenome Genome Integrator (PheGenI) tool was enriched for neoplasm-associated traits. The neoplasm genes were characterized using diverse bioinformatics tools for pathways, gene ontology, genome-wide association, protein expression and functional class. Chemogenomics analysis was performed using the canSAR protein annotation tool. Results: The neoplasm-associated traits segregated into 1,305 genes harboring 2,837 single nucleotide polymorphisms (SNPs). Also identified were 65 open reading frames (ORFs) encompassing 137 SNPs. The neoplasm genes and the associated SNPs were classified into distinct tumor types. Protein expression in the secretome was seen for 913 of the neoplasm-associated genes, including 17 novel uncharacterized ORFs. Druggable proteins, including enzymes, transporters, channel proteins and receptors, were detected. Thirty-four novel druggable lead genes emerged from these studies, including seven cancer lead targets. Chemogenomics analysis using the canSAR protein annotation tool identified 168 active compounds (<1 μM) for the neoplasm genes in the body fluids. Among these, 7 most active lead compounds with drug-like properties (1-600 nM) were identified for the cancer lead targets, encompassing enzymes and receptors. Conclusion: Over seventy percent of the neoplasm trait-associated genes were detected in the body fluids, such as ascites, blood, tear, milk, semen, urine, etc. Ligand-based druggabililty analysis helped establish lead prioritization. The association of these proteins with diverse cancer types and other diseases provides a framework to develop novel diagnosis and therapy targets.*

Correspondence to: Dr. Ramaswamy Narayanan, Department of Biological Sciences, Charles E. Schmidt College of Science, Florida Atlantic University, Boca Raton, FL 33431, U.S.A. Tel: +1 5612972247, Fax: +1 5612973859, e-mail: rnarayan@fau.edu

Key Words: Biomarkers, body fluids, chemoinformatics, chemogenomics, clinical variations, druggable targets, expression quantitative loci, gene ontology, genome-wide association studies, leukemia and lymphoma, open reading frames, phenome-genome association, secretome, rule of five, tumors.

A vast amount of genome-wide association studies (GWAS)-based datasets is becoming available for mining the genome for disease association (1-10). Discovery of novel molecular targets for diverse diseases is greatly aided by the Phenome to Genome analysis tools. The National Center for Biotechnology Information (NCBI) Phenome Genome Integrator bioinformatics tool (PheGenI) offers an effective approach to decipher a gene's polymorphic association with a disease phenotype (11). The association evidence, together with the Expression Quantitative Trait Loci (eQTL) analysis (12), allows us to prioritize molecular targets for rational drug discovery.

The recent availability of numerous protein expression analysis tools has expanded our capability to monitor the protein levels in diverse normal and tumor tissues (13-19). Tools, such as the human protein reference database (HPRD), Multi-omics profiling expression database (MOPED) and Proteomics database (Proteomics DB) encompass protein expression datasets from a large number of body fluids. Discovery of molecular targets that can be detected in the body fluids (the cancer secretome) offers an advantage for biomarker lead prioritization efforts (20, 21). Changes in the target gene expression level can be readily monitored in response to cancer progression or therapy in a non-invasive manner.

Further, novel diagnostic and response-to-therapy indicator proteins may emerge from a database of neoplasm-associated genes that can be readily detected in the body fluids.

Major druggable class of proteins for small molecular weight compounds in the human genome includes cluster of differentiation (CD) markers, enzymes, ion channel proteins, G-protein-coupled receptors, nuclear receptors and transporters (16, 22). Currently, only 640/22,000 proteins present in the

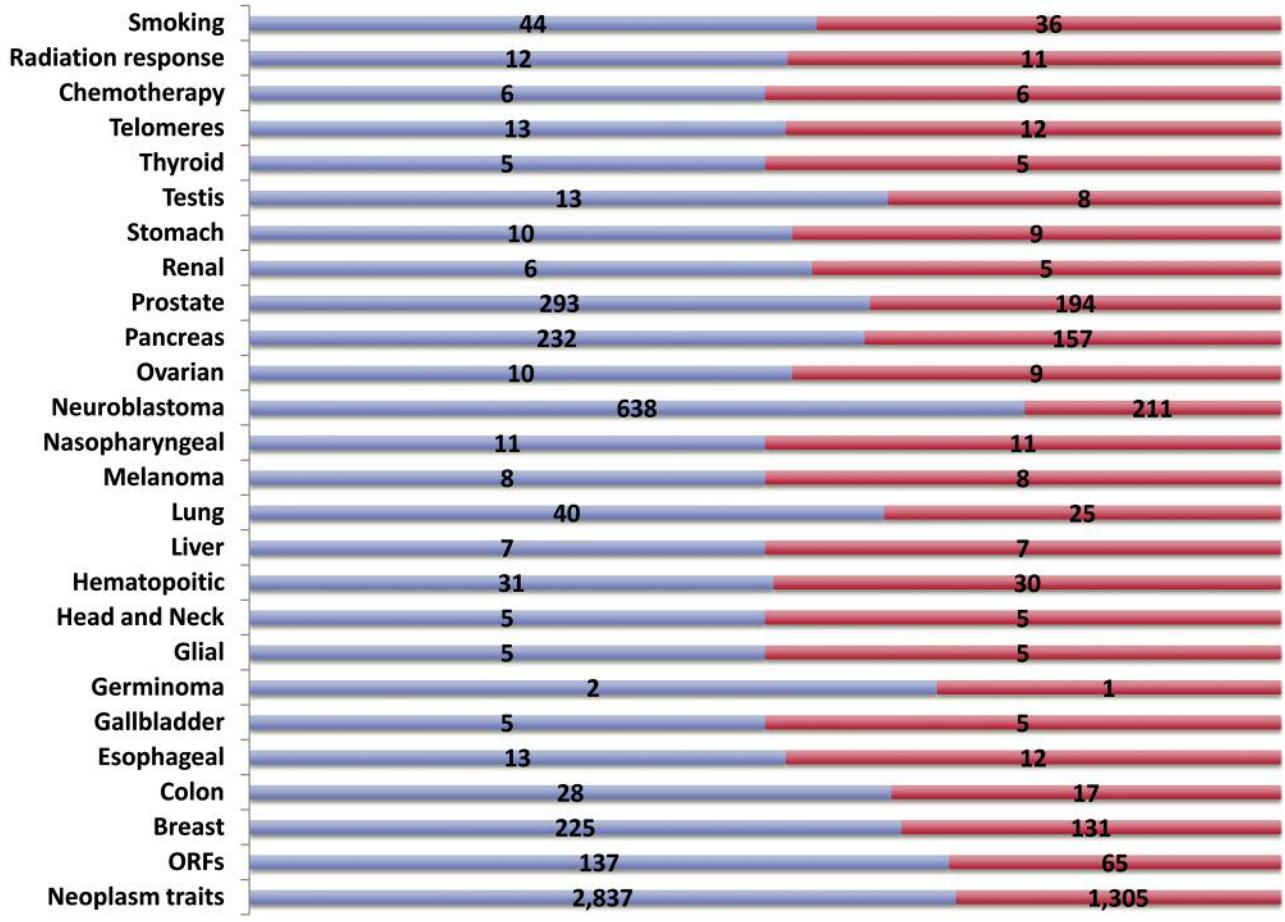

Figure 1. *Neoplasm-associated traits in the human genome. The NCBI Phenome-Genome Integrator was used to enrich neoplasm-associated traits. For the indicated traits, the number of genes (red) and the number of associated SNPs (blue) are shown. p-Values $<1\times10^{-5}$. The SNP class included the exon, intron, near gene and untranslated region.*

human genome are targeted by the Federal Drug Administration (FDA)-approved drugs. Additional targets are clearly needed to provide a basis for cancer drug discovery. The 3D structures of the human proteins can be readily mined for ligand-based druggability prediction using chemogenomics approaches (23, 24). A cancer-oriented protein annotation tool, canSAR, provides tools to mine the cancer proteome for druggableness with links to the chEMBL repository of compounds (25-28). By utilizing such an approach, recently novel molecular targets were identified for diverse diseases, including diabetes (29), Ebola virus disease (30), neurodegenerative diseases (31) and pancreatic cancer (32).

To aid in the discovery of novel cancer-related targets and facilitate drug discovery efforts, a database of genes related to the neoplasm-associated traits was established. Protein expression profile of these genes in diverse body fluids was established. Druggable class of proteins from the secretome was analyzed using chemogenomics approaches. Active lead compounds (<1 μM) with drug-like property were identified. These results open up novel opportunities for cancer drug discovery efforts, as well as for the development of new biomarkers.

Materials and Methods

The bioinformatics and proteomics tools used in the study have been described previously (32-34). The protein annotation and chemical structure-based mining was performed using the canSAR integrated knowledgebase 2.0 (26, 35). The browse canSAR section was used and the neoplasm-associated proteins were batch-analyzed for protein annotations, 3D structures, compounds and bioactivity details. The canSAR compounds link for genes has diverse filters, such as activity and assay types, concentrations, molecular weight, rule of five (RO5) violations, prediction of oral bioavailability and toxicophores. The protein 3D structure information was obtained from the Swiss Protein Database (36). The chemical structures were obtained from the chEMBL (28). Comprehensive gene annotation for the neoplasm- associated genes was established using the GeneCards

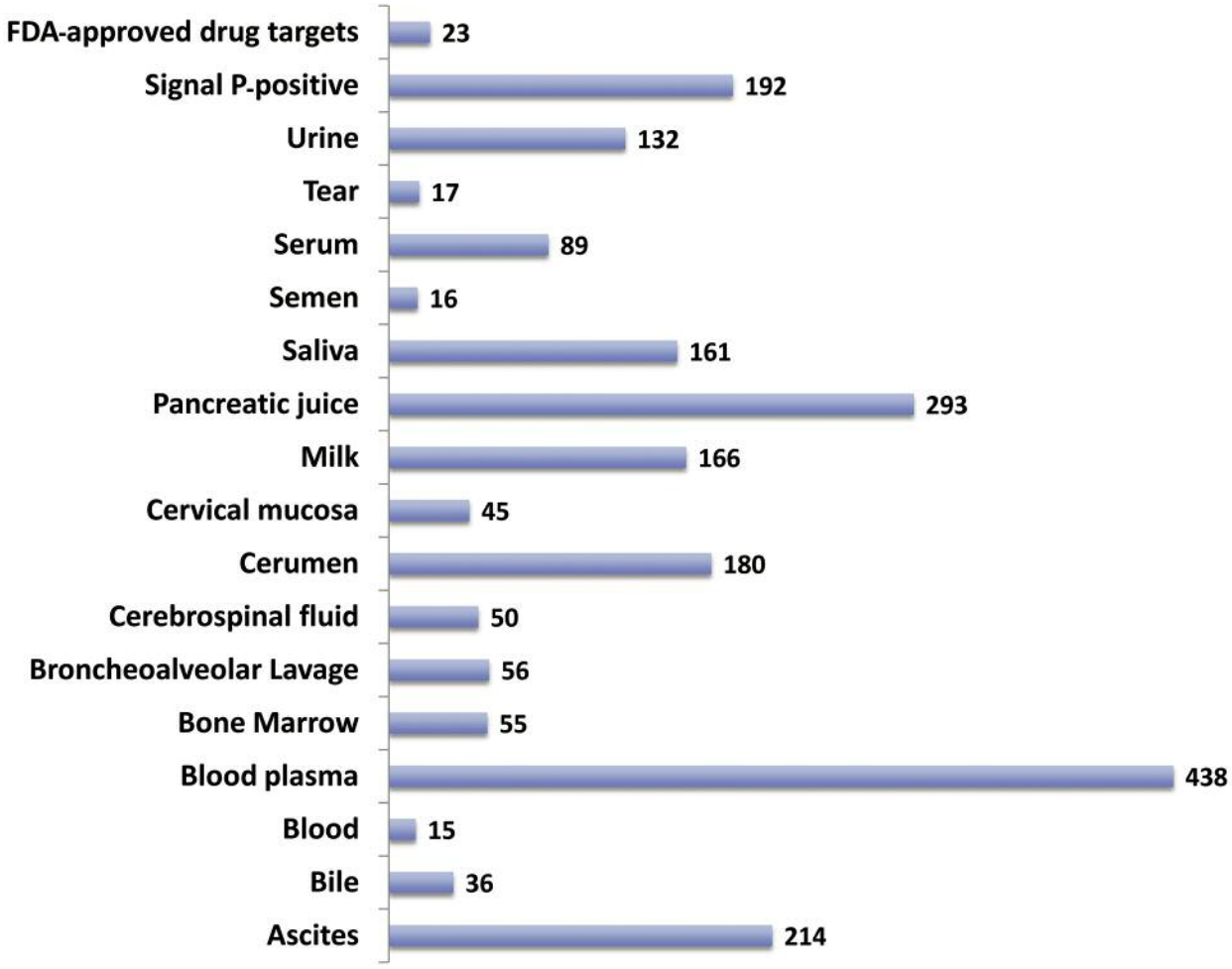

Figure 2. *Cancer proteome expression in diverse human body fluids. The protein expression in diverse body fluids was inferred from the Multi-Omics Protein Expression Database, the Proteomics DB and the Human Protein reference database. The numbers indicate the number of genes detected for each of the body fluids.*

(37), the DAVID functional annotation tool (38) and the UniProt (39) databases. Protein expression was verified using the HPRD (13), the human protein map (HPM) (14), Proteomics DB (17), the MOPED (18- 19) and the human protein atlas (HPA) (16, 40).

Putative drug hits were filtered from the canSAR datasets for the neoplasm-associated genes using the Lipinski's rule of five (also known as Pfizer's rule of five), RO5. The RO5 is a rule of thumb to evaluate druggableness or to determine whether a compound with a certain pharmacological or biological activity possesses properties that would make it a likely orally active drug in humans (41- 42). Highest stringency was chosen for the RO5 violation (value=0). Drugs with half-maximal inhibitory concentration (IC_{50}) values, inhibitory activities and inhibitory constant (Ki) values are chosen for the CanSAR output. The chemical structures were verified using the chEMBL tool (27). Toxicophore negative was chosen to filter the hits for toxicity associated compound structures (43). FDA approved listing of drugs were obtained from the DrugBank database (22).

Results

Cancer polymorphic traits in the human genome. The NCBI PheGenI genetic association studies tool was used to establish an initial database of neoplasm-associated traits. Among the diverse association evidence in the GWAS database, neoplasm-associated genes and the single nucleotide polymorphisms (SNPs) were enriched (Figure 1). The human genome segregated into 1,305 neoplasm-associated genes encompassing 2,837 polymorphic SNPs. Sixty five previously uncharacterized open reading frames (ORFs) were also identified in these studies. These novel ORFs, part of the "Dark Matter" proteome have been recently characterized in the context of cancer and other diseases (34, 44). The neoplasm traits were further classified into individual tumor types, risk factor and response to

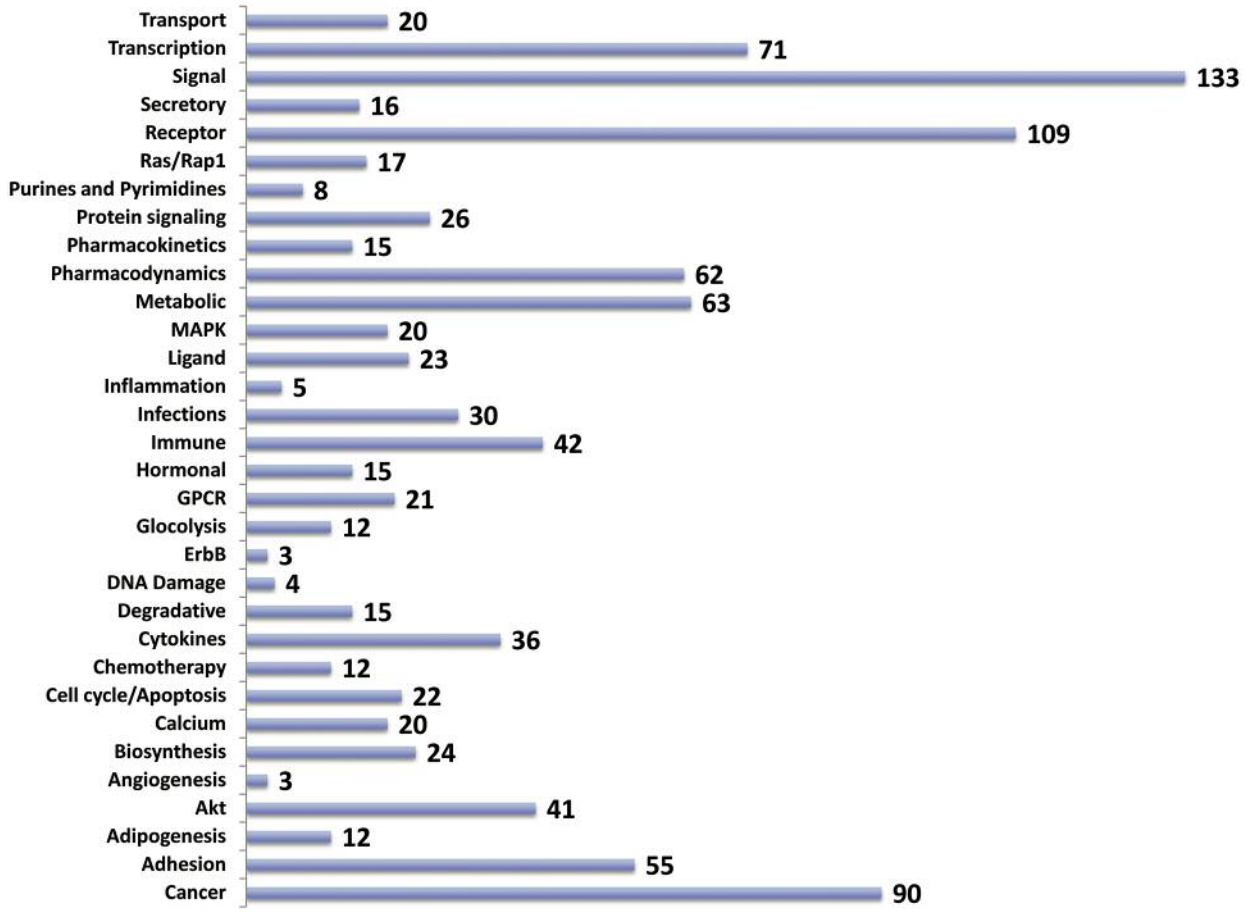

Figure 3. *Pathway mapping of the cancer proteome from the body fluids. The neoplasm-associated proteins were analyzed for gene pathways using the GeneALaCart and the DAVID functional annotation tools. The numbers indicate the number of genes associated with the indicated pathways.*

therapy. The largest number of associated genes and SNPs was seen in neuroblastoma, prostate, pancreatic and breast carcinomas. Association evidence was also seen in telomere function, smoking behavior and response to chemotherapy and radiation. This database of cancer subtype-related association provided a framework for detailed bioinformatics and proteomics characterization.

Enrichment of cancer proteome in the body fluids. The majority of the neoplasm-associated genes (n=1,305) are well-characterized proteins. However, the data for these proteins exist in diverse databases making the lead gene prioritization for cancer drug discovery often difficult. It was reasoned that the protein biomarkers detectable in diverse body fluids (the cancer secretome) might offer a diagnostic and response to therapy indicator potential. Hence, the protein expression databases, the HPRD, the MOPED and the Proteomics DB, which have proteome expression datasets from normal and cancer patient-derived body fluids, were batch-analyzed for the neoplasm-associated genes (Figure 2). A large number of the neoplasm-associated proteins (913/1,305) were detected in diverse body fluids. Among these, 192 proteins had signal peptide sequence predicted by the Signal P program (45-46). The remaining proteins (n=721) belong to the non-classical secretary pathways. Blood plasma, pancreatic juice and the ascitic fluid showed the maximum number of proteins. In addition, neoplasm proteins were detected in cerumen (earwax), milk, saliva, tear and urine.

Pathway analysis. In order to develop a rationale for novel cancer therapeutic targets discovery, a comprehensive pathway mapping was undertaken for the neoplasm-associated proteins. The GeneALaCart and the Database for Annotation, Visualization and Integrated Discovery (DAVID) functional annotation tools were used and batch-

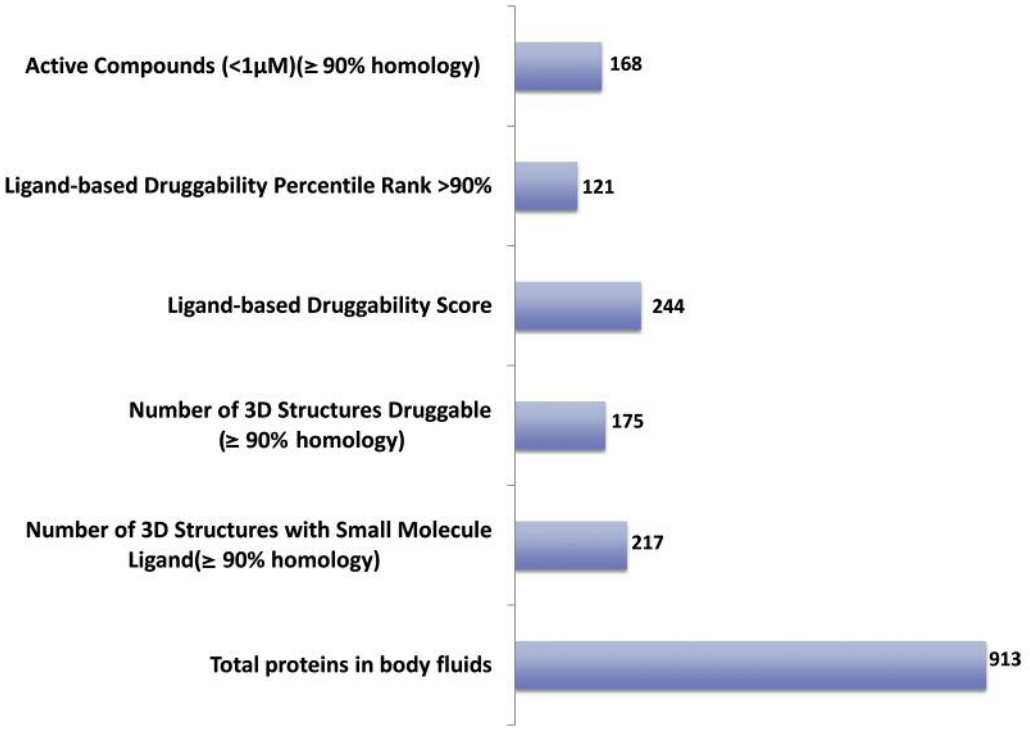

Figure 4. *Druggability analysis of the cancer proteome from the body fluids. The canSAR protein annotation tool was used to batch-analyze the neoplasm-associated proteins from the body fluids. A summary of the output based on the druggable 3D structures of the proteins and the ligand-based druggability indication is shown. The numbers indicate the number of genes for the criteria shown. The filter was chosen at >90% homology for the hits. Only the active compounds for the proteins (<1 μM in bioactivity) are included.*

analyzed to cluster the pathways implicated with the cancer proteome (Figure 3). The pathway data was merged from the output from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, the Interactome pathway, the Biosystem pathways, the Tocris pathway, the Thomson Reuters pathway and the PharmGKB pathway tools. The pathways segregated into cell growth regulation (angiogenesis, apoptosis, cell cycle, DNA damage and degradative), metabolic (biosynthesis, glycolytic, hormonal and purines and pyrimidines), signaling (Akt, cytokines, ErbB, MAPK, Ras/RaP1 and protein), druggable targets (adhesion, ligands, transporters, receptors,and G protein–coupled receptors (GPCRs), pharmacokinetics and pharmacodynamics, chemotherapy and neoplasm-related diseases (inflammation, infection and immune). These results were also verified from the gene ontology analysis from the canSAR and the DAVID functional annotation datasets. Cell localization ontology categorized neoplasm-associated genes into the membrane (n=516), nuclear (n=446), cytosol (n=416), cytoskeletal (n=81), endoplasmic reticulum (n=87), Golgi (n=81), mitochondria (n=77), receptors (n=233), extracellular (n=206), neuron (n=123), axonal (n=81) and vesicle (n=126).

Druggable proteins in body fluids: the cancer secretome. In order to develop an understanding over the druggableness of the neoplasm-associated proteins, genes were classified into classes of coding and non-coding genes. A comparison of all neoplasm-associated genes with the genes identified in the body fluids is shown in Table I. The protein coding genes clustered into druggable class, including cell adhesion molecules, ion channel proteins, enzymes, GPCRs and other receptors and transporters. Enzymes were the largest class of proteins detected in the body fluids (n=144). In addition, 17 uncharacterized ORFs, pseudogenes, RNA genes and linc RNAs were also identified in the body fluids. Utilizing a recent dataset from the HPA (16), these body fluid proteins were further investigated. The list was filtered into authentic secreted proteins with Signal P prediction (n=192) and current FDA-approved drug targets (n=23).

Druggablity profile of cancer secretome. To facilitate the discovery of new cancer therapeutics, the canSAR integrated protein annotation tool was batch-analyzed for the cancer proteome from the body fluids (Figure 4). Using 3D structural evidence and ligand-based druggability ranking (>90% confidence), 121/913 body fluid proteins were

Table I. *Protein classes of the neoplasm-associated genes.*

Protein class	Neoplasm-associated genes	Body fluid genes
Neoplasm-associated proteins	1350	913
Antisense non-protein coding	5	0
Antisense protein coding	1	0
Apoptosis	56	47
Blood group	4	4
Cell adhesion	**53**	**44**
Channel proteins	**19**	**6**
Cytoskeletal	39	8
Endogenous ligands	23	15
Enzymes	**214**	**144**
Factors	97	47
G protein–coupled receptors (GPCR)	**43**	**19**
Immunoglobulins	42	28
Interleukins	10	8
linc RNA	12	1
Lipoproteins	7	3
Open reading frames (ORFs)	65	17
Protein coding	983	908
Pseudogenes	249	4
Receptors	69	31
RNA genes	34	1
Transcription factors	71	47
Transporters	**6**	**2**
Secreted proteins (Signal P)	192	192
FDA-approved drug targets	41	23

The genes associated with neoplasm traits were batch-analyzed using the GeneALaCart Met Analysis tool and classified into major protein classes. The protein expression in the body fluids was established using MOPED and Proteomics DB tools. Druggable class of proteins is shown in bold.

predicted to be capable of binding to a ligand. Twenty percent of the body fluid proteins (175/913) were predicted to be druggable based on 3D structures. Active small-molecular-weight compounds were identified (<1 μM bioactivity) for 168 neoplasm-associated protein targets.

Using a high-stringency definition of druggability percent score (>90%); RO5 violation score (zero); bioactivity cut off (<1 μM); and lack of toxicophore groups, thirty-two secretome proteins were identified as putative drug lead targets (Table II). These lead proteins encompass enzymes (alcohol dehydrogenese, aldo-keto reductase, amine oxidase, demethylase, glucokinase, serine/threonine kinases, lipoxygenease and thymidine phosphorylase), receptors (dopamine receptor, insulin-like growth factor receptor, stem cell growth factor receptor, melanocyte-stimulating hormone receptor and purinoceptor), coagulation factor XIII and cytochrome P450 286 protein.

Cancer lead targets and putative compounds. Seven out of these 32 proteins showed strong Phenome-Genome association evidence in diverse neoplasms: a downstream effector of Cdc42 in cytoskeletal reorganization (CDC42BP) with large B-cell diffused lymphoma, (47); Casein kinase I isoform alpha (CSNK1A1) involved in Wnt signaling with esophageal cancer (48); a Serine/threonine-protein kinase (CHEK2) with stomach neoplasms (49); an intronless melanocyte-stimulating hormone G-protein coupled receptor (MC1R), which is a genetic risk factor for melanoma and non-melanoma skin cancer (50, 51) and basal cell carcinoma (52); a Lysine-specific demethylase 4C (KDM4C), which is an indicator *for* response to radiation (53); a S-methyl-5'-thioadenosine phosphorylase (MTAP), which is co-deleted in diverse tumors along with the tumor suppressor *P16* gene (51) and P2Y purinoceptor 12 (P2RY12), a G-protein coupled receptor with neuroblastoma (54).

These seven lead proteins also had active drug-like compounds (nanomolar bioactivity, RO5 violation value of zero, molecular weight <500 and lack of toxicophore structures) in the chEMBL chemical repository (Table III). A compound against the target, protein, check point kinase 2, CHEK2 (canSAR # 404540|CHEMBL574737), is currently a clinical candidate (55).

Mining the cancer-oriented databases, such as the NCBI ClinVar, the catalogue of somatic mutations in cancer (COSMIC) and the cBioPortal provided additional supporting evidence implicating other tumor types for these genes. Pathogenic clinical variations were seen for the lead genes in malignant melanomas (56-57), neoplastic syndromes, hereditary, familial cancer of breast, Li-Fraumeni syndrome 2 (58), diaphyseal medullary stenosis with malignant fibrous histiocytoma (59) and platelet-type bleeding disorder 8 (60). The COSMIC database showed somatic mutations for diverse tumors and the top three tumor types with mutations are shown (Table III). The cBioPortal Meta analysis tool's mutational assessor was used to identify high impact missense mutations for distinct tumor types.

Discussion

The human proteome consists of over 22,000 proteins and various isoforms associated with these proteins (61). Detailed knowledge of these proteins at the level of variations, polymorphisms, gene ontology, motifs and domains, gene expression at mRNA and protein levels and disease relevance exist across diverse bioinformatics databases. Despite the large number of proteins, only 620 proteins are the basis of mechanism-based FDA approved drugs (DrugBank listing). These target proteins predominantly involve four protein families, such as enzymes, transporters, ion channels and receptors called druggable genes (62-67). The current drugs largely encompass antagonists or agonists for these protein targets. Gene ontology prediction tools indicate that over 70% of the drug targets are membrane-bound or secreted (16).

Table II. *Neoplasm-associated lead proteins*.

Gene Name	Gene description	PheGenI association	Expression in body fluids
ADH1C	Alcohol dehydrogenase 1B		Blood plasma, pancreatic juice
AKR1C2	Aldo-keto reductase family 1 member C2		Bile, earwax, semen
AKT1	RAC-alpha serine/threonine-protein kinase		Milk
AKT2	RAC-beta serine/threonine-protein kinase		Semen
ALOX5	Arachidonate 5-lipoxygenase	Blood pressure	Ascites, blood plasma, saliva, semen
CDC42BPA	Serine/threonine-protein kinase MRCK alpha	Monocytes	Ascites, pancreatic juice
CDC42BPB	**Serine/threonine-protein kinase MRCK beta**	**Lymphoma, Large B-cell, Diffuse**	**Semen, urine**
CDC42BPG	Serine/threonine-protein kinase MRCK gamma	Leprosy	Blood plasma
CHEK2	**Serine/threonine-protein kinase Chk2**	**Stomach Neoplasms|Optic Disk**	**Blood plasma, proximal fluid**
CSNK1A1	**Casein kinase I isoform alpha**	**Esophageal Neoplasms**	**Ear wax, semen**
CYP2B6	Cytochrome P450 2B6	Smoking	Pancreatic juice
DRD1	D(1A) dopamine receptor		Ear wax
F13A1	Coagulation factor XIII A chain	Alzheimer Disease|Walking|Lipoproteins, VLDL|Lipids|Triglycerides	Blood plasma, brancheoalveolar lavage, ear wax, serum, saliva,semen
GCK	Glucokinase	Glucose|Hemoglobin A, Glycosylated	Blood plasma
IGF1R	Insulin-like growth factor 1 receptor	Body height|Sleep|Respiratory function tests	Blood plasma, semen
ITPR1	Inositol 1,4,5-trisphosphate receptor type 1	Triglycerides|Cholesterol	Blood plasma, serum, pancreatic juice, semen
KDM4C	**Lysine-specific demethylase 4C**	**Response to radiation|Body Mass Index**	**Blood plasma**
KIT	Mast/stem cell growth factor receptor Kit		Blood plasma
MAOA	Amine oxidase [flavin-containing] A		Pancreatic juice
MC1R	**Melanocyte-stimulating hormone receptor**	**Carcinoma, Basal Cell|Hair Color|Melanosis**	**Blood plasma**
MTAP	**S-methyl-5'-thioadenosine phosphorylase**	**Melanoma**	**Ascites, ear wax, semen,** blood plasma, milk, proximal fluid, serum
MTOR	Serine/threonine-protein kinase mTOR	Corneal topography	Pancreatic juice
P2RY12	**P2Y purinoceptor 12**	**Neuroblastoma**	**Pancreatic juice**
PIK3C2A	Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit alpha	Schizophrenia	Blood plasma, pancreatic juice
PIK3C2B	Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta	Echocardiography|Subcutaneous Fat|Cholesterol, LDL|Abdominal Fat|Body Weights and Measures	Blood plasma
PIM1	Serine/threonine-protein kinase pim-1		Bone marrow, blood plasma
PLK2	Serine/threonine-protein kinase PLK2		Blood plasma
PRKD1	Serine/threonine-protein kinase D1	Blood pressure	Pancreatic juice
RAF1	RAF proto-oncogene serine/threonine-protein kinase	Cardiomegaly	
RIOK1	Serine/threonine-protein kinase RIO1		Ear wax
SMG1	Serine/threonine-protein kinase SMG1		Pancreatic juice
TBK1	Serine/threonine-protein kinase TBK1		Pancreatic juice, semen
TYMP	Thymidine phosphorylase	Erythrocyte indices	Ascites, blood plasma, broncheoalveolar lavage, ear wax, semen

Lead neoplasm-associated secretome proteins were identified using the ligand-based druggability percentile rank (>90%) and with active compounds (<1 μM). Genes with neoplasm association evidence are shown in bold. Protein expression from MOPED and Proteomes DB analysis is shown.

The cancer therapeutics, which is moving toward personalized medicine (68), requires additional drug targets. Reasoning that the GWAS databases can provide an attractive starting point for developing a list of druggable cancer targets, mining of the NCBI Phenome-Genome Integrator database was undertaken. The PheGenI tool from the GWAS database identified a total of 1,305 genes associated with distinct neoplasm traits. Largely, these genes encompassed protein-coding genes (983/1,305); however, non-protein coding sequences, including long intergenic RNAs, linc RNAs (n=12), pseudogenes (n=249) and antisense RNAs (n=5) were also part of this list of genes. A significant number of druggable genes (enzymes, receptors, channel proteins, transporters and cell adhesion molecules) were identified in the study (n=404). Furthermore, 192 putative secreted proteins were identified in the neoplasm-associated gene list.

Table III. *Drug-like cancer lead compounds. Lead neoplasm -associated secretome proteins were identified using the ligand-based druggability percentile rank (>90%) and with active compounds (<1 μM). Most active compounds are shown with canSAR ID # and IC_{50} values. The chEMBL IDs are shown in parentheses. Clinical candidate is bolded. Neoplasm association is underlined. Pathogenic clinical variations from NCBI ClinVar are shown. COSMIC mutations are shown for top three tumor types. cBioPortal functional impact score using mutational assessor is shown for high impact missense mutations.*

Gene name	Gene description	PheGenI association	Expression in body fluids	Number of CanSar compounds <1 μM	Most active compound (IC_{50} bioactivity, RO5 violation zero, No toxiocophore, Mwt <500). CanSar #, IC_{50} (CHEMBL #)	NCBI Function (RefSeq summary)	NCBI Clinical variations	COSMIC mutations	cBioPortal
CDC42BPB	Serine/ threonine-protein kinase MRCK beta	Lymphoma, Large B-cell, Diffuse	Semen, urine	5	998028, 300 nM (CHEMBL1908393)	This gene encodes a member of the serine/threonine protein kinase family. The encoded protein contains a Cdc42/Rac-binding p21 binding domain resembling that of PAK kinase. The kinase domain of this protein is most closely related to that of myotonic dystrophy kinase-related ROK. Studies of the similar gene in rat suggested that this kinase may act as a downstream effector of Cdc42 in cytoskeletal reorganization.	Malignant melanoma	Stomach, skin, large intestine	Melanoma, glioblastoma
CHEK2	Serine/threonine-protein kinase Chk2	Stomach neoplasms| Optic disk	Blood plasma, proximal fluid	186	404540, 10 nM (CHEMBL574737) Clinical candidate	In response to DNA damage and replication blocks, cell cycle progression is halted through the control of critical cell cycle regulators. The protein encoded by this gene is a cell cycle checkpoint regulator and putative tumor suppressor. Also, mutations in this gene are thought to confer a predisposition to sarcomas, breast cancer, and brain tumors.	Neoplastic Syndromes, Hereditary, Familial cancer of breast, Li-Fraumeni syndrome 2 PMID 11719428	Central nervous system, large intestine, endometrium	Stomach, uterine, ovary, lung, esophagus, cervix, Head and neck, bladder
CSNK1A1	Casein kinase I isoform alpha	Esophageal neoplasms	Ear wax, semen	43	1051989, 48 nM (CHEMBL2010872)	Casein kinases are operationally defined by their preferential utilization of acidic proteins such as caseins as substrates. Participates in Wnt signaling.	Malignant melanoma	Meningioma, bone, endometrium	Uterine, melanoma, breast, lung, liver, colon
KDM4C	Lysine-specific demethylase 4C	Response to radiation |Body Mass Index	Blood plasma	1	763993, 600 nM (CHEMBL1230640)	This gene is a member of the Jumonji domain 2 (JMJD2) family. The encoded protein is a trimethylation-specific demethylase, and converts specific trimethylated histone residues to the dimethylated	Malignant melanoma	Stomach, endometrium, skin	Adreno-cortical, thyroid, colon, medullo-blastoma,

Table III. *Continued*

Table III. *Continued*

Gene name	Gene description	PheGenI association	Expression in body fluids	Number of CanSar compounds <1 μM	Most active compound (IC_{50} bioactivity, RO5 violation zero, No toxiocophore, Mwt <500). CanSar #, IC_{50} (CHEMBL #)	NCBI Function (RefSeq summary)	NCBI Clinical variations	COSMIC mutations	cBioPortal
						form. This enzymatic action regulates gene expression and chromosome segregation. Chromosomal aberrations and changes in expression of this gene may be found in tumor cells.			uterine, Head and neck
MC1R	Melanocyte-stimulating hormone receptor	Carcinoma, Basal Cell Hair Color Melanosis	Blood plasma	394	830142, 1.3 nM (CHEMBL373821)	This intronless gene encodes the receptor protein for melanocyte-stimulating hormone (MSH). The encoded protein, a seven pass trans-membrane G protein coupled receptor, controls melanogenesis. This receptor is a major determining factor in sun sensitivity and is a genetic risk factor for melanoma and non-melanoma skin cancer.	Cutaneous malignant melanoma 5, risk factor, PMID 8894704, Uv-induced skin damage, susceptibility to, PMID 16463023	Stomach, endo-metrium, Esophagus	Colon, renal cell carcinoma
MTAP	S-methyl-5'-thioadenosine phosphorylase	Melanoma	Ascites, ear wax, semen, blood plasma, milk, proximal fluid, serum	67	786551, 1.7 nM (CHEMBL1243250)	This gene encodes an enzyme that plays a major role in polyamine metabolism and is important for the salvage of both adenine and methionine. The encoded enzyme is deficient in many cancers because this gene and the tumor suppressor p16 gene are co-deleted.	Diaphyseal medullary stenosis with malignant fibrous histiocytoma PMID 22464254, 4713573, 8781110	CNS, skin, urinary track (CNV, loss)	Melanoma, renal cell carcinoma
P2RY12	P2Y purinoceptor 12	Neuro-blastoma	Pancreatic juice	722	1116271, 79 nM (CHEMBL2402259)	The product of this gene belongs to the family of G-protein coupled receptors. This family has several receptor subtypes with different pharmacological selectivity, which overlaps in some cases, for various adenosine and uridine nucleotides. This receptor is involved in platelet aggregation, and is a potential target for the treatment of thromboembolisms and other clotting disorders. Mutations in this gene are implicated in bleeding disorder, platelet type 8 (BDPLT8).	Platelet-type bleeding disorder 8 PMID 12578987	CNS, Cervix, lung (CNV Gain)	Melanoma, adreno-cortical carcinoma

Cancer cells secrete numerous proteins by both classical and non-classical secretory pathways (21, 45). The cancer secretome includes the extracellular matrix components, as well as all the proteins that are released from a given type of cancer cells, such as growth factors, cytokines, adhesion molecules, shed receptors and proteases (20, 68). The secreted proteins in diverse body fluids offer novel biomarker and drug therapy targets. From the neoplasm-associated traits, 913 proteins were identified in diverse body fluids encompassing druggable targets. Only 23 out of these proteins are currently FDA-approved targets (16). Thus, it is distinctly possible that additional biomarkers and drug targets can emerge from the database of the cancer secretome generated in this study.

Using chemoinformatics approaches, 33 of the cancer secretome proteins, encompassing enzymes and receptors, were predicted as druggable. Bioactive compounds (<1 μM) targeting these proteins were identified in the canSAR/ chEMBL databases. These 33 protein targets provide a drug discovery rationale for cancer. Furthermore, the discovery of drug-like bioactive compounds targeting seven of these lead secretome proteins provides an immediate starting point for novel cancer therapeutics. The involvement of the neoplasm-associated proteins with other diseases, such as nicotine addiction, Alzheimer's disease, diabetes, cardiac, hematological, metabolic and respiratory diseases, leprosy and schizophrenia opens-up novel biomarker and therapeutic opportunities for these diseases as well.

In summary, the results presented in this study demonstrate the power of chemogenomics approaches for rational cancer drug discovery. Mining the cancer secretome for neoplasm-associated traits is likely to lead to the discovery of new molecular entities for diagnosis and therapy. The lead compounds identified in the study can be rapidly tested in cell culture and pre-clinical models for efficacy.

Conflicts of Interest

None.

Data Availability

The detailed data of this study as a supplemental Table is available upon request.

Acknowledgements

This work was supported in part by the Genomics of Cancer Fund, Florida Atlantic University Foundation. The Authors would like to thank the canSAR gene annotation tool for valuable datasets and Jeanine Narayanan for editorial assistance.

References

1 Wu X, Scelo G, Purdue MP, Rothman N, Johansson M, Ye Y, Wang Z, Zelenika D, Moore LE, Wood CG, Prokhortchouk E, Gaborieau V, Jacobs KB, Chow WH, Toro JR, Zaridze D, Lin J, Lubinski J, Trubicka J, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Jinga V, Bencko V, Slamova A, Holcatova I, Navratilova M, Janout V, Boffetta P, Colt JS, Davis FG, Schwartz KL, Banks RE, Selby PJ, Harnden P, Berg CD, Hsing AW, Grubb RL, 3rd, Boeing H, Vineis P, Clavel-Chapelon F, Palli D, Tumino R, Krogh V, Panico S, Duell EJ, Quiros JR, Sanchez MJ, Navarro C, Ardanaz E, Dorronsoro M, Khaw KT, Allen NE, Bueno-de-Mesquita HB, Peeters PH, Trichopoulos D, Linseisen J, Ljungberg B, Overvad K, Tjonneland A, Romieu I, Riboli E, Stevens VL, Thun MJ, Diver WR, Gapstur SM, Pharoah PD, Easton DF, Albanes D, Virtamo J, Vatten L, Hveem K, Fletcher T, Koppova K, Cussenot O, Cancel-Tassin G, Benhamou S, Hildebrandt MA, Pu X, Foglio M, Lechner D, Hutchinson A, Yeager M, Fraumeni JF, Jr., Lathrop M, Skryabin KG, McKay JD, Gu J, Brennan P and Chanock SJ: A: Genome-wide association study identifies a novel susceptibility locus for renal cell carcinoma on 12p11.23. Human molecular genetics *21(2)*: 456-462, 2012.

2 Kiyotani K, Mushiroda T, Tsunoda T, Morizono T, Hosono N, Kubo M, Tanigawara Y, Imamura CK, Flockhart DA, Aki F, Hirata K, Takatsuka Y, Okazaki M, Ohsumi S, Yamakawa T, Sasa M, Nakamura Y and Zembutsu H: A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese. Human molecular genetics *21(7)*: 1665-1672, 2012.

3 Yoon KA, Park JH, Han J, Park S, Lee GK, Han JY, Zo JI, Kim J, Lee JE, Takahashi A, Kubo M, Nakamura Y and Lee JS: A genome-wide association study reveals susceptibility variants for non-small cell lung cancer in the Korean population. Human molecular genetics *19(24)*: 4948-4954, 2010.

4 Benjamin EJ, Dupuis J, Larson MG, Lunetta KL, Booth SL, Govindaraju DR, Kathiresan S, Keaney JF, Jr., Keyes MJ, Lin JP, Meigs JB, Robins SJ, Rong J, Schnabel R, Vita JA, Wang TJ, Wilson PW, Wolf PA and Vasan RS: Genome-wide association with select biomarker traits in the Framingham Heart Study. BMC medical genetics *Suppl 1*: S11, 2007.

5 Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, Mentch FD, Qiu H, Glessner JT, Thomas KA, Frackelton EC, Chiavacci RM, Imielinski M, Monos DS, Pandey R, Bakay M, Grant SF, Polychronakos C and Hakonarson H: A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS genetics *7(9)*: e1002293, 2011.

6 Elliott KS, Zeggini E, McCarthy MI, Gudmundsson J, Sulem P, Stacey SN, Thorlacius S, Amundadottir L, Gronberg H, Xu J, Gaborieau V, Eeles RA, Neal DE, Donovan JL, Hamdy FC, Muir K, Hwang SJ, Spitz MR, Zanke B, Carvajal-Carmona L, Brown KM, Australian Melanoma Family Study I, Hayward NK, Macgregor S, Tomlinson IP, Lemire M, Amos CI, Murabito JM, Isaacs WB, Easton DF, Brennan P, PanScan C, Barkardottir RB, Gudbjartsson DF, Rafnar T, Hunter DJ, Chanock SJ, Stefansson K and Ioannidis JP: Evaluation of association of HNF1B variants with diverse cancers: collaborative analysis of data from 19 genome-wide association studies. PloS one *5(5)*: e10858, 2010.

7 Gandhi S and Wood NW: Genome-wide association studies: the key to unlocking neurodegeneration? Nature neuroscience *13(7)*: 789-794, 2010.

8 Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, Kawaguchi T, Tsunoda T, Watanabe M, Takeda A, Tomiyama H, Nakashima K, Hasegawa K, Obata F, Yoshikawa T, Kawakami H, Sakoda S, Yamamoto M, Hattori N, Murata M, Nakamura Y and Toda T: Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nature genetics *41(12)*: 1303-1307, 2009.

9 Sim X, Ong RT, Suo C, Tay WT, Liu J, Ng DP, Boehnke M, Chia KS, Wong TY, Seielstad M, Teo YY and Tai ES: Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. PLoS genetics *7(4)*: e1001363, 2011.

10 Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, Paisan-Ruiz C, Lichtner P, Scholz SW, Hernandez DG, Kruger R, Federoff M, Klein C, Goate A, Perlmutter J, Bonin M, Nalls MA, Illig T, Gieger C, Houlden H, Steffens M, Okun MS, Racette BA, Cookson MR, Foote KD, Fernandez HH, Traynor BJ, Schreiber S, Arepalli S, Zonozi R, Gwinn K, van der Brug M, Lopez G, Chanock SJ, Schatzkin A, Park Y, Hollenbeck A, Gao J, Huang X, Wood NW, Lorenz D, Deuschl G, Chen H, Riess O, Hardy JA, Singleton AB and Gasser T: Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nature genetics *41(12)*: 1308-1312, 2009.

11 Ramos EM, Hoffman D, Junkins HA, Maglott D, Phan L, Sherry ST, Feolo M and Hindorff LA: Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. Eur J Hum Genet *22(1)*: 144-147, 2014.

12 Liang L, Morar N, Dixon AL, Lathrop GM, Abecasis GR, Moffatt MF and Cookson WO: A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. Genome research *23(4)*: 716-726, 2013.

13 Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R and Pandey A: Human Protein Reference Database--2009 update. Nucleic acids research *37(Database issue)*: D767-772, 2009.

14 Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LD, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A, Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang TC, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TS, Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H and Pandey A: A draft map of the human proteome. Nature *509(7502)*: 575-581, 2014.

15 Lawrence RT and Villen J: Drafts of the human proteome. Nature biotechnology *32(8)*: 752-753, 2014.

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

17 Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese JH, Bantscheff M, Gerstmair A, Faerber F and Kuster B: Mass-spectrometry-based draft of the human proteome. Nature *509(7502)*: 582-587, 2014.

18 Kolker E, Higdon R, Haynes W, Welch D, Broomall W, Lancet D, Stanberry L and Kolker N: MOPED: Model Organism Protein Expression Database. Nucleic acids research *40(Database issue)*: D1093-1099, 2012.

19 Montague E, Stanberry L, Higdon R, Janko I, Lee E, Anderson N, Choiniere J, Stewart E, Yandl G, Broomall W, Kolker N and Kolker E: MOPED 2.5 – an integrated multi-omics resource: multi-omics profiling expression database now includes transcriptomics data. Omics: a journal of integrative biology *18(6)*: 335-343, 2014.

20 Karagiannis GS, Pavlou MP and Diamandis EP: Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology. Molecular oncology *4(6)*: 496-510, 2010.

21 Paltridge JL, Belle L and Khew-Goodall Y: The secretome in cancer progression. Biochimica et biophysica acta *1834(11)*: 2233-2241, 2013.

22 Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, Adamjee S, Dame ZT, Han B, Zhou Y and Wishart DS: DrugBank 4.0: shedding new light on drug metabolism. Nucleic acids research *42(Database issue)*: D1091-1097, 2014.

23 Jacoby E: Computational chemogenomics. Wiley Interdisciplinary Reviews: Computational Molecular Science. *1(1)*: 57-67, 2011.

24 Flight MH: Chemogenomics: A change of tactic. Nature reviews Drug discovery *7(2)*: 122-124, 2008.

25 Bulusu KC, Tym JE, Coker EA, Schierz AC and Al-Lazikani B: canSAR: updated cancer research and drug discovery knowledgebase. Nucleic acids research *42(Database issue)*: D1040-1047, 2014.

26 Halling-Brown MD, Bulusu KC, Patel M, Tym JE and Al-Lazikani B: canSAR: an integrated cancer public translational research and drug discovery resource. Nucleic acids research *40(Database issue)*: D947-56, 2012.

27 Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, Kruger FA, Light Y, Mak L, McGlinchey S, Nowotka M, Papadatos G, Santos R and Overington JP: The ChEMBL bioactivity database: an update. Nucleic acids research *42(Database issue)*: D1083-1090, 2014.

28 Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B and Overington JP: ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic acids research *40(Database issue)*: D1100-1107, 2012.

29 Delgado AP, Brandao, P, and Narayanan R: Diabetes Associated Genes from the Dark Matter of the Human Proteome. MOJ Proteomics Bioinform *b;1(4)*: 00020-28, 2014.
30 Narayanan R: Ebola-associated genes in the human genome: Implications for novel targets. MOJ Proteomics Bioinform *1(5)*: 00032-41, 2014.
31 Narayanan R: Neurodegenerative diseases: Phenome to genome analysis. MOJ Proteomics Bioinform 1(6): 00033-43, 2014.
32 Narayanan R: Phenome-Genome Association Studies of Pancreatic Cancer: New Targets for Therapy and Diagnosis. Cancer genomics & proteomics 12(1): 9-19, 2015.
33 Delgado A, Hamid S, Brandao P and Narayanan R: A novel transmembrane glycoprotein cancer biomarker present in the x chromosome. Cancer genomics & proteomics *11(2)*: 81-92, 2014.
34 Delgado A, Chapado MJ, Brandao P, Hamid S, Narayanan R. Atlas of the Open reading Frames in human diseases: Dark matter of the human genome. MOJ Proteomics Bioinform *2(1)*: 00036-49, 2015.
35 Narayanan R: Druggableness of the Ebola-associated genes in the human genome: Chemoinformatics approaches. MOJ Proteomics Bioinform *2(2)*: 2015, In press.
36 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE: The Protein Data Bank. Nucleic acids research *28(1)*: 235-242, 2000.
37 Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, Nativ N, Bahir I, Doniger T, Krug H, Sirota-Madi A, Olender T, Golan Y, Stelzer G, Harel A and Lancet D: GeneCards Version 3: the human gene integrator. Database 2010; 2010.
38 Huang DW, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protocols *4(1)*: 44-57, 2008.
39 Magrane M and Consortium U: UniProt Knowledgebase: a hub of integrated protein data. Database 2011; 2011.
40 Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S, Wernerus H, Bjorling L and Ponten F: Towards a knowledge-based Human Protein Atlas. Nature biotechnology *28(12)*: 1248-1250, 2010.
41 Lipinski CA: Lead- and drug-like compounds: the rule-of-five revolution. Drug discovery today Technologies *1(4)*: 337-341, 2004.
42 Lipinski CA, Lombardo F, Dominy BW and Feeney PJ: Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced drug delivery reviews *46(1-3)*: 3-26, 2001.
43 Oprea TI, Davis AM, Teague SJ and Leeson PD: Is there a difference between leads and drugs? A historical perspective. Journal of chemical information and computer sciences *41(5)*: 1308-15, 2001.
44 Delgado AP, Brandao P, Chapado MJ, Hamid S and Narayanan R: Open reading frames associated with cancer in the dark matter of the human genome. Cancer genomics & proteomics *11(4)*: 201-213, 2014.
45 Hathout Y: Approaches to the study of the cell secretome. Expert review of proteomics *4(2)*: 239-248, 2007.
46 Petersen TN, Brunak S, von Heijne G and Nielsen H: SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods *8(10)*: 785-786, 2011.
47 Zhang J, Grubor V, Love CL, Banerjee A, Richards KL, Mieczkowski PA, Dunphy C, Choi W, Au WY, Srivastava G, Lugar PL, Rizzieri DA, Lagoo AS, Bernal-Mizrachi L, Mann KP, Flowers C, Naresh K, Evens A, Gordon LI, Czader M, Gill JI, Hsi ED, Liu Q, Fan A, Walsh K, Jima D, Smith LL, Johnson AJ, Byrd JC, Luftig MA, Ni T, Zhu J, Chadburn A, Levy S, Dunson D and Dave SS: Genetic heterogeneity of diffuse large B-cell lymphoma. Proceedings of the National Academy of Sciences of the United States of America *110(4)*: 1398-1403, 2013.
48 Wu C, Hu Z, He Z, Jia W, Wang F, Zhou Y, Liu Z, Zhan Q, Liu Y, Yu D, Zhai K, Chang J, Qiao Y, Jin G, Liu Z, Shen Y, Guo C, Fu J, Miao X, Tan W, Shen H, Ke Y, Zeng Y, Wu T and Lin D: Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. Nature genetics *43(7)*: 679-684, 2011.
49 Abnet CC, Freedman ND, Hu N, Wang Z, Yu K, Shu XO, Yuan JM, Zheng W, Dawsey SM, Dong LM, Lee MP, Ding T, Qiao YL, Gao YT, Koh WP, Xiang YB, Tang ZZ, Fan JH, Wang C, Wheeler W, Gail MH, Yeager M, Yuenger J, Hutchinson A, Jacobs KB, Giffen CA, Burdett L, Fraumeni JF Jr., Tucker MA, Chow WH, Goldstein AM, Chanock SJ and Taylor PR: A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. Nature genetics *42(9)*: 764-767, 2010.
50 Barrett JH, Iles MM, Harland M, Taylor JC, Aitken JF, Andresen PA, Akslen LA, Armstrong BK, Avril MF, Azizi E, Bakker B, Bergman W, Bianchi-Scarra G, Bressac-de Paillerets B, Calista D, Cannon-Albright LA, Corda E, Cust AE, Debniak T, Duffy D, Dunning AM, Easton DF, Friedman E, Galan P, Ghiorzo P, Giles GG, Hansson J, Hocevar M, Hoiom V, Hopper JL, Ingvar C, Janssen B, Jenkins MA, Jonsson G, Kefford RF, Landi G, Landi MT, Lang J, Lubinski J, Mackie R, Malvehy J, Martin NG, Molven A, Montgomery GW, van Nieuwpoort FA, Novakovic S, Olsson H, Pastorino L, Puig S, Puig-Butille JA, Randerson-Moor J, Snowden H, Tuominen R, Van Belle P, van der Stoep N, Whiteman DC, Zelenika D, Han J, Fang S, Lee JE, Wei Q, Lathrop GM, Gillanders EM, Brown KM, Goldstein AM, Kanetsky PA, Mann GJ, Macgregor S, Elder DE, Amos CI, Hayward NK, Gruis NA, Demenais F, Bishop JA, Bishop DT and Geno T: MELC Genome-wide association study identifies three new melanoma susceptibility loci. Nature genetics *43(11)*: 1108-1113, 2011.
51 Bishop DT, Demenais F, Iles MM, Harland M, Taylor JC, Corda E, Randerson-Moor J, Aitken JF, Avril MF, Azizi E, Bakker B, Bianchi-Scarra G, Bressac-de Paillerets B, Calista D, Cannon-Albright LA, Chin AWT, Debniak T, Galore-Haskel G, Ghiorzo P, Gut I, Hansson J, Hocevar M, Hoiom V, Hopper JL, Ingvar C, Kanetsky PA, Kefford RF, Landi MT, Lang J, Lubinski J, Mackie R, Malvehy J, Mann GJ, Martin NG, Montgomery GW, van Nieuwpoort FA, Novakovic S, Olsson H, Puig S, Weiss M, van Workum W, Zelenika D, Brown KM, Goldstein AM, Gillanders EM, Boland A, Galan P, Elder DE, Gruis NA, Hayward NK, Lathrop GM, Barrett JH and Bishop JA: Genome-wide association study identifies three loci associated with melanoma risk. Nature genetics *41(8)*: 920-925, 2009.
52 Nan H, Xu M, Kraft P, Qureshi AA, Chen C, Guo Q, Hu FB, Curhan G, Amos CI, Wang LE, Lee JE, Wei Q, Hunter DJ and Han J: Genome-wide association study identifies novel alleles associated with risk of cutaneous basal cell carcinoma and squamous cell carcinoma. Human molecular genetics 20(18): 3718-24, 2011.

53 Niu N, Qin Y, Fridley BL, Hou J, Kalari KR, Zhu M, Wu TY, Jenkins GD, Batzler A, Wang L. Radiation pharmacogenomics: a genome-wide association approach to identify radiation response biomarkers using human lymphoblastoid cell lines. Genome Research *20(11)*: 1482-92, 2010.

54 Maris JM, Mosse YP, Bradfield JP, Hou C, Monni S, Scott RH, Asgharzadeh S, Attiyeh EF, Diskin SJ, Laudenslager M, Winter C, Cole KA, Glessner JT, Kim C, Frackelton EC, Casalunovo T, Eckert AW, Capasso M, Rappaport EF, McConville C, London WB, Seeger RC, Rahman N, Devoto M, Grant SF, Li H and Hakonarson H: Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. The New England journal of medicine *358(24)*: 2585-2593, 2008.

55 Matthews TP, Jones AM and Collins I: Structure-based design, discovery and development of checkpoint kinase inhibitors as potential anticancer therapies. Expert opinion on drug discovery *8(6)*: 621-640, 2013.

56 Valverde P, Healy E, Sikkink S, Haldane F, Thody AJ, Carothers A, Jackson IJ and Rees JL: The Asp84Glu variant of the melanocortin 1 receptor (MC1R) is associated with melanoma. Human molecular genetics *5(10)*: 1663-1666, 1996.

57 Nakayama K, Soemantri A, Jin F, Dashnyam B, Ohtsuka R, Duanchang P, Isa MN, Settheetham-Ishida W, Harihara S and Ishida T: Identification of novel functional variants of the melanocortin 1 receptor gene originated from Asians. Human genetics *119(3)*: 322-330, 2006.

58 Lee SB, Kim SH, Bell DW, Wahrer DC, Schiripo TA, Jorczak MM, Sgroi DC, Garber JE, Li FP, Nichols KE, Varley JM, Godwin AK, Shannon KM, Harlow E and Haber DA: Destabilization of CHK2 by a missense mutation associated with Li-Fraumeni Syndrome. Cancer research *61(22)*: 8062-8067, 2001.

59 Camacho-Vanegas O, Camacho SC, Till J, Miranda-Lorenzo I, Terzo E, Ramirez MC, Schramm V, Cordovano G, Watts G, Mehta S, Kimonis V, Hoch B, Philibert KD, Raabe CA, Bishop DF, Glucksman MJ and Martignetti JA: Primate genome gain and loss: a bone dysplasia, muscular dystrophy, and bone cancer syndrome resulting from mutated retroviral-derived MTAP transcripts. American journal of human genetics *90(4)*: 614-627, 2012.

60 Cattaneo M, Zighetti ML, Lombardi R, Martinez C, Lecchi A, Conley PB, Ware J and Ruggeri ZM: Molecular bases of defective signal transduction in the platelet P2Y12 receptor of a patient with congenital bleeding. Proceedings of the National Academy of Sciences of the United States of America *100(4)*: 1978-1983, 2003.

61 Pertea M and Salzberg SL: Between a chicken and a grape: estimating the number of human genes. Genome biology *11(5)*: 206, 2010.

62 Dixon SJ and Stockwell BR: Identifying druggable disease-modifying gene products. Current opinion in chemical biology *13(5-6)*: 549-555, 2009.

63 Griffith M, Griffith OL, Coffman AC, Weible JV, McMichael JF, Spies NC, Koval J, Das I, Callaway MB, Eldred JM, Miller CA, Subramanian J, Govindan R, Kumar RD, Bose R, Ding L, Walker JR, Larson DE, Dooling DJ, Smith SM, Ley TJ, Mardis ER and Wilson RK: DGIdb: mining the druggable genome. Nature methods *10(12)*: 1209-1210, 2013.

64 Hopkins AL and Groom CR: The druggable genome. Nature reviews Drug discovery 1(9):727-30, 2002.

65 Katt WP and Cerione RA: Glutaminase regulation in cancer cells: a druggable chain of events. Drug discovery today *19(4)*: 450-457, 2013.

66 Russ AP and Lampel S: The druggable genome: an update. Drug discovery today *10(23-24)*: 1607-1610, 2005.

67 Sioud M and Leirdal M: Druggable signaling proteins. Methods in molecular biology *361*: 1-24, 2007.

68 Mendelsohn J, Tursz T, Schilsky RL, and Lazar V: WIN Consortium – challenges and advances. Nature reviews Clinical oncology *8(3)*: 133-134, 2011.

Received February 18, 2015
Revised February 28, 2015
Accepted March 3, 2015