Mining Exosomal Genes for Pancreatic Cancer Targets

AMY MAKLER and RAMASWAMY NARAYANAN

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

Abstract. Background: Exosomes, cell-derived vesicles encompassing lipids, DNA, proteins coding genes and noncoding RNAs (ncRNAs) are present in diverse body fluids. They offer novel biomarker and drug therapy potential for diverse diseases, including cancer. Materials and Methods: Using gene ontology, exosomal genes database and GeneCards metadata analysis tools, a database of cancer-associated protein coding genes and ncRNAs (n=2,777) was established. Variant analysis, expression profiling and pathway mapping were used to identify putative pancreatic cancer exosomal gene candidates. Results: Five hundred and seventy-five protein-coding genes, 26 RNA genes and one pseudogene directly associated with pancreatic cancer were identified in the study. Nine open reading frames (ORFs) encompassing enzymes, apoptosis and transcriptional regulators, and secreted factors and five cDNAs (enzymes) emerged from the analysis. Among the ncRNA class, 26 microRNAs (miRs), one pseudogene, one long noncoding RNA (LNC) and one antisense gene were identified. Furthermore, 22 exosome-associated protein-coding targets (a cytokine, enzymes, membrane glycoproteins, receptors, and a transporter) emerged as putative leads for pancreatic cancer therapy. Seven of these protein-coding targets are FDA-approved and the drugs-based on these could provide repurposing opportunities for pancreatic cancer. Conclusion: The database of exosomal genes established in this study provides a framework for developing novel biomarkers and drug therapy targets for pancreatic cancer.

Pancreatic ductal adenocarcinoma is predicted to overtake breast cancer and become the third leading cause of cancer-related deaths in the United States (1). Pancreatic cancer has a very poor prognosis in most patients and seriously lacks effective therapies. Early diagnosis continues to be a challenge in clinics due to advanced-stage presentation. Metastasis and resistance to current chemotherapeutics further contributes to the poor prognosis (2). Most therapies for pancreatic cancer still revolve around the use of ineffective cytotoxic drugs. Novel drug targets and early-stage biomarkers are urgently needed for this untreated cancer.

Exosomes are 40-150 nm extracellular vesicles (EVs) released by all cell types (3, 4). Exosomes are highly heterogeneous (5) and likely reflect the phenotype of the cell that generate them. Similar to cells, exosomes are composed of a lipid bi-layer and contain all known molecular constituents of a cell, including proteins, microRNAs (miRs), mRNA, and DNA (6-8). This lipid bi-layer offers high stability in body fluids, making the exosomes highly attractive targets for diagnostic markers (9). Exosomal secretion involves either a constitutive release via the trans-Golgi network or by an inducible release mechanism (10). While the precise function of the exosomes is unclear, it is suggested they may be involved in removal of cellular waste (11).

The levels of exosomes in cancer patients are higher than the levels found in normal patients (12-14). This phenomenon may be due to altered physiology in diseased organs. Tumor-derived exosomes have anti-tumorigenic properties including induction of apoptosis and enhancement of anti-tumor immunity as well as pro-tumorigenic properties such as tumor growth, angiogenesis and progression (14-16). The exosomal proteins, nucleic acids and noncoding RNAs detectable in diverse body fluids offer novel biomarker potential for cancer (17-20). For early detection of pancreatic cancer, secreted biomarkers could make a significant improvement over the current options (6, 21).

As an initial step towards developing a comprehensive knowledgebase of pancreatic cancer exosomes, a working database of cancer-associated exosomal genes was generated using bioinformatics approaches. From this database of cancer exosomal genes, pancreatic cancer targets encompassing
protein-coding genes, open reading frames (ORFs) and non-coding RNAs, (ncRNAs) including antisense, pseudogenes, long noncoding RNAs (LNC) and microRNAs (miRs) were identified. Enrichment analysis involving somatic mutations, mRNA expression, genome-wide association, protein expression and pancreatic cancer pathways enabled the identification of pancreatic cancer lead candidate genes. Druggable classes of proteins including enzymes, membrane glycoproteins, receptors, and transporters were identified. Seven protein-coding genes from these leads are current target for FDA approved drugs. The targets identified in the study offer a strong rationale for novel biomarkers discovery for the diagnosis and therapy of pancreatic cancer.

Materials and Methods

Knowledge databases. Disease and knowledge-oriented databases MalaCards (22), DisGeNET (23), UCSC genome browser (24) and European Bioinformatics Institute (http://www.ebi.ac.uk/) were used to obtain general information. The Gene Ontology tool, QuickGO (25), was used to classify exosome-associated genes. Exosome encyclopedia, Exocarta (26), was used to develop an initial working database of exosomal protein-coding genes, ncRNAs, secreted products in diverse body fluids and individual cancer-related genes. Comprehensive bioinformatics analysis of the database of exosomal genes was performed using the GeneAlacart tool from the LifeMap GeneCards Suite (27). The GeneAnalytics Gene Analysis tool (27) was used to categorize results into lists of matched tissues, cells, diseases, pathways, compounds and gene ontology (GO) terms to enhance gene set interpretation. Associations between genes and diseases/phenotypes based on shared pathways, interaction networks and paralogy relations were inferred using the VarElect NGS Phenotyper of the GeneCards Suite (27). VarElect utilizes the Deep LifeMap Knowledgebase to infer direct, as well as indirect, links between genes and phenotypes. Indirect association between genes and disease are based on shared pathways and interaction networks analysis.

Other bioinformatics analyses. Protein expression of the candidate genes was investigated using the Human Proteome Map (28) and the Proteomics DB (29). Genome-wide association was verified using the NCBI Phenome-Genome Integrator (PheGenI) (30) and the Genome-Wide Association Studies Catalogue, GWAS (31). Clinical variations were verified using the Clinvar database (32). Protein motifs and domains were verified using the Prosite-Expasy (33). The protein-related information was obtained from the UniProtKB (34). Cancer-related databases COSMIC (35), cBIOPortal (36), International Cancer Genome Consortium (ICGC) (37) and Oncomine (38) were used to verify cancer relevance of the Human Genome Nomenclature Committee was used to verify the gene symbols used in the database (39).

Uncharacterized ORF proteins were compared against the dark matter ORF database previously described (40). Various ncRNA databases MIRMine (http://guanlab.ccb.med.umich.edu/mirmine/), MIR Cancer (41), mir2Disease (42), MIRDB (43), MIRTarBase (44), miRFocus (http://mirfocus.org), LNC2Cancer (45), IncRNA and Disease (46), IncRNA DB (47) and StarBase (48) were used to verify long noncoding RNAs (Lnc RNAs), long intergenic RNAs (Linc RNAs), micro RNAs (miRs), pseudogenes and antisense RNAs. Pancreatic cancer expression was verified using the Array express from European Bioinformatics Institute (49), pancreatic cancer database (http://www.pancreaticcancerdatabase.org) and Pancreatic expression database (50). Other bioinformatics tools used in the study have been described earlier (51, 52).

MicroRNA nomenclature. A capitalized "miR-" refers to the mature form of the miRNA, while the uncapsulated "mir-" refers to the primary miRNA and the pri-miRNA, and "MIR" refers to the gene that encodes them (53).

Results

Identification of exosomal genes by gene ontology. Gene ontology provides a valuable approach to identify a broad class of genes based on functional query. An initial database of exosome-associated genes was developed using the QuickGo gene ontology annotation tool encompassing known protein-coding genes and uncharacterized Open Reading Frames (ORF). Classes of exosomal proteins (cytoplasmic, nuclear, extracellular, secreted, assembly, biogenesis and regulators) were categorized (Figure 1). In addition to 2,777 known protein-coding genes, uncharacterized cDNAs (n=6), ORFS (n=9) were identified. A majority of the exosomal proteins were extracellular (2748/2777). This initial database of exosomal proteins provided the framework for cancer target discovery.

Datamining the ExoCarta database. ExoCarta, a vesicle encyclopedia (Compendium on extracellular vesicles) is a powerful consolidated exosome database encompassing datasets and links to other bioinformatics resources for protein coding genes (mRNAs and proteins), ncRNAs and lipid particles. The entire database (V5, July 2015) was downloaded and filtering options were used to classify output of genes into classes of coding and noncoding genes (Figure 2). In addition to the protein-coding genes (1,028) and mRNAs (1,767) the noncoding miRs (1,308) were the largest class of genes obtained from the ExoCarta database (Figure 2). This database was cross-verified with the QuickGo-based working database of exosomal proteins and a consolidated database was generated. The consolidated database provided the starting point for target(s) discovery by additional datamining.

Exosomal secretome genes. Secreted proteins and noncoding genes (ncRNAs) present in diverse body fluids offer a biomarker potential for early diagnosis and pharmacogenomics. Hence it was of interest to sub-classify the database of exosomal genes into expression profile, based on detection in the body fluids (Figure 3). Exosomal gene expression was seen in amniotic fluids, ascites, milk, serum, saliva, semen etc. Urinary secretome encompassed the largest number of exosomal genes (n=1,505). These secreted exosomal proteins and ncRNAs would provide the basis for biomarker discoveries for diverse cancers.
Cancer profile of the exosomal genes. Identification of exosomal genes present in individual cancer types is an attractive starting point for cancer-specific target discovery. Individual cancer-related exosomal genes were identified using the Exocarta tool and the cancer profile of the exosomal genes was established (Figure 4). Exosomal genes were present in diverse cancer types including sarcomas, carcinomas and leukemias. The largest number of hits was seen for colon (n=1,615), hepatocellular (n=1,138), neuroblastoma (n=1,279), ovarian (n=2,032), prostate (n=1,048) and squamous carcinoma (n=904). Rat pancreatic cancer genes identified in the study were used to identify the human homologues (n=165). The cancer type-related set of exosomal genes provide an opportunity for cancer-specific target discovery.

Pancreatic cancer exosomal targets. The GeneCards Meta analysis tool was used to develop pancreatic cancer exosomal genes. The output data was filtered for protein coding genes, RNA genes and pseudogenes (Figure 5). The pancreatic cancer genes were further subjected to variant analysis using the VarElect tool and genes predicted to be directly related were identified. The list of genes included protein coding genes (n=575), ncRNAs (n=26) and pseudogene (n=1). The coding genes included druggable classes of proteins (adhesion molecules, enzymes, receptors, transporters, channel proteins), growth factors, cytokines and chemokines, oncogenes and tumor suppressor genes.

Enrichment of pancreatic cancer-associated exosomal protein coding genes. The pancreatic cancer-associated list of exosomal genes shown in Figure 5 was subjected to additional enrichment analysis using the VarElect and GeneAnalytic tool of the GeneCards Suite. Exosomal protein-coding genes related to pancreatic cancer pathways, tissue specific expression, secretome nature and somatic mutations were identified (Figure 6). Twenty-two protein-coding targets (cytokines, enzymes, membrane

---

Figure 1. Exosomal landscape. The Gene Ontology tool QuickGo (EBI) was used to enrich exosomal proteins using various filters. The number of genes for each of the classes of exosomes is shown.

Figure 2. Exosomal proteins and non-coding genes. The ExoCarta database was classified into indicated class of coding and noncoding genes. The numbers indicate the number of genes for each class of genes.
glycoproteins, receptors, transporter) were shared across pancreatic cancer pathways, expression, somatic mutations and were in the secretome. These 22 exosomal proteins offer lead potential for drug discovery.

**Exosomal ncRNAs in pancreatic cancer.** The pancreatic cancer-related ncRNAs identified in Figure 5 were further subjected to variant analysis using the VarElect tool. Twenty-four miRs, one pseudogene, one LNC RNA and one antisense gene were identified as directly related to pancreatic cancer. Two miRs (mir206 and mir30A) were indirectly related to pancreatic cancer (Table I). These IncRNAs and miRs were subjected to comprehensive datamining using diverse ncRNA and miR databases as indicated in the Materials and Methods Section. Fifteen of these miRs target 19 pancreatic cancer exosomal protein-coding genes identified in the study. The
lncRNAs, H19 and HOTAIR, were up-regulated in pancreatic cancers (lnc2cancer and lncRNA database). In addition, fifteen MiRs (MIR100, MIR10b, MIR125B1, MIR145, MIR155, MIR196A1, MIR200A, MIR200C, MIR203A, MIR21, MIR210, MIR212, MIR221, MIR222 and MIR23B) were up-regulated and four MiRs (MIR126, MIR141, MIR373 and MIR375) were down-regulated in pancreatic cancer. Among the miRs identified, four of the MiRs (MIR21, MIR34a, MIR155, and MIR221) were also causally linked to pancreatic cancer as shown by multiple databases (mir2dsdisease, pancreatic cancer expression database, and miR cancer). Furthermore, eight of the ncRNAs (H19, MIR126, MIR141, MIR200a, MIR200c, MIR203a, MIR210, MIR34a, MIR373 and MIR375) were amplified or deleted in over 10% of the pancreatic cancer patients analyzed, n=109 (cBioPortal).

The MiRs (MIR200a, MIR200c, MIR203a, MIR210, and MIR34a) showed a correlation of gene amplification (as measured by cBioportal) with an up-regulation of gene expression in the pancreatic cancers (pancreatic cancer expression database). Expression of the miRs was detected in blood, saliva, semen, serum, sperm, and plasma. These exosomal and secreted lncRNAs and the miRs offer the potential to develop a panel of biomarkers for pancreatic cancer diagnosis.

Novel pancreatic cancer ORF protein targets. QuickGo gene ontology analysis identified (see Figure 1) six uncharacterized cDNAs, two of which shared the same symbol and was therefore considered a single gene, and nine ORFS. Reasoning that these uncharacterized genes may offer a potential for novel target discovery, these genes were subjected to a comprehensive bioinformatics analysis (Table II). The novel protein targets were classified into enzymes, secretome, transcriptional regulator and transmembrane proteins. Expression in diverse body fluids was detected for the ORFs. Genetic association studies revealed an association with asthma, cancer, diabetes, inflammation, obesity, osteoporosis, bone density, metabolic syndrome X, chronic kidney disease and substance-related disorders. Nine of the uncharacterized proteins (C11orf54, C16orf69, C17orf80, C19orf18, C1orf123, C5orf46, C6orf120, EXOSC2 and MARK3) were up-regulated and three of the RNA-processing enzymes (EXOSC8, EXOSC9 and EXOSC10) were down-regulated in pancreatic cancer. Furthermore, eight of these uncharacterized proteins (C17orf80, C19orf18, C1orf123, C2orf16, C6orf120, EXOSC2, EXOSC10 and MARK3) were found to be altered in more than 5% of the pancreatic cancer patients (ICGC dataset). Transcriptome and proteome analysis revealed that the ORFs (C19orf18, C16orf89, C2orf16, C5orf46 and C6orf120) had expression in limited tissues and were absent in the pancreas.

Putative pancreatic cancer exosomal lead genes. The 575 exosomal genes directly related to pancreatic cancer identified by variant analysis (Figure 5), were further
Table I. Pancreatic cancer exosomal ncRNAs.

| ncRNAs (Symbols) | Description | Expression in body fluids (Exocarta) | miRFocus - Lead Protein-Coding Targets |
|------------------|-------------|--------------------------------------|-------------------------------------|
| H19 | H19, Imprinted Maternally Expressed Transcript | Blood, saliva, serum, sperm, plasma | CASP7, IGFBP3, SMO |
| HOTAIIR | HOX Transcript Antisense RNA | | |
| HSP90AA2P | Heat Shock Protein 90 Alpha Family Class A Member 2, Pseudogene | | |
| MIR100 | MicroRNA 100 | Blood, saliva, semen, serum, sperm, plasma | |
| MIR10B | MicroRNA 10b | Blood, saliva, semen, serum, sperm, plasma | |
| MIR125B1 | MicroRNA 125b-1 | Saliva, semen, sperm | |
| MIR126* | MicroRNA 126 | Blood, saliva, semen, serum, sperm, plasma | |
| MIR141* | MicroRNA 141 | Blood, saliva, semen, serum, sperm | PTEN |
| MIR145 | MicroRNA 145 | Blood, saliva, semen, serum, sperm | |
| MIR155* | MicroRNA 155 | Blood, saliva, semen, serum, sperm, plasma | MET, TBHS1 |
| MIR193B | MicroRNA 193b | Blood, saliva, semen, serum, sperm | PTEN |
| MIR196A1 | MicroRNA 196a-1 | Blood, saliva, semen, serum, sperm, plasma | CASP3 |
| MIR200A | MicroRNA 200a | Blood, saliva, semen, serum, sperm | |
| MIR200C | MicroRNA 200c | Blood, saliva, semen, serum, sperm | |
| MIR203A | MicroRNA 203a | Blood, saliva, semen, serum, sperm, plasma | BIRC5, TNF |
| MIR206 | MicroRNA 206 | Blood, saliva, serum, sperm, plasma | MET |
| MIR211* | MicroRNA 21 | Blood, saliva, semen, serum, sperm, plasma | FASLG, MMP2, MMP9, PTEN, TGFBI |
| MIR210 | MicroRNA 210 | Blood, saliva, semen, serum, sperm | |
| MIR221* | MicroRNA 221 | Blood, semen, serum, sperm, plasma | CASP3, PTEN |
| MIR222 | MicroRNA 222 | Blood, saliva, semen, serum, sperm, plasma | |
| MIR23B | MicroRNA 23b | Blood, saliva, semen, serum, sperm, plasma | MET, PTEN, SRC |
| MIR29A | MicroRNA 29a | Blood, saliva, semen, serum, sperm, plasma | PTEN |
| MIR29C | MicroRNA 29c | Blood, semen, serum, plasma | |
| MIR33A | MicroRNA 33a | Blood, saliva, semen, serum, sperm, plasma | TBS1, BAX, MET |
| MIR37A* | MicroRNA 37a | Blood, semen, serum, sperm, plasma | MAP2K1, MET, MST1, RALB |
| MIR37B* | MicroRNA 37b | Semen, sperm | |
| MIRLET7A1 | MicroRNA Let-7a-1 | Blood, saliva, semen, serum, sperm, plasma | CASP3, IGFBP3, THBS1 |

The exosomal ncRNAs identified in the study were analyzed using pancreatic cancer expression, mutations and ncRNA databases. Up-regulated ncRNAs in pancreatic cancer are shown bolded. Over 10% of the patients showing gene amplifications or deletions are underlined. The miR targets for the 22 protein coding genes were identified by miRFocus tool. *Down-regulated in pancreatic cancers; †Causally linked.

enriched using filters encompassing somatic mutations, pathways, secretome nature and expression in the pancreatic juice. Twenty-two putative exosomal lead proteins emerged from these studies. These proteins encompassed druggable classes (enzymes, transporters and receptors), a cytokine, tumor suppressor genes and proteins involved in apoptosis, cell proliferation and differentiation, angiogenesis and genes involved in nucleotide and protein binding (Table III).

Fifteen proteins ATP binding cassette subfamily B member 1 (ABCB1), BCL2 associated X, apoptosis regulator (BAX1), baculoviral IAP repeat containing 5 (BIRC5), caspase 3, apoptosis-related cysteine protease (CASP3), caspase 8, apoptosis-related cysteine protease (CASP8), CD82 antigen (CD82), Fas ligand (FASLG), Insulin like growth factor binding protein 3 (IGFBP3), MET proto-oncogene, receptor tyrosine kinase (MET), matrix metalloproteinase 2 (MMP2), matrix metalloproteinase 9 (MMP9), smoothered, frizzled class receptor (SMO), SRC proto-oncogene, non-receptor tyrosine kinase (SRC), transforming growth factor beta 1 (TGFBI) and tumor necrosis factor (TNF) were up-regulated in pancreatic cancer. Some of the up-regulated proteins (BIRC5, FASLG, SMO and TNF) were also found to be down-regulated in distinct patient samples suggesting patient-specific variations. Genetic association studies revealed that the proteins BIRC5, FASLG, SMO and TNF had association with multiple cancers including pancreatic neoplasm. These four proteins also showed restricted expression in multiple tissues and were absent in the normal pancreas. The ICGC dataset analysis showed that the proteins ABCB1, MET, MMP2, Phosphatase and Tensin Homolog (PTEN) and TGFBI harbored mutations in more than 10% of the pancreatic
cancer patients. Interestingly, nineteen of the twenty-two pancreatic cancer exosomal proteins were also targets for the exosomal miRs identified in the study.

Comparison of these 22 protein targets against the DrugBank list of targets identified seven Federal Drug Administration (FDA) approved targets for diverse cancers. These targets include Mitogen-activated protein kinase kinase 1.

MAPK1 (metastatic melanoma, unresectable melanoma), MET (Advanced renal cell carcinoma, progressive metastatic medullary thyroid cancer), MMP2 (angiogenesis and metastasis inhibitor), MMP9 (angiogenesis and metastasis inhibitor), SMO (Locally advanced basal cell carcinoma, metastatic basal cell carcinoma), SRC (Chronic myelogenous leukemia) and TNF (multiple myeloma).

**Discussion**

Extracellular vesicular exosomes are becoming attractive targets for diagnosis and therapy and may provide a basis for drug delivery. Secretion of proteins and miRs in the exosomes provides a framework for biomarkers development. The field

| Uncharacterized proteins/cDNAs | Symbol | Motifs/Function | Secreted in | Exosomal miR Targets (miRFocus) |
|-------------------------------|--------|----------------|-------------|--------------------------------|
| C11orf52                      | C11orf52 | Extracellular exosome | Urine, milk | hsa-mir-375 |
| C11orf54                      | C11orf54* | Ester hydrolase/protein, zinc ion binding | Cerumen, saliva, seminal plasma, urine | |
| C16orf89                      | C16orf89* | Gonadotropin, beta chain/protein homodimerization activity | Cerumen, cerebrospinal fluid, proximal fluid, urine | |
| C17orf80                      | C17orf80* | TM Domain/cell migration-inducing | | hsa-mir-125b-5p, hsa-mir-155-5p |
| C19orf18                      | C19orf18* | Signal peptide, secreted transmembrane | | |
| C1orf123                      | C1orf123* | Signal peptide, secreted transmembrane | | |
| C2orf16                       | C2orf16 | P-S-E-R-S-H-S repeats containing DNA pol viral N superfamily/enzyme, transcriptional regulator | | Milk |
| C5orf46                       | C5orf46* | Skin and saliva secreted protein, extracellular exosome | | Saliva |
| C6orf120                      | C6orf120* | Signal peptide/may be involved in induction of apoptosis in CD4(+) T-cells, but not CD8(+) T-cells or hepatocytes. Exosome_RNA_bind1/Hydrolase | | Hepatocyte secretome |
| cDNA FLJ33193                | EXOSC2* | | | hsa-mir-30a-3p, hsa-mir-155-5p |
| cDNA FLJ55260, highly similar to Exosome complex exonuclease RRP4 (EC 3.1.13.-) | EXOSC8* | ExoRNase_PH_dom1/Exosome RNA processing | | |
| cDNA FLJ55260, highly similar to Exosome complex exonuclease RRP4 (EC 3.1.13.-) | EXOSC9* | ExoRNase_PH_dom1/RNA processing | | |
| cDNA FLJ50618, highly similar to Exosome component 10 | EXOSC10* | RNaseH-like_dom3'-5' exonuclease | | hsa-mir-30a-5p, hsa-mir-193b-3p, hsa-mir-222-3p |
| MARK3* | | | | |
| MARK3* | | | | |

The uncharacterized ORF proteins were characterized using the Dark matter ORF database. The cDNAs were analyzed using the UniProtKB database. Approved symbols, protein motifs and function are shown. Disease association was inferred from DisGeNet, MalaCards, GWAS catalogue and the NCBI PheGenI databases. *Up-regulated in pancreatic cancer, ‡down-regulated in pancreatic cancer; over 5% of the patients showing gene amplifications or deletions are underlined. Genes with restricted expression and lack of expression in normal pancreas are in bold.
of exosomes research is continuing to evolve rapidly. In the last twenty years, the number of peer-reviewed publications on exosomes has increased exponentially. The precise function of the exosomes in normal cell physiology and in pathogenesis of diseased phenotype is unclear. Accumulating knowledgebases of exosomes in cancer (54), neurological (55), metabolic and cardiac disease (56), immune function (57) and virus infections (58) is beginning to provide a strong rationale for intervention and therapy for diverse diseases.

Increasing results demonstrate the relevance of exosomal proteins and miRs in the progression of pancreatic cancer (59-61). In pancreatic ductal adenocarcinoma (PDAC) patients, serum-derived exosomes possess significantly higher glypican-1 (GPC1) in 100% cases when compared to the exosomes from healthy donors (12). MET proto-oncogene, receptor tyrosine kinase (MET) is implicated in chemoresistance in PDAC and offers a therapeutic target (62). The PRSS1 gene intron mutations may be a common perturbation in chronic pancreatitis of pancreatic cancer (63).

In PDAC, the involvement of different microRNAs including miR-10a, miR-21, miR-34a miR-150, miR-203, miR-210, miR-212-3p let-7, miR-744, miR-1246 and others are demonstrated in several studies (64). Exosomal microRNA such as miR-212-3p, miR-203, miR-21 have been shown to enhance invasion, modulate immune response and drug resistance (65, 66).

In addition to miRs, long non-coding RNAs (lncRNA) were also identified in the study. One of the lncRNAs, HOX Transcript Antisense RNA (HOTAIR) has been shown to be strongly up-regulated in the saliva of pancreatic cancer patients (67). The H19, imprinted maternally expressed transcript (H19) is involved in promotion of pancreatic cancer metastasis.
perhaps by involving epithelial-mesenchymal transition (68). Polymorphisms in the H19 gene is associated with risk for bladder cancer development (69). The pseudogene, heat shock protein 90 alpha family class A member 2, pseudogene (HSP90AA2P) is significantly associated with perineural invasion and may provide a biomarker (70).

The 22 lead candidate protein-coding genes identified in this study include several druggable classes such as enzymes (protease, metallopeptidases, kinases, GTPase and caspases); receptor (Frizzled Class) and transporters (ATP binding cassette). In addition, cytokines, growth and angiogenic factors were also identified.

Seven of these targets (MAPK1, MET, MMP2, MMP9, SM0, SRC and TNF) are FDA approved drug targets and the drugs are already used in the clinic for diverse cancer therapy. These drugs can be readily repurposed for pancreatic cancer treatment. The other fifteen proteins identified in the study offer new opportunities for cancer-related drug discovery efforts.

Nine exosomal ORF proteins and five cDNAs were identified in the study. These novel uncharacterized genes included enzymes (RNA processing exosomal enzymes), transcriptional regulators and secreted molecules. The cancer association of the C16orf19 (Lung cancer, renal cell carcinoma), C5orf46 (Renal cell carcinoma), EXOSC2 (Liver neoplasms) and MARK3 (brain stem astrocytic neoplasm) provides additional tumor markers potential. The C5orf46 skin and saliva secreted extracellular exosomal protein may provide a noninvasive diagnostic potential (71).

The 22 exosomal protein-coding genes identified in the study in turn were regulated by the exosomal miRs (Tables I and III). This suggests a network of gene regulation for these pancreatic cancer exosomal targets. Further investigations with these genes might offer a potential for development of multiple markers for diagnosis.

Pancreatic cancer diagnosis and therapy needs urgent molecular targets. The poor prognosis of pancreatic cancer due to advanced-stage presentation in most patients necessitates the discovery of novel, reliable, and non-invasive or minimally invasive biomarkers for early detection. The database of pancreatic cancer-related secreted exosomal genes and ncRNAs identified in this study could offer clinical relevance in early detection, improved prognosis and development of targeted therapies. The database of genes developed in the study will be made available freely to investigators as supplemental table upon request.

In summary, the results demonstrate the usefulness of bioinformatics approaches to pancreatic cancer target discovery. The database of pancreatic cancer exosomal proteins and miRs established in the study provides a starting point for discovery of novel diagnostics and therapeutics. A panel of pancreatic cancer related biomarkers encompassing the protein coding genes and ncRNAs can be tested for clinical validation.

Conflicts of Interest

The Authors declare no conflicts of interest.

Data Availability

Complete data is available upon request.

Acknowledgements

This work was supported in part by the Genomics of Cancer Fund, Florida Atlantic University Foundation. The Authors thank Jeanine Narayanan for editorial assistance.

References

1 Siegel RL, Miller KD and Jemal A: Cancer statistics. CA Cancer J Clin 66(1): 7-30, 2016.
2 Zheng X, Carstens JL, Kim J, Scheible M, Kaye J, Sugimoto H, Wu CC, LeBluV VS and Kalluri R: Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature 527(7579): 525-530, 2015.
3 Kahlert C and Kalluri R: Exosomes in tumor microenvironment influence cancer progression and metastasis. J Mol Med (Berl) 91(4): 431-437, 2013.
4 Raposo G and Stoorvogel W: Extracellular vesicles: exosomes, microvesicles, and friends. J Cell Biol 200(4): 373-383, 2013.
5 Koval J, Arras G, Colombo M, Jouve M, Morath JP, Primidal-Bengtson B, Dingli F, Loew D, Tkach M and Thery C: Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. Proc Natl Acad Sci USA 113(8): E968-977, 2016.
6 Kahlert C, Melo SA, Protopopov A, Tang J, Seth S, Koch M, Zhang J, Weitz J, Chin L, Futreal A and Kalluri R: Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. J Biol Chem 289(7): 3869-3875, 2014.
7 Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ and Lotvall JO: Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol 9(6): 654-659, 2007.
8 Balaj L, Lessard R, Dai L, Cho YJ, Pomeroy SL, Breakfield XO and Skog J: Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. Nat Commun 2: 180, 2011.
9 Keller S, Ridinger J, Rupp AK, Janssen JW and Altevogt P: Body fluid derived exosomes as a novel template for clinical diagnostics. J Transl Med 9; 86, 2011.
10 Record M, Subra C, Silvente-Poirat S and Poirat M: Exosomes as intercellular signalosomes and pharmacological effectors. Biochem Pharmacol 81(10): 1171-1182, 2011.
11 Thery C, Zitvogel L and Amigorena S: Exosomes: composition, biogenesis and function. Nat Rev Immunol 2(8): 569-579, 2002.
12 Melo SA, Luecke LB, Kahlert C, Fernandez AF, Gammon ST, Kaye J, LeBlue VS, Mittendorf EA, Weitz J, Rahbari N, Reissfelder C, Pilarsky C, Fraga MF, Piwnica-Worms D and Kalluri R: Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. Nature 523(7559): 177-182, 2015.
Cancer Genomics & Proteomics 14: 161-172 (2017)

13 Zhou B, Xu JW, Cheng YG, Gao JY, Hu SY, Wang L, and Zhan HX: Early detection of pancreatic cancer: Where are we now and where are we going? Int J Cancer, 2017. doi: 10.1002/ijc.30670. [Epub ahead of print]

14 Kalluri R: The biology and function of exosomes in cancer. J Clin Invest 126(4): 1208-1215, 2016.

15 Zhang X, Yuan X, Shi H, Wu L, Qian H and Xu W: Exosomes in cancer: small particle, big player. Journal of Hematology & Oncology 8(1): 83, 2015.

16 Karagiannis GS, Pavlou MP and Diamandis EP: Cancer secretomes reveal pathophysiological pathways in cancer molecular oncology. Mol Oncol 4(6): 496-510, 2010.

17 San Lucas FA, Allenson K, Bernard V, Castillo J, Kim DU, Ellis K, Ehli EA, Davies GE, Petersen JL, Li D, Wolff R, Katz M, Varadachary G, Wistuba I, Maitra A and Alvarez H: Minimally invasive genomic and transcriptomic profiling of visceral cancers by next-generation sequencing of circulating exosomes. Ann Oncol 27(4): 635-641, 2016.

18 Kadota T, Yoshioka Y, Fujita Y, Kuwano K, and Ochiya T: Extracellular vesicles in lung cancer-From bench to bedside. Semin Cell Dev Biol 71048-9521(17): 30140-30144, 2017.

19 Rabinowits G, Gercel-Taylor C, Day JM, Taylor DD and Rabbits J: The biology and function of exosomes in cancer. J Pancreat Disord Ther 10(1): 42-46, 2009.

20 Majem B, Rigau M, Reventos J and Wong DT: Non-coding RNAs in saliva: emerging biomarkers for molecular diagnostics. Int J Mol Sci 16(4): 8676-8698, 2015.

21 Patel GK, Patton MC, Singh S, Khusman M and Singh AP: Pancreatic Cancer Exosomes: Shedding Off for a Meaningful Journey. Pancreat Disord Ther 6(2): e148, 2016.

22 Rappaport N, Twik M, Nativ N, Stelzer G, Bahir I, Stein Tl, Safran M and Lancet D: MalaCards: A Comprehensive Automatically-Mined Database of Human Diseases. Curr Protoc Bioinformatics 47(1.24): 1-19, 2014.

23 Pinoero J, Queralt-Rosinach N, Bravo A, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F and Furlong LI: DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database 2015: baw028, 2015.

24 Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, Dresser TR, Fujita PA, Gurvavado L, Haussler M, Harte RA, Heitner S, Hinrichs AS, Learned K, Lee BT, Li CH, Raney BJ, Rhead B, Rosenblum KR, Sloan CA, Speir ML, Zweig AS, Haussler D, Kuhn RM and Kent WJ: The UCSC Genome Browser database: 2014 update. Nucleic Acids Res 42: D764-770, 2014.

25 Huntley RP, Sawford T, Mutwoto-Muellenen P, Shpytysyna A, Bonilla C, Martin MJ and O Donovan C: The GOA database: gene Ontology annotation updates for 2015. Nucleic Acids Res 43: D1057-1063, 2015.

26 KeertiKumar S, Chisanga D, Arijaratne D, Al Saffar H, Anand S, Zhao K, Samuel M, Pathan M, Ioanis G, Chinthakumar N, Gangoda L and Mathivanan S: ExoCarta: A Web-Based Compendium of Exosomal Cargo. J Mol Biol 428(4): 688-692, 2016.

27 Safran M, Dalal I, Alexander J, Rosen N, Iny Stein T, Shmosh M, Nativ N, Bahir I, Doniger T, Krug H, Sirota-Madi A, Olender T, Golany O, Stelzer G, Harel A and Lancet D: GeneCards Version 3: the human gene integrator. Database 2010: baq030, 2010.

28 Kim MS, Pinto SM, Getnet D, Nirujiogu RS, Manda SS, Chaerkady R, Maduguun A, 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, Radhakrishna A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasmurthy 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, Sirdeshmuk 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.

29 Wilhelm M, Schlegl J, Ahne H, Moghadas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Buttmann L, Jesselto M, Mathieson T, Lemeer S, Schnathbaum K, Reimer U, Wenschu H, Mollenhauer M, Slotta-Huspenina J, Boese HJ, Bantscheff M, Gerstmaur A, Faerber F and Kuster B: Mass-spectrometry-based draft of the human proteome. Nature 509(7502): 582-587, 2014.

30 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.

31 Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flice P, Manolio T, Hindorff L and Parkinson H: The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42(Database issue): D1001-1006, 2014.

32 Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM and Maglott DR: ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42(Database issue): D980-985, 2014.

33 Arimoto P, Jonnalagedda M, Arnold K, Baratin D, Csatari D, Evans B, Der Colas D, Flegel V, Fortier A, Gasteiger E, Grosshans A, Hernandez C, Ioannidis V, Kuznetsov D, Lichten R, Moretti S, Mostaguir K, Rosinchi N, Rossier G, Xenarios I and Stockinger H: ExPaSy: SIB bioinformatics resource portal. Nucleic Acids Res 40(Web Server issue): W597-603, 2012.

34 UniProt C: Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res 41(Database issue): D43-47, 2013.

35 Forbes SA, Bindal N, Bamford S, Cole C, Kok KY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR and Futreal PA: COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 39(Database issue): D945-950, 2011.

36 Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C and Schultz N: The eBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2(5): 401-404, 2012.

37 International Cancer Genome Census, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, Guttmacher A, Guyer M, Hemsley FM, Jennings JL, Kerr D, Klatt P, Kolar P, Kusada J, Lane DP, Laplace F, Youyong L, Nettekoven G, Ozenberger B, Peterson J, Rao TS, Remacle J, Schafer AJ, Shibata T, Stratton MR, Vockley JG, Watanabe K, Yang H, Yuen MM, Knoppers BM, Bobrow M,
