Am J Clin Exp Urol 2019;7(5):327-340
www.ajceu.us /ISSN:2330-1910/AJCEU0102864

Original Article

Bioinformatics analyses of publicly available NEPCa datasets

Siyuan Cheng, Xiuping Yu

Department of Biochemistry and Molecular Biology, LSU Health-Shreveport, Shreveport, USA

Received September 26, 2019; Accepted October 13, 2019; Epub October 15, 2019; Published October 30, 2019

Abstract: Gene expression profiles are valuable resources for the identification of key players that driver disease progression. However, neuroendocrine prostate cancer (NEPCa) specimens are rare, limiting research on this aggressive disease. In this study, we generated a 12-gene signature of NEPCa and used this signature to differentiate NEPCa from prostate adenocarcinoma (AdPCa) samples in publicly available datasets. From these samples, we identified genes that were differentially expressed in NEPCa and AdPCa. Gene ontology and network analyses revealed key players in the pathogenesis of NEPCa, including E2Fs, members of MHC class II, and factors involved in neuron differentiation, neurogenesis, and stem cell signaling. In conclusion, we identified a 12-gene signature of NEPCa and found pathways that are important for the pathologic development of NEPCa.

Keywords: Neuroendocrine, prostate cancer, signature

Introduction

Prostate cancer (PCa) is the most diagnosed non-cutaneous cancer among American men. In 2019, 174,650 men in the US are predicted to be diagnosed with PCa [1]. Although most patients diagnosed with localized PCa can be treated successfully with surgery or radiotherapy, when PCa develops into castrate-resistant prostate cancer (CRPCa), the five-year survival rate drops to below 30% [2, 3]. During CRPCa progression, an aggressive neuroendocrine phenotype (NEPCa) arises in about 25% PCa cases and ensues with high mortality [4]. Understanding the mechanisms that drive NEPCa development could reveal new therapeutic targets and lead to novel strategies for treatment.

Gene expression profiles, including both microarray and RNA-sequencing data, are valuable resources for understanding the changes in transcriptomes during cancer progression. However, due to the fact that late-stage PCa patients are usually not biopsied, NEPCa specimens are rare. It is possible that some of the un-labeled PCa samples in publicly available datasets are NEPCa. However, due to the lack of exclusive biomarkers for NEPCa, these unlabeled NEPCa samples are not easily identified. In this study, we generated a 12-gene signature of NEPCa based on information collected from literature. We applied this signature to publicly available datasets and identified additional NEPCa samples. We then identified and analyzed the genes that were differentially expressed in NEPCa vs AdPCa samples. Our analyses revealed key players in the pathological process of NEPCa.

Methods and results

Data mining and heatmap construction

All RNA microarray data were downloaded from Gene Expression Omnibus (GEO) [5]. Beltran Neuroendocrine Prostate Cancer, SU2C, and the Cancer Cell Line Encyclopedia RNA-Seq datasets were downloaded from cBioPortal [6] and CCLE website [7]. Bioinformatics analyses were conducted by using R studio [8]. The heatmaps and unsupervised clusters were generated by using the pheatmap package [9]. Differentially expressed genes in RNA sequencing datasets were selected by using edgeR package based on a cutoff point of 2-fold change and *P*-Value less than 0.05 [10]. Differentially expressed genes in RNA microarray datasets

were selected by using Limma package [11]. Protein-protein interaction networks were generated by using Cytoscape [12], which is based on STRING database [13]. Network modules were selected by using the Cytoscape built-in APP MCODE [14]. Gene ontology (GO), Kyoto encyclopedia of genes and genomes (KEGG), and upstream regulator analyses were conducted by using ShinyGO V0.60 [15] and Ingenuity Pathway Analysis (IPA) [16].

Generation of a 12-gene signature of NEPCa

Based on literature reviews, we came up with multiple lists of genes involved in NEPCa progression. We adjusted the NEPCa signature genes by applying them to the Beltran Neuroendocrine Prostate Cancer dataset, which contains forty-nine histologically confirmed CRPCa samples including both prostate adenocarcinoma (AdPCa) and NEPCa [17]. First, we performed principal component analysis (PCA) to explore this dataset. As shown in **Figure 1A**, while 45 samples localized together on the PCA plot, 4 samples positioned far away from the majority samples, suggesting that these 4 specimens are not closely related to the other PCa tissues. These 4 samples were removed from the cohort and not used in further analyses.

The mRNA expression data of the rest 45 samples were extracted and then subjected to unsupervised cluster analysis. As a result, we selected a list of 12 genes to define the NEPCa phenotype, including nine up- genes (CHGA, SYP, SRRM4 [18, 30], FOXA2 [19], INSM1 [20], EZH2 [21], ASCL1 [22], NCAM1 [23], SOX2 [24]) and three down- genes (AR, REST, and SPDEF) [25]. Among the up-regulated genes, chromogranin A (CHGA) and synaptophysin (SYP) are common neuronal markers, SRRM4 a mRNA splicing factor [26], FOXA2 a pioneer transcription factor [19], INSM1 a transcription repressor [20], EZH2 a key component of polycomb repressive complex 2 [27], ASCL1 a master neuronal transcription factor [28], NCAM1 a neural cell adhesion molecule [23], and SOX2 a marker of neural stem cells [29]. Among the down-regulated genes, AR controls prostate epithelial cell differentiation, REST represses the expression of neuronal genes, and SPDEF is an Ets transcription factor that regulates the differentiation of epithelial cells [25]. Most of the 12 genes are functionally involved in controlling cell identity and differentiation.

Visualization of the differential expression of NEPCa signature genes

Heatmaps were generated to visualize the differential expression of the 12 genes in NEPCa vs AdPCa. As shown in **Figure 1B**, the 45 samples in the Beltran-cohort clearly formed two major clusters. Cluster 1 contained all the histologically confirmed NEPCa and cluster 2 AdPCa.

We also applied the 12-gene list to the Cancer Cell Line Encyclopedia dataset, which contains RNA-seq data of more than one thousand cancer cell lines including eight (one NE and seven non-NE) PCa cell lines. The mRNA expression data of these 12 genes were extracted from the dataset and subjected to unsupervised cluster analysis [7]. As shown in **Figure 1C**, H660, the only NEPCa cell line, was clearly separated from the other PCa cell lines.

We then applied the 12-gene signature to three datasets of prostatic PDXs. GSE66187 is an array-based dataset that contains 24 PCa PDXs (including histologically confirmed 4 NEPCa and 20 AdPCa) [17]. As shown in **Figure 2A**, the 4 NEPCa samples were clustered into one group and the 20 AdPCa into another one.

GSE32967 is another array-based dataset that contains microarray data from 22 samples representing 3 of each AdPCa and NEPCa PDXs [31]. As shown in **Figure 2B**, the 22 samples formed two major clusters, the NEPCa (n = 14) and the AdPCa (n = 8) clusters. This cluster pattern is concordant with the histological classification.

GSE41192 is also an array-based dataset containing microarray data from 32 samples representing 3 PDXs of NEPCa, 3 clinical specimens of unknown histology, and 26 PDXs of AdPCa [32]. Again, all the 3 NEPCa samples were grouped together following the unsupervised cluster analysis and the other samples fell into the AdPCa group.

Taken together, these results indicate that the 12-gene NEPCa signature could aid us to differentiate NEPCa from AdPCa.

Selecting NEPCa samples from publicly available datasets

Using the 12-gene NEPCa signature, we looked for NEPCa samples from publicly available datasets.

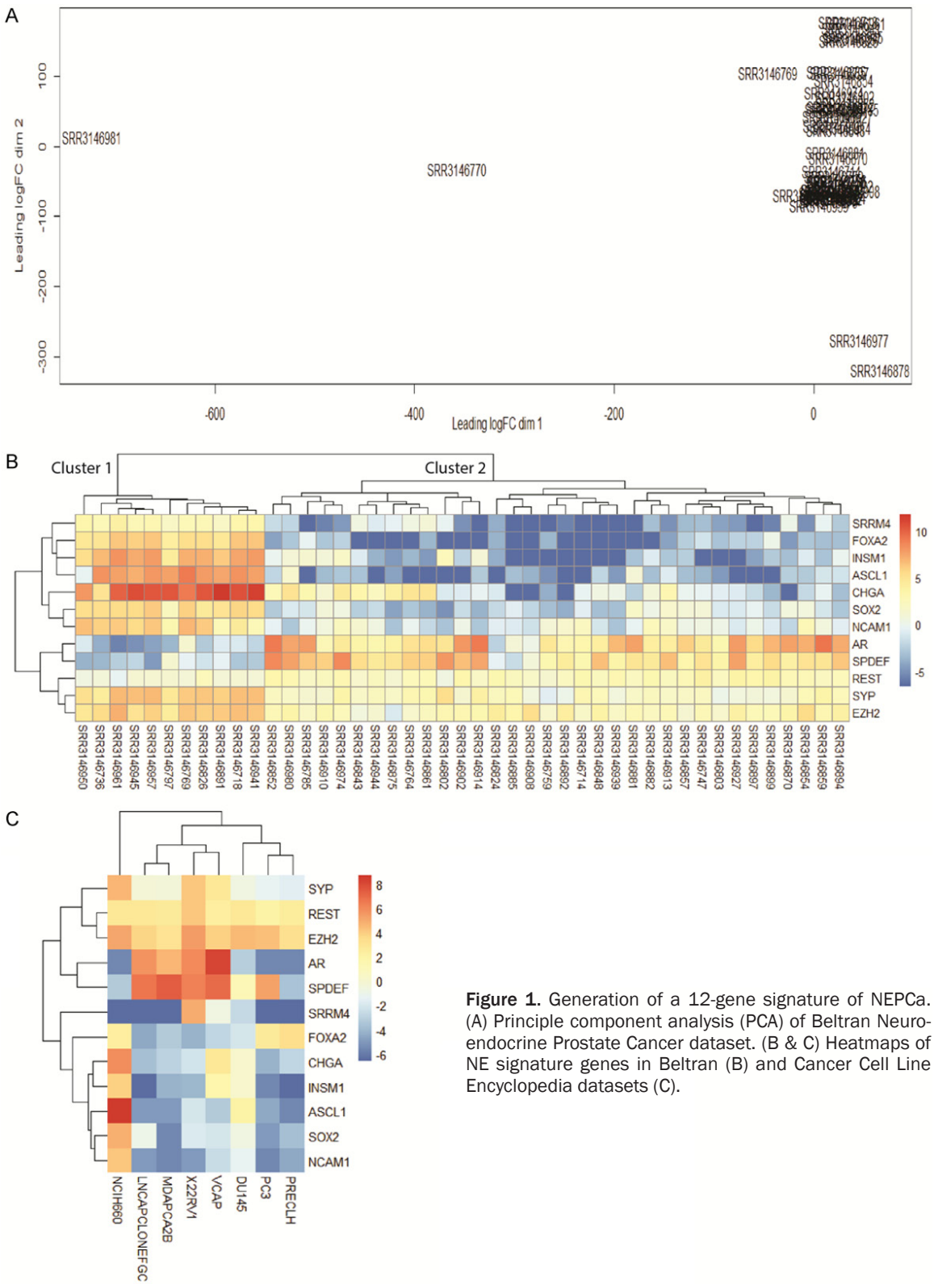

Figure 1. Generation of a 12-gene signature of NEPCa. (A) Principle component analysis (PCA) of Beltran Neuroendocrine Prostate Cancer dataset. (B & C) Heatmaps of NE signature genes in Beltran (B) and Cancer Cell Line Encyclopedia datasets (C).

First, we applied this signature to the SU2C dataset, which contains 444 CRPCa and matched normal tissues [33]. As shown in **Figure 3A**, 5 samples formed a distinct cluster,

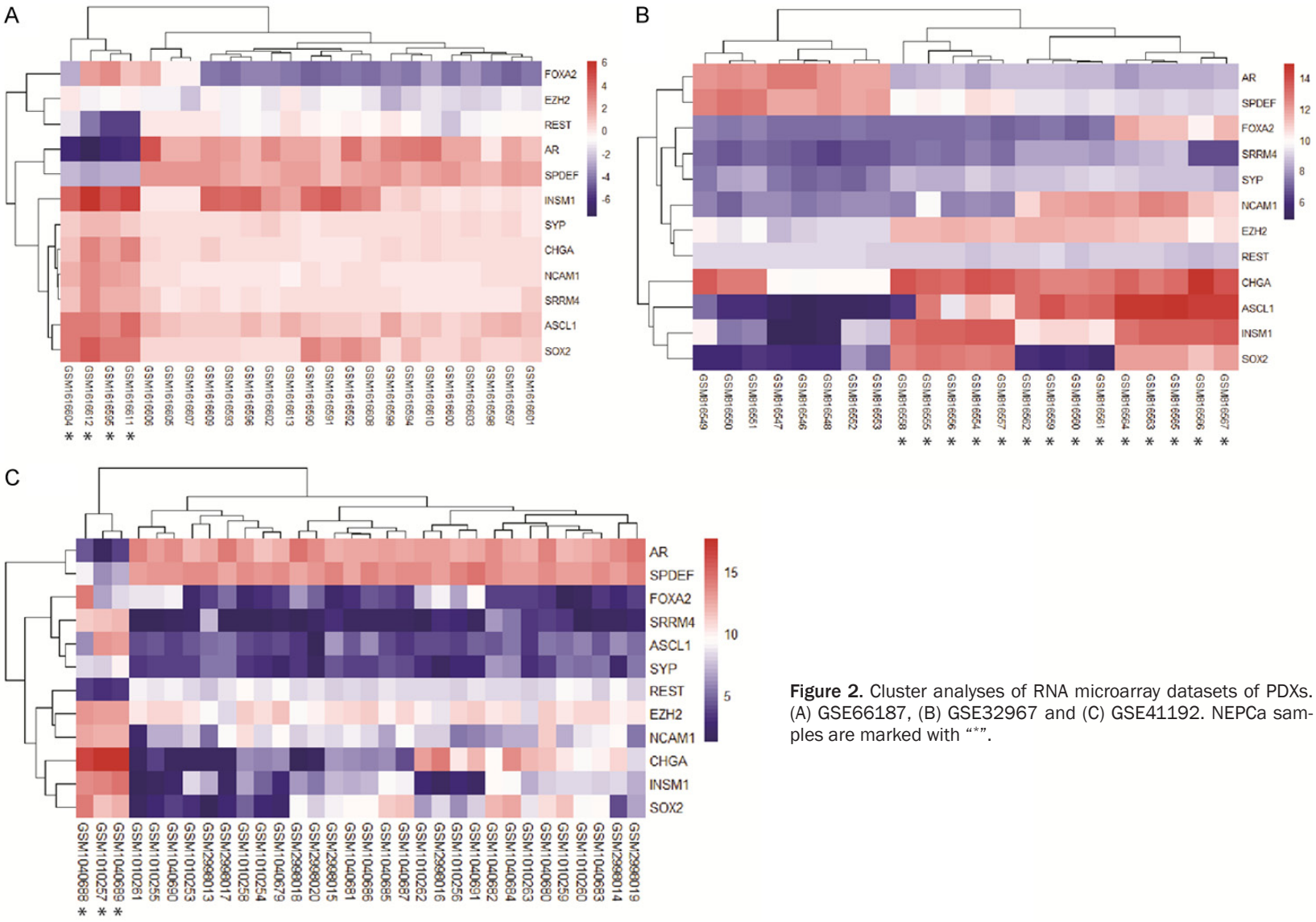

Figure 2. Cluster analyses of RNA microarray datasets of PDXs. (A) GSE66187, (B) GSE32967 and (C) GSE41192. NEPCa samples are marked with "*".

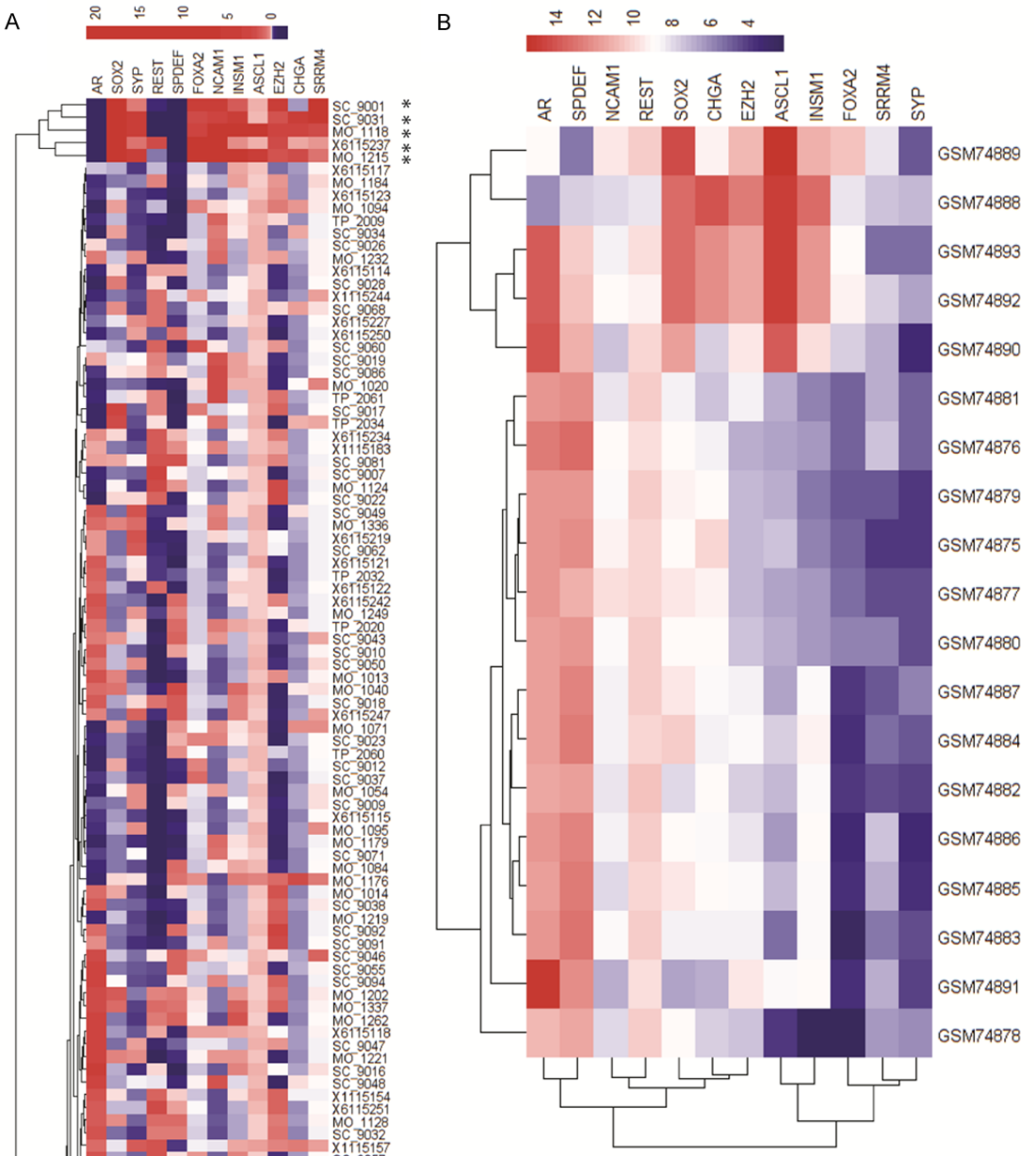

Figure 3. Selecting NEPCa samples in (A) SU2C RNA sequencing dataset and (B) GSE3325 RNA microarray dataset.

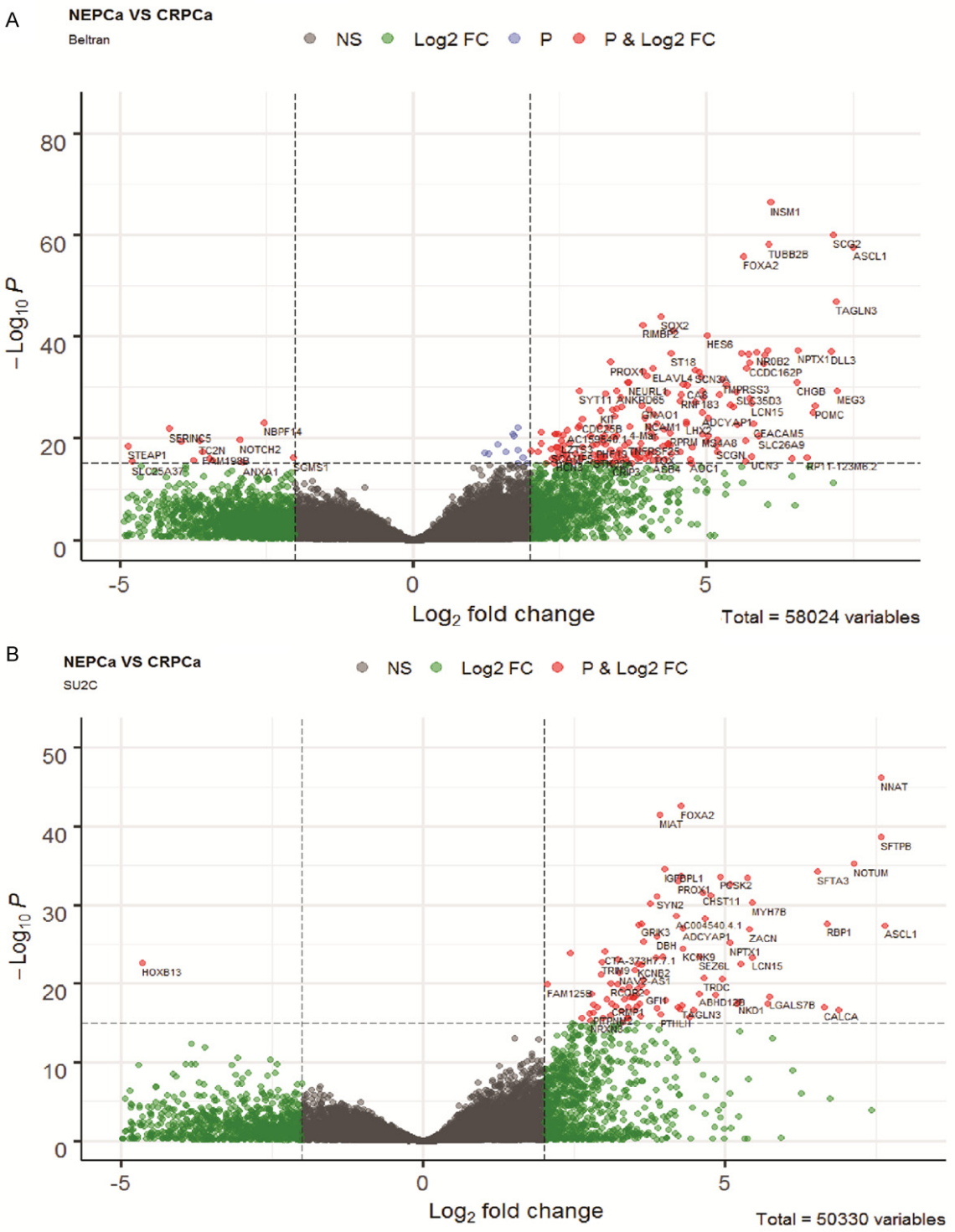

Figure 4. Volcano plot of differentially expressed genes between NEPCa and AdPCa in (A) Beltran and (B) SU2C datasets.

separated from the other samples in this cohort. These 5 samples were selected as NEPCa.

GSE3325 is a microarray dataset that contains 19 prostatic samples including 6 benign prostate, 7 primary PCa, and 6 metastatic PCa tis-

Table 1. Gene Ontology (GO) analysis of common DE genes in Beltran and SU2C datasets

GO Category	Enrichment FDR	Genes in list
Neuron projection development	1.37E-11	122
Neuron differentiation	7.62E-17	172
Neuron development	2.01E-12	137
Neurogenesis	5.94E-18	198
Nervous system development	6.92E-23	277
Brain development	1.24E-10	99
Head development	1.37E-11	106
Generation of neurons	5.94E-18	189
Central nervous system development	4.43E-11	124
Cell proliferation	1.16E-11	214
Cell projection organization	3.86E-13	175
Cell division	3.91E-14	95
Cell development	3.25E-12	221
Cell cycle process	3.63E-11	158
Cell cycle phase transition	9.36E-11	87
Cell cycle	3.63E-11	196
Embryonic organ morphogenesis	3.63E-11	55
Mitotic sister chromatid segregation	6.42E-11	37
Mitotic cell cycle process	3.34E-16	127
Mitotic cell cycle phase transition	4.93E-11	84
Mitotic cell cycle	2.48E-15	138
Animal organ morphogenesis	2.80E-10	125
Animal organ development	2.69E-11	328
Plasma membrane bounded cell projection organization	7.54E-13	171
Organelle localization	1.54E-10	94
Regulation of biological quality	8.91E-15	382
Movement of cell or subcellular component	2.01E-12	215
Regulation of localization	7.56E-11	264
Locomotion	2.57E-11	194
Anatomical structure morphogenesis	1.47E-10	254

sues [34]. As shown in **Figure 3B**, 5 of 6 metastatic PCa samples formed a cluster together and were designated as NE-positive tumors, including 2 AR^- and 3 AR^+ samples. We considered the AR^- samples NEPCa and the AR^+ ones double positive tumors.

Identification of differentially expressed genes and pathway analyses

We first analyzed Beltran and SU2C datasets to identify the differentially expressed (DE) genes (≥ 2 fold, $P < 0.05$) between NEPCa and CRPCa. Both datasets contain RNAseq data of clinical specimens. Volcano plots were generated to visualize the top DE genes in these datasets (**Figure 4A** and **4B**). Also, we merged the DE gene-lists and identified a total of 1396 DE genes that were shared by both datasets. Among these, the most noticeable ones were SCG2 and TUBB2B. SCG2 encodes a neuroendocrine secretory protein and TUBB2B a beta isoform of tubulin, which is highly expressed in brain. The expression levels of both genes were higher in NEPCa compared to AdPCa.

To explore the top biological functions and pathways associated with the differentially expressed genes, we conducted gene ontology (GO) enrichment analysis on the 1396 genes. GO analysis indicated that the top enriched pathways were neurogenesis, neuron related pathways, cell proliferation, and cell cycle related pathways (**Table 1**).

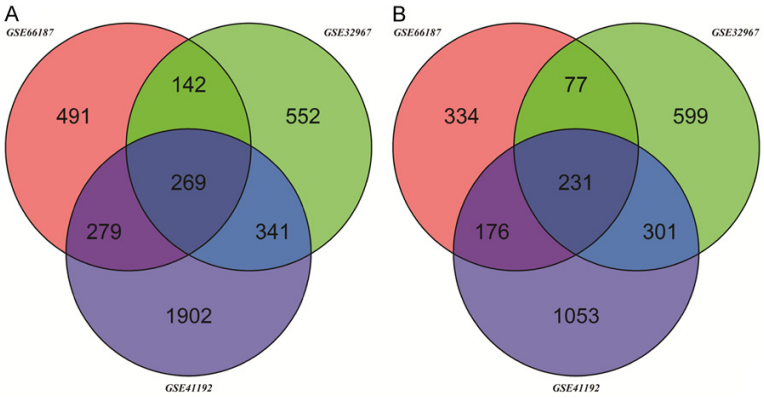

Figure 5. Venn diagram of common up- (A) and down- (B) genes in NEPCa compared to AdPCa in three RNA microarray datasets.

Table 2. Gene Ontology (GO) analysis of up-regulated genes in NEPCa in RNA microarray data

GO Category	Enrichment FDR	Genes in list
Neuron migration	0.000195	6
Neuron differentiation	0.000148	15
Regulation of nervous system development	0.000244	12
Central nervous system development	0.000148	13
Endocrine system development	0.000101	6
Neuron development	0.00029	13
Neurogenesis	4.04E-05	18
Neural nucleus development	8.98E-05	5
Nervous system development	1.30E-05	23
Brain development	0.000224	11
Tissue development	0.000285	18
Telencephalon development	7.92E-05	8
Regulation of timing of cell differentiation	0.000184	3
Regulation of multicellular organismal process	0.000118	24
Regulation of multicellular organismal development	9.45E-06	22
Regulation of developmental process	4.04E-05	23
Regulation of development, heterochronic	0.000218	3
Regulation of cell differentiation	0.000109	18
Pallium development	8.16E-05	7
Locomotory behavior	2.29E-05	8
Generation of neurons	6.44E-05	17
Forebrain neuron fate commitment	0.000148	3
Forebrain development	0.000101	9
Cerebral cortex development	0.000101	6
Cellular developmental process	0.000101	29
Cell differentiation	7.14E-05	29

To increase the number of samples, we included array-based expression data derived from PDXs. Although the number of genes in array-based datasets are limited by the probes on the chips, data generated using PDXs have certain advantages. For example, clinical NEPCa specimens often contain mixed NEPCa and AdPCa components whereas established PDXs tumors exhibit either predominant NEPCa or AdPCa, thus providing more accurate transcriptome information than clinical tissues when they are used to identify the DE genes in NEPCa vs AdPCa. Therefore, we conducted differential gene expression analyses on the three array-based datasets derived from PDXs. To overcome the gene number limitation as well as the high variation in microarray datasets, we first identified the DE genes between NEPCa and AdPCa from each dataset (GSE66187, GSE32967 and GSE41192 excluding clinical specimens) and then merged the lists of DE genes to generate the common up- or down- regulated gene lists. We identified a total of 269 up- and 231 down- regulated genes in NEPCa (**Figure 5A** and **5B**). GO analysis was applied to these two gene lists. The up-regulated genes were enriched in neurogenesis (ASCL1, NKX2-1, PAX6, HMGB3, ASPM, ESR1, POU3-F2, FOXA2, TUBB2B, CHRNA3, NPPB, etc.) and cell proliferation related pathways (BTG3, FGF9, IGFBP2, NELL1, SHOX2, FANCA, CENPF, KIF2C, etc.) (**Table 2**). And the down-regulated genes were enriched in androgen receptor (AR) signaling pathway (NKX3-1, AR, and PMEPA, etc.) and epithelial cell proliferation pathway (CCND1, RB1, NKX3-1, AR, TACSTD, etc.) (**Table 3**).

Then the common DE genes in both RNA sequencing datasets and microarray datasets were merged to generate the up- and down- DE gene lists that represent the transcriptome dif-

Table 3. Gene Ontology (GO) analysis of down-regulated genes in NEPCa in RNA microarray data

GO Category	Enrichment FDR	Genes in list
Diphosphoinositol polyphosphate metabolic process	0.009865	2
Negative regulation of stress fiber assembly	0.009865	3
Androgen receptor signaling pathway	0.010656	4
Negative regulation of actin filament bundle assembly	0.011785	3
Positive regulation of GTPase activity	0.020823	8
Diadenosine polyphosphate metabolic process	0.021976	2
Salivary gland morphogenesis	0.021976	3
Regulation of stress fiber assembly	0.021976	4
Substrate-dependent cell migration, cell extension	0.023445	2
Regulation of epithelial cell proliferation involved in prostate gland development	0.023445	2
Salivary gland development	0.023445	3
Positive regulation of smooth muscle cell migration	0.023445	3
Negative regulation of osteoblast differentiation	0.023445	3

Table 4. Common up-regulated genes in all NEPCa samples

ZDHHC22	CHGB	FAM92A1	KIT	NRXN1	RHEBL1	SYT11
ACTL6B	CHRNA3	FBXO5	KNTC1	NUF2	RIMKLA	SYT13
AP3B2	CKAP2L	FSD1L	KPNA2	NUSAP1	RNF183	TAGLN3
ARHGEF2	CKS1B	GAB2	LHX2	PCNA	RPRM	TEAD1
ARNTL2	CPT1C	GAD1	LMO3	PCSK1	RUNX1T1	TEKT2
ASCL1	CRMP1	GLS2	LPHN3	PCSK2	S100A6	TIMELESS
ASPM	CRTAC1	GPSM2	MAPRE2	PEG10	SCG2	TMEM108
ATP1B1	CTHRC1	HELLS	MARK1	PELI2	SEZ6	TMEM74
BTBD3	DDX11	HES6	MCM2	PGAP1	SEZ6L	TOP2A
CA10	DEK	HOXB3	MCOLN3	PGBD1	SH3GL2	TOX
CA8	DLL3	HOXD8	MELK	PGBD5	SIX2	TPX2
CABP7	DNAJC6	IGFBPL1	MIAT	PHACTR3	SLC35D3	TRIM9
CACNA2D3	DNMT1	INA	MMP16	PHYHIPL	SLC36A4	TUBA1A
CAMK1D	DONSON	INSM1	MYT1	PLP2	SLITRK6	TUBB2B
CCDC88A	DPYSL3	ISL1	NAB1	PLS1	SMC2	TYMS
CCNA2	DPYSL5	ITGB3BP	NCAM1	PRC1	SMC4	UBE2C
CDC20	DUSP26	IVNS1ABP	NELL1	PRMT8	SNAP25	UBE2T
CDC25B	DUSP4	JAKMIP2	NEUROD1	PRTFDC1	SNCAIP	UHRF1
CDCA8	DUSP6	JAM3	NHS	PSIP1	SSX2IP	USP1
CDKAL1	E2F7	KCNH2	NKX2-1	PSRC1	STIL	
CDKN2D	ELAVL3	KIF14	NOL4	RAB3C	STMN1	
CELSR2	ELAVL4	KIF18A	NPTX1	RFC4	STMN3	
CENPF	ESPL1	KIF18B	NRCAM	RGS16	STX1A	
CEP152	ETV5	KIF1A	NRM	RGS7	STXBP1	
CHGA	FAM64A	KIF2C	NRSN1	RHBDL3	SYN2	

ferences between AdPCa and NEPCa in all the samples analyzed. A total of 169 and 135 genes were identified as the common up- (**Table 4**) or down- (**Table 5**) regulated genes in NEPCa, respectively. Moreover, we applied Ingenuity Pathway Analysis (IPA) to identify the key upstream regulators of these genes and listed them in **Table 6** (for up-genes) and **Table 7** (for down- genes). Among these upstream regulators, it is interesting to note that NEPCa specimens exhibited increased expression of neuronal master transcription factors (ASCL1, NEUROD1, and NEUROG3) and decreased expression of REST, a transcription factor that represses neuronal differentiation, and AR, a key modulator of prostatic differentiation. Also, the expression of TP53 and RB1 decreased in NEPCa, which is consistent with previous reports [35]. Additionally, we found that E2F1, CCND1, and FOXM1 were among the list of upstream-regulators and their expressions increased in NEPCa. Given their important roles in cell cycle regulation, these factors

Table 5. Common down-regulated genes in all NEPCa samples

ACACA	CPNE4	FKBP5	MESP1	SLC22A5	TP53INP1
ACPP	CRAT	FOLH1	MYBPC1	SLC25A37	TP53TG1
ACSL1	CREB3L4	FOLH1B	NAP1L2	SLC2A10	TRIB1
ACSM3	CROT	GATA2	NEDD4L	SLC30A4	TRPM4
ADRB1	CRYL1	GLUD1	NFIX	SLC35F2	TRPV6
ADRB2	CSGALNACT1	GNMT	NKX3-1	SLC44A4	VIPR1
AIM1	CUX2	GRHL2	NME4	SLC45A3	VSTM2L
ALDH1A3	CWH43	GUCY1A3	OAZ3	SLC7A8	YAP1
ALDH6A1	CYP1B1	H2AFJ	ORAI3	SPDEF	ZBTB16
AMD1	DCXR	HEBP1	PDE9A	SRXN1	ZCCHC6
AQP3	DHRS7	HERPUD1	PDLIM5	ST6GALNAC1	
AR	DIAPH2	HOMER2	PLA2G2A	STARD3NL	
ARHGAP6	DNASE2B	HOXB13	PLA2G7	STEAP1	
ATP2C2	DPP4	HPN	PMEPA1	STEAP2	
BANK1	EFNA1	ICAM3	PPAP2A	STEAP4	
BCAS1	EFNA4	IDH1	PPP3CA	TACSTD2	
BLVRB	EMB	IL1R1	RAB27B	TC2N	
BMPR1B	ENDOD1	IL6R	RAB3D	THRB	
C1orf116	ENTPD5	INPP5A	RDH11	TMCC3	
C6orf132	EPHX2	ITGB5	RGS10	TMEFF2	
CAB39L	EPN3	KCNN2	RNF135	TMEM144	
CAMKK2	ERGIC1	KLK2	RNF144B	TMEM192	
CGREF1	FAM198B	LATS2	SERINC5	TMEM87A	
COBLL1	FKBP11	MAPKAPK3	SLC10A7	TMPRSS2	
COLEC12	ZG16B	MCCC2	SLC12A8	TNFRSF10B	

Table 6. Key regulators of common up-regulated genes

Upstream Regulator	Activation z-score	*p*-value of overlap
ZBTB17	-3.988	2.24E-18
CCND1	2.714	1.55E-11
E2F4		2.64E-10
TP53	-1.956	3E-10
CREB1	-0.302	6.16E-10
E2F1	3.071	1.93E-09
ASCL1	2.611	2.52E-09
REST	-2.417	3.65E-08
TCF3	-1.8	6.05E-08
FOXM1	3.102	8.98E-08
ATN1		9.64E-08
TBX2	2.828	0.000000134
NEUROD1	2.201	0.000000527
RRP1B		0.00000233
MITF	3.273	0.00000256
POU3F2		0.00000329
HES1	-2.395	0.00000706
RB1	-1.858	0.0000076
TP73	0.181	0.0000146
NEUROG3	2.412	0.0000258

could be the therapeutic targets of NEPCa. Finally, among the upstream regulators, HES1 expression decreased. HES1 is a direct downstream target of Notch signaling. This suggests that Notch signaling is repressed in NEPCa.

Network analysis

Protein-protein interactions (PPIs) network analysis is another way to identify the key players among a list of genes. We applied PPIs network analysis to the common DE genes of Beltran and SU2C datasets. Seven modules were identified as NEPCa-related PPI networks. Module 1 (**Figure 6A**) had the highest number of genes and GO analysis (KEGG) indicated that genes in this module were enriched in cell cycle regulation. Transcription factor binding motif enrichment analysis was conducted by using ShinyGO V0.60. The results revealed that E2Fs were the potential regulators of this group of genes. Genes in module 2 were enriched in GPCR signaling (**Figure 6B**). SSTs/SSTRs could be the upstream regulator of this subgroup. Module 3 contained genes in MHC class II family (**Figure 6C**). The expression of these genes decreased in NEPCa vs AdPCa in both Beltran and SU2C datasets. Genes in module 4 were related to cell secretion and exocytosis (**Figure 6D**) and genes in module 5 were enriched in neuron differentiation and neurogenesis processes (**Figure 6E**). Some of our NE signature genes were located inside module 5,

Table 7. Key regulators of common down-regulated genes

Upstream Regulator	Activation z-score	*p*-value of overlap
AR	-2.259	2.04E-10
TP53	-2.542	7.31E-02
ESR1	-0.847	3.40E-02
CTNNB1	-2.414	1.26E-03
CCND1	2.804	1.10E-06
ERG		6.48E-08

including FOXA2 and ASCL1. Promotor analysis indicated that these genes could be potentially regulated by C2H2 ZF family transcription factors EGR1 and ZNF263. Genes in module 6 were enriched in stem cell pluripotency, embryo development and nervous system development (**Figure 6F**), suggesting that genes in this module could endow NEPCa cells stem-like features. HOXB3 and HOXB5-7, the neuron specific homeobox genes, were also positioned into this module. Their expression levels increased in NEPCa compared to AdPCa, consistent with the acquisition of neuronal features of NEPCa. In contrast, the expression of HOXB13, which is specifically expressed in prostate, was lost in NEPCa, indicating the loss of prostate differentiation in NEPCa. Genes in module 7 were related to epithelium development (**Figure 6G**), including NKX3.1 and SPDEF, both of which are highly expressed in prostate epithelia. The expression of both genes decreased in NEPCa.

Discussion

In this study, we generated a 12-gene NEPCa signature. By applying this signature to PCa samples with known histology, we confirmed its usefulness in differentiating NEPCa from AdPCa. We then used this 12-gene signature to identify NEPCa samples from publicly available datasets. Additionally, using gene expression profiles, we identified the differentially expressed genes in NEPCa vs AdPCa and conducted PPI network, GO, and IPA analyses. These analyses indicated an enrichment of neuronal pathways in NEPCa.

Additionally, PPIs network and GO analyses of DE genes provide important insight for the identification of key modules and hub genes involved in the pathogenesis of NEPCa. For example, genes in module 1 (**Figure 6A**) are related to cell cycle regulation. Our analyses indicated that the transcription of this group of genes can be regulated by E2Fs, a family of transcription factors that control cell cycle progression [36]. An important downstream target of E2Fs is EZH2, a critical component of chromatin modification complex [27]. Previous studies have shown that the expression of EZH2 increases in NEPCa [21] and that E2Fs can directly regulate EZH2 expression [37, 38]. The identification of the E2F module in our study provides another piece of evidence supporting the emerging critical role of the E2F/EZH2 axis in NEPCa progression.

Our analyses also unearthed some novel pathways involved in NEPCa. For example, our analysis revealed a decrease in the expression of multiple members of the MHC class II family in NEPCa (**Figure 6C**). The MHC class II members are primarily expressed in antigen-presenting cells [39], which function to initiate immune responses. The decreased expression of these genes may reflect a suppressed immune response in NEPCa. This provides a rationale for testing whether restoring immune function can be an effective treatment for NEPCa, even though the overall successful rate of cancer immunotherapy is low in PCa.

Moreover, we found an interesting switch in the expression of Homeobox (HOX) genes. HOXs are master transcription factors that regulate cell fate determination and organogenesis during early embryo development. HOXB13 is the most posterior HOX gene, highly expressed in prostate [40]. It has been suggested that the expression of HOXB13 can be used as a marker for the identification of the prostate origin of metastatic cancer cells [40]. We found that the HOXB13 expression level decreased in NEPCa. In contrast, HOXB3 and HOXB5-7, the neuron-related HOXs, exhibited increased expression levels in NEPCa. This is in line with the hypothesis of a linage switch from prostatic epithelia to neuronal cells.

In summary, we analyzed publicly available datasets and identified the differentially expressed genes and key players in NEPCa development. The re-organized datasets that contain NEPCa samples could facilitate further studies on NEPCa and lead to the identification of new therapeutic targets.

Disclosure of conflict of interest

None.

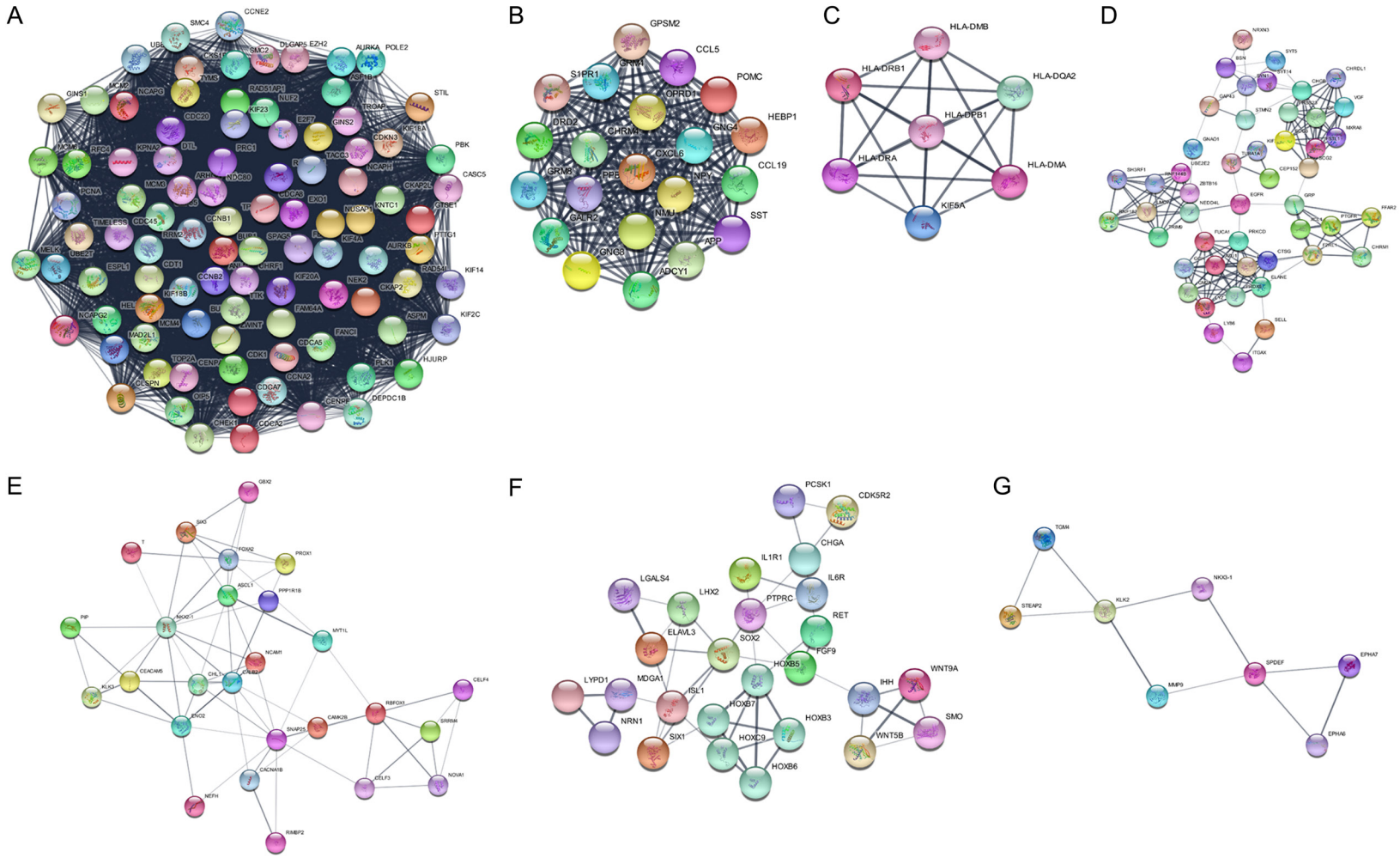

Figure 6. PPIs subnetwork analysis. (A) Cell cycle related (B) GPCR signaling related (C) MHC class II family (D) cell secretion and exocytosis related (E) neuron differentiation related (F) stem cell pluripotency related and (G) epithelium development related PPIs networks.

Address correspondence to: Xiuping Yu, Department of Biochemistry and Molecular Biology, LSU Health-Shreveport, Shreveport, USA. Tel: 615-775-6692; E-mail: xyu@lsuhsc.edu

References

[1] Siegel RL, Miller KD and Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019; 69: 7-34.

[2] Saad F and Hotte SJ. Guidelines for the management of castrate-resistant prostate cancer. Can Urol Assoc J 2010; 4: 380-384.

[3] Society AC. Survival rates for prostate cancer. 2019; https://www.cancer.org/cancer/prostate-cancer/detection-diagnosis-staging/survival-rates.html#written_by. Accessed 7-24-2019, 2019.

[4] Beltran H, Tomlins S, Aparicio A, Arora V, Rickman D, Ayala G, Huang J, True L, Gleave ME, Soule H, Logothetis C and Rubin MA. Aggressive variants of castration-resistant prostate cancer. Clin Cancer Res 2014; 20: 2846-2850.

[5] Edgar R, Domrachev M and Lash AE. Gene expression omnibus: ncbi gene expression and hybridization array data repository. Nucleic Acids Res 2002; 30: 207-210.

[6] Cerami E, Gao J, Dogrusoz U, Gross BE, Onur Sumer S, Arman Aksoy B, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C and Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012; 2: 401-404.

[7] Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER 3rd, Barretina J, Gelfand ET, Bielski CM, Li H, Hu K, Andreev-Drakhlin AY, Kim J, Hess JM, Haas BJ, Aguet F, Weir BA, Rothberg MV, Paolella BR, Lawrence MS, Akbani R, Lu Y, Tiv HL, Gokhale PC, de Weck A, Mansour AA, Oh C, Shih J, Hadi K, Rosen Y, Bistline J, Venkatesan K, Reddy A, Sonkin D, Liu M, Lehar J, Korn JM, Porter DA, Jones MD, Golji J, Caponigro G, Taylor JE, Dunning CM, Creech AL, Warren AC, McFarland JM, Zamanighomi M, Kauffmann A, Stransky N, Imielinski M, Maruvka YE, Cherniack AD, Tsherniak A, Vazquez F, Jaffe JD, Lane AA, Weinstock DM, Johannessen CM, Morrissey MP, Stegmeier F, Schlegel R, Hahn WC, Getz G, Mills GB, Boehm JS, Golub TR, Garraway LA and Sellers WR. Next-generation characterization of the cancer cell line encyclopedia. Nature 2019; 569: 503.

[8] Team R. RStudio: integrated development environment for R2019.

[9] Kolde R. Pheatmap: pretty heatmaps. R Package Version 2012; 61: 915.

[10] Robinson MD, McCarthy DJ and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010; 26: 139-140.

[11] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43: e47.

[12] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.

[13] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ and von Mering C. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47: D607-D613.

[14] Bader GD and Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003; 4: 2.

[15] Ge S and Jung D. ShinyGO: a graphical enrichment tool for animals and plants. bioRxiv 2018: 315150.

[16] Krämer A, Green J, Pollard J Jr and Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics 2014; 30: 523-530.

[17] Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, Marotz C, Giannopoulou E, Chakravarthi BV, Varambally S, Tomlins SA, Nanus DM, Tagawa ST, Van Allen EM, Elemento O, Sboner A, Garraway LA, Rubin MA and Demichelis F. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat Med 2016; 22: 298-305.

[18] Zhang X, Coleman IM, Brown LG, True LD, Kollath L, Lucas JM, Lam HM, Dumpit R, Corey E, Chéry L, Lakely B, Higano CS, Montgomery B, Roudier M, Lange PH, Nelson PS, Vessella RL and Morrissey C. SRRM4 expression and the loss of REST activity may promote the emergence of the neuroendocrine phenotype in castration-resistant prostate cancer. Clin Cancer Res 2015; 21: 4698-4708.

[19] Mirosevich J, Gao N, Gupta A, Shappell SB, Jove R and Matusik RJ. Expression and role of Foxa proteins in prostate cancer. Prostate 2006; 66: 1013-1028.

[20] Xin Z, Zhang Y, Jiang Z, Zhao L, Fan L, Wang Y, Xie S, Shangguan X, Zhu Y, Pan J, Liu Q, Huang Y, Dong B and Xue W. Insulinoma-associated protein 1 is a novel sensitive and specific marker for small cell carcinoma of the prostate. Hum Pathol 2018; 79: 151-159.

[21] Clermont PL, Lin D, Crea F, Wu R, Xue H, Wang Y, Thu KL, Lam WL, Collins CC, Wang Y and Helgason CD. Polycomb-mediated silencing in

neuroendocrine prostate cancer. Clin Epigenetics 2015; 7: 40.
[22] Rapa I, Ceppi P, Bollito E, Rosas R, Cappia S, Bacillo E, Porpiglia F, Berruti A, Papotti M and Volante M. Human ASH1 expression in prostate cancer with neuroendocrine differentiation. Mod Pathol 2008; 21: 700-707.
[23] Yao JL, Madeb R, Bourne P, Lei J, Yang X, Tickoo S, Liu Z, Tan D, Cheng L, Hatem F, Huang J and Anthony di Sant'Agnese P. Small cell carcinoma of the prostate: an immunohistochemical study. Am J Surg Pathol 2006; 30: 705-712.
[24] Yu X, Cates JM, Morrissey C, You C, Grabowska MM, Zhang J, DeGraff DJ, Strand DW, Franco OE, Lin-Tsai O, Hayward SW and Matusik RJ. SOX2 expression in the developing, adult, as well as, diseased prostate. Prostate Cancer Prostatic Dis 2014; 17: 301-309.
[25] Beltran H, Tomlins S, Aparicio A, Arora V, Rickman D, Ayala G, Huang J, True L, Gleave ME, Soule H, Logothetis C and Rubin MA. Aggressive variants of castration-resistant prostate cancer. Clin Cancer Res 2014; 20: 2846-2850.
[26] Li Y, Donmez N, Sahinalp C, Xie N, Wang Y, Xue H, Mo F, Beltran H, Gleave M, Wang Y, Collins C and Dong X. SRRM4 drives neuroendocrine transdifferentiation of prostate adenocarcinoma under androgen receptor pathway inhibition. Eur Urol 2017; 71: 68-78.
[27] Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS and Zhang Y. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 2002; 298: 1039-1043.
[28] Tang BL. The potential of targeting brain pathology with Ascl1/Mash1. Cells 2017; 6: 26.
[29] Sancho-Martinez I, Baek SH and Izpisua Belmonte JC. Lineage conversion methodologies meet the reprogramming toolbox. Nat Cell Biol 2012; 14: 892-899.
[30] Li Y, Zhang Q, Lovnicki J, Chen R, Fazli L, Wang Y, Gleave M, Huang J and Dong X. SRRM4 gene expression correlates with neuroendocrine prostate cancer. Prostate 2019; 79: 96-104.
[31] Tzelepi V, Zhang J, Lu JF, Kleb B, Wu G, Wan X, Hoang A, Efstathiou E, Sircar K, Navone NM, Troncoso P, Liang S, Logothetis CJ, Maity SN and Aparicio A. Modeling a lethal prostate cancer variant with small-cell carcinoma features. Clin Cancer Res 2012; 18: 666-677.
[32] Lin D, Dong X, Wang K, Wyatt AW, Crea F, Xue H, Wang Y, Wu R, Bell RH, Haegert A, Brahmbhatt S, Hurtado-Coll A, Gout PW, Fazli L, Gleave ME, Collins CC and Wang Y. Identification of DEK as a potential therapeutic target for neuroendocrine prostate cancer. Oncotarget 2015; 6: 1806-1820.
[33] Dan R, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, Beltran H, Abida WM, Bradley RK, Vinson J, Cao X, Vats P, Kunju LP, Hussain M, Feng FY, Tomlins SA, Cooney KA, Smith DC, Brennan C, Siddiqui J, Mehra R, Chen Y, Rathkopf DE, Morris MJ, Solomon SB, Durack JC, Reuter VE, Gopalan A, Gao J, Loda M, Lis RT, Bowden M, Balk SP, Gaviola G, Sougnez C, Gupta M, Yu EY, Mostaghel EA, Cheng HH, Mulcahy H, True LD, Plymate SR, Dvinge H, Ferraldeschi R, Flohr P, Miranda S, Zafeiriou Z, Tunariu N, Mateo J, Perez Lopez R, Demichelis F, Robinson BD, Schiffman MA, Nanus DM, Tagawa ST, Sigaras A, Eng KW, Elemento O, Sboner A, Heath EI, Scher HI, Pienta KJ, Kantoff P, de Bono JS, Rubin MA, Nelson PS, Garraway LA, Sawyers CL and Chinnaiyan AM. Integrative clinical genomics of advanced prostate cancer. Cell 2015; 161: 1215-1228.
[34] Varambally S, Yu J, Laxman B, Rhodes DR, Mehra R, Tomlins SA, Shah RB, Chandran U, Monzon FA, Becich MJ, Wei JT, Pienta KJ, Ghosh D, Rubin MA and Chinnaiyan AM. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. Cancer Cell 2005; 8: 393-406.
[35] Ku SY, Rosario S, Wang Y, Mu P, Seshadri M, Goodrich ZW, Goodrich MM, Labbé DP, Gomez EC, Wang J, Long HW, Xu B, Brown M, Loda M, Sawyers CL, Ellis L and Goodrich DW. Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. Science 2017; 355: 78-83.
[36] Bracken AP, Pasini D, Capra M, Prosperini E, Colli E and Helin K. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. EMBO J 2003; 22: 5323-5335.
[37] Conway E, Healy E and Bracken AP. PRC2 mediated H3K27 methylations in cellular identity and cancer. Curr Opin Cell Biol 2015; 37: 42-48.
[38] Coe BP, Thu KL, Aviel-Ronen S, Vucic EA, Gazdar AF, Lam S, Tsao MS and Lam WL. Genomic deregulation of the E2F/Rb pathway leads to activation of the oncogene EZH2 in small cell lung cancer. PLoS One 2013; 8: e71670.
[39] Jones EY, Fugger L, Strominger JL and Siebold C. MHC class II proteins and disease: a structural perspective. Nat Rev Immunol 2006; 6: 271-282.
[40] Brechka H, Bhanvadia RR, VanOpstall C and Vander Griend DJ. HOXB13 mutations and binding partners in prostate development and cancer: Function, clinical significance, and future directions. Genes Dis 2017; 4: 75-87.