Nanopore Long-Read Transcriptomics Profiling Reveals Gene Expression Signatures of Mouse Tumor Endothelial Cells 2H-11

Yuanyuan Tian
Jilin University
Jiao Zhao
Jilin University
Ju Huang
Jilin University
Haiying Zhang
Jilin University
Fushun Ni
Jilin University
Rong Tang
Jilin University
Jing Shao
Jilin University
Yilei Li
Jilin University
Wei Li (liwei2006@jlu.edu.cn)
Jilin University

Research Article

Keywords: Tumor endothelial cells(TECs), Transcriptomic, Differentially expressed genes (DEGs), Signaling pathway

Posted Date: December 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1129675/v1

Abstract

Background:Tumor endothelial cells (TECs) play an indispensable role in tumor growth and metastasis. Compared with normal endothelial cells (NECs), TECs exhibit unique phenotypic and functional heterogeneity in terms of metabolism, genetics, and transcriptomics. It is not only the key to coordinate tumor angiogenesis, but also an important factor of immune regulation in the tumor microenvironment. In recent years, the role of TECs in tumor metabolism and invasion has been continuously reported. However, the research on the mechanism behind the complex functions of TECs is still at the basic stage. We use Oxford Nanopore Technology (ONT) three-generation full-length transcriptome sequencing to detect all genetic structural changes in the transcriptome of mouse TECs 2H-11 and mouse NECs SVEC4-10.

Results: In Tumor endothelial cells 2H-11,1847genes are up-regulated and 1202 genes are down-regulated. According to the Gene ontology (GO) enrichment analysis of differentially expressed genes (DEGs), we found that different functional trends related to metabolic processes, developmental processes, localization, immune system processes, and locomotion are the main reasons for the differences. DEGs are mainly enriched in signal pathways related to cancer, immunity and metabolism, involving Pathways in cancer,Antigen processing and presentation , Proteoglycans in cancer, Focal adhesion, MAPK signaling pathway ,Protein digestion and absorption、ECM-receptor interaction、PI3K-Akt signaling pathway and Glutathione metabolism. We also obtained the structural variation of transcripts such as alternative splicing, gene fusion, and alternative polyadenylation and accurately quantified the expression of the transcript. Some of our results have been confirmed in other documents. But other data have not been reported yet, which is the focus of our future exploration.

Conclusion: We try to use transcriptomics and bioinformatics methods to characterize tumor endothelial cell-related genes and signaling pathways.It could help better understand the molecular mechanisms of tumor endothelial cells involved in tumorigenesis and development. DEGs in key pathways may be potential diagnostic markers or therapeutic targets of TECs. Our data also provide useful genetic resources for improving the genome and transcriptome annotations of TECs and NECs.

Background

Tumor Endothelial Cells (TECs) originate from normal vascular endothelial cells around the tumor. However, due to the long-term exposure to the tumor microenvironment(TME), the cell phenotype, function and even genes have changed, so they are different from normal endothelial cells(NECs).The TME changes, such as hypoxia and chronic growth factor stimulation, could induce a series of endothelial dysfunctions including irregular diameters, fragility, leakiness and abnormal blood flow .TECs play an indispensible role in tumor growth and metastasis[1–6]. More and more evidences show that compared with NECs, TECs exhibit unique phenotype and functional heterogeneity in metabolism, genetics and transcriptomics, making them proliferate in cells, Migration ability, gene expression profile, and response to growth factors and several drugs are all different to a certain extent [7, 8]. The functions

of TECs are multifaceted. Compared with NECs, TECs have significantly altered morphology and genetic phenotype, including chromosomal structural changes and mutations. TECs exhibit a stem cell-like origin so they are the key to coordinate tumor angiogenesis [7, 9, 10]. TECs are also recognized transcription factors associated with tumor aggressiveness and known to affect angiogenesis [11, 12].Moreover,TECs have evolved into key mediators of immune regulation in TME. The research data of Katerina Rohlenova et al. indicate that TECs may play a role in tumor immune surveillance and represents the first line of defense of immune cells in the TME [13].TECs can act as antigen-presenting cells by themselves to be associated with T cell initiation, activation and proliferation [14, 15, 16]. In addition, TECs are essential for the formation of tertiary lymphoid structure (TLS) in tumors, which has recently been related to the treatment response of checkpoint antibody therapy [17]. Katerina Rohlenova et al. studied the heterogeneity of metabolic gene expression characteristics of different endothelial cells at the single-cell level. And they found that TECs have the most obvious metabolic transcriptome diversity. The TECs phenotype expresses different metabolic transcriptome characteristics [13]. Lambrechts et al. proved that compared with NECs, TECs have a high transcriptional activity, and the RNA content increased four times, which may be caused by the high metabolic demand of nucleotide biosynthesis or glycolysis [12]. More and more evidences indicate that TECs may be involved in tumor progression and metastasis [7, 10], which implies their possible prognosis and predictive potential.

However, due to the technical difficulty of TECs isolation and the limited selection of the depth of single-cell spectrum cell populations, the genetic characterization of TECs have been at the basic stage for a long time,so that the mechanism behind the complex functions of TECs have not been studied in depth, while the current knowledge storage cannot provide strong theoretical support for TECs for the clinical cancer treatment of specific drug targets.

In recent years, with the development of high-throughput sequencing technology, the cost of sequencing has continued to decline, and transcriptome sequencing has become the main method for studying gene expression. However, due to factors such as read length and sensitivity, traditional second-generation sequencing technologies have short assembled genes and incomplete transcript structures. In addition, in eukaryotes, due to the structural variation of genes, there are events such as alternative splicing. They could produce a variety of transcripts, resulting in a relatively low accuracy of transcript reconstruction assembled by second-generation transcriptome sequencing.We use Oxford Nanopore Technology (ONT) three-generation full-length transcriptome sequencing to detect all genetic structural changes in the transcriptome of tumor endothelial cells (TECs) 2H -11 and normal endothelial cells (NECs) SVEC4-10. Nanopore sequencing is a new generation of nanopore-based single-molecule real-time electrical signal sequencing technology for full-length transcriptome sequencing. It is also known as isoform sequencing (Iso-Seq) ,which could directly obtain 5′ to 3′ high-quality full-length transcript sequences that based on the third-generation sequencing platform [18, 19, 20]. Our results accurately identify the structural variations of the transcript, such as alternative splicing, gene fusion, alternative polyadenylation, etc.we also accurately quantify the expression of the transcript (mRNA or polyA + lncRNA). We examined differentially expressed genes (DEGs), gene ontology (GO), and functional pathway analysis. We try to use transcriptomics and bioinformatics methods to characterize the signaling pathways and genes

related to tumor endothelial cells, and to better understand the molecular mechanisms of tumor endothelial cells involved in tumorigenesis and development.

Results

Data description

The long-read transcriptome data presented here were from mouse normal endothelial cells SVEC4-10 and mouse tumor endothelial cells 2H-11 culture under laboratory condition,with three samples each. (Sample H1, H2, H3, from TECs 2H-11 and Sample S1, S2, S3 from NECs SVEC4-10). Totally, 7,145,224, 5,738,625 and 8,308,634 clean reads were generated from three samples of 2H-11, with an average read length of 1186 bp、1047bp and 916 bp and an N50 of 1560bp,1419 bp and 1219 bp. 5,487,487, 4,739,190 and 21,496,454 clean reads were generated from three samples of SVEC4-10, with an average read length of 1191 bp,1188bp and 1086 bp and an N50 of 1527bp,1500 bp and 1380 bp, respectively (Table. 1).

Table 1
Overview of Nanopore all sample clean reads

Sample ID	Read Num	Base Num	N50	Mean Length	Max Length	Mean Q score
H1	7,145,224	8,475,179,085	1560	1186	28,619	Q12
H2	5,738,625	6,011,317,116	1419	1047	25,972	Q12
H3	8,308,634	7,613,609,482	1219	916	60,662	Q12
S1	5,487,487	6,537,183,635	1527	1191	487,578	Q12
S2	4,739,190	5,634,114,176	1500	1188	89,959	Q12
S3	21,496,454	23,346,764,762	1380	1086	171,171	Q12

For both 2H-11 and SVEC4-10、the length of clean reads was distributed among 1 kb~10+ kb, and the most abundant length was 1 kb (Fig. 1A and B).In addition, the average quality (Q) scores of majority of clean reads of 2H-11 and SVEC4-10 were Q13 (Fig. 2A and B). After discarding rRNA,6,905,496 , 5,509,170 , 7,939,379 , 5,312,938 , 4,535,096 and 20,822,758 clean reads were gained, and among them 89.10%, 88.57%, 88.66%, 88.96%, 87.27% and 89.23% were identified as being full-length(Table. 2).

Table 2
Summary of full-length clean reads

Sample ID	Number of Clean Read (except rRNA)	Number of full-length reads	Full-Length Percentage (FL%)
H1	6,905,496	6,152,877	89.10%
H2	5,509,170	4,879,500	88.57%
H3	7,939,379	7,038,880	88.66%
S1	5,312,938	4,726,331	88.96%
S2	4,535,096	3,957,596	87.27%
S3	20,822,758	18,581,090	89.23%

The length of full-length clean reads from 2H-11 and SVEC4-10 was distributed among 1 kb~10+ kb; the most abundant lengths for both were 1 kb (Fig. 3A and B).After removing redundant reads, the lengths of remaining full-length transcripts of both 2H-11 and SVEC4-10 were ranged from 1 kb to 9 kb, with the largest group of 1 kb (Fig. 4A and B).

DEG identification and selection

Transcripts of differentially expressed genes (DEGs) were characterized using isoform sequencing (Iso-Seq) analyses of 2H-11 cells and SVEC4-10 cells. The results revealed obvious differences between 2H-11 cells groups and the SVEC4-10 groups. We discovered 1847 up-regulated and 1202 down-regulated DEGs in the 2H-11 cells. The hierarchical clustering heat map, MA plot, and volcano plots were generated to represent the up- and down-regulated genes (logFC ±2 and $p<0.001$). Fig. 5A represents the heatmap of up- and down-regulated genes in red and green, respectively. The volcano plot (Fig. 5B) and the MA plot (Fig. 5C) visualize the differences between measurements taken in SVEC4-10 cells and 2H-11cells DEGs.

Gene annotation and functional classification of DEGs

All the DEGs were uploaded to the GO Enrichment Analysis tool and database for annotation. The biological processes (BPs), cellular components (CCs), molecular functions (MFs), and pathways were predicted in the significantly enriched GO terms of the differentially express genes (Fig. 6). The enrichment trend of some DEGs in these GO secondary functions is different from the enrichment trend of all genes. It is important to analyze whether these functions are related to the difference.

The most DEGs are involved in various BPs, such as cellular process (GO:0009987), single-organism process (GO:0044699), metabolic process (GO:0008152), biological regulation (GO:0065007), response to stimulus (GO:0050896), multicellular organismal process (GO:0032501), signaling (GO:0023052), cellular component organization or biogenesis (GO:0071840), developmental process (GO:0032502) and localization (GO:0051179) functions. Among these BPs, some DEGs are involved in multi-organism process (GO:0051704; 191genes), immune system process (GO:0002376; 214genes), locomotion

(GO:0040011; 161genes) and biological adhesion (GO:0022610; 114 genes), growth (GO:0040007; 61genes) and behavior (GO:0007610; 58genes).The enrichment trend of these DEGs in these GO secondary functions is different from the enrichment trend of all genes ,so these functions are related to differences.

The most DEGs are involved in various CCs, such as cell (GO:0005623), cell part (GO:0044464), organelle (GO:0043226), membrane (GO:0016020) ,membrane part (GO:0044425), organelle part (GO:0044422), macromolecular complex (GO:0032991) and membrane-enclosed lumen (GO:0031974) functions.Among these CCs, these DEGs are involved in the inextracellular region (GO:0005576; 225genes), extracellular region part (GO:0044421; 169genes), synapse (GO:0045202; 47genes), cell junction (GO:0030054; 65genes) and synapse part (GO:0044456; 31genes).The enrichment trend of these DEGs in these GO secondary functions is different from the enrichment trend of all genes, so these functions are related to differences.

The most DEGs are involved in various MFs, such as binding (GO:0005488) and catalytic activity (GO:0003824) functions. Some DGEs are involved in the antioxidant activity (GO:0016209; 13genes), chemoattractant activity (GO:0042056; 7genes), and chemorepellent activity (GO:0045499; 8genes).The enrichment trend of these DEGs in these GO secondary functions is different from the enrichment trend of all genes so these functions are related to differences.

Pathway analysis

Pathway analysis helps elucidate data from canonical prior structured knowledge in the form of pathways. It allows finding distinct cell processes, diseases, or signaling pathways that are statistically associated with the selection of DEGs. It is always used to analyze whether the DEGs are significantly different in a certain pathway (over-presentation). Pathway significant enrichment analysis takes pathway in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database as the unit, and applies hypergeometric test to find pathways that are significantly enriched in DEGs compared with the entire gene background. The results of enrichment analysis of the DEGs are shown in the Fig. 7 The Fig. 7A shows the top 20 pathways with the smallest significant Q value.The Fig. 7B indicates the network view of DEGs pathways in 2H-11 and SVEC4-10.

They are involved in various KEGG pathways, such as the Renin secretion (ko04924; 39 genes), Pancreatic secretion(ko04972; 40 genes), Fluid shear stress and atherosclerosis (ko05418; 37 genes), Human papillomavirus infection (ko05165; 73 genes), Rap1 signaling pathway (ko04015; 44 genes), Pathways in cancer (ko05200; 88 genes), Antigen processing and presentation(ko04612; 26 genes),Proteoglycans in cancer (ko05205; 40 genes), Focal adhesion (ko04510 41 genes), MAPK signaling pathway (ko04010; 52 genes), Protein digestion and absorption□ko04974; 23 genes),

ECM-receptor interaction(ko04512; 24 genes), Human cytomegalovirus infection□ko05163; 46 genes□, Type I diabetes mellitus□ko04940; 19 genes□, Phagosome□ko04145; 39 genes□, PI3K-Akt signaling

pathway□ko04151; 70 genes□, Allograft rejection□ko05330; 18 genes□, Glutathione metabolism□ko00480; 17 genes□, Graft-versus-host disease□ko05332; 18 genes) and Toxoplasmosis□ko05145; 22 genes).

Here, we mainly show the signal pathways of some genes related to our research. There are a large number of DEGs enriched in Pathways in cancer (ko05200; 88 genes; Table. 3) and PI3K-Akt signaling pathway (ko04151; 70 genes; Table. 4). A relatively large number of DEGs are enriched in Proteoglycans in cancer (ko05205; 40 genes; Table. 5), Focal adhesion (ko04510; 41 genes; Table. 6) and MAPK signaling pathway (ko04010; 52 genes; Table. 7). There is also a small number of genes enriched in Antigen processing and presentation (ko04612; 26 genes; Table. 8), Protein digestion and absorption (ko04974; 23genes; Table. 9), ECM-receptor interaction (ko04512; 24genes; Table. 10**)** and Glutathione metabolism (ko00480; 17 genes; Table. 11).

Table 8
Genes involved Antigen processing and presentation pathway. (p-value = 0.00011321866609603849; rich factor =2.1770077367)

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000035929	H2-Q4	0.004570336	-1.537433763	down
ENSMUSG00000024401	Tnf	9.88E-09	4.921767952	up
ONT.10206	ONT.10206	1.16E-07	-3.858448298	down
ENSMUSG00000024308	Tapbp	5.22E-27	-2.26916257	down
ONT.21272	ONT.21272	0.007145755	1.664993518	up
ENSMUSG00000073411	H2-D1	1.30E-15	-1.634302503	down
ENSMUSG00000060550	H2-Q7	8.82E-05	-1.084974005	down
ENSMUSG00000006611	Hfe	0.003328646	1.733234191	up
ENSMUSG00000059970	Hspa2	0.000972136	-1.211418313	down
ENSMUSG00000021477	Ctsl	2.43E-25	-1.747916795	down
ENSMUSG00000090877	Hspa1b	5.02E-10	3.433264195	up
ENSMUSG00000022216	Psme1	1.10E-29	-2.01144189	down
ENSMUSG00000067212	H2-T23	5.69E-24	-2.877428307	down
ENSMUSG00000037649	H2-DMa	3.11E-09	2.724576686	up
ENSMUSG00000037321	Tap1	7.27E-07	-1.763033392	down
ENSMUSG00000091971	Hspa1a	3.55E-12	5.290791162	up
ENSMUSG00000053835	H2-T24	0.000633033	-2.850276057	down
ENSMUSG00000060802	B2m	1.27E-23	-1.720470365	down
ENSMUSG00000061232	H2-K1	3.87E-26	-2.025866713	down
ENSMUSG00000092243	Gm7030	0.001190601	-1.912033582	down
ONT.6732	ONT.6732	1.82E-05	-2.524698523	down
ENSMUSG00000079197	Psme2	1.21E-13	-1.611409427	down
ENSMUSG00000073402	Gm8909	0.000921921	-2.408210504	down
ONT.22061	ONT.22061	8.03E-08	3.399291437	up
ENSMUSG00000056116	H2-T22	1.63E-23	-1.966873441	down

Table 9
Genes involved Protein digestion and absorption pathway. (p-value = 0.000288132612123349; rich factor = 2.17062146892655)

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000048126	Col6a3	9.48E-34	8.237114364	up
ENSMUSG00000004098	Col5a3	7.53E-06	2.343294368	up
ENSMUSG00000001435	Col18a1	1.90E-09	3.471710732	up
ENSMUSG00000031273	Col4a6	5.24E-16	-3.063792325	down
ENSMUSG00000029163	Emilin1	1.05E-13	5.538029232	up
ENSMUSG00000056174	Col8a2	4.13E-05	-3.614177874	down
ENSMUSG00000032332	Col12a1	0.009306321	1.305619343	up
ENSMUSG00000001918	Slc1a5	2.74E-07	1.266888836	up
ENSMUSG00000054342	Kcnn4	0.001042811	2.792801361	up
ENSMUSG00000027966	Col11a1	9.47E-09	4.925101847	up
ENSMUSG00000026576	Atp1b1	2.01E-15	-2.7595007	down
ENSMUSG00000026837	Col5a1	1.22E-06	1.149962411	up
ENSMUSG00000027820	Mme	1.98E-07	-4.569286255	down
ENSMUSG00000040907	Atp1a3	1.23E-21	5.920207798	up
ENSMUSG00000031274	Col4a5	4.62E-06	-1.233139489	down
ENSMUSG00000029661	Col1a2	1.27E-06	1.30884117	up
ENSMUSG00000001119	Col6a1	1.36E-10	-1.735704123	down
ENSMUSG00000025650	Col7a1	4.29E-06	2.572385689	up
ENSMUSG00000020241	Col6a2	7.99E-12	-2.197425083	down
ENSMUSG00000031502	Col4a1	3.11E-11	-1.438161228	down
ENSMUSG00000024053	Emilin2	4.99E-17	5.955417654	up
ENSMUSG00000068196	Col8a1	2.15E-15	2.664904277	up
ENSMUSG00000000958	Slc7a7	0.000135723	3.515025572	up

Table 10
Genes involved ECM-receptor interaction pathway. (p-value = 0.000348331597598506; rich factor = 2.10448476549127)

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000048126	Col6a3	9.48E-34	8.237114364	up
ENSMUSG00000041936	Agrn	2.72E-06	-1.629566318	down
ENSMUSG00000028364	Tnc	4.22E-12	2.223658118	up
ENSMUSG00000023885	Thbs2	3.24E-31	8.006758795	up
ENSMUSG00000001281	Itgb7	7.71E-05	2.925093206	up
ENSMUSG00000038486	Sv2a	0.001262385	-1.707618944	down
ENSMUSG00000027111	Itga6	3.34E-07	-1.825360992	down
ENSMUSG00000015647	Lama5	3.23E-16	-2.641045482	down
ENSMUSG00000031273	Col4a6	5.24E-16	-3.063792325	down
ENSMUSG00000001507	Itga3	6.06E-10	-1.777849405	down
ENSMUSG00000040998	Npnt	4.08E-08	-3.24447338	down
ENSMUSG00000034687	Fras1	3.50E-09	-4.652846498	down
ENSMUSG00000017009	Sdc4	6.03E-09	-1.518175322	down
ENSMUSG00000031274	Col4a5	4.62E-06	-1.233139489	down
ENSMUSG00000029661	Col1a2	1.27E-06	1.30884117	up
ENSMUSG00000028047	Thbs3	2.46E-10	2.24046441	up
ENSMUSG00000001119	Col6a1	1.36E-10	-1.735704123	down
ENSMUSG00000005087	Cd44	0.000121501	1.123040059	up
ENSMUSG00000042284	Itga1	1.01E-07	-2.184817206	down
ENSMUSG00000022817	Itgb5	3.96E-08	2.06121516	up
ENSMUSG00000020241	Col6a2	7.99E-12	-2.197425083	down
ENSMUSG00000031502	Col4a1	3.11E-11	-1.438161228	down
ENSMUSG00000039952	Dag1	1.18E-05	-1.128763692	down
ENSMUSG00000029304	Spp1	4.87E-133	5.690870707	up

Table 11
Genes involved in Glutathione metabolism pathway. (p-value = 0.000621008; rich factor = 2.366449275)

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000018339	Gpx3	6.25E-15	6.07451606	up
ENSMUSG00000032348	Gsta4	3.32E-39	-2.885542505	down
ENSMUSG00000058135	Gstm1	4.63E-17	2.362659014	up
ENSMUSG00000074183	Gsta1	0.001087574	2.431146261	up
ENSMUSG00000038155	Gstp2	5.86E-09	2.328303859	up
ENSMUSG00000008540	Mgst1	1.70E-12	-1.768900074	down
ENSMUSG00000074604	Mgst2	2.62E-05	2.166938591	up
ENSMUSG00000039062	Anpep	2.81E-05	3.75229385	up
ENSMUSG00000028597	Gpx7	8.31E-74	4.851252354	up
ENSMUSG00000026688	Mgst3	3.90E-17	2.527299541	up
ENSMUSG00000001663	Gstt1	1.06E-12	2.740719466	up
ENSMUSG00000028961	Pgd	3.43E-07	1.073122031	up
ENSMUSG00000029864	Gstk1	1.31E-15	5.024276188	up
ENSMUSG00000027603	Ggt7	5.36E-05	1.792209318	up
ENSMUSG00000025934	Gsta3	8.73E-17	-6.331940412	down
ENSMUSG00000011179	Odc1	9.84E-12	-1.805655481	down
ENSMUSG00000027890	Gstm4	0.006745519	1.774064896	up

Gene Set Enrichment Analysis(GSEA)

GSEA results showed that the differentially expressed gene set was enriched for some functional gene networks that are clearly associated with endothelial cells, such as neuron differentiation (13 genes), bicellular tight junction (10 genes), andapical plasma membrane (22 genes), filopodium (8 genes), double-stranded RNA binding (10 genes), RNA-DNA hybrid ribonuclease activity (19 genes) and guanyl nucleotide binding (31 genes), as well as some novel signal pathways such as Ribosome biogenesis in eukaryotes (40 genes), Cytosolic DNA-sensing pathway (17 genes) ,Renin secretion (32 genes) and Pancreatic secretion (32 genes) (Fig. 8).

Alternative splicing profiles

The main types of gene alternative splicing (AS) are as follows: (A) Exon skipping; (B) Alternative 3'splicing site; (C) Mutually exclusive exon; (D) Alternative 5'splice site; (E) Intron retention. From the

analysis results of the Astalavista software, We collected statistics on the occurrence of the above 5 alternative splicing events in the transcripts of 6 samples, and obtained a statistical graph of the number of predicted alternative splicing events in each sample.(Fig. 9) After analysing the data of 2H-11 and SVEC4-10, we detected Exon skipping was the most frequent AS events and Mutually exclusive exon was the rarest AS events. The prevalence of different AS events was similar between 2H-11 and SVEC4-10, indicating relevant pathogenesis of these two types of cells. It is worthy that one gene might possess several alternative splicing patterns.For each sample and every possible splice event, percent-splice-in (PSI) value was calculated, which is the ratio of normalized read counts indicating inclusion of a transcript element over the total normalized reads for that event (both inclusion and exclusion reads). Because the number of samples>=4, PSI-Sigma uses delta PSI>0.1 and Pvalue<0.01 to filter by default to get the filtered differential alternative splicing events.(Table. 12) Changes in average PSI values when comparing groups of samples mean a shift in splicing patterns between the groups or a splice event.

Table 12
Differentially spliced genes (DSGs) between 2H-11 and SVEC4-10.

GeneID	Target_Exon	Event_Type	ΔPSI (%)	T-test_p-value	FDR(BH)
Gm20425	9:103189546-103189594	A3SS	93.96	2.76E-08	3.29E-07
H2-K1	17:33996477-33996503	A3SS	75.73	3.38E-05	5.31E-05
Zscan21	5:138117836-138117920	SES	62.35	2.07E-05	6.94E-05
Gm20425	9:103189594-103189644	TSS|A5SS	61.62	6.72E-05	7.99E-04
Gm20425	9:103189594-103189642	TSS|A5SS	61.62	6.72E-05	7.99E-04
Gm20425	9:103189594-103189595	A5SS	61.62	6.72E-05	7.99E-04
Gm20425	9:103189546-103189594	A3SS	60.74	3.22E-06	3.83E-05
Npm1	11:33157037-33157048	A3SS	55.35	7.34E-08	8.22E-08
Arl14ep	2:106969298-106969558	A3SS	42.95	1.40E-03	2.81E-03
H2-K1	17:33996477-33996503	TSS|A3SS	41.31	1.70E-06	2.68E-06
Ube2t	1:134962610-134962786	TSS|A5SS	38.81	6.13E-04	1.11E-03
Eif4a2	16:23112351-23113166	TSS|A3SS	37.09	1.76E-03	2.59E-03
Fancl	11:26403322-26403378	SES	35.54	2.95E-03	3.29E-03
Mpnd	17:56009330-56009586	SES	34.69	1.26E-03	2.02E-03
Nmrk1	19:18632000-18632081	TSS|A5SS	34.09	1.84E-04	3.13E-04
Mtg1	7:140138380-140138380	A5SS	33.03	1.85E-04	9.90E-04
4833420G17Rik	13:119472139-119472166	A3SS	32.56	5.86E-03	7.24E-03
Apol9b	15:77733626-77733796	SES	32.1	3.58E-05	5.03E-05
Flywch1	17:23770235-23770332	SES	31.53	7.81E-03	1.19E-02
Eif4a2	16:23112351-23112461	SES	30.77	1.76E-03	2.59E-03
Target Exon: Genomic coordinates of the alternative exon.					
Event Type: Category of the splicing event					
A3SS: Alternative 3'splice site. A5SS: Alternative 5's plice site.					
SES: Single exon skipping. TSS:Transcription start site(Alternative 5'first site)					
ΔPSI (%): the average difference of PSI values in group 2H-11 and group SVEC4-10					

T-test p-value: p-value derived from two-sample t-test.

FDR (BH): false discovery rate based on the p-values.

Differentially spliced genes (DSGs) between 2H-11 and SVEC4-10 were analysed and the top 20 significantly altered AS events were summarized in Table. 12. A3SS(Alternative 3’splice site) and SES(Single exon skipping) were dominant AS types, and several genes (Gm20425, H2-K1, Zscan21, Npm1) demonstrated two AS events with opposite preference in 2H-11 and SVEC4-10. It is obvious that most AS patterns were more active in TECs than NECs. Moreover, the results demonstrated significant difference of AS patterns among different stages of 2H-11 and SVEC4-10 (all FDR<0.001), which indicate that certain alternative splicing patterns might show significant differences between TECs and NECs.

Alternative polyadenylation (APA) analysis

Polyadenylation refers to the covalent linkage of polyadenylic acid to messenger RNA (mRNA) molecules. In the process of protein biosynthesis, this is part of the way to produce mature mRNA ready for translation. In eukaryotes, polyadenylation is a mechanism that interrupts mRNA molecules at their 3'ends. The polyadenylic acid tail (or poly A tail) protects mRNA from exonuclease attack.And it is very important for the termination of transcription, export of mRNA from the nucleus and translation. The alternative polyadenylation (APA) of precursor mRNA may contribute to the diversity of the transcriptome and the coding ability of the genome and the regulatory mechanism of genes. We use TAPIS pipeline to identify APA.[27] The APA identified by each sample is shown in Fig. 10A and B.Moreover,we use DREME to analyze the sequence of 50bp upstream of the polyA site of all transcripts and the identified motif is shown in Fig. 10C and D.

Find fusion transcript

Use the consensus sequence before de-redundancy to screen each sample for fusion transcripts according to the following conditions:(1) must map to 2 or more loci; (2) minimum coverage for each loci is 5% and minimum coverage in bp is >= 1 bp; (3) The total length of all loci compared must account for more than 95% of the total length of the transcript; (4) The distance between the two loci must be more than 10kb.The fusion transcripts obtained from 6 samples ranged from 33 to 108.

Sequence prediction of coding region of new gene

The TransDecoder (v3.0.0) software is based on the length of the Open Reading Frame (ORF), the log-likelihood score, the comparison of the amino acid sequence and the protein domain sequence of the Pfam database, etc. It can identify reliable potential coding sequence (Coding Sequence, CDS) from the transcript sequence. Use TransDecoder software to predict the coding region sequence and its corresponding amino acid sequence of the new transcript obtained. A total of 19,021 orf were obtained this time, of which 9,954 were complete orf. The predicted sequence length distribution of the entire ORF region encoding protein is shown in the Fig. 11.

Transcription factor analysis

Transcription factors are proteins that can bind to a specific nucleotide sequence upstream of a gene. These proteins can regulate the binding of RNA polymerase to the DNA template, thereby regulating gene transcription. Animal transcription factor identification uses the animal transcription factor database-animalTFDB 3.0 [12]. This project predicts a total of 2,047 transcription factors from the new transcripts obtained. We only show the top 20 transcription factors Family information .Statistics on the number of different types of transcription factors are shown in Fig. 12.

long noncoding RNA (LncRNA) prediction

Because lncRNA does not encode protein, it is possible to determine whether the transcript is an lncRNA by screening the transcript for coding potential. Four methods of CPC[13] analysis, CNCI[14] analysis, CPAT[15], pfam[16] protein domain analysis were used to predict the newly discovered transcripts for lncRNA. A total of 954 lncRNA transcripts were predicted by the four methods. In order to visually display the analysis results, the noncding transcripts identified by the above 4 analysis softwares are added to the 4 analysis results to take the intersection. Then according to the position of the lncRNA on the reference genome annotation information (gff), the lncRNA is classified as shown in Fig. 13.

Discussion

In recent years, a few studies have used popular high-throughput gene expression technologies, microarrays and serial analysis of gene expression (SAGE) [34–39] to have a more complete understanding of individual TECs functions. The detection method of analyzing the entire gene expression profile or the global transcriptome at the single-cell level has gradually attracted attention [40–43]. In the field of tumor angiogenesis, some single-cell RNA sequencing (scRNA-seq) studies have reported a descriptive list of previously known TECs phenotypes [44, 45]. Transcription Omics studies revealed only a limited number of ECs (endothelial cells) phenotypes [46–49]. Powerful single-cell transcriptomics studies can now characterize EC phenotypes in more detail and can modify traditional views on tumor endothelium.We use the ONT three-generation full-length transcriptome to detect all gene structural changes in the transcriptome.Moreover, it accurately identify structural variations in transcripts such as alternative splicing, gene fusion, selective polyadenylation and other transcripts (mRNA or polyA+ lncRNA) expression was accurately quantified. Differentially expressed gene sets were obtained and differential analysis was performed, including functional annotation and pathway analysis.

We provide comprehensive transcriptomic analysis of mouse tumor endothelial cell 2H-11 and mouse normal endothelial cell SVEC4-10 data sets. This method can provide gene expression profiles for endothelial cell phenotype changes caused by various factors in the tumor microenvironment. We screened a total of 3049 differentially expressed genes (DEGs) in TECs 2H-11 and NECs SVEC4-10. Cell metabolism, immune system processes, cell development and positioning, biological adhesion and other biological functions have different trends in DEGs and all expressed genes, suggesting that these functions may be the main reason for the differences. These functions are reflected in a wealth of signaling pathways, including Pathways in cancer, PI3K-Akt signaling pathway, Focal adhesion, MAPK signaling pathway, Antigen processing and presentation, ECM-receptor interaction and Glutathione

metabolism. It is comprehensively found that TECs play a vital role in promoting tumor progression in various ways such as induceing tumor angiogenesis,changing cell metabolism and participating in immune processes.

The literature reported the characteristic gene expression profile of breast cancer-derived ECs at the single-cell level, indicating that extracellular matrix (ECM) related genes play a key role in tumor-derived ECs. And it has been determined that some of these genes can be used as general TECs biomarkers. In addition, their DEGs are also enriched in extracellular matrix metabolism, vascular smooth muscle, drug metabolism, cancer pathways, and axon guidance related signaling pathways [50].

Our research also detected that DEGs are enriched in signal pathways such as ECM-receptor interaction, Glutathione metabolism, and Pathways in cancer. These results broaden the different roles of DEGs of TECs in the extracellular matrix. It also provides a new direction in cell metabolism. In cancer-related, we have obtained 88 differential genes. These genes can be screened again according to relevant experimental methods and verified through specific experiments. Exploring the mechanism behind them is the content of our next research. Cantelmo et al. found that TECs have high glucose metabolism, which can shunt intermediate products to nucleotide synthesis. And they have identified a large number of metabolic genes expression profiles as metabolic targets in TECs [51]. Immature endothelial cells and apical cells have stronger glycolytic gene characteristics. It is speculated that this may be caused by a worse nutrient deficiency environment and more abundant metabolic change signals [52].In our results, 17 DEGs are enriched in the signal pathway of glutathione metabolism. The role of glutathione metabolism in tumorigenesis and development has been extensively studied.However, there are few studies related to tumor endothelial cells. Our research gives the possibility of exploring glutathione metabolism in TECs, and guides the exploration of TECs in cell metabolism direction. The findings of Bakanovic et al. showed that ETBR signal transduction blocked the adhesion of T cells to tumor endothelium. After restoring ETBR in vitro, the adhesion of T cells was restored and the ability of T cells to home to tumors was also significantly enhanced. Up-regulation of ETBR has been observed in many cancers, including breast and ovarian cancer, and up-regulation of ETBR indicates unfavorable results [53–55]. The specific mechanism of tumor endothelial ETBR up-regulation is not yet clear, but it may be related to the process of neovascularization. In our results, a total of 26 DEGs are enriched in the signal pathway of antigen processing and transmission, which suggests that TECs can be used as an antigen presenting cell to participate in the initiation and activation of T cells. This function is likely to be involved in a certain process of up-regulation of ETBR in tumor endothelium. According to reports in the literature, BMP2 secreted by liver cancer cells promotes cell proliferation, migration, invasion and angiogenesis [56]. In our results, BMP2k, BMP1 and BMP4 were found to be significantly DEGs in 2H-11 and SVEC4-10. These genes were enriched in TGF-beta and Transcriptional misregulatio signaling pathways respectively, and participated in tumor angiogenesis and other processes. In the literature studying whether the intratumoral microenvironment of head and neck squamous cell carcinoma (HNSC) is highly immunosuppressive, it was found that the abnormal expression of TECs in HNSC. And the secreted HSPA12B can be partially absorbed by macrophages through OLR1, leading to subsequently activate the PI3K/Akt/mTOR signaling pathway and increase the expression of M2 markers [57]. In our results, there

are a total of 70 DEGs enriched in the PI3K/Akt signaling pathway. We can screen the genes related to HSPA12B by methods such as gene fusion or gene co-expression, and provide research for elucidating the crosstalk mechanism between tumor endothelial cells (TECs) and Tumour-associated macrophages (TAM). The data helps to better understand the formation of the immunosuppressive microenvironment within the tumor [57]. All of our results confirm some views in the current research field from the direction of transcriptomics, but more importantly, we have discovered more related genes and used bioinformatics methods to predict their main functional roles and participation signal pathways.

Our results provide the possibility to study tumor endothelial cells in cell metabolism, participation in immune system processes, tumor angiogenesis and other mechanisms behind the occurrence and development of tumors. Our goal is to continue to use specific experiments and bioinformatics methods to verify our results in the later stages.

More and more evidences show that abnormal alternative splicing (AS) is a common event in the occurrence and progression of cancer [58]. Abnormal pre-mRNA alternative splicing has been widely accepted as a new contributor to cancer development [59, 60]. Although many cancer-specific mRNA isotypes have been identified, our understanding of the spectrum of alternative splicing events and their functional pathways is far behind. With the rapid development of high-throughput sequencing and bioinformatics methods, it is possible to more comprehensively reveal AS in cancer. Different splicing patterns of a gene lead to multiple subtypes, which makes alternative splicing and its regulatory mechanism in cancer more complicated [61, 62]. We obtained the alternative splicing types of 2H-11 and SVEC4-10 through full-length transcriptome sequencing.The main types of gene alternative splicing events are as follows: (A) Exon skipping; (B) Alternative 3'splicing site; (C) Mutually exclusive exon; (D) Alternative 5'splice site; (E) Intron retention (Fig. 9). After analysing the data of 2H-11 and SVEC4-10, we detected Exon skipping was the most frequent AS events and Mutually exclusive exon was the rarest AS events. The prevalence of different AS events was similar between TECs and NECs, indicating relevant pathogenesis of these two types of cells. It is worthy that one gene might possess several alternative splicing patterns. Differentially spliced genes (DSGs) between 2H-11 and SVEC4-10 were analysed and the top 20 significantly altered AS events were summarized in Table 12. A3SS(Alternative 3’splice site) and SES(Single exon skipping) were dominant AS types, and several genes (Gm20425, H2-K1, Zscan21, Npm1) demonstrated two AS events with opposite preference in TECs and NECs. According to reports, the five most widely used spontaneously metastatic immune activity models (AT-84, SCC VII, MOC2, MOC2-10, and 4MOC2) were sequenced and a comprehensive genome analysis was performed to analyze the patient’s tumors and utility evaluation. The results found that these five models have genetic variations in 10 cancer markers and 14 signaling pathways and mechanisms (metabolism, epigenetics, and immune evasion), with similar degrees in patients. Immune deficiencies in HLA-A (H2-Q10, H2-Q4, H2-Q7, and H2-K1), Pdcd1, Tgfb1, Il2ra, Il12a, Cd40, and Tnfrsf14 were identified [63]. In our results, in addition to the same alternative splicing subtypes H2-Q4, H2-Q7, H2-k1, we also detected the following alternative splicing subtypes H2-D1, H2-T2, H2-T22, H2-T24, H2-DMa. It indicated that the subtypes obtained by alternative splicing of TECs may play an important role in the tumor progression of immunodeficiency,

and the newly detected subtypes are likely to play an importment role with the known subtypes or independently play a functional role that has not yet been studied.

It is obvious that most AS patterns were more active in TECs than NECs. Moreover, the results demonstrated significant difference of AS patterns among different stages of TECs and NECs (all FDR<0.001).It reveals that certain alternative splicing patterns might show significant differences between TECs and NECs.In cancer vessels, TECs express several atypical splicing isoforms not expressed (or expressed at low levels) in NECs, which could represent putative targets for anti-angiogenic therapy [64]. Many of our results have not been studied in cancer and their biological functions are unclear. Therefore, alternative promoters, alternative terminators, and misregulation of exon skipping of these genes are promising research directions for elucidating the complex mechanisms of tumorigenesis and development using TECs as an entry point in the future.

Conclusion

In this study, we have shown the feasibility of developing transcriptomics profiling from viable single endothelial cells. Nanopore long-read transcriptome analysis shows the gene expression profile between tumor endothelial cells (TECs) and normal endothelial cells (NECs). Our results also accurately identify the structural variations of the transcript, such as alternative splicing, gene fusion, alternative polyadenylation, etc. We also accurately quantified the expression of transcripts and made simple predictions for new genes, LncRNA and transcription factors. Our research helps to understand the potential genes and pathways that may target TECs in future treatment research and expand the understanding of TECs progression.In addition to of this, our current results have limitations. So our next plan is to perform functional verification and mechanism clarification of some results based on the joint application of specific experimental techniques and bioinformatics. Although the TECs we used were derived from mouse, they have high homology between mouse and humans. It is completely meaningful for us to use mouse for experimental design.Understanding cell heterogeneity and regulatory changes under disease conditions is the basis for successful medical treatment development. Our transcriptomics analysis provides a wealth of candidate molecules for the development of targeted methods to counter the tumor-promoting behavior caused by TECs. Including increased tumor metabolism, tumor angiogenesis, changes in tumor microenvironment immune regulation and promotion of tumor metastasis, etc.

Methods

Preparation of 2H-11 and SVEC4-10 samples

The 2H-11(ATCC CRL-2163) Mouse tumor vascular endothelial cells line and SVEC4-10(ATCC CRL-2181) Mouse normal vascular endothelial cells line were purchased from the American Type Culture Collection (ATCC Manassas, VA, USA). The 2H-11 cells are grown in Dulbecco's Modified Eagle's Medium (DMEM; ATCC 302002)containing 10% fetal bovine serum (Gibco; Thermo Fisher Scientific,Inc.) at 37°C and 5%

CO2. SVEC4-10 cells are grown in DMEM (PM150210) containing 10% FBS(164210-500)and 1% P/S(PB180120) at 37°C and 5% CO2.

RNA isolation, cDNA library construction and Nanopore sequencing

Firstly, total RNA of 2H-11 three samples and SVEC4-10 three samples were respectively isolated using TRizol reagent (Thermo Fisher, Shanghai, China) on dry ice.1ug total RNA was prepared for cDNA libraries using cDNA-PCR Sequencing Kit (SQK-PCS109) protocol provided by Oxford Nanopore Technologies(ONT). Briefly ,the template switching activity of reverse transcriptases enrich for full-length cDNAs and add defined PCR adapters directly to both ends of the first-strand cDNA. And following cDNA PCR for 14 circles with LongAmp Tag (NEB). The PCR products were then subjected to ONT adaptor ligation using T4 DNA ligase (NEB). Agencourt XP beads was used for DNA purification according to ONT protocol. The final cDNA libraries were added to FLO-MIN109 flowcells and run on PromethION platform at Biomarker Technology Company (Beijing, China).

Oxford Nanopore Technologies Long Read Processing and remove redundant

The original data format of Nanopore sequencing off-machine data is the second-generation fast5 format that contains all the original sequencing signals. After base calling is performed through the Guppy software in the MinKNOW2.2 software package, fast5 format data will be converted to fastq format for subsequent quality control analysis.

(1) After the original fastq data is further filtered for short fragments and low-quality reads, the total Clean Data is obtained. Clean reads were filtered with minimum average read Q score=6 and minimum read length=500bp; ribosomal RNA were discarded after mapping to rRNA database.

(2) According to the principle of cDNA sequencing, Full-length non-chemiric (FLNC) transcripts were determined by searching for primer at both ends of reads.

(3) Clusters of FLNC transcripts were obtained after mapping to reference genome with mimimap2, and consensus isoforms were gained after polishing within each cluster by pinfish (https://github.com/nanoporetech/pinfish); consensus sequences were mapped to the reference genome (assembly AAP 1.0) using minimap2.

(4) Mapped reads were further collapsed by cDNA Cupcake package (https://github.com/Magdoll/cDNA_Cupcake) with min-coverage=85% and min-identity=90%; and alignments with differences only in the 5'exons are merged for alternative splicing analysis. Integrating the de-redundant results of each sample, we finally got 65,462 non-redundant transcript sequences.

Gene functional annotation and Quantification of gene expression levels and Differential expression analysis

Gene function was annotated based on the following databases: NR (NCBI non-redundant protein sequences);Pfam (Protein family); KOG/COG/eggNOG (Clusters of Orthologous Groups of proteins); Swiss-Prot (A manually annotated and reviewed protein sequence database); KEGG (Kyoto Encyclopedia of Genes and Genomes) and GO (Gene Ontology).Full length reads were mapped to the reference transcriptome sequence. Reads with match quality above 5 were further used to quantify. Expression levels were estimated by reads per gene/transcript per 10,000 reads mapped.Use the DESeq2 R package (1.6.3) to analyze the differential expression of six groups of three samples each of 2H-11 and SVEC4-10. DESeq2 provide statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with a FDR < 0.01 and fold change ≥ 2 found by DESeq2 were assigned as differentially expressed.

Functional enrichment analysis

Gene Ontology (GO) enrichment analysis of the differentially expressed genes (DEGs) was implemented by the GOseq R packages based Wallenius non-central hyper-geometric distribution[21], which can adjust for gene length bias in DEGs.

Kyoto Encyclopedia of Genes and Genomes (KEGG) [22] is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (http://www.genome.jp/kegg/). We used KOBAS [23] software to test the statistical enrichment of differential expression genes in KEGG pathways.

Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA)can perform enrichment analysis on all genes based on the expression of all genes without prior experience. General difference analysis usually only focuses on some significant up-regulated or down-regulated genes, and this will miss some genes that are not significantly differentially expressed but have important biological significance. The GSEA does not set a difference threshold and can detect weak but consistent trends. In this project, the GSEA-P software, MSigDB 1.0 [24]is used for GSEA analysis, the KEGG pathway and the gene set of the BP, CC, and MF branches of GO are used as the gene set of interest, and the log2FC of each difference group is used as the background gene set. Score to analyze the enrichment of the gene set of interest, and finally control pvalue or FDR as a significantly enriched gene set. Here, according to the analysis results of GSEA, we select the first 5 GO nodes/pathway of the enrichment as the enrichment map.

Structure analysis and Transcription factors prediction

Transcripts were validated against known reference transcript annotations with gffcompare[25]. Alternative splicing events including Intron retention(IR), Exon skipping(ES), Alternate Donor site (AD), Alternate acceptor site (AA) and Mutually exclusive exon (MEE) were identified by the AStalavista tool

[26]. Alternative polyadenylation analysis was conducted with TAPIS [27]. The reliable potential Coding Sequence (CDS) is obtained by Trans Decoder (v3.0.0) software[28] based on the length of Open Reading Frame (ORF), Log-likelihood Score, amino acid sequence and protein domain sequence of Pfam database. A total of 19,021 ORFs were obtained, including 9,954 complete ORFs.Animal transcription factors were identified from animal TFDB [29].

lncRNA analysis

Four computational approaches include CPC[30] /CNCI[31] /CPAT[32] /Pfam [33]/ were combined to sort non-protein coding RNA candidates from putative protein-coding RNAs in the transcripts.Putative protein-coding RNAs were filtered out using a minimum length and exon number threshold. Transcripts with lengths more than 200 nt and have more than two exons were selected as lncRNA candidates and further screened using CPC/CNCI/CPAT/Pfam that have the power to distinguish the protein-coding genes from the non-coding genes.

Statistical analysis

Data are expressed as the mean ± standard deviation. Two-tailed Student's t-test was used to assess the difference between two groups.One way analysis of variance followed by Dunnett's multiple comparisons test was used to determine differences between groups. $P < 0.05$ was considered to indicate a statistically significant difference.Statistical analysis was performed using SPSS software version 13.0 (SPSS, Inc., Chicago, IL, USA).

The following tables appear at the end of the pathway analysis in the results.

Table 3
Genes involved pathways in cancer (p-value = 4.22E-05; rich factor =1.51466523307125)

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000019464	Ptger1	0.004481278	1.555387272	up
ENSMUSG00000015312	Gadd45b	7.63E-13	-1.620740262	down
ENSMUSG00000038668	Lpar1	6.99E-12	2.464831326	up
ENSMUSG00000026875	Traf1	1.82E-06	4.328917412	up
ENSMUSG00000000126	Wnt9a	9.65E-14	-3.687846665	down
ENSMUSG00000027669	Gnb4	1.77E-24	3.156972597	up
ENSMUSG00000033446	Lpar6	4.30E-06	-1.47441176	down
ENSMUSG00000026167	Wnt10a	4.62E-15	6.140910757	up
ENSMUSG00000021253	Tgfb3	1.49E-19	-2.811490772	down
ENSMUSG00000040033	Stat2	1.73E-22	-2.753359624	down
ENSMUSG00000022996	Wnt10b	7.53E-06	4.14003287	up
ENSMUSG00000045005	Fzd5	0.003148121	-1.951220358	down
ENSMUSG00000024620	Pdgfrb	4.43E-25	7.387113993	up
ENSMUSG00000063594	Gng8	4.41E-07	4.504601408	up
ENSMUSG00000032348	Gsta4	3.32E-39	-2.885542505	down
ENSMUSG00000015957	Wnt11	3.01E-08	-3.714233803	down
ENSMUSG00000024521	Pmaip1	3.67E-07	4.210814343	up
ENSMUSG00000000184	Ccnd2	3.29E-06	-1.043125026	down
ENSMUSG00000023206	Il15ra	0.006162277	2.689922495	up
ENSMUSG00000001552	Jup	8.24E-10	-1.648099303	down
ENSMUSG00000003031	Cdkn1b	1.34E-05	-1.238122802	down
ENSMUSG00000021835	Bmp4	4.69E-14	-5.516947478	down
ENSMUSG00000027111	Itga6	3.34E-07	-1.825360992	down
ENSMUSG00000046532	Ar	1.07E-06	4.399699337	up
ENSMUSG00000021070	Bdkrb2	3.22E-11	5.434714104	up
ENSMUSG00000017057	Il13ra1	0.000470141	1.068079931	up
ENSMUSG00000002068	Ccne1	0.001002611	1.159325874	up

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000031659	Adcy7	6.37E-10	2.251734087	up
ENSMUSG00000031712	Il15	0.000643098	3.250151516	up
ENSMUSG00000032000	Birc3	5.59E-14	3.201901557	up
ENSMUSG00000030849	Fgfr2	4.92E-05	-2.886756219	down
ENSMUSG00000056234	Ncoa4	2.57E-30	-2.81089371	down
ENSMUSG00000058135	Gstm1	4.63E-17	2.362659014	up
ENSMUSG00000039239	Tgfb2	1.67E-09	-1.997547364	down
ENSMUSG00000031861	Lpar2	3.44E-06	-1.629340094	down
ENSMUSG00000031380	Vegfd	0.001566715	1.687775564	up
ENSMUSG00000074183	Gsta1	0.001087574	2.431146261	up
ENSMUSG00000061353	Cxcl12	6.90E-71	-4.592347083	down
ENSMUSG00000000782	Tcf7	0.000800291	1.88438868	up
ENSMUSG00000015647	Lama5	3.23E-16	-2.641045482	down
ENSMUSG00000038155	Gstp2	5.86E-09	2.328303859	up
ENSMUSG00000031273	Col4a6	5.24E-16	-3.063792325	down
ENSMUSG00000031304	Il2rg	1.75E-12	4.751579325	up
ENSMUSG00000001507	Itga3	6.06E-10	-1.777849405	down
ENSMUSG00000029337	Fgf5	6.46E-12	-4.845054065	down
ENSMUSG00000057614	Gnai1	1.92E-14	-4.954249953	down
ENSMUSG00000020250	Txnrd1	1.01E-06	-1.048797208	down
ENSMUSG00000008540	Mgst1	1.70E-12	-1.768900074	down
ENSMUSG00000031740	Mmp2	4.21E-37	5.134447584	up
ENSMUSG00000037225	Fgf2	0.007804604	-1.43756031	down
ONT.12871	ONT.12871	2.55E-05	-1.614334219	down
ENSMUSG00000074604	Mgst2	2.62E-05	2.166938591	up
ENSMUSG00000026104	Stat1	9.13E-57	-3.233512584	down
ENSMUSG00000003849	Nqo1	1.03E-05	1.102132904	up
ENSMUSG00000023067	Cdkn1a	3.96E-16	1.788555402	up

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000071042	Rasgrp3	0.009886116	2.313913324	up
ENSMUSG00000007659	Bcl2l1	9.06E-07	-1.290628098	down
ENSMUSG00000027208	Fgf7	3.10E-16	5.907005533	up
ENSMUSG00000021367	Edn1	9.63E-22	-3.748692778	down
ENSMUSG00000031565	Fgfr1	5.36E-23	5.719292741	up
ENSMUSG00000026688	Mgst3	3.90E-17	2.527299541	up
ENSMUSG00000001663	Gstt1	1.06E-12	2.740719466	up
ENSMUSG00000025739	Gng13	0.007734179	2.307657326	up
ENSMUSG00000038146	Notch3	3.06E-08	4.332684491	up
ENSMUSG00000006728	Cdk4	3.24E-09	-1.201435977	down
ENSMUSG00000058126	Tpm3-rs7	1.51E-08	4.141210391	up
ENSMUSG00000031274	Col4a5	4.62E-06	-1.233139489	down
ENSMUSG00000019789	Hey2	0.002142255	-3.125274981	down
ENSMUSG00000000489	Pdgfb	4.79E-08	-3.083130531	down
ENSMUSG00000029231	Pdgfra	4.37E-05	2.007407497	up
ENSMUSG00000032487	Ptgs2	0.000378606	1.242751519	up
ENSMUSG00000005413	Hmox1	1.39E-28	-2.218249722	down
ENSMUSG00000009376	Met	5.03E-08	-1.669939982	down
ENSMUSG00000024778	Fas	0.000845907	-1.111953014	down
ENSMUSG00000025856	Pdgfa	1.92E-06	1.546746373	up
ENSMUSG00000030748	Il4ra	8.93E-11	2.003146402	up
ENSMUSG00000030827	Fgf21	3.11E-05	-2.071979935	down
ENSMUSG00000032246	Calml4	6.89E-13	1.99463605	up
ENSMUSG00000004791	Pgf	3.86E-05	2.979696008	up
ENSMUSG00000020431	Adcy1	1.04E-07	-2.851126799	down
ENSMUSG00000031502	Col4a1	3.11E-11	-1.438161228	down
ENSMUSG00000027381	Bcl2l11	0.002811121	-1.254929202	down
ENSMUSG00000025934	Gsta3	8.73E-17	-6.331940412	down

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000043004	Gng2	5.92E-20	3.418575448	up
ENSMUSG00000026923	Notch1	2.50E-10	4.719561581	up
ENSMUSG00000032766	Gng11	9.35E-26	6.592055622	up
ENSMUSG00000057967	Fgf18	8.98E-11	-5.337256823	down
ENSMUSG00000027890	Gstm4	0.006745519	1.774064896	up

Table 4
Genes involved PI3K-Akt signaling pathway. (p-value = 0.000572211189699301; rich factor =1.47360346291882)

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000048126	Col6a3	9.48E-34	8.237114364	up
ENSMUSG00000038668	Lpar1	6.99E-12	2.464831326	up
ENSMUSG00000027669	Gnb4	1.77E-24	3.156972597	up
ENSMUSG00000033446	Lpar6	4.30E-06	-1.47441176	down
ENSMUSG00000028364	Tnc	4.22E-12	2.223658118	up
ENSMUSG00000024620	Pdgfrb	4.43E-25	7.387113993	up
ENSMUSG00000023885	Thbs2	3.24E-31	8.006758795	up
ENSMUSG00000063594	Gng8	4.41E-07	4.504601408	up
ENSMUSG00000000184	Ccnd2	3.29E-06	-1.043125026	down
ONT.22189	ONT.22189	4.39E-06	2.178374899	up
ENSMUSG00000003031	Cdkn1b	1.34E-05	-1.238122802	down
ENSMUSG00000001281	Itgb7	7.71E-05	2.925093206	up
ONT.1713	ONT.1713	0.001840856	-1.836943132	down
ENSMUSG00000021457	Syk	3.90E-17	6.397263609	up
ENSMUSG00000027111	Itga6	3.34E-07	-1.825360992	down
ENSMUSG00000025915	Sgk3	2.51E-08	-2.689074051	down
ENSMUSG00000002068	Ccne1	0.001002611	1.159325874	up
ENSMUSG00000048482	Bdnf	1.22E-12	-1.886705573	down
ONT.9836	ONT.9836	0.001699125	-1.736891074	down
ENSMUSG00000006445	Epha2	0.000415402	-1.008969954	down
ENSMUSG00000030849	Fgfr2	4.92E-05	-2.886756219	down
ENSMUSG00000028466	Creb3	1.59E-06	-1.058144164	down
ENSMUSG00000031861	Lpar2	3.44E-06	-1.629340094	down
ENSMUSG00000031380	Vegfd	0.001566715	1.687775564	up
ENSMUSG00000038648	Creb3l2	7.93E-07	-1.396213632	down
ENSMUSG00000003070	Efna2	0.001110511	-2.009317279	down

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000015647	Lama5	3.23E-16	-2.641045482	down
ENSMUSG00000031273	Col4a6	5.24E-16	-3.063792325	down
ENSMUSG00000031304	Il2rg	1.75E-12	4.751579325	up
ENSMUSG00000001507	Itga3	6.06E-10	-1.777849405	down
ENSMUSG00000029337	Fgf5	6.46E-12	-4.845054065	down
ENSMUSG00000048915	Efna5	7.43E-15	-2.822568609	down
ONT.5605	ONT.5605	0.006964006	-1.031582078	down
ENSMUSG00000037225	Fgf2	0.007804604	-1.43756031	down
ONT.5337	ONT.5337	1.49E-13	4.618450161	up
ONT.12871	ONT.12871	2.55E-05	-1.614334219	down
ENSMUSG00000023067	Cdkn1a	3.96E-16	1.788555402	up
ENSMUSG00000007659	Bcl2l1	9.06E-07	-1.290628098	down
ENSMUSG00000027208	Fgf7	3.10E-16	5.907005533	up
ENSMUSG00000031565	Fgfr1	5.36E-23	5.719292741	up
ENSMUSG00000025739	Gng13	0.007734179	2.307657326	up
ENSMUSG00000006728	Cdk4	3.24E-09	-1.201435977	down
ENSMUSG00000031274	Col4a5	4.62E-06	-1.233139489	down
ENSMUSG00000029661	Col1a2	1.27E-06	1.30884117	up
ENSMUSG00000000489	Pdgfb	4.79E-08	-3.083130531	down
ENSMUSG00000029231	Pdgfra	4.37E-05	2.007407497	up
ENSMUSG00000028047	Thbs3	2.46E-10	2.24046441	up
ENSMUSG00000079092	Prl2c2	2.56E-05	2.351883915	up
ENSMUSG00000028019	Pdgfc	8.87E-10	2.061535439	up
ENSMUSG00000001119	Col6a1	1.36E-10	-1.735704123	down
ENSMUSG00000024777	Ppp2r5b	1.03E-10	-1.998562884	down
ENSMUSG00000042284	Itga1	1.01E-07	-2.184817206	down
ENSMUSG00000009376	Met	5.03E-08	-1.669939982	down
ENSMUSG00000022817	Itgb5	3.96E-08	2.06121516	up

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000032006	Pdgfd	9.19E-05	3.757978313	up
ENSMUSG00000025856	Pdgfa	1.92E-06	1.546746373	up
ENSMUSG00000030748	Il4ra	8.93E-11	2.003146402	up
ENSMUSG00000030827	Fgf21	3.11E-05	-2.071979935	down
ENSMUSG00000020241	Col6a2	7.99E-12	-2.197425083	down
ENSMUSG00000004791	Pgf	3.86E-05	2.979696008	up
ENSMUSG00000031502	Col4a1	3.11E-11	-1.438161228	down
ENSMUSG00000027381	Bcl2l11	0.002811121	-1.254929202	down
ENSMUSG00000043004	Gng2	5.92E-20	3.418575448	up
ONT.14228	ONT.14228	0.001694666	2.705282492	up
ENSMUSG00000032766	Gng11	9.35E-26	6.592055622	up
ENSMUSG00000020901	Pik3r5	3.23E-06	3.594723947	up
ENSMUSG00000057967	Fgf18	8.98E-11	-5.337256823	down
ONT.7383	ONT.7383	8.73E-06	3.619241787	up
ENSMUSG00000020108	Ddit4	1.02E-09	1.799740624	up
ENSMUSG00000029304	Spp1	4.87E-133	5.690870707	up

Table 5
Genes involved Proteoglycans in cancer. (p-value = 0.000202842293116734; rich factor =1.78895291310168)

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000000126	Wnt9a	9.65E-14	-3.687846665	down
ENSMUSG00000030774	Pak1	1.14E-13	2.644006802	up
ENSMUSG00000026167	Wnt10a	4.62E-15	6.140910757	up
ENSMUSG00000022996	Wnt10b	7.53E-06	4.14003287	up
ENSMUSG00000045005	Fzd5	0.003148121	-1.951220358	down
ENSMUSG00000015957	Wnt11	3.01E-08	-3.714233803	down
ENSMUSG00000024998	Plce1	8.56E-19	-6.195320166	down
ENSMUSG00000024401	Tnf	9.88E-09	4.921767952	up
ENSMUSG00000038086	Hspb2	4.81E-07	1.20152208	up
ENSMUSG00000021822	Plau	0.00212055	1.360346051	up
ENSMUSG00000004266	Ptpn6	0.005149339	1.475199289	up
ENSMUSG00000021477	Ctsl	2.43E-25	-1.747916795	down
ENSMUSG00000035273	Hpse	0.000133412	3.523055696	up
ENSMUSG00000031714	Gab1	0.000510204	-1.524537327	down
ENSMUSG00000024486	Hbegf	0.000665058	1.079615461	up
ENSMUSG00000039239	Tgfb2	1.67E-09	-1.997547364	down
ENSMUSG00000069601	Ank3	0.000425512	2.366113746	up
ENSMUSG00000019907	Ppp1r12a	1.36E-05	-1.084887677	down
ENSMUSG00000052397	Ezr	1.46E-110	-3.973529447	down
ENSMUSG00000031740	Mmp2	4.21E-37	5.134447584	up
ENSMUSG00000007655	Cav1	7.82E-15	-1.891462588	down
ENSMUSG00000034220	Gpc1	2.78E-09	2.149554812	up
ENSMUSG00000037225	Fgf2	0.007804604	-1.43756031	down
ENSMUSG00000025351	Cd63	1.20E-06	-1.02254876	down
ENSMUSG00000023067	Cdkn1a	3.96E-16	1.788555402	up
ENSMUSG00000009621	Vav2	6.68E-06	-1.590740892	down

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000031565	Fgfr1	5.36E-23	5.719292741	up
ENSMUSG00000017009	Sdc4	6.03E-09	-1.518175322	down
ENSMUSG00000029661	Col1a2	1.27E-06	1.30884117	up
ENSMUSG00000005087	Cd44	0.000121501	1.123040059	up
ENSMUSG00000009376	Met	5.03E-08	-1.669939982	down
ENSMUSG00000032470	Mras	8.16E-08	-1.418686184	down
ENSMUSG00000004864	Mapk13	5.17E-19	6.650712814	up
ENSMUSG00000024778	Fas	0.000845907	-1.111953014	down
ENSMUSG00000022817	Itgb5	3.96E-08	2.06121516	up
ENSMUSG00000046223	Plaur	1.02E-15	-1.57649484	down
ENSMUSG00000020044	Timp3	9.79E-54	-3.078712357	down
ENSMUSG00000019929	Dcn	5.35E-96	-4.556672091	down
ENSMUSG00000007805	Twist2	0.000163699	2.475389749	up
ENSMUSG00000030102	Itpr1	3.37E-07	-1.619609064	down

Table 6
Genes involved Focal adhesion. (p-value = 0.000254614724043645; rich factor =1.75609810479376)

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000048126	Col6a3	9.48E-34	8.237114364	up
ENSMUSG00000030774	Pak1	1.14E-13	2.644006802	up
ENSMUSG00000021823	Vcl	1.99E-09	-1.584165501	down
ENSMUSG00000028364	Tnc	4.22E-12	2.223658118	up
ENSMUSG00000020469	Myl7	6.17E-18	-6.054355828	down
ENSMUSG00000024620	Pdgfrb	4.43E-25	7.387113993	up
ENSMUSG00000023885	Thbs2	3.24E-31	8.006758795	up
ENSMUSG00000000184	Ccnd2	3.29E-06	-1.043125026	down
ENSMUSG00000001281	Itgb7	7.71E-05	2.925093206	up
ENSMUSG00000027111	Itga6	3.34E-07	-1.825360992	down
ENSMUSG00000032000	Birc3	5.59E-14	3.201901557	up
ENSMUSG00000031380	Vegfd	0.001566715	1.687775564	up
ENSMUSG00000015647	Lama5	3.23E-16	-2.641045482	down
ENSMUSG00000031273	Col4a6	5.24E-16	-3.063792325	down
ENSMUSG00000001507	Itga3	6.06E-10	-1.777849405	down
ENSMUSG00000067818	Myl9	4.09E-61	10.27294181	up
ENSMUSG00000019907	Ppp1r12a	1.36E-05	-1.084887677	down
ENSMUSG00000007655	Cav1	7.82E-15	-1.891462588	down
ENSMUSG00000006219	Fblim1	2.35E-19	2.113325297	up
ENSMUSG00000009621	Vav2	6.68E-06	-1.590740892	down
ENSMUSG00000031274	Col4a5	4.62E-06	-1.233139489	down
ENSMUSG00000029661	Col1a2	1.27E-06	1.30884117	up
ENSMUSG00000000489	Pdgfb	4.79E-08	-3.083130531	down
ENSMUSG00000029231	Pdgfra	4.37E-05	2.007407497	up
ENSMUSG00000028047	Thbs3	2.46E-10	2.24046441	up
ENSMUSG00000028019	Pdgfc	8.87E-10	2.061535439	up

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000029860	Zyx	8.09E-10	-1.474719245	down
ENSMUSG00000035133	Arhgap5	0.00030879	-1.343751853	down
ENSMUSG00000001119	Col6a1	1.36E-10	-1.735704123	down
ENSMUSG00000042284	Itga1	1.01E-07	-2.184817206	down
ENSMUSG00000009376	Met	5.03E-08	-1.669939982	down
ENSMUSG00000022817	Itgb5	3.96E-08	2.06121516	up
ENSMUSG00000032006	Pdgfd	9.19E-05	3.757978313	up
ENSMUSG00000025856	Pdgfa	1.92E-06	1.546746373	up
ENSMUSG00000026594	Ralgps2	3.59E-07	-1.55487962	down
ENSMUSG00000020241	Col6a2	7.99E-12	-2.197425083	down
ENSMUSG00000004791	Pgf	3.86E-05	2.979696008	up
ENSMUSG00000031502	Col4a1	3.11E-11	-1.438161228	down
ENSMUSG00000022438	Parvb	8.24E-38	8.588755469	up
ENSMUSG00000039114	Nrn1	4.70E-19	6.657679888	up
ENSMUSG00000029304	Spp1	4.87E-133	5.690870707	up

Table 7
Genes involved MAPK Signaling pathway. (p-value= 0.000273002393812116; rich factor =1.6312226985099)

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000015312	Gadd45b	7.63E-13	-1.620740262	down
ENSMUSG00000030774	Pak1	1.14E-13	2.644006802	up
ENSMUSG00000021253	Tgfb3	1.49E-19	-2.811490772	down
ENSMUSG00000024620	Pdgfrb	4.43E-25	7.387113993	up
ENSMUSG00000020623	Map2k6	5.62E-05	3.13258605	up
ENSMUSG00000024401	Tnf	9.88E-09	4.921767952	up
ENSMUSG00000026072	Il1r1	2.19E-13	2.038495544	up
ENSMUSG00000002307	Daxx	1.07E-06	-1.198496457	down
ENSMUSG00000031392	Irak1	0.007458992	-1.059587506	down
ENSMUSG00000037337	Map4k1	3.17E-08	4.337362115	up
ENSMUSG00000019960	Dusp6	2.58E-34	-3.02063234	down
ENSMUSG00000048482	Bdnf	1.22E-12	-1.886705573	down
ENSMUSG00000059970	Hspa2	0.000972136	-1.211418313	down
ENSMUSG00000006445	Epha2	0.000415402	-1.008969954	down
ENSMUSG00000090877	Hspa1b	5.02E-10	3.433264195	up
ENSMUSG00000030849	Fgfr2	4.92E-05	-2.886756219	down
ENSMUSG00000039239	Tgfb2	1.67E-09	-1.997547364	down
ENSMUSG00000031383	Dusp9	0.001945323	2.931905617	up
ONT.16231	ONT.16231	8.53E-47	9.288250485	up
ENSMUSG00000031380	Vegfd	0.001566715	1.687775564	up
ENSMUSG00000003070	Efna2	0.001110511	-2.009317279	down
ENSMUSG00000034765	Dusp5	0.000420241	-2.032967264	down
ENSMUSG00000029337	Fgf5	6.46E-12	-4.845054065	down
ENSMUSG00000048915	Efna5	7.43E-15	-2.822568609	down
ENSMUSG00000037225	Fgf2	0.007804604	-1.43756031	down
ENSMUSG00000071042	Rasgrp3	0.009886116	2.313913324	up

#ID	gene_name	FDR	log2FC	regulated
ENSMUSG00000091971	Hspa1a	3.55E-12	5.290791162	up
ENSMUSG00000027208	Fgf7	3.10E-16	5.907005533	up
ENSMUSG00000071369	Map3k5	0.000528465	-3.02026962	down
ENSMUSG00000031565	Fgfr1	5.36E-23	5.719292741	up
ENSMUSG00000000531	Grasp	0.00863678	2.107104805	up
ENSMUSG00000000489	Pdgfb	4.79E-08	-3.083130531	down
ENSMUSG00000029231	Pdgfra	4.37E-05	2.007407497	up
ENSMUSG00000028989	Angptl7	1.77E-30	-8.520019771	down
ENSMUSG00000031530	Dusp4	2.01E-06	-1.643448406	down
ENSMUSG00000028019	Pdgfc	8.87E-10	2.061535439	up
ENSMUSG00000027500	Stmn2	7.03E-08	4.716455948	up
ENSMUSG00000009376	Met	5.03E-08	-1.669939982	down
ENSMUSG00000032470	Mras	8.16E-08	-1.418686184	down
ENSMUSG00000004864	Mapk13	5.17E-19	6.650712814	up
ENSMUSG00000024778	Fas	0.000845907	-1.111953014	down
ENSMUSG00000032006	Pdgfd	9.19E-05	3.757978313	up
ENSMUSG00000025856	Pdgfa	1.92E-06	1.546746373	up
ENSMUSG00000026594	Ralgps2	3.59E-07	-1.55487962	down
ENSMUSG00000030827	Fgf21	3.11E-05	-2.071979935	down
ENSMUSG00000004791	Pgf	3.86E-05	2.979696008	up
ENSMUSG00000032577	Mapkapk3	1.68E-05	4.028619132	up
ENSMUSG00000024948	Map4k2	2.15E-05	-1.807154667	down
ENSMUSG00000057967	Fgf18	8.98E-11	-5.337256823	down
ENSMUSG00000025408	Ddit3	7.21E-22	-1.916912092	down
ENSMUSG00000020866	Cacna1g	0.008139772	-1.901413319	down
ENSMUSG00000024242	Map4k3	0.000124131	-1.253268683	down

Abbreviations

TECs: Tumor endothelial cells; NECs:Normal endothelial cells;

ONT: Oxford Nanopore Technology; GO:Gene ontology;

DEGs: Differentially expressed genes; TME:tumor microenvironment;

Iso-Seq:Isoform sequencing; FPKM: Fragments per kilobase of transcripts per million

FDR: False discovery rate; FC: fold change;

BPs: Biological Processes; CCs: Cellular Components ; MFs: Molecular Functions ;

KEGG : Kyoto Encyclopedia of Genes and Genomes;

GSEA: Gene Set Enrichment Analysis; ES: enrichment score;AS: alternative splicing;

PSI: percent-splice-in; A3SS: Alternative 3'splice site;

A5SS: Alternative 5's plice site; SES: Single exon skipping ;

TSS:Transcription start site (Alternative 5'first site)

DSGs:Differentially spliced genes; mRNA:messenger RNA;

lncNRA: long noncoding RNAs (lncRNAs)

Poly A tail: Polyadenylic acid tail; APA: Alternative polyadenylation;

ORF: Open Reading Frame; CDS: Coding Sequence;

SAGE: serial analysis of gene expression; scRNA-seq: single-cell RNA sequencing;

ECs: enthelial cells; ECM: extracellular matrix;

HNSC: head and neck squamous cell carcinoma;

TAM: Tumour-associated macrophages;

ATCC:American Type Culture Collection;

DMEM: Dulbecco's Modified Eagle's Medium;

FLNC: Full-length non-chemiric; NR:NCBI non-redundant protein sequences

Pfam: Protein family

KOG/COG/eggNOG: Clusters of Orthologous Groups of proteins

Swiss-Prot: A manually annotated and reviewed protein sequence database

IR: Intron retention; ES: Exon skipping; AD: Alternate Donor site

AA: Alternate acceptor site; MEE: Mutually exclusive exon

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated during the current study are available in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE191094)

Competing interests

The authors declare that they have no competing interests.

Funding

The present study was supported by the Fundamental Research Funds for the Central Universities, JLU and the Natural Science Foundation of Jilin Province, China (Grant No. 20170623093-06TC20190201091JC).

Authors' contributions

Yuanyuan Tian drafted the manuscript. Jiao Zhao design the study. Ju Huang analyzed the data. Haiying Zhang, Fushun Ni and Rong Tang contributed to revising and editing the manuscript. All authors have read and approved the manuscript, and ensure that this is the case.

Acknowledgements

Not applicable.

Author details

The Key Laboratory of Pathobiology, Ministry of Education, The College of Basic Medical Sciences, Jilin University, Changchun, Jilin 130021, China

References

1. Dudley AC. Tumor endothelial cells. Cold Spring Harb Perspect Med. 2012; 2:a006536.
2. Cameron Ivan,Short Nicholas,Sun LuZhe,Hardman W Elaine. Endothelial cell pseudopods and angiogenesis of breast cancer tumors. Cancer Cell International. 2005; 5(1):17.
3. Kessenbrock K, Plaks V, Werb Z. Matrix Metalloproteinases: Regulators of the Tumor Microenvironment. Cell. 2010; 141:52–67.
4. Ilmer M, Horst D. Pancreatic CSCs and microenvironment. Genes Cancer. 2015; 6:365–6.
5. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nature Medicine. 2013; 19:1423–37.
6. V ander Jeught K, Bialkowski L, Daszkiewicz L, et,al. Targeting the tumor microenvironment to enhance antitumor immune responses. Oncotarget. 2015; 6:1359–81.
7. Kyoko Hida,Nako Maishi,Yu Sakurai,Yasuhiro Hida,Hideyoshi Harashima. Heterogeneity of tumor endothelial cells and drug delivery. Advanced Drug Delivery Reviews. 2016; 99 (Pt B):140–147.
8. Nagl L, Horvath L, Pircher A, Wolf D. Tumor Endothelial Cells (TECs) as Potential Immune Directors of the Tumor Microenvironment - New Findings and Future Perspectives. Front Cell Dev Biol. 2020 Aug 19;8:766.
9. Brad St. Croix; Carlo Rago; Victor Velculescu; et,al. Genes Expressed in Human Tumor Endothelium. Science, 2000, 289(5482):1197–1202.
10. Nako Maishi; Dorcas A. Annan; Hiroshi Kikuchi; Yasuhiro Hida; Kyoko Hida. Tumor Endothelial Heterogeneity in Cancer Progression. Cancers, 2019, 11(10):1511–1511.
11. Lewis L Michelle; Edwards Meredith C; Meyers Zachary R; et,al. Replication Study: Transcriptional amplification in tumor cells with elevated c-Myc. eLife, 2018, 7
12. Lambrechts Diether; Wauters Els; Boeckx Bram; Aibar Sara; et,al. Phenotype molding of stromal cells in the lung tumor microenvironment. Nature Medicine, 2018, 24(4):1277–1289.
13. Katerina Rohlenova; Jermaine Goveia; Melissa García-Caballero; Abhishek Subramanian; et,al..Single-Cell RNA Sequencing Maps Endothelial Metabolic Plasticity in Pathological Angiogenesis. Cell Metabolism, 2020, 31(4):862-877.e14.
14. Buckanovich Ronald J; Facciabene Andrea; Kim Sarah; Benencia Fabian; et,al. Endothelin B receptor mediates the endothelial barrier to T cell homing to tumors and disables immune therapy.Nature Medicine, 2008, 14(1):28–36.
15. Kambayashi Taku, Laufer Terri M. Atypical MHC class II-expressing antigen-presenting cells: can anything replace a dendritic cell? Nature Reviews Immunology, 2014, 14(11):719–30.
16. Jermaine Goveia; Katerina Rohlenova; Federico Taverna; Lucas Treps; et,al. An Integrated Gene Expression Landscape Profiling Approach to Identify Lung Tumor Endothelial Cell Heterogeneity and Angiogenic Candidates. Cancer Cell, 2020, 37(3):421–421.

17. Sautès-Fridman Catherine; Petitprez Florent; Calderaro Julien; Fridman Wolf Herman. Tertiary lymphoid structures in the era of cancer immunotherapy. Nature Reviews Cancer, 2019, 19(6):307–325.
18. Deamer D, Akeson M, Branton D: Three decades of nanopore sequencing. Nat Biotechnol 2016, 34:518–524.
19. Magi Alberto; Semeraro Roberto; Mingrino Alessandra; Giusti Betti; D'Aurizio Romina. Nanopore sequencing data analysis: state of the art, applications and challenges. Briefings in Bioinformatics, 2017, 1–17.
20. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 2016, 17:239.
21. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2):R14.
22. Kanehisa M, Araki M, Goto S, Hattori M, et,al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008 Jan;36(Database issue):D480-4.
23. Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics. 2005 Oct 1;21(19):3787-93.
24. Subramanian Aravind; Tamayo Pablo; Mootha Vamsi K; et,al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(43):15545–50.
25. Gffcompare: classify, merge, tracking and annotation of GFF files by comparing to a reference annotation GFF.
26. Foissac S, Sammeth M. ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Research 2007, 35(Web Server issue):W297-9.
27. Abdel-Ghany Salah E; Hamilton Michael; Jacobi Jennifer L; Ngam Peter; Devitt Nicholas; Schilkey Faye; Ben-Hur Asa; Reddy Anireddy S N. A survey of the sorghum transcriptome using single-molecule long reads. Nature Communications, 2016, 7:11706.
28. TransDecoder: Find Coding Regions Within Transcripts.
29. Hu Hui; Miao Ya-Ru; Jia Long-Hao; Yu Qing-Yang; Zhang Qiong; Guo An-Yuan. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Research, 2019, 47(D1):D33-D38.
30. Kong Lei; Zhang Yong; Ye Zhi-Qiang; Liu Xiao-Qiao; Zhao Shu-Qi; Wei Liping; Gao Ge. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Research, 2007, 35(Web Server issue):W345-9.
31. Sun Liang; Luo Haitao; Bu Dechao; Zhao Guoguang; Yu Kuntao; Zhang Changhai; Liu Yuanning; Chen Runsheng; Zhao Yi. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Research, 2013, 41(17):e166.

32. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model.Nucleic Acids Res. 2013 Apr 1;41(6):e74.
33. Mistry Jaina; Chuguransky Sara; Williams Lowri; Qureshi Matloob; et,al. Pfam: The protein families database in 2021. Nucleic Acids Research, 2020, 49(D1):D412-D419.
34. St Croix B, Rago C, Velculescu V, Traverso G, Romans KE, Montgomery E, Lal A, Riggins GJ, Lengauer C, Vogelstein B, Kinzler KW. Genes expressed in human tumor endothelium. Science. 2000; 289:1197–202.
35. Ghilardi C, Chiorino G, Dossi R, Nagy Z, Giavazzi R, Bani M. Identification of novel vascular markers through gene expression profiling of tumor-derived endothelium. BMC Genomics. 2008; 9:201.
36. Parker Belinda S; Argani Pedram; Cook Brian P; Liangfeng Han; et,al.Alterations in vascular gene expression in invasive breast carcinoma. Cancer Research, 2004, 64(21):7857–66.
37. Bhati R, Patterson C, Livasy CA, Fan C, Ketelsen D, Hu Z, Reynolds E, Tanner C, Moore DT, Gabrielli F, Perou CM, Klauber-DeMore N. Molecular characterization of human breast tumor vascular cells. Am J Pathol. 2008; 172:1381–90.
38. Pepin F, Bertos N, Laferriere J, Sadekova S, Souleimanova M, et,al. Gene-expression profiling of microdissected breast cancer microvasculature identifies distinct tumor vascular subtypes. Breast Cancer Res. 2012; 14:R120.
39. Hu M, Polyak K. Serial analysis of gene expression. Nat Protoc. 2006;1:1743–60.
40. Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, Zhou A, Eyob H, Balakrishnan S, Wang CY, Yaswen P, Goga A, Werb Z. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature. 2015; 526:131–5.
41. Anoop P Patel, Itay Tirosh, John J Trombetta, et,al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014; 344(6190):1396–1401.
42. Abdelmohsen K, Panda AC, De S, Grammatikakis I, Kim J, Ding J, Noh JH, Kim KM, Mattison JA, de Cabo R, Gorospe M. Circular RNAs in monkey muscle: agedependent changes. Aging (Albany NY). 2015; 7:903–10.
43. Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, Cvejic A, Teichmann SA. Power analysis of single-cell RNA-sequencing experiments. Nature Methods. 2017; 14:381–387.
44. Lambrechts Diether; Wauters Els; Boeckx Bram; Aibar Sara; et,al. Phenotype molding of stromal cells in the lung tumor microenvironment. Nature Medicine, 2018, 24(4):1277–1289.
45. Zhao, Q, Eichten, A, Parveen, A, Adler, C, Huang, Y, Wang, W, Ding, Y, Adler, A, Nevins, T, Ni, M. Single-cell transcriptome analyses reveal endothelial cell heterogeneity in tumors and changes following antiangiogenic treatment. Cancer Res. 2018, 78, 2370–2382.
46. Coppiello G, Collantes M, Sirerol-Piquer MS, Vandenwijngaert S, et,al. Meox2/Tcf15 heterodimers program the heart capillary endothelium for cardiac fatty acid uptake. Circulation. 2015 Mar 3;131(9):815-26.

47. Raluca Marcu; Yoon Jung Choi; Jun Xue; Chelsea L. Fortin; et,al. Human Organ-Specific Endothelial Cell Heterogeneity. iScience, 2018, 4:20–35.
48. Nolan DJ, Ginsberg M, Israely E, Palikuqi B, et,al. Molecular signatures of tissue-specific microvascular endothelial cell heterogeneity in organ maintenance and regeneration. Dev Cell. 2013 Jul 29;26(2):204–219.
49. Sabbagh MF, Heng JS, Luo C, Castanon RG, Nery JR, Rattner A, Goff LA, Ecker JR, Nathans J. Transcriptional and epigenomic landscapes of CNS and non-CNS vascular endothelial cells. Elife. 2018 Sep 6;7:e36187.
50. Sun Zhengda; Wang Chih-Yang; Lawson Devon A; et,al. Single-cell RNA sequencing reveals gene expression signatures of breast cancer-associated endothelial cells. Oncotarget, 2018, 9(13):10945–10961.
51. Anna Rita Cantelmo; Lena-Christin Conradi; Aleksandra Brajic; et,ala. Inhibition of the Glycolytic Activator PFKFB3 in Endothelium Induces Tumor Vessel Normalization, Impairs Metastasis, and Improves Chemotherapy. Cancer Cell, 2016, 30(6):968–985.
52. Costas A. Lyssiotis; Alec C. Kimmelman. Metabolic Interactions in the Tumor Microenvironment. Trends in Cell Biology, 2017, 27(11):863–875.
53. Buckanovich Ronald J; Facciabene Andrea; Kim Sarah; et,al. Endothelin B receptor mediates the endothelial barrier to T cell homing to tumors and disables immune therapy. Nature Medicine, 2008, 14(1):28–36.
54. Wülfing Pia; Diallo Raihanatou; Kersting Christian; Wülfing Christian; et,al. Expression of endothelin-1, endothelin-A, and endothelin-B receptor in human breast cancer and correlation with long-term follow-up. Clinical Cancer Research, 2003, 9(11):4125-31.
55. Grimshaw Matthew J; Hagemann Thorsten; Ayhan Ayse; Gillett Cheryl E; Binder Claudia; Balkwill Frances R. A role for endothelin-2 and its receptors in breast tumor cell invasion. Cancer Research, 2004, 64(7):2461–8.
56. Feng Peng-Cheng; Ke Xing-Fei; Kuang Hui-Lan; Pan Li-Li; Ye Qiang; Wu Jian-Bing. BMP2 secretion from hepatocellular carcinoma cell HepG2 enhances angiogenesis and tumor growth in endothelial cells via activation of the MAPK/p38 signaling pathway. Stem Cell Research & Therapy, 2019, 10(1):237.
57. Zhou J, Zhang A, Fan L. HSPA12B Secreted by Tumor-Associated Endothelial Cells Might Induce M2 Polarization of Macrophages via Activating PI3K/Akt/mTOR Signaling. Onco Targets Ther. 2020 Sep 11;13:9103–9111.
58. Munkley Jennifer; Livermore Karen; Rajan Prabhakar; Elliott David J. RNA splicing and splicing regulator changes in prostate cancer pathology. Human genetics 2017; 136(9):1143–1154
59. Paronetto MP, Passacantilli I, Sette C. Alternative splicing and cell survival: from tissue homeostasis to disease. Cell death and differentiation. 2016; 23(12):1919–1929.
60. Narayanan SP, Singh S, Shukla S. A saga of cancer epigenetics: linking epigenetics to alternative splicing. The Biochemical journal. 2017; 474(6):885–896

61. Scotti MM, Swanson MS. RNA mis-splicing in disease. Nature reviews Genetics. 2016; 17(1):19–32.
62. Xiaowei Song; Zhenyu Zeng; Huanhuan Wei; Zefeng Wang. Alternative splicing in cancers: From aberrant regulation to new therapeutics. Seminars in Cell & Developmental Biology, 2018, 75:13–22.
63. Li Hui; Ngan HoiLam; Liu Yuchen; Chan Helen Hoi Yin; et,al. Comprehensive Exome Analysis of Immunocompetent Metastatic Head and Neck Cancer Models Reveals Patient Relevant Landscapes. Cancers, 2020, 12(10):2935–2935.
64. Di Matteo Anna; Belloni Elisa; Pradella Davide; et,al. Alternative splicing in endothelial cells: novel therapeutic opportunities in cancer angiogenesis. Journal of Experimental & Clinical Cancer Research, 2020, 39(1):275–275.

Figures

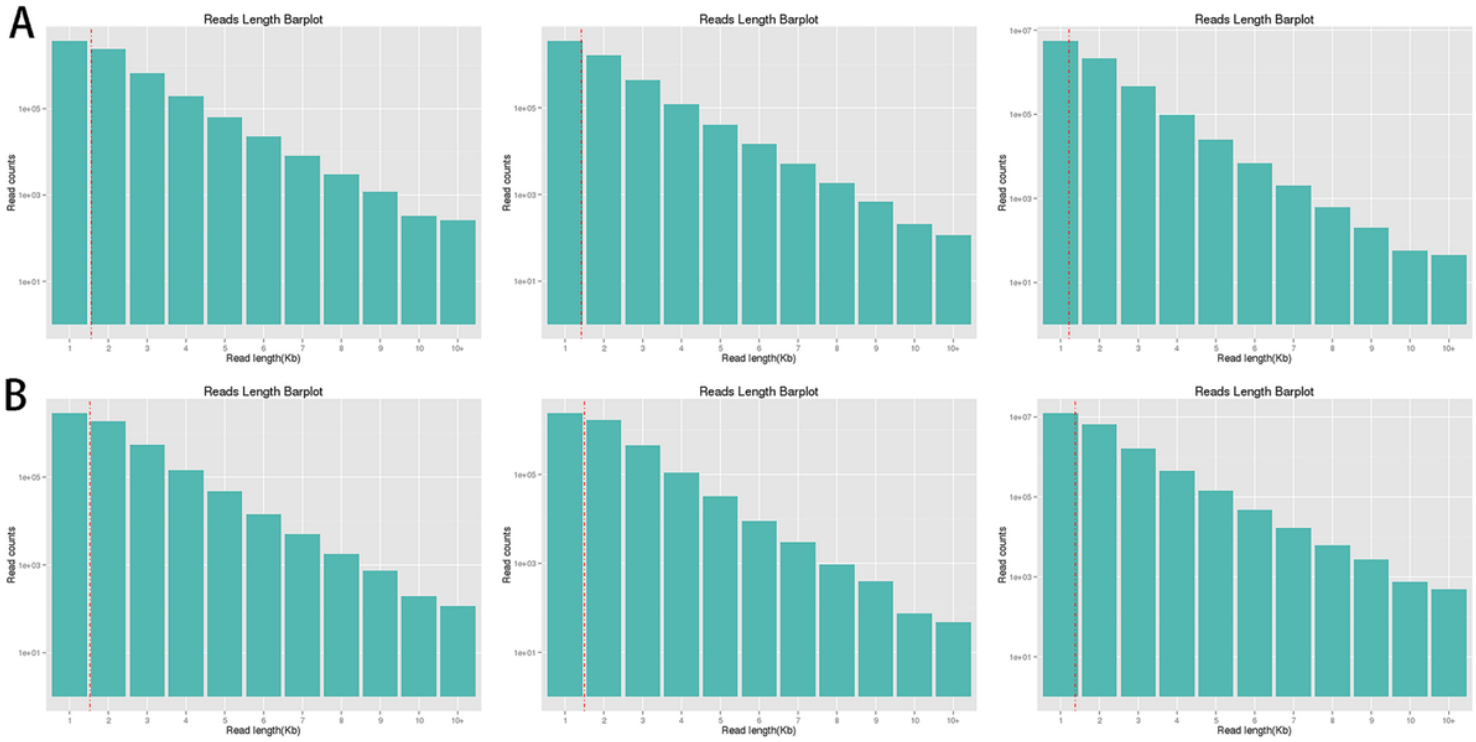

Figure 1

Length distribution of clean reads generated from Nanopore sequencing. **A** Clean reads of 2H-11. **B** Clean reads of SVEC4-10.

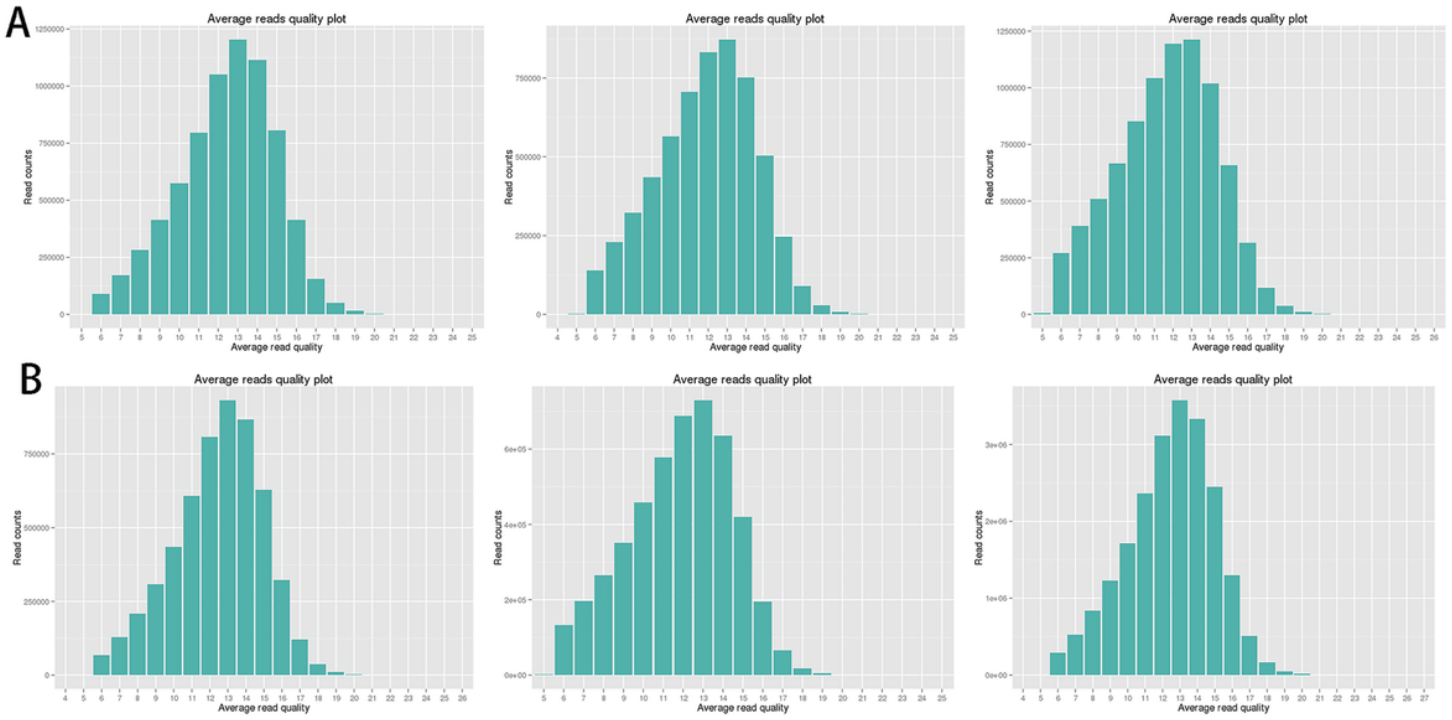

Figure 2

Average quality distribution of clean reads derived from Nanopore sequencing. **A** Clean reads of 2H-11.**B** Clean reads of SVEC4-10.

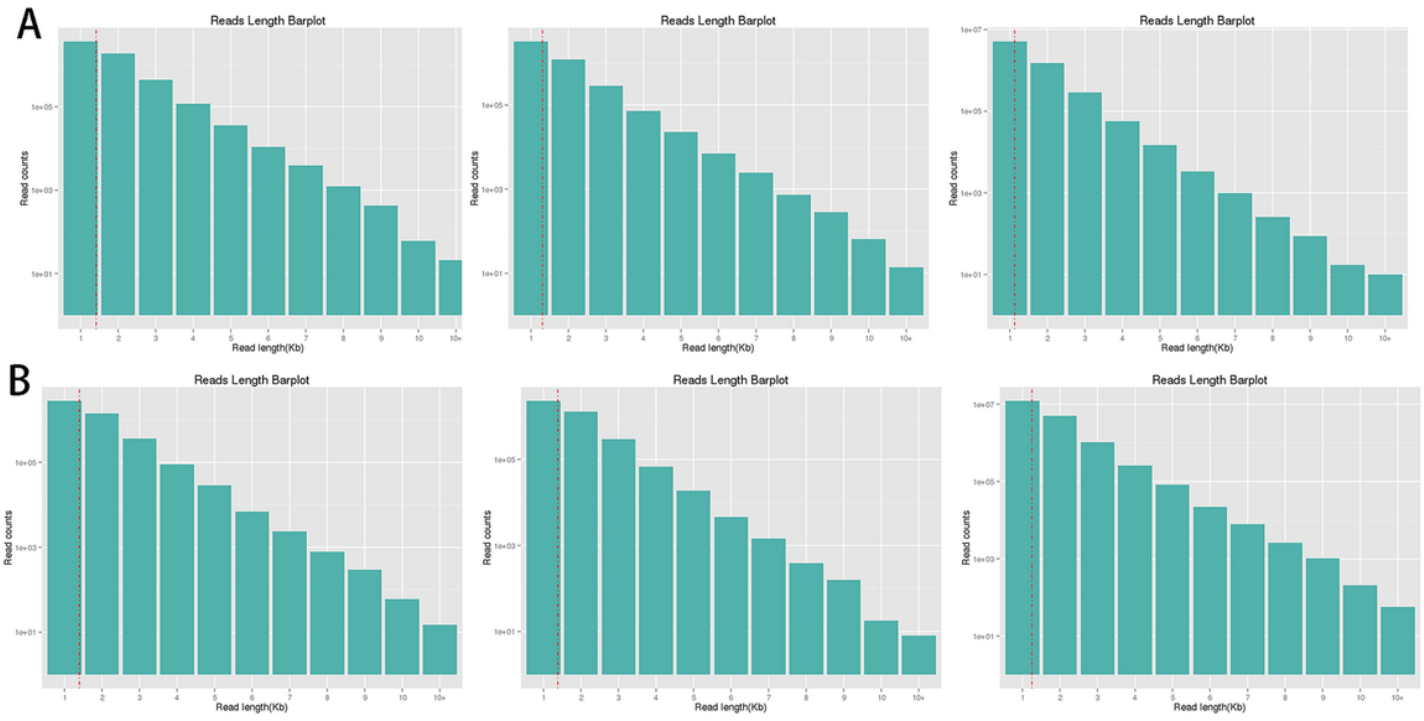

Figure 3

Length distribution of full-length reads. **A** Full-length reads of 2H-11. **B** Full-length reads of SVEC4-10.

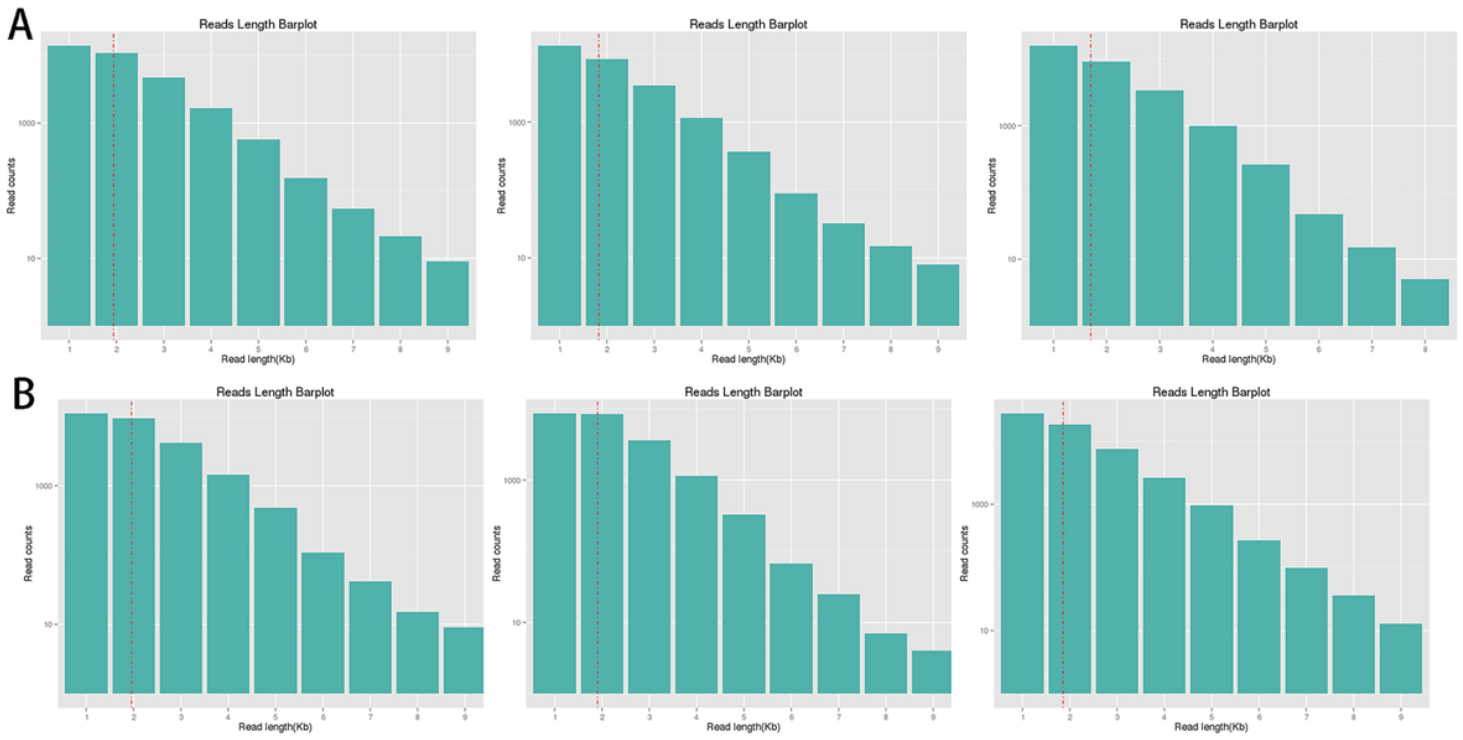

Figure 4

Length distribution of full-length transcripts after removing redundant reads. A Redundant reads-removed full-length transcripts of 2H-11. B Redundant reads-removed full-length transcripts of SVEC4-10.

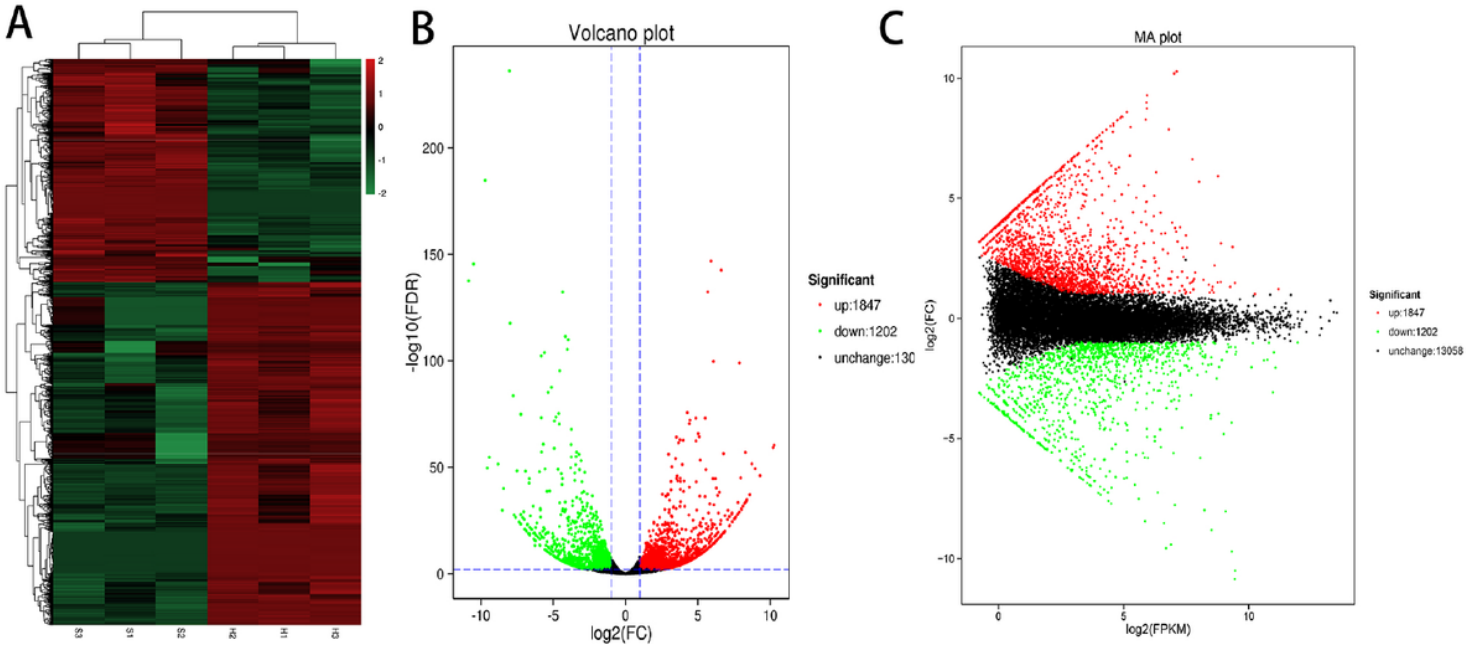

Figure 5

Heat map, Volcano and MA plots. **A** Numbers under the x-axis indicated different sample and clustering of sample, the y-axis indicated DEGs and hierarchical clustering of DEGs. Hierarchical clustering of all

DEGs based on normalized FPKM (Fragments per kilobase of transcripts per million) values. Red indicates higher expression, and green indicates lower expression. S1, S2 and S3 indicate three biological replications of SVEC4-10 cells. H1, H2 and H3 indicate three biological replications of 2H-11 cells. **B** The volcano plot was constructed by plotting the negative log of the log10 FDR (false discovery rate)) value on the y-axis.This results in data points with low log10 FDRvalues (highly significant) appearing toward the top of the plot. The x-axis is the log2FC(fold change) between the two cells (SVEC4-10 and 2H-11). **C** MA plot visualizes the differences between measurements taken in SVEC4-10 cells and 2H-11cells DEGs, by transforming the data into M (log ratio), A (mean average) scales and log2FPKM (fragments per million), log2FC, then plotting these values. The green dots represent down-regulated DEGs, the red dots represent up-regulated DEGs, and the black dots represent non-differentially expressed genes.

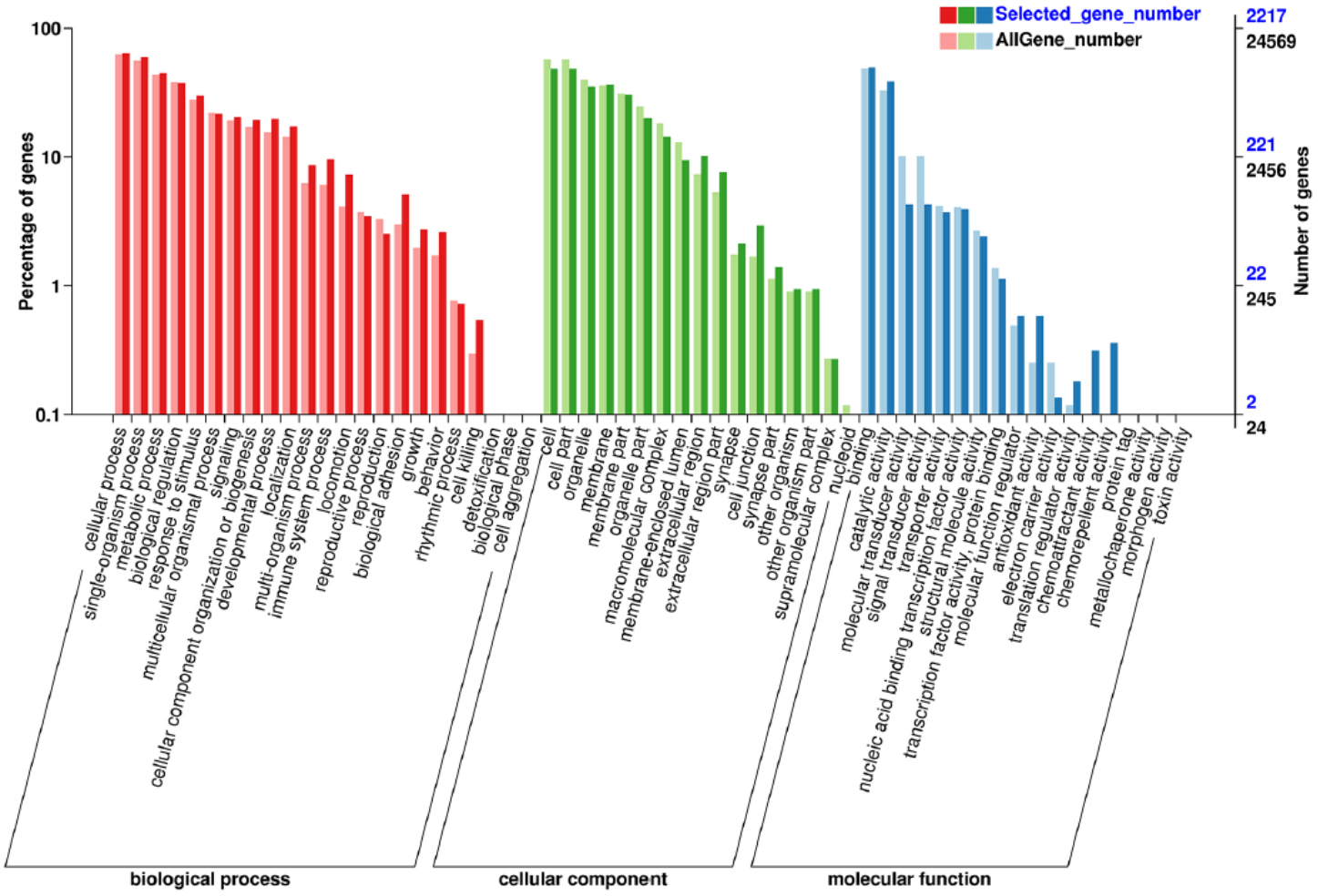

Figure 6

Differentially expressed genes in GO annotation classification statistics map.The abscissa is the GO classification, the left side of the ordinate is the percentage of the number of transcripts, and the right side is the number of transcripts. This figure shows the transcript enrichment of each secondary function

of GO in the background of differential expression transcripts and the background of all transcripts, reflecting the status of each secondary function in the two contexts, and the description of the secondary functions with obvious differences in proportions. The enrichment trend of differentially expressed transcripts is different from that of all transcripts, so we can focus on analyzing whether this function is related to differences.

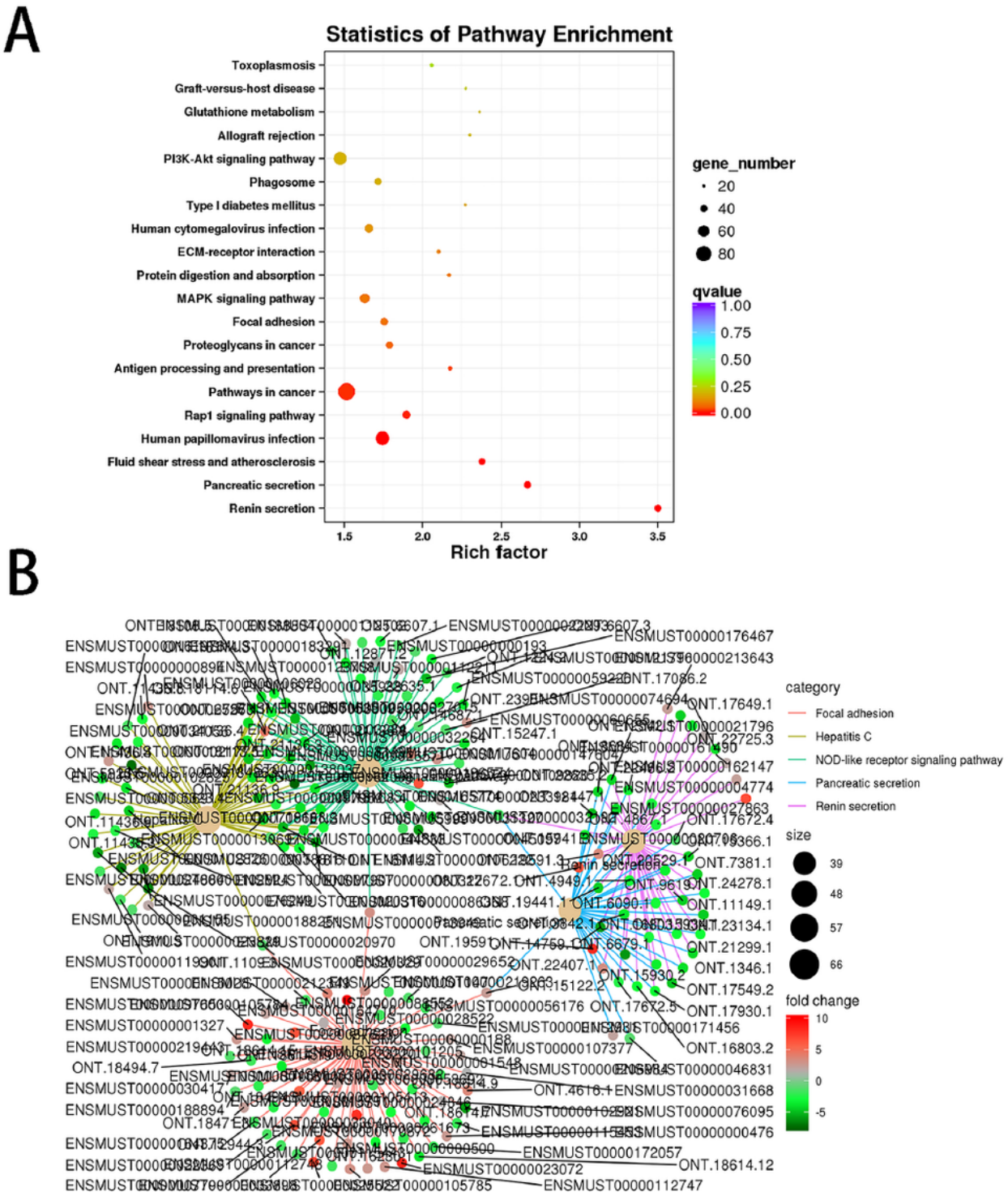

Figure 7

Differentially expressed genes KEGG pathway enrichment statistics map. **A** The y-axis represents the enriched KEGG pathways and the horizontal axis represents the rich factor. The higher the rich factor, the more significant the enrichment. The diameter of the sphere represents the number of unigenes in each pathway. The colour indicated correct P value (q value) on multiple hypothesis testing. The lower the q value, the more reliable the testing.**B** The colors of the radial lines represent the types of different pathways (pink: Focal adhesion; yellow: Hepatitis C; green: NOD-like receptor signaling pathway; blue: Pancreatic secretion; purple: Renin secretion).The size of the center circle represents the number of genes enriched in the pathway.The small circles of different colors represent the fold change(FC) between 2H-11 and SVEC4-10.

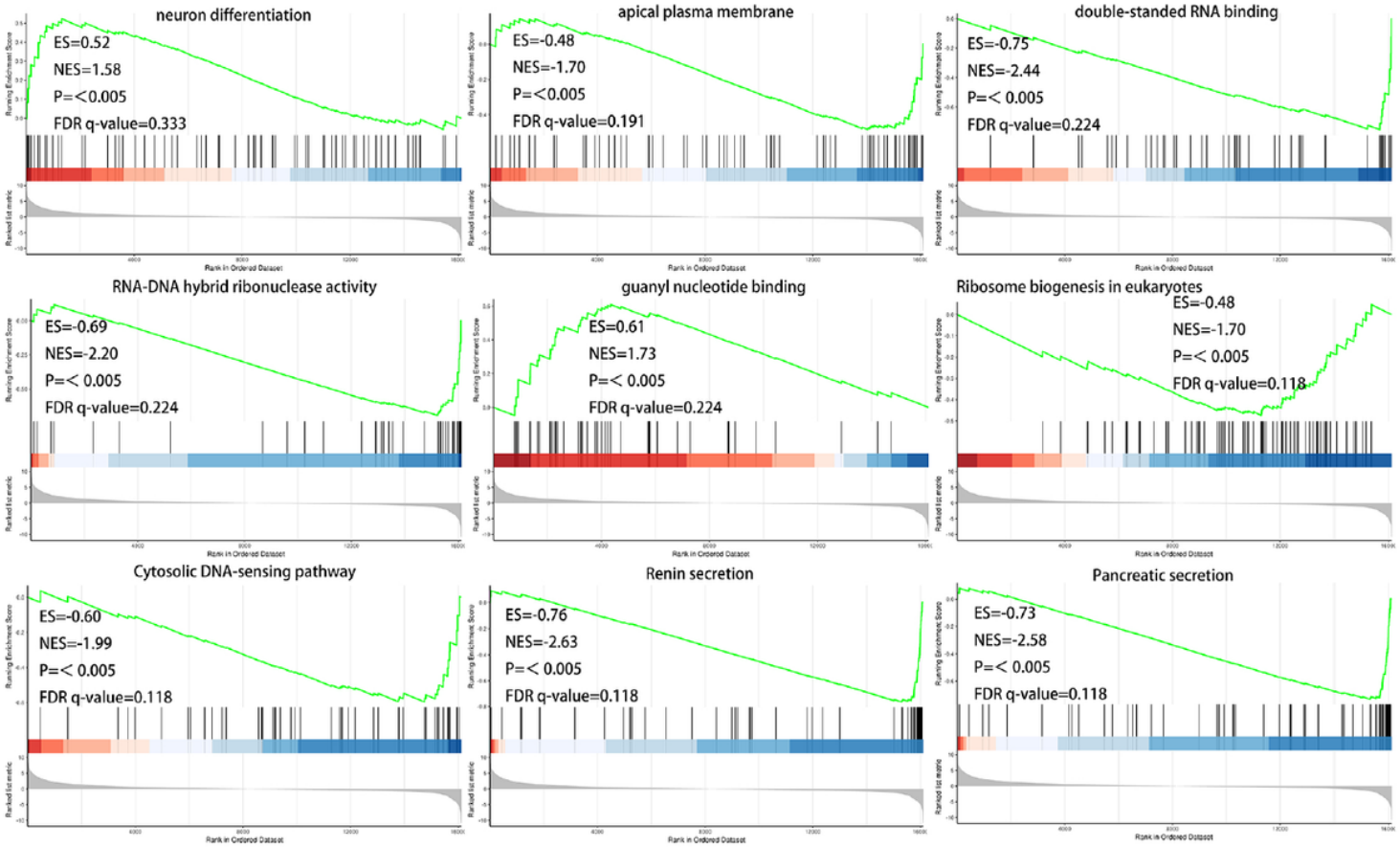

Figure 8

GSEA enriches genes into multiple functional gene networks**.**The title of the graph represents the name of GO Term/KEGG Pathway; the abscissa represents the position information of the sorted gene set; the black vertical line in the middle part of the graph on the abscissa represents the genes in the GO Term/KEGG Pathway; the bottom part of the graph is the distribution of Rank values of all genes in the gene set (the distribution of log2FC).The ordinate represents the sorting quantity (log2FC); the top part of the figure is the line chart of the gene enrichment score (ES); the ordinate is the dynamic ES, the green curve represents the ES of the gene set at each position, the peak value of the green curve represents the ES value of the gene set.

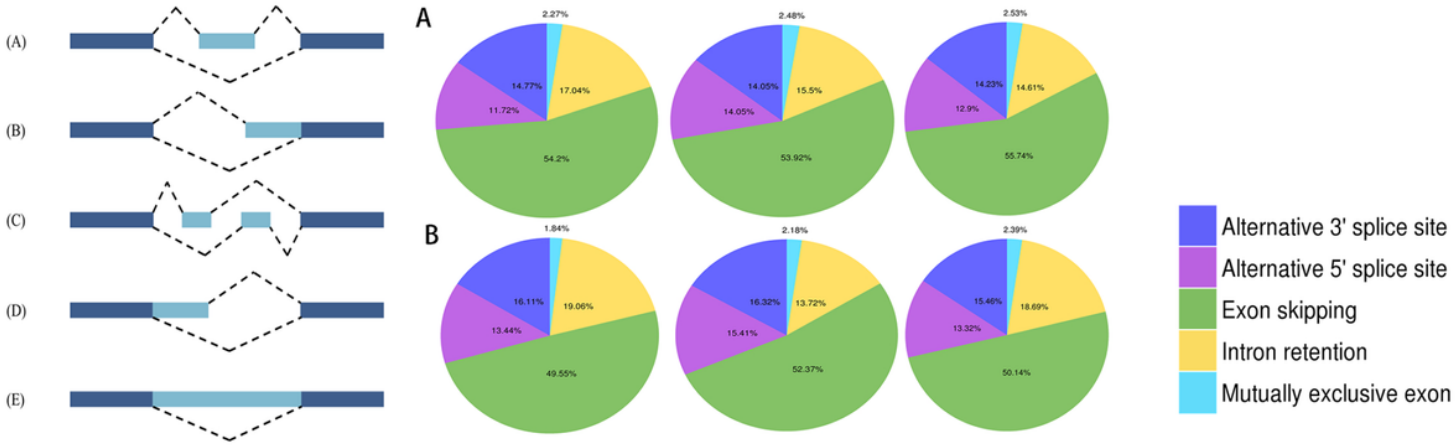

Figure 9

Alternative splicing(AS) types and the statistic pie chart of AS events. (A) Exon skipping; (B) Alternative 3’splice site (C)Mutually exclusive exon (D) Alternative 5’splice site (E) Intron retention.**A** Percentage of each class of alternative splicing event in the six samples of 2H-11.**B** Percentage of each class of alternative splicing event in the six samples of SVEC4-10.

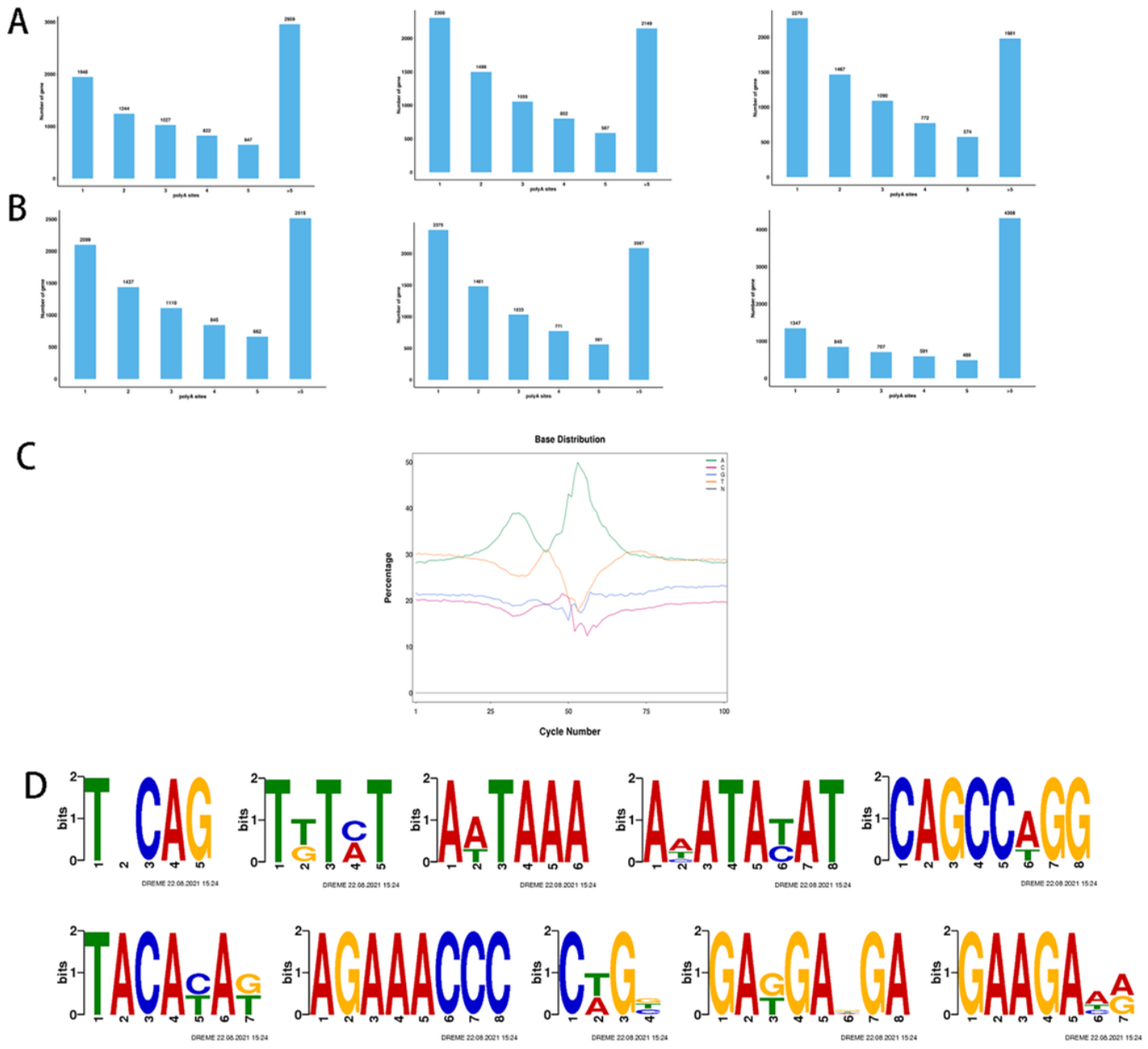

Figure 10

Alternative polyadenylation (APA). **A** Distribution of the number of polyadenylation sites of genes in 2H-11. **B** Distribution of the number of polyadenylation sites of genes in SVEC4-10. X-coordinate: number of polyadenylation sites; Y-coordinate: number of genes. **C** APA upstream and downstream base distribution map. **D** Motif map upstream of polyA site.

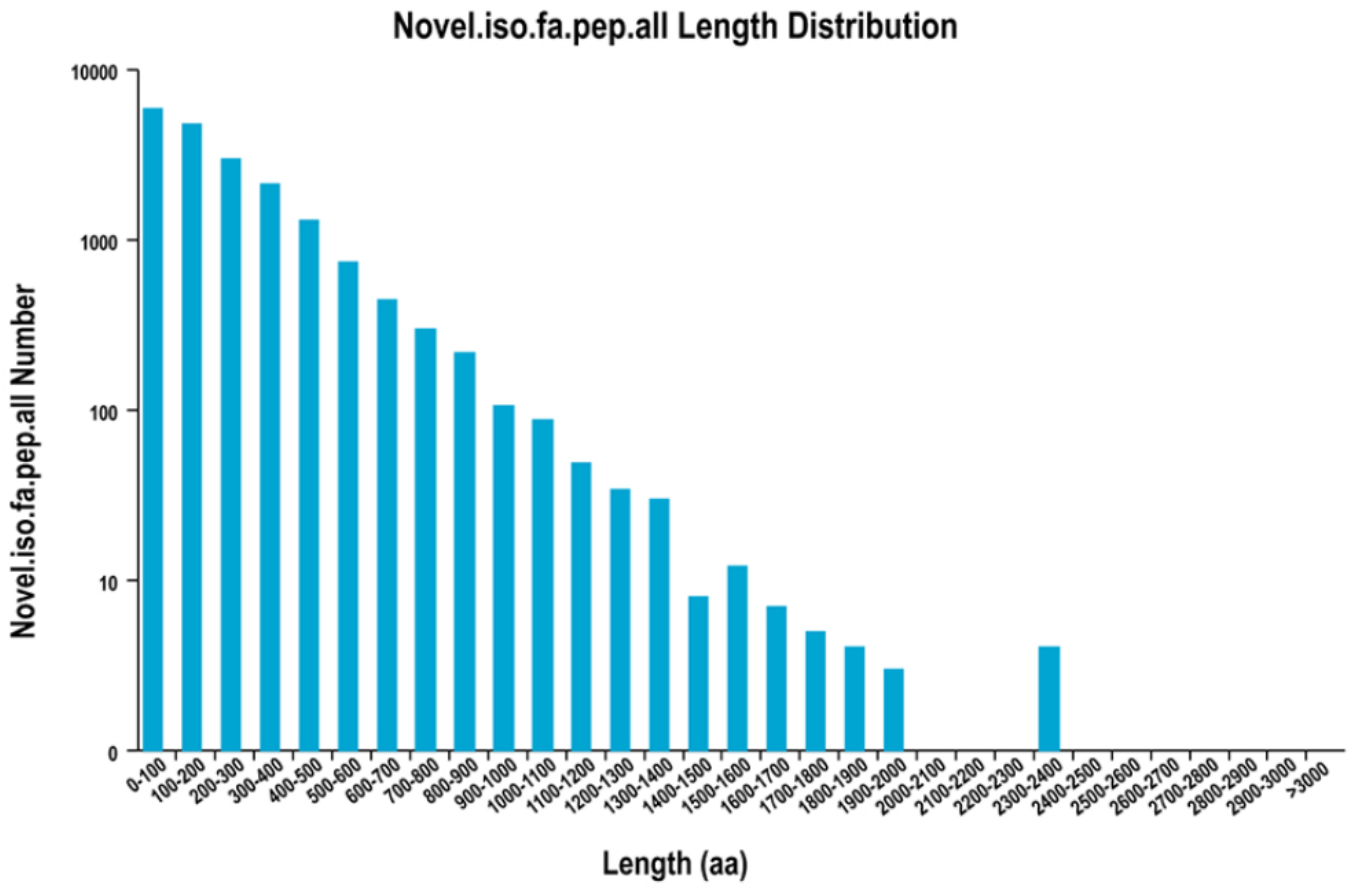

Figure 11

Predicted CDS-encoded protein length distribution map.

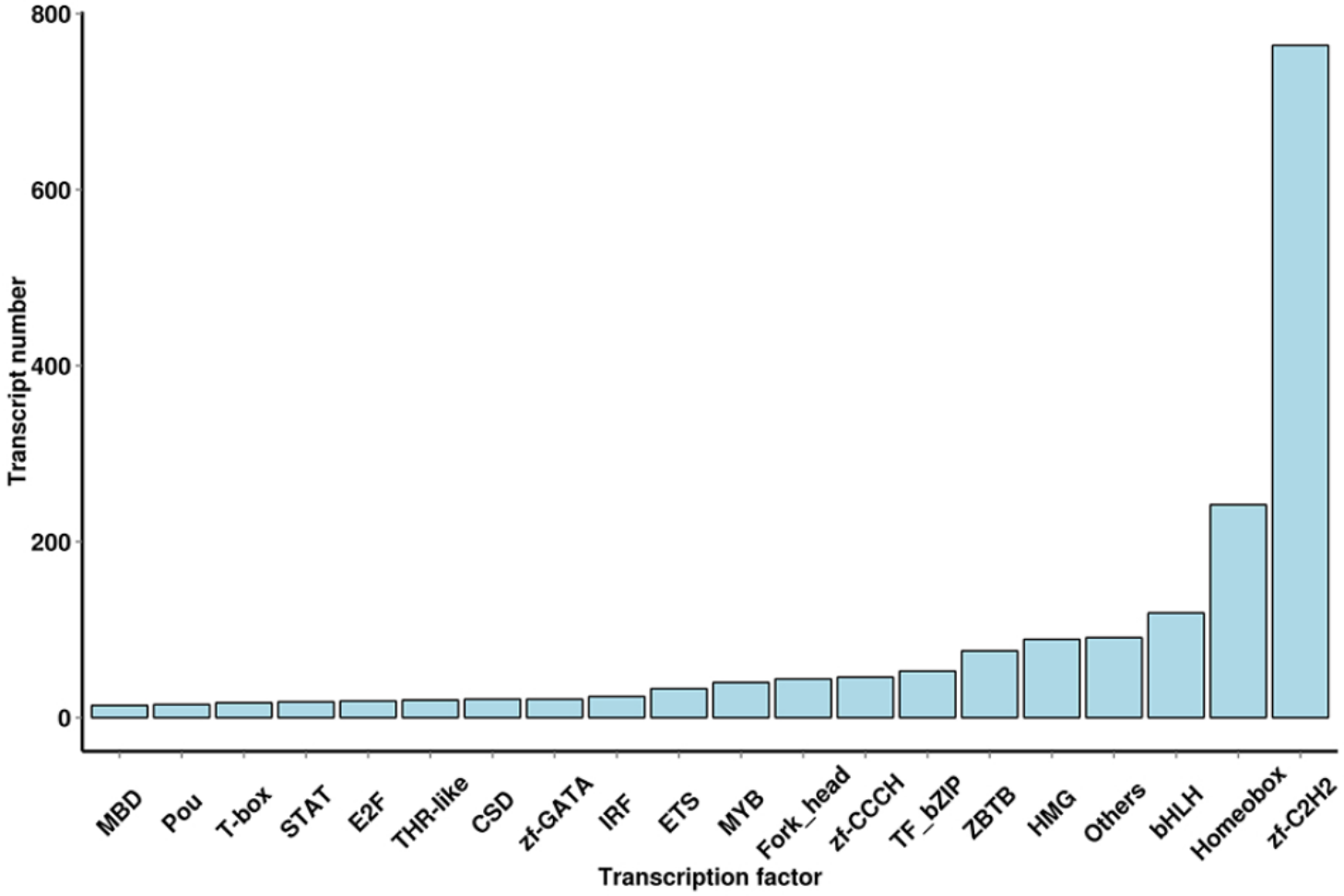

Figure 12

Transcription factor type distribution. The abscissa is the classification of transcription factor families, and the ordinate is the number of transcription factors.

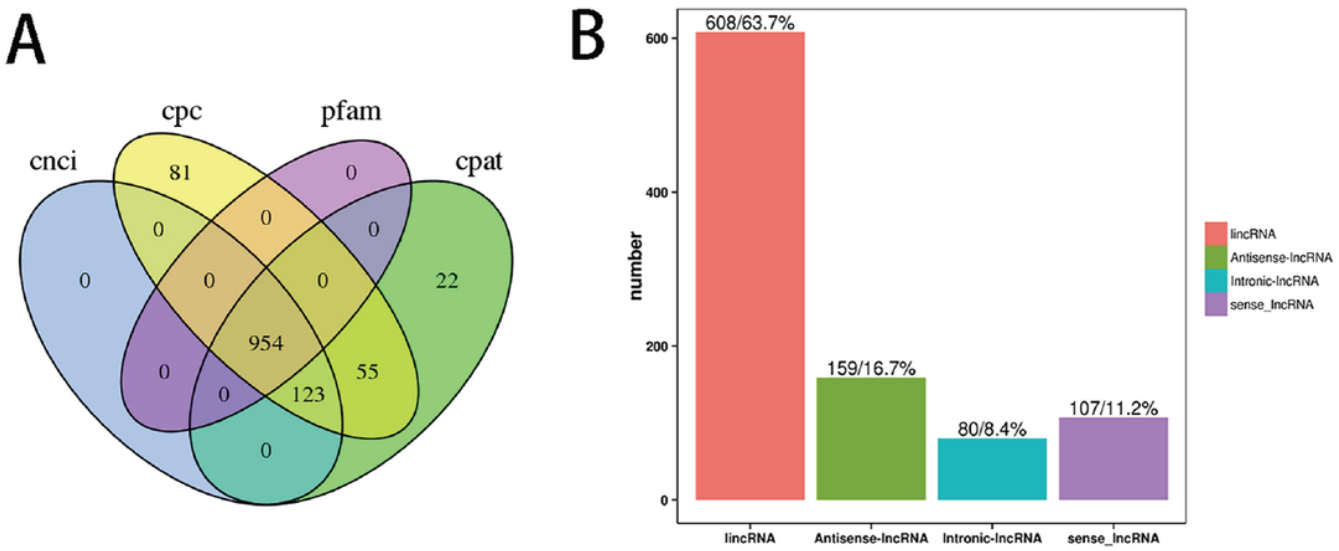

Figure 13

Identification of lncRNA transcripts. **A** Venn graph of lncRNA prediction by four steps, including CPAT, CNCI, CPC, and Pfam. **B** Position classification map obtained according to the position of lncRNA on the reference genome annotation information (gff).