Single-cell transcriptomics and Whole-genome sequencing reveal tumor landscape in breast carcinosarcoma

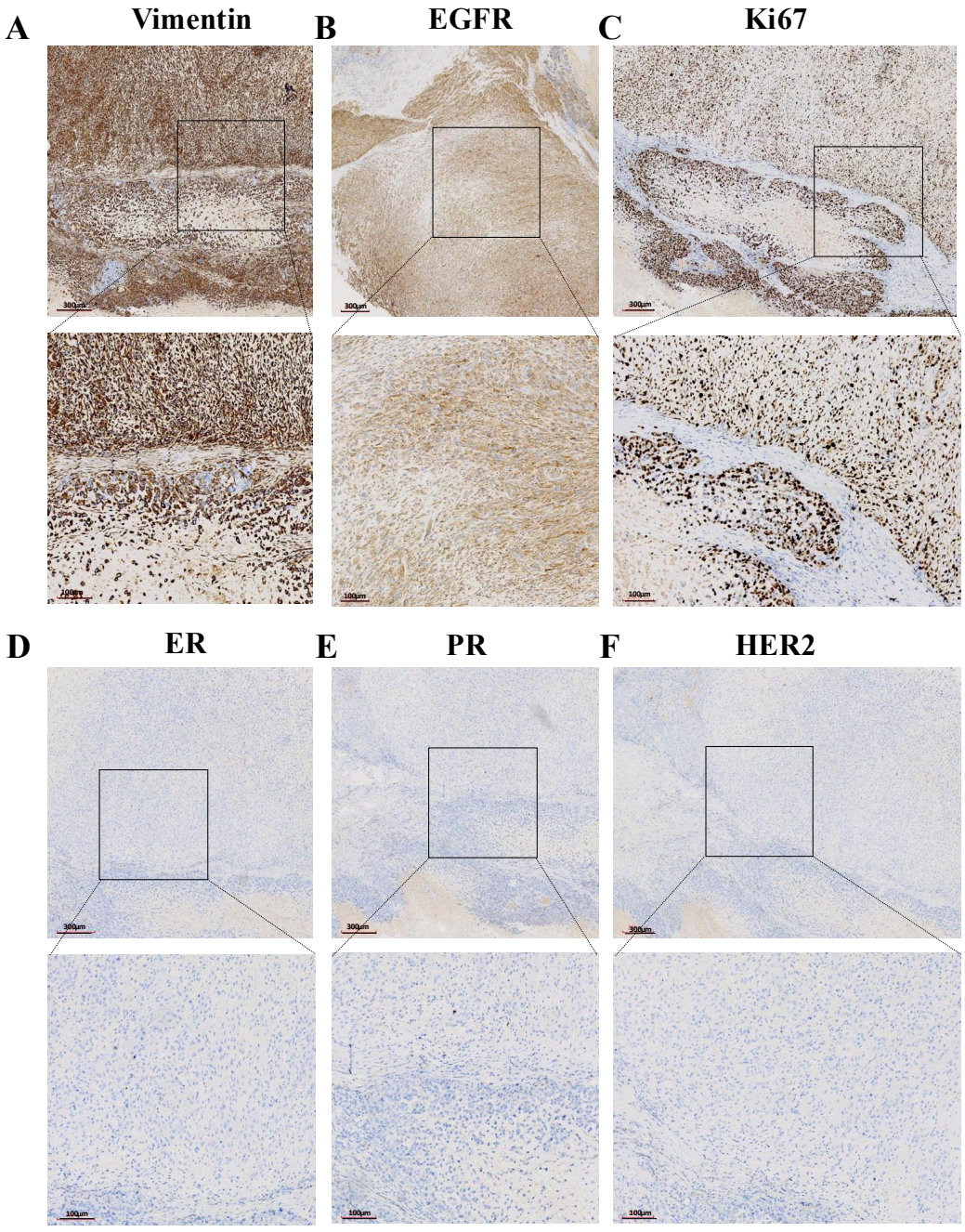

Figure S1 Photomicrographic immunohistochemical images of the tumor cells from breast carcinosarcoma. Representative immunohistochemical images for (**A**)Vimentin, (**B**)epidermal growth factor receptor (EGFR) and (**C**)Ki-67 were positive in tumor cells, with a Ki-67 positivity rate of 90%. Representative immunohistochemical images for (**D**)estrogen receptor (ER),

(**E**)progesterone receptor (PR) and **(F**)human epidermal growth factor receptor 2 (HER2) were negative (scale bar: 300 μm, 100 μm).

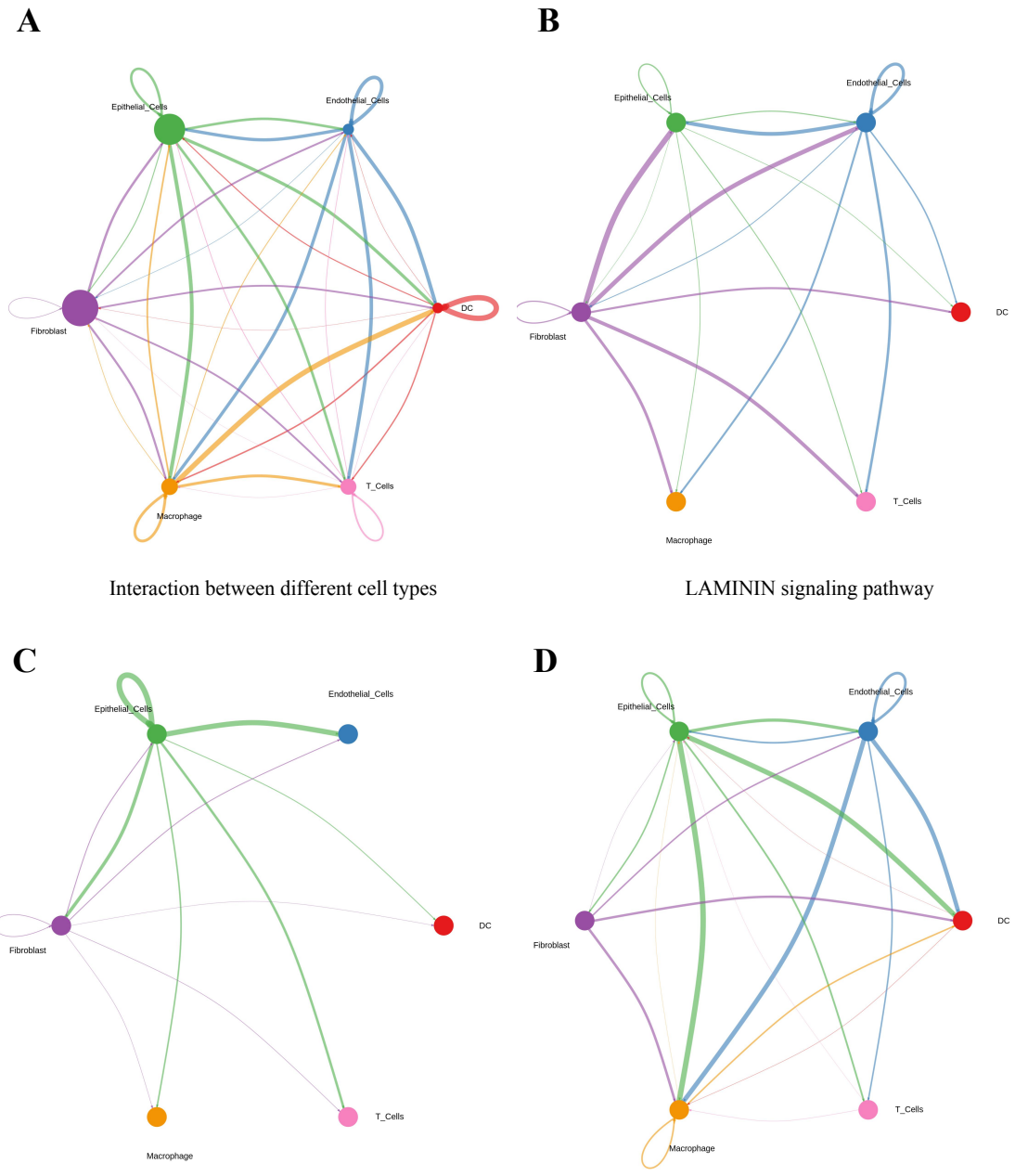

Figure S2 Cell-cell interactions in breast carcinosarcoma. (**A**) Interactions among different cell types (Interaction weights/strength). (**B-D**) Signaling pathways for communication between epithelial cells and mesenchymal cells based on bioinformatics (LAMININ signaling pathway, MK

signaling pathway and APP signaling pathway).

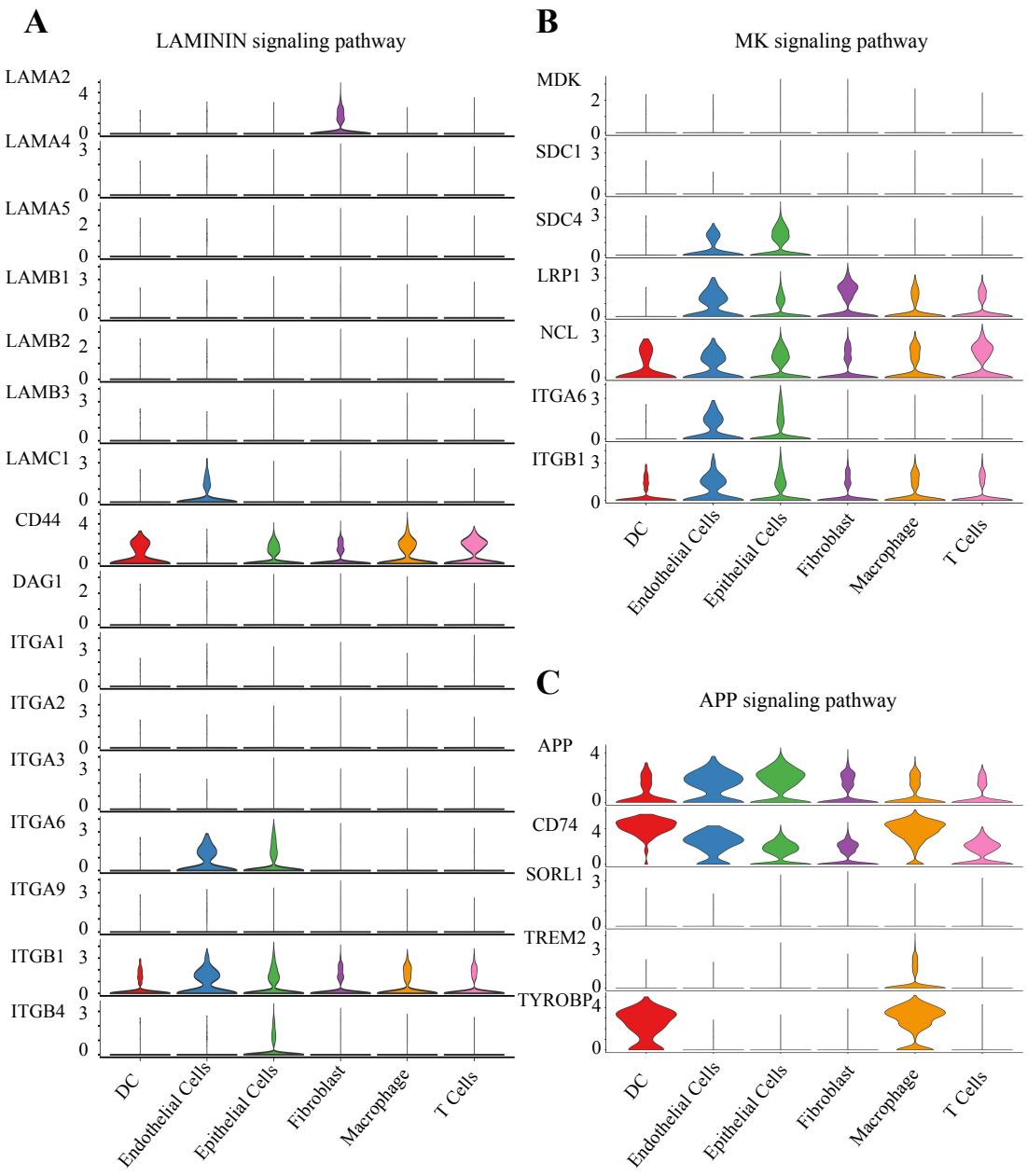

Figure S3 Violin diagram of the expression of marker genes for signaling pathways in communication between epithelial cells and mesenchymal cells. (**A**） LAMININ signaling pathway. (**B**) MK signaling pathway. (**C**) APP signaling pathway.

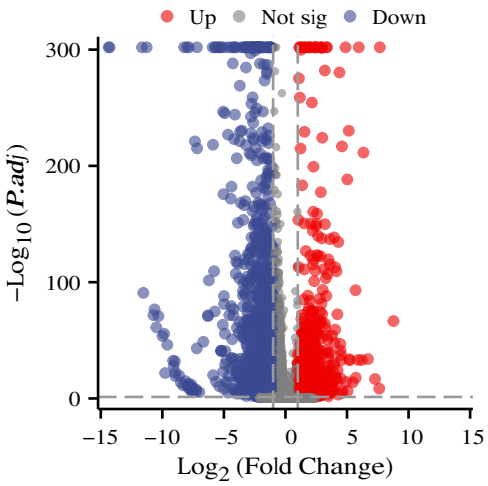

Figure S4 Differential gene expression between breast carcinosarcoma and adjacent normal tissue.

Tumor

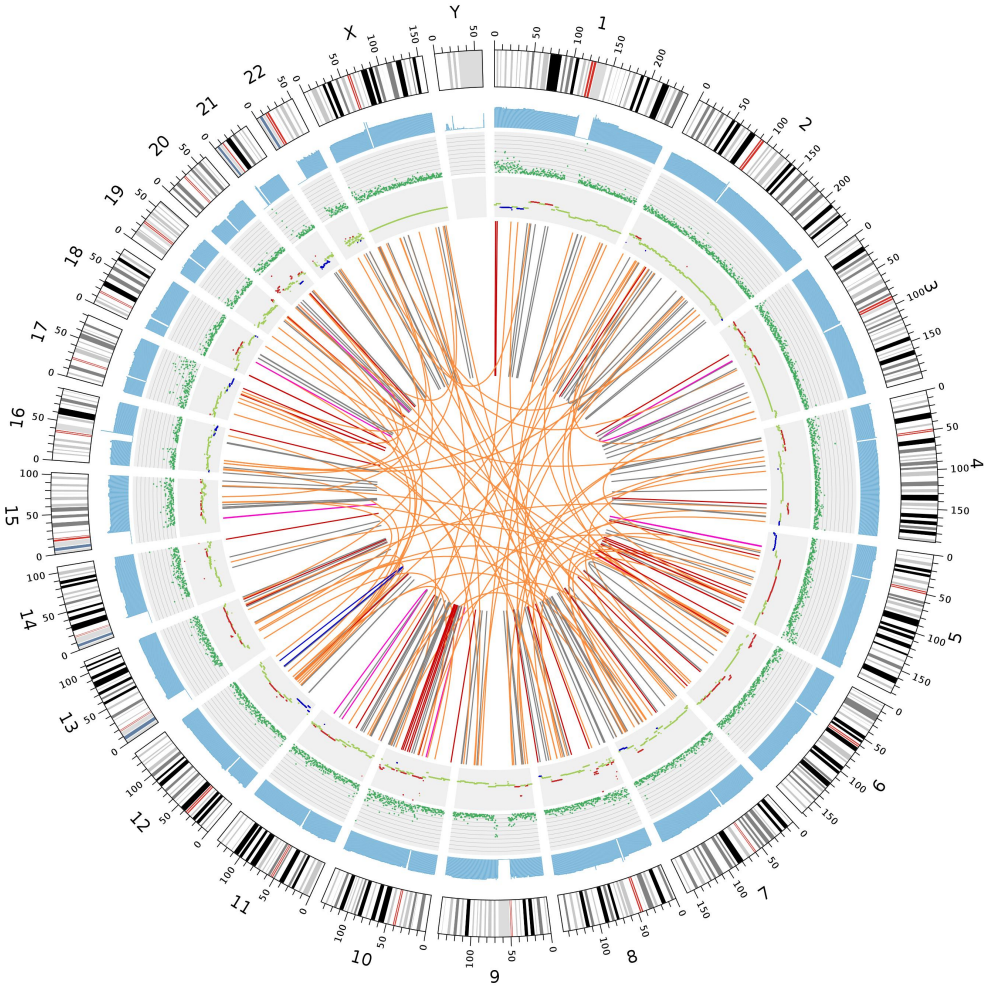

Figure S5 Somatic variation of tumor

Track 1 displays chromosomal ideograms with chromosome numbers and karyotype features; Track 2 shows average sequencing coverage in 0.5 Mbp bins (blue); Track 3 presents SNP/InDel density (green dots, calculated as variant counts per 0.5 Mbp normalized to 1 Mbp); Track 4 visualizes CNV status (red, amplification, blue, deletion, green, normal); Track 5 summarizes filtered SVs at exonic/splice regions with distinct colors for CTX, INS, DEL, ITX and INV. All plot text has been enlarged for optimal print resolution.

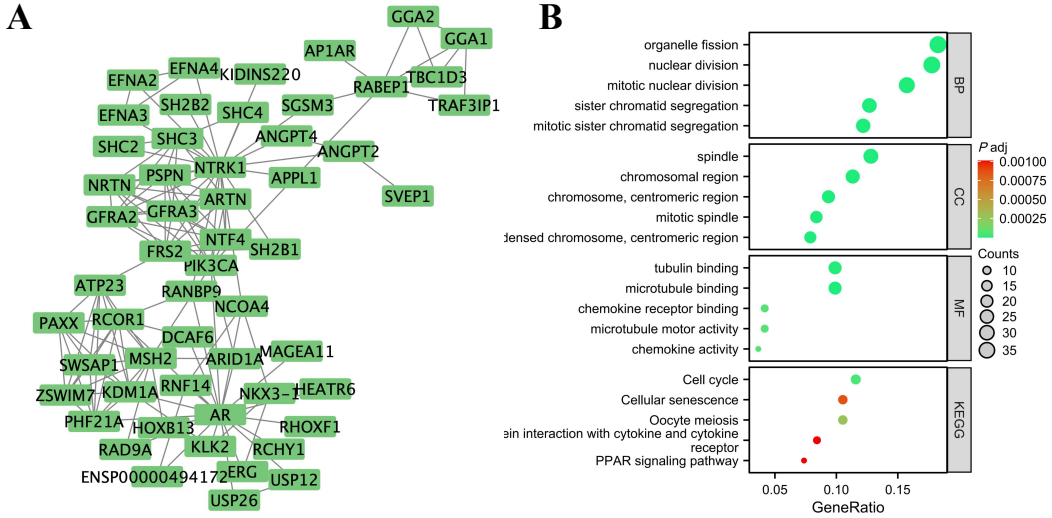

Figure S6 Protein–protein interaction network and functional enrichment analysis. (**A**) Protein–protein interaction network(interacting score at 0.40). (**B**) Functional enrichment analysis.

Table S1 The number of SNPs on different regions of the genome and coding regions. CDS, coding sequences; SNP, single nucleotide polymorphism; UTR, untranslated region; ncRNA, noncoding ribonucleic acid.

Sample	Normal	Tumor
CDS	21825	22745
synonymous_SNP	11104	11555
missense_SNP	10142	10588
stopgain	97	98
stoploss	11	13
startloss	24	25

unknown	447	466
intronic	1231434	1290578
UTR3	31764	33286
UTR5	6004	6284
splicing	621	658
ncRNA_exonic	14304	14967
ncRNA_intronic	229268	239926
ncRNA_UTR3	0	0
ncRNA_UTR5	0	0
ncRNA_splicing	164	173
upstream	22163	23066
downstream	23565	24616
intergenic	1889517	1976564
Total	3469799	3631984

Table S2 The number of SNPs in different regions of the genome. SNP, single nucleotide polymorphism; Het, heterozygote; Hom, homozygote; TS, transformation; TV, transmutation; dbSNP, the single nucleotide polymorphism.

Sample	Normal	Tumor
Total	3469799	3631984
Het	1914267	2107321
Hom	1555532	1524663
ts	2333678	2440829
tv	1136121	1191155
ts/tv	2.05	2.05
dbSNP percentage	3441542(99.19%)	3603255(99.21%)
novel	28257	28729
novel_ts	15711	16177
novel_tv	12546	12552
novel_ts/tv	1.25	1.29

Table S3 The number of INDELs on different regions of the genome and coding regions. CDS, coding sequences; SNP, single nucleotide polymorphism; UTR, untranslated region; ncRNA, noncoding ribonucleic acid.

Sample	Normal	Tumor
CDS	625	643
frameshift deletion	72	78
frameshift insertion	59	58
nonframeshift deletion	206	212
nonframeshift insertion	174	184
stopgain	8	9
stoploss	2	1
startloss	4	2
unknown	100	99
intronic	419781	421786
UTR3	11474	11634
UTR5	1140	1120
splicing	422	419
ncRNA_exonic	2369	2437
ncRNA_intronic	66119	67455
ncRNA_UTR3	0	0
ncRNA_UTR5	2	2
ncRNA_splicing	104	100
upstream	7323	7184
downstream	8717	8528
intergenic	524203	531242
Total	1041959	1052249

Table S4 The number of INDELs in different regions of the genome. INDELs, insertions and deletions; Het, heterozygote; Hom, homozygote; dbSNP, the single nucleotide polymorphism.

Sample	Normal	Tumor
Total	1041959	1052249
Het	796795	819346
Hom	245164	232903
dbSNP percentage	820178(78.71%)	847264(80.52%)
novel	221781	204985

Table S5 The number of Somatic SNVs in different regions of the genome. SNVs, single nucleotide variants; CDS, coding sequence; SNP, single nucleotide polymorphism; UTR, untranslated region;

ncRNA, noncoding ribonucleic acid.

Sample	Tumor
CDS	71
synonymous_SNP	26
missense_SNP	40
stopgain	2
stoploss	1
unknown	2
intronic	2989
UTR3	85
UTR5	23
splicing	4
ncRNA_exonic	52
ncRNA_intronic	760
ncRNA_UTR3	0
ncRNA_UTR5	0
ncRNA_splicing	1
upstream	77
downstream	57
intergenic	5082
Others	9
Total	9210

Table S6 The number of Somatic INDELs in different regions of the genome. INDELs, insertions and deletions; ncRNA, noncoding ribonucleic acid.

Sample	Tumor
CDS	14
frameshift_deletion	7
frameshift_insertion	1
nonframeshift_deletion	6
nonframeshift_insertion	0
stopgain	0
stoploss	0
unknown	0
intronic	3879
UTR3	113
UTR5	14
splicing	6

ncRNA_exonic	30
ncRNA_intronic	603
ncRNA_UTR3	0
ncRNA_UTR5	0
ncRNA_splicing	1
upstream	62
downstream	72
intergenic	4671
Others	1
Total	9466

Table S7 Tumor mutational burden (SNP&InDel)

Sample	Num_mutations	Num_nonsy	TMB_nonsy
Tumor	57	33	0.8919

Table S8 Analysis results of predisposing genes. Ref, reference; Alt, alternative.

Hugo_Symbol	Chromosome	Ref_allele	Alt_allele	Variant_Classification
MITF	3	A	G	Missense_Mutation
BIRC6	2	G	A	Missense_Mutation
SRGAP1	12	C	G	Missense_Mutation
TP53	17	G	T	Missense_Mutation
CTCF	16	A	C	Missense_Mutation
TJP1	15	G	A	Missense_Mutation
STAG2	X	A	T	Missense_Mutation
POLG	15	A	G	Missense_Mutation
CSMD3	8	T	C	Missense_Mutation
NTRK3	15	C	T	Missense_Mutation
CTNNA1	5	C	T	Missense_Mutation
NTRK1	1	G	T	Splice_Site
ARID1B	6	C	T	Missense_Mutation
PAX7	1	G	C	Missense_Mutation
PTPRT	20	T	C	Missense_Mutation
ETV4	17	G	A	Missense_Mutation
CANT1	17	C	T	Missense_Mutation
BCORL1	X	G	A	Missense_Mutation
BRCA1	17	C	T	Missense_Mutation
CSDE1	1	A	C	Missense_Mutation
KMT2A	11	G	T	Missense_Mutation
TAF1	X	G	C	Missense_Mutation
KDR	4	G	A	Missense_Mutation

ZFHX3	16	CGCCACCGCCGCC GCCGCCGCCACT	C	In_Frame_Del
SVEP1	9	TAAA	T	Splice_Site
RB1	13	TA	T	Splice_Site
MSH2	2	TA	T	Splice_Site
IL6ST	5	T	TAA	Splice_Site
RABEP1	17	GTAGTGTTTGGAA TTTTCTGTTCATA	G	Splice_Site
TSC1	9	G	GAAA	Splice_Site
MED23	6	T	TAAA	Splice_Site
POLE	12	G	GA	Splice_Site
AR	X	TGCA	T	In_Frame_Del
ZFHX3	16	T	TTGCTGCTGC	In_Frame_Ins

Table S9 Analysis results of driving genes. Ref, reference; Alt, alternative.

Hugo_Symbol	Chromosome	Ref_allele	Alt_allele	Variant_Classification
HOXA11	7	G	C	Missense_Mutation
RALGAPA1	14	AT	A	Frame_Shift_Del