Method

1 Experimental Procedure

1.1 Evaluation of DNA quality

The quality of isolated genomic DNA was verified by using these three methods in combination:

1) DNA degradation and contamination were monitored on 1% agarose gels;
2) DNA concentration was measured by Qubit® DNA Assay Kit in Qubit® 3.0 Flurometer(Invitrogen, USA).

1.2 Library Preparation

The exome sequences were efficiently enriched from 0.4 μg genomic DNA using Agilent liquid capture system (Agilent SureSelect Human All Exon V6) according to the manufacturer' s protocol. Firstly, qualified genomic DNA was randomly fragmented to an average size of 180-280bp by Covaris S220 sonicator. Remaining overhangs were converted into blunt ends via exonuclease polymerase activities. Secondly, DNA fragments were end repaired and phosphorylated, followed by A-tailing and ligation at the 3' ends with paired-end adaptors. DNA fragments with ligated adapter molecules on both ends were selectively enriched in a PCR reaction. After PCR reaction, libraries hybridize with liquid phase with biotin labeled probe, then use magnetic beads with streptomycin to capture the exons of genes. Captured libraries were enriched in a PCR reaction to add index tags to prepare for sequencing. Products were purified using AMPure XP system (Beckman Coulter, Beverly, USA), DNA concentration was measured by Qubit®3.0 Flurometer (Invitrogen, USA), libraries were analyzed for size distribution by NGS3K/Caliper and quantified by real-time PCR (3 nM). At last, DNA library were sequenced on Illumina for pairedend 150bp reads

1.3 Clustering & Sequencing

2 Bioinformatics Analysis Pipeline

2.1. Data Quality Control

2.1.1. Raw data

The original fluorescence image files obtained from Illumina platform are transformed to short reads (Raw data) by base calling and these short reads are recorded in FASTQ format, which contains sequence information and corresponding sequencing quality information.

2.1.2. Evaluation of data (Data quality control)

Sequence artifacts, including reads containing adapter contamination, low-quality nucleotides and unrecognizable nucleotide (N), undoubtedly set the barrier for the subsequent reliable bioinformatics analysis. Hence quality control is an essential step and applied to guarantee the meaningful downstream analysis.

All the downstream bioinformatics analyses were based on the high quality clean data, which were retained after these steps. At the same time, QC statistics including total reads number, raw data, raw depth, sequencing error rate and percentage of reads with Q30 (the percent of bases with phred-scaled quality scores greater than 30) were calculated and summarized.

2.2. Reads Mapping to Reference Sequence

Valid sequencing data is mapped to the reference genome (GRCh37/hg19/ GRCh38) by BurrowsWheeler Aligner (BWA) software to get the original mapping result in BAM format. Subsequently, Samtools and Sambamba are spectively utilized to sort bam files, do duplicate-marking to generate final bam file. If one or one pair read(s) has multiple mapping positions, the strategy adopted by BWA are to select the best one, if there are multi best mapping position, we randomly pick one. Mapping step is very difficult due to mismatches, including true mutation and sequencing error, and duplicates resulted from PCR amplification. These duplicate reads are uninformative and shouldn' t be considered as evidence for variants. Sambamba is employed to mark these duplicates so that we will ignore them in the following analysis.

2.3. Variant detection

SAMtools mpileup and bcftools were used to do variant calling and identify SNP, InDels.

2.4. Somatic mutation calling

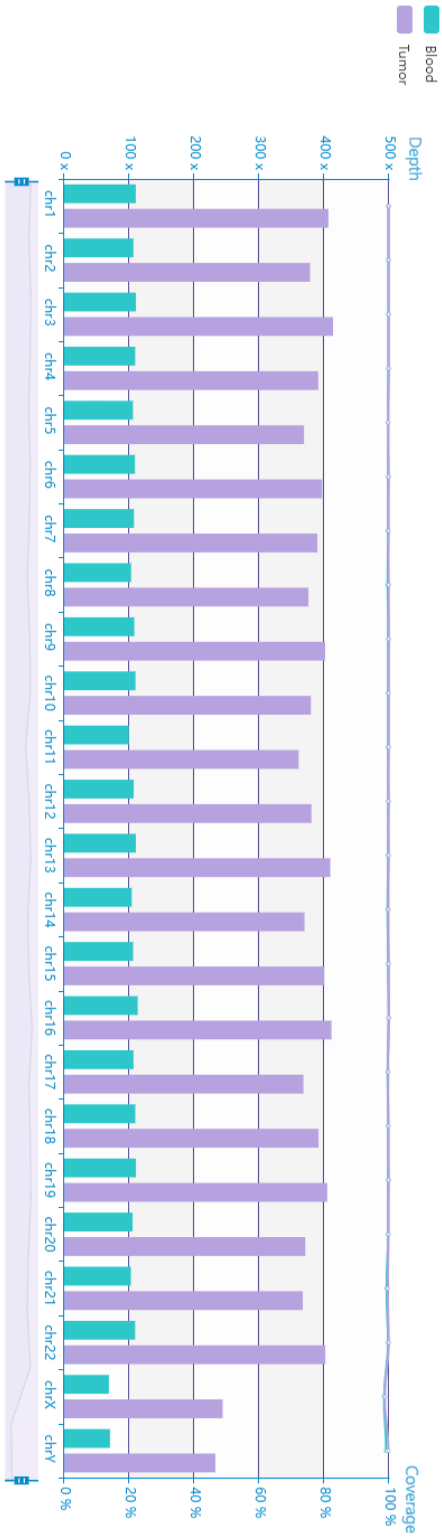
The somatic SNV was detected by muTect, the somatic InDel by Strelka, and Control-FREEC was used to detect somatic CNV

2.5. Annotation

ANNOVAR is performed to do annotation for VCF file obtained in the previous step. The variant position, variant type, conservative prediction and other information are obtained at this step through a variety of databases, such as dbSNP, 1000 Genome, esp6500, GnomAD, CADD, HGMD and COSMIC, and so on. Since we are interested in exonic variants, gene transcript annotation databases, such as Consensus CDS, RefSeq, Ensemble and UCSC, are also applied for annotation to determine amino acid alternation. Functional annotation is very important because the link between genetic variation and cancer can be found. In this step, the databases mainly include GO, KEGG, Reactome, Biocarta, PID, and so on.

Result

1 Sequencing depth and coverage



2 The outcome of somatic mutation

2.1 Single nucleotide variant(SNV)

Sample	Xuezheng_Tumor
CDS	3
synonymous_SNP	1
missense_SNP	2
stopgain	0
stoploss	0
unknown	0
intronic	53
UTR3	6
UTR5	1
splicing	2
ncRNA_exonic	7
ncRNA_intronic	10
ncRNA_UTR3	0
ncRNA_UTR5	0
ncRNA_splicing	1
upstream	3
downstream	2
intergenic	23
Others	0
Total	111

2.2 Insertion and Deletion(InDel)

Sample	Xuezheng_Tumor
CDS	0
frameshift_deletion	0
frameshift_insertion	0
nonframeshift_deletion	0
nonframeshift_insertion	0
stopgain	0
stoploss	0
unknown	0
intronic	0
UTR3	0
UTR5	0
splicing	0
ncRNA_exonic	0
ncRNA_intronic	0
ncRNA_UTR3	0
ncRNA_UTR5	0

ncRNA_splicing	0
upstream	0
downstream	0
intergenic	0
Others	0
Total	0

SNV list

Priority	L	L	L	L	L	L	L	L	L	L	L	L	L
CHROM	chr6	chr1	chr7	chr9	chr15	chr12	chr22	chr2	chr7	chr15	chr17	chr17	chr11
POS	55759124	161513933	65763276	41648075	20408438	15934	24657860	129980059	49960980	101975947	41105795	41105831	1097926
ID	.	rs409763	rs36809443 3	.	rs11235571 7	rs7970380 49	rs2001116 41	rs6750772	rs9639891	rs1998372 86	rs7565458 01	rs7539260 8	.
REF	T	G	T	T	A	T	A	G	C	C	A	A	C
ALT	C	A	C	C	G	G	G	C	G	G	G	C	G
QUAL	.	.	.	.	.	.	.	.	.	.	.	.	.
FILTER	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
GeneName	BMP5	FCGR2A	CCT6P1	FAM27E2	HERC2P3	LOC100288778	POM121L10P	RAB6C-AS1	RNF216P1	WASH3P	KRTAP4-9	KRTAP4-9	MUC2
Description	Bone morphogenetic protein-5	Fc fragment of IgG, low affinity IIa, receptor for (CD32)	.	.	.	.	.	.	.	.	.	.	Mucin 2, intestinal/tracheal
Func	splicing	splicing	ncRNA_splicing	ncRNA_exonic	ncRNA_exonic	ncRNA_exonic	ncRNA_exonic	ncRNA_exonic	ncRNA_exonic	ncRNA_exonic	exonic	exonic	exonic
ExonicFunc	.	.	.	.	.	.	.	.	.	.	missense SNV	missense SNV	synonymous SNV
Gene	NM_021073	NM_001136219,NM_021642	NR_003110	NR_103714	NR_036432	NR_130745	NR_024593	NR_036537	NR_015449,NR_023384,NR_023385	NR_003659	NM_001146041	NM_001146041	NM_002457

GeneDetail	NM_021073:exon6:c.1105-9A>G	NM_001136219:exon6:c.780+1G>A;NM_021642:exon6:c.777+1G>A	NR_003110:exon12:c.1372-5T>C	.	.	.	.	.	.	.	.	.	.
AAChange	.	.	.	.	.	.	.	.	.	.	KRTAP4-9:NM_001146041:exon1:c.A407G:p.K136R	KRTAP4-9:NM_001146041:exon1:c.A443C:p.N148T	MUC2:NM_002457:exon31:c.C7614G:p.T2538T

Genecode33	ENST00000370830.4	ENST00000461298.1,ENST00000367972.8,ENST00000271450.11,ENST00000537821.2,ENST00000459885.1,ENST00000471026.5,ENST00000467525.5,ENST0	ENST00000633579.1,ENST00004345 0.6,ENST00004422 66.2,ENST00004431 08.6	.	ENST00000437318.5,ENST00004284 53.5,ENST00000426501.5,ENST00004305 98.5	ENST00000706.3	ENST00000436994.1	ENST00000 10061.4,ENST00000624615.1,ENST00000412425.1	ENST00000471244.1,ENST00004053 96.2,ENST00000404006.5,ENST00004071 84.5,ENST00000404404.5,ENST00004949 47.5,ENST00000403969.5	ENST00000 58784.5,ENST00000560956.1,ENST00000557932.5	ENST00000 17453.1,ENST00000391415.1	ENST00000 39141 5.1,ENST00006 17453.1	ENST00000613187.1,ENST00000361558.7
cpgIslandExt	.	.	.	.	.	.	.	CpG:54	.	.	.	.	.
cytoBand	6p12.1	1q23.3	7q11.21	9p11.2	15q11.1	12p13.33	22q11.23	2q21.1	7p22.1	15q26.3	17q21.2	17q21.2	11p15.5
wgRna	.	.	.	.	.	.	.	.	.	.	.	.	.
genomicSuperDups	.	Score=0.97162;Name=chr1:161591131	Score=0.98027 1;Name=chr7:65062896	Score=0.99104 7;Name=chr9:61857542	Score=0.95094 5;Name=chr2:133376270	Score=0.98263 1;Name=chrX:15602021 6	Score=0.98292 2;Name=chr22:24250602	Score=0.96443 5;Name=chr2:131346148	Score=0.93379 4;Name=chr1:279617 45	Score=0.99 1969;Name=chr1:10000	Score=0.94 5624;Name=chr17:41081136	Score=0.94 5624;Name=chr17:41081136	Score=0.925343;Name=chr11:1095529
Repeat	Score=22;Name="418312 8:Simple_repeat"	.	.	.	.	Score=12;Name="933443:Simple_repeat"	.	.	.	.	.	.	.

avsnp150	.	rs409763	rs36809443	.	rs11235717	rs79703804 9	rs20011641	rs6750772	rs9639891	rs199837286	rs7565458 01	rs7539 2608	.
clinvar_20190305	.	.	.	.	.	.	.	.	.	.	.	.	.
gwasCatalognew	.	.	.	.	.	.	.	.	.	.	.	.	.
1000g2015aug_Chinese	.	.	.	.	.	.	.	.	0.24	.	.	.	.
1000g2015aug_eas	.	.	.	.	.	.	.	.	0.25	.	.	.	.
1000g2015aug_all	.	.	.	.	.	.	.	.	0.14	.	.	.	.
esp6500siv2_all	.	.	.	.	.	.	.	.	.	.	.	.	.
GnomAD_exome_AF_popmax	.	0	.	.	.	.	.	.	.	##########		0.001	.
GnomAD_exome_AF_eas	.	0	.	.	.	.	.	.	.	3E−04	3E−04	0.001	.
GnomAD_genome_AF_popmax	.	0	0	.	0.31	.	0	0.119	0.11	0.028	0.001	0.005	.
GnomAD_genome_AF_eas	.	0.01	0	.	0.3	.	0	0.136	0.26	0.007	0.005	0.005	.
NovoDb_WES	.	0.02	.	.	.	.	.	.	.	.	.	.	.
NovoDb_WGS	.	.	.	.	0.2	.	.	0.109	0.25	.	0.004	0.004	.
dbscSNV_SCORE	0.0003,0	1.0000,0.742	.	.	.	.	.	.	.	.	.	.	.
Interpro_domain	.	.	.	.	.	.	.	.	.	.	.	.	.
SIFT	.	.	.	.	.	.	.	.	.	.	1.0,T	0.899,T	.
Polyphen2_HDIV	.	.	.	.	.	.	.	.	.	.	0.0,B	0.0,B	.
Polyphen2_HVAR	.	.	.	.	.	.	.	.	.	.	0.0,B	0.0,B	.
LRT	.	.	.	.	.	.	.	.	.	.	0.421,N	0.002,N	.
MutationTaster	.	0.999,D	.	.	.	.	.	.	.	.	1,N	1,N	.
MutationAssessor	.	.	.	.	.	.	.	.	.	.	1.29,N	3.205,N	.
FATHMM	.	.	.	.	.	.	.	.	.	.	2.2,T	6.59,T	.

CADD_raw	.	1.752,14.71	.	.	.	.	.	.	.	.	−2.089,0.001	−2.107,0.001	.	
SiPhy_29way_logOdds	.	7.87	.	.	.	.	.	.	.	.	7.549	2.384	.	
gerp++gt2	.	2.16	2.1	.	.	2.51	.	.	.	.	.	.	.	
REVEL	.	.	.	.	.	.	.	.	.	0.194	0.013	0.029	.	
cosmic70	.	.	.	.	.	.	.	.	.	ID=COSM3932020;OCCURRENCE=3(urinary_tract)	.	.	.	
ExAC_ALL	.	0	.	.	.	.	.	.	.	.	2E−04	0.001	.	
ExAC_EAS	.	0.01	.	.	.	.	.	.	.	.	0.001	4E−04	.	
INFO	DB;SOMATIC;VT=SNP	DB;SOMATIC;VT=SNP	SOMATIC;VT=SNP	SOMATIC;VT=SNP	DB;SOMATIC;VT=SNP	SOMATIC;VT=SNP	DB;SOMATIC;VT=SNP	DB;SOMATIC;VT=SNP	DB;SOMATIC;VT=SNP	DB;SOMATIC;VT=SNP	SOMATIC;VT=SNP	SOMATIC;VT=SNP	SOMATIC;VT=SNP	
FORMAT	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	GT:AD:BQ:DP:FA:SS	
Xuezheng_Blood	0:27,0:.:27:0.00:0	0:37,1:.:38:0.026:0	0:58,1:.:59:0.017:0	0:8,0:.:8:0.00:0	0:14,0:.:14:0.00:0	0:12,0:.:10:0.00:0	0:15,0:.:15:0.00:0	0:40,1:.:41:0.024:0	0:9,0:.:9:0.00:0	0:10,0:.:7:0.00:0	0:35,1:.:37:0.028:0	0:40,1:.:41:0.024:0	0:25,0:.:16:0.00:0	
Xuezheng_Tumor	0/1:187,22:37:210:0.105:2	0/1:163,18:37:182:0.099:2	0/1:115,12:37:127:0.094:2	0/1:0,15:37:15:1.00:2	0/1:22,5:37:27:0.185:2	0/1:13,5:37:18:0.278:2	0/1:35,8:37:43:0.186:2	0/1:164,13:37:177:0.073:2	0/1:29,10:37:39:0.256:2	0/1:43,9:35:52:0.173:2	0/1:112,13:36:125:0.104:2	0/1:134,22:36:156:0.141:2	0/1:46,7:37:53:0.132:2	
Ori_REF	T	G	T	T	A	T	A	G	C	C	A	A	C	
Ori_ALT	C	A	C	C	G	G	G	C	G	G	G	C	G	
shared_hom	0	0	0	0	0	0	0	0	0	0	0	0	0	0
shared_het	0	0	0	0	0	0	0	0	0	0	0	0	0	0

OMIM	Bone morphogenetic protein−5, 112265	{Lupus nephritis, susceptibility to}, 152700 (3), Autosomal dominant; {Malaria, severe, susceptibility to}, 611162 (3); {Pseudomonas aeruginosa,	.	.	.	.	.	.	.	.	.	.	Mucin 2, intestinal/tracheal, 158370
GWAS_Pubmed_pValue	.	.	.	.	.	.	.	.	.	.	.	.	.
HGMD_Disease_ID	.	.	.	.	.	.	.	.	.	.	.	.	.
HGMD_Mutation_URL	.	.	.	.	.	.	.	.	.	.	.	.	.

PubMedID	11580864|2263636|10677507|1427904|13393316|8630036|5319761	19915573|15367919|2565886|2965389|2414215|1683644|19011614|22608500|9843982|1301170|2521732|2139735|17005690|8636449|2971615|19965803|10675363|264362	.	.	.	.	.	.	.	.	.	.	9722984|1980995|2703501|1985113|16754877|17058067|2265829|8975711|15081123|24072822|7864825|10636731|1885763|11872843
GO_BP	.	.	.	.	.	.	.	.	.	.	.	.	DIGESTION;SYSTEM_PR
GO_CC	.	.	.	.	.	.	.	.	.	.	.	.	PROTEINACEOUS_EXTRACELLULAR_MATRIX;EXTRACELLULAR_REGION;EXTRACELLULAR_REGION_PA
GO_MF	.	.	.	.	.	.	.	.	.	.	.	.	STRUCTURAL_MOLECULE_ACTIVITY;EXTRACELLULAR_MATRIX_S

KEGG_PATHWAY	KEGG_HEDGEHOG_SIGNALING_PATHWAY;KEGG_TGF_BETA_SIGNALING_PATHWAY	KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS;KEGG_LEISHMANIA_INFECTION;KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	.	.	.	.	.	.	.	.	.	KEGG_VIBRIO_CHOLERAE_INFECTION
PID_PATHWAY	.	PID_PTP1BPATHWAY;PID_INTEGRIN2_PATHWAY	.	.	.	.	.	.	.	.	.	.
BIOCARTA_PATHWAY	BIOCARTA_ALK_PATHWAY	.	.	.	.	.	.	.	.	.	.	.

REACTOME_PATHWAY	REACTOME_O_LINKED_GLYCOSYLATION_OF_MUCINS;REACTOME_TERMINATION_OF_O_GLYCAN_BIOSYNTHESIS;REACTOME_METABOLISM_OF_PROTEINS;R