**全基因组测序**

**结构变异分析报告**

|  |  |
| --- | --- |
| **客户单位:** | **{{ proj\_info.customer\_institute }}** |
| **报告单位:** | **北京希望组生物科技有限公司** |
| **项目编号:** | **{{ proj\_info.proj\_id }}** |
| **报告撰写人:** | **{{ proj\_info.bioinfo\_engineer }}** |
| **审 核 人:** | **{{ proj\_info.supervisor }}** |
| **报告日期:** | **{{ proj\_info.report\_date }}** |

目录

[1 项目概况 1](#_Toc523334331)

[2 技术背景 1](#_Toc523334332)

[2.1 Nanopore测序技术简介 1](#_Toc523334333)

[2.2 结构变异简介 2](#_Toc523334334)

[3 项目流程 4](#_Toc523334335)

[3.1 实验流程 4](#_Toc523334336)

[3.1.1 DNA样本检测 5](#_Toc523334337)

[3.1.2 构建测序文库 5](#_Toc523334338)

[3.1.3 上机测序 6](#_Toc523334339)

[3.2 生物信息分析 7](#_Toc523334340)

[3.2.1 Nanopore测序下机数据统计过滤 7](#_Toc523334341)

[3.2.2 结构变异检测 7](#_Toc523334342)

[3.2.3 结构变异注释 7](#_Toc523334343)

[3.2.4 疾病相关基因分析 7](#_Toc523334344)

[4 分析结果 8](#_Toc523334345)

[4.1 数据质控和统计 8](#_Toc523334346)

[4.2 序列比对和统计 9](#_Toc523334347)

[4.3 结构变异检测和统计 9](#_Toc523334348)

[4.4 结构变异注释 10](#_Toc523334349)

[4.5 相关基因调研 14](#_Toc523334350)

[4.6 重点关注的SV 14](#_Toc523334351)

[5 文件交付及格式说明 15](#_Toc523334352)

[5.1 软件及参数 15](#_Toc523334353)

[5.2 文件列表 15](#_Toc523334354)

[参考文献 16](#_Toc523334355)

# 1 项目概况

# 2 技术背景

## 2.1 Nanopore测序技术简介

从早期的Sanger测序到第二代高通量测序、再到现在的单分子测序技术（第三代高通量测序），DNA测序技术一直推动着生命科学研究的发展。

当第二代高通量测序技术进入成熟阶段后，读长过短、PCR扩增带来的偏向性等问题开始日益凸显；作为基因组学上新的转折点，单分子实时测序技术以第三代高通量测序技术（Third-generation Sequencing ）的身份开始进入科研应用。

**MinION：首个纳米孔测序仪**

Oxford Nanopore 成功研发了世界上首例的纳米孔 DNA 测序仪，MinION。MinION 是一款便携式、实时、长读长、低成本设备，旨在让任何人，不论是在科学研究、教育，还是一系列实际应用中，例如疾病/病原体监测、环境监测、食物链监测、自我量化，甚至微重力生物学等领域，都能进行便捷的生物分析。MinION一次可以使用一个流动槽（Flow cell），单个Flow cell上有512个纳米孔通道，平均48h内可以产生10-20 Gbp的测序数据。

自 2015 年开始销售以来，MinION 已被分布在70 余个国家的科学家们广泛使用，助力于他们的科研工作，实现了在传统实验室内和实地环境下的众多应用。

**GridION X5：通量更高的第三代测序平台**

GridION X5是一款紧凑型台式系统，它是基于目前的MinION Flow cell设计的，可同时运行和分析多达五个Flow cells，平均48h内可以产生50-100 Gbp的测序数据。目前的化学试剂和软件版本能够在GridION X5运行期间生成最高50 Gb的数据，并且计算模块能够实时分析数据。希望组是国内最先引进GridION X5测序平台的公司（ONT官网报道），于2017年9月正式提供GridION测序服务。

**PromethION：目前通量最高的第三代测序平台**

PromethION是专门为大型测序研究项目或测序服务商提供的高通量纳米孔测序仪，可以同时运行48个Flow cells，每个Flow cell上有3000个纳米孔通道，单个Flow cell平均48h内可以产生50-120 Gbp的数据。PromethION目前开放了24个Flow cells的插槽，平均48h内可以产生1.2-6 Tbp的测序数据。2018年7月2日，希望组正式成为国内首家PromethION测序服务提供商（ONT官网报道）。



Figure 2.1 Oxford Nanopore测序平台

依托GridION平台和PromethION平台，希望组在承诺高标准的交付指标的同时，将进一步大幅压缩项目服务周期以满足呈指数增长的三代测序服务市场。

## 2.2 结构变异简介

结构变异（SV）指的是长度大于50 bp的基因组变异，主要包括插入（insertion）、缺失（deletion）、重复（duplication）、倒位（inversion）、易位（translocation）1。如图Figure 2.2，其中插入可以分为新序列插入（novel sequence insertion）和转座元件插入（Mobile-element insertion），重复可以分为串联重复（tandem duplication）和散在重复（interspersed duplication）。此外还有复杂类型的结构变异（complex structure variation）1。与SNP和小的indel影响的碱基数目相比，SV影响的碱基数目更多2。结构变异与多种人类疾病相关，如癌症2、自闭症3和阿尔兹海默症4等。

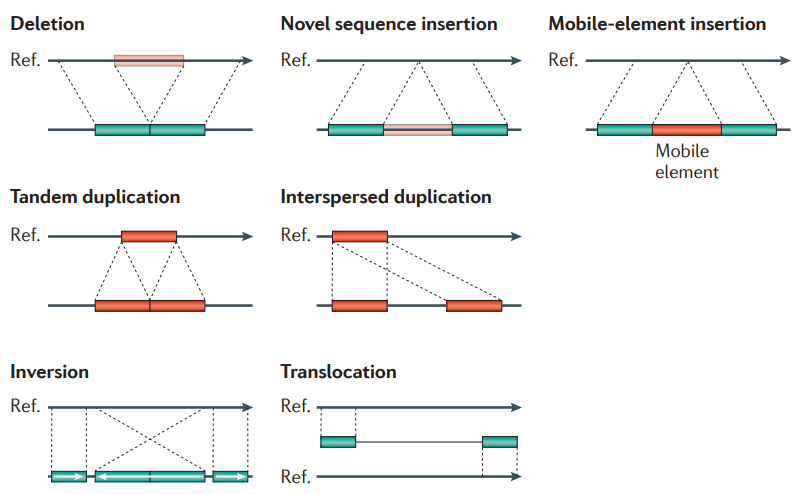


Figure 2.2 结构变异的类型1

目前二代测序检测SV的方法主要有四种5，分别是依赖测序深度（Read Depth，RD）、Paired-reads（PR） mapping、Split-reads（SR） mapping和从头组装（AS）（Figure 2.3）。由于二代测序的读长较短（100-300 bp），存在错误比对、GC偏好性、无法跨越长片段重复区等问题，二代测序在检测结构变异时的敏感性低（从10%到70%），同时假阳性率高（高达89%）6。

三代测序技术可以运用测序深度（Read Depth，RD）、Split-reads（SR） mapping和从头组装（AS）的方法检测SV。相对于二代测序的短reads，三代测序的读长更长（Pacbio和Nanopore测序的平均读长均超过10 kp），因而可以弥补许多二代测序的缺陷，检测SV的敏感性和阳性率更高（10-30X的人Pacbio和Nanopore检测的敏感性约为80%，阳性率大于80%）6。同时三代测序可以检测复杂类型的结构变异。最近，研究者们运用三代测序发现了可以导致糖原贮积症7和Carney综合征的结构变异8，发现了SKBR3细胞系中HER2基因的复杂duplication现象2。

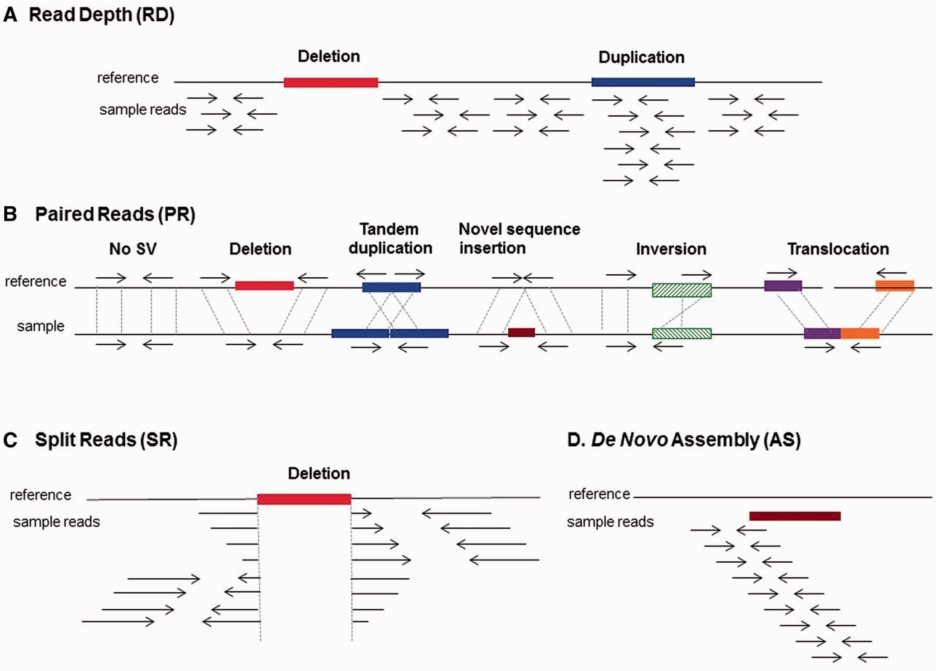


Figure 2.3 高通量测序检测SV的方法

# 3 项目流程

## 3.1 实验流程

第三代测序技术对样品DNA的要求非常高，希望组根据客户提供的DNA进行严格的样品检测，建库以及上机测序，每个环节都严格把控，以求测序结果的准确性和长度得到保障。

希望组采用磁珠提取、Qiagen试剂盒、Binano提取等多种方法提取高质量的DNA，构建Nanopore测序文库，利用Oxford Nanopore Technology测序仪GridION或PromethION对DNA进行单分子实时测序，获得原始测序数据。

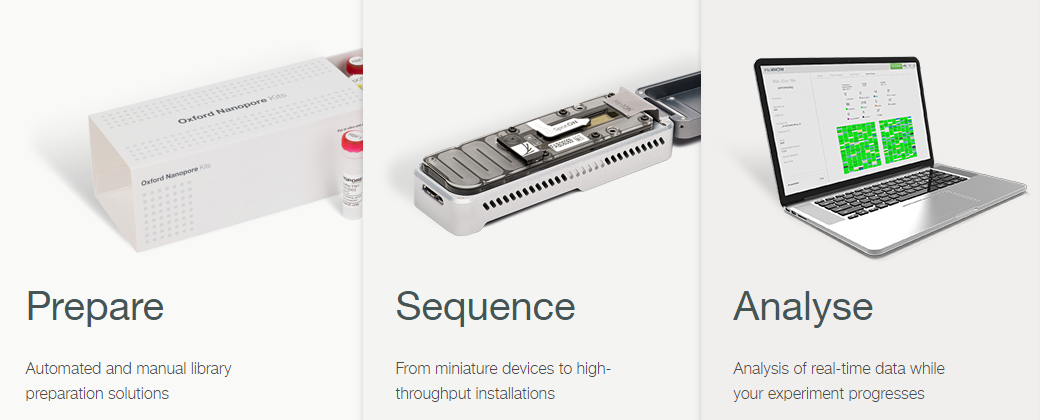


Figure 3.1 Nanopore测序分析流程

### 3.1.1 DNA样本检测

我们采用以下4种方法检测DNA是否合格：

（1）样品的外观性状是否含有异物

（2）0.75%琼脂糖电泳检测是否样品有降解以及DNA片段大小

（3）Nanodrop检测DNA纯度(OD260/280在1.8；260/230在2.0-2.2之间)

（4）Qubit精确地对DNA进行定量

### 3.1.2 构建测序文库

样本质检合格后，用Megaruptor 对基因组DNA 进行随机打断；利用磁珠富集、纯化大片段DNA；使用BluePippin全自动核酸回收仪将大片段进行切胶回收；对片段化的DNA 进行损伤修复；纯化后在DNA 片段两端进行末端修复并进行加A连接反应；纯化后，连接测序接头，最后用Qubit 精确地对建好的DNA文库进行定量检测。

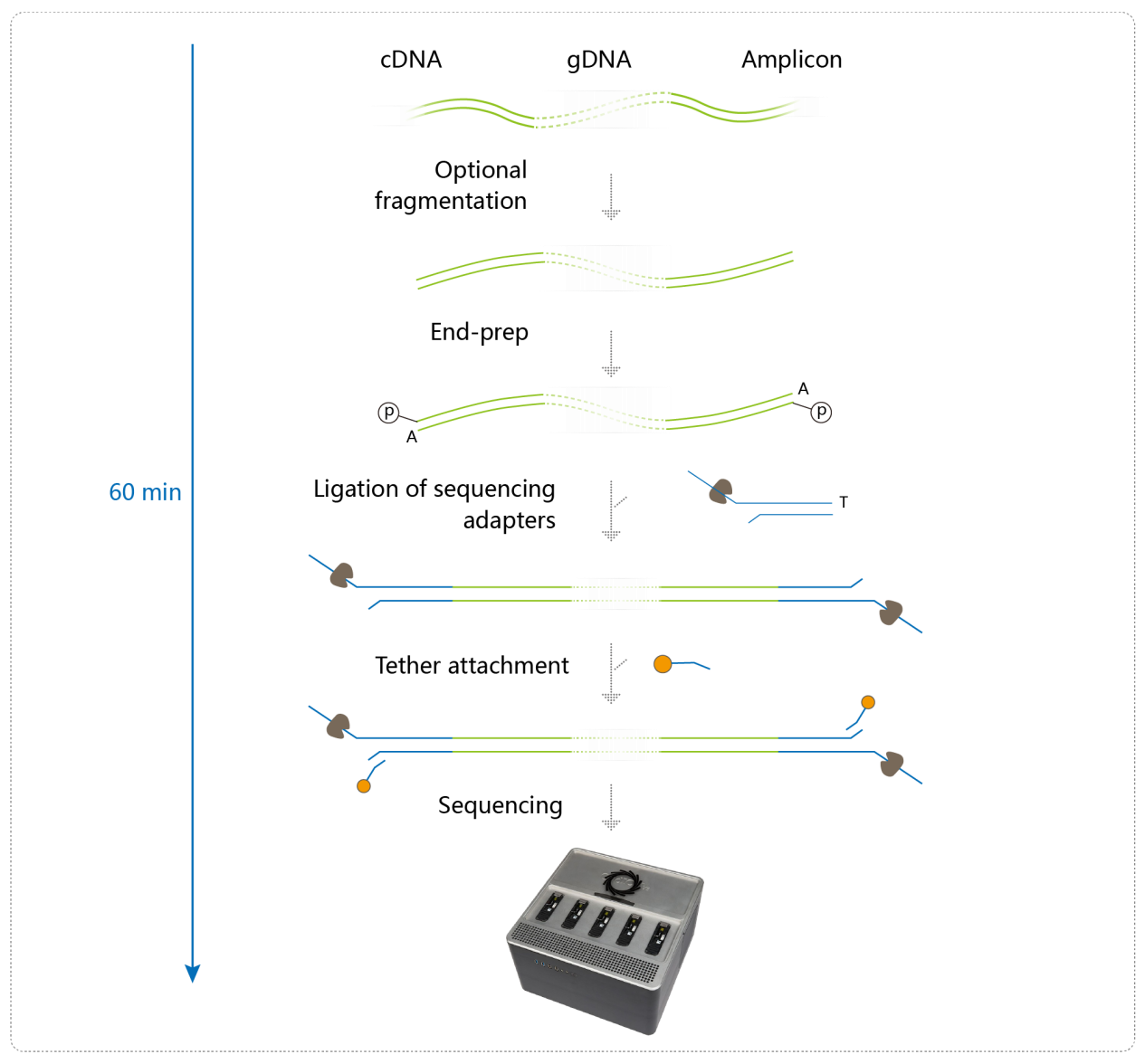


Figure 3.2 Nanopore 测序建库流程

### 3.1.3 上机测序

构建完成的测序文库，使用GridION或PromethION进行测序。

## 3.2 生物信息分析

### 3.2.1 Nanopore测序下机数据统计过滤

对Nanopore测序下机数据进行统计并按照reads平均qScore值大于或等于7进行过滤。

### 3.2.2 结构变异检测

我们使用NGMLR-Sniffles流程进行结构变异检测。该软件流程可以分析的主要结构变异类型有缺失（DEL），插入（INS），重复（DUP），倒位（INV），易位（TRA）。

### 3.2.3 结构变异注释

采用ANNOVAR进行结构变异的基因、转录因子结合位点、基因组大片的重复（segmental duplication）、CpG岛、MicroRNA结合位点等注释。按照结构变异与公共数据库结构变异的相互重叠度大于或等于50%（插入（INS）与公共数据库的INS的距离小于或等于1000bp），进行1000 genome phase3、DGV gold standard CNV、dbVar nstd37和Decipher结构变异注释。按照上述的标准，搜索希望组收录的Nanopore结构变异数据库，并进行注释。

### 3.2.4 疾病相关基因分析

根据疾病的表型调研OMIM、HPO和Clinvar等数据库，并阅读相关文献，确定疾病相关基因。根据上述注释结果，重点分析疾病相关基因的结构变异。

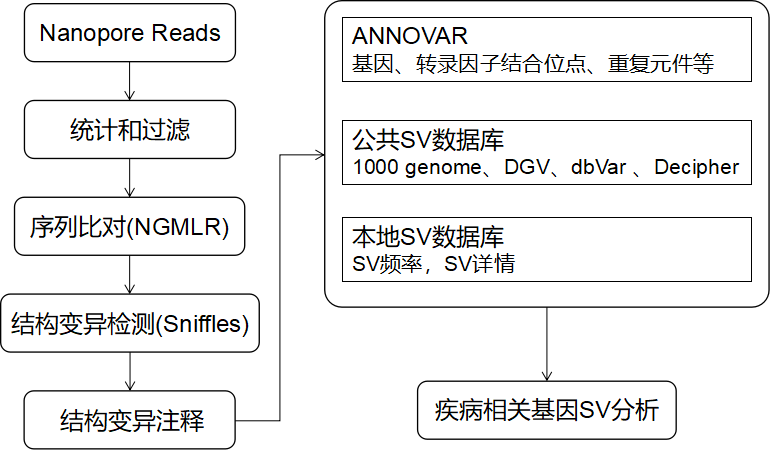


Figure 3.3 生物信息分析流程

# 4 分析结果

## 4.1 数据质控和统计

Nanopore测序的下机数据为fast5格式，主要储存了reads的电信号、名称、测序时间、测序所用的试剂盒型号等信息。下机数据首先经过Nanopore提供的MinKnow软件进行basecalling，将电信号解析为碱基序列，并输出常见的fastq格式文件。我们对fastq文件进行过滤，去除平均质量小于7的reads。

通过质控的数据统计如Table 4.1，Reads长度分布图如Figure 4.1。

Table 4.1 下机数据统计表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sample ID | Total Bases | Reads num | Mean length | Max length | N50 | Depth |
| {{proj\_info.sample\_id}} | {{fastq\_qc.num\_bases}} | {{fastq\_qc.num\_of\_reads}} | {{fastq\_qc.average\_length }} | {{fastq\_qc.max\_length}} | {{fastq\_qc.N50}} | {{fastq\_qc.depth}} |

注：Total Bases：总碱基数目；Reads num：测序reads数目；Mean length: reads平均长度；Max length: reads最大长度；N50：reads 长度N50，指的是按照reads的长度从大到小排序，当reads累计碱基数目大于或等于测序总碱基数目50%时read的长度，Depth：测序深度，基因组大小按照3 Gbp计算。

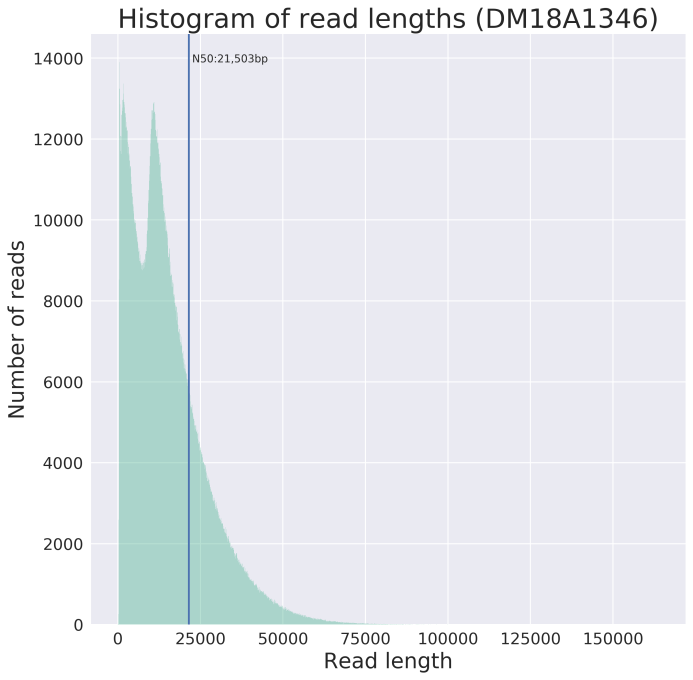


Figure 4.1 Reads长度分布图

## 4.2 序列比对和统计

对NGMLR比对的结果统计reads和bases的比对情况，如Table 4.2所示。

Table 4.2 比对结果统计

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sample ID | total reads | mapped reads | total bases | mapped bases | error rate |
| {{proj\_info.sample\_id}} | {{bam\_stat.raw\_total\_sequences}} | {{bam\_stat.reads\_mapped}} | {{bam\_stat.total\_length}} | {{bam\_stat.bases\_mapped}} | {{bam\_stat.error\_rate}} |

注：错误率是错误比对碱基数与比对上所有碱基数（基于更准确的cigar值计算）的比值；另外，reads比对率是{{bam\_stat.reads\_map\_rate}}，碱基比对率是{{bam\_stat.bases\_map\_rate}}（比对率是比对上的数目与总数目的比值）。

## 4.3 结构变异检测和统计

对Sniffles检测出的结构变异，按照支持结构变异的reads数目大于或等于4进行筛选。5种主要类型的结构变异数目统计如Table 4.3，结构变异数目统计柱状图如Figure 4.2，长度在30 bp-1 Mbp之间的结构变异长度分布图如Figure 4.3。

Table 4.3 结构变异数目统计

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sample ID | DEL | DUP | INS | INV | TRA | Total |
| {{proj\_info.sample\_id}} | {{sv\_num.DEL}} | {{sv\_num.DUP}} | {{sv\_num.INS}} | {{sv\_num.INV}} | {{sv\_num.TRA}} | {{sv\_num.Total}} |

注：DEL代表Deletion； INS代表Insertion； DUP代表Duplication；INV代表Inversion；TRA代表Translocation

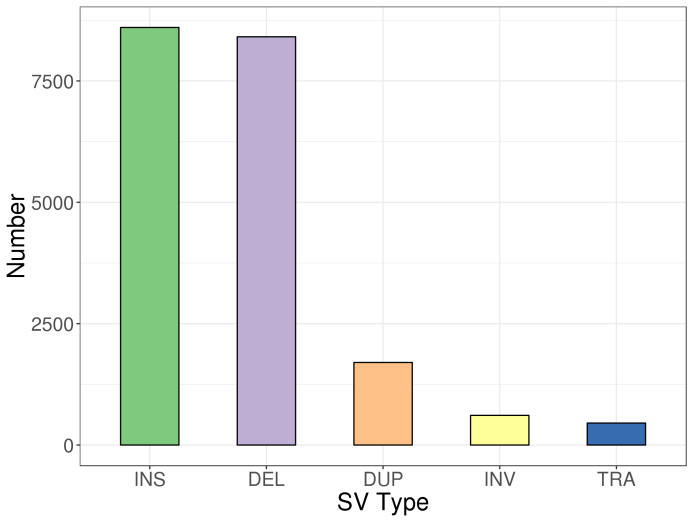


Figure 4.2 结构变异数目柱状图

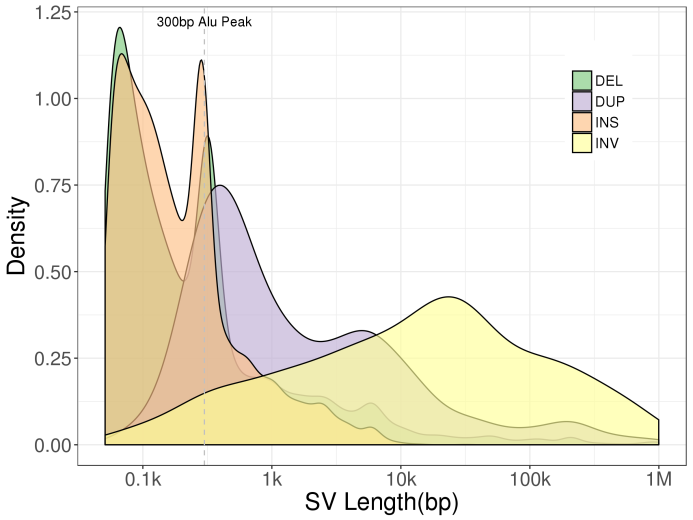


Figure 4.3 结构变异长度分布图

## 4.4 结构变异注释

结构变异注释主要有三个方面的作用：1. 与健康人公共结构变异数据库（DGV，1000 genome SV）比较过滤掉人群中高频的结构变异；2. 与疾病公共结构变异数据库（dbVar，Decipher等）比较鉴别疾病相关的结构变异；3. 结构变异的基因、重复元件、调控元件等注释有助于分析结构变异的功能。4. 与本地人群结构变异数据库比较，过滤平台相关的结构变异bias。结构变异注释结果说明如下（详细的注释结果见附件）：

|  |  |  |
| --- | --- | --- |
| 编号 | 名称 | 说明 |
| 1 | Chr | 染色体编号 |
| 2 | Start | 变异起始位置 |
| 3 | End | 变异终止位置 |
| 4 | SVTYPE | 结构变异类型，包括DEL，INS，INV，DUP，TRA等 |
| 5 | SVID | 结构变异编号 |
| 6 | SVLEN | 结构变异长度 |
| 7 | RE | Reads Evidence，支持结构变异的reads数目 |
| 8 | INFO | 结构变异检测VCF INFO，包括断点的标准差、支持该结构变异的reads的名称等信息 |
| 9 | GrandSV\_Frequency | 希望组本地SV数据库，SV频率 |
| 10 | GrandSV\_Variant | 希望组本地SV数据库，结构变异的染色体位置及类型，格式为 chr:start-end,type，如果该结构变异为TRA，那么显示断点位置，例如：chr1:123456-123457,TRA；如果注释到数据库中的多个结构变异，那么用“；”将它们分开 |
| 11 | 1000g\_SV | 1000 genome 结构变异注释，结构变异的染色体位置及类型，格式为：chr:start-end,type，如果注释到数据库中的多个结构变异，那么用“；”将它们分开 |
| 12 | 1000g\_subtype | 1000 genome 结构变异注释，结构变异的亚类，包括：ALU, CNV, DEL, DEL\_ALU, DEL\_HERV, DEL\_LINE1, DEL\_SVA, DUP, INS, INV, LINE1, SVA |
| 13 | 1000g\_AF | 1000 genome 结构变异注释，结构变异在人群中的频率，如果注释到数据库中的多个结构变异，那么用“；”将频率分开 |
| 14 | 1000g\_EAS\_AF | 1000 genome 结构变异注释，结构变异在东亚人群中的频率，如果注释到数据库中的多个结构变异，那么用“；”将频率分开 |
| 15 | 1000g\_EUR\_AF | 1000 genome 结构变异注释，结构变异在欧洲人群中的频率，如果注释到数据库中的多个结构变异，那么用“；”将频率分开 |
| 16 | 1000g\_AFR\_AF | 1000 genome 结构变异注释，结构变异在非洲人群中的频率，如果注释到数据库中的多个结构变异，那么用“；”将频率分开 |
| 17 | 1000g\_AMR\_AF | 1000 genome 结构变异注释，结构变异在美洲混血人群中的频率，如果注释到数据库中的多个结构变异，那么用“；”将频率分开 |
| 18 | 1000g\_SAS\_AF | 1000 genome 结构变异注释，结构变异在南亚人群中的频率，如果注释到数据库中的多个结构变异，那么用“；”将频率分开 |
| 19 | dgv\_SV | dgv gold standard CNV注释，结构变异的染色体位置及类型，格式为：chr:start-end,type，如果注释到数据库中的多个结构变异，那么用“；”将它们分开 |
| 20 | dgv\_AF | dgv gold standard CNV注释，结构变异在人群中的频率，如果注释到数据库中的多个结构变异，那么用“；”将频率分开 |
| 21 | dgv\_sample\_size | dgv gold standard CNV注释，研究的样本大小，如果注释到数据库中的多个结构变异，那么用“；”将它们分开 |
| 22 | dbVar\_SV | dbVar 结构变异注释，结构变异的染色体位置及类型，格式为：chr:start-end,type，如果注释到数据库中的多个结构变异，那么用“；”将它们分开 |
| 23 | dbVar\_CLNSIG | dbVar 结构变异注释，疾病相关结构变异的临床分级，如果注释到数据库中的多个结构变异，那么用“；”将它们分开 |
| 24 | dbVar\_PHENO | dbVar 结构变异注释，疾病相关结构变异的表型，如果注释到数据库中的多个结构变异，那么用“；”将它们分开 |
| 25 | Decipher\_region | Decipher 单倍剂量不足区域，格式为：chrom:start-end,WILD |
| 26 | Decipher\_HI | Decipher 预测的单倍剂量不足值，格式为：基因|HI\_Score|HI\_Percentage，HI\_Score 为预测的单倍剂量不足值，越大表示该基因越容易表现单倍剂量不足，HI\_Percentage，越小表示该基因越容易表现单倍剂量不足，一般， HI\_Percentage 0-10%表明该基因极易出现单倍剂量不足，HI\_Percentage 90-100% 表明该基因不易出现单倍剂量不足。 |
| 27 | Func.refGene | 结构变异区域的基因功能元件，包括exonic, splicing, intronic, ncRNA, UTR5, UTR3, upstream, downstream, intergenic |
| 28 | Gene.refGene | 结构变异区域的基因。 |
| 29 | GeneDetail.refGene | 描述UTR、splicing、ncRNA\_splicing或intergenic区域的变异情况。当Func列的值为exonic、ncRNA\_exonic、intronic、ncRNA\_intronic、upstream、downstream、upstream；downstream、ncRNA\_UTR3、ncRNA\_UTR5时，该列为空；当Func列的值为exonic;splicing时，表示该位点位于某些转录本的exonic区，另一些转录本的splicing区，这种情况下，GeneDetail会给出该位点对于转录本splicing的影响，例如，NM\_152486:exon3:c.232C>T，表示该变异位于转录本NM\_152486上，exon3表示第3个外显子，c.232C>T表示cDNA的232bp处发生由C到T的突变；当Func列的值为intergenic时，该列格式为dist=1322;dist=12414，表示该变异位点距离两侧基因的距离，对于结构变异此项为“.” |
| 30 | ExonicFunc.refGene | 外显子区的SNV or InDel变异类型（SNV的变异类型包括synonymous\_SNV, missense\_SNV, stopgain, stopgloss和unknown；InDel的变异类型包括frameshift insertion, frameshift deletion, stopgain, stoploss, nonframeshift insertion, nonframeshift deletion和unknown），对于结构变异此项为“.” |
| 31 | AAChange.refGene | 氨基酸改变，只有当Func列为exonic或exonic;splicing时，该列才有结果。按照每个转录本进行注释（例如，AIM1L:NM\_001039775:exon2:c.C2768T:p.P923L，其中，AIM1L表示该变异所在的基因名称，NM\_001039775表示该变异所在的转录本ID，exon2表示该变异位于转录本的第二个外显子上，c.C2768T表示该变异引起cDNA在第2768位上由C突变为T，p.P923L表示该变异引起蛋白序列在第923位上的氨基酸由Pro变为Leu）,再如，FMN2:NM\_020066:exon1:c.160\_162del:p.54\_54del，表示该变异引起cDNA的第160到162位发生删除，p.54\_54del表示该变异引起蛋白序列在第54位上的氨基酸删除，对于结构变异此项为“.” |
| 32 | OMIM\_id | OMIM 编号 |
| 33 | hpo\_or\_dis | HPO疾病描述 |
| 34 | chpo\_or\_dis | CHPO 疾病中文描述 |
| 35 | inheritance | 疾病遗传方式 |
| 36 | cytoBand | 该区域所处的染色体区段（利用Giemas染色观察得到的）。如果变异位点跨过多个区段，用短横线连接 |
| 37 | dgvMerged | dgv数据库中与该区域有重叠的结构变异编号。 |
| 38 | tfbsConsSites | 基于transfac矩阵数据库(v7.0)，计算所有转录因子结合位点在人/小鼠/大鼠比对中的保守分值，当结合位点的分值达到阈值时，认为该位点在人/小鼠/大鼠中保守。该列给出的是该变异位点所在的保守转录因子结合位点的位置和分值，即Name和Score。Name是结合位点处的motif名称，这些motif能够被转录因子识别，例如V$CDPCR3\_01，利用一些在线服务器（如MSigDB）能够查询这个motif能够被哪些转录因子识别；Score是该结合位点的保守分值 |
| 39 | genomicSuperDups | genomicSuperDups：检测该变异位点是否位于重复片段（segmental duplication）中。重复区域中检测到的遗传变异大多数是由于序列比对错误造成的，所以被注释到segmental duplications的变异需要谨慎对待，很可能是假阳性位点。给出两个值，一是Name，表示基因组中与该变异位点所在区域相似的片段的位置；二是Score，表示两个相似片段的序列一致性。例如，Score=0.994828;Name=chr19:60000，表示chr19:60000所在片段跟该变异位点所在片段相似，序列一致性为0.994828，范围0~1 |
| 40 | rmsk | 重复序列注释信息，重复序列来源于RepeatMasker注释。例如,Name="1385:(CACCC)n(Simple\_repeat)"。Name由两部分构成，一部分(CACCC)n是repeat的名称，另一部分Simple是repeat的类别。只要有注释信息，就表明该变异位于散在重复序列或低复杂度序列中；这些区域容易出现比对错误，所以该区域的变异位点可靠性不高 |
| 41 | wgEncodeBroadHmmGm12878HMM | chromHMM 预测的增强子，启动子和抑制子注释，Gm12878细胞系，详见chromHMM状态说明 |
| 42 | wgEncodeBroadHmmH1hescHMM | chromHMM 预测的增强子，启动子和抑制子注释，H1hesc细胞系，详见chromHMM状态说明 |
| 43 | wgEncodeBroadHmmHepg2HMM | chromHMM 预测的增强子，启动子和抑制子注释，Hepg2细胞系，详见chromHMM状态说明 |
| 44 | wgEncodeBroadHmmHuvecHMM | chromHMM 预测的增强子，启动子和抑制子注释，Huvec细胞系，详见chromHMM状态说明 |
| 45 | wgEncodeBroadHmmK562HMM | chromHMM 预测的增强子，启动子和抑制子注释，K562细胞系，详见chromHMM状态说明 |
| 46 | cpgIslandExt | CpG岛预测结果，注释结果为CpG岛名称，如CpG:116（116是该CpG岛中CG二核苷酸的数目），CpG岛是指在基因组的某些区域，CpG保持或高于正常概率，与甲基化修饰及基因表达调控相关 |
| 47 | targetScanS | UCSC提供TargetScanS注释数据库，库中包含在3‘UTR中保守的microRNA结合位点。此项给出microRNA靶点的信息，一是score ，是该靶点的分值，有分值即认为此位点在人中保守；二是Name，是作用于该靶点的microRNA名称。例如，Score=62;Name=KRAS:miR-181:1，表示该靶点的分值是62，其位于KRAS基因的3‘UTR中，受到该变异位点影响的microRNA是miR-181:1 |
| 48 | wgRna | 基于miRBase和snoRNABase，对变异位点相关的microRNA和snoRNA进行注释，给microRNA和snoRNA的基因名称 |

## 4.5 相关基因调研

## 4.6 重点关注的SV

# 5 文件交付及格式说明

## 5.1 软件及参数

使用NGMLR软件进行Nanopore reads 与参考基因组hg19比对，比对参数选用该软件预设的用于Nanopore reads的参数：-x ont；

使用Sniffles软件进行SV检测，选定最小reads支持数为1（--min\_support 1），后面根据统计需求选择不同的reads过滤条件。

## 5.2 文件列表

{{proj\_info.sample\_id}} # 根目录

├── 0.fast5 # H5格式下机数据

├── 1.fastq # 原始 fastq 和summary文件

├── 2.bam # mapping 文件

├── 3.sv # 建构变异VCF及注释结果

# 参考文献

[1]. Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. Nature Reviews Genetics, 12(5), 363.

[2]. Nattestad, M., Goodwin, S., Ng, K., Baslan, T., Sedlazeck, F. J., Rescheneder, P., ... & Tseng, E. (2018). Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line. Genome research.

[3]. Hedges, D. J., Hamilton-Nelson, K. L., Sacharow, S. J., Nations, L., Beecham, G. W., Kozhekbaeva, Z. M., ... & Jaworski, J. M. (2012). Evidence of novel fine-scale structural variation at autism spectrum disorder candidate loci. Molecular autism, 3(1), 2.

[4]. Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerrière, A., Vital, A., ... & Dubas, F. (2006). APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nature genetics, 38(1), 24.

[5]. Escaramís, G., Docampo, E., & Rabionet, R. (2015). A decade of structural variants: description, history and methods to detect structural variation. Briefings in functional genomics, 14(5), 305-314.

[6]. Sedlazeck, F. J., Reschender, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., & Schatz, M. C. (2018). Accurate detection of complex structural variations using single molecule sequencing. Nature methods.

[7]. Miao, H., Zhou, J., Yang, Q., Liang, F., Wang, D., Ma, N., ... & Zhang, Q. (2018). Long-read sequencing identified a causal structural variant in an exome-negative case and enabled preimplantation genetic diagnosis. bioRxiv, 326496.

[8]. Merker, J. D., Wenger, A. M., Sneddon, T., Grove, M., Zappala, Z., Fresard, L., ... & Montgomery, S. B. (2018). Long-read genome sequencing identifies causal structural variation in a Mendelian disease. Genetics in Medicine, 20(1), 159.