

# 高通量测序领域常用名词解释大全

## 一、 学科分类名词

### 1. 什么是基因组学

基因组学 ( genomics ), 研究生物基因组和如何利用基因的一门学问。用于概括涉及基因作图、测序和整个基因组功能分析的遗传学分支。该学科提供基因组信息以及相关数据系统利用, 试图解决生物, 医学, 和工业领域的重大问题。

### 2. 什么是功能基因组学

功能基因组学( Functional genomics )又往往被称为后基因组学( Post genomics ), 它利用结构基因组所提供的信息和产物, 发展和应用新的实验手段, 通过在基因组或系统水平上全面分析基因的功能, 使得生物学研究从对单一基因或蛋白质得研究转向多个基因或蛋白质同时进行系统的研究。这是在基因组静态的碱基序列弄清楚之后转入对基因组动态的生物学功能学研究。研究内容包括基因功能发现、基因表达分析及突变检测。基因的功能包括: 生物学功能, 如作为蛋白质激酶对特异蛋白质进行磷酸化修饰; 细胞学功能, 如参与细胞间和细胞内信号传递途径; 发育上功能, 如参与形态建成等。采用的手段包括经典的减法杂交, 差示筛选, cDNA 代表差异分析以及 mRNA 差异显示等, 但这些技术不能对基因进行全面系统的分析, 新的技术应运而生, 包括基因表达的系统分析 ( serial analysis of geneexpression, SAGE ), cDNA 微阵列( cDNA microarray ), DNA 芯片( DNA chip )和序列标志片段显示( sequence tagged fragments display)。

### 3. 什么是比较基因组学

比较基因组学(Comparative Genomics)是基于基因组图谱和测序基础上, 对已知的基因和基因组结构进行比较, 来了解基因的功能、表达机理和物种进化的学科。利用

模式生物基因组与人类基因组之间编码顺序上和结构上的同源性，克隆人类疾病基因，揭示基因功能和疾病分子机制，阐明物种进化关系，及基因组的内在结构。

#### **4. 什么是表观遗传学**

表观遗传学是研究基因的核苷酸序列不发生改变的情况下，基因表达了可遗传的变化的一门遗传学分支学科。表观遗传的现象很多，已知的有 DNA 甲基化 ( DNAmethylation )，基因组印记 ( genomicimprinting )，母体效应 ( maternaleffects )，基因沉默 ( genesilencing )，核仁显性，休眠转座子激活和 RNA 编辑 ( RNA editing ) 等。

#### **5. 什么是计算生物学**

计算生物学是指开发和应用数据分析及理论的方法、数学建模、计算机仿真技术等。当前，生物学数据量和复杂性不断增长，每 14 个月基因研究产生的数据就会翻一番，单单依靠观察和实验已难以应付。因此，必须依靠大规模计算模拟技术，从海量信息中提取最有用的数据。

## **二、 技术分类名词**

### **1. 什么是 Sanger 法测序**

Sanger 法测序利用一种 DNA 聚合酶来延伸结合在待定序列模板上的引物。直到掺入一种链终止核苷酸为止。每一次序列测定由一套四个单独的反应构成，每个反应含有所有四种脱氧核苷酸三磷酸(dNTP)，并混入限量的一种不同的双脱氧核苷三磷酸(ddNTP)。由于 ddNTP 缺乏延伸所需要的 3-OH 基团，使延长的寡聚核苷酸选择性地 G、A、T 或 C 处终止。终止点由反应中相应的双脱氧而定。每一种 dNTPs 和 ddNTPs 的相对浓度可以调整，使反应得到一组长几百至几千碱基的链终止产物。它们具有共同的起始点，但终止在不同的核苷酸上，可通过高分辨率变性凝胶电泳分离大

小不同的片段，凝胶处理后可用 X-光胶片放射自显影或非同位素标记进行检测。

## 2. 什么是高通量测序

高通量测序技术 ( High-throughput sequencing , HTS ) 是对传统 Sanger 测序 ( 称为一代测序技术 ) 革命性的改变, 一次对几十万到几百万条核酸分子进行序列测定, 因此在有些文献中称其为下一代测序技术(next generation sequencing , NGS ) 足见其划时代的改变, 同时高通量测序使得对一个物种的转录组和基因组进行细致全貌的分析成为可能, 所以又被称为深度测序(Deep sequencing)。

## 3. 什么是单分子测序技术

单分子实时测序技术 SMRT 技术是基于边合成边测序的思想, 以 SMRT 芯片为载体进行测序反应。 SMRT 芯片是一种带有很多零模式波导孔 ( zero-mode waveguides , ZMW ) ( 厚度为 100 nm ) 的金属片, ZMW 是一种直径只有几十个纳米的孔, 由于其底部上的小孔短于激光的单个波长, 导致激光无法直接穿过小孔, 而会在小孔处发生光的衍射, 形成局部发光的区域, 即为荧光信号检测区, 该区域内锚定有 DNA 聚合酶。测序时将基因组的 DNA 打断成许多小的片段, 制成液滴后将其分散到不同的 ZMW 纳米孔中。当 ZMW 孔底部聚合反应发生时, 被不同荧光标记的核苷酸会在小孔的荧光探测区域中被聚合酶滞留数十毫秒, 荧光标记会在激光束( Excitation ) 的激发下发出荧光, 根据荧光的种类就可以判定 dNTP 的种类, 反应完成后荧光标记会被聚合酶切除而弥散出 ZMW 小孔, 其它未参与合成的 dNTP 由于没进入荧光信号检测区而不进入荧光信号检测区。 SMRT 测序时, 样品准备过程涉及到样品 DNA 的打断、末端补齐、连接接头、测序这几个步骤。测序中需要的样品量很少, 样品准备中所用的试剂也很少, 而且测序过程中省去扫描和洗涤的过程, 所以测序所花的时间较少。 Pacific Biosciences ( 单分子实时测序 ) 是正统的三代测序仪. ZS Genetics 和 Oxford

Nanopore 等纳米孔单分子测序技术则是更加革命性的三代测序技术甚至是第四代技术。

### **三、 产品分类名词**

#### **1. 什么是基因组重测序 ( Genome Re-sequencing )**

全基因组重测序是对基因组序列已知的个体进行基因组测序，并在个体或群体水平上进行差异性分析的方法。随着基因组测序成本的不断降低，人类疾病的致病突变研究由外显子区域扩大到全基因组范围。通过构建不同长度的插入片段文库和短序列、双末端测序相结合的策略进行高通量测序，实现在全基因组水平上检测疾病关联的常见、低频、甚至是罕见的突变位点，以及结构变异等，具有重大的科研和产业价值。

#### **2. 什么是基因组 de novo 测序**

de novo 测序也称为从头测序：其不需要任何现有的序列资料就可以对某个物种进行测序，利用生物信息学分析手段对序列进行拼接，组装，从而获得该物种的基因组图谱。获得一个物种的全基因组序列是加快对此物种了解的重要捷径。随着新一代测序技术的飞速发展，基因组测序所需的成本和时间较传统技术都大大降低，大规模基因组测序渐入佳境，基因组学研究也迎来新的发展契机和革命性突破。利用新一代高通量、高效率测序技术以及强大的生物信息分析能力，可以高效、低成本地测定并分析所有生物的基因组序列。

#### **3. 什么是外显子测序 ( whole exon sequencing )**

外显子组测序是指利用序列捕获技术将全基因组外显子区域 DNA 捕捉并富集后进行高通量测序的基因组分析方法。外显子测序相对于基因组重测序成本较低，对研究已知基因的 SNP、Indel 等具有较大的优势，但无法研究基因组结构变异如染色体断裂重组等。

#### **4. 什么是转录组测序**

转录组学( transcriptomics )是在基因组学后新兴的一门学科,即研究特定细胞在某一功能状态下所能转录出来的所有 RNA ( 包括 mRNA 和非编码 RNA ) 的类型与拷贝数。Illumina 提供的 mRNA 测序技术可在整个 mRNA 领域进行各种相关研究和新的发现。mRNA 测序不对引物或探针进行设计,可自由提供关于转录的客观和权威信息。研究人员仅需要一次试验即可快速生成完整的 poly-A 尾的 RNA 完整序列信息,并分析基因表达、cSNP、全新的转录、全新异构体、剪接位点、等位基因特异性表达和罕见转录等最全面的转录组信息。简单的样品制备和数据分析软件支持在所有物种中的 mRNA 测序研究。

#### **5. 什么是表达谱测序**

基因表达谱(gene expression profile): 指通过构建处于某一特定状态下的细胞或组织的非偏性 cDNA 文库,大规模 cDNA 测序,收集 cDNA 序列片段、定性、定量分析其 mRNA 群体组成,从而描绘该特定细胞或组织在特定状态下的基因表达种类和丰度信息,这样编制成的数据表就称为基因表达谱

#### **6. 什么是 small RNA 测序**

Small RNA ( micro RNAs、siRNAs 和 pi RNAs ) 是生命活动重要的调控因子,在基因表达调控、生物个体发育、代谢及疾病的发生等生理过程中起着重要的作用。Illumina 能够对细胞或者组织中的全部 Small RNA 进行深度测序及定量分析等研究。实验时首先将 18-30 nt 范围的 Small RNA 从总 RNA 中分离出来,两端分别加上特定接头后体外反转录做成 cDNA 再做进一步处理后,利用测序仪对 DNA 片段进行单向末端直接测序。通过 Illumina 对 SmallRNA 大规模测序分析,可以从中获得物种全基因组水平的 miRNA 图谱,实现包括新 miRNA 分子的挖掘,其作用靶基因的

预测和鉴定、样品间差异表达分析、 miRNAs 聚类 and 表达谱分析等科学应用。成熟的 microRNA ( miRNA ) 是 17~24nt 的单链非编码 RNA 分子 , 通过与 mRNA 相互作用影响目标 mRNA 的稳定性及翻译 , 最终诱导基因沉默 , 调控着基因表达、细胞生长、发育等生物学过程。基于第二代测序技术的 microRNA 测序 , 可以一次性获得数百万条 microRNA 序列 , 能够快速鉴定出不同组织、不同发育阶段、不同疾病状态下已知和未知的 microRNA 及其表达差异 , 为研究 microRNA 对细胞进程的作用及其生物学影响提供了有力工具。

## **7. 什么是 Chip-seq**

染色质免疫共沉淀技术 ( Chromatin Immunoprecipitation , ChIP ) 也称结合位点分析法 , 是研究体内蛋白质与 DNA 相互作用的有力工具 , 通常用于转录因子结合位点或组蛋白特异性修饰位点的研究。将 ChIP 与第二代测序技术相结合的 ChIP-Seq 技术 , 能够高效地在全基因组范围内检测与组蛋白、转录因子等互作的 DNA 区段。ChIP-Seq 的原理是 : 首先通过染色质免疫共沉淀技术 ( ChIP ) 特异性地富集目的蛋白结合的 DNA 片段 , 并对其进行纯化与文库构建 ; 然后对富集得到的 DNA 片段进行高通量测序。研究人员通过将获得的数百万条序列标签精确定位到基因组上 , 从而获得全基因组范围内与组蛋白、转录因子等互作的 DNA 区段信息。

## **8. 什么是 RIP-seq**

RNA Immunoprecipitation 是研究细胞内 RNA 与蛋白结合情况的技术 , 是了解转录后调控网络动态过程的有力工具 , 能帮助我们发现 miRNA 的调节靶点。这种技术运用针对目标蛋白的抗体把相应的 RNA-蛋白复合物沉淀下来 , 然后经过分离纯化就可以对结合在复合物上的 RNA 进行测序分析。RIP 可以看成是普遍使用的染色质免疫沉淀 ChIP 技术的类似应用 , 但由于研究对象是 RNA-蛋白复合物而不是 DNA-蛋白复

合物，RIP 实验的优化条件与 ChIP 实验不太相同（如复合物不需要固定，RIP 反应体系中的试剂和抗体绝对不能含有 RNA 酶，抗体需经 RIP 实验验证等等）。RIP 技术下游结合 microarray 技术被称为 RIP-Chip，帮助我们更高通量地了解癌症以及其它疾病整体水平的 RNA 变化。

## 9. 什么是宏基因组

宏基因组 (Magenomics) 研究的对象是整个微生物群落。相对于传统单个细菌研究来说，它具有众多优势，其中很重要的两点：(1) 微生物通常是以群落方式共生于某一小生境中，它们的很多特性是基于整个群落环境及个体间的相互影响的，因此做 Metagenomics 研究比做单个个体的研究更能发现其特性；(2) Metagenomics 研究无需分离单个细菌，可以研究那些不能被实验室分离培养的微生物。

## 10. 什么是 DNA 甲基化

DNA 甲基化是指在 DNA 甲基化转移酶的作用下，在基因组 CpG 二核苷酸的胞嘧啶 5'碳位共价键结合一个甲基基团。正常情况下，人类基因组“垃圾”序列的 CpG 二核苷酸相对稀少，并且总是处于甲基化状态，与之相反，人类基因组中大小为 100—1000 bp 左右且富含 CpG 二核苷酸的 CpG 岛则总是处于未甲基化状态，并且与 56% 的人类基因组编码基因相关。人类基因组序列草图分析结果表明，人类基因组 CpG 岛约为 28890 个，大部分染色体每 1 Mb 就有 5—15 个 CpG 岛，平均值为每 Mb 含 10.5 个 CpG 岛，CpG 岛的数目与基因密度有良好的对应关系[9]。由于 DNA 甲基化与人类发育和肿瘤疾病的密切关系，特别是 CpG 岛甲基化所致抑癌基因转录失活问题，DNA 甲基化已经成为表观遗传学和表观基因组学的重要研究内容。

## 11. 什么是基因组印记

基因组印记(又称遗传印记)是指基因根据亲代的不同而有不同的表达。印记基因的

存在能导致细胞中两个等位基因的一个表达而另一个不表达。基因组印记是一正常过程，此现象在一些低等动物和植物中已发现多年。印记的基因只占人类基因组中的少数，可能不超过 5%，但在胎儿的生长和行为发育中起着至关重要的作用。基因组印记病主要表现为过度生长、生长迟缓、智力障碍、行为异常。目前在肿瘤的研究中认为印记缺失是引起肿瘤最常见的遗传学因素之一。

## **四、 信息分析名词**

### **1. 什么是 Read**

高通量测序平台产生的序列标签就称为 read，测序产生的每一条连续的序列称为一个 Read。如果是 Pair-end 测序，分别从一个片段的 3' 和 5' 端进行测序，会产生两个 Reads。如果是 Single-end 测序，只从一端测序，一个片段只会产生一个 Read。比如一条 lane 产生 100M 的 reads，如果是 PE100 测序，产量就是 20G，如果是 SE 测序，产量就是 10G。

### **2. 什么是 Contig**

拼接软件基于 reads 之间的 overlap 区，拼接获得的序列称为 Contig(重叠群)。

### **3. 什么是 Scaffold**

基因组 de novo 测序，通过 reads 拼接获得 Contigs 后，往往还需要构建 454 Paired-end 库或 Illumina Mate-pair 库，以获得一定大小片段（如 3Kb、6Kb、10Kb、20Kb）两端的序列。基于这些序列，可以确定一些 Contig 之间的顺序关系，这些先后顺序已知的 Contigs 组成 Scaffold。

### **4. 什么是 Contig N50**

Reads 拼接后会获得一些不同长度的 Contigs。将所有的 Contig 长度相加，能获得一个 Contig 总长度。然后将所有的 Contigs 按照从长到短进行排序，如获得



Contig 1 ,Contig2 ,Contig 3.....Contig 25。将 Contig 按照这个顺序依次相加 , 当相加的长度达到 Contig 总长度的一半时 , 最后一个加上的 Contig 长度即为 Contig N50。举例 : Contig1+Contig 2+ Contig 3+Contig 4=Contig 总长度\*1/2 时 , Contig 4 的长度即为 ContigN50。 Contig N50 可以作为基因组拼接的结果好坏的一个判断标准。

## **5. 什么是 Scaffold N50**

Scaffold N50 与 Contig N50 的定义类似 。 Contigs 拼接组装获得一些不同长度的 Scaffolds。将所有的 Scaffold 长度相加 ,能获得一个 Scaffold 总长度。然后将所有的 Scaffolds 按照从长到短进行排序 , 如获得 Scaffold 1 , Scaffold 2 , Scaffold3.....Scaffold 25。将 Scaffold 按照这个顺序依次相加 , 当相加的长度达到 Scaffold 总长度的一半时 , 最后一个加上的 Scaffold 长度即为 Scaffold N50。举例 : Scaffold 1+Scaffold2+ Scaffold 3 +Scaffold 4 +Scaffold 5=Scaffold 总长度\*1/2 时 , Scaffold 5 的长度即为 Scaffold N50。 Scaffold N50 可以作为基因组拼接的结果好坏的一个判断标准。

## **6. 什么是基因组注释**

基因组注释(Genome annotation) 是利用生物信息学方法和工具,对基因组所有基因的生物学功能进行高通量注释,是当前功能基因组学研究的一个热点。基因组注释的研究内容包括基因识别和基因功能注释两个方面。基因识别的核心是确定全基因组序列中所有基因的确切位置。

## **7. 什么是测序深度**

测序深度是指测序得到的总碱基数与待测基因组大小的比值。假设一个基因大小为 2M , 测序深度为 10X , 那么获得的总数据量为 20M。覆盖度是指测序获得的序列占

整个基因组的比例。由于基因组中的高 GC、重复序列等复杂结构的存在，测序最终拼接组装获得的序列往往无法覆盖所有的区域，这部分没有获得的区域就称为 Gap。例如一个细菌基因组测序，覆盖度是 98%，那么还有 2% 的序列区域是没有通过测序获得的。测序深度 (Sequencing Depth)：测序得到的碱基总量 (bp) 与基因组大小 (Genome) 的比值，它是评价测序量的指标之一。测序深度与基因组覆盖度之间是一个正相关的关系，测序带来的错误率或假阳性结果会随着测序深度的提升而下降。重测序的个体，如果采用的是双末端或 Mate-Pair 方案，当测序深度在 10~15X 以上时，基因组覆盖度和测序错误率控制均得以保证。假设一个基因大小为 2M，测序深度为 10X，那么获得的总数据量为 20M。覆盖度是指测序获得的序列占整个基因组的比例。由于基因组中的高 GC、重复序列等复杂结构的存在，测序最终拼接组装获得的序列往往无法覆盖所有的区域，这部分没有获得的区域就称为 Gap。例如一个细菌基因组测序，覆盖度是 98%，那么还有 2% 的序列区域是没有通过测序获得的。

## 8. 什么是测序的覆盖度

对于 coverage，由于大片段拼接的 gap (空白或者缺口)、测序读长有限、重复序列等问题的存在，测序分析后组装得到的基因组序列通常无法完全覆盖所有区域，覆盖度就是最终得到的结果占整个基因组的比例。例如一个人的基因组测序，覆盖度为 98.5%，那么说明该基因组还有 1.5% 的区域通过我们的组装和分析无法得到；对于 depth，就是被测基因组上单个碱基被测序的平均次数，比如某样本的测序深度为 30X，那么就是说该样本的基因组上每一个单碱基平均被测序 (或者说读取) 了 30 次，注意，是平均。当然了，depth 也有最大和最小值，这个都可以由信息分析得到。其实也就是为了提高准确率什么的，一般 15X 就差不多了。

## 9. 什么是 SNP、SNV

单核苷酸多态性 single nucleotide polymorphism , SNP 或单核苷酸位点变异 SNV。个体间基因组 DNA 序列同一位置单个核苷酸变异(替代、插入或缺失)所引起的多态性。不同物种、个体基因组 DNA 序列同一位置上的单个核苷酸存在差别的现象。有这种差别的基因座、DNA 序列等可作为基因组作图的标志。人基因组上平均约每 1000 个核苷酸即可能出现 1 个单核苷酸多态性的变化, 其中有些单核苷酸多态性可能与疾病有关, 但可能大多数与疾病无关。单核苷酸多态性是研究人类家族和动植物品种遗传变异的重要依据。在研究癌症基因组变异时, 相对于正常组织, 癌症中特异的单核苷酸变异是一种体细胞突变 ( somatic mutation ), 称做 SNV。

## **10. 什么是 indel**

基因组上小片段 ( >50bp ) 的插入或缺失, 形同 SNP/SNV。

## **11. 什么是基因组拷贝数变异**

基因组拷贝数变异 ( copy number variation , CNV ) 是基因组变异的一种形式, 通常使基因组中大片段的 DNA 形成非正常的拷贝数量。例如人类正常染色体拷贝数是 2, 有些染色体区域拷贝数变成 1 或 3, 这样, 该区域发生拷贝数缺失或增加, 位于该区域内的基因表达量也会受到影响。如果把一条染色体分成 A-B-C-D 四个区域, 则 A-B-C-C-D/A-C-B-C-D/A-CC-B-C-D/A-B-D 分别发生了 C 区域的扩增及缺失, 扩增的位置可以是连续扩增如 A-B-C-CD 也可以是在其他位置的扩增, 如 A-C-B-C-D。

## **12. 什么是基因组结构变异**

染色体结构变异 ( structure variation , SV ) 是指在染色体上发生了大片的变异。主要包括染色体大片段的插入和缺失 ( 引起 CNV 的变化 ), 染色体内部的某块区

域发生翻转颠换,两条染色体之间发生重组( inter-chromosome trans-location )等。

一般 SV 的展示利用 Circos 软件。

### **13. 什么是 Segment duplication**

一般称为 SD 区域,串联重复是由序列相近的一些 DNA 片段串联组成。串联重复在人类基因多样性的灵长类基因中发挥重要作用。在人类染色体 Y 和 22 号染色体上,有很大的 SD 序列。

### **14. 什么是 genotype and phenotype**

基因型与表型一般指某些单核苷酸位点变异与表现形式间的关系。

### **15. 什么是 UniGene**

UniGene 是以自动化的方式,对于每一个新进入到 GeneBank 的序列,进行序列相似性分析,如果可以找到可能是来自于同一个基因的基因组( cluster ),则将次序列归入到这一个基因组,如果找不到,则成立一个新的基因组。据估计,人类的基因约有八万到十万个左右,而在 UniGenes 中的所有人类序列中,经过上述方式加以分组之后,在 1998 您 6 月,已得到的超过四万三千个独特的基因组( unique gene clusters ),其中大约六千余个具有已知的基因。

### **16. 什么是 RPKM、 FPKM**

RPKM 是指每 1 百万个 map 上的 reads 中 map 到外显子的每 1K 个碱基上的 reads 个数。假如有 1 百万个 reads 映射到了人的基因组上,那么具体到每个外显子呢,有多少映射上了呢,而外显子的长度不一,那么每 1K 个碱基上又有多少 reads 映射上去如果对应特定基因的话,那么就是每 1000000 mapped 到该基因上的 reads 中每 kb 有多少是 mapped 到该基因上的 exon 的 read 映射到外显子上总的 reads 个数。这个是映射到某个区域上的 reads 个数,这个区域或者

是已知注释的基因或者跨两个外显子的边界或者是某个基因已经注释的转录本的内含子、外显子。对于真核生物来说，外显子和它们自己内部的关系由某类型的 mRNA 来注释。

$$\text{RPKM (exon)} = 10^9 * \text{exon\_tag\_count} / (\text{total\_tag\_count} * \text{exon\_size})$$
$$\text{RPKM (gene)} = 10^9 * \text{gene\_tag\_count} / (\text{total\_tag\_count} * \text{canonical\_transcript\_size})$$

Mortazavi et al. (2008) Nature Methods

举例：比如对应到该基因的 read 有 1000 个，总 reads 个数有 100 万，而该基因的外显子总长为 5kb，那么它的 RPKM 为： $10^9 * 1000(\text{reads 个数}) / 10^6(\text{总 reads 个数}) * 5000(\text{外显子长度}) = 200$  或者： $1000(\text{reads 个数}) / 1(\text{百万}) * 5(\text{K}) = 200$  这个值反映基因的表达水平。FPKM (fragments per kilobase of exon per million fragments mapped)。FPKM 与 RPKM 计算方法基本一致。不同点就是 FPKM 计算的是 fragments，而 RPKM 计算的是 reads。Fragment 比 read 的含义更广，因此 FPKM 包含的意义也更广，可以是 pair-end 的一个 fragment，也可以是一个 read。

## 17. 什么是转录本重构

转录组本重构是指用测序的数据进行转录本组装拼接。有两种组装方式：1，de-novo 构建；2，有参考基因组重构。其中 de-novo 组装是指在不依赖参考基因组的情况下，将有 overlap 的 reads 连接成一个更长的序列，经过不断的延伸，拼成一个个的 contig 及 scaffold。常用工具包括 velvet，trans-ABYSS，Trinity 等。有参考基因组重构，是指先将 read 比对回到基因组上，然后在基因组通过 reads 覆盖度，junction 位点的信息等得到转录本，常用工具包括 scripture、cufflinks。

## 18. 什么是基因融合

基因融合 (genefusion) 是指将基因组位置不同的两个基因中的一部分或全部整合到一起，形成新的基因，称作融合基因，或嵌合体基因。该基因有可能翻译出融合或嵌

合体蛋白。