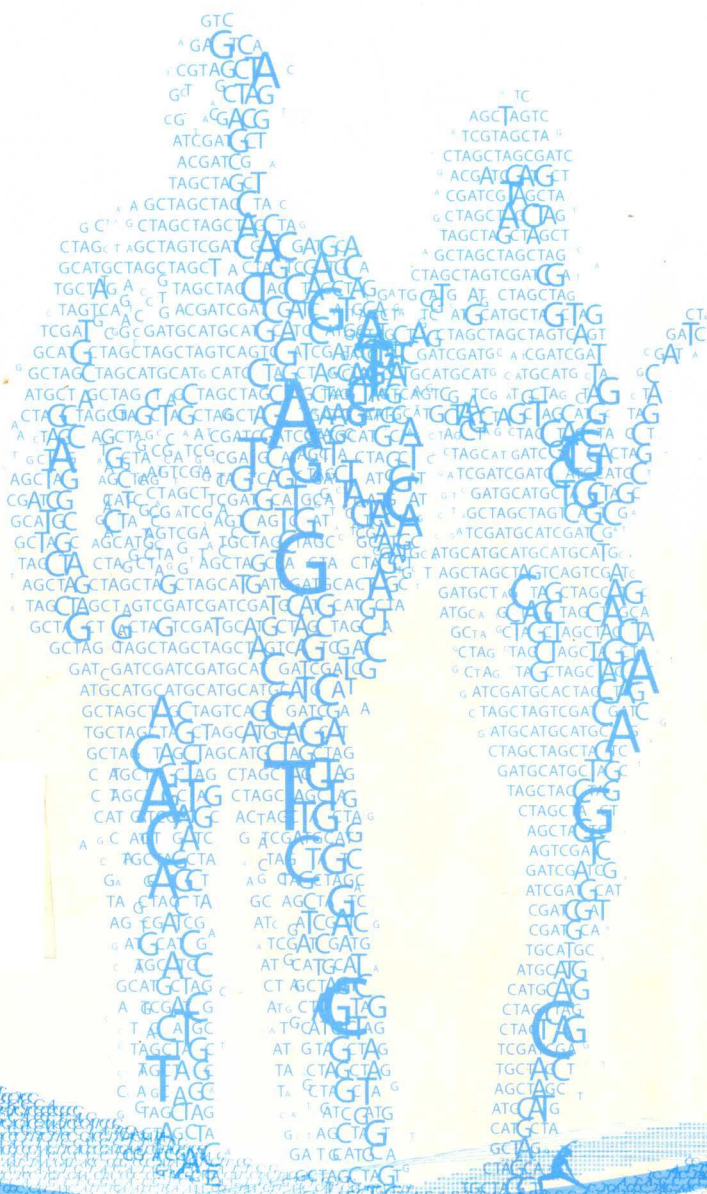


基于高通量技术的 人类疾病组学研究



进入21世纪后，以Roche公司的454技术，Illumina公司的Solexa技术和ABI公司的SOLiD技术为代表的新一代测序技术相继问世并迅猛发展，新一代测序技术越来越多的应用于解决生物学问题尤其是人类疾病的研究，使得疾病的研究模式发生了重大转变。从过去以假说为导向的研究模式转变为如今以数据为导向，现代化、工业化的研究模式，极大地加速了人们对于疾病的认识，为疾病的研究、诊断、预防及治疗提供了更为有效的手段。新一代测序技术的成功应用为疾病研究带来了新的机遇。

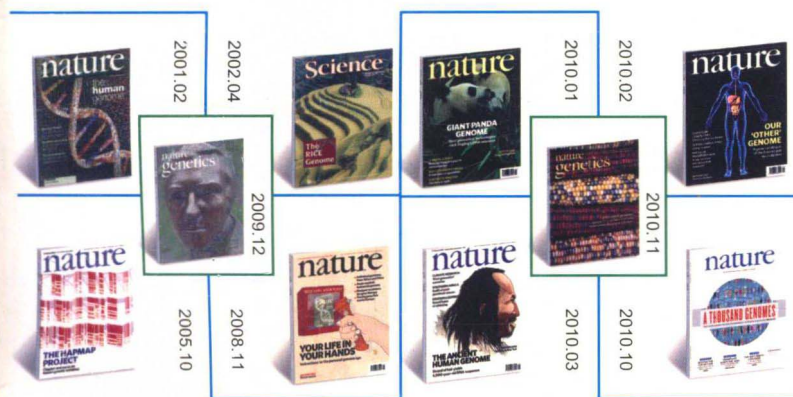
继人类基因组计划完成之后，国际单倍体型（HapMap）计划和千人基因组计划相继开展并积累了海量的遗传变异数据，尤其是2008年启动的千人基因组计划，它绘制出迄今为止最详尽的、最有医学应用价值的人类基因组遗传多态性图谱。这些遗传变异数据为更深层次上了解种族之间、个体之间的基因差异，以及为各种疾病的关联分析提供详细的基础数据，将极大地推动种群遗传学、人类疾病研究、比较基因组学及药物基因组学等研究。人类基因研究进入了一个划时代的新阶段。

华大基因通过高通量测序技术在DNA水平，RNA水平、表观遗传学水平及宏基因组水平对疾病展开全方位的研究，并结合质谱技术开展蛋白质组水平的研究，甚至是对各组学平台进行贯穿研究，由此得出的海量数据，与疾病相关的临床信息（发病年龄、治疗史、疗效、家族史、易感性等）相结合，为全面揭示疾病的生物学机制奠定了重要基础。

华大基因作为世界领先的基因组研究中心，拥有一流的高通量测序及生物信息分析能力，愿与相关专家一道致力于疾病的遗传背景研究，共同推动人类健康与医学研究的进步。

华大发表的顶级杂志文章

34 Nature group & Science publications.



- ▶ A Map of Human Genome Variation from Population-Scale Sequencing
Nature Oct. 2010
- ▶ Diversity of Human Copy Number Variation and Multicopy Genes
Science Oct. 2010

- ▶ The Chicken Genome
Nature Dec. 2004
- ▶ Domestication of Silkworm by Large Scale Resequencing
Science Oct. 2009
- ▶ Human Pan-Genome by Sequence Assembly
Nature Biotechnology Jan. 2010
- ▶ Epigenomics Map of Silkworm by Methylation Study
Nature Biotechnology May. 2010
- ▶ Genes for High Altitudes by Exome Sequencing
Science Jul. 2010
- ▶ Resequencing of 200 Human Exomes Identifies an Excess of Low-Frequency Non-Synonymous Coding Variants
Nature Genetics Oct. 2010
- ▶ Genome-Wide Patterns of Genetic Variation Among Elite Maize Inbred Lines
Nature Genetics Oct. 2010
- ▶ Resequencing of 31 Wild and Cultivated Soybean Genomes Identifies Patterns of Genetic Diversity and Selection
Nature Genetics Nov. 2010

人类疾病组学
研究策略
孟德尔遗传病研究
复杂疾病研究

29

疾病多水平研
究—测序技术
应用

DNA 水平
RNA水平
表观遗传水平
蛋白水平

1

人类进化与
群体遗传学
研究

47

孔凡

孔凡

孔凡

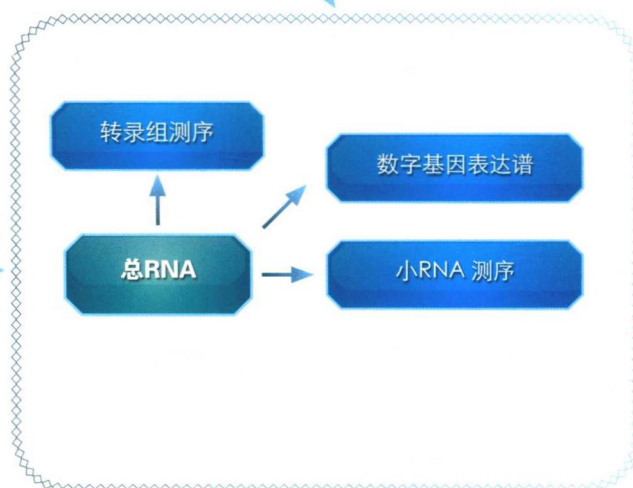
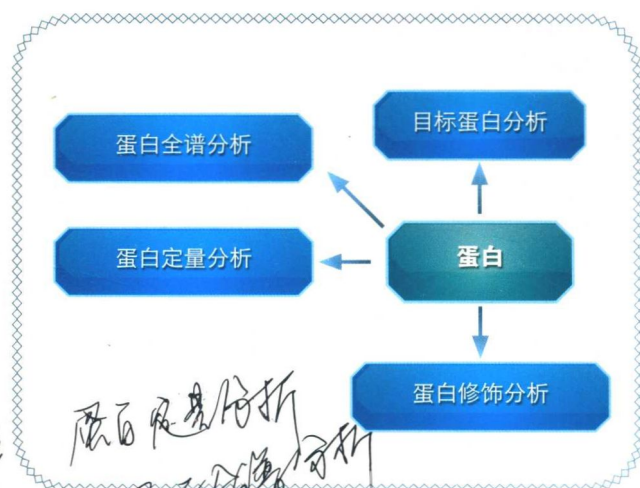
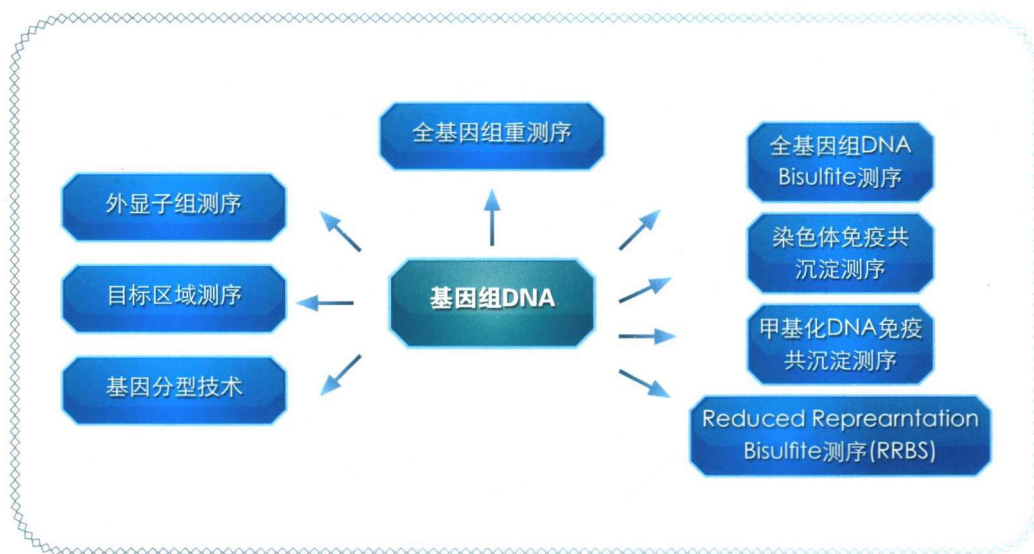
疾病多水平研究—测序技术应用

疾病多水平研究

测序技术应用

测序技术应用

疾病多水平研究—测序技术应用



DNA 水平

全基因组重测序

全基因组重测序(whole genome resequencing)是对基因组序列已知的不同个体进行全基因组测序,并在个体或群体水平上进行差异性分析的方法。

华大基因将短读长序列(Short-Reads)、双末端(Paired-End)和不同长度的插入片段(Insert-Size)技术相结合进行测序,更为全面地挖掘基因序列差异和结构变异,实现基因型多样性分析、遗传进化分析以及致病和易感性基因等的筛选。随着第二代测序技术的发展与普及,全基因组重测序已成为疾病研究、人类遗传和群体进化研究领域最为有效的方法之一。

技术优势:

- ☆可检测到基因组内较大的结构性变异及基因拷贝数变化等
- ☆可全面发掘新的和稀有的遗传变异
- ☆更加全面、可靠的基因变异信息
- ☆数据偏差小、高均一性,真实反应样本基因组信息
- ☆技术流程成熟

应用领域:

- ☆人类疾病与临床研究
- ☆药物基因组学研究
- ☆人类遗传学研究
- ☆群体遗传学及考古遗传学研究
- ☆比较基因组学研究

工作流程:

提取基因组DNA并随机打断,电泳回收所需长度的DNA片段,并加上接头进行cluster制备,最后上机测序。利用末端配对方法对插入片段两端进行测序,除获得序列信息外,还有利于大片段结构变异的挖掘,减少重复序列对组装的影响。

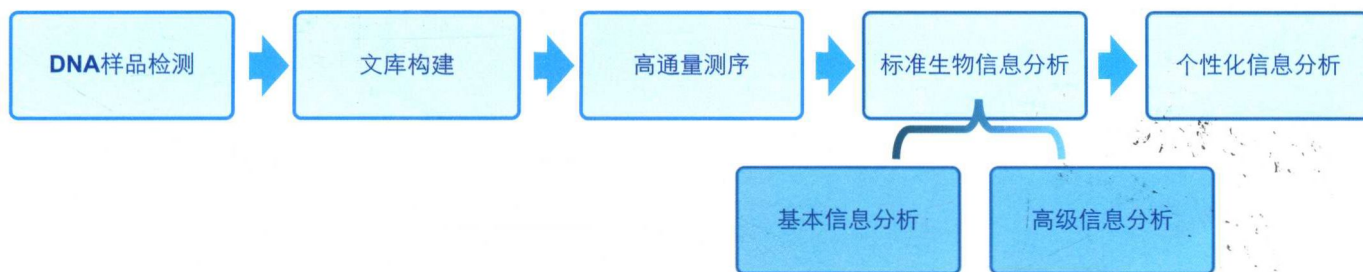


图1 全基因组重测序工作流程

信息分析：

在全基因组重测序中，通过序列间的比对，可以寻找到大量的单核苷酸多态性（Single Nucleotide Polymorphism, SNP），插入缺失（Insertion/Deletion, InDel）和结构变异（Structure Variation, SV）位点，利用生物信息分析软件对基因组间的差异进行注释（图2）。此外，根据基因组差异，在群体水平上还可进行基因组间进化关系和遗传变异多态性分析，如分析人群差异和群体历史，获得研究群体的连锁不平衡和单体型图（Haplotype）等数据。

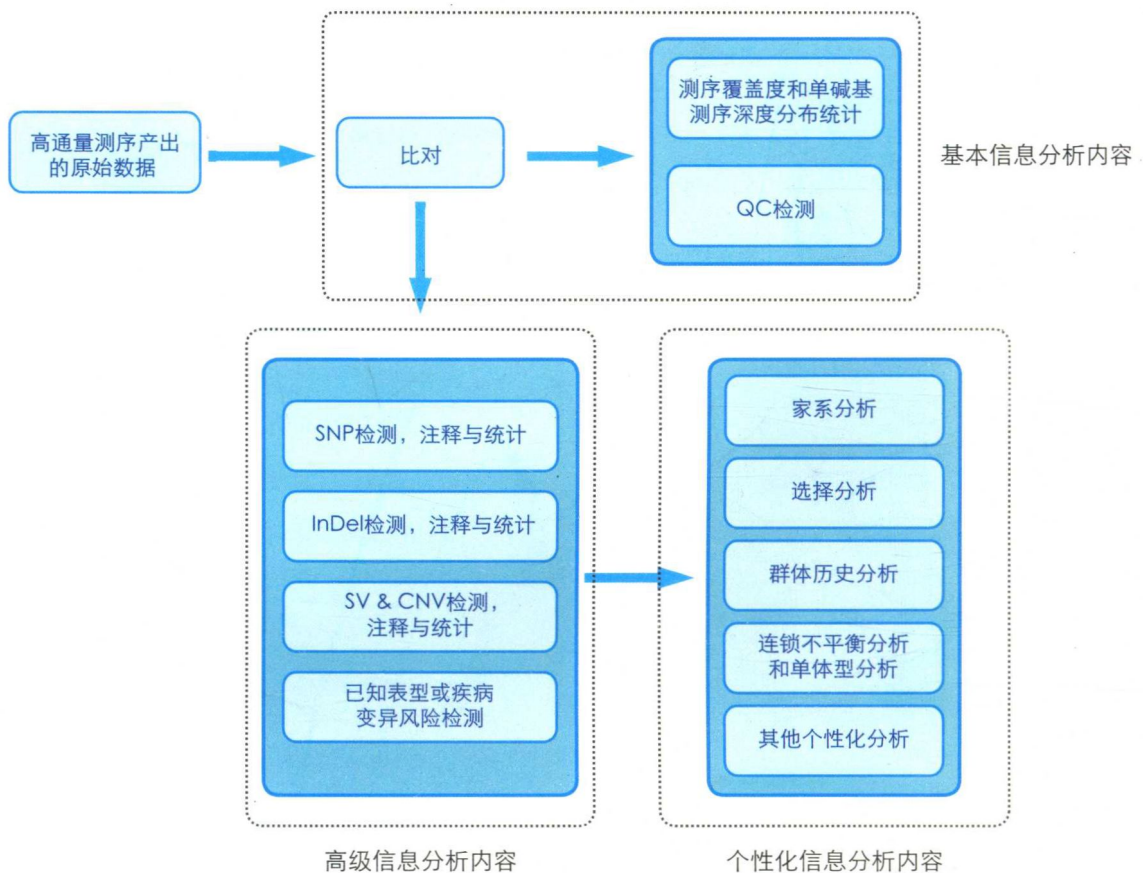


图2 人类全基因组重测序生物信息学分析流程

华大基因成功案例：

1. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*, 2010, 467: 1061-1073.
2. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, 2010, 463: 757-762.
3. The diploid genome sequence of an Asian individual. *Nature*. 6 November 2008; 456: 60-65.
4. The International HapMap Project. *Nature*, 2003, 426: 789-796.
5. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409: 860-921.

外显子组测序

外显子组测序技术 (Exome sequencing) 是利用Agilent SureSelect 或 NimbleGen SeqCap EZ 人全外显子组液相等富集系统捕获外显子区域DNA, 并结合第二代测序技术进行测序分析。

外显子组测序是一种新型的基因组分析技术, 只需测定 1% 的基因组序列, 就涵盖了与个体表型相关的大部分功能性变异。它以更快的方法确定致病基因及突变位点, 将有助于加速筛选发现单基因疾病, 癌症等复杂疾病的致病基因和易感基因等。

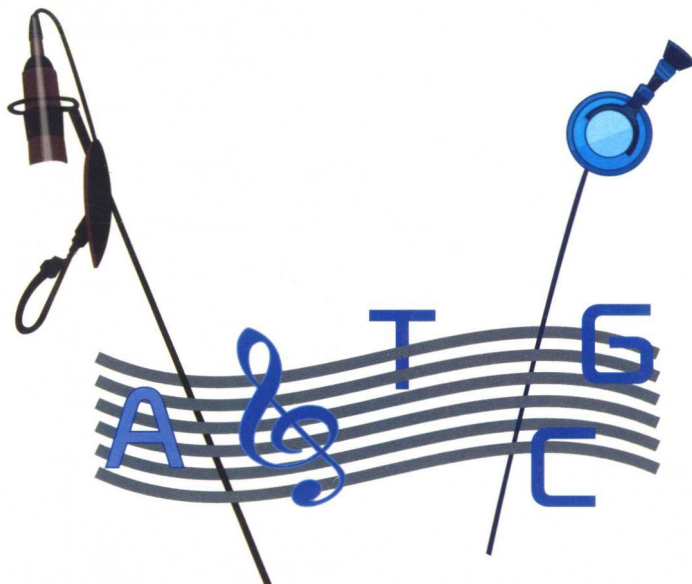
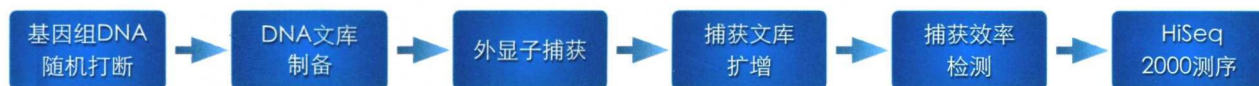
技术优势:

- ☆可发现外显子区的绝大部分的疾病相关变异
- ☆可发现常见变异和罕见变异, 并发掘新的变异位点
- ☆只对基因组的约1%测序, 更加经济、高效

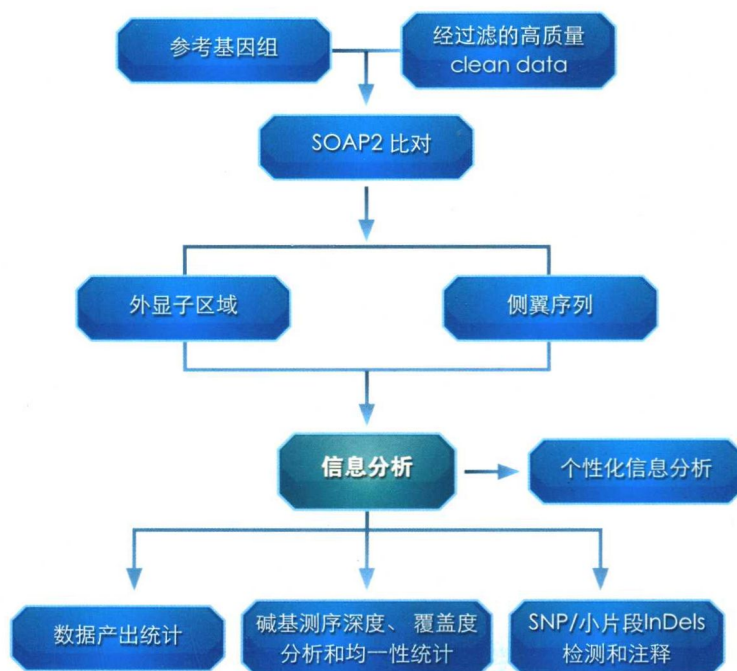
应用领域:

- ☆检测发现致病位点或易感位点的疾病机理研究
- ☆筛选发现疾病诊断或药物治疗的分子Marker
- ☆大规模群体遗传与进化、种群性状研究

工作流程:



信息分析:



捕获平台:

华大基因目前主要采用Agilent SureSelect和NimbleGen SeqCap EZ 人外显子靶向序列富集系统（图3、4）。这两个系统都采用液相系统进行高特异性和高覆盖率的外显子区域捕获。其优化的高通量操作流程，严格的内部质量控制，已成功适用于上千样本的外显子组测序研究。

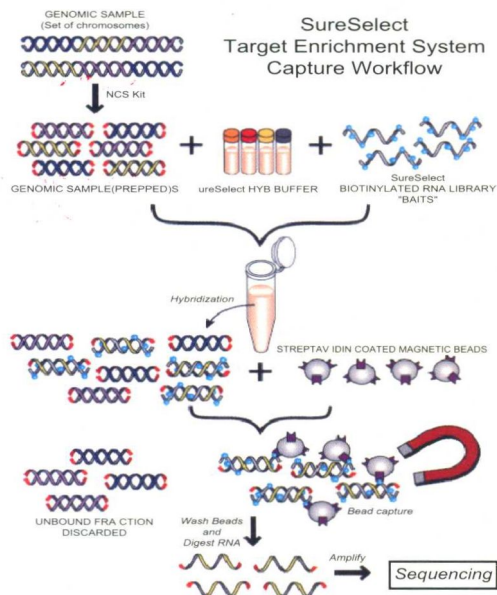


图3 Agilent SureSelect外显子捕获

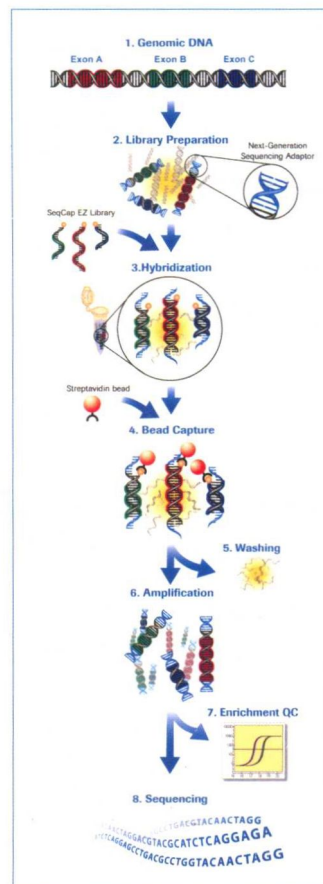


图4 NimbleGen SeqCap EZ外显子捕获

表 1 NimbleGen和Agilent外显子捕获平台探针的覆盖区域

Design	Capture Targets Mb (Regions covered by probes)	Database Used to Select Primary Targets
NimbleGen SeqCap EZ Exome v1	34	CCDS Jun2008 + miRbase v11, Apr 2008
NimbleGen SeqCap EZ Exome v2	44	CCDS Sep 2009 + miRbaes v14, Sep2009+RefSeq Jan2010
Agilent SureSelect Human All Exon v1(developed with Broad)	38	CCDS Sep 2008+miRBase V13
Agilent SureSelect Human All Exon v2(developed with Broad)	44	V1+additional RefSeq contents including CCDS sep2009
Agilent SureSelect Human All Exon 50Mb(developed with Sanger)	50	CCDS sep2009+miRBase V14+GENCODE+Sanger

华大基因成功案例

1. Li Y, Vinckenbosch N, Tian G, et al. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. **Nat Genet**, 2010, 42(11): 969-972.
2. Xin Yi, Yu Liang, Emilia Huerta-Sanchez, et al. Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. **Science**, 2010, 329(5987): 75-78.
3. Jun Ling Wang, Xu Yang, Kun Xia, et al. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. **Oxford Journals Medicine Brain**, 2010, 12(133): 3510-3518.

查看原文

捕获平台的原理

及优缺点



目标区域测序

目标区域测序(Target region sequencing)是针对感兴趣的基因组区域(如人的X染色体外显子区域等)设计并定制探针,与基因组DNA进行芯片杂交(NimbleGen Sequence Capture array)或溶液杂交(Agilent Sure-Select System),然后富集目标区域DNA,再利用新一代高通量测序平台进行测序。适用于单基因疾病、复杂疾病及癌症等多种人类疾病的研究。

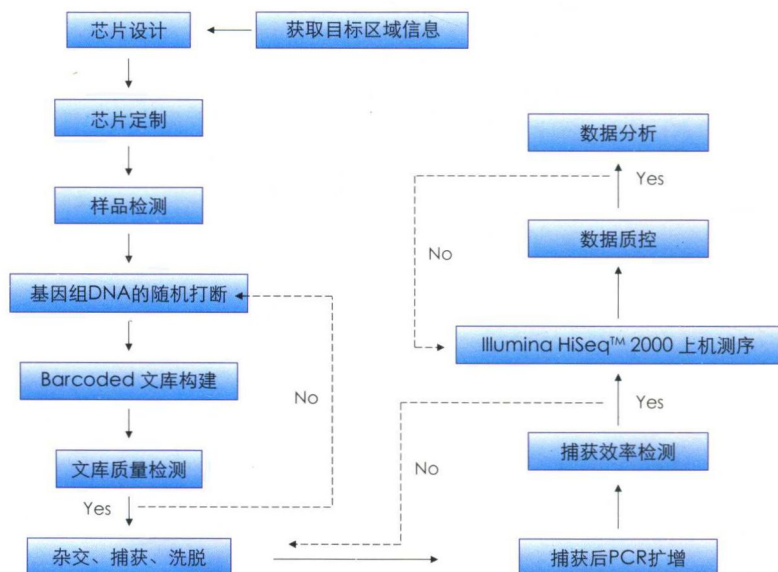
技术优势:

- ☆对感兴趣的基因组区域进行遗传变异位点的寻找
- ☆对特定区域进行深入研究
- ☆缩短研究周期-加快文章发表与加速临床应用
- ☆更高通量-适合大样本量的研究

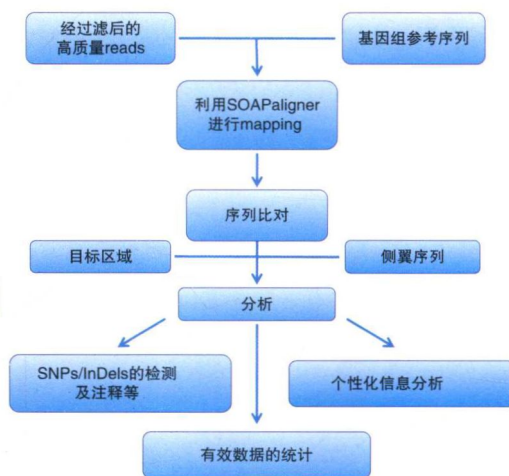
应用领域:

- ☆候选致病位点/基因的验证
- ☆新的致病位点/基因的寻找
- ☆疾病诊断
- ☆药物靶点设计
- ☆药物适合人群筛选
- ☆为代谢通路研究提供遗传学基础等

技术流程:



信息分析:



基因分型技术

单核苷酸多态性与结构变异是人群中最基本的遗传变异，这些变异是个体之间对于疾病易感性和对于疾病治疗效果不同的主要原因。基因分型技术（genotyping）通过比较不同个体的基因组序列发现SNP和结构变异，同时将这些变异与人类疾病进行关联分析，从而找到疾病相关的遗传变异。

技术优势：

- ☆可进行细胞遗传学分析，全基因组基因分型和拷贝数检测
- ☆可进行SNP、疾病相关位点的检测
- ☆个性化设计定制芯片，更灵活
- ☆技术成熟，数据可信度高
- ☆高通量，低成本

应用领域：

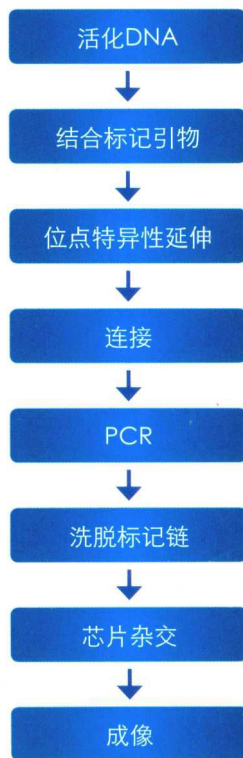
- ☆寻找与癌症及其他复杂疾病相关的致病基因
- ☆应用临床诊断与个性化治疗
- ☆大规模群体遗传与进化、种群性状研究

工作流程：

华大基因主要使用以下两种基因分型芯片：Infinium HD 芯片和GoldenGate 定制芯片，两种芯片的技术流程如下图：



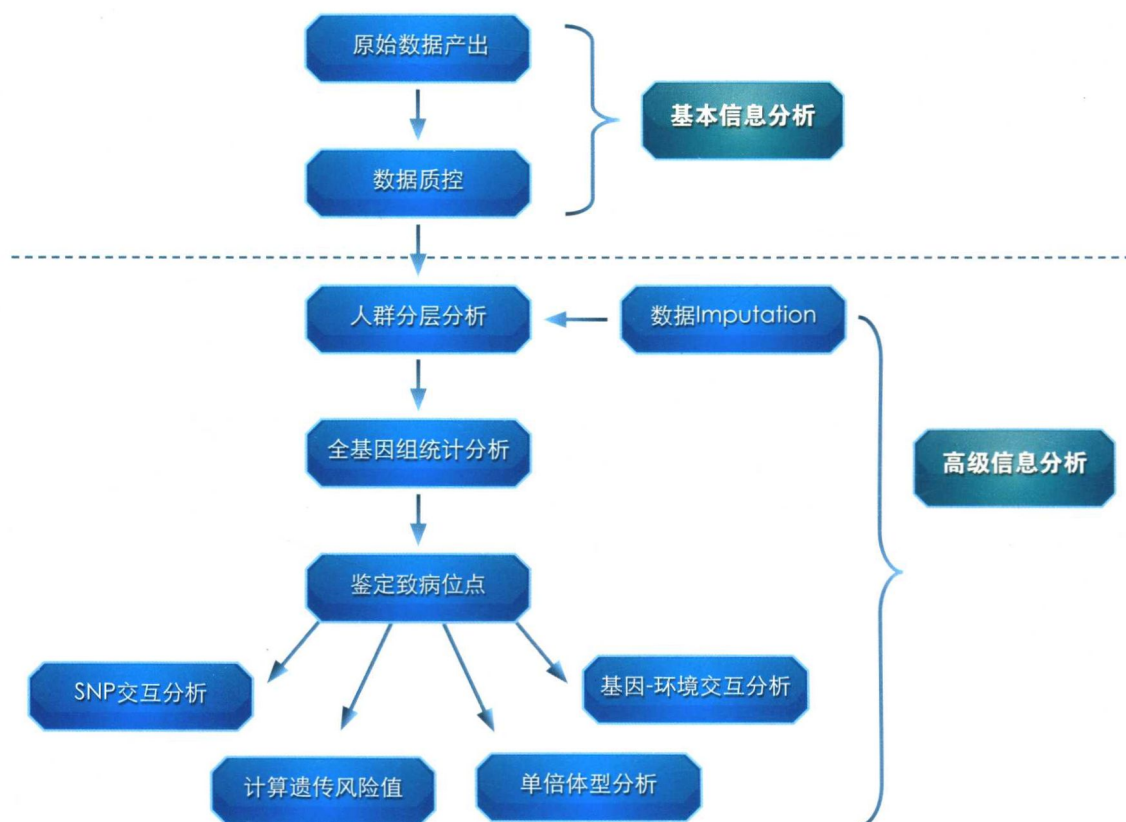
Infinium HD Assay 流程



GoldenGate Assay 流程

华大基因主要使用iScan扫描仪对各分型芯片进行扫描。iScan扫描仪，集高效能的激光器，光学检测功能于一体，具有高精度和无可比拟的通量产率。再加上强大的计算硬件和丰富的应用软件综合而成的计算服务平台，为海量数据的分析、处理与存储提供了坚实基础。

信息分析：



RNA水平

转录组测序

转录组测序(transcriptome resequencing) 即特定组织或细胞在某一功能状态下转录出来的所有RNA的总和, 包括mRNA和非编码RNA。转录组测序是基因结构和功能研究的基础。基于新一代高通量测序技术的转录组测序, 已广泛应用于生物医学研究、临床研究和药物研发等。

技术优势:

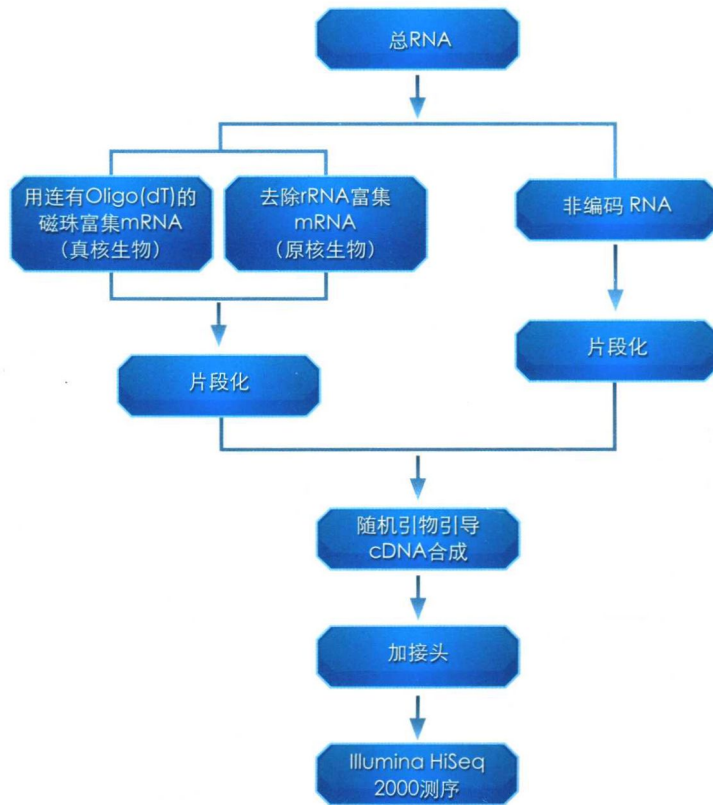
检测能力	无偏向性转录组测序; 不依赖于已知信息; 重复性好, 无需技术重复
检测信号	数字信号
检测范围	几乎所有转录本
检测精度	从几个到数十万拷贝精确计数
分辨率	可以检测单个碱基差异、基因家族中相似基因以及可变剪接造成的不同转录本的表达

应用领域:

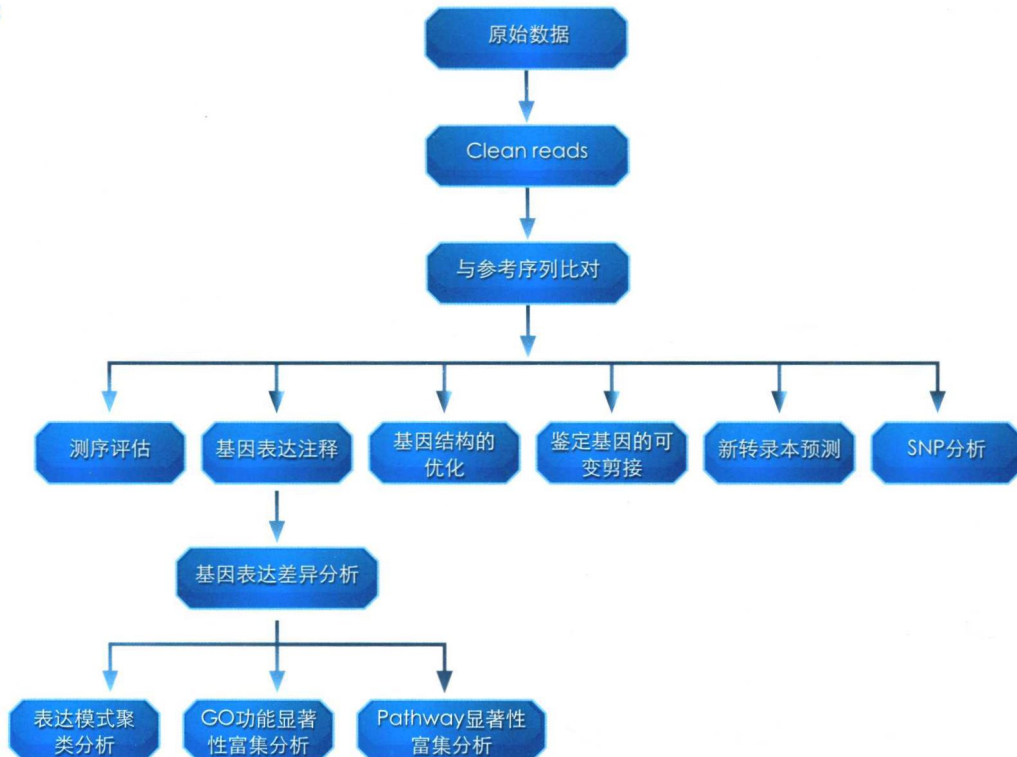
- ☆功能基因的筛选
- ☆分子标志和药物靶点的筛选鉴定
- ☆药理与毒理学研究
- ☆生物代谢调控研究
- ☆疾病分型研究



工作流程：



信息分析：



数字基因表达谱升级版

数字基因表达谱升级版(RNA-Seq(Quantification))是用来研究某一生物对象在特定生物过程中基因表达差异的技术,该技术结合了转录组测序建库的实验方法与DGE (Digital Gene Expression Tag Profiling) 的信息分析手段,具有定量准、可重复性高、检测范围宽、成本低等特点。

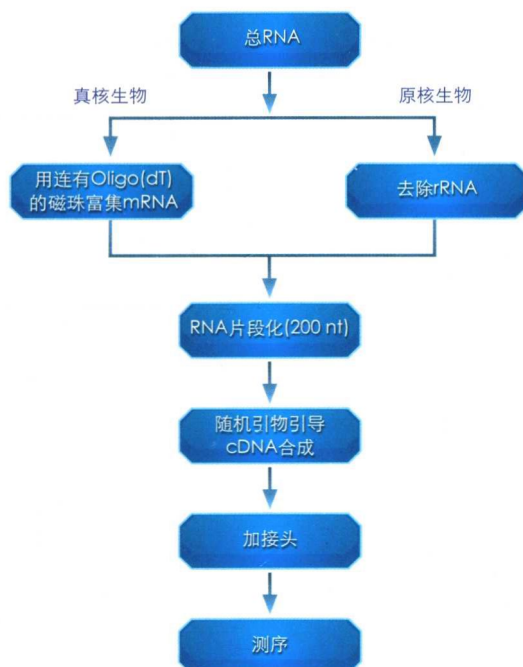
技术优势:

- ☆检测基因数比基因芯片多25%
- ☆定量准,可重复性高(重复相关系数 ≥ 0.99)
- ☆数字化信号,无背景噪音,无交叉杂交
- ☆高、低丰度基因均可检测
- ☆不受研究物种限制,模式生物和非模式生物均可检测
- ☆数据可与时俱进,即随数据库更新而更新
- ☆数据与基因芯片数据可无缝衔接,即RNA-Seq数据可用于基因芯片分析软件进行进一步信息挖掘

应用领域:

- ☆差异基因表达谱研究
- ☆遗传疾病机理研究
- ☆药理学或毒理学机制研究
- ☆全基因组表达图谱绘制

实验流程:



信息分析:

