

基因组学重点

第一章 基因组

1. **基因 (gene)**: 产生一条多肽链或功能性 RNA 所必须的全部核苷酸序列。
2. **基因组 (genome)**: 细胞或生物体中, 一套完整单倍体的遗传物质的总和。
3. **基因组学 (genomics)**: 涉及基因组作图、测序和整个基因组功能分析的一门学科。
4. **C 值 (C Value)**: 每一种生物中其单倍体基因组的 DNA 总量。
5. **C 值矛盾 / C 值悖论**: C 值和生物结构或组成的复杂性不一致的现象。

6. 基因组的序列组成:

- ① **单拷贝序列 (单一序列)**: 基因主要位于单一序列
- ② **轻度重复序列**
- ③ **中度重复序列**: 短散在序列、长散在序列
- ④ **高度重复序列**:

7. 编码 RNA 的基因:

rRNA 基因、tRNA 基因、scRNA 基因、snRNA 基因、snoRNA 基因、小分子干扰 RNA

8. **联合基因 (union of genomic sequence)**: 基因组中连续的 DNA 序列, 编码一组关联的彼此重叠的功能产物。

9. **假基因 (pseudogene)**: 具有与功能基因相似的序列, 但由于有许多突变以致失去了原有的功能, 所以假基因是没有功能的基因。

来源: 一般认为是由 mRNA 反转录成 cDNA, 然后整合在基因组中 (不含有内含子)。

假基因能否表达 (有功能), 为什么? 有些假基因是可以转录的, 特别是起源于重复基因的假基因和获得启动子的加工的假基因; 转录的假基因产生残缺的蛋白, 失去了原有的功能, 但可能产生新的功能。

第二章 遗传图绘制

1. **基因组作图 (genomic mapping)**: 确定界标或基因在构成基因组的各条染色体上的位置, 以及染色体上各个界标或基因之间的相对距离, 绘制遗传连锁图或物理图。

2. **遗传作图 (genetic mapping)**: 采用遗传学分析方法将基因或其他 DNA 顺序标定在染色体上构建连锁图 (单位: 厘摩 cM)。

3. **物理作图 (physical mapping)**: 采用分子生物学技术直接将 DNA 分子标记、基因或克隆标定在基因组的实际位置所构建的位置图 (单位: 厘镭 cR 或碱基对 bp)。

3. **基因标记**: **分类**: 形态标记、细胞学标记、生化标记、DNA 分子标记

特点 (缺点):

- ① 数量较多, 受环境影响小。
- ② 高等生物如脊椎动物和显花植物等, 可用作标记的基因十分有限, 许多性状都涉及多基

因。

③高等生物基因组中存在大量的基因间隔区，纯粹用基因作为标记将在遗传图谱中留下大片的无标记区段。

④只有部分基因其等位基因成员可以通过常规试验予以区分，因而产生的遗传图是不完整的，必需寻找其他有效的标记。

4. DNA 分子标记分为哪三大类：

①**限制性片段长度多态性 (restriction fragment length polymorphisms, RFLP)**：由于同源染色体同一区段 DNA 序列的差异，当用限制酶处理时，可产生长度不同的限制性片段。

②**简单序列长度多态性 (simple sequence length polymorphisms, SSLPs)**：可变排列的简单重复序列，即重复次数不一，在染色体的同一座位重复序列拷贝数不同。

③**单核苷酸多态性 (single nucleotide polymorphisms, SNPs)**：在人类两个个体之间大约每 500-1000bp 就有一个碱基差异，如果一个碱基位置发生的变异在 1%以上的人群存在，这个位点就被定义为 SNP 位点。

DNA 分子标记优点：

①不受时间和环境的限制；②遍布整个基因组，数量无限；③不影响性状表达；④自然存在的变异丰富，多态性好；⑤共显性，能鉴别纯合体和杂合体

5. **厘摩 CM**：同一染色体上不同基因之间的相隔距离就是根据两个基因之间在减数分裂时发生的交换频率确定的，其单位定义为 1%的交换率。

6. **重组热点 (recombination hot spot)**：染色体的某些位点之间比其他位点之间有更高的交换率。

7. 模式生物的连锁分析方法：

①**有性杂交实验** 根据需要有计划地实施杂交方案进行连锁分析，如果蝇、老鼠以及玉米、水稻等动植物的遗传学实验。

②**系谱分析** 不能进行有计划的遗传试验，只能收集家系成员的相关资料进行连锁分析，主要涉及人类以及多年生的树木等。

③**DNA 转移** 不发生减数分裂的生物，如细菌与病毒基因组的连锁分析。

8. 细菌的遗传作图：

①**接合 (conjugation)** 两个细菌机械接触，其中一个细菌（供体）将 DNA 转移到另一个细菌（受体）中。

②**转化 (transformation)** 供体细胞释放一段 DNA，经受体细胞摄取后整合到基因组中，可借助抗性培养基筛选重组克隆。

③**转导 (transduction)** 以噬菌体为媒介，将长度可达 50kb 的 DNA 片段从供体细胞转移到受体细胞。

第三章 物理图绘制

1. 遗传图的缺陷（为什么绘制物理图）：

- ①遗传图分辨率有限
- ②遗传图覆盖面较低
- ③遗传图分子标记的排列有时会出现差错

2. 物理作图的方法：

- ①**限制性作图（restriction mapping）**：将限制性酶切位点标定在 DNA 分子的相对位置。
- ②**基于克隆的基因组作图（clone-based mapping）**：根据克隆的 DNA 片段之间的重叠顺序构建重叠群，绘制物理连锁图。
- ③**荧光原位杂交（fluorescence in situ hybridization, FISH）**：将荧光标记的探针与染色体杂交确定分子标记的所在位置。
- ④**序列标签位点（STS）作图**：通过 PCR 或分子杂交将小段 DNA 顺序定位在基因组的 DNA 区段中。

3. 大分子 DNA 克隆载体：

（1）YAC（酵母人工染色体）：

组成部分：着丝粒；端粒；自主复制序列

缺点：

- i）存在高比例的嵌合体，即一个 YAC 克隆含有两个本来不相连的独立片段。
- ii）部分克隆子不稳定，在转代培养中可能会发生缺失或重排。
- iii）插入 DNA 片段的分离和纯化困难，因为 YAC 与酵母染色体具有相似的结构。
- iv）转化效率低。

（2）BAC（细菌人工染色体）：

优点：

- i）BAC 载体源于大肠杆菌中天然的 F 质粒，F 质粒与一般质粒有两点不同：单拷贝复制；相对分子质量大
- ii）F 质粒衍生的 BAC 载体具有较大的克隆容量，可达 30kb 以上，而且比较稳定。
- iii）因每个细胞只有一份 BAC 拷贝，克隆的 DNA 片段不会因重复发生嵌合问题。
- iv）BAC 的另一优点是，可采取类似制备质粒的方法直接提取克隆的 DNA 在大规模测序时便于机械化操作。

4. 指纹作图的方法：

（1）**限制性带型（restriction pattern）指纹**：用不同限制性酶处理样品，经凝胶分离产生 DNA 条带。如果 2 个克隆产生的 DNA 条带有部分是相同的，说明它们含有重叠的顺序。

（2）**重复序列 DNA（repetitive DNA）指纹**：将不同克隆的 DNA 限制性片段电泳后转移到杂交膜中，用一种或几种基因组范围的重复顺序作为探针与之杂交，如出现相同的杂交带型，可判断这 2 个克隆是重叠的。

（3）**重复顺序 DNA PCR（repetitive DNA PCR）或分散重复顺序 PCR（interspersed repeat**

element PCR, IPE-PCR) 指纹: 分散重复顺序在基因组中的密度很高, 2 个相邻重复顺序之间的单一顺序长度是不一致的, 设计 1 对与重复顺序互补的引物, 对检测的克隆进行 PCR 扩增可产生一系列单一顺序 DNA 条带, 这是一种特殊的指纹。如果 2 个克隆具有相同的 PCR 产物, 说明它们含有重叠的顺序。

(4) STS 作图 (STS content mapping) 指纹: 当需将某一克隆重叠群锚定到现有的已具有 STS 标记的物理图上时, 这一方法特别有效。因为 STS 是单一顺序, 并且序列已知。根据选定的某个 STS 序列设计引物, 可对大量的单个克隆进行 PCR 检测, 凡是能扩增出条带的克隆均含有顺序重叠的插入子。

5. 荧光原位杂交 (fluorescent in situ hybridization, FISH): 用特殊荧光素标记探针 DNA, 变形成单链后与变形后的染色体或细胞核靶 DNA 杂交。在荧光显微镜下观察并记录结果。

6. 使用 STS 的条件:

(1) 是一段已知的序列, 可据此设计 PCR 引物来检测不同 DNA 片断中是否存在这一序列。

(2) STS 在染色体上必须是独一无二的。如果在基因组中有多个位点出现, 作图数据将含混不清。

7. 寻找 STS 的方法:

(1) **表达顺序标签 (expressed sequence tag, EST):** 是一些从 cDNA 克隆中找到的小段顺序。因为细胞中的 mRNA 来自编码蛋白质基因, 因此 cDNA 代表了 mRNA 所在细胞中表达的基因。

(2) **SSLP:** 具有多态性并已在连锁分析中定位的 SSLP 具有特别的价值, 因其可直接建立遗传图与物理图之间的联系。

(3) **随机基因组序列:** 从克隆的 DNA 中随机测序可获得有用的序列, 也可从数据库中寻找所感兴趣的某些顺序。

8. 辐射杂种作图: 将人体细胞暴露在 3000~8000 rad 剂量的 X 射线中可引起染色体随机断裂, 辐射的 X 射线剂量越大, 产生的染色体片段越小。经强辐射处理的人体细胞很快死亡, 但若在辐射后立即将处理过的细胞与未辐射的仓鼠或其他鼠类细胞融合, 有些人体细胞染色体片段将会整合到鼠类染色体中进行扩增, 因此又称为辐射与融合基因转移。

原理: 基因组中单一序列标签位点 (STS) 都有独一无二的组成, 在染色体上的位置是确定的。位于染色体上近邻排列的 STS 是机械连接的, 在外力作用下断裂 DNA 或部分酶切降解 DNA 时, 2 个不同 STS 出现在同一片段的机会取决于它们在基因组中的相对位置。位置靠近的 STS 有很大机会出现在同一 DNA 片段中。相距越远, 出现在不同 DNA 片段中的概率越大。当两个片段含有同一 STS 序列时, 可以确认这两个片段彼此重叠。

方法:

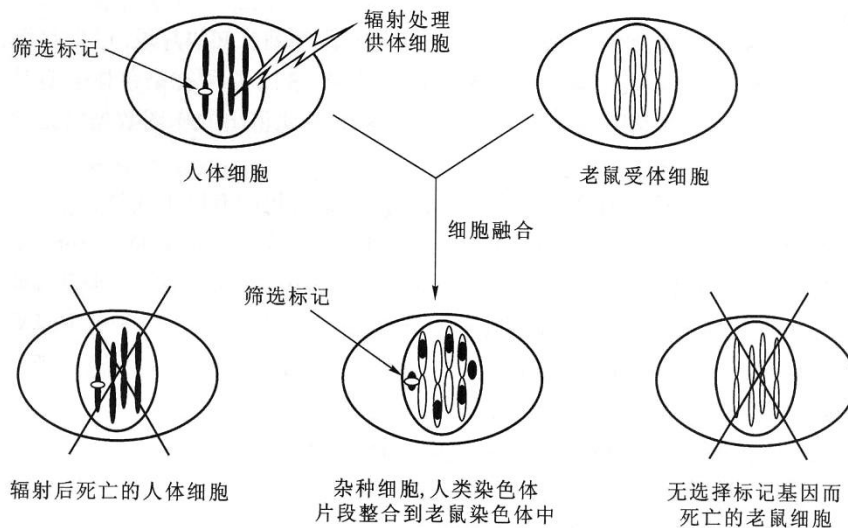


图 3.10 辐射杂种

作为供体的人体细胞含有标记基因, 经辐射处理后染色体被打断成长度不一的片段。X 射线辐射剂量越高, 产生的片段越小。辐射处理的人体细胞与仓鼠细胞融合后产生的辐射杂种可在筛选培养基上存活

第四章 基因组测序与序列组装

1. DNA 测序方法（第一代）：链终止法测序、化学降解法测序

2. DNA 测序的方法——链终止法

基本原理：

- ①通过合成与单链 DNA 互补的多核苷酸链, 由于合成的互补链可在不同位置随机终止反应, 产生只差一个核苷酸的 DNA 分子, 从而来读取待测 DNA 分子的顺序。
- ②利用脱氧核苷三磷酸 (dNTP) 的类似物双脱氧核苷三磷酸 (ddNTP) 取代正常的底物。

链终止法的技术要点：

- ①制备相同的单链模板 DNA
- ②将引物 (primer) 退火 (anneal)
- ③在 DNA 多聚酶催化下合成新链
- ④在链终止反应中加入少量链终止 ddNTP

链终止法对 DNA 多聚酶的要求：

①高酶活性

如果测序所用的多聚酶与模板结合能力强, 在终止核苷酸掺入新链之前不会脱离模板, 即不会提前终止反应。

②无 5' → 3' 外切核酸酶活性 大多数 DNA 多聚酶都有外切核酸酶活性, 5' → 3' 核酸酶活性可将新合成 DNA 链的 5' 端除去核苷酸从而改变链的长度, 给顺序的阅读造成困难与误差。

③无 3' → 5' 外切核酸酶活性 原因与②相同。

3、什么是序列间隙、物理间隙？如何填补这两类间隙

序列间隙 (sequence gap)：指测序时遗漏的序列，这些序列仍然保留在尚未挑选的克隆中。

解决方法：

首先收集所有已知间隙两侧的序列，然后根据间隙两侧的序列设计专一性探针，从基因组文库中筛选阳性克隆。由于每个间隙两侧的序列是已知的，可据此设计专一性 PCR 引物，随后两两配对直接从筛选到的阳性克隆中扩增，再进行克隆和测序。

物理间隙 (physical gap)：指构建基因组文库时被丢失的 DNA 序列，它们从已有的克隆群体中永久性的消失。

解决方法：

①利用一个不同的载体重新构建一个基因文库

②利用 PCR 方法，将不同重叠群末端序列作为引物两两配组扩增基因组 DNA

5. 作图法测序和鸟枪法测序有什么区别？

作图法：构建克隆群（遗传、物理图谱）、需要几年的时间、得到精细图谱

鸟枪法：不需背景信息、时间短、需要大型计算机、得到的是草图

6. 基因组鸟枪法测序要构建大小不同的插入片段克隆文库，原因何在？

①任何一种载体都会因某些插入片段与宿主菌的不兼容而不能扩增，使一些 DNA 片段丢失。不同类型的载体遭遇的不兼容是不相同的，因而第二种文库可以保留在第一种文库中可能失去的克隆片段。

②多种质粒文库也增加了克隆片段的总长，扩大了覆盖面。

③10kb 文库的构建有助于在 2kb 质粒来源的两端测序序列组装时校正由重复序列产生的差错

④Fosmid 和 BAC 文库的构建可以使下片段 DNA 文库组建的重叠群在大分子克隆中有效而准确地归并与整合，避免了在全基因组范围内直接进行重叠群排序所产生的错误，可提高序列组装的效率，保证序列组装的可信度。

第五章 基因组序列注释

1. 搜寻基因方法：

①根据开放读码框预测基因

②非编码序列、内含子

③密码子偏爱性

④外显子—内含子边界

⑤上游控制顺序

⑥软件预测

2. Northern 印记 (Northern blotting) : 在进行杂交实验时, 从样品中纯化的 RNA 经琼脂糖凝胶电泳分离, 然后转移到杂交膜上的过程。

3. 动物园杂交 (zoo-blotting) : 如果某一物种的 DNA 顺序与来自另一亲缘物种的 DNA 片段杂交产生阳性信号, 该区段可能含有 1 个或多个基因, 这种方法又称为动物园杂交。

4. 实验确认基因的实验: Northern 印记 动物园杂交

5. 基因功能的检测

①基因失活是基因功能分析的主要手段

②基因剔除

③基因打靶技术

④基因的过表达用于功能检测

⑤反义 RNA

6. 怎样预测基因功能?

同源搜索设计基因功能

根据蛋白质结构域预测基因功能

7. 组学包括: 基因组学 (genomics)、转录组学 (transcriptomics)、小 RNA 组学 (ribo-gnomics)、剪接组学 (spliceomics)、蛋白质组学 (proteomics)、代谢组学 (metabolomics)

8. 基因本体 (Gene Ontology, GO) : 是一个词汇表, 具有树形结构, 可以电脑操作和动态搜索。

三大基因本体定义 (内容) : 细胞组分 (cellular component)、分子功能 (molecular function)、生物学过程 (biological process)

9. 宏基因组/宏基因组学 (metagenome/metagenomics) : 指研究一类在特殊的或极端的环境下共栖生长微生物的混合基因组。

10. 人类第二基因组 (human second genome) : 人类体内的微生物有 1000 多种, 其遗传信息的总和称为人类微生物组 (human microbiome)

11. 转录组: 是在某一特定的条件下单个或一组细胞所具有的 mRNA 总和。

第十章 基因组表观遗传

1. 表观遗传 (epigenetic) : 所谓表观遗传就是不基于 DNA 差异的核遗传。即细胞分裂过程中, DNA 序列不变的前提下, 全基因组的基因表达调控所决定的表型遗传, 涉及染色质重编程、整体的基因表达调控 (如隔离子, 增强子, 弱化子, DNA 甲基化, 组蛋白修饰等功能), 及基因型对表型的决定作用。

2. 表观遗传机制:

DNA 甲基化 (DNA methylation)

基因组印记 (genomic imprinting)

染色质重塑 (chromatin remodeling)

位置效应 (position effect)

3. 位置效应: 因染色体不同区段的结构不同而影响基因表达的现象。

4. 座位控制区 (locus control region, LCR): 仅涉及一些特定的 DNA 顺序的位置效应, 它们位于一组基因的上游, 可以控制染色质的结构变化。

5. 绝缘子 (insulator): 这种可以阻止邻近位置激活或失活效应的序列。

6. 副突变 (paramutation): 在杂合子中一等位基因影响同一座位上另一等位基因的表型。

7. DNA 甲基化怎样调控基因表达?

DNA 甲基化与基因表达在 2 种水平控制基因的活性, 即局部范围影响单个启动子与增强子的功能及在大范围如整条染色体和基因组水平影响许多基因的表达。

8. 基因组印记 (genomic imprinting): 或亲代印记, 因为等位基因来自上一代不同性别的亲本而使表达模式发生可遗传变化的现象。

9. 核小体相位 (nucleosome phasing) (核小体定位): 指一段顺序确定的 DNA 与核小体核心八聚体的可变结合方式。

10. 表观遗传密码 (epigenetic code): 是一种假说, 表示每个真核细胞所具有的表观遗传修饰的总和, 由所有组蛋白密码和 DNA 甲基化修饰组成。

12. 组蛋白密码 (Histone code): 组蛋白密码是一种假说, 其含义是, 组蛋白的化学修饰可部分地调控储存在 DNA 中的遗传信息的表达。

第十三章 基因组进化的模式

1. 原基因组 (protogenome): 由原始核酶 RNA 复制形成的能自我复制和指令简单生化反应的体系。

2. 新基因的产生有哪些方式?

① 基因加倍之后的趋势, 这类基因基本保持原有的基因功能, 但往往获得了新的表达模式, 这是新基因产生的主要方式

② 外显子或结构域洗牌 (真核生物约 19% 的基因产生与外显子洗牌)

③ 逆转录及其随后的趋异或重拍

④ 外源基因的水平转移

⑤ 非编码序列转变为编码序列

3. 外显子洗牌 (exon shuffling): 不同的结构域加倍或重组, 产生具有创新功能的基因。

4. 基因冗余: 一条染色体上出现一个基因的很多复份(复本)。

5. 从进化的角度看, 基因冗余包括?

- ①年轻的重复基因
- ②正在向新功能基因过渡的重复基因
- ③保留着部分功能重叠的重复基因

DemonYukiku