

1. Homology（同源性）：

- 同源性是指两个或多个基因序列、蛋白质或其他生物分子在进化上的相似性。同源基因通常由共同的祖先基因演化而来。在基因组学中，研究同源性有助于理解基因的功能和进化关系。

2. SNP（Single Nucleotide Polymorphism，单核苷酸多态性）：

- SNP指的是在某个基因组位置上的单个核苷酸（A、T、C或G）的变异。这些变异是人类和其他物种遗传多样性的主要来源，并且可以用于疾病关联研究和个体化医疗。

3. ENCODE计划（Encyclopedia of DNA Elements）：

- ENCODE计划是一个国际研究项目，旨在识别并分类人类基因组中的所有功能元素，包括基因、调控序列和其他功能性区域。该项目帮助科学家更全面地理解基因组的结构和功能。

4. Phenotype（表型）：

- 表型是指生物体的可观察特征或性状，如身高、眼睛颜色或疾病状态。表型是基因型（遗传信息）和环境相互作用的结果。

5. UTR（Untranslated Region，非翻译区）：

- UTR指的是位于蛋白质编码序列之前（5' UTR）或之后（3' UTR）的基因组区域。这些区域虽然不编码蛋白质，但对基因的表达调控和mRNA的稳定性具有重要作用。

6. SBS测序（Sequencing by Synthesis，合成测序）：

- SBS测序是一种DNA测序技术，通过在合成新的DNA链时逐个添加并检测核苷酸来确定DNA序列。这种方法通过监测在DNA合成过程中发生的化学变化来识别每个新增加的核苷酸。

7. Shotgun测序（霰弹枪测序）：

- 霰弹枪测序是一种将长的DNA片段随机分解成许多小片段，然后分别进行测序的技术。之后，通过计算机软件将这些小片段重组，以得到完整的DNA序列。这种方法特别适用于大型基因组的测序。

8. Scaffold（支架）：

- 在基因组组装中，支架指的是由多个序列片段（通常是联接子）组成的更大片段。这些片段通过已知的空白区域（通常是通过插入序列估计的）连接在一起，以辅助构建更完整的基因组序列。

9. Transcriptome（转录组）：

- 转录组是指在特定细胞、组织或生物体在特定条件下的所有RNA分子的总和，包括mRNA、rRNA、tRNA和非编码RNA。转录组分析有助于理解基因在不同条件下如何被表达和调控。

10. Metabolomics（代谢组学）：

- 代谢组学是研究生物体内所有代谢物的组合及其变化的科学。通过分析细胞或组织中的代谢产物，代谢组学可以帮助科学家了解生物过程和疾病状态。

简答题

1. HGP计划的四个图谱:

- 连锁图(Linkage map):表示基因或DNA标记在染色体上的相对位置与遗传距离的图谱。
- 物理图:以STS位点(Sequence-Tagged Sites)为图标,表示基因组的物理大小或标记距离的图谱。
- 转录图:所有基因的转录本(transcript)序列的总和。
- DNA全序列图:目标是获得基因组的完整DNA序列。

2. CLIP-seq的原理、建库步骤和应用场景:

- 原理:结合免疫沉淀和高通量测序,通过紫外线使RNA结合蛋白和RNA形成共价键,然后进行免疫沉淀、测序和分析,以鉴定RNA结合蛋白与RNA的结合位点。
- 建库步骤:细胞固定和裂解,RNA部分降解,免疫沉淀,提取RNA并制备测序库,进行高通量测序。
- 应用场景:用于研究RNA结合蛋白与RNA之间的相互作用,揭示RNA结合蛋白的功能和作用机制。

3. SBC测序法没有广泛应用的原因:

- 芯片制作成本高昂,难以批量化生产。
- 测序读长较短,难以进行后续分析。
- 测序通量较低,难以满足大规模项目的需求。
- 测序准确度不够高,错误率较大。
- 相比NGS技术,SBC在成本、效率、准确性等方面均显劣势。

4. HGP对生命科学领域的贡献:

- 建立了大规模国际科研合作的典范。
- 推动了测序技术的进步和生物信息学的发展。
- 揭示了人类遗传信息蓝图,加深了对生命奥秘的理解。
- 为后续的功能基因组学研究奠定了基础。
- 带动了诸如ENCODE、1000基因组等相关项目的开展。
- 为精准医学、个体化治疗提供了遗传基础。
- 对医学、农业等领域也产生了深远影响。

问答题举例

1. 疾病诊断和个体化医疗

- 目的：诊断遗传性疾病，为个体提供定制化治疗方案。
- 方法：使用全基因组测序技术（Whole Genome Sequencing, WGS）或全外显子组测序技术（Whole Exome Sequencing, WES）来分析个体的基因组。通过这些技术，医生可以识别与特定疾病相关的遗传变异。
- 应用实例：在乳腺癌治疗中，通过检测BRCA1和BRCA2基因的突变，可以确定患者是否适合接受特定的药物治疗，如PARP抑制剂。

2. 进化生物学

- 目的：了解物种的进化历史和生物多样性。
- 方法：利用基因组比较分析（Comparative Genomics），通过比较不同物种的基因组序列，寻找进化上的共同特征和差异。
- 应用实例：通过比较人类与黑猩猩的基因组，科学家能够识别出影响认知发展和大脑功能的关键遗传因素。

3. 农业改良

- 目的：提高作物产量、抗病性和营养价值；改善畜牧业。

- 方法：使用基因编辑技术，如**CRISPR-Cas9**，对作物或牲畜的基因组进行精确修改。此外，通过基因关联研究（**Genome-Wide Association Studies, GWAS**）识别影响特定性状的遗传标记。
- 应用实例：在水稻中引入耐盐性基因，以提高其在盐碱地生长的能力；或通过基因编辑增强某些作物的营养成分，如富硒大米。

4. 微生物学研究

- 目的：了解病原体的基因组结构和功能，开发新的治疗方法。
- 方法：使用宏基因组学（**Metagenomics**）技术分析微生物群落的基因组。对病原体进行深度测序，以确定其遗传变异和抗药性机制。
- 应用实例：在抗生素耐药性研究中，通过分析病原菌的基因组变异，识别导致抗药性的特定基因，从而帮助开发新的抗生素。

5. 生态遗传学

- 目的：研究物种如何适应不同的环境条件。
- 方法：利用群体基因组学（**Population Genomics**）分析物种在不同环境下的遗传变异。通过表观遗传学（**Epigenetics**）研究环境如何影响基因表达。
- 应用实例：研究气候变化对北极熊基因组的影响。通过分析在温暖和冷冻环境下北极熊的基因表达模式，科学家能够了解这些动物如何适应极端的气候条件。

6. 基因表达调控网络分析

- 目的：理解基因如何在不同条件下被调控以及它们之间的相互作用，从而揭示基因功能和疾病机制。
- 方法：使用深度学习模型，如卷积神经网络（**CNN**）或递归神经网络（**RNN**），分析大规模转录组数据。这些模型能够从复杂的基因表达数据中学习并识别模式，预测基因之间的相互作用。
- 应用实例：利用深度学习分析癌症患者的转录组数据，识别特定基因表达模式与癌症发展的关联，从而为癌症诊断和治疗提供新见解。

7. 基因组变异与疾病相关性研究

- 目的：确定哪些基因变异与特定疾病相关，为疾病预防和治疗提供靶点。
- 方法：运用深度学习算法，如变分自编码器（**VAE**）或生成对抗网络（**GAN**），分析大量基因组数据，特别是**SNP**数据。这些算法能够从海量的遗传数据中提取有用信息，预测疾病风险。
- 应用实例：通过分析成千上万个个体的基因组数据，深度学习模型可以识别出与阿尔茨海默病等神经退行性疾病相关的特定基因变异。

8. 个性化药物反应预测

- 目的：预测个体对特定药物的反应，以实现个体化的医疗治疗。
- 方法：利用深度学习技术，如深度神经网络（**DNN**），分析基因组数据和患者的临床信息。这些模型可以从复杂的数据集中学习并预测个体对特定药物的反应和副作用。

- **应用实例：**在心脏病治疗中，通过分析患者的基因组数据和历史医疗记录，深度学习模型可以预测患者对某些抗凝血药物的反应，比如华法林。这有助于医生选择最适合患者的药物，并调整适当的剂量，以减少副作用和提高治疗效果。