了解基本概念和idea：

1. 什么是DNA/RNA/Protein, 什么是SNP （SNV），INDEL, CNV, mRNA，miRNA, gene expression 等
2. 什么是GWAS； candidate gene； pathway；quality control；Fine-mapping, locus zoom (region plot)；network analysis; association vs. causal?
3. 统计模型：regression model，survival analysis，meta-analysis，差异表达，co-expression，gene set enrichment analysis, pathway analysis, eQTL, polygenetic risk score, pharmacogenetics (personalized medicine) 等
4. 公开的数据库有哪些，比如TCGA，Oncomine，GEO等
5. 如何编程（参考 生信技能树），可视化结果

练习任务：

1. 找到糖尿病（GWAS）关联的的基因（位点，top SNP）？
2. 找乳腺癌（breast cancer）的风险基因有哪些？
3. 找到clopidogrel 代谢相关的基因有哪些？

生物信息研究问的问题：

1. 针对什么病症的科学研究？为什么要做这方面的研究？
   1. 比如每年发病人数多少？多少死亡？花费如何？
   2. 目前该病难以预防或者治疗吗？难点在哪？
2. 发病的特征是什么？ 有什么特殊性？个体间差异如何？易感人群有哪些？
   1. 比如发病率如何？传染性或者遗传性如何?
   2. 发病前，发病后有什么特征？
   3. 不同人之间病症等有没有差异？差异是什么？原因是什么？
   4. 哪些人容易得病?目前有没有预防措施？预防做得怎么样？
   5. 目前有哪些治疗方案？包括用药，手术等？各自效果如何？
   6. 该病是不是复发性的？多长时间容易复发？复发后有什么干预手段？
3. **研究的科学目的是什么，研究内容和科学意义分别是什么？**
4. **数据来源 （materials等）和分析方法（method）**
   1. 样本量多大？什么人群的数据？数据质量如何？
   2. 临床数据有没有？ 重要的风险因子是什么？
   3. 遗传数据有没有？
      1. 测序：
         1. 测序内容有哪些？DNA 全基因组？ 外显子？Array？甲基化？mRNA？miRNA？启动子？Copy number variants? Indels? 蛋白表达？
         2. 测序仪器有哪些？测序价格，测序速度如何？
         3. 测序后得到的原始数据格式是怎样的？
      2. 数据分析方法有哪些？
         1. 有没有被广泛认可并使用的方法？ （看综述和原文章）
         2. 用了哪些软件？
      3. **数据分析的pipeline**
         1. 数据清洗，整理等
         2. 数据标准化
         3. 数据分析软件列表
         4. 每一步分析数据的详细流程和code
         5. 数据分析的输出和解读，可视化等
      4. 重复已经发表的数据分析
         1. 重复发表的文章数据分析
         2. 挖掘数据库资源
      5. 有没有可能改进数据分析方法，如何改进
         1. 原始方法的限制在哪
         2. 有没有可能在原来方法上面做出改进，比如提高运行速度，比如使用其他模型等等
5. 风险预测模型如何建立（临床 + 遗传等）？
   1. 分析的表型（phenotype）是什么？
   2. 运用什么分析方法：关联性分析？聚类分析？等
   3. 精准医疗模型如何建立
6. 得病后如何治疗，有哪些治疗方案？ 可否做精准（个性化）治疗？
7. 未来发展方向？