生物信息分析入门

来源：

<http://mp.weixin.qq.com/s?__biz=MjM5MDIxNTQ4MA==&mid=212994084&idx=1&sn=760fff894aba9251741799bdd4a2c8d6&scene=2#rd>

生物信息学是生命科学研究的重大前沿领域，未来将占据生命科学研究的半壁江山。已经有越来越多的小伙伴投入到生物信息的学习中，但是入门难、深入慢、摸不到方向等都成为持续学习的拦路虎。

**谁可以来学**

什么是生物信息学？现在已有的定义都非常宽泛，例如这个：“生物信息学利用应用数学、信息学和计算机科学的方法研究生物学的问题”。感觉好像是非生物学背景的人从事的实践。

实际上，当你真正从事生物信息工作或研究时，可能会发现你需完成的事情大致分为三类：一是开发、设计生物信息学方法和技术，构建数据库；二是用成熟的方法、技术和数据库，来研究自己的问题；三是两者结合。

**由此可见，生物学背景、数学背景、计算机背景、统计学背景等，对生物信息感兴趣，有意向从事生物信息方向的工作或科研的人都可以参与生物信息学的学习。**

**你需要学什么**

从上面关于生物信息学的定义和具体从事的实践工作，可以看出，要胜任生物信息学方面的研究或工作，需要具备以下几方面的知识：**1、计算机基础知识；2、生物学基础知识；3、生物信息基础知识。**

1、计算机方面，你需要掌握linux操作基础，学会Perl语言和R语言。C语言和java也是不错的编程语言。但在生物信息领域使用比较多的还是Perl和R。现在使用Python人也越来越多。

2、非生物背景的人，需要特别加强了解生物学方面的知识。系统地学习一两门生物学课程是非常必要的，同时还需要根据研究或工作需要及时补充学习相关知识。

3、生物信息方面，首先需要了解生物信息学的一些基本概念和研究内容等，这方面可以选择一两本生物信息学教材来学习；其次学习掌握测序、数据库、数据格式等基础知识，学习关于生物信息的研究领域，如全基因组、外显子、目标区域捕获等，以及生物信息的应用领域，如肿瘤筛查、产前诊断、个体化医疗等。这方面可以查看一些大牛的综述进行学习。

a) 生信基础软件(blast，blat，fastqc，blast，clustw，phylip等)  
b) NGS variant calling相关软件（bwa，samtools，picard，GATK，VarScan，beagle等）  
c) 基因组相关软件（velvet，SOAPdenovo2，repeatmasker, Glimmer, orthMCL等）  
d) 转录组相关软件(bowtie,，trinity，tophat，cufflinks，cuffdiff，DEseq， SOAPfuse等）

**初学者必读的经典书籍**

**导论**  
生物信息学陈铭主编，科学出版社  
Fundamental Concepts of Bioinformatics (USA) Dan E. Krane & Michael L.Raymer (2003) Pearson Education Post-genome Informatics Minoru Kanehisa （金久时）(2001) Oxford University Press Introduction to Bioinformatics Teresa K. Attwood etc. (1999 ) Prentice Hall   
**专著**  
Bioinformatics: Sequence and Genome Analysis (USA) David W. MountComputational Molecular Biology: An Algorithmic Approach, Pavel A. Pevzner(2000) MIT Press Statistical Analysis of Gene Expression Microarray Data Terry Speed (2003)Chapman & Hall/CRC

**每天必上的Bioinformatics网站**

**数据库**

NCBI   
http://www.ncbi.nlm.nih.gov/   
Ensembl   
http://www.ensembl.org/   
Gene Ontology   
http://www.geneontology.org/

UCSC   
http://genome.ucsc.edu/

Genecard

http://www.genecards.org/

UniprokUniprok

http://www.uniprot.org/

**主要的生物信息学期刊网站**

bioinfomatics   
http://bioinformatics.oxfordjournals.org/archive/   
BMC Bioinfomatics   
http://www.biomedcentral.com/bmcbioinformatics/   
PLoS Computational Biology   
http://www.ploscompbiol.org/home.action   
Nucleic Acids Research   
http://nar.oxfordjournals.org/

Genome Research

http://genome.cshlp.org/

SCI-hub

http://sci-hub.org

**植物**

http://www.arabidopsis.org/   
http://www.mirbase.org/

**统计**

http://stats.stackexchange.com/

http://www.biostatistic.net/portal.php

**中文**  
http://www.plob.org/

http://www.yunbio.com/（云生物）

**论坛**

BioBB

http://www.bioinformatics.org/pipermail/bbb/

BioC

https://stat.ethz.ch/mailman/listinfo/bioconductor

BioPerl

http://www.bioperl.org/wiki/Mailing\_lists

BioPython

http://lists.open-bio.org/pipermail/biopython/

BioRuby

http://lists.open-bio.org/pipermail/bioruby/

BioStar

http://biostar.stackexchange.com/

Blue Obelisk

http://blueobelisk.shapado.com/

CCP4 bulletin board

https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=CCP4BB

gmx-users

http://lists.gromacs.org/pipermail/gmx-users/

MetaOptimize

http://metaoptimize.com/qa/

Molecular Station

http://www.molecularstation.com/

Protocol Online

http://www.protocol-online.org/

R-help

https://stat.ethz.ch/mailman/listinfo/r-help

Stackoverflow

http://stackoverflow.com/

Stats

http://stats.stackexchange.com/

GATK Support Forum

http://gatkforums.broadinstitute.org/

**其他**

https://david.ncifcrf.gov/home.jsp

http://www.genome.jp/kegg/   
http://www.uniprot.org/  
http://www.ebi.ac.uk/   
http://www.expasy.org/   
SEQanswers   
http://www.seqanswers.com/   
BioStar   
http://www.biostars.org/