

实验 3.1 数据库查询与文献检索

(1 学时)

一、实验背景

在科研实践中，高效查询数据库与文献库至关重要。在本项目中，多个实验环节依赖数据库的使用和文献查询。在本次实验中，我们选择生物信息学常用数据库进行基本功能和查询方法的介绍。

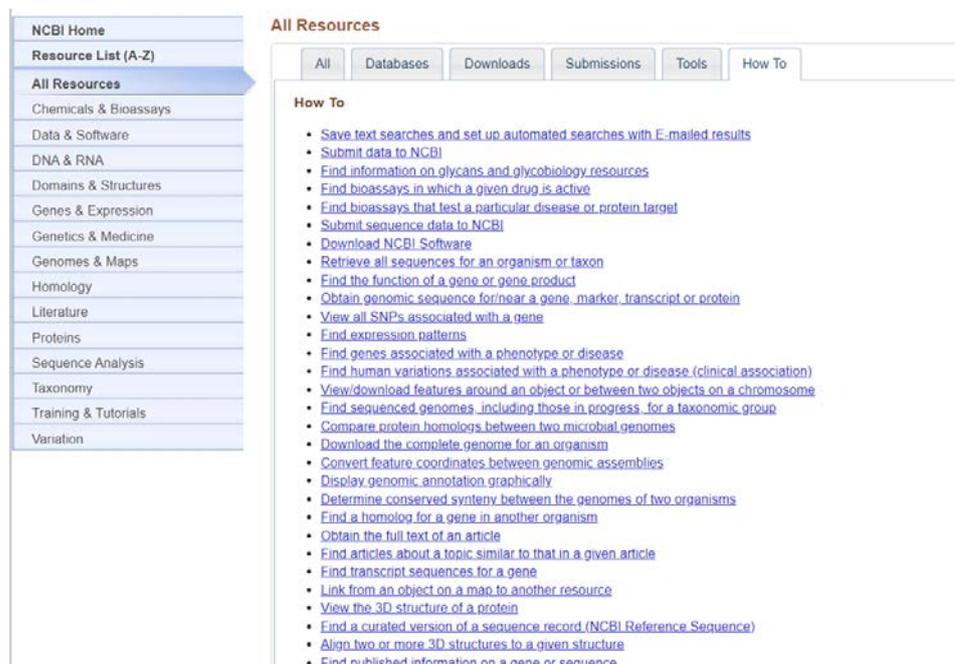
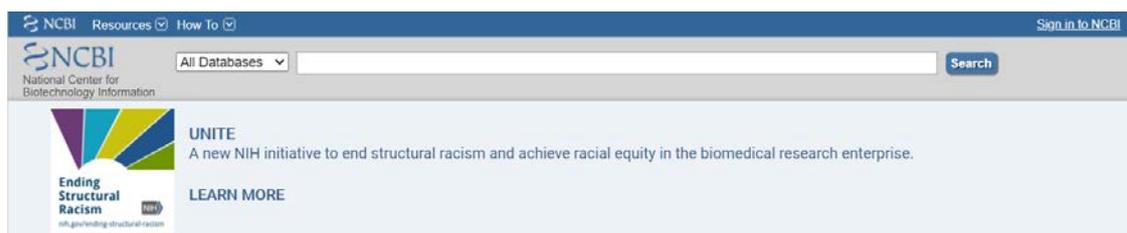
二、实验目标

掌握生信常用数据库的基本功能和查询方法

三、数据库介绍

第一部分 NCBI 数据库简介

NCBI 数据库，全称 National Center for Biotechnology Information (国家生物技术信息中心)，是隶属美国国家卫生研究所的国家医学图书馆(NLM)的分部，拥有包括 PubMed、PubMed Central 和 GenBank 在内的大约 40 个在线文献和分子生物学数据库。数据库资源涵盖基因组、变异、基因和基因表达、蛋白质、小分子和生物信息学分析等关键词，汇集生物分子结构、元件和数据分析工具。以下是 NCBI 查询界面，可登录点击，了解 NCBI 查询方法及功能清单：



第二部分 NCBI 文献检索

文献检索是科研工作绕不开的内容, NCBI 的文献资源包括: NLM Catalog、PubMed、PubMed Central (PMC)、Bookshelf 以及补充资源 MeSH、PubMed DTD、JATS、NIHMS。最常用的是 PubMed

以下是文献检索步骤:

1. 选择 PubMed
2. 在搜索框输入关键词, 以 OCA2 基因为例
3. 通过增加关键词缩小范围

两种常用方式:

- 1) 布尔逻辑检索符: AND (与)、OR (或)、NOT (非)

注意, 逻辑字符要大写

以 OCA2 基因 rs1800414 位点为例进行检索:

- 2) “Advanced” 高级检索, 通过不同的字段来增加检索条件

4. 跟踪进展
可注册 NCBI 账户, 然后点击 Create alert
5. 设置显示的模式
 - 1) 显示是否按时间、相似度等排序, 是否显示摘要
 - 2) 根据文章类型、是否提供免费全文、出版日期等在文章搜索界面左侧边框当中尝试勾选设置不同参数
6. 保存检索结果

第三部分 其他常用数据库

除 NCBI 外, 其他生信数据分析常用综合性数据库还包括以下几个:

1. EMBL [<https://www.ebi.ac.uk/ena>]
欧洲分子生物学实验室 EMBL(The European Molecular Biology Laboratory)
2. DDBJ [<https://www.ddbj.nig.ac.jp/index-e.html>]

DDBJ(DNA Data Bank of Japan), 于 1984 年建立, 是世界三大 DNA 数据库

之一，与 NCBI 的 GenBank，EMBL 的 EBI 数据库共同组成国际 DNA 数据库

3. CNGB [https://db.cngb.org/]

中国国家数据库（China National GeneBank）位于深圳大鹏新区，是继世界三大数据库之后的全球第四大国家级数据库。它是中国首个，也是唯一一个国家基因库，相对于全球另外三个基因库而言，国家基因库样品保存的规模、存储量和可访问的数据量皆是全球最大。

4. BIGD [https://bigd.big.ac.cn/]

中国国家基因组科学数据中心 生命与健康大数据中心 (National Genomics Data Center BIG Data Center)

第四部分 重要的表型数据库——OMIM

OMIM 数据库，全称 Online Mendelian Inheritance in Man（在线人类孟德尔遗传数据库）

OMIM 数据库包括：

1.gene entry 基因条目；2.allelic variations 等位基因变异；3.gene map 基因图谱；4.phenotypic series 表型系列；5.phenotype entry 表型条目；6.clinical synopsis 临床提要；7.external links 外部链接

基因型与表型的关系：

以 OCA2 为例，可按以下路径查询：

<https://www.omim.org/entry/611409?search=OCA2&highlight=oca2>

HGNC Approved Gene Symbol: **OCA2**

Cytogenetic location: **15q12-q13** Genomic coordinates (GRCh38): **15:27,719,007-28,099,341** (from NCBI)

Gene-Phenotype Relationships

Location	Phenotype Clinical Synopses	Phenotype MIM number	Inheritance	Phenotype mapping key
15q12-q13	[Skin/hair/eye pigmentation 1, blond/brown hair]	227220	AR	3
	[Skin/hair/eye pigmentation 1, blue/nonblue eyes]	227220	AR	3
	Albinism, brown oculocutaneous	203200	AR	3
	Albinism, oculocutaneous, type II	203200	AR	3

location 代表相关基因在染色体中的位置；phenotype 代表基因相关的表型；phenotype MIM number 代表表型的 MIM 编号；inheritance 代表遗传，是指该基因的遗传类型，如 AR 是指常染色体隐性遗传，AD 是常染色体显性遗传，SMu 是指体细胞突变，鼠标点击缩写符号就会出现不同缩写代表的具体含义；phenotype mapping key 代表表型映射关键，3 代表该疾病的分子基础是已知的；Gene/Locus 代表对应的基因或位点；Gene/Locus MIM number 代表对应的基因或

基因座 MIM 编号。点 location 还能显示该位置在同一个染色体相邻的基因列表以及引发的各种疾病。

PheneGene Graphics 

TEXT

▼ Description

The **OCA2** gene encodes a protein that corresponds to the 'pink-eyed dilution' (p) mouse mutant. The gene product plays a role in regulating the pH of melanosomes (Yuasa et al., 2007). 

▼ Cloning and Expression

Gardner et al. (1992) isolated mouse cDNA clones from the p locus from murine melanoma and melanocyte libraries. The deduced 833-residue protein has a molecular mass of 92 kD. Gardner et al. (1992) obtained the human counterpart of the murine p cDNA by screening a human melanoma cDNA library with a fragment of mouse genomic DNA. The predicted amino acid sequence of the human gene product showed 84% identity from amino acids 283 to 414 of the predicted mouse protein. 

Rinchik et al. (1993) demonstrated that the human cDNA DN10, linked to the p locus in mice, identifies the human homolog (P) of the mouse p gene, and appears to encode an integral membrane transporter protein. The human P protein is an 838-amino acid polypeptide that contains 12 putative transmembrane domains and exhibits structural homology to transporters of small organic molecules. 

Lee et al. (1995) noted that the 838-residue P protein contains 12 transmembrane domains arranged similarly to various transporters and appears to be an integral membrane protein of melanosomes. Sequence comparisons suggested to Lee et al. (1995) that the P protein is a member of a family of

根据以上文献线索，即可查询该基因与表型之间关系的研究进展。

实验 3.2 遗传特征解读

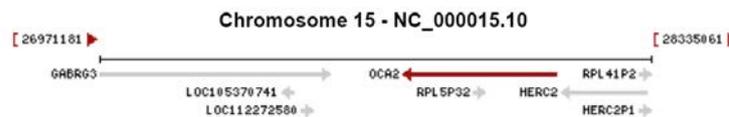
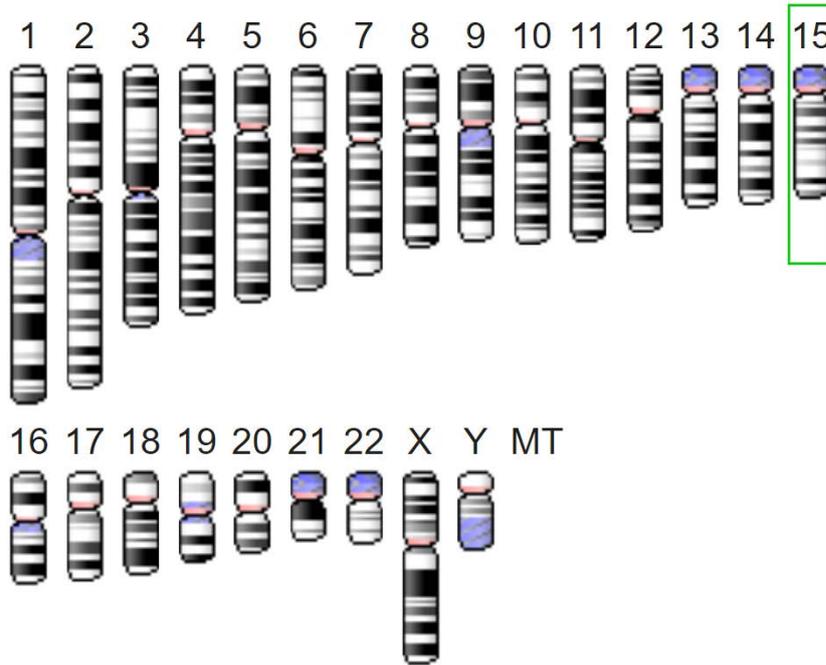
(3 学时)

一、实验目标

本次实验需查阅数据库和文献，对变异检测结果进行解读。

我们以东亚人先天肤色深浅度突变位点 rs1800414 为例，介绍查阅 NCBI 和 OMIM 数据库对位点进行解读的步骤和方法，解读结果涵盖以下要点：

1、该位点在基因组和基因序列中的位置；



位点：The T/C-27951890 (rs1800414) (参考序列版本：GRch38)

2、该位点对基因表达及性状的影响：

TT：黑色素前体跨膜转运蛋白表达正常，酪氨酸酶合成转运正常，黑色素合成顺畅，先天肤色深；

CC：黑色素前体跨膜转运蛋白表达量少，酪氨酸酶合成和转运受阻，黑色素合成少，先天肤色浅；

TC：黑色素前体跨膜转运蛋白表达量以及酪氨酸酶合成转运量一般，黑色素合成居于中等水平，先天肤色介于深色和浅色之间，表现为一般肤色。

二、实验步骤

第一部分 NCBI 数据库检索

检索 rs1800414 位点信息, 获取该位点在基因组和基因区域的具体位置信息：



Search results

Items: 5

rs1800414 [*Homo sapiens*]

1.

Variant type: **SNV**
 Alleles: **T>A,C** [Hide Flanks]

```

AAGATGAATGCCAGGGACAAACGAATTGAGGAAAAACATGAAGATAACAA
ATCCCAACACTGTCAGGCATTTGGCGAGCAGAATCCCCTCAGATATCCTA
[T/A/C]
GCTGTAAGAGAGAAAACCACAGCTCATTACTCTGCACAACCTTCTGACTC
CTGCAGCGTGTATAACCTACGCAACTGGAATCACGACAGATTTCACGAG
  
```

Chromosome: **15:27951891 (GRCh38)**
15:28197037 (GRCh37)

Canonical SPDI: NC_000015.10:27951890:T:A,NC_000015.10:27951890:T:C
 Gene: OCA2 (Varview)
 Functional Consequence: missense_variant,genic_downstream_transcript_variant,coding_sequence_variant
 Clinical significance: benign
 Validated: by frequency,by alfa,by cluster
 MAF: C=0.003892/500 (ALFA)
 C=0.00027/1 (TWINSUK)
 C=0.000446/2 (Estonian)
 ...more

HGVS: NC_000015.10:g.27951891T>A, NC_000015.10:g.27951891T>C,
 NC_000015.9:g.28197037T>A, NC_000015.9:g.28197037T>C,
 NG_009846.1:g.152422A>T, NG_009846.1:g.152422A>G, NM_000275.3:c.1844A>T,
 NM_000275.3:c.1844A>C, NM_000275.3:c.1844A>T, NM_000275.3:c.1844A>C
 ...more

[PubMed](#) [LitVar](#)

打开页面左下角 PubMed 文献库，选择与 rs1800414 位点及其所在基因、表型等相关的文献：

16 results

1 The distinctive geographic patterns of common pigmentation variants at the OCA2 gene.
Cite Kidd KK, Pakstis AJ, Donnelly MP, Bulbul O, Cherni L, Gurkan C, Kang L, Li H, Yun L, Paschou P, Meiklejohn KA, Haigh E, Speed WC.
Share Sci Rep. 2020 Sep 22;10(1):15433. doi: 10.1038/s41598-020-72262-6. PMID: 32963319 [Free PMC article.](#)

2 Distribution of variants in multiple vitamin D-related loci (DHCR7/NADSYN1, GC, CYP2R1, CYP11A1, CYP24A1, VDR, RXR α and RXR γ) vary between European, East-Asian and Sub-Saharan African-ancestry populations.
Cite Jones P, Lucock M, Chaplin G, Jablonski NG, Veysey M, Scarlett C, Beckett E.
Share Genes Nutr. 2020 Mar 13;15(1):5. doi: 10.1186/s12263-020-00663-3. PMID: 32169032 [Free PMC article.](#)

3 Application of partial least squares in exploring the genome selection signatures between populations.
Cite Sun H, Zhang Z, Olasege BS, Xu Z, Zhao Q, Ma P, Wang Q, Pan Y.
Share Heredity (Edinb). 2019 Mar;122(3):288-293. doi: 10.1038/s41437-018-0121-y. Epub 2018 Jul 26. PMID: 30050061 [Free PMC article.](#)

4 Genome-wide association study of pigimentary traits (skin and iris color) in individuals of East Asian ancestry.
Cite Rawofi L, Edwards M, Krithika S, Le P, Cha D, Yang Z, Ma Y, Wang J, Su B, Jin L, Norton HL, Parra EJ.
Share PeerJ. 2017 Nov 2;5:e3951. doi: 10.7717/peerj.3951. eCollection 2017. PMID: 29109912 [Free PMC article.](#)

8 Association study confirms the role of two OCA2 polymorphisms in normal skin pigmentation variation in East Asian populations.
Cite Eaton K, Edwards M, Krithika S, Cook G, Norton H, Parra EJ.

第二部分 文献解读

阅读文献，并形成以下位点解读报告

1、性状背景：人皮肤色素相关背景介绍

人类的肤色是一种高度可遗传的表型特征，不同地理位置、种族的人群肤色是在长期自然选择和人群迁移、混合下形成的，这使得不同人群的肤色表型既有一些共同的遗传基础，又有鲜明的局部遗传差异。亚洲人的典型特征之一是有黄色皮肤，但人群中肤色深浅也有较大差异。



黑色素是决定人类肤色的主要色素，它由酪氨酸衍生而来，分为黑色真黑色

素和黄红色脱黑色素两种不同类型。胚胎发育时期的神经嵴细胞分化成黑色素细胞，进一步发育成为皮肤黑色素细胞，产生类似溶酶体的细胞器——黑素体，负责合成和储存黑色素。黑素体发育成熟后会向黑色素细胞周围的角质细胞中转移。不同个体皮肤角质层黑素体中黑色素的组成、丰度和分布不同，导致不同个体呈现出的肤色深浅各不相同。

2、变异位点影响性状效应的解读要点

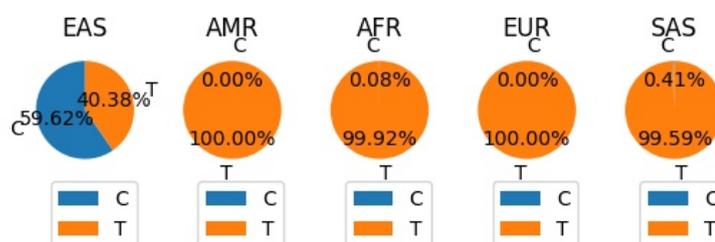
1) OCA2 基因负责编码跨膜转运蛋白，参与黑色素前体的跨膜转运，同时一定程度上通过加工和转运酪氨酸酶，控制黑色素细胞中真黑色素的含量，从而影响肤色深浅。OCA2 基因位于第 15 号染色体长臂的第 13 号区域，简称为 15q13。

2) rs1800414 变异位点位于 15 号染色体上 27951890 号位置，落在 OCA2 基因第 17 号外显子区，为非同义突变，引起组氨酸（密码子 CAT/CAC）变为精氨酸（CGT/CGC），导致转运蛋白或酪氨酸酶合成加工相关的蛋白结构发生改变，黑色素前体转运及合成受阻。

3) rs1800414 位点 T/C 突变频率在欧洲人群中 <0.3%，在中国汉族人群中高于 50%，该变异类型在非洲人群中缺失。（OCA2 基因位于 DNA 负链，与基因组正链碱基型互补）

(<http://www.uniprot.org/>). The geographic distribution of this SNP across world populations (data from HGDP database: <http://hgdp.uchicago.edu>) showed that the derived G allele of rs1800414 is highly prevalent in most East Asian populations (G allele > 50% in Han Chinese), as well as in American Indians, but is in low frequency in A-A populations (G allele = 16.0% (A-A from China); 21.3% (A-A from Cambodia)), totally absent in Africans, and absent or extremely rare in western Europeans (G allele < 0.3%) (fig. 2B). This piece of evidence suggests that rs1800414 is likely an East Asian specific mutation, consistent with the proposed local selection on this SNP and an independent evolution of skin pigmentation in the area (Edwards, et al. 2010).

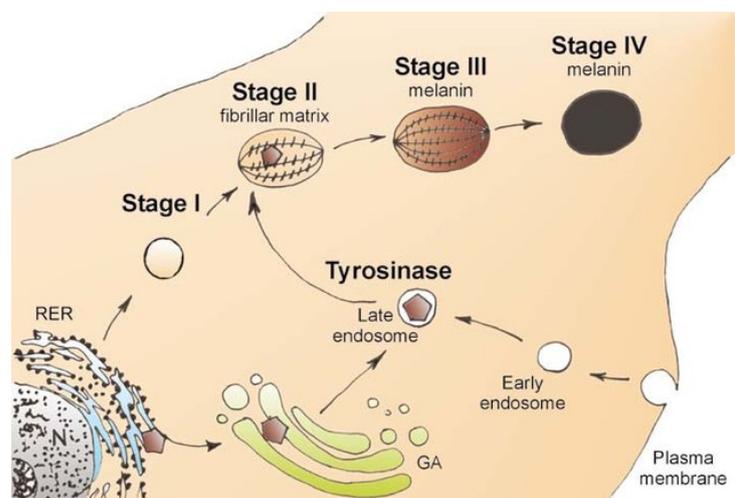
4) 千人基因组计划中 rs1800414 位点在不同人群中检测到的突变频率：



3、变异位点影响 OCA2 基因功能进而影响表型的机制解读

OCA2 基因功能并不单一，全长 344.44kb，包含 32 个不同的内含子，可形成 9 种不同的可变剪切类型，编码至少 8 种不同的蛋白质亚型，分别含不同转运

结构域和跨膜结构域。其中包括一种完整的膜蛋白，有 12 个跨膜结构域，功能类似通道或转运蛋白，在黑素体的生物发生过程中起重要作用，还能一定程度上通过运输和加工酪氨酸酶黑色素合成过程中的限速酶)，控制黑素细胞中真黑色素的含量。动物实验证实，OCA2 基因敲低，一是导致小鼠黑素细胞中的转运蛋白不表达，二是导致小鼠皮肤中的酪氨酸酶以及与之密切相关的两种蛋白 TRP-1 和 TRP-2 表达量显著下降。总之，OCA2 基因从合成和转运两方面影响黑色素的形成。



以上为黑素细胞中黑素体的发育路径：Stage I->Stage II->Stage III->Stage IV，黑素体中的黑色素累积，发育成熟后向黑素细胞周围的皮肤细胞中转移。在黑素体发育过程中，酪氨酸酶催化酪氨酸的形成，为合成黑色素供应原料。

4、延伸探究：人皮肤色素沉着相关基因研究进展概述

当前，已有大量小鼠和其他模式生物相关的研究可用来验证与人类色素沉着相关的基因功能，国际小鼠表型联盟已经鉴定出 34 个基因，这些基因与 224 个色素沉着表型相关。34 个基因主要分为两类，一类调节黑素细胞前体的细胞增殖、迁移和分化，包含 SOX10、PAX3、MITF、EDN3、EDNRB、KIT 和 KITLG 等关键基因；另外一类基因参与黑素细胞中黑色素的合成和转运，包含 TYR、TYR 1、DCT、OCA2、SLC24A5、SLC45A2、MC1R、ASIP、GPR143 和 PMEL 等。这些基因除了会影响肤色，也会影响头发或眼睛的颜色。

5、主要参考文献及数据库

- 1、Yuanqing Feng, Michael A McQuillan, Sarah A Tishkoff, Evolutionary genetics of skin pigmentation in African populations, Human Molecular Genetics, 2021, ddab007, <https://doi.org/10.1093/hmg/ddab007>
- 2、Jannot, AS., Meziani, R., Bertrand, G. et al. Allele variations in the OCA2 gene (pink-eyed-dilution locus) are associated with genetic susceptibility to melanoma. Eur J Hum Genet 13, 913–920 (2005). <https://doi.org/10.1038/sj.ejhg.5201415>
- 3、Edwards M, Bigham A, Tan J, Li S, Gozdzik A, Ross K, Jin L, Parra EJ. Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. PLoS Genet. 2010 Mar

5;6(3):e1000867. doi: 10.1371/journal.pgen.1000867. PMID: 20221248; PMCID: PMC2832666.

4、 <https://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=human&q=OCA2>

5、 Rosemlat S, Durham-Pierre D, Gardner JM, Nakatsu Y, Brilliant MH, Orlow SJ. Identification of a melanosomal membrane protein encoded by the pink-eyed dilution (type II oculocutaneous albinism) gene. Proc Natl Acad Sci U S A. 1994 Dec 6;91(25):12071-75. doi: 10.1073/pnas.91.25.12071. PMID: 7991586; PMCID: PMC45378.

三、预期实验结果

自选位点，参考以上实验步骤形成解读报告。