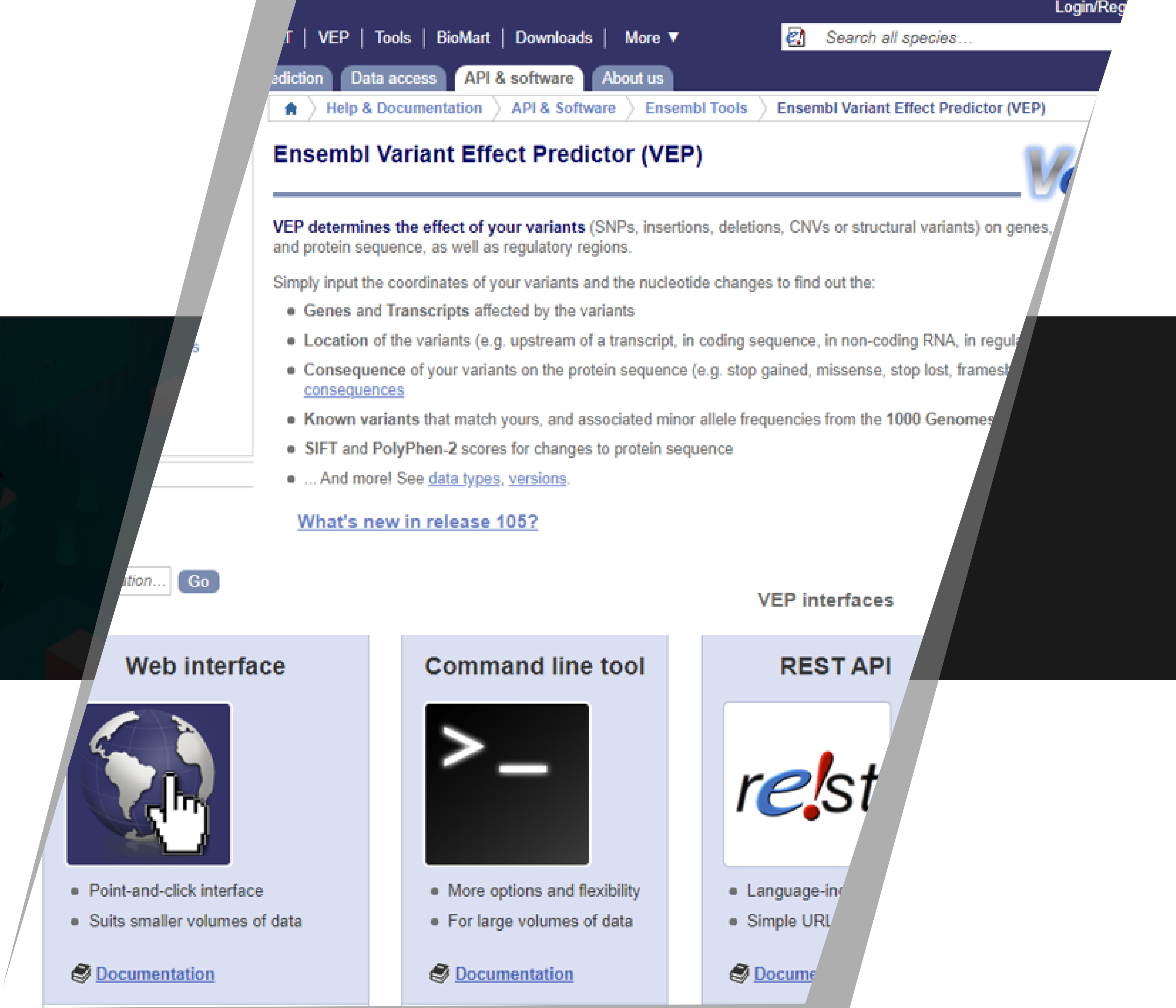
解螺旋

Ensembl数据库使用教程

科研技能 单元课 01

- 针对特定科研技能的专项突破

@小糖球 参与制作



Ensembl Variant Effect Predictor (VEP)

VEP determines the effect of your variants (SNPs, insertions, deletions, CNVs or structural variants) on genes, and protein sequence, as well as regulatory regions.

Simply input the coordinates of your variants and the nucleotide changes to find out the:

- **Genes** and **Transcripts** affected by the variants
- **Location** of the variants (e.g. upstream of a transcript, in coding sequence, in non-coding RNA, in regul
- **Consequence** of your variants on the protein sequence (e.g. stop gained, missense, stop lost, framesh consequences
- **Known variants** that match yours, and associated minor allele frequencies from the **1000 Genomes**
- **SIFT** and **PolyPhen-2** scores for changes to protein sequence
- ... And more! See data types, versions.

What's new in release 105?

VEP interfaces

Web interface

- Point-and-click interface
- Suits smaller volumes of data

Documentation

Command line tool

- More options and flexibility
- For large volumes of data

Documentation

REST API

- Language-in
- Simple URL

Docume

课程目录

1. 数据库介绍
   - 1.1 数据库介绍
   - 1.2 数据库的网站结构
   - 1.3 Ensembl/NCBI Map Viewer/UCSC区别
   - 1.4 Ensembl数据来源、注释及注释文件
   - 1.5 Ensembl主要功能
2. 检索
3. 数据获取
4. Biomart
5. 版本转换

Ensembl数据库简介

■ 基本情况介绍

➢ 简介

Ensembl 是一个由英国 Sanger 研究所 Wellcome 基金会（ WTSI）和欧洲分子生物学实验室所属分部欧洲生物信息学研究所（ EMBI-EBI）共同协作运营的一个项目，其目标在于自动的基因组注释，并把这些注释与其他有用的生物数据整合起来，通过网络公开给所有人使用。

➢ 网址

https://www.ensembl.org/index.html

➢ 发展历史

- Ensembl 数据库网站开始于 July 2000， 是一个真核生物基因组注释项目，其侧重于脊椎动物的基因组数据，但也包含了其他生物，如线虫，酵母，拟南芥和水稻等。
- 2009年，Ensembl基因组项目启动，为植物、真菌、无脊椎动物后生动物、细菌和原生生物基因组提供了特定的门户网站。这些目的是提供分类学参考点，给出可以理解基因的进化背景，以及所有主要非脊椎动物实验生物、重要农业物种、病原体和载体的覆盖范围。
- 2020年，Ensembl在Ensembl和Ensembl基因组网站上支持了超过500000个基因组，包括快速发布（可快速访问新注释的基因组）和新冠病毒-19（可访问SARS-CoV-2基因组）。
- 近年来， 随着时间推移， 越来越多的基因组数据已经被添加到了 Ensembl，同时 Ensembl 可用数据的范围也扩展到了比较基因组学、变异，以及调控数据。

Ensembl数据库简介

■ 网站站点及查询内容

- 主站点

  脊椎动物 GRCh37.p13：https://www.ensembl.org/info/website/tutorials/grch37.html

  GRCh38.p13：https://www.ensembl.org/Homo_sapiens/Info/Index?db

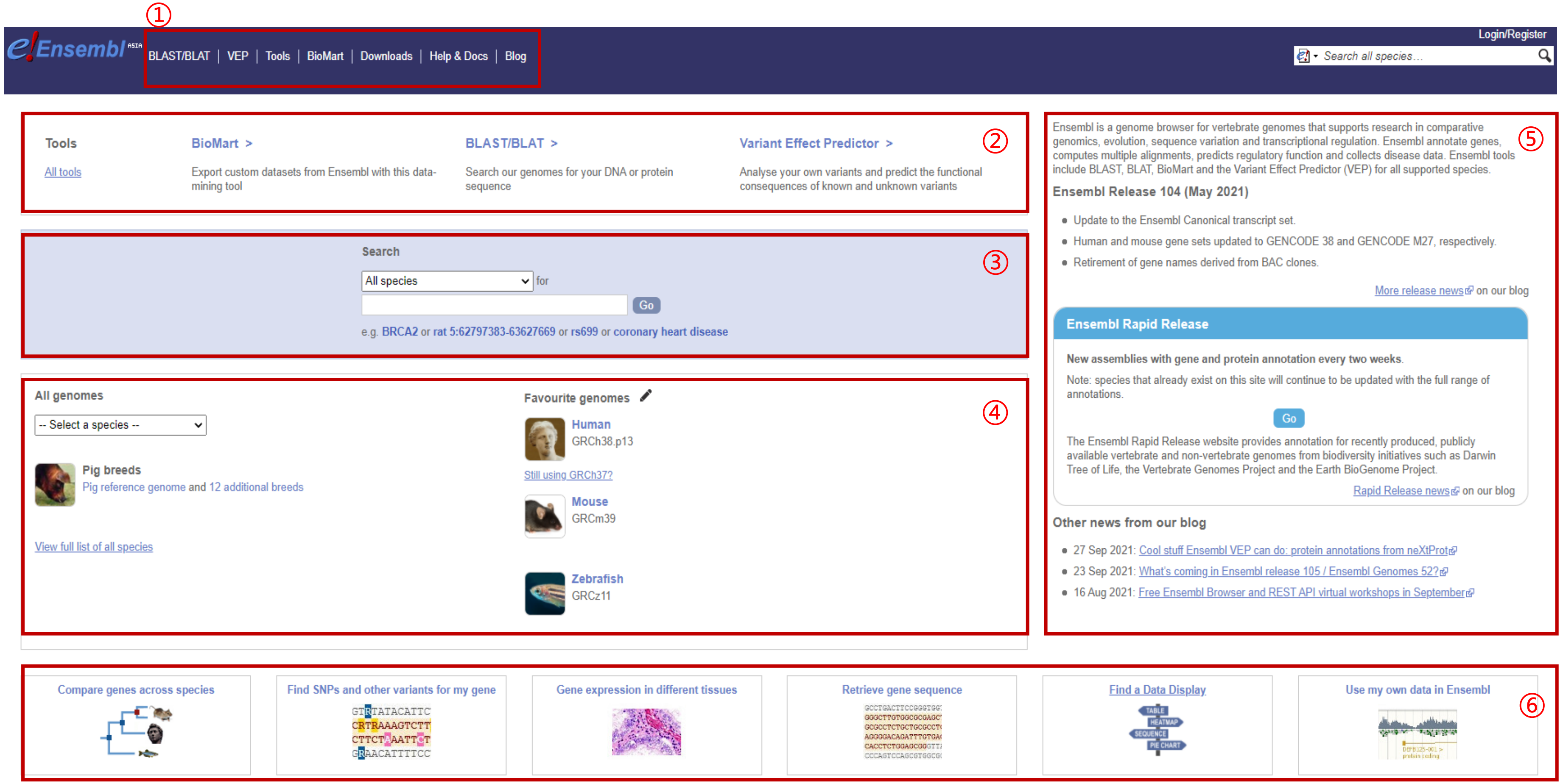
- 其余站点
  - COVID-19：https://covid-19.ensembl.org/index.html
  - 快速检索界面：https://rapid.ensembl.org/index.html
  - 原生生物：https://protists.ensembl.org/index.html
  - 植物：https://plants.ensembl.org/index.html
  - 真菌：https://fungi.ensembl.org/index.html
  - 非脊椎动物（昆虫类、线虫类）：https://metazoa.ensembl.org/index.html
  - 细菌：https://bacteria.ensembl.org/index.html
- 查询内容
  - 可以查找物种、DNA、转录物、蛋白质搜索相关条目的入口，网站分为5大标签，分别为genome、chromosome region、gene、transcirpt、 variation。
  - 可以通过关键词查询，也可以用BLAST进行相似序列的搜索，通过BioMart导出序列和基因信息
  - 也可以显示各种染色体，可以在染色体水平上选择感兴趣的位点，逐层放大浏览整个基因组。
  - 查看不同物种的SNP
  - 查看比对到Ensembl基因上的mRNA或蛋白的序列位置
  - 上传自己数据

■ 主要功能

- 对真核生物的基因组进行自动注释
  - 对基因组的转录调控区域的自动注释
  - 对比较基因组学，也就是跨物种尺度的序列的多态性保守度等等进行的一个注释。
  - 对基因组中的多态性位点进行了定位。
  - 对某基因进行功能注释
  - 对某基因突变进行结构注释
  - 某基因在组织中表达情况
- BLAST和BLAT主要用于序列的短序列比对
- Biomart 主要是对数据进行下载
- VEP 主要用来对序列其中的多态性位点进行一个功能性的注释

Ensembl数据库主页面

在浏览器打开Ensembl数据库的网址，可以跳转至以下主页面。



Tools
All tools

BioMart >
Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >
Search our genomes for your DNA or protein sequence

Variant Effect Predictor >
Analyse your own variants and predict the functional consequences of known and unknown variants

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes
-- Select a species --

Pig breeds
Pig reference genome and 12 additional breeds

View full list of all species

Favourite genomes

Human
GRCh38.p13

Still using GRCh37?

Mouse
GRCm39

Zebrafish
GRCz11

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 104 (May 2021)

- Update to the Ensembl Canonical transcript set.
- Human and mouse gene sets updated to GENCODE 38 and GENCODE M27, respectively.
- Retirement of gene names derived from BAC clones.

More release news on our blog

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.

Note: species that already exist on this site will continue to be updated with the full range of annotations.

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

Rapid Release news on our blog

Other news from our blog

- 27 Sep 2021: Cool stuff Ensembl VEP can do: protein annotations from neXtProt
- 23 Sep 2021: What's coming in Ensembl release 105 / Ensembl Genomes 52?
- 16 Aug 2021: Free Ensembl Browser and REST API virtual workshops in September

① 导航栏
- BLAST: 短序列对比
- Tools: 在线工具
- Downloads：下载数据
- VEP:对序列其中的多态性位点进行功能性注释
- Biomart: 下载数据
- Help: 帮助文档
- Blog: 博客

② 工具栏
- Tools: 所有在线工具
- Biomart: 下载数据
- BLAST: 短序列对比
- Variant Effect Predictor:突变效应预测

③ 搜索栏

④ 物种选择栏
- 选择物种
- 点击浏览所有物种
- 可以编辑自己常用的物种快捷链接

⑤ 新闻通知栏

⑥ 其他数据库链接
- 比较基因组学
- 基因组中的多态性位点定位
- 基因在不同组织中的表达
- 查找基因序列
- 可视化查找数据
- 自己数据在线分析

Ensembl数据库使用教程

https://www.ensembl.org/index.html

科研技能 单元课

www.helixlife.cn/

Ensembl数据库主站点

在浏览器打开Ensembl数据库的主站点，或在主界面的Favorite Genome中点击human，即可跳转以下界面。

Human (GRCh38.p13) ▾

- 基因组版本
- 检索框

Search Human (Homo sapiens)

Search all categories ▾ | Search… | Go

e.g. BRCA2 or 17:63992802-64038237 or rs699 or osteoarthritis

- 检索类别及检索框

Genome assembly: GRCh38.p13 (GCA_000001405.28) ➢ 基因组版本

- More information and statistics • 此版本基因组统计数据
- Download DNA sequence (FASTA) • 可下载的DNA序列（FASTA格式）
- Convert your data to GRCh38 coordinates • 数据版本转化
- Display your data in Ensembl • 个人数据上传入口

Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart ▾ Go • 转换版本

- 核型 — View karyotype
- Example region
- 变异
- 基因树 — Example gene tree

Gene annotation ➢ 基因注释，包括编码及非编码序列、剪切体、cDNA、蛋白序列、非编码RNAs

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

- More about this genebuild • 基因注释所包含的信息
- Download FASTA files for genes, cDNAs, ncRNA, proteins • 下载FASTA格式基因注释信息
- Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins • 下载GTF、GFF3格式基因注释信息
- Update your old Ensembl IDs • 更新

- 基因 — Example gene
- Example transcript
- 转录本
- 变异

Comparative genomics ➢ 比较基因组学——同系物、基因树、多物种全基因相比对

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

- More about comparative analysis • 比较基因组学所包含的内容
- Download alignments (EMF) • 可下载信息

Variation ➢ 变异——变异、疾病、表型

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes

- More about variation in Ensembl • 变异包含的所有信息
- Download all variants (GVF) • 下载变异信息
- Variant Effect Predictor • 变异效应预测

- Example variant
- Example phenotype • 表型
- Example structural variant • 结构变异

- 调控特征

Regulation ➢ 参与调控信息——包含DNA甲基化、转录因子结合位置、组蛋白修饰、增强子、微阵列注释等

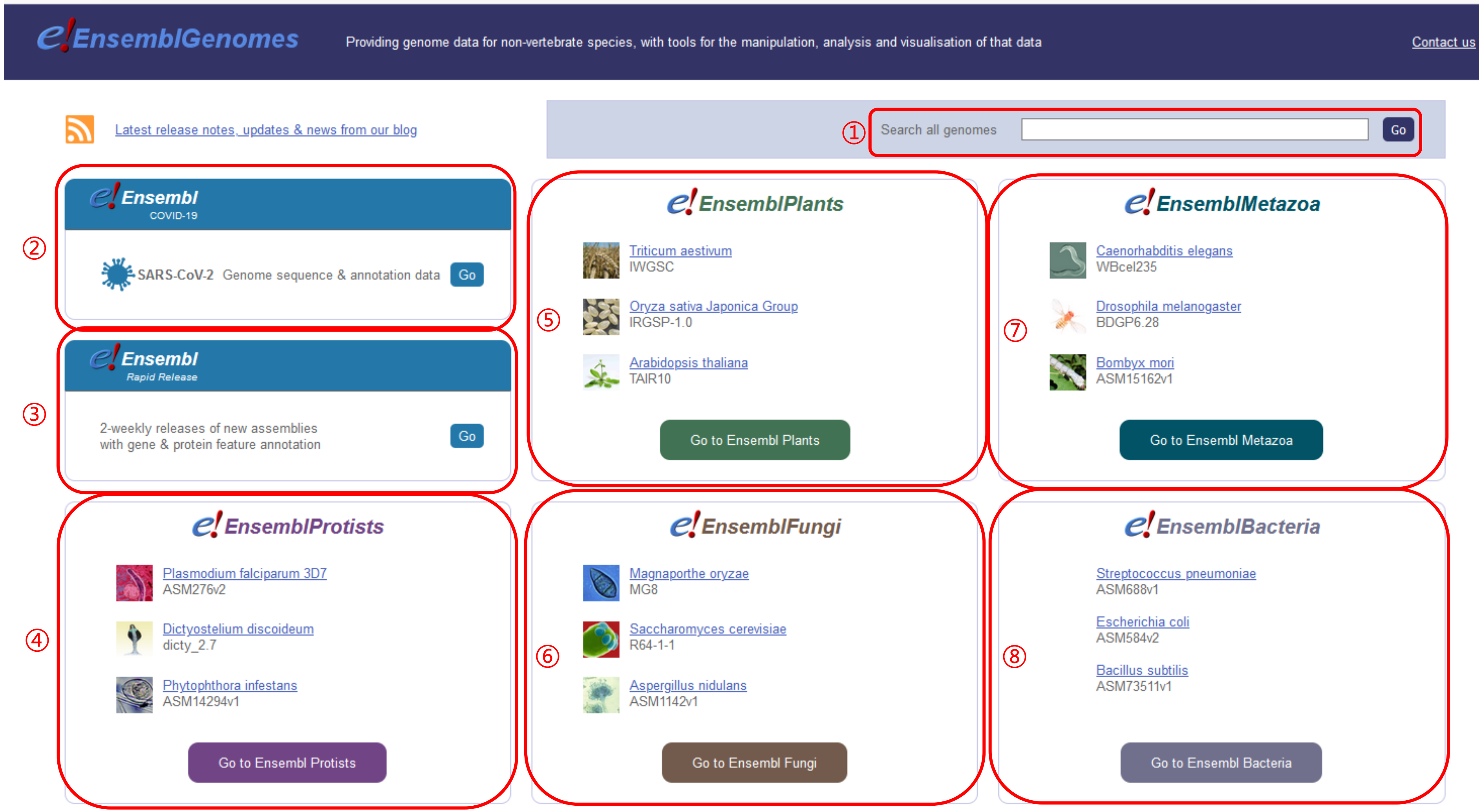
What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and microarray annotations.

- More about the Ensembl regulatory build and microarray annotation • 参与调控信息所包含的信息
- Experimental data sources • 实验数据来源
- Download all regulatory features (GFF) • 所有调控信息下载（GFF格式）

- Example regulatory feature
- ENCODE data in Ensembl
- 编码数据

Ensembl数据库其余网站站点

点击链接http://ensemblgenomes.org/链接到Ensembl数据库姊妹网站，即可跳转以下界面或在主界面最下边our sister site点击链接进入。



EnsemblGenomes Providing genome data for non-vertebrate species, with tools for the manipulation, analysis and visualisation of that data Contact us

Latest release notes, updates & news from our blog

① Search all genomes Go

② Ensembl COVID-19
SARS-CoV-2 Genome sequence & annotation data Go

③ Ensembl Rapid Release
2-weekly releases of new assemblies with gene & protein feature annotation Go

⑤ EnsemblPlants
Triticum aestivum IWGSC
Oryza sativa Japonica Group IRGSP-1.0
Arabidopsis thaliana TAIR10
Go to Ensembl Plants

⑦ EnsemblMetazoa
Caenorhabditis elegans WBcel235
Drosophila melanogaster BDGP6.28
Bombyx mori ASM15162v1
Go to Ensembl Metazoa

④ EnsemblProtists
Plasmodium falciparum 3D7 ASM276v2
Dictyostelium discoideum dicty_2.7
Phytophthora infestans ASM14294v1
Go to Ensembl Protists

⑥ EnsemblFungi
Magnaporthe oryzae MG8
Saccharomyces cerevisiae R64-1-1
Aspergillus nidulans ASM1142v1
Go to Ensembl Fungi

⑧ EnsemblBacteria
Streptococcus pneumoniae ASM688v1
Escherichia coli ASM584v2
Bacillus subtilis ASM73511v1
Go to Ensembl Bacteria

①在所有基因组中检索
②COVID-19基因组
③快捷入口，数据2周更新
④原生生物基因组
⑤植物基因组
⑥真菌基因组
⑦非脊椎动物基因组
⑧细菌基因组

Ensembl & NCBI Map Viewer & UCSC

- Ensembl、NCBI Map Viewer、UCSC是最为常用基因组检索数据库。

- **Map viewer:** 是NCBI 网站上提供的一个非常有用的寻找基因的工具，通过Map viewer你可以了解你感兴趣基因在基因组中所处的位置、基因序列、内含子及外显子的排列、基因的细胞遗传学图、EST.SNP等等许多有用的信息。点击NCBI首页的Map viewer，即可进入Map viewer。

- **UCSC**:是由University of California Santa Cruz（UCSC）基因组学研究所的跨部门团队Genome Bioinformatics Group开发和维护。该网站已发展为包括大量的脊椎动物和模型生物的装配体和注释信息，以及用于查看、分析和下载数据的大量工具。

- Ensembl 与NCBI Map Viewer和UCSC最大区别表现在以下5点:
  - Ensembl的基因数据集是依据mRNA和蛋白序列的数据信息自动注释的。数据来源为新的基因组数据，UniProt/SwissProt和UniProt/TrEMBL的蛋白序列，NCBI的RefSeq里的DNA和蛋白序列和EMBL的cDNA序列。
  - Ensembl是一个开源（Perl API ）的全自动的基因注释软件系统，很多网站都采用Ensembl这套软件系统。
  - Ensembl拥有其特有的BioMart功能。BioMart可以依据设定的要求对基 因组进行条件性检索，检索的结果可以以图表的形式给出。
  - 与其它数据库相整合，比如DAS。
  - 基因组间的比较分析。

Ensembl数据来源与注释

■ 数据来源

Ensembl是常用真核生物参考基因组来源之一，对通用基因的注释有两种：1）Ensembl GeneBuild（自动化注释），速度快、实时更新，在不同物种上均适用；2）Wellcome基金会的 Havana (VEGA)小组的注释（手工注释），速度慢但是准确，它依据的都是已经验证过的mRNA和蛋白序列来注释，比较费时。

➢ Havana (VEGA)小组的注释有以下6种：

- Protein coding： 包括开放阅读框 (ORF)
- Processed transcript：没有开放阅读框（ORF）
- Pseudogene：假基因，是指脱氧核糖核酸（DNA）的碱基序列中，一段与其他生物体内已知的基因序列非常相似的片段，但因移码突变或者无义突变破坏了ORF，无法编码蛋白质，发挥原有的基因功能
- IG gene：免疫球蛋白家族基因
- TR Gene：T细胞受体基因
- TEC (To be Experimentally Confirmed)：实验证实

■ 注释文件

可通过Ensembl ftp进行下载，注释集通常包括了许多的特征信息，如protein coding genes、ncRNA、repeat features、 mRNA可变剪接等，因此下载之后要根据项目需求过滤出自己需要的特征信息。

➢ 一个物种在Ensembl ftp上有多个注释文件，文件包括gff3和gtf两种格式（两者的差别和转换，请参考此链接内容http://blog.nextgenetics.net/?e=27 ），以gtf为例。

- *.gtf结尾的文件，全部的注释结果，一般会使用这个文件。
- *.chr.gtf结尾的文件，包括了所有染色体序列上的注释结果，也包括线粒体染色体的注释信息，但不包括没有被组装成染色体的序列上的注释信息。每条染色体也有单独的注释文件。
- *.abinitio.gtf结尾的文件，为用Genescan和abinitio基因预测工具生成的注释信息，一般是denovo注释的结果。

➢ 人类和小鼠基因组的GTF文件与GENCODE计划发布的gene set文件相同。

➢ The GENCODE project 的目标为对人类和小鼠基因组提供高质量的注释信息和实验确证。

➢ The GENCODE gene sets被其他项目作为参考而广泛使用（如 1000 Genomes）。

课程目录

1 数据库介绍

2 检索

- 2.1 按关键词搜索
- 2.2 按物种检索
- 2.3 按疾病名检索
- 2.4 BLAST/BLAT序列相似性查找

3 数据获取

4 Biomart

5 版本转换

Ensembl检索方法---关键词检索

1. 在浏览器打开Ensembl数据库的主页面，在搜索框中输入关键词，且在All species下拉框中选择物种，点击GO，即可进行相关检索。这里我们以BRCA2为例进行实际操作。

Export custom datasets from Ensembl with this data-mining tool

Search our genomes for your DNA or protein sequence

Analyse your own variants and predict the functional consequences of known and unknown variants

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

- 检索关键词示例，基因名、染色体位置、疾病名称

BRCA2

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

Ensembl检索方法---关键词检索

❷ 信息检索的页面如下图所示。

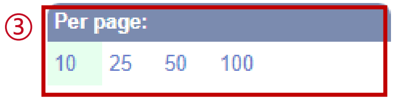
①

Restrict category to:			
Gene	•	基因	1429
Transcript	•	转录本	3655
Variant	•	突变	34
Phenotype	•	表型	7
GeneTree	•	基因树	135
Marker	•	分子标志	1
ProbeFeature	•	探针特征	9
Clones & Regions	•	克隆区域	1
Help & Docs	•	帮助文档	12
Protein Domain	•	蛋白结构域	[illegible]

②

Restrict species to:			
Human	•	人	108
Mouse	•	鼠	975
Zebrafish	•	斑马鱼	24
Abingdon island giant tortoise	•	黑陆龟	14
African green monkey	•	非洲绿猴	[illegible]
African ostrich	•	非洲鸵鸟	[illegible]

... 266 more species ...



③

Per page:

10 25 50 100

BRCA2

7198 results match BRCA2

Ensembl号

④

BRCA2 (Human Gene)
ENSG00000139618 13:32315086-32400268:1
BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]
LRG_293 (LRG display in Ensembl gene record; description: Locus Reference Genomic record for BRCA2,) is an external reference matched to Gene ENSG00000139618
Variant table • Phenotypes • Location • External Refs. • Regulation • Orthologues • Gene tree

• 本词条可链接的条目

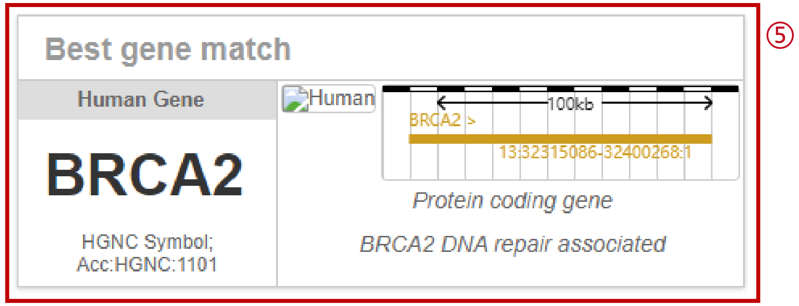
BRCA2-207 (Human Transcript) • 类别
ENST00000614259 13:32316461-32400268:1 • 染色体位置
BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101].
Location • External Refs. • cDNA seq. • Exons • Variant table • Protein seq. • Population • Protein summary

BRCA2-210 (Human Transcript)
ENST00000680887 13:32316072-32400268:1
BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101].
Location • External Refs. • cDNA seq. • Exons • Variant table • Protein seq. • Population • Protein summary

BRCA2-209 (Human Transcript)
ENST00000666593 13:32380007-32394933:1
BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101].
Location • External Refs. • cDNA seq. • Exons • Variant table • Protein seq. • Population • Protein summary

BRCA2-208 (Human Transcript)
ENST00000665585 13:32356428-32398233:1
BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101].
Location • External Refs. • cDNA seq. • Exons • Variant table • Protein seq. • Population • Protein summary

BRCA2-201 (Human Transcript)
ENST00000380152 13:32315508-32400268:1



⑤

Best gene match

Human Gene

BRCA2

HGNC Symbol;
Acc:HGNC:1101

Protein coding gene

BRCA2 DNA repair associated

①结果分类

②物种类别入口

③调整每页显示结果数

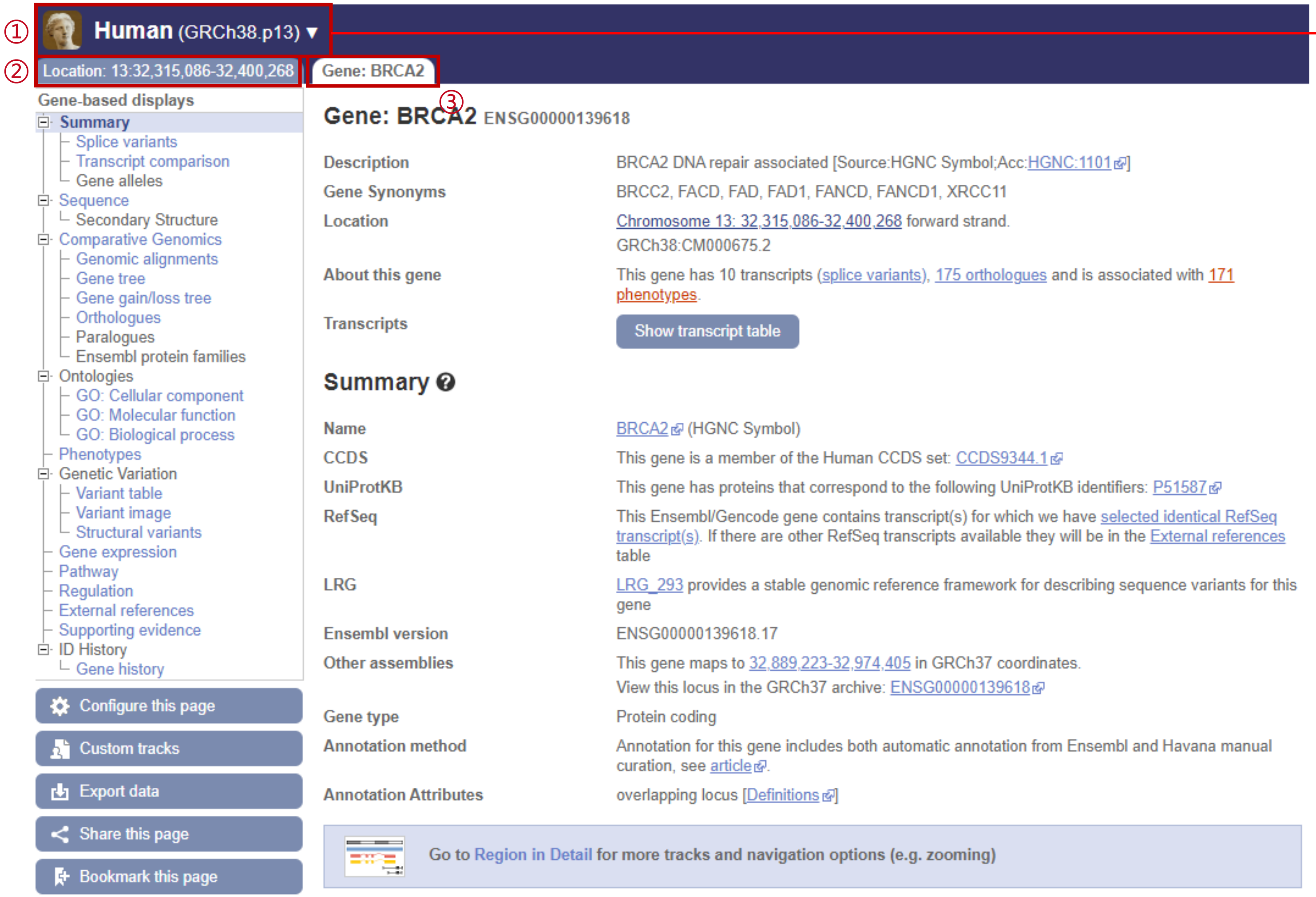
④检索结果列表

⑤最佳匹配项链接

接下来我们以第一条人BRCA2基因为例进一步讲解，点击BRCA2(Human Gene)，可跳转至新界面。

Ensembl检索方法---关键词检索

❸ 点击第一个链接后，结果界面如下图所示。



① **Human** (GRCh38.p13) ▾

② Location: 13:32,315,086-32,400,268

Gene: BRCA2 ③

Gene-based displays
- Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Cellular component
  - GO: Molecular function
  - GO: Biological process
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history

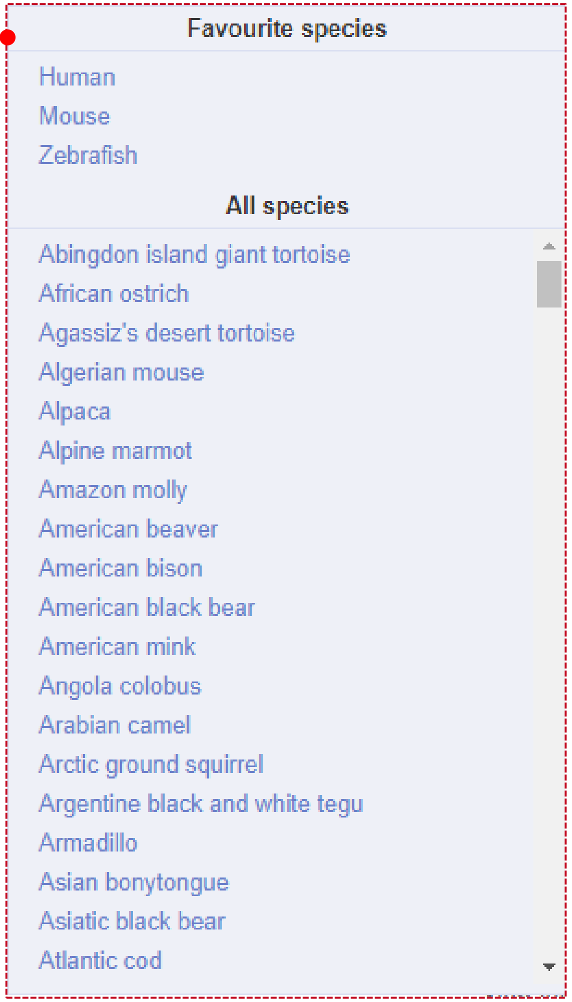
Gene: BRCA2 ENSG00000139618

Description	BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]
Gene Synonyms	BRCC2, FACD, FAD, FAD1, FANCD, FANCD1, XRCC11
Location	Chromosome 13: 32,315,086-32,400,268 forward strand. GRCh38:CM000675.2
About this gene	This gene has 10 transcripts (splice variants), 175 orthologues and is associated with 171 phenotypes.
Transcripts	Show transcript table

Summary

Name	BRCA2 (HGNC Symbol)
CCDS	This gene is a member of the Human CCDS set: CCDS9344.1
UniProtKB	This gene has proteins that correspond to the following UniProtKB identifiers: P51587
RefSeq	This Ensembl/Gencode gene contains transcript(s) for which we have selected identical RefSeq transcript(s). If there are other RefSeq transcripts available they will be in the External references table
LRG	LRG_293 provides a stable genomic reference framework for describing sequence variants for this gene
Ensembl version	ENSG00000139618.17
Other assemblies	This gene maps to 32,889,223-32,974,405 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000139618
Gene type	Protein coding
Annotation method	Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.
Annotation Attributes	overlapping locus [Definitions]

Go to Region in Detail for more tracks and navigation options (e.g. zooming)



Favourite species

Human
Mouse
Zebrafish

All species

Abingdon island giant tortoise
African ostrich
Agassiz's desert tortoise
Algerian mouse
Alpaca
Alpine marmot
Amazon molly
American beaver
American bison
American black bear
American mink
Angola colobus
Arabian camel
Arctic ground squirrel
Argentine black and white tegu
Armadillo
Asian bonytongue
Asiatic black bear
Atlantic cod

①点击可切换物种

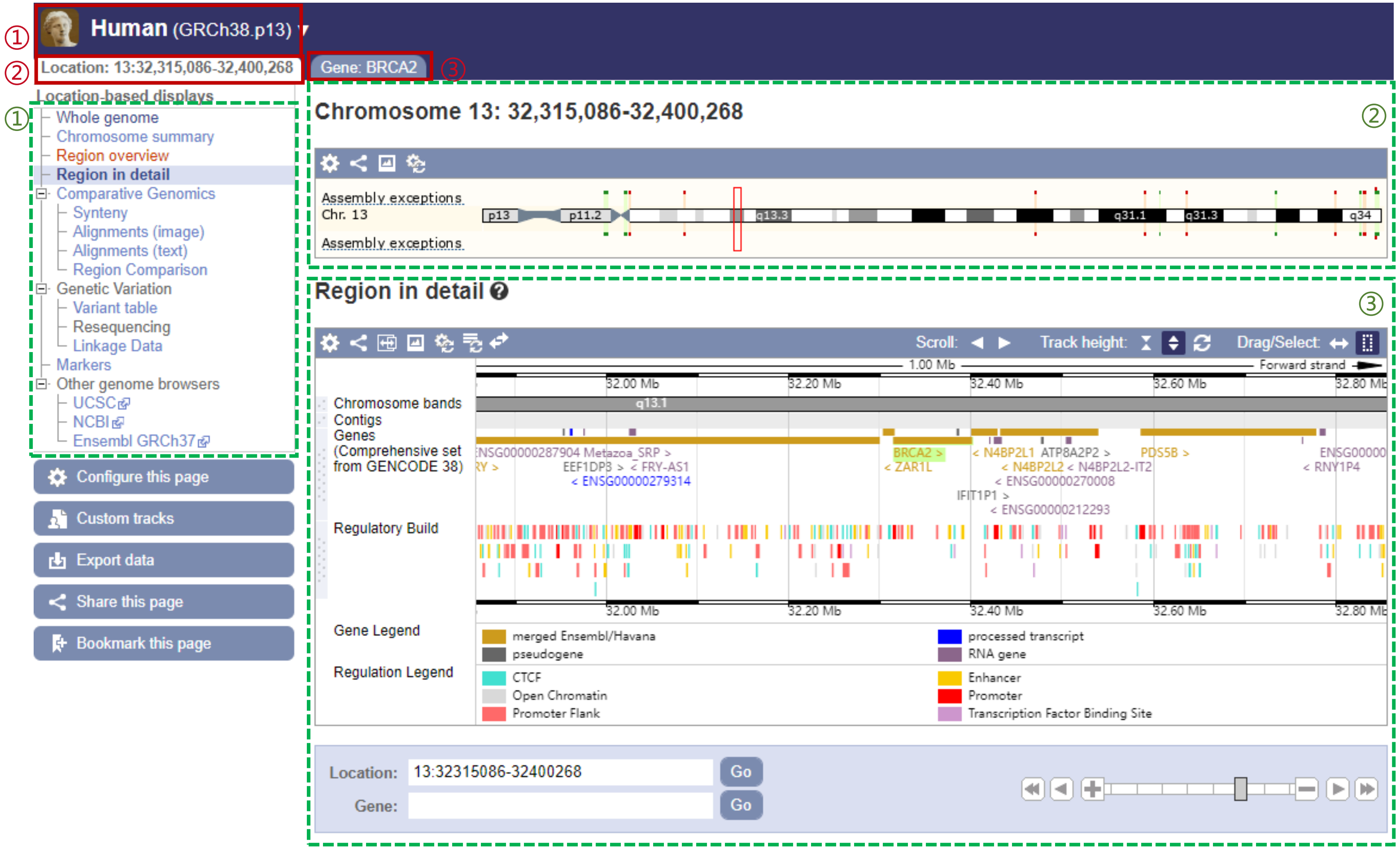
②点击显示染色体定位信息

③点击查看基因详情

后面的PPT将对针对②和③分别进行详细讲解。

Ensembl检索方法---关键词检索

4 在结果界面，点击②进入BRCA2的染色体定位信息，其界面如下面所示



在BRCA2的染色体定位信息界面，可以看到以下信息——

①左侧导航栏

- 全基因组信息（Whole genome）
- 染色体概况（Chromosome summary）
- 区域概览（Region overview）
- 区域详情（Region in detail）
- 比较基因组学（Comparative Genomics）
- 遗传变异（Genetic Variation）
- 标志物（Markers）
- 其他数据库链接（Other genome browsers）

②BRCA2所定位染色体的详细信息

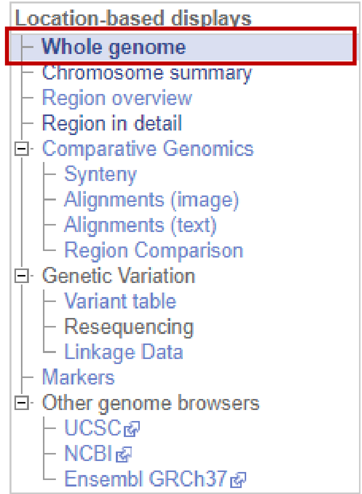
③内容展示区

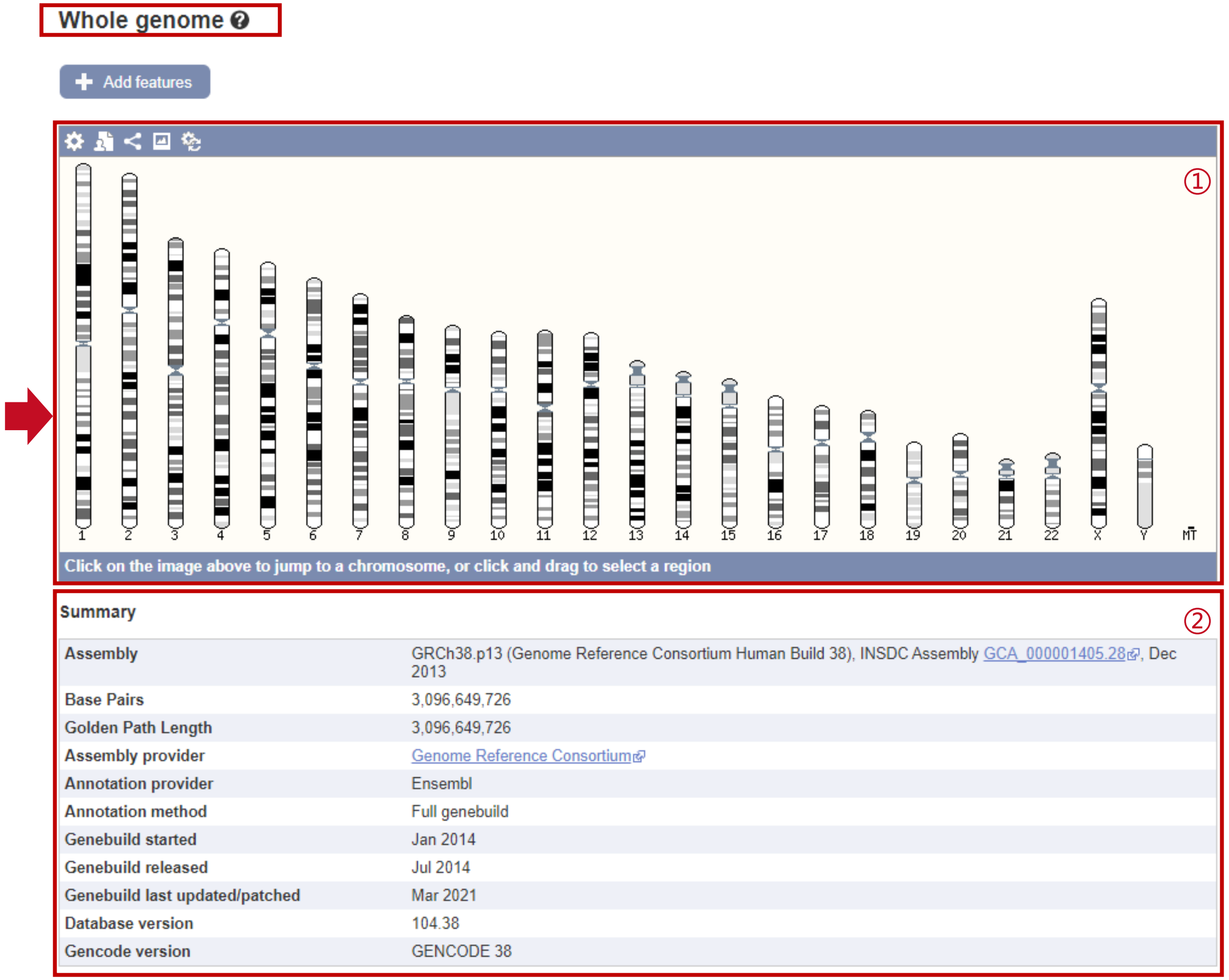
- 点击左侧导航栏相应链接，右侧内容亦会出现变化。

后续PPT将对导航栏各个按钮所展示的内容进行详细介绍

Ensembl检索方法---关键词检索

- BRCA2的染色体定位信息结果界面
  - 在左侧导航栏中点击Whole genome，其右侧内容展示区将如下图所示。





Assembly	GRCh38.p13 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.28, Dec 2013
Base Pairs	3,096,649,726
Golden Path Length	3,096,649,726
Assembly provider	Genome Reference Consortium
Annotation provider	Ensembl
Annotation method	Full genebuild
Genebuild started	Jan 2014
Genebuild released	Jul 2014
Genebuild last updated/patched	Mar 2021
Database version	104.38
Gencode version	GENCODE 38

①染色体信息图谱

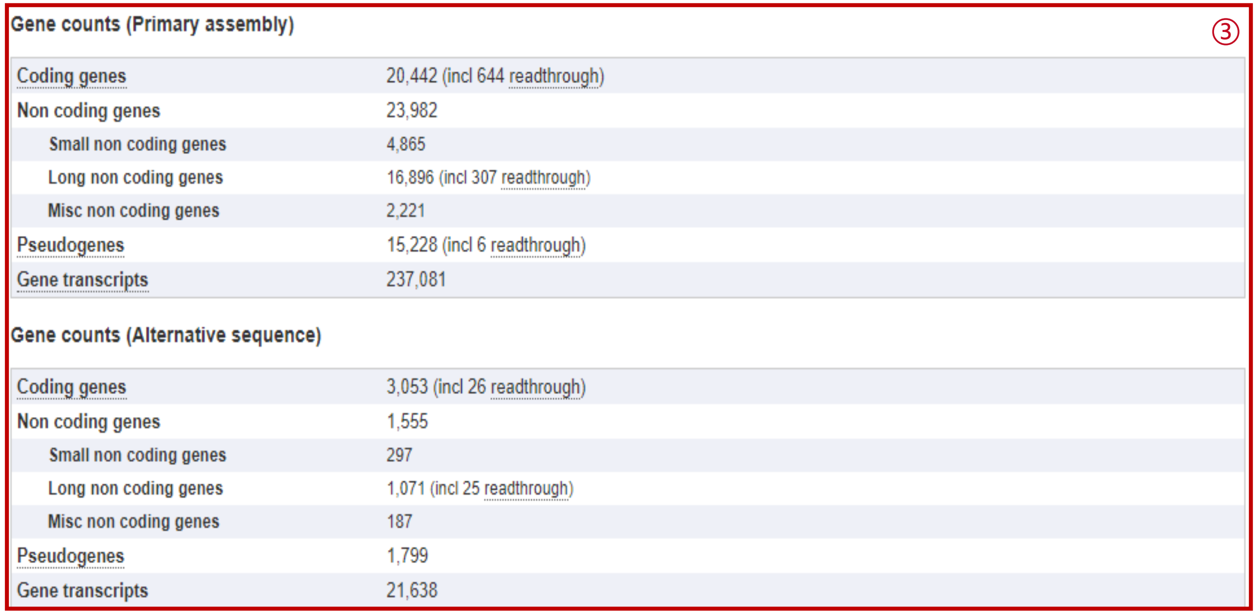
点击染色体号可进一步查看某个染色体信息，也可以选择某区域进一步查看详细信息

②染色体概况

- 版本信息（Assembly）
- 碱基数（Base Pairs）
- 基因组全长（Golden Path Length）
- 版本来源（Assembly provider）
- 注释来源（Annotation provider）
- 注释方法（ Annotation method）
- 基因集创建时间（Genebuild started）
- 基因集发布时间（Genebuild released）
- 基因集最新版本更新时间（Genebuild last updated/patched)
- 数据库版本(Database version)
- 基因组版本(Gencode version)

Ensembl检索方法---关键词检索

- BRCA2的染色体定位信息结果界面
  - 在左侧导航栏中点击Whole genome，其右侧内容展示区将如下图所示——接上页PPT。



③

Gene counts (Primary assembly)

Coding genes	20,442 (incl 644 readthrough)
Non coding genes	23,982
Small non coding genes	4,865
Long non coding genes	16,896 (incl 307 readthrough)
Misc non coding genes	2,221
Pseudogenes	15,228 (incl 6 readthrough)
Gene transcripts	237,081

Gene counts (Alternative sequence)

Coding genes	3,053 (incl 26 readthrough)
Non coding genes	1,555
Small non coding genes	297
Long non coding genes	1,071 (incl 25 readthrough)
Misc non coding genes	187
Pseudogenes	1,799
Gene transcripts	21,638

④

Other

Genscan gene predictions	51,756
Short Variants	714,562,852
Structural variants	6,768,792

③基因数量（Primary assembly & Alternative sequence）

- 编码基因数（Coding genes）
- 非编码基因数（ Non Coding genes ）
- 假基因（Pseudogenes）
- 基因转录本（Gene transcripts）

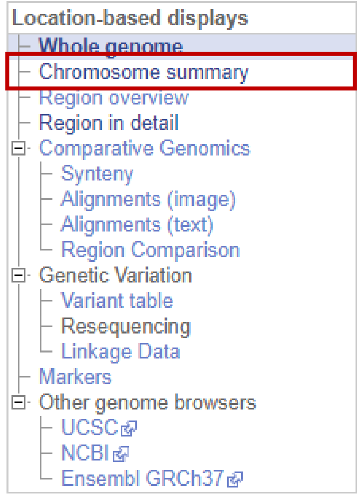
④其它信息

- Genscan预测基因（Genscan gene predictions）
- 短变异体（Short Variants）
- 结构变异（Structural variants）

Ensembl检索方法---关键词检索

■ BRCA2的染色体定位信息结果界面

➢ 在左侧导航栏中点击Chromosome summary，其右侧内容展示区将如下图所示



Chromosome Statistics

Length (bps)	114,364,328
Coding genes	322 (incl. 7 readthrough)
Non coding genes	656
Small non coding genes	98
Long non coding genes	483 (incl. 8 readthrough)
Misc non coding genes	75
Pseudogenes	395
Short Variants	23,960,588

①染色体概况

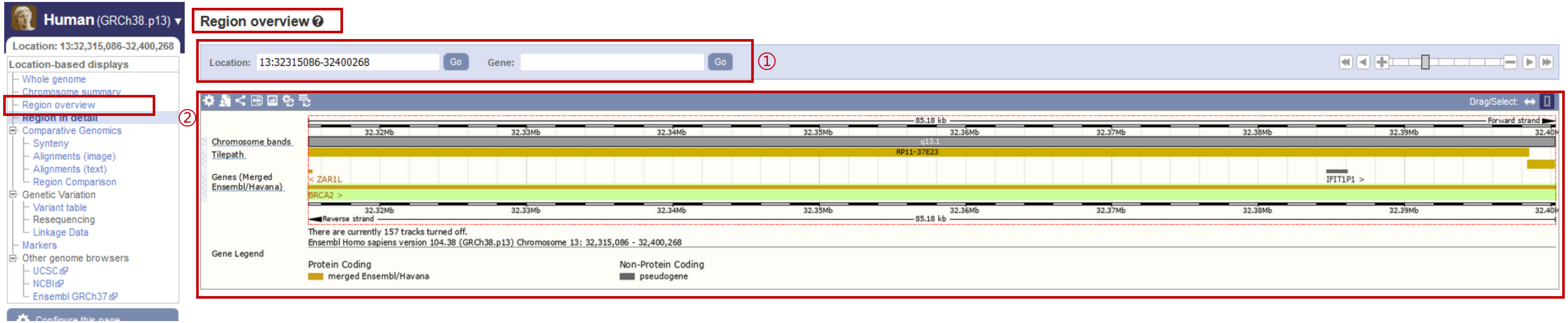
- 染色体号（Chromosome）
- 蛋白编码基因（Protein Coding Genes）
- 短链非编码基因（Short Non Coding Genes）
- 长链非编码基因（ Long Non Coding Genes ）
- 假基因（Pseudogenes）
- GC比例（GC repeats）
- 变异数（Variations ）
- 染色体基本信息统计（Chromosome Statistics）

②染色体具体数据

- 长度（ Length ）
- 编码基因（Coding genes）
- 非编码基因（Non coding genes）
- 假基因（Pseudogenes）
- 短变异体（Short Variants）

Ensembl检索方法---关键词检索

- BRCA2的染色体定位概况信息结果界面
  - 在左侧导航栏中点击Region overview，其右侧内容展示区将如下图所示



Location: 13:32,315,086-32,400,268

Location: 13:32315086-32400268

There are currently 157 tracks turned off.

Ensembl Homo sapiens version 104.38 (GRCh38.p13) Chromosome 13: 32,315,086 - 32,400,268

①染色体定位起始位点

②染色体定位具体数据

- 染色体条带（Chromosome bands）
- 染色体位置（Tilepath）
- 基因（Genes）
- 基因说明（Gene Legend）

Ensembl检索方法---关键词检索

- BRCA2的染色体定位详细信息结果界面
  - 在左侧导航栏中点击Region in detail，其右侧内容展示区将如下图所示

Chromosome 13: 32,315,086-32,400,268

Region in detail

Location: 13:32315086-32400268

Showing 93 of 141 features, due to track being limited to 40 rows by default - click to show more

Showing all 25 features - click to show fewer

①感兴趣区域，可以拖动更改感兴趣的区间，也可以通过点击染色体更改

②点击可设置 减少或增加显示信息

③该区域为我们搜索的BRCA2基因

蓝色代表有转录本，但不一定有功能

黄色代表可靠的基因

灰色代表假基因

淡粉色开放性染色质

粉色启动子侧翼序列

亮黄色代表增强子

红色代表启动子

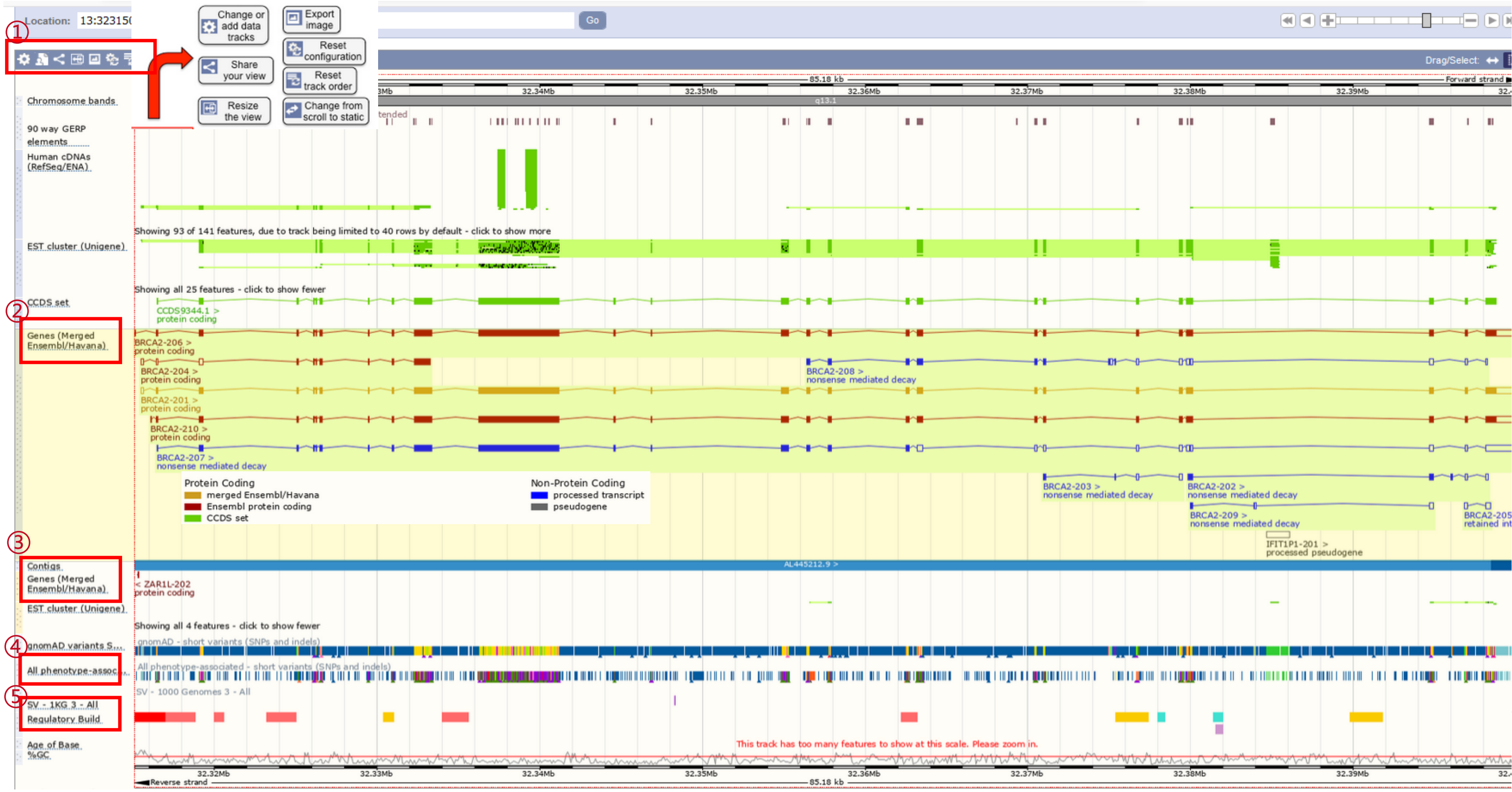
藕色转录因子结合位点

褐色代表RNA对应的基因

④对区域3的放大，下一页对该部分进行说明

Ensembl检索方法---关键词检索

- BRCA2的染色体定位详细信息结果界面
  - 接上图4的内容，如下图所示



①各图标代表的含义，其中export可以下载不同分辨率及格式的图片

②基因及不同转录本，我们需根据颜色进一步判断哪些转录本有功能：其中，黄色、红色、绿色代表有编码功能，蓝色及灰色代表无编码功能

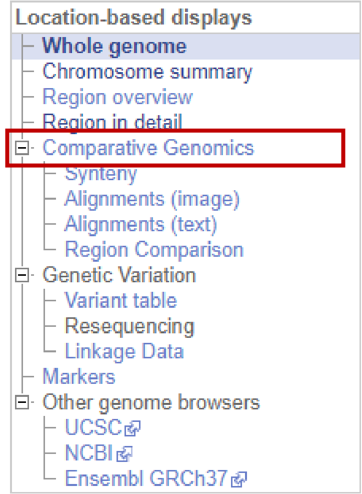
③千人基因组计划SNP信息

④与表型相关的SNP信息，点击图中竖线可显示具体表型信息

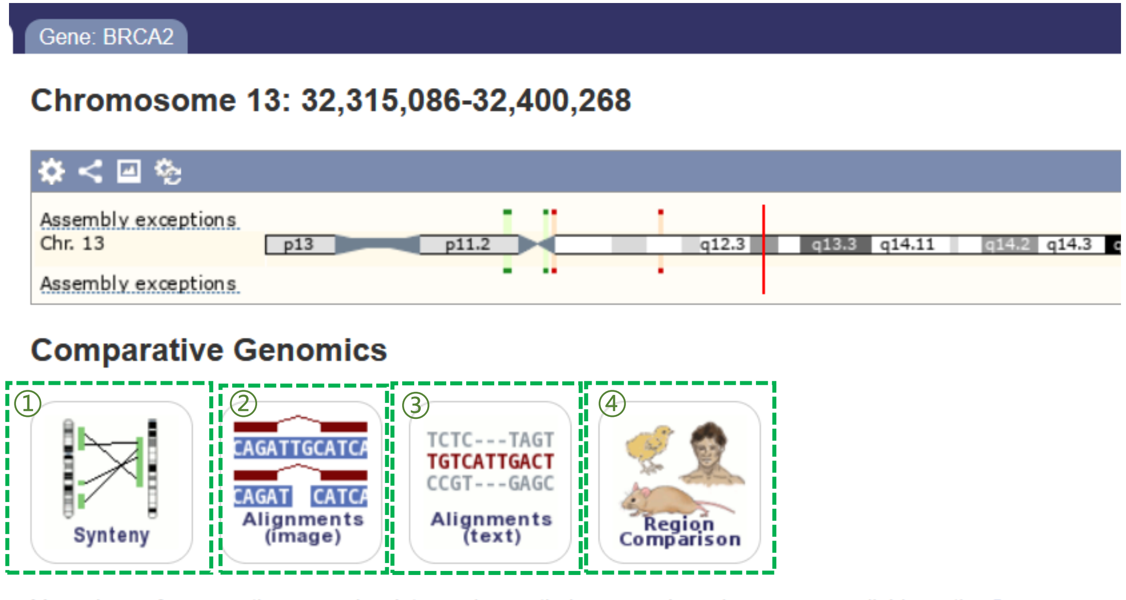
⑤调控信息注释，其中红色代表启动子，淡粉色代表启动子侧翼序列，黄色代表增强子，蓝色代表CTCF序列，藕色代表转录因子结合位点

Ensembl检索方法---关键词检索

- ■ BRCA2的染色体定位信息结果界面
  - ➢ 在左侧导航栏中点击comparative genomics，其右侧内容展示区将如下图所示。



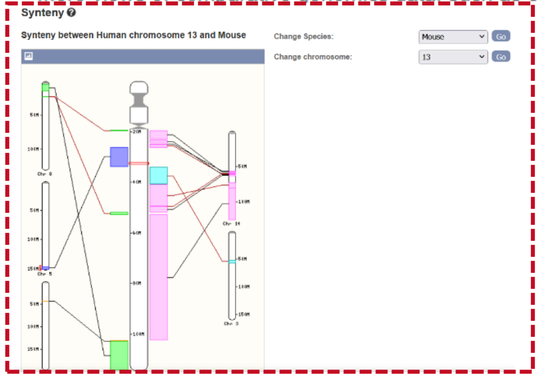
Location-based displays
- Whole genome
- Chromosome summary
- Region overview
- Region in detail
- ⊟ Comparative Genomics
  - Synteny
  - Alignments (image)
  - Alignments (text)
  - Region Comparison
- ⊟ Genetic Variation
  - Variant table
  - Resequencing
  - Linkage Data
- Markers
- ⊟ Other genome browsers
  - UCSC
  - NCBI
  - Ensembl GRCh37



Chromosome 13: 32,315,086-32,400,268

Comparative Genomics

More views of comparative genomics data, such as orthologues and paralogues, are available on the Gene page.



Synteny between Human chromosome 13 and Mouse

在BRCA2的比较基因组信息界面，可以看到以下信息——

①同线性(Synteny)

- 点击synteny即可进入同线性比较页面，如下图红色虚线框中所示
- 在change species下拉框中可以选择比较的物种
- 在change chromosome中选择比较染色体
- 图片中间长条代表的是感兴趣的物种染色体，在此例中为人第十三号染色体
- 两侧细条代表比较物种的染色体
- 粉红色区域代表同线性区域
- 黑色线条连接的是同向同线性区域
- 棕褐色线条链接的是反向同线性区域

②图片形式展示BRCA2比对序列(Alignments (images))

③文字形式展示BRCA2比对序列(Alignments (text))

④区域比较（Region comparison)

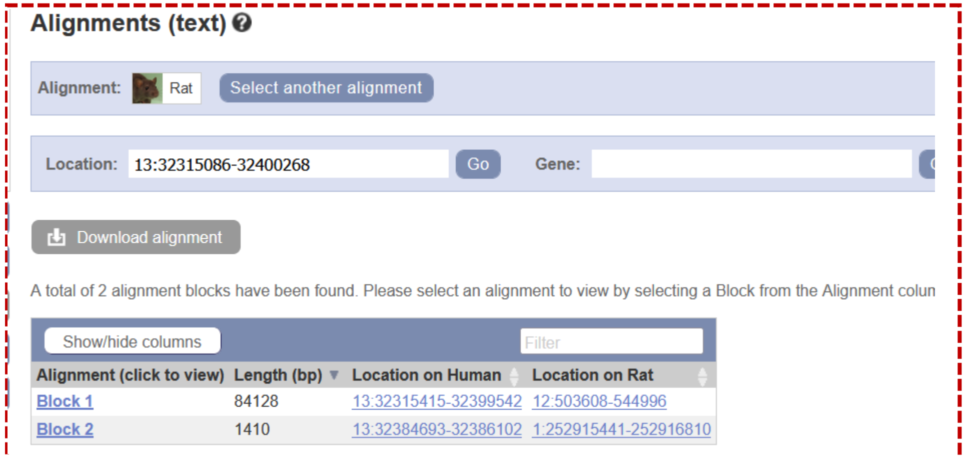
后续PPT将对链接打开的页面进行详细介绍

Ensembl检索方法---关键词检索

- BRCA2的染色体定位信息结果界面

➢ 分别点击上图②③④，其内容展示区将如下图所示。

②



③

Alignments (text)

Alignment: Rat Select another alignment

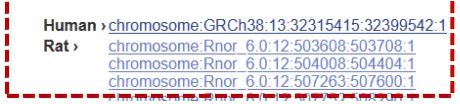
Location: 13:32315086-32400268 Go **Gene:**

Download alignment

A total of 2 alignment blocks have been found. Please select an alignment to view by selecting a Block from the Alignment colun

Show/hide columns Filter

Alignment (click to view)	Length (bp)	Location on Human	Location on Rat
Block 1	84128	13:32315415-32399542	12:503608-544996
Block 2	1410	13:32384693-32386102	1:252915441-252916810



Human › chromosome:GRCh38:13:32315415:32399542:1
Rat › chromosome:Rnor_6.0:12:503608:503708:1
chromosome:Rnor_6.0:12:504008:504404:1
chromosome:Rnor_6.0:12:507263:507600:1

④

②图片形式展示BRCA2比对序列(Alignments (images))

- 上下两个蓝色框中内容分别代表感兴趣物种（人）和选择比较的物种（大鼠）染色体定位信息
- 染色体倒位或中断用三角形标注，单击三角形标注可了解详细信息

③文字形式展示BRCA2比对序列

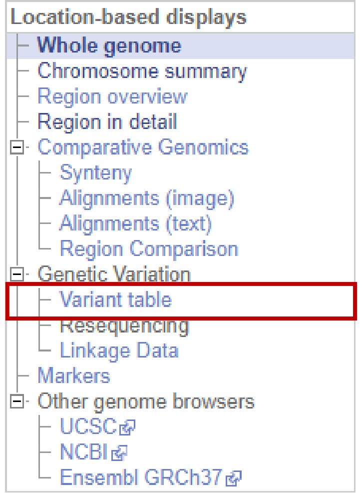
- 两个物种序列比较的文字版页面，分别点击列表中block可进一步详细查看具体内容
- 点击Block1进入下边红色框中页面内容
- 第一行代表人基因组及染色体、基因起始位置具体信息
- 第二行代表大鼠，点击可连接到图片版本

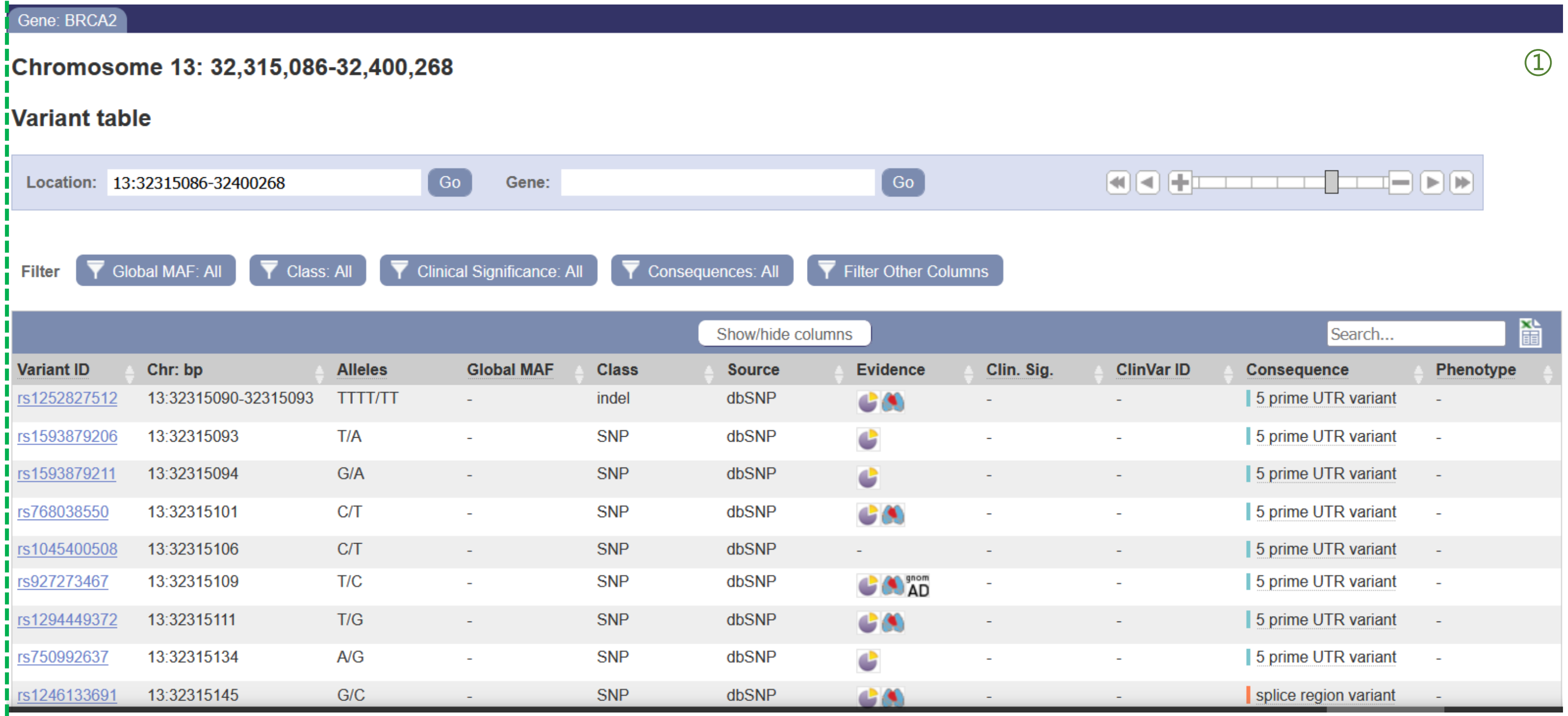
④区域比较（Region comparision)

- 人类第13号染色体32315415-32399542基因信息对应大鼠12号染色体对应区段的基因信息
- 绿色阴影连接对应的基因序列

Ensembl检索方法---关键词检索

- BRCA2的染色体定位信息结果界面
  - 在左侧导航栏中点击Variant table，其右侧内容展示区将如下图所示。





Chromosome 13: 32,315,086-32,400,268

Variant table

Variant ID	Chr: bp	Alleles	Global MAF	Class	Source	Evidence	Clin. Sig.	ClinVar ID	Consequence	Phenotype
rs1252827512	13:32315090-32315093	TTTT/TT	-	indel	dbSNP		-	-	5 prime UTR variant	-
rs1593879206	13:32315093	T/A	-	SNP	dbSNP		-	-	5 prime UTR variant	-
rs1593879211	13:32315094	G/A	-	SNP	dbSNP		-	-	5 prime UTR variant	-
rs768038550	13:32315101	C/T	-	SNP	dbSNP		-	-	5 prime UTR variant	-
rs1045400508	13:32315106	C/T	-	SNP	dbSNP	-	-	-	5 prime UTR variant	-
rs927273467	13:32315109	T/C	-	SNP	dbSNP	gnomAD	-	-	5 prime UTR variant	-
rs1294449372	13:32315111	T/G	-	SNP	dbSNP		-	-	5 prime UTR variant	-
rs750992637	13:32315134	A/G	-	SNP	dbSNP		-	-	5 prime UTR variant	-
rs1246133691	13:32315145	G/C	-	SNP	dbSNP		-	-	splice region variant	-

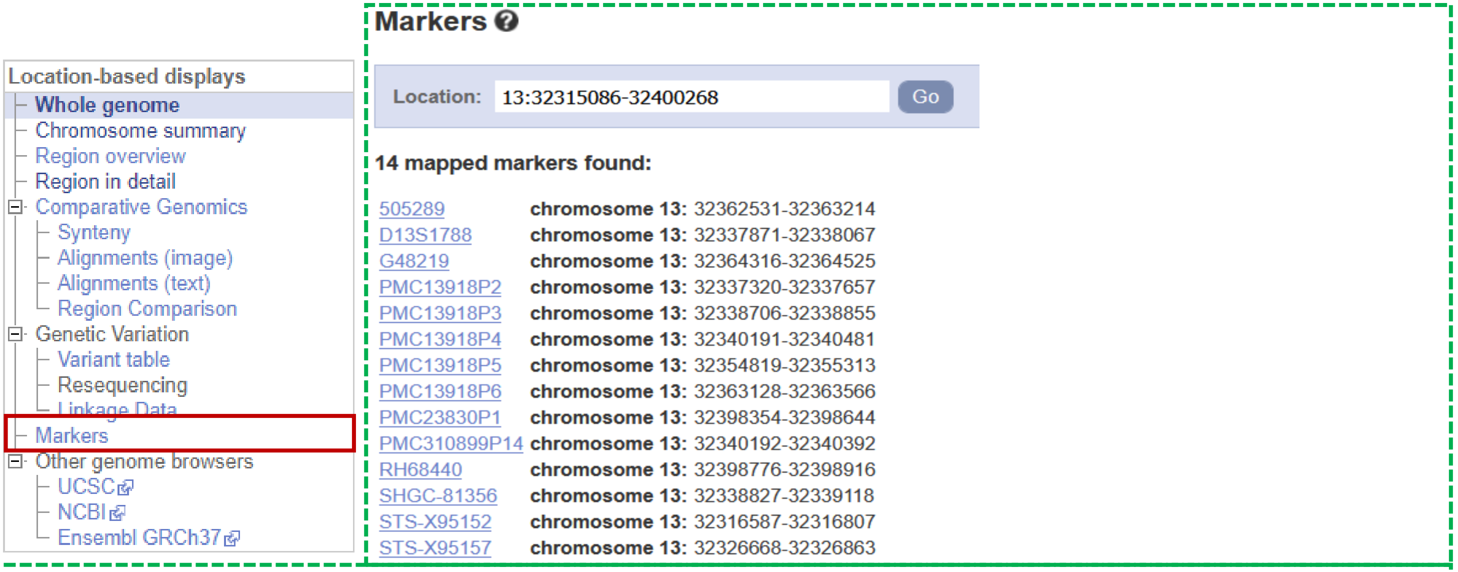
①变异信息

具体每个含义在变异部分将详细介绍

Ensembl检索方法---关键词检索

■ BRCA2的染色体定位信息结果界面

➢ 在左侧导航栏中点击markers，其右侧内容展示区将如下图①所示，进一步点击列表中marker具体内容如图② 所示。

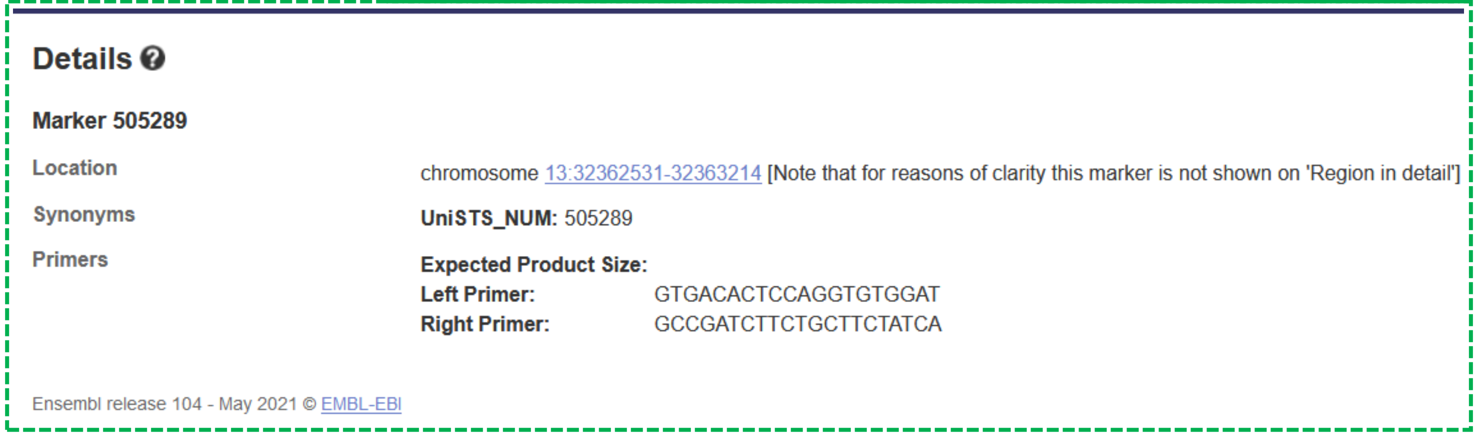


①

Markers

14 mapped markers found:

505289	**chromosome 13:** 32362531-32363214
D13S1788	**chromosome 13:** 32337871-32338067
G48219	**chromosome 13:** 32364316-32364525
PMC13918P2	**chromosome 13:** 32337320-32337657
PMC13918P3	**chromosome 13:** 32338706-32338855
PMC13918P4	**chromosome 13:** 32340191-32340481
PMC13918P5	**chromosome 13:** 32354819-32355313
PMC13918P6	**chromosome 13:** 32363128-32363566
PMC23830P1	**chromosome 13:** 32398354-32398644
PMC310899P14	**chromosome 13:** 32340192-32340392
RH68440	**chromosome 13:** 32398776-32398916
SHGC-81356	**chromosome 13:** 32338827-32339118
STS-X95152	**chromosome 13:** 32316587-32316807
STS-X95157	**chromosome 13:** 32326668-32326863



②

Details

Marker 505289

Location	chromosome 13:32362531-32363214 [Note that for reasons of clarity this marker is not shown on 'Region in detail']
Synonyms	**UniSTS_NUM:** 505289
Primers	**Expected Product Size:** **Left Primer:** GTGACACTCCAGGTGTGGAT **Right Primer:** GCCGATCTTCTGCTTCTATCA

Ensembl release 104 - May 2021 © EMBL-EBI

①标志物（Markers）

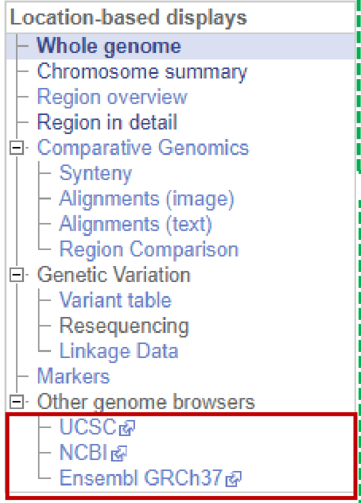
- 列表中显示该染色体区域的标志物
- 点击任一marker即可链接到详细信息页面，如图②

②点击marker 505289进入详细信息页面

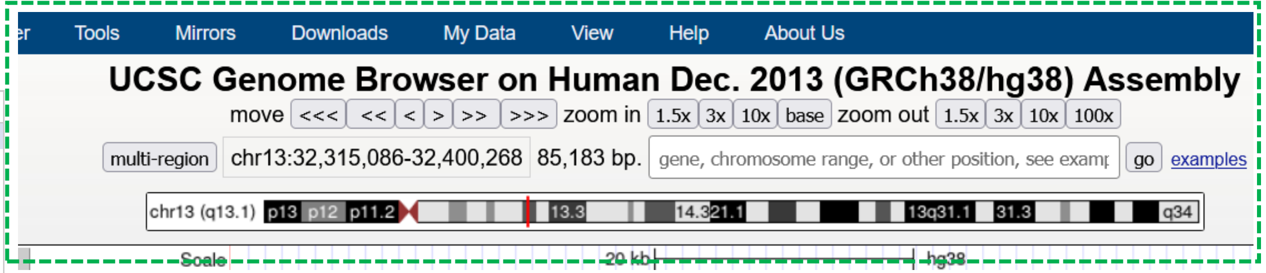
- 染色体定位信息
- 其它名称
- 引物信息

Ensembl检索方法---关键词检索

- ■ BRCA2的染色体定位信息结果界面
  - ➢ 在左侧导航栏中点击Other genome browsers，其右侧内容展示区将如下图所示。



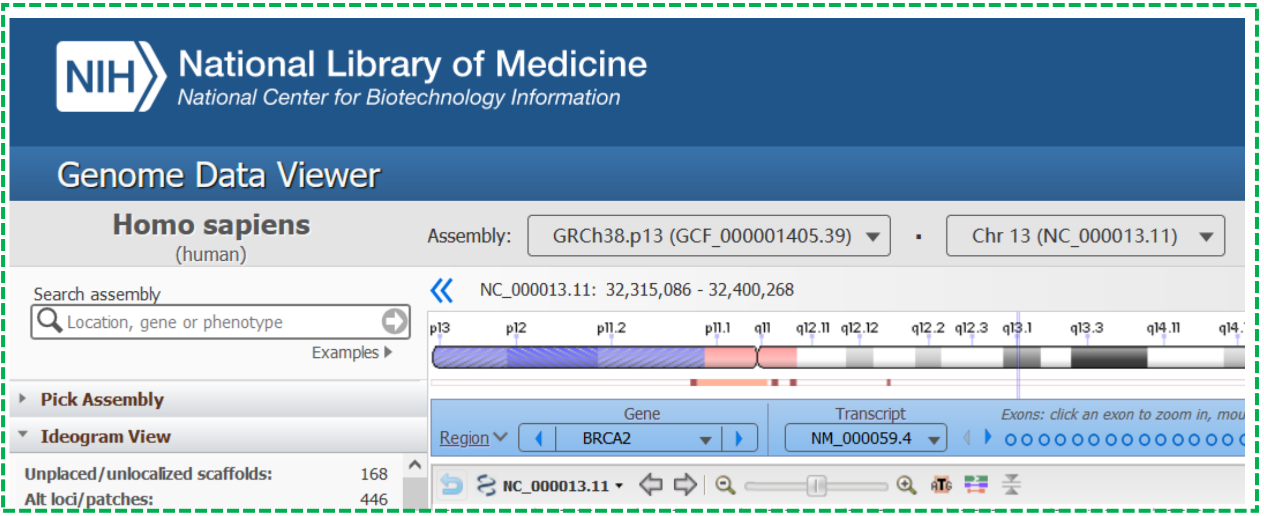
①



UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

chr13:32,315,086-32,400,268 85,183 bp. gene, chromosome range, or other position, see examp

②



Homo sapiens
(human)

Assembly: GRCh38.p13 (GCF_000001405.39) · Chr 13 (NC_000013.11)

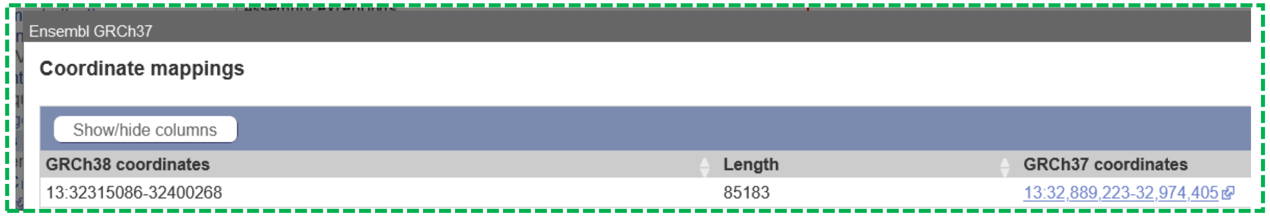
NC_000013.11: 32,315,086 - 32,400,268

Gene: BRCA2 Transcript: NM_000059.4

Exons: click an exon to zoom in, mou

Unplaced/unlocalized scaffolds: 168
Alt loci/patches: 446

③



Coordinate mappings

GRCh38 coordinates	Length	GRCh37 coordinates
13:32315086-32400268	85183	13:32,889,223-32,974,405

①点击左侧USCS进入UCSC主界面

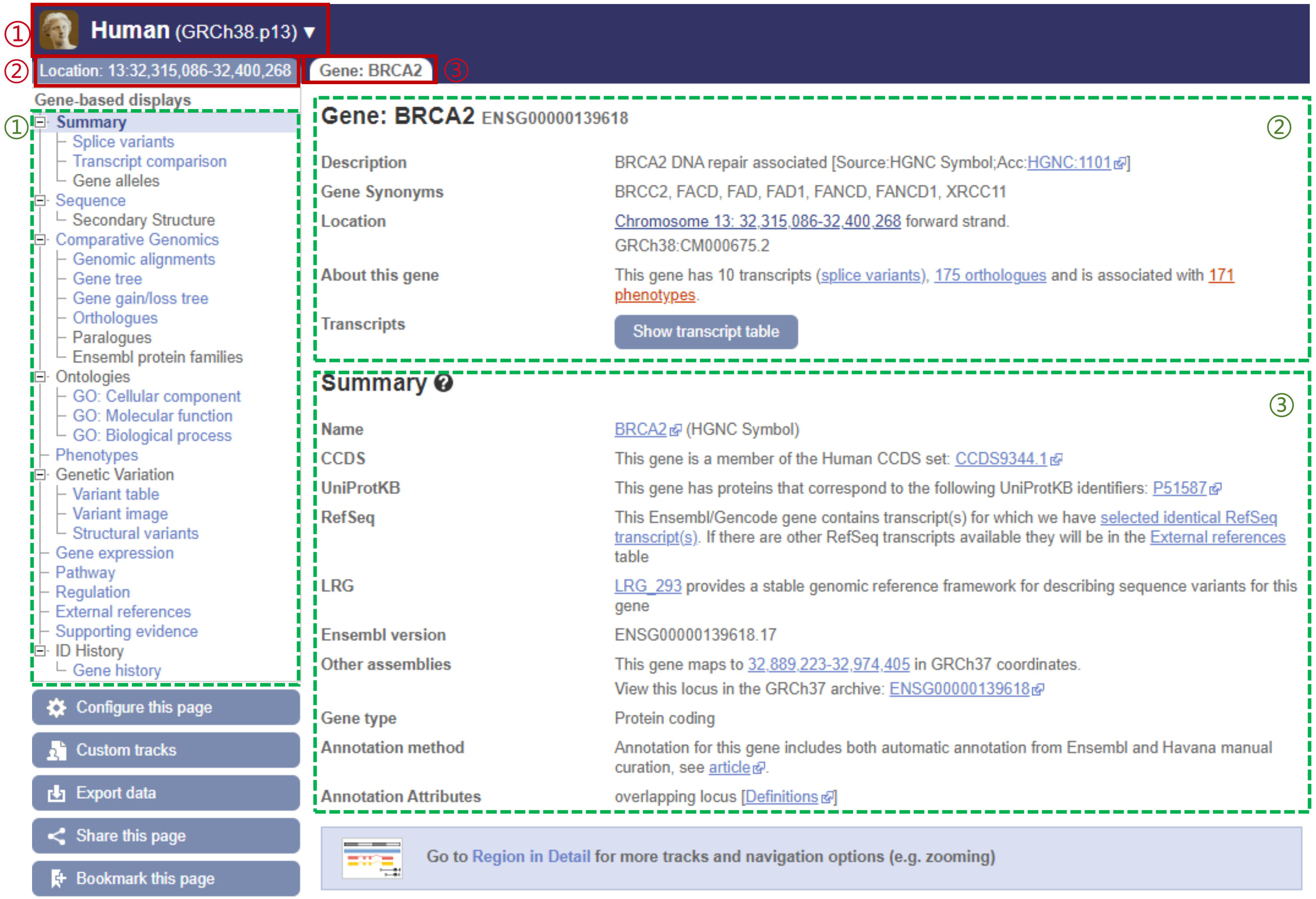
②点击左侧NCBI进入NCBI主界面

③点击左侧Ensembl GRCh37显示具体区域信息

- 具体各界面主要内容请详见各数据库单元课教程

Ensembl检索方法---关键词检索

5 在结果界面，点击③可了解BRCA2的基因信息，其界面如下面所示



① **Human** (GRCh38.p13) ▾

② Location: 13:32,315,086-32,400,268 Gene: BRCA2 ③

Gene-based displays

①
- Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Cellular component
  - GO: Molecular function
  - GO: Biological process
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history

Configure this page
Custom tracks
Export data
Share this page
Bookmark this page

Gene: BRCA2 ENSG00000139618 ②

Description	BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]
Gene Synonyms	BRCC2, FACD, FAD, FAD1, FANCD, FANCD1, XRCC11
Location	Chromosome 13: 32,315,086-32,400,268 forward strand. GRCh38:CM000675.2
About this gene	This gene has 10 transcripts (splice variants), 175 orthologues and is associated with 171 phenotypes.
Transcripts	Show transcript table

Summary ❓ ③

Name	BRCA2 (HGNC Symbol)
CCDS	This gene is a member of the Human CCDS set: CCDS9344.1
UniProtKB	This gene has proteins that correspond to the following UniProtKB identifiers: P51587
RefSeq	This Ensembl/Gencode gene contains transcript(s) for which we have selected identical RefSeq transcript(s). If there are other RefSeq transcripts available they will be in the External references table
LRG	LRG_293 provides a stable genomic reference framework for describing sequence variants for this gene
Ensembl version	ENSG00000139618.17
Other assemblies	This gene maps to 32,889,223-32,974,405 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000139618
Gene type	Protein coding
Annotation method	Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.
Annotation Attributes	overlapping locus [Definitions]

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

在BRCA2的基因界面，可以看到以下信息——

①左侧导航栏

- Summary（基础信息）
- Sequence（序列信息）
- Comparative Genomic（比较基因组）
- Ontologies（同源基因）
- Genetic Variation（遗传变异）
- ID History（ID历史）

②BRCA2基因的简单描述

- Description（描述）
- Gene Synonyms（基因别名）
- Location（染色体定位）
- About this gene（关于该基因）
- Trascripts（转录本），点击Show Transcript table可查看详细信息，详看下张PPT。

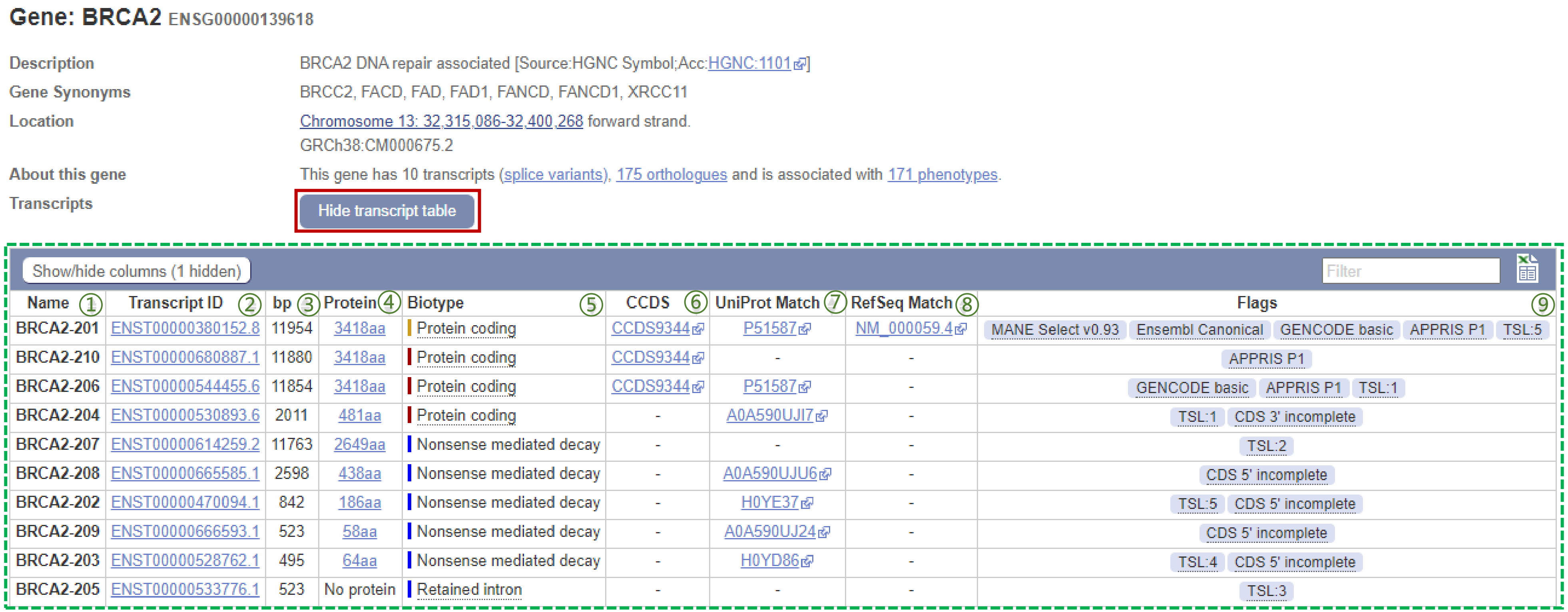
③右侧内容展示区

- 点击左侧导航栏相应链接，右侧内容亦会出现变化

后续PPT将对导航栏各个按钮所展示的内容进行详细介绍

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 点击Show Transcript table可查看详细信息



Gene: BRCA2 ENSG00000139618

Description	BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]
Gene Synonyms	BRCC2, FACD, FAD, FAD1, FANCD, FANCD1, XRCC11
Location	Chromosome 13: 32,315,086-32,400,268 forward strand. GRCh38:CM000675.2
About this gene	This gene has 10 transcripts (splice variants), 175 orthologues and is associated with 171 phenotypes.
Transcripts	Hide transcript table

Name ①	Transcript ID ②	bp ③	Protein ④	Biotype ⑤	CCDS ⑥	UniProt Match ⑦	RefSeq Match ⑧	Flags ⑨
BRCA2-201	ENST00000380152.8	11954	3418aa	Protein coding	CCDS9344	P51587	NM_000059.4	MANE Select v0.93 Ensembl Canonical GENCODE basic APPRIS P1 TSL:5
BRCA2-210	ENST00000680887.1	11880	3418aa	Protein coding	CCDS9344	-	-	APPRIS P1
BRCA2-206	ENST00000544455.6	11854	3418aa	Protein coding	CCDS9344	P51587	-	GENCODE basic APPRIS P1 TSL:1
BRCA2-204	ENST00000530893.6	2011	481aa	Protein coding	-	A0A590UJI7	-	TSL:1 CDS 3' incomplete
BRCA2-207	ENST00000614259.2	11763	2649aa	Nonsense mediated decay	-	-	-	TSL:2
BRCA2-208	ENST00000665585.1	2598	438aa	Nonsense mediated decay	-	A0A590UJU6	-	CDS 5' incomplete
BRCA2-202	ENST00000470094.1	842	186aa	Nonsense mediated decay	-	H0YE37	-	TSL:5 CDS 5' incomplete
BRCA2-209	ENST00000666593.1	523	58aa	Nonsense mediated decay	-	A0A590UJ24	-	CDS 5' incomplete
BRCA2-203	ENST00000528762.1	495	64aa	Nonsense mediated decay	-	H0YD86	-	TSL:4 CDS 5' incomplete
BRCA2-205	ENST00000533776.1	523	No protein	Retained intron	-	-	-	TSL:3

①名称

②转录本ID

- 点击Ensembl号可查看转录本信息，见下页PPT

③ 转录本长度

④ 蛋白氨基酸量

⑤表型

- protein coding编码蛋白功能 nonsense mediated decay无意衰变 retained intron保留的内含子

⑥CCDS序列

⑦UniProt匹配信息

⑧ 参考文献

⑨ 所用Flag情况

- ENST00000380152.8代表的含义：

Ensembl ID遵循Ensembl “Genebuild”过程中所有标识符的规定。这些标识符包括：

ENS代表Ensembl 作为开头

G代表基因

T代表转录本

E代表外显子

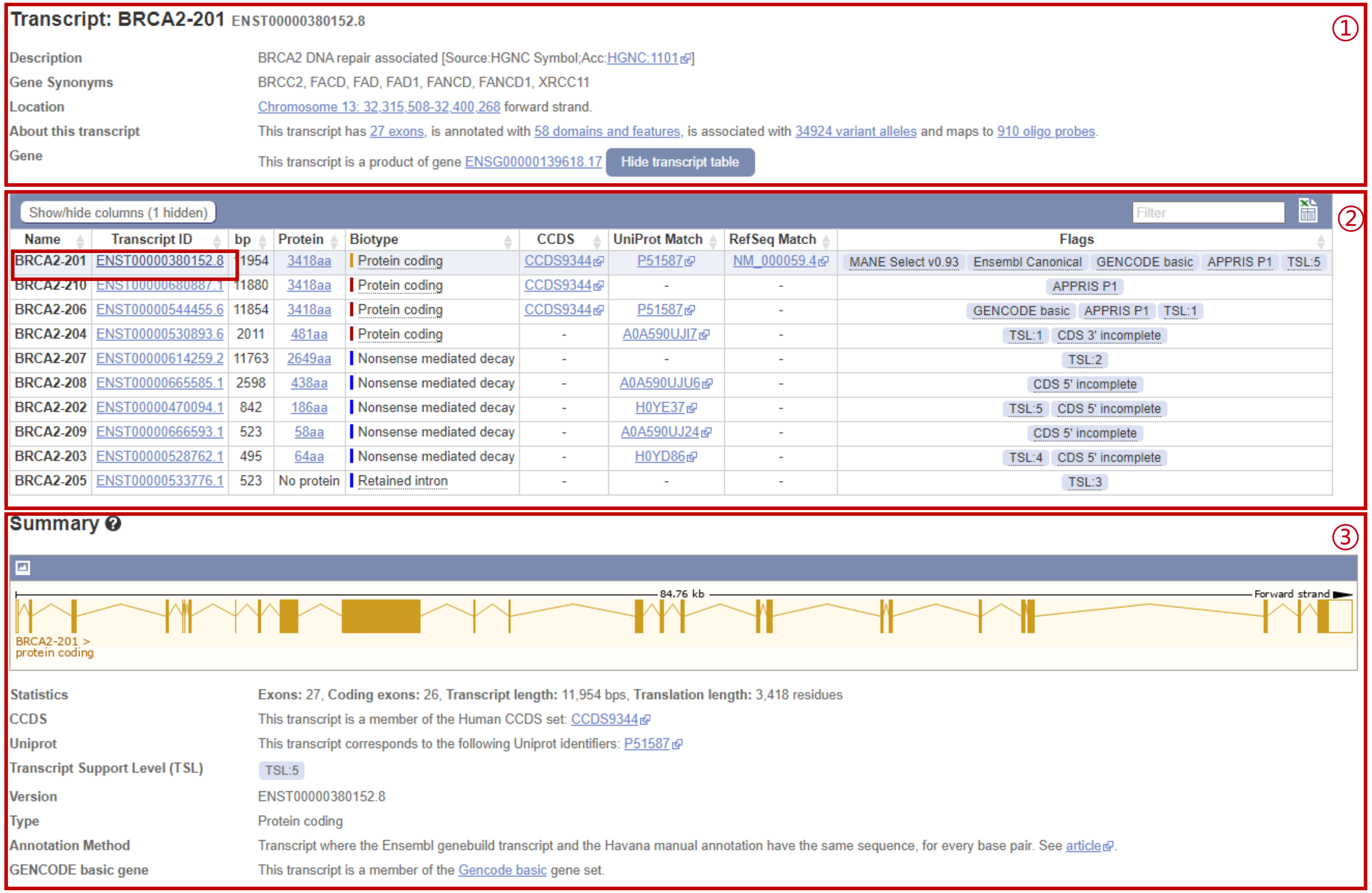
P代表肽段

F代表家族

后边11位数字是稳定的，即如果没有相应的基因序列或注释的变化，它不会在新版本发布时发生变化

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 点击Transcript table中的第一个转录本信息，可跳转至以下界面。



Transcript: BRCA2-201 ENST00000380152.8

①

Description	BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]
Gene Synonyms	BRCC2, FACD, FAD, FAD1, FANCD, FANCD1, XRCC11
Location	Chromosome 13: 32,315,508-32,400,268 forward strand.
About this transcript	This transcript has 27 exons, is annotated with 58 domains and features, is associated with 34924 variant alleles and maps to 910 oligo probes.
Gene	This transcript is a product of gene ENSG00000139618.17 Hide transcript table

②

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt Match	RefSeq Match	Flags
BRCA2-201	ENST00000380152.8	11954	3418aa	Protein coding	CCDS9344	P51587	NM_000059.4	MANE Select v0.93 Ensembl Canonical GENCODE basic APPRIS P1 TSL:5
BRCA2-210	ENST00000680887.1	11880	3418aa	Protein coding	CCDS9344	-	-	APPRIS P1
BRCA2-206	ENST00000544455.6	11854	3418aa	Protein coding	CCDS9344	P51587	-	GENCODE basic APPRIS P1 TSL:1
BRCA2-204	ENST00000530893.6	2011	481aa	Protein coding	-	A0A590UJI7	-	TSL:1 CDS 3' incomplete
BRCA2-207	ENST00000614259.2	11763	2649aa	Nonsense mediated decay	-	-	-	TSL:2
BRCA2-208	ENST00000665585.1	2598	438aa	Nonsense mediated decay	-	A0A590UJU6	-	CDS 5' incomplete
BRCA2-202	ENST00000470094.1	842	186aa	Nonsense mediated decay	-	H0YE37	-	TSL:5 CDS 5' incomplete
BRCA2-209	ENST00000666593.1	523	58aa	Nonsense mediated decay	-	A0A590UJ24	-	CDS 5' incomplete
BRCA2-203	ENST00000528762.1	495	64aa	Nonsense mediated decay	-	H0YD86	-	TSL:4 CDS 5' incomplete
BRCA2-205	ENST00000533776.1	523	No protein	Retained intron	-	-	-	TSL:3

Summary

③

Statistics	**Exons:** 27, **Coding exons:** 26, **Transcript length:** 11,954 bps, **Translation length:** 3,418 residues
CCDS	This transcript is a member of the Human CCDS set: CCDS9344
Uniprot	This transcript corresponds to the following Uniprot identifiers: P51587
Transcript Support Level (TSL)	TSL:5
Version	ENST00000380152.8
Type	Protein coding
Annotation Method	Transcript where the Ensembl genebuild transcript and the Havana manual annotation have the same sequence, for every base pair. See article.
GENCODE basic gene	This transcript is a member of the Gencode basic gene set.

① 转录本信息

- 描述(Description)
- 基因别名(Gene synonyms)
- 位置(Location)
- 关于此转录本(About this transcript)
- 基因概述(Gene)

② 转录本列表

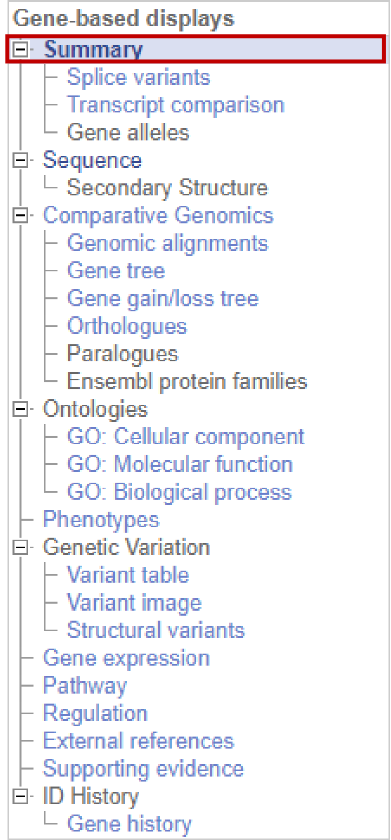
- 具体内容详见上页PPT

③转录本信息总结

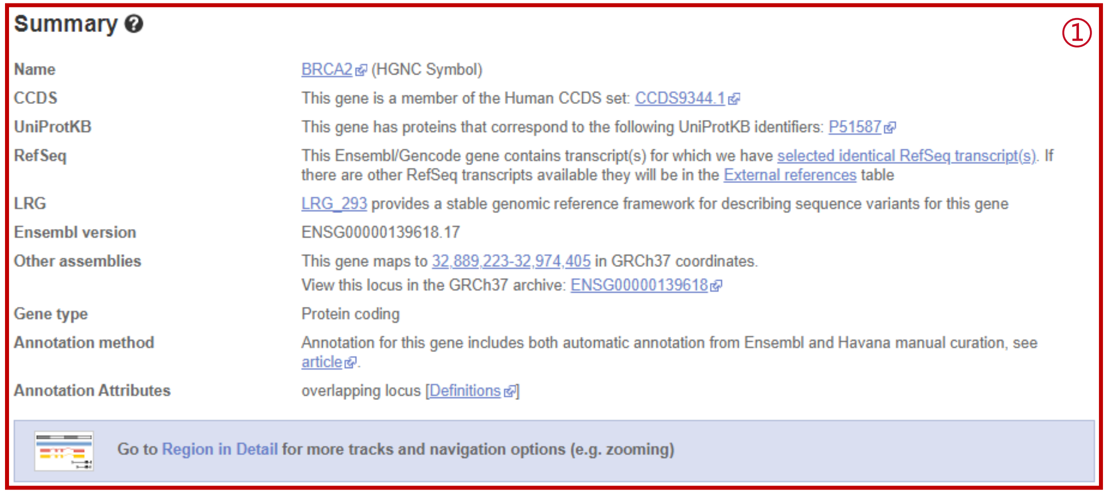
- 统计信息(Statistics)
- CCDS序列(CCDS)
- Uniprot参考信息(Uniprot)
- 转录本证据等级(Transcript Support Level)
- 转录本版本(Version)
- 转录本表型(Type)
- 注释方法(Annotation Method)
- GENCODE基因(GENCODE basic gene)

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击Summary，其右侧内容展示区将如下图所示。



Gene-based displays
- ⊟ **Summary**
  - Splice variants
  - Transcript comparison
  - Gene alleles
- ⊟ Sequence
  - Secondary Structure
- ⊟ Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- ⊟ Ontologies
  - GO: Cellular component
  - GO: Molecular function
  - GO: Biological process
- Phenotypes
- ⊟ Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ⊟ ID History
  - Gene history

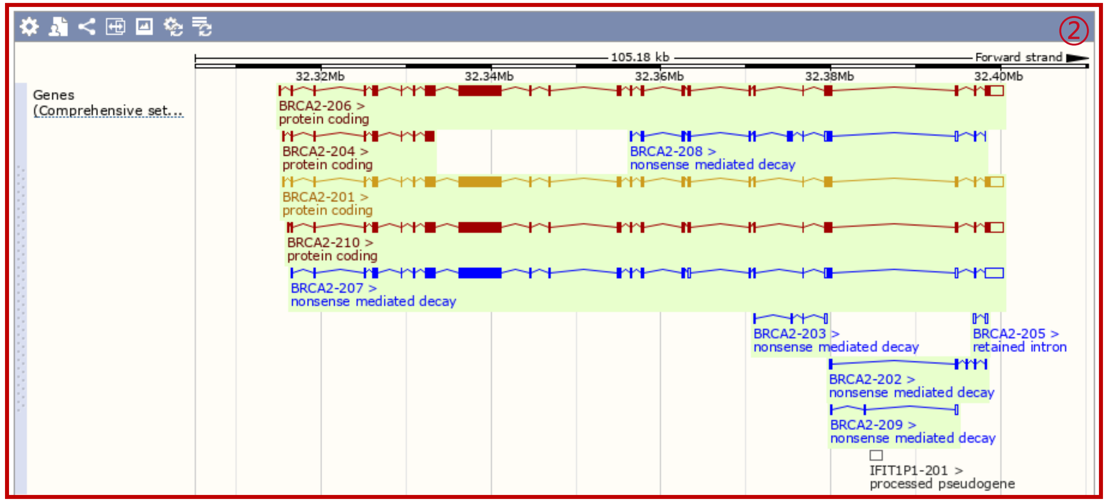


Summary ?

①

Name	BRCA2 (HGNC Symbol)
CCDS	This gene is a member of the Human CCDS set: CCDS9344.1
UniProtKB	This gene has proteins that correspond to the following UniProtKB identifiers: P51587
RefSeq	This Ensembl/Gencode gene contains transcript(s) for which we have selected identical RefSeq transcript(s). If there are other RefSeq transcripts available they will be in the External references table
LRG	LRG_293 provides a stable genomic reference framework for describing sequence variants for this gene
Ensembl version	ENSG00000139618.17
Other assemblies	This gene maps to 32,889,223-32,974,405 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000139618
Gene type	Protein coding
Annotation method	Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.
Annotation Attributes	overlapping locus [Definitions]

Go to Region in Detail for more tracks and navigation options (e.g. zooming)



① 概述

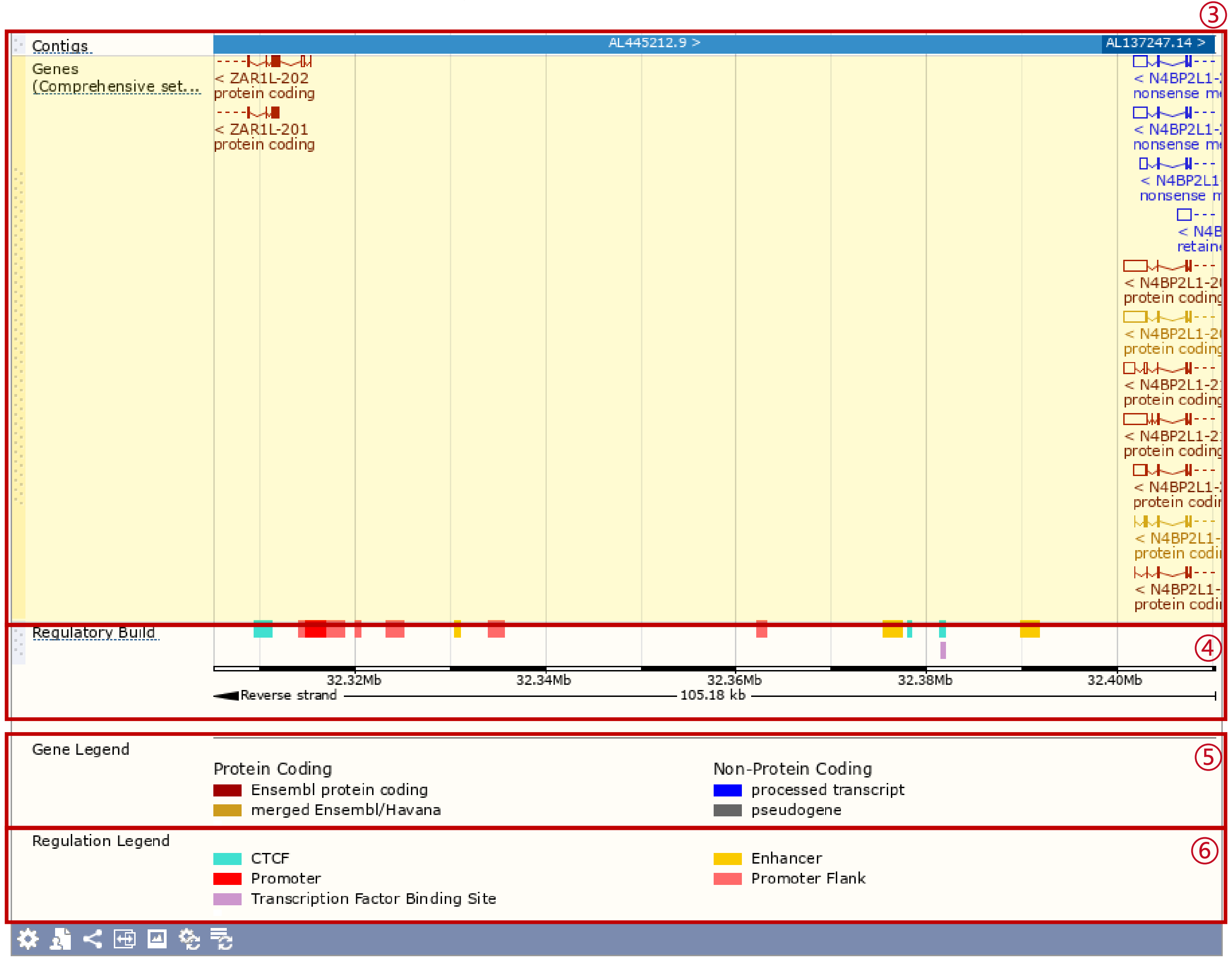
- 名称(Name)
- CCDS号
- UniProtKB链接
- 参考文献（RefSeq）
- LRG（LRG数据信息）
- Ensembl 号&Other assemblies
- 基因类型（Gene type）
- 注释方法（Annotation method）
- 注释属性（Annotation Attributes）

② 基因结构及序列信息

- 每个转录本的基因结构，外显子由框表示，连接框的线为内含子，填充框反应编码序列，未填充框反应UTRs（未翻译区）
- 红色：转录本是蛋白质编码，来自Ensembl或VEGA/Havana
- 黄色：转录本是Ensembl与VEGA/Havana协商而定
- 蓝色、灰色或粉红色：转录本属于非编码区

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击Summary，其右侧内容展示区将如下图所示——接上页PPT。



③ 基因结构及序列信息

- 每个转录本的基因结构，外显子由框表示，连接框的线为内含子，填充框反应编码序列，未填充框反应UTRs（未翻译区）
- 红色：转录本是蛋白质编码，来自Ensembl或VEGA/Havana
- 黄色：转录本是Ensembl与VEGA/Havana协商而定
- 蓝色、灰色或粉红色：转录本属于非编码区

④ 调节结构

- 浅蓝：CTCT区 黄色：增强子
- 红色：启动子 粉红色：启动子侧翼序列
- 藕粉：转录因子结合位点

⑤ 基因标识

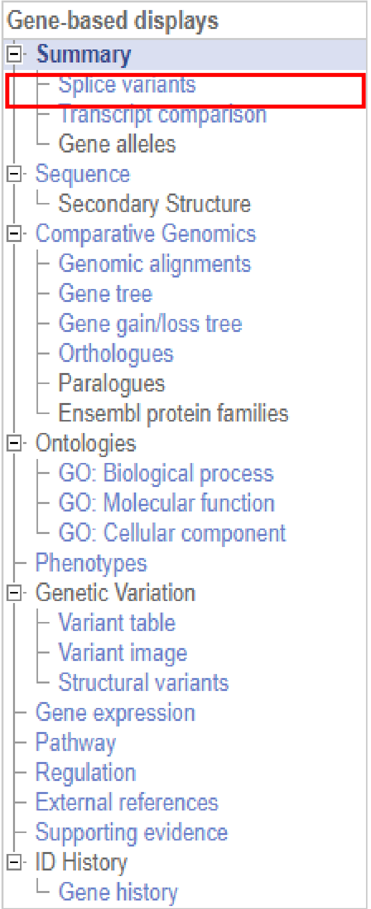
- 红色：转录本是蛋白质编码，来自Ensembl或VEGA/Havana
- 黄色：转录本是Ensembl与VEGA/Havana协商而定
- 蓝色、灰色或粉红色：转录本属于非编码区

⑥ 调节标识

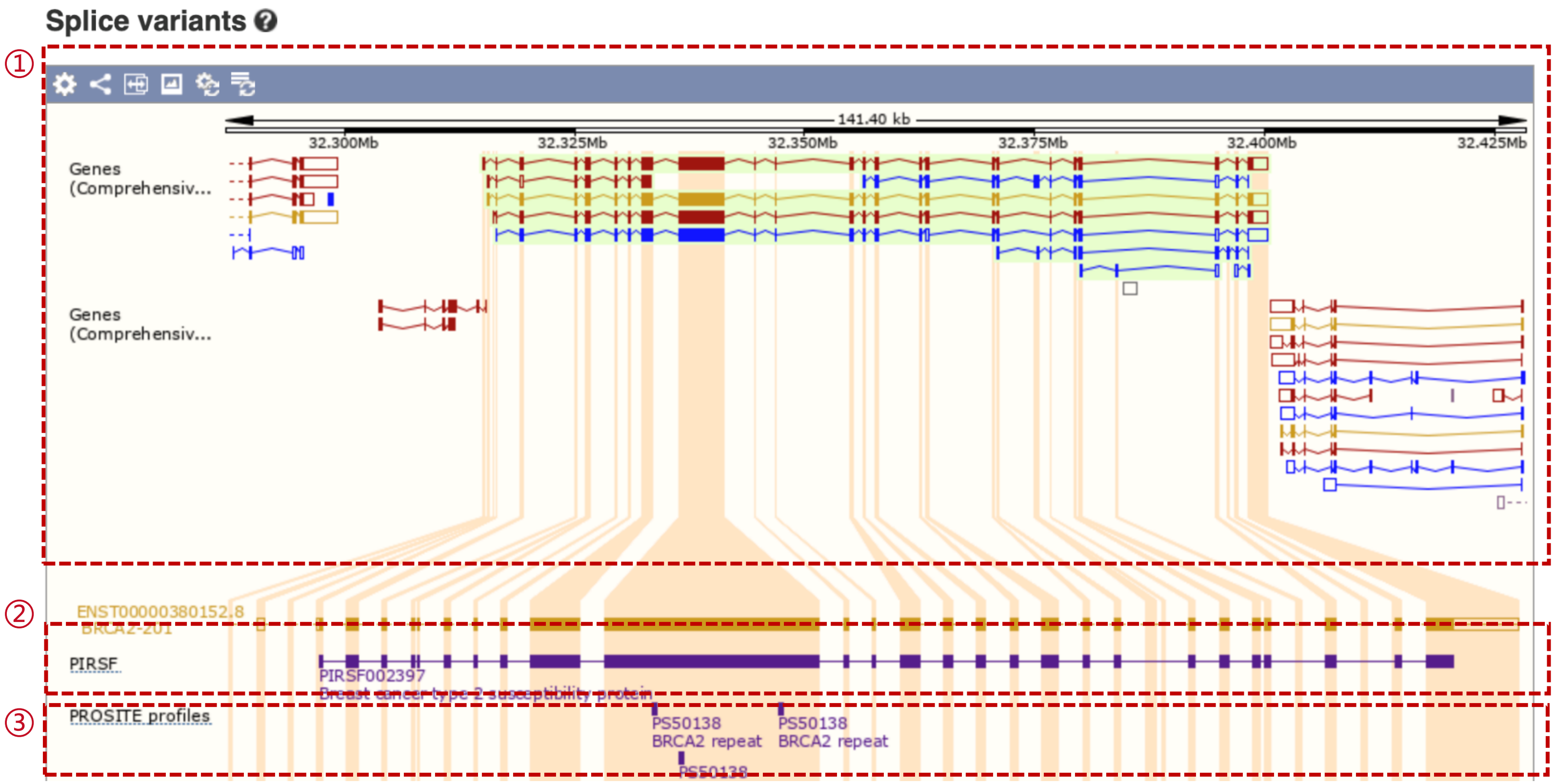
- 浅蓝：CTCF区 黄色：增强子
- 红色：启动子 粉红色：启动子侧翼序列
- 藕粉：转录因子结合位点

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击Splice variants，其右侧内容展示区将如下图所示。



- Gene-based displays
  - Summary
    - Splice variants
    - Transcript comparison
    - Gene alleles
  - Sequence
    - Secondary Structure
  - Comparative Genomics
    - Genomic alignments
    - Gene tree
    - Gene gain/loss tree
    - Orthologues
    - Paralogues
    - Ensembl protein families
  - Ontologies
    - GO: Biological process
    - GO: Molecular function
    - GO: Cellular component
  - Phenotypes
  - Genetic Variation
    - Variant table
    - Variant image
    - Structural variants
  - Gene expression
  - Pathway
  - Regulation
  - External references
  - Supporting evidence
  - ID History
    - Gene history



①可变剪切信息

- 每个转录本的基因结构，外显子由框表示，连接框的线为内含子，填充框反应编码序列，未填充框反应UTRs（未翻译区）
- 红色：转录本是蛋白质编码，来自Ensembl或VEGA/Havana
- 黄色：转录本是Ensembl与VEGA/Havana协商而定

②转录本对应的蛋白信息

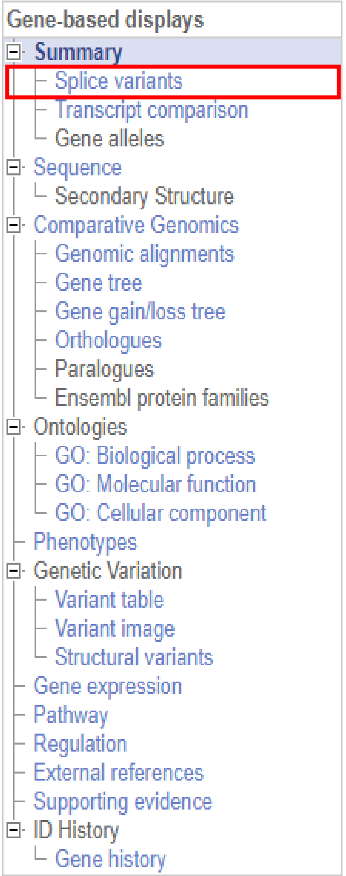
- PIRSF数据库来源蛋白信息

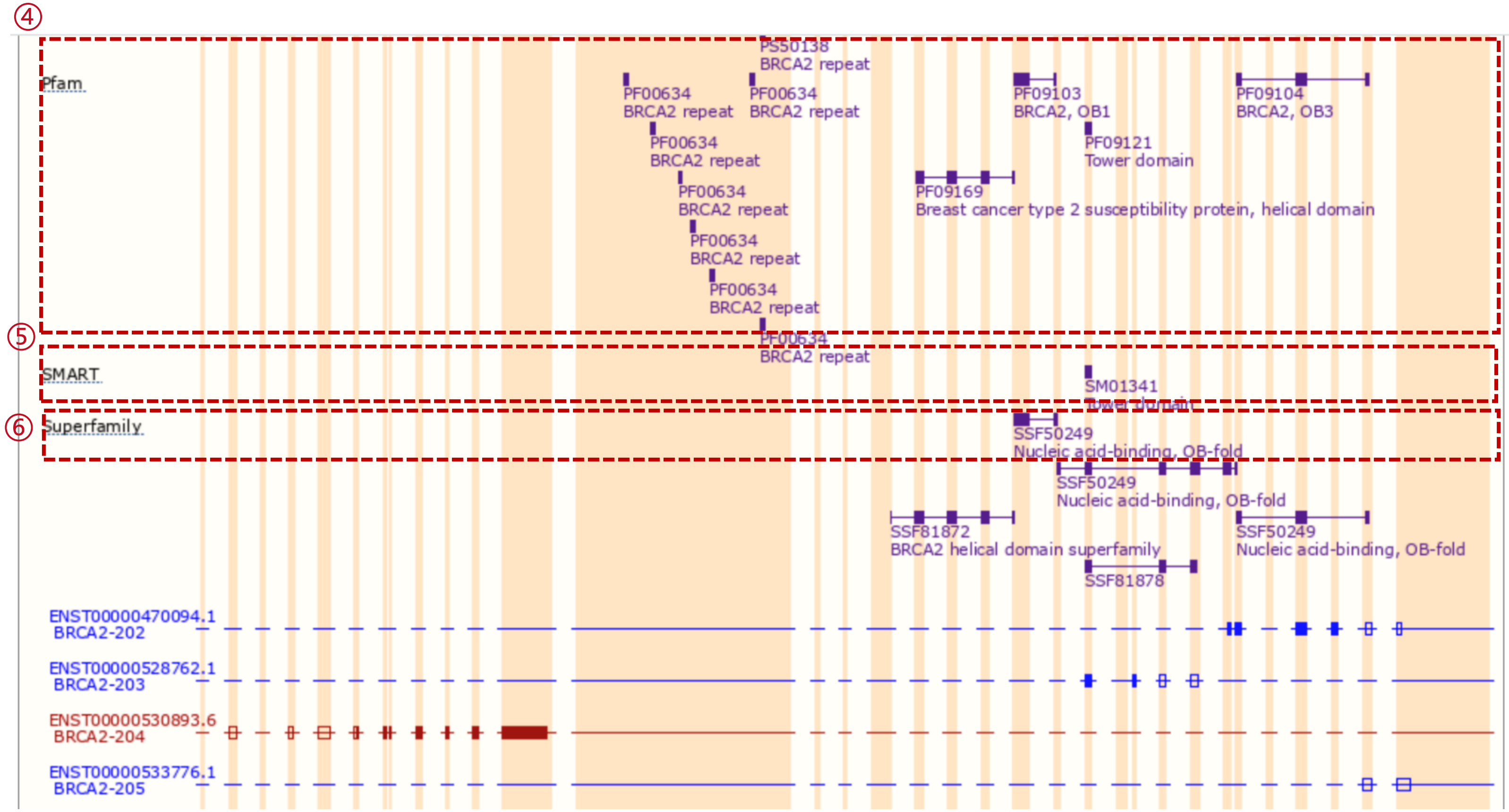
③PROSITE数据库来源的蛋白质位点和模型信息

- PROSITE提供一级蛋白结构信息

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - ➤ 在左侧导航栏中点击Splice variants，其右侧内容展示区将如下图所示。





④pfam数据库对应的蛋白信息

- Pfam数据库是一系列蛋白质家族的集合,其中每一个蛋白家族都以多序列比对和隐马尔科夫模型的形式来表示。提供二级蛋白结构信息

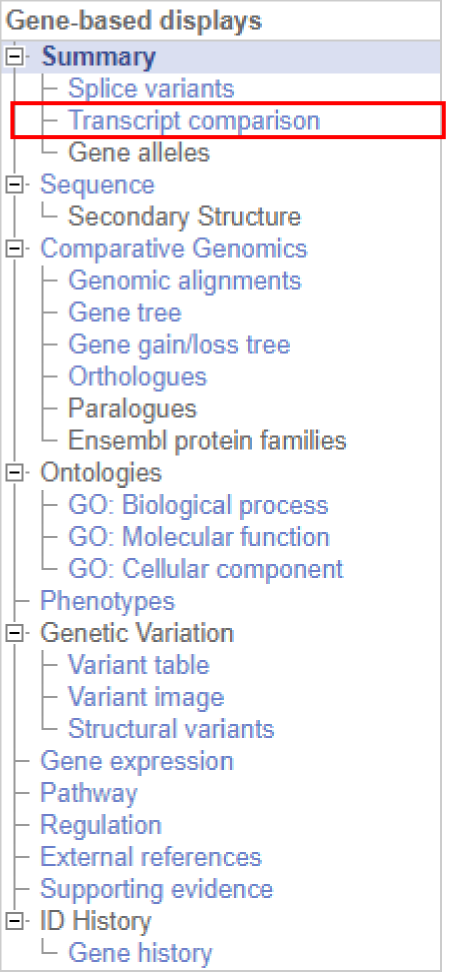
⑤SMART数据库提供的蛋白结构域信息

- SMART是蛋白结构域的数据库,该数据库最新版本为v8,收录了1300多个蛋白结构域信息,覆盖了来自uniprot, ensembl等多个数据库的蛋白。

⑥Superfamily数据库蛋白信息

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 点击左侧导航栏中点击Transcript comparison，可进行不同转录本的比较。

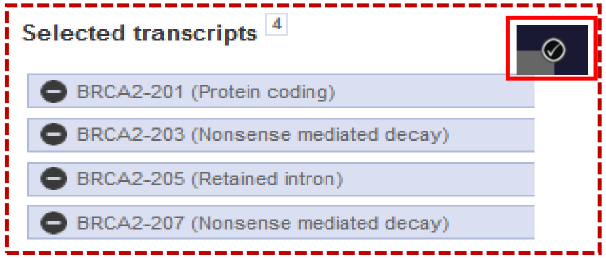


①

No transcripts selected

You must select transcripts using the "Select transcripts" button from menu on the left hand side of this page, or by clicking here.

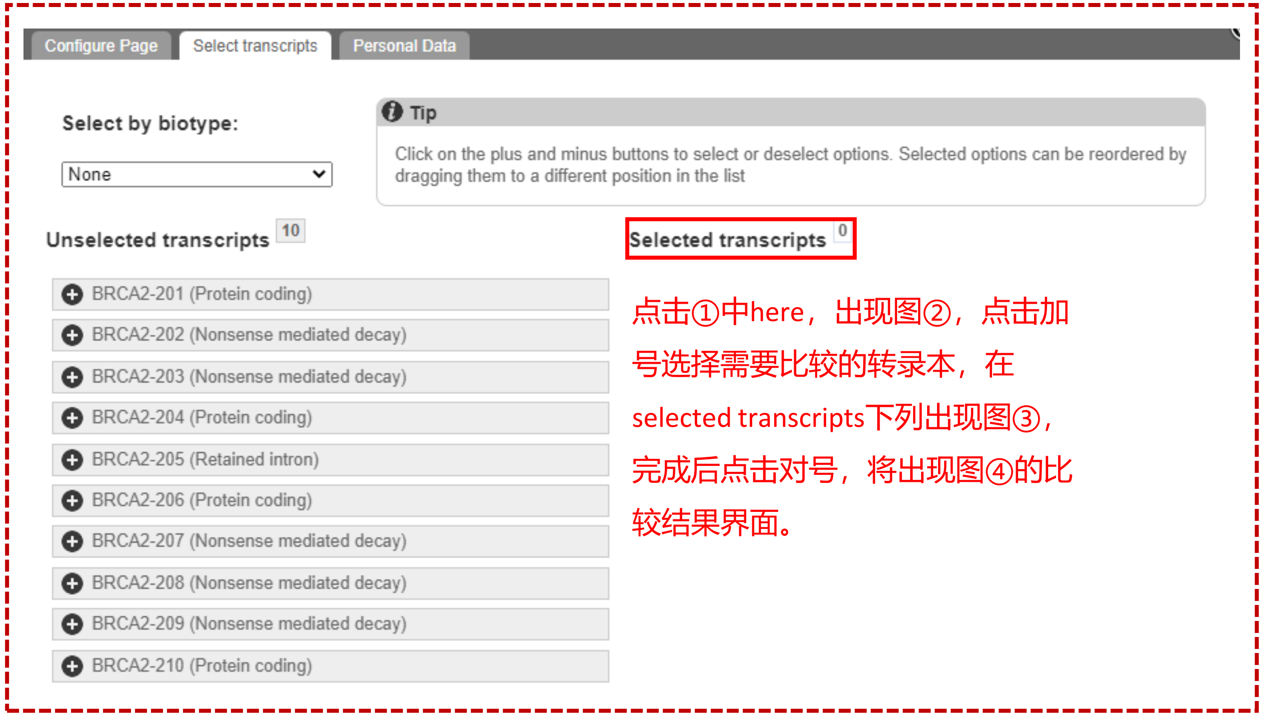
③



Selected transcripts 4

- BRCA2-201 (Protein coding)
- BRCA2-203 (Nonsense mediated decay)
- BRCA2-205 (Retained intron)
- BRCA2-207 (Nonsense mediated decay)

②



Select by biotype: None

Tip

Click on the plus and minus buttons to select or deselect options. Selected options can be reordered by dragging them to a different position in the list

Unselected transcripts 10

- BRCA2-201 (Protein coding)
- BRCA2-202 (Nonsense mediated decay)
- BRCA2-203 (Nonsense mediated decay)
- BRCA2-204 (Protein coding)
- BRCA2-205 (Retained intron)
- BRCA2-206 (Protein coding)
- BRCA2-207 (Nonsense mediated decay)
- BRCA2-208 (Nonsense mediated decay)
- BRCA2-209 (Nonsense mediated decay)
- BRCA2-210 (Protein coding)

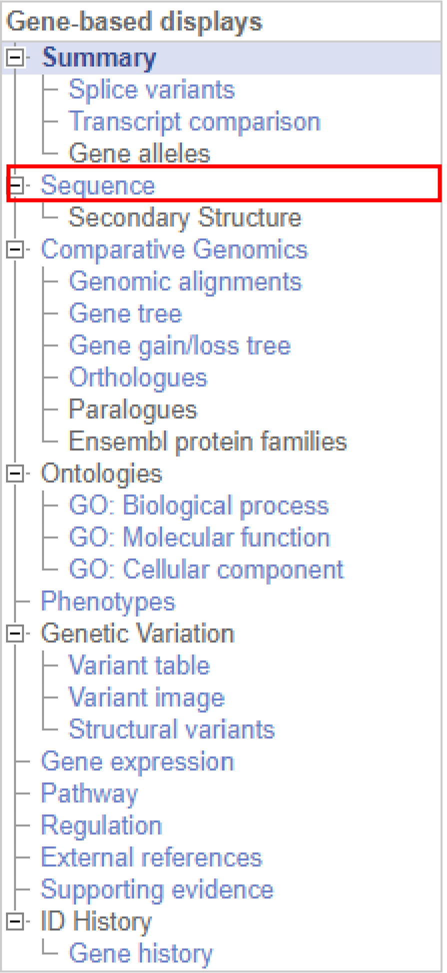
Selected transcripts 0

点击①中here，出现图②，点击加号选择需要比较的转录本，在selected transcripts下列出现图③，完成后点击对号，将出现图④的比较结果界面。

④

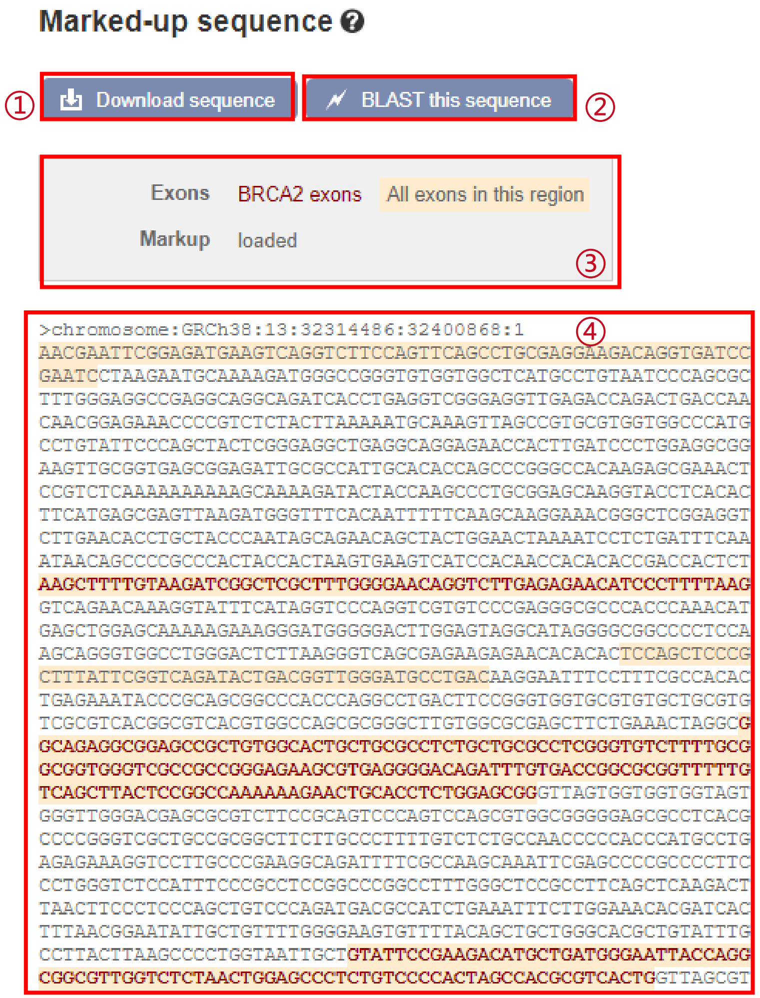
Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击sequence，其右侧内容展示区将如下图所示。



Gene-based displays

- **Summary**
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history



Marked-up sequence

Exons BRCA2 exons All exons in this region

Markup loaded

```
>chromosome:GRCh38:13:32314486:32400868:1
AACGAATTCGGAGATGAAGTCAGGTCTTCCAGTTCAGCCTGCGAGGAAGACAGGTGATCC
GAATCCTAAGAATGCAAAAGATGGGCCGGGTGTGGTGGCTCATGCCTGTAATCCCAGCGC
TTTGGGAGGCCGAGGCAGGCAGATCACCTGAGGTCGGGAGGTTGAGACCAGACTGACCAA
CAACGGAGAAACCCCGTCTCTACTTAAAAATGCAAAGTTAGCCGTGCGTGGTGGCCCATG
CCTGTATTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAACCACTTGATCCCTGGAGGCGG
AAGTTGCGGTGAGCGGAGATTGCGCCATTGCACACCAGCCCGGGCCACAAGAGCGAAACT
CCGTCTCAAAAAAAAAAGCAAAAGATACTACCAAGCCCTGCGGAGCAAGGTACCTCACAC
TTCATGAGCGAGTTAAGATGGGTTTCACAATTTTTCAAGCAAGGAAACGGGCTCGGAGGT
CTTGAACACCTGCTACCCAATAGCAGAACAGCTACTGGAACTAAAATCCTCTGATTTCAA
ATAACAGCCCCGCCCACTACCACTAAGTGAAGTCATCCACAACCACACACCGACCACTCT
AAGCTTTTGTAAGATCGGCTCGCTTTGGGGAACAGGTCTTGAGAGAACATCCCTTTTAAG
GTCAGAACAAAGGTATTTCATAGGTCCCAGGTCGTGTCCCGAGGGCGCCCACCCAAACAT
GAGCTGGAGCAAAAAGAAAGGGATGGGGGACTTGGAGTAGGCATAGGGGCGGCCCCTCCA
AGCAGGGTGGCCTGGGACTCTTAAGGGTCAGCGAGAAGAGAACACACACTCCAGCTCCCG
CTTTATTCGGTCAGATACTGACGGTTGGGATGCCTGACAAGGAATTTCCTTTCGCCACAC
TGAGAAATACCCGCAGCGGCCCACCCAGGCCTGACTTCCGGGTGGTGCGTGTGCTGCGTG
TCGCGTCACGGCGTCACGTGGCCAGCGCGGGCTTGTGGCGCGAGCTTCTGAAACTAGGCG
GCAGAGGCGGAGCCGCTGTGGCACTGCTGCGCCTCTGCTGCGCCTCGGGTGTCTTTTGCG
GCGGTGGGTCGCCGCCGGGAGAAGCGTGAGGGGACAGATTTGTGACCGGCGCGGTTTTTG
TCAGCTTACTCCGGCCAAAAAAGAACTGCACCTCTGGAGCGGGTTAGTGGTGGTGGTAGT
GGGTTGGGACGAGCGCGTCTTCCGCAGTCCCAGTCCAGCGTGGCGGGGGAGCGCCTCACG
CCCCGGGTCGCTGCCGCGGCTTCTTGCCCTTTTGTCTCTGCCAACCCCCACCCATGCCTG
AGAGAAAGGTCCTTGCCCGAAGGCAGATTTTCGCCAAGCAAATTCGAGCCCCGCCCCTTC
CCTGGGTCTCCATTTCCCGCCTCCGGCCCGGCCTTTGGGCTCCGCCTTCAGCTCAAGACT
TAACTTCCCTCCCAGCTGTCCCAGATGACGCCATCTGAAATTTCTTGGAAACACGATCAC
TTTAACGGAATATTGCTGTTTTGGGGAAGTGTTTTACAGCTGCTGGGCACGCTGTATTTG
CCTTACTTAAGCCCCTGGTAATTGCTGTATTCCGAAGACATGCTGATGGGAATTACCAGG
CGGCGTTGGTCTCTAACTGGAGCCCTCTGTCCCCACTAGCCACGCGTCACTGGTTAGCGT
```

①点击下载序列,点击后出现图⑤界面，见下张PPT

②通过Blast将该序列与数据库中其他序列进行比对，点击后出现图⑥界面，见下张PPT

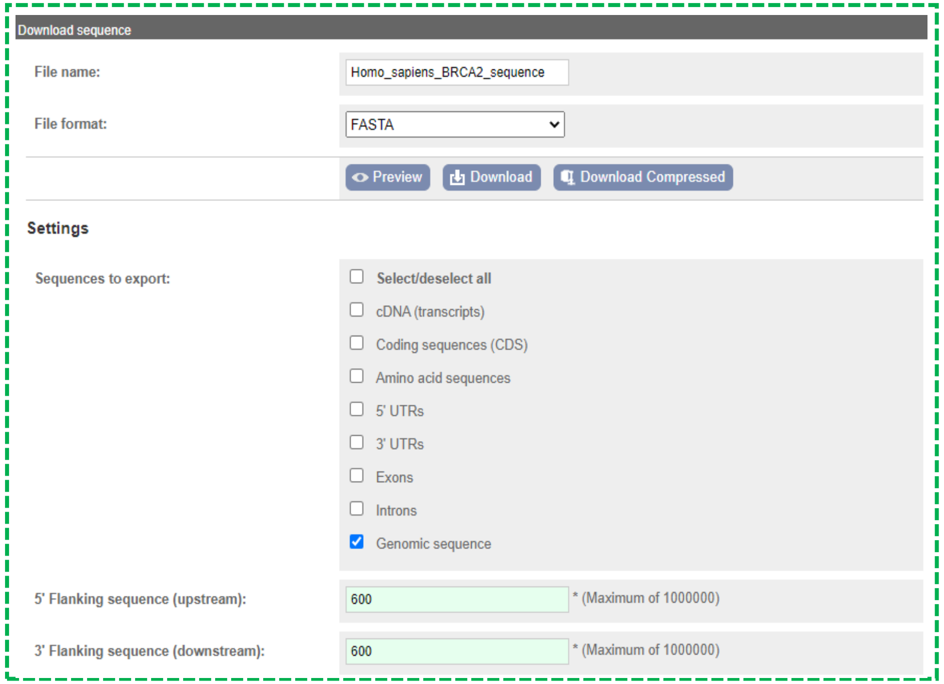
③外显子信息：紫红色为BRCA2外显子，黄底为该区域所有外显子信息

④序列信息

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击sequence，其右侧内容展示区将如下图所示——接上页PPT。

⑤



Download sequence

File name: Homo_sapiens_BRCA2_sequence

File format: FASTA

Settings

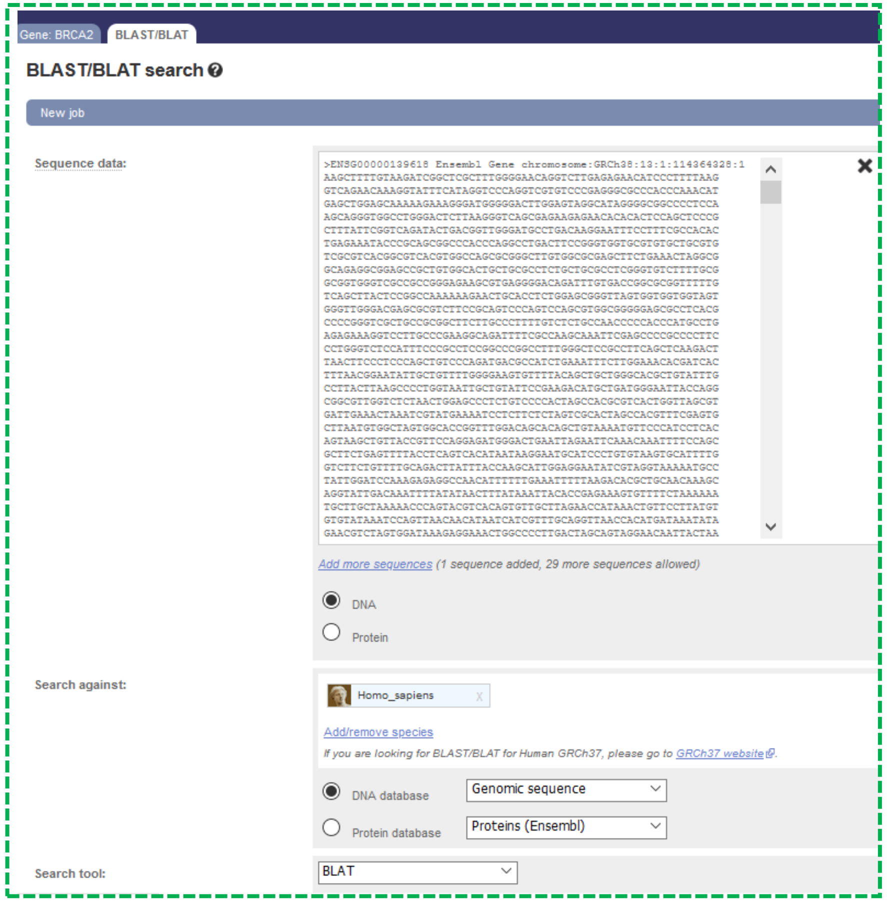
Sequences to export:

- ☐ Select/deselect all
- ☐ cDNA (transcripts)
- ☐ Coding sequences (CDS)
- ☐ Amino acid sequences
- ☐ 5' UTRs
- ☐ 3' UTRs
- ☐ Exons
- ☐ Introns
- ☑ Genomic sequence

5' Flanking sequence (upstream): 600 * (Maximum of 1000000)

3' Flanking sequence (downstream): 600 * (Maximum of 1000000)

⑥



BLAST/BLAT search

Sequence data:

Add more sequences (1 sequence added, 29 more sequences allowed)

● DNA
○ Protein

Search against: Homo_sapiens

Add/remove species

If you are looking for BLAST/BLAT for Human GRCh37, please go to GRCh37 website.

● DNA database: Genomic sequence
○ Protein database: Proteins (Ensembl)

Search tool: BLAT

图⑤界面：

- 文件名称（File name）
- 文件格式（File format）
- 设置（Settings）
- 5’ 端侧翼序列（5’ flanking sequence）
- 3’ 端侧翼序列（3’ flanking sequence）

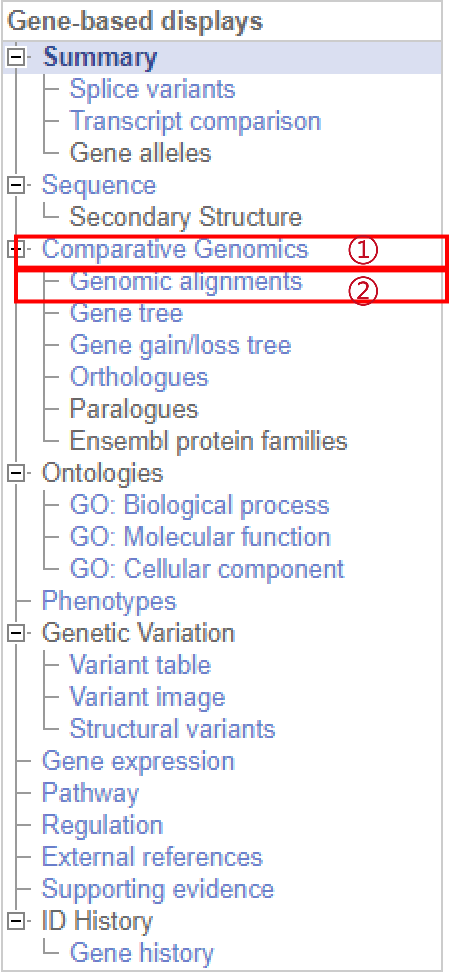
图⑥界面——常见的Blast界面

- 序列详细内容（sequence data）
- 搜索范围 （Search against）
- 搜索工具 （Search tool）

点击该界面下方的RUN按钮就可以进行该序列的比对。

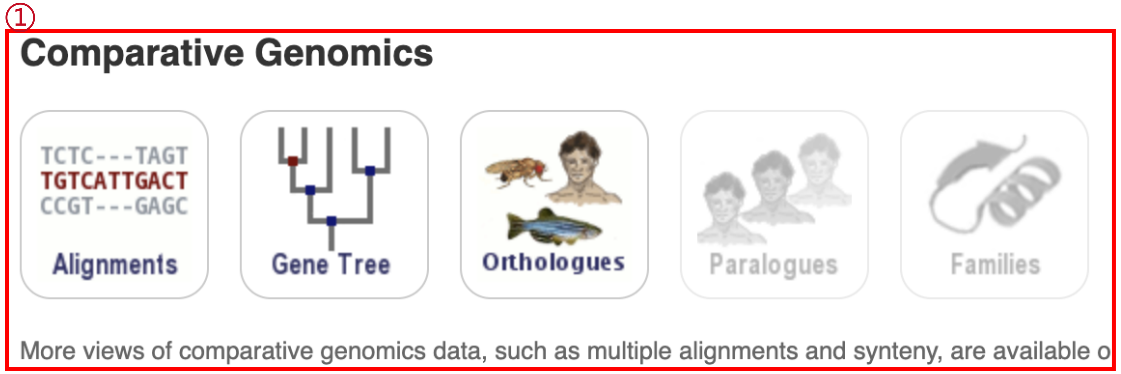
Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击Comparative genomics，其右侧内容展示区将如下图所示。



Gene-based displays
- Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics ①
  - Genomic alignments ②
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history

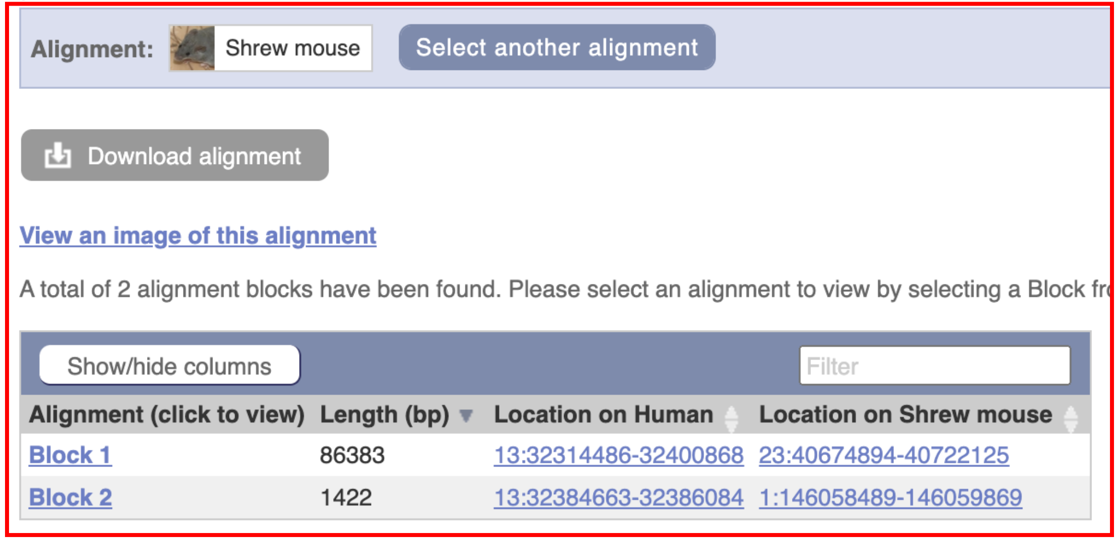
①



Comparative Genomics

More views of comparative genomics data, such as multiple alignments and synteny, are available o

②



View an image of this alignment

A total of 2 alignment blocks have been found. Please select an alignment to view by selecting a Block fr

Alignment (click to view)	Length (bp)	Location on Human	Location on Shrew mouse
Block 1	86383	13:32314486-32400868	23:40674894-40722125
Block 2	1422	13:32384663-32386084	1:146058489-146059869

①点击左侧Comparative genomics 出现此框信息

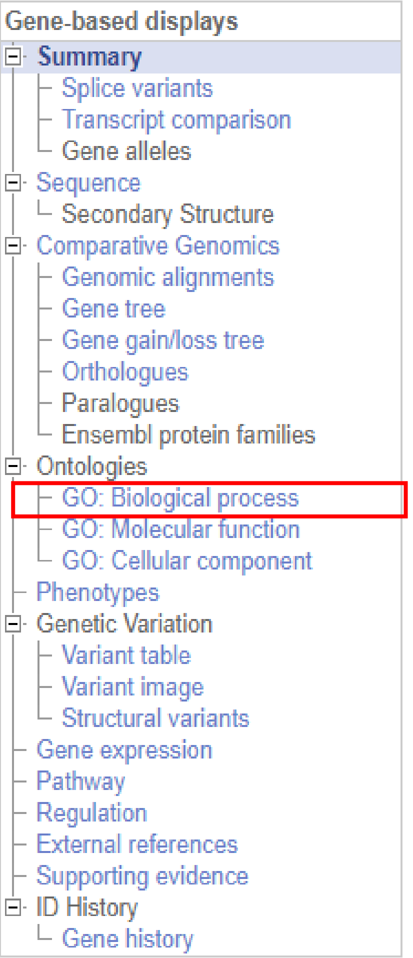
- 序列对比（Alignments)
- 基因树（Gene tree)
- 不同物种间的同源序列（orthologues)
- 旁系同源（Paralogues)
- 家系（families)

②点击左侧导航栏中Alignment进入该界面

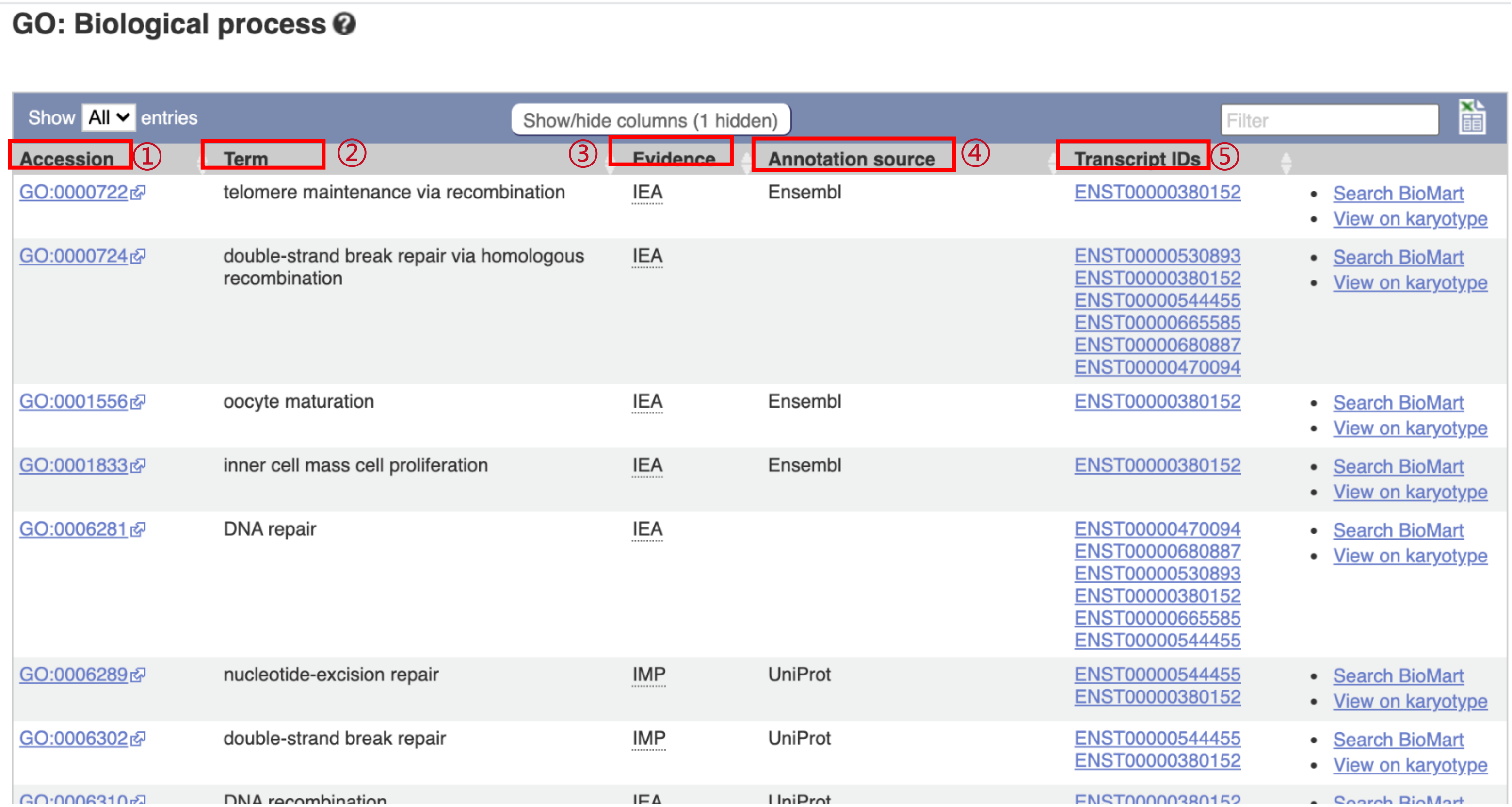
- 点击alignment后需要根据个人需求选择比较物种
- 结果信息包含发现的相似序列数量、长度、在选择的物种一中定位，在比对物种中的定位信息

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击GO: biological process，其右侧内容展示区将如下图所示。



Gene-based displays
- Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history



GO: Biological process

Show All entries | Show/hide columns (1 hidden) | Filter

Accession ①	Term ②	③ Evidence	Annotation source ④	Transcript IDs ⑤	
GO:0000722	telomere maintenance via recombination	IEA	Ensembl	ENST00000380152	Search BioMart View on karyotype
GO:0000724	double-strand break repair via homologous recombination	IEA		ENST00000530893 ENST00000380152 ENST00000544455 ENST00000665585 ENST00000680887 ENST00000470094	Search BioMart View on karyotype
GO:0001556	oocyte maturation	IEA	Ensembl	ENST00000380152	Search BioMart View on karyotype
GO:0001833	inner cell mass cell proliferation	IEA	Ensembl	ENST00000380152	Search BioMart View on karyotype
GO:0006281	DNA repair	IEA		ENST00000470094 ENST00000680887 ENST00000530893 ENST00000380152 ENST00000665585 ENST00000544455	Search BioMart View on karyotype
GO:0006289	nucleotide-excision repair	IMP	UniProt	ENST00000544455 ENST00000380152	Search BioMart View on karyotype
GO:0006302	double-strand break repair	IMP	UniProt	ENST00000544455 ENST00000380152	Search BioMart View on karyotype
GO:0006310	DNA recombination	IEA	UniProt	ENST00000380152	Search BioMart

①GO序列号，点击可快速查看详细信息

②具体生物过程名称

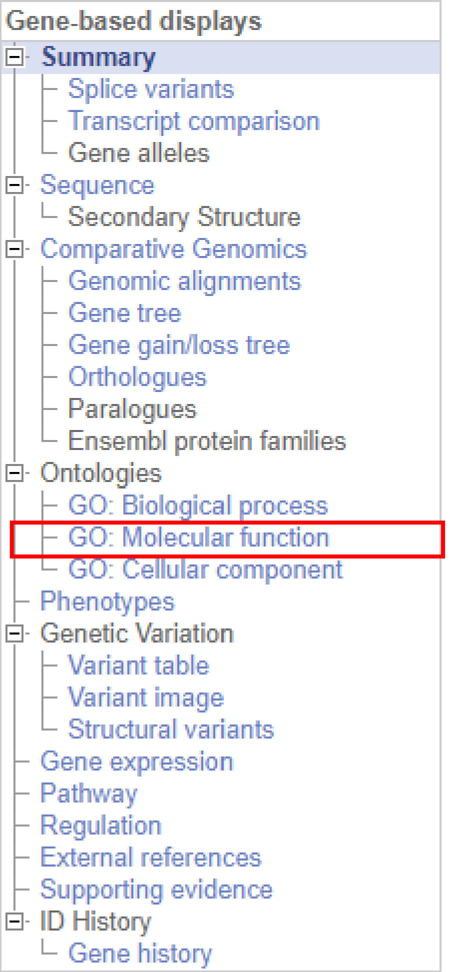
③证据来源：EXP来源于推测，IMP：来源于突变表型，鼠标放至名称上即可显示证据来源

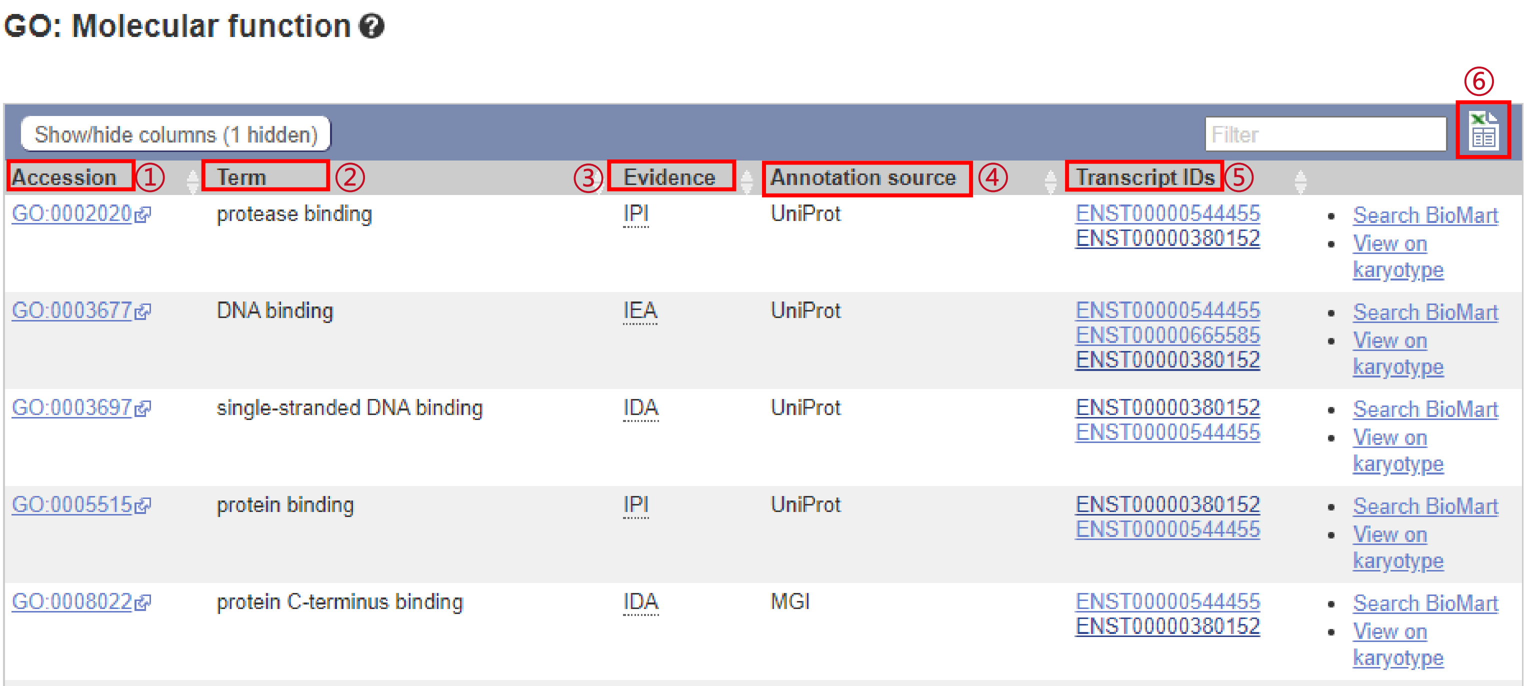
④注释信息来源

⑤转录本ID

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击GO：Molecular function，其右侧内容展示区将如下图所示。





GO: Molecular function

Accession ①	Term ②	③ Evidence	Annotation source ④	Transcript IDs ⑤	
GO:0002020	protease binding	IPI	UniProt	ENST00000544455 ENST00000380152	Search BioMart View on karyotype
GO:0003677	DNA binding	IEA	UniProt	ENST00000544455 ENST00000665585 ENST00000380152	Search BioMart View on karyotype
GO:0003697	single-stranded DNA binding	IDA	UniProt	ENST00000380152 ENST00000544455	Search BioMart View on karyotype
GO:0005515	protein binding	IPI	UniProt	ENST00000380152 ENST00000544455	Search BioMart View on karyotype
GO:0008022	protein C-terminus binding	IDA	MGI	ENST00000544455 ENST00000380152	Search BioMart View on karyotype

①GO序列号，点击可快速查看详细信息

②具体分子功能名

③证据来源：

EXP来源于实验，IDA：通过直接分析推测，鼠标放至名称上即可显示证据来源

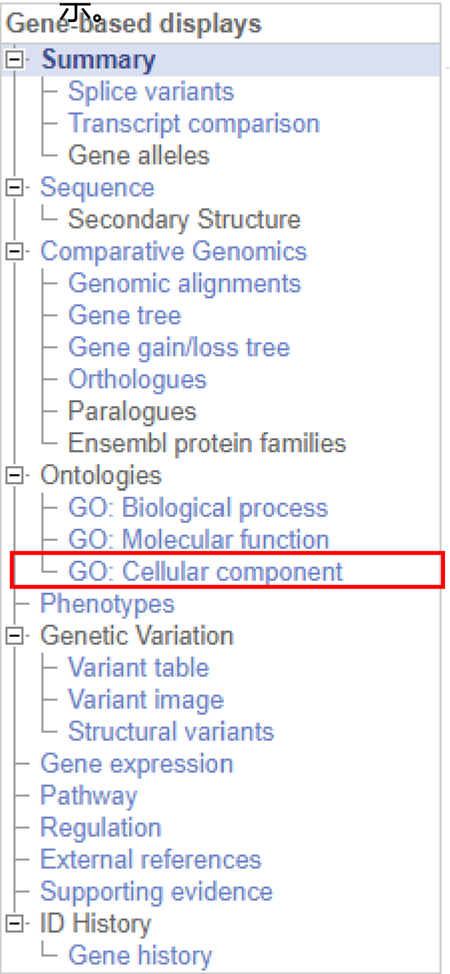
④注释信息来源

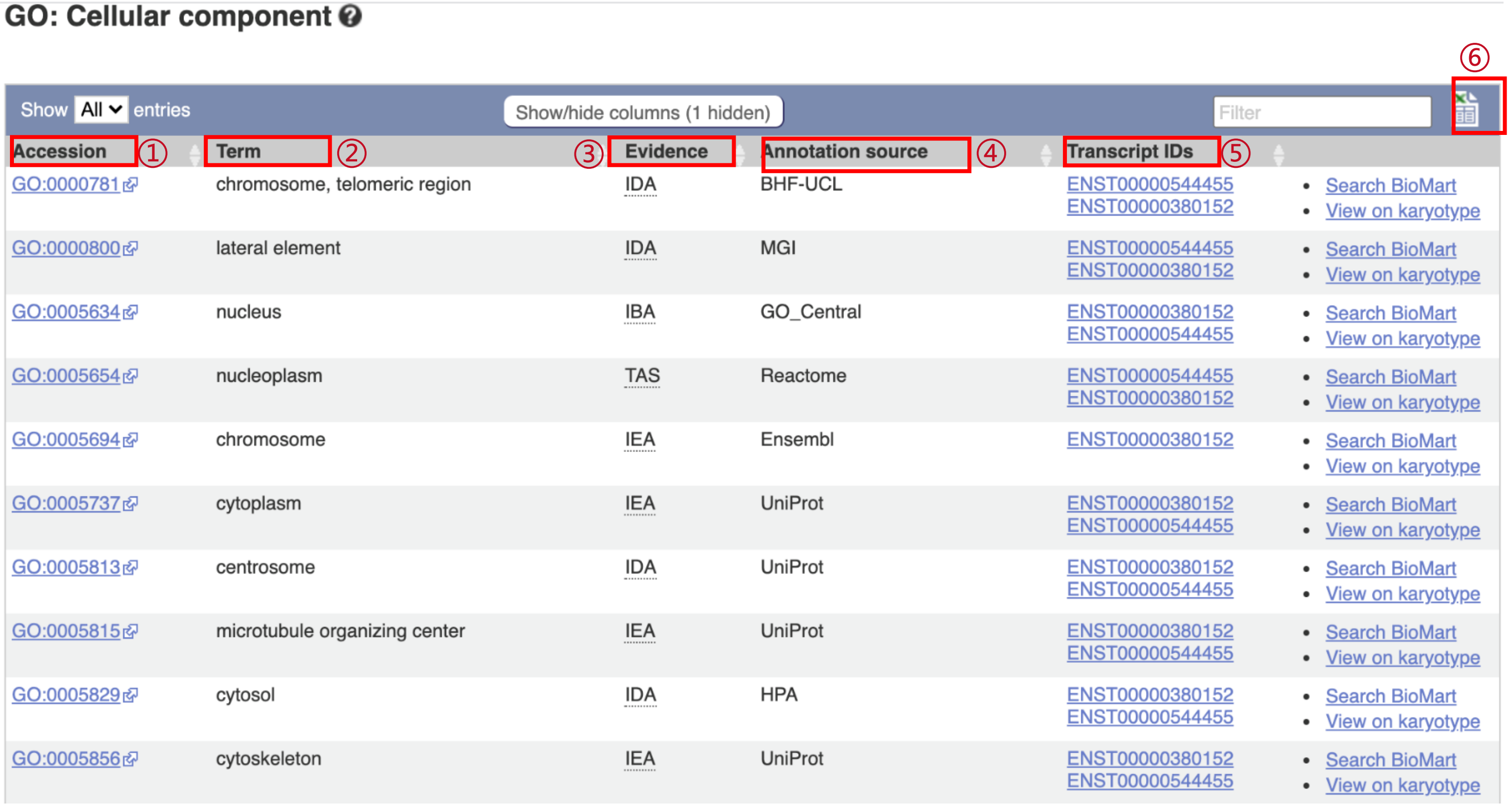
⑤转录本ID

⑥下载

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击GO：Cellular component，其右侧内容展示区将如下图所示。





Accession	Term	Evidence	Annotation source	Transcript IDs	
GO:0000781	chromosome, telomeric region	IDA	BHF-UCL	ENST00000544455 ENST00000380152	Search BioMart View on karyotype
GO:0000800	lateral element	IDA	MGI	ENST00000544455 ENST00000380152	Search BioMart View on karyotype
GO:0005634	nucleus	IBA	GO_Central	ENST00000380152 ENST00000544455	Search BioMart View on karyotype
GO:0005654	nucleoplasm	TAS	Reactome	ENST00000544455 ENST00000380152	Search BioMart View on karyotype
GO:0005694	chromosome	IEA	Ensembl	ENST00000380152	Search BioMart View on karyotype
GO:0005737	cytoplasm	IEA	UniProt	ENST00000380152 ENST00000544455	Search BioMart View on karyotype
GO:0005813	centrosome	IDA	UniProt	ENST00000380152 ENST00000544455	Search BioMart View on karyotype
GO:0005815	microtubule organizing center	IEA	UniProt	ENST00000380152 ENST00000544455	Search BioMart View on karyotype
GO:0005829	cytosol	IDA	HPA	ENST00000380152 ENST00000544455	Search BioMart View on karyotype
GO:0005856	cytoskeleton	IEA	UniProt	ENST00000380152 ENST00000544455	Search BioMart View on karyotype

①GO序列号，点击可快速查看详细信息

②具体细胞组分名

③证据来源：

IDA：通过直接分析推测，鼠标放至名称上即可显示证据来源

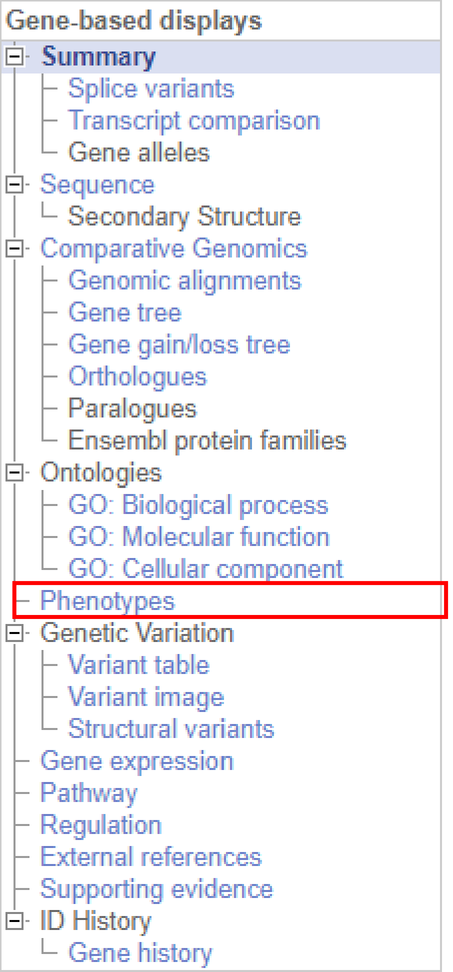
④注释信息来源

⑤转录本ID

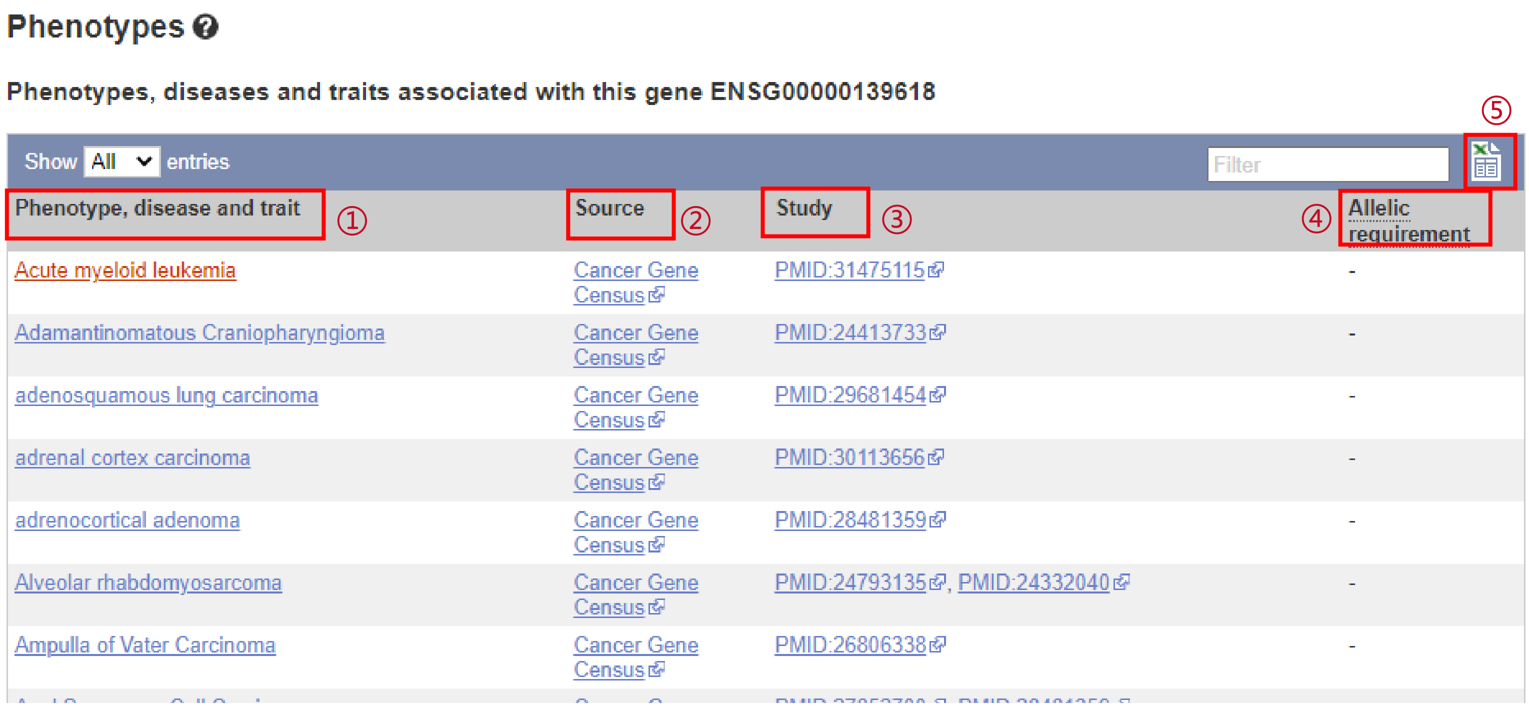
⑥下载

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击Phenotypes，其右侧内容展示区将如下图所示。



Gene-based displays
- Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history



Phenotypes

Phenotypes, diseases and traits associated with this gene ENSG00000139618

Phenotype, disease and trait ①	Source ②	Study ③	④ Allelic requirement
Acute myeloid leukemia	Cancer Gene Census	PMID:31475115	-
Adamantinomatous Craniopharyngioma	Cancer Gene Census	PMID:24413733	-
adenosquamous lung carcinoma	Cancer Gene Census	PMID:29681454	-
adrenal cortex carcinoma	Cancer Gene Census	PMID:30113656	-
adrenocortical adenoma	Cancer Gene Census	PMID:28481359	-
Alveolar rhabdomyosarcoma	Cancer Gene Census	PMID:24793135, PMID:24332040	-
Ampulla of Vater Carcinoma	Cancer Gene Census	PMID:26806338	-

①表型、疾病：可根据自己研究方向选择进一步查看详细内容

②来源

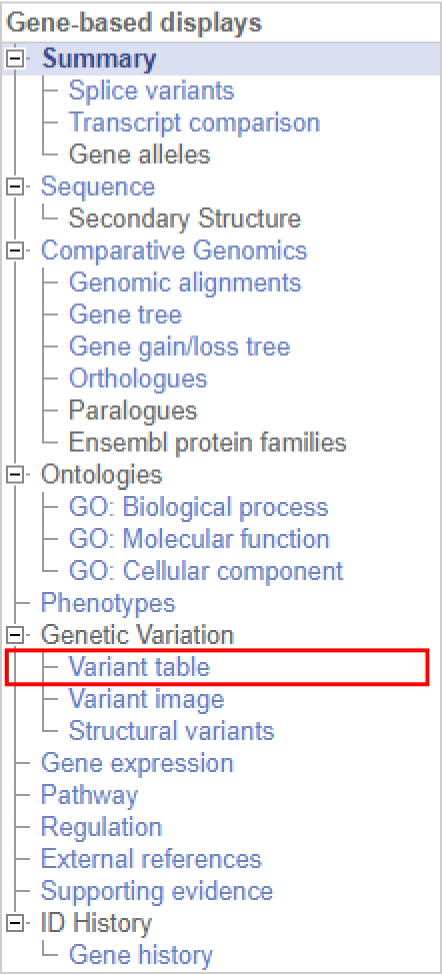
③最开始证实疾病相关的文章链接

④与疾病相关的等位基因状态

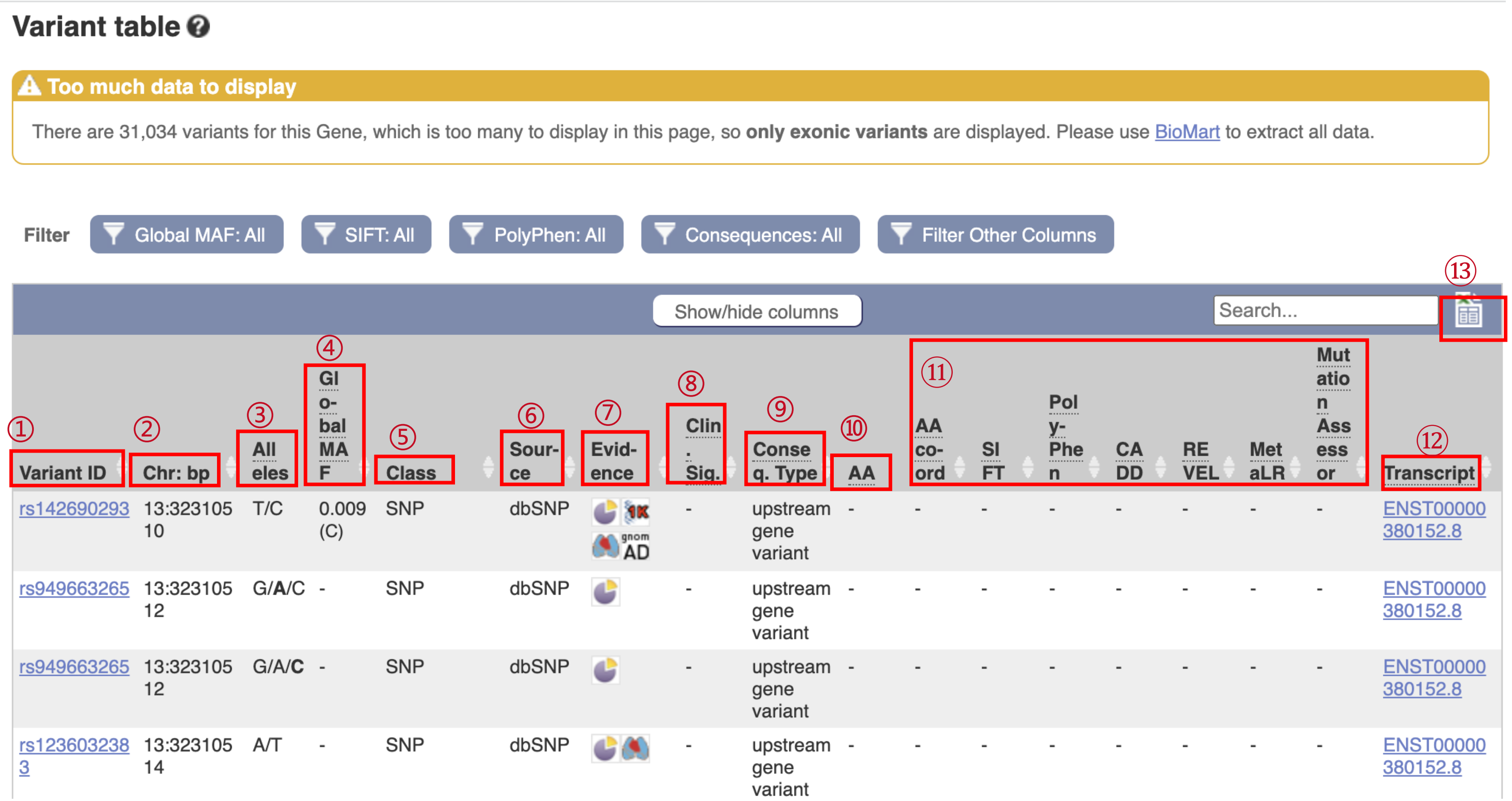
⑤下载

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击variant table，其右侧内容展示区将如下图所示。



Gene-based displays
- Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history



Variant table

Too much data to display

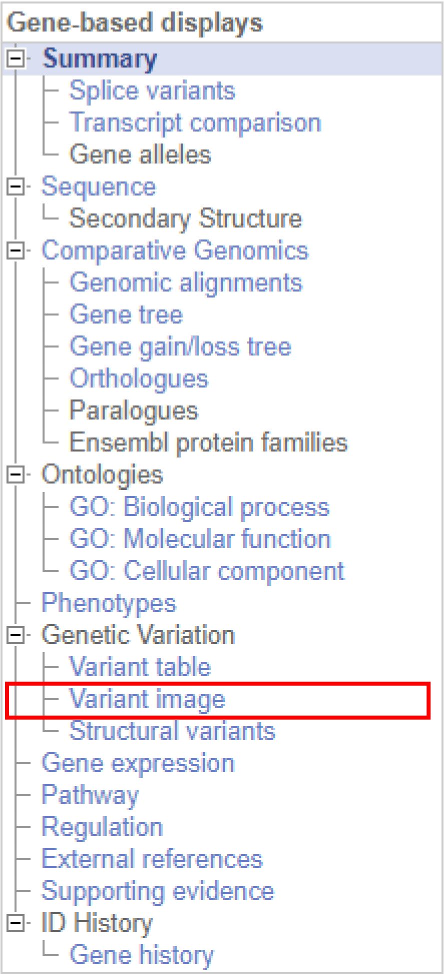
There are 31,034 variants for this Gene, which is too many to display in this page, so **only exonic variants** are displayed. Please use BioMart to extract all data.

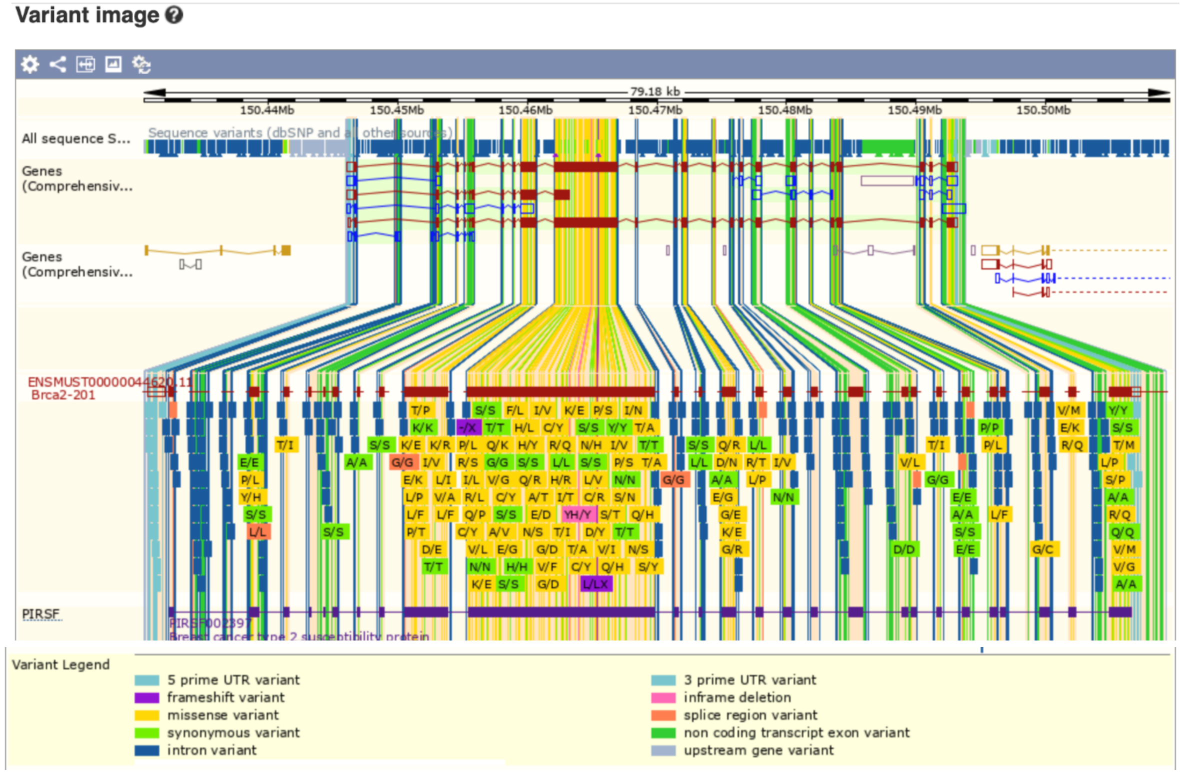
Variant ID	Chr: bp	Alleles	Global MAF	Class	Source	Evidence	Clin. Sig.	Conseq. Type	AA	AA co-ord	SIFT	PolyPhen	CADD	REVEL	MetaLR	Mutation Assessor	Transcript
rs142690293	13:32310510	T/C	0.009 (C)	SNP	dbSNP	1K gnomAD	-	upstream gene variant	-	-	-	-	-	-	-	-	ENST00000380152.8
rs949663265	13:32310512	G/A/C	-	SNP	dbSNP		-	upstream gene variant	-	-	-	-	-	-	-	-	ENST00000380152.8
rs949663265	13:32310512	G/A/**C**	-	SNP	dbSNP		-	upstream gene variant	-	-	-	-	-	-	-	-	ENST00000380152.8
rs1236032383	13:32310514	A/T	-	SNP	dbSNP		-	upstream gene variant	-	-	-	-	-	-	-	-	ENST00000380152.8

①变异ID号
②染色体定位
③等位基因
④第二位常见突变等位基因突变频率
⑤类别
⑥来源
⑦证据等级
⑧临床意义
⑨结果类型
⑩氨基酸残基数
⑪预测突变与功能的链接，点击可进入
⑫转录本ID
⑬下载

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击variant image，其右侧内容展示区将如下图所示。

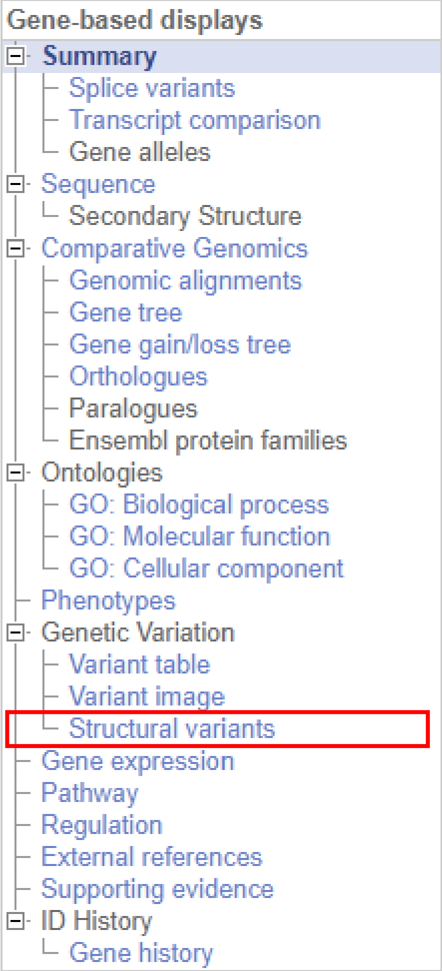




- 目前对于人类物种不展示变异图片，其他物种不受影响，接下来以mouse BRCA2基因为例
- 每个转录本的基因结构，外显子由框表示，连接框的线为内含子，填充框反应编码序列，未填充框反应UTRs（未翻译区）
- 蓝色：5’或3’UTR变异
- 紫色：移码突变
- 粉色：框内缺失突变
- 黄色：错义突变
- 橙色：可变剪切区域
- 浅绿：同义突变
- 深绿：非编码转录外显子突变
- 蓝色：内含子突变
- 淡紫色：上游基因变异

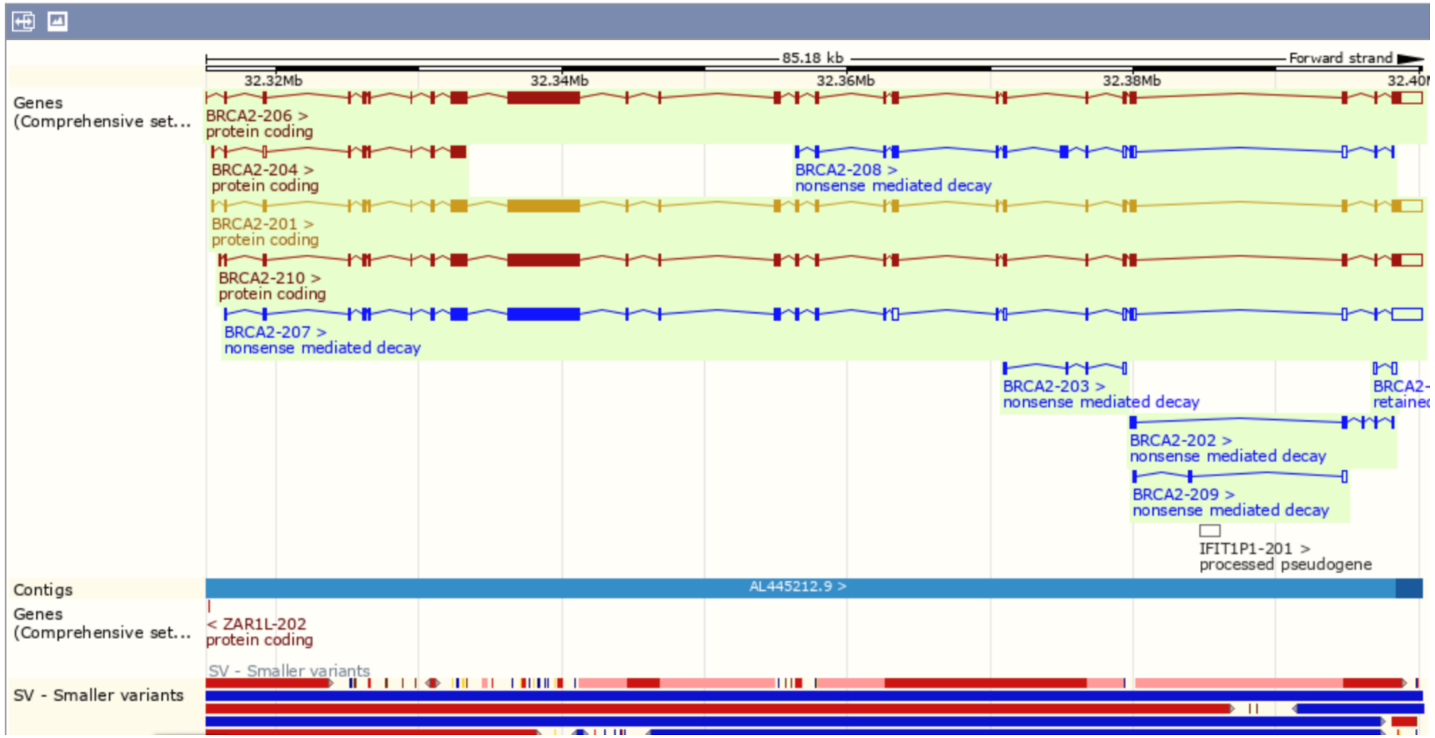
Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击Structural variants，其右侧内容展示区将如下图所示。



Gene-based displays
- Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history

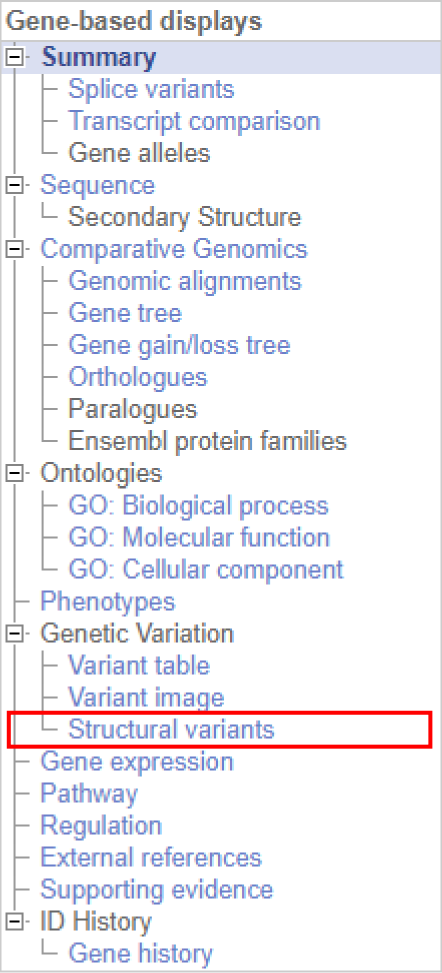
Structural variants

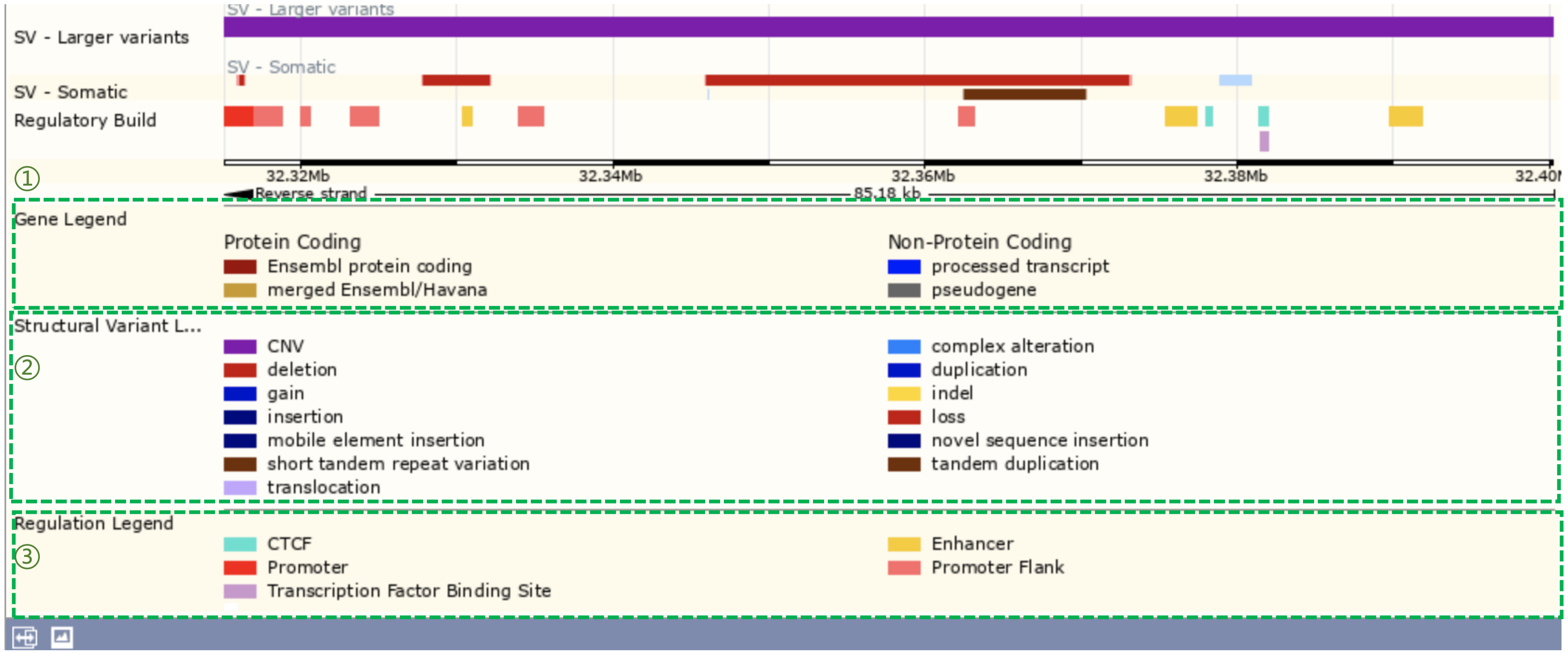


- 每个转录本的基因结构，外显子由框表示，连接框的线为内含子，填充框反应编码序列，未填充框反应UTRs（未翻译区）

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击Structural variants,其右侧内容展示区将如下图所示,接上页PPT。





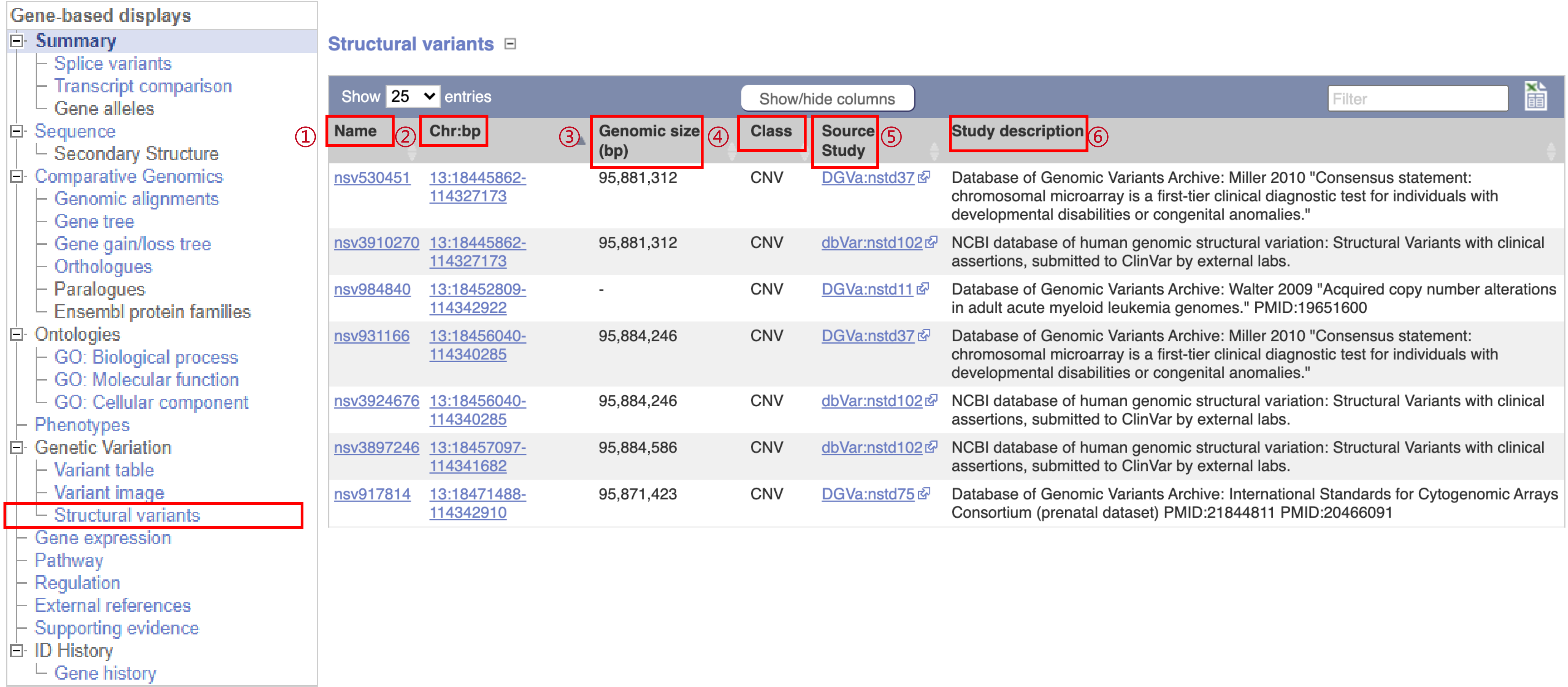
- ①酒红色：转录本是蛋白质编码，来自Ensembl或VEGA/Havana
- 黄色：转录本是Ensembl与VEGA/Havana协商而定
- 蓝色灰色：转录本属于非编码区

- ②紫色：CNV 浅蓝色：复杂变异
- 酒红：缺失 蓝色：重复
- 深蓝：获得 黄色：插入缺失
- 深蓝：插入 红色：缺失
- 藏青：移动元素插入 新序列插入
- 棕色：短串联重复变异 串联重复变异 藕粉色：易位

- ③浅蓝：CTCF区 黄色：增强子
- 红：启动子 浅红：启动子侧翼序列
- 藕色：转录因子结合位点

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击structual variants，其右侧内容展示区将如下图所示:接上页PPT。



Gene-based displays
- Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history

Structural variants

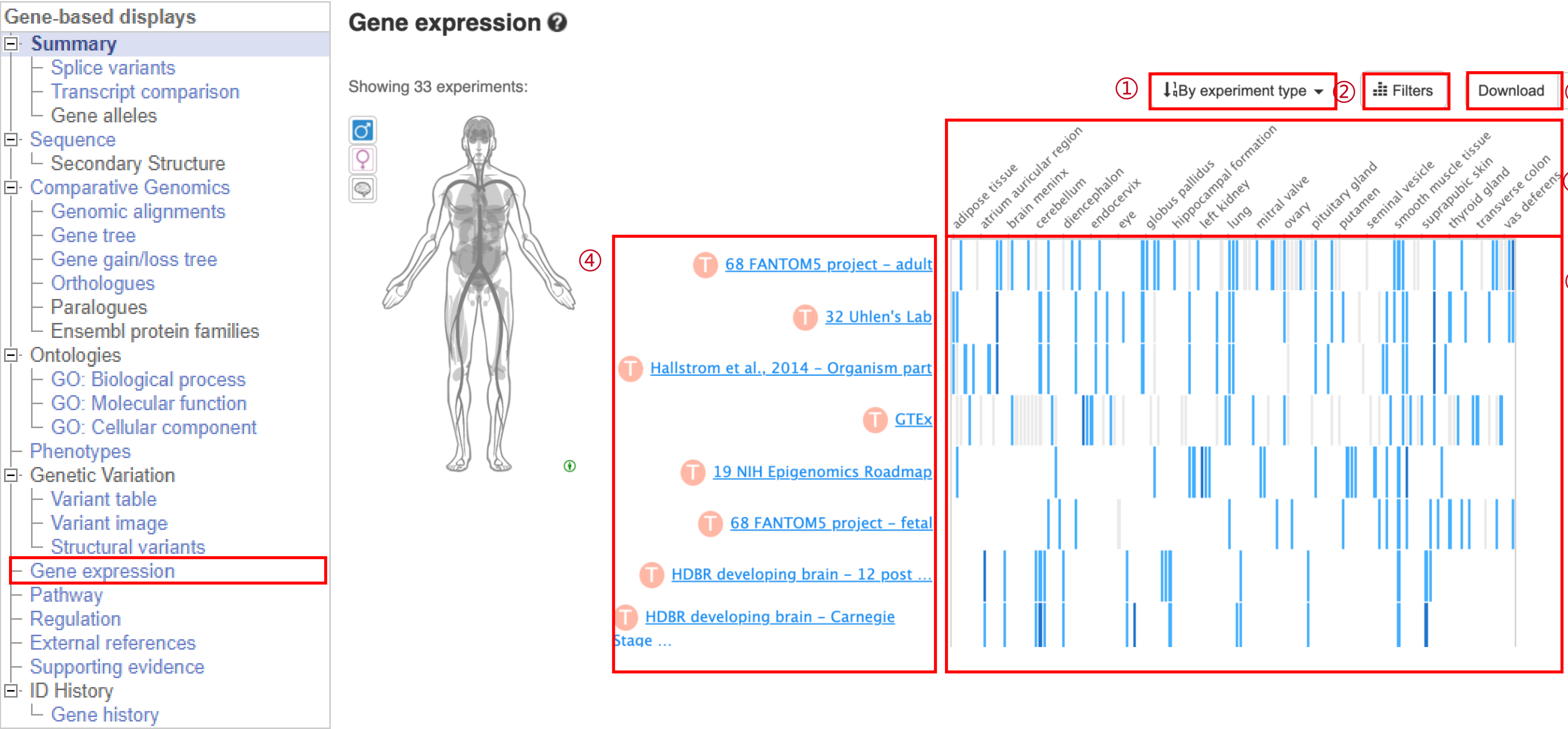
Show 25 entries　Show/hide columns　Filter

① Name	② Chr:bp	③ Genomic size (bp)	④ Class	⑤ Source Study	⑥ Study description
nsv530451	13:18445862-114327173	95,881,312	CNV	DGVa:nstd37	Database of Genomic Variants Archive: Miller 2010 "Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies."
nsv3910270	13:18445862-114327173	95,881,312	CNV	dbVar:nstd102	NCBI database of human genomic structural variation: Structural Variants with clinical assertions, submitted to ClinVar by external labs.
nsv984840	13:18452809-114342922	-	CNV	DGVa:nstd11	Database of Genomic Variants Archive: Walter 2009 "Acquired copy number alterations in adult acute myeloid leukemia genomes." PMID:19651600
nsv931166	13:18456040-114340285	95,884,246	CNV	DGVa:nstd37	Database of Genomic Variants Archive: Miller 2010 "Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies."
nsv3924676	13:18456040-114340285	95,884,246	CNV	dbVar:nstd102	NCBI database of human genomic structural variation: Structural Variants with clinical assertions, submitted to ClinVar by external labs.
nsv3897246	13:18457097-114341682	95,884,586	CNV	dbVar:nstd102	NCBI database of human genomic structural variation: Structural Variants with clinical assertions, submitted to ClinVar by external labs.
nsv917814	13:18471488-114342910	95,871,423	CNV	DGVa:nstd75	Database of Genomic Variants Archive: International Standards for Cytogenomic Arrays Consortium (prenatal dataset) PMID:21844811 PMID:20466091

①名称
②染色体定位
③基因组大小
④类别
⑤来源
⑥研究描述

Ensembl检索方法---关键词检索

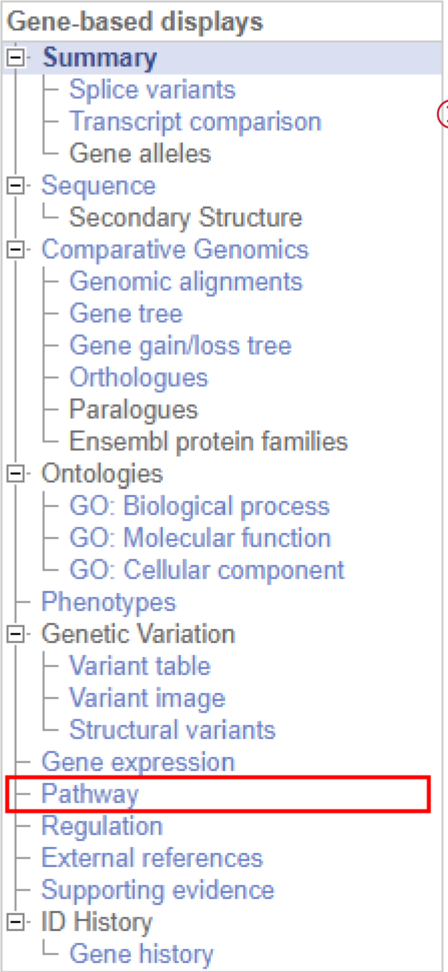
- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击Gene expression，其右侧内容展示区将如下图所示。



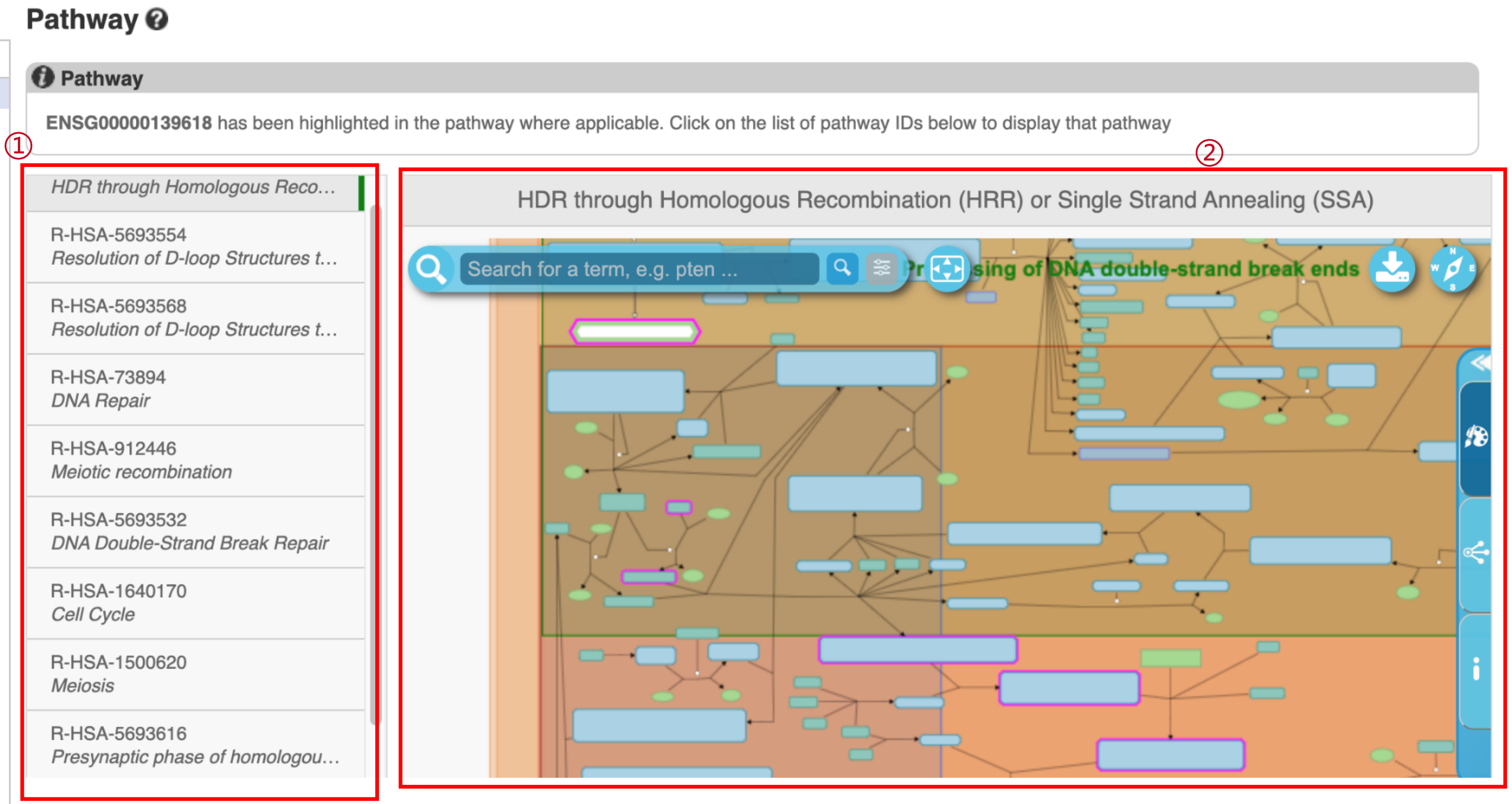
①选择排序方式
②根据组织器官过滤显示项目
③下载
④显示的试验来源
⑤不同组织器官
⑥不同组织器官表达情况
P代表蛋白组学数据
T代表转录组数据
深蓝:高表达
蓝色：中表达
浅蓝：低表达
灰色：低于cutoff值
空白：无数据

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - ➢ 在左侧导航栏中点击Pathway，其右侧内容展示区将如下图所示。



Gene-based displays
- **Summary**
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history



Pathway

Pathway

ENSG00000139618 has been highlighted in the pathway where applicable. Click on the list of pathway IDs below to display that pathway

①

HDR through Homologous Reco…

R-HSA-5693554
Resolution of D-loop Structures t…

R-HSA-5693568
Resolution of D-loop Structures t…

R-HSA-73894
DNA Repair

R-HSA-912446
Meiotic recombination

R-HSA-5693532
DNA Double-Strand Break Repair

R-HSA-1640170
Cell Cycle

R-HSA-1500620
Meiosis

R-HSA-5693616
Presynaptic phase of homologou…

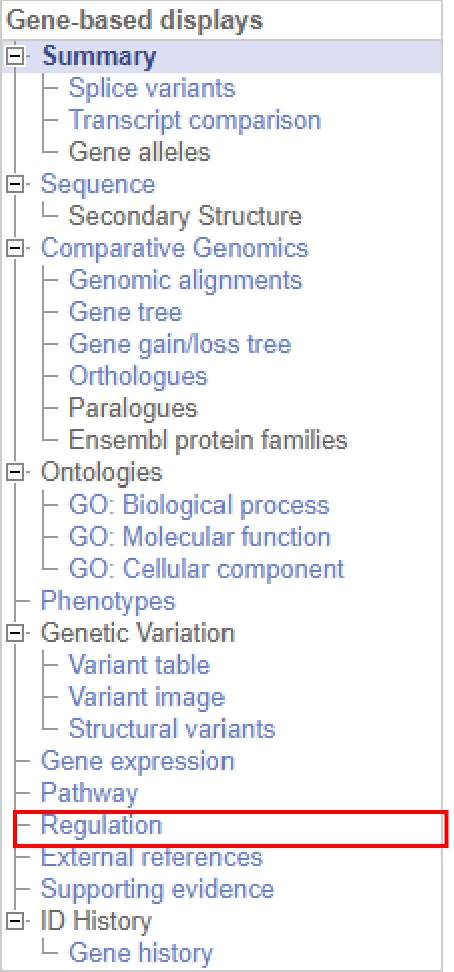
②

HDR through Homologous Recombination (HRR) or Single Strand Annealing (SSA)

①靶基因参与的信号通路
②可视化展现通路具体内容

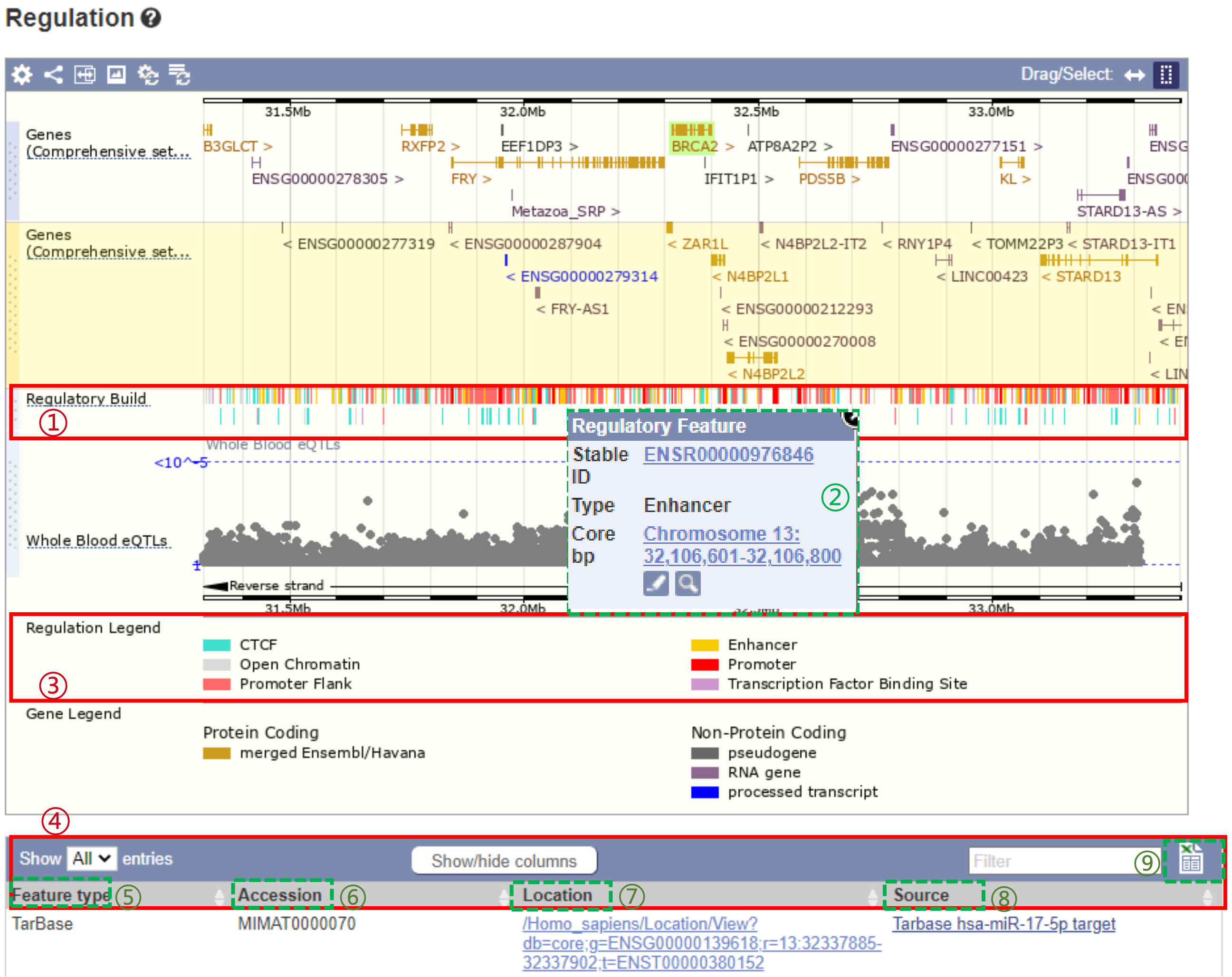
Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击regulation，其右侧内容展示区将如下图所示。



Gene-based displays

- ⊟ Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- ⊟ Sequence
  - Secondary Structure
- ⊟ Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- ⊟ Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- ⊟ Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ⊟ ID History
  - Gene history



Regulation

Regulatory Feature

Stable ID: ENSR00000976846

Type: Enhancer

Core bp: Chromosome 13: 32,106,601-32,106,800

②

Regulation Legend

- CTCF
- Open Chromatin
- Promoter Flank
- Enhancer
- Promoter
- Transcription Factor Binding Site

Gene Legend

Protein Coding
- merged Ensembl/Havana

Non-Protein Coding
- pseudogene
- RNA gene
- processed transcript

Show All entries　Show/hide columns　Filter

Feature type	Accession	Location	Source
TarBase	MIMAT0000070	/Homo_sapiens/Location/View?db=core;g=ENSG00000139618;r=13:32337885-32337902;t=ENST00000380152	Tarbase hsa-miR-17-5p target

①点击某竖条可显示具体信息见②

②详细信息

转录调节特征

固定ID

类型（Type）

核心染色体定位

③不同颜色代表不同含义

蓝色：CTCF区

红色：启动子侧翼序列

紫色：表观修饰序列

黄色：增强子

淡紫色：转录因子结合位点

④文字展示界面

⑤特征类型

⑥序列号

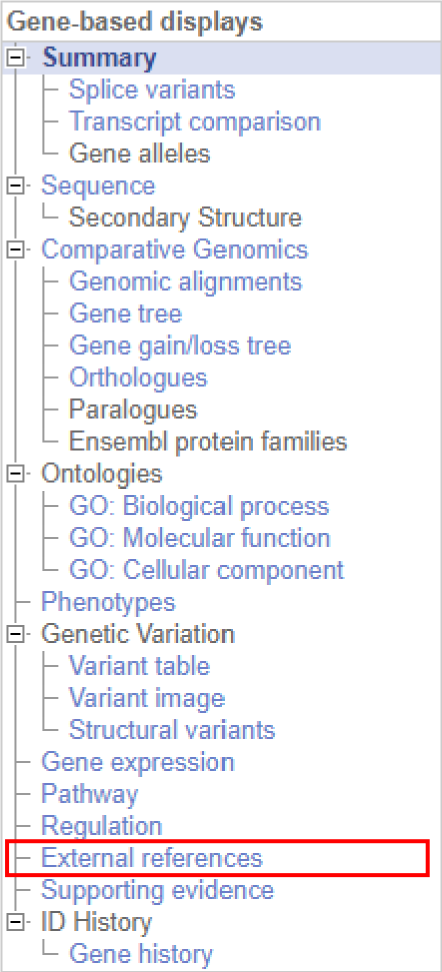
⑦染色体上位置

⑧来源

⑨下载数据

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - 在左侧导航栏中点击External references，其右侧内容展示区将如下图所示。



External references

This gene corresponds to the following database identifiers:

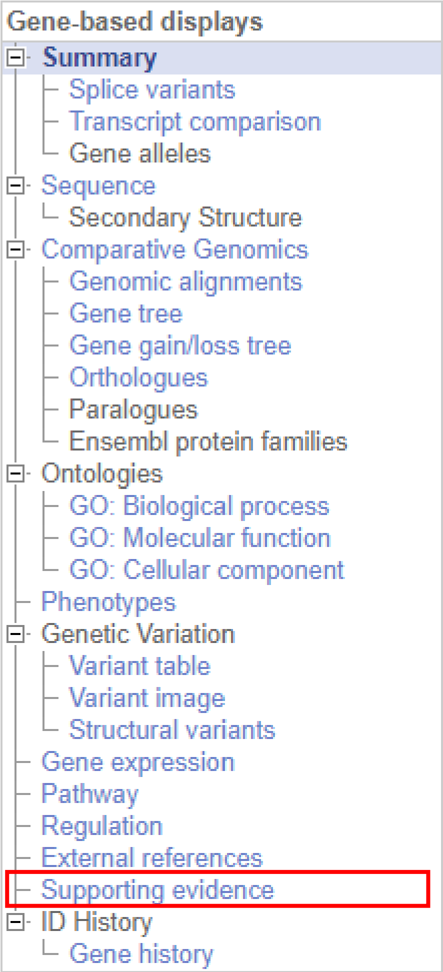
External database	Database identifier
Expression Atlas	ENSG00000139618 [view all locations]
HGNC Symbol	BRCA2 BRCA2 DNA repair associated [view all locations]
LRG	LRG_293 Locus Reference Genomic record for BRCA2 [view all locations]
MIM gene	BRCA2 GENE; BRCA2 [*600185] BRCA2 GENE; BRCA2;;FANCD1 GENE; FANCD1 [view all locations]
MIM morbid	BREAST CANCER [#114480] BREAST CANCER;;BREAST CANCER, FAMILIAL BREAST CANCER, FAMILIAL MALE, INCLUDED [view all locations] BREAST-OVARIAN CANCER, FAMILIAL, SUSCEPTIBILITY TO, 2; BROVCA2 BREAST CANCER, FAMILIAL, SUSCEPTIBILITY TO, 2, INCLUDED [#612555] BREAST-OVARIAN CANCER, FAMILIAL, SUSCEPTIBILITY TO, 2; BROVCA2 BREAST CANCER, FAMILIAL, SUSCEPTIBILITY TO, 2, INCLUDED;;OVARIAN CANCER, FAMILIAL, SUSCEPTIBILITY TO, 2, INCLUDED [view all locations] FANCONI ANEMIA, COMPLEMENTATION GROUP D1; FANCD1 [#605724] FANCONI ANEMIA, COMPLEMENTATION GROUP D1; FANCD1;;FAD1 [view all locations] GLIOMA SUSCEPTIBILITY 3; GLM3 [#613029] GLIOMA SUSCEPTIBILITY 3; GLM3 [view all locations] MEDULLOBLASTOMA; MDB MEDULLOBLASTOMA, DESMOPLASTIC, INCLUDED [#155255] MEDULLOBLASTOMA; MDB MEDULLOBLASTOMA, DESMOPLASTIC, INCLUDED;;MEDULLOBLASTOMA WITH EXTENSIVE NODULARITY, INCLUDED; MBEN, INCLUDED [view all locations] PANCREATIC CANCER, SUSCEPTIBILITY TO, 2 [#613347] PANCREATIC CANCER, SUSCEPTIBILITY TO, 2;;PNCA2 [view all locations] PROSTATE CANCER [#176807] PROSTATE CANCER [view all locations]

①外部数据库
包含Atlas、HGNC、LRG、MIM等多个数据库
②每个外部数据库所展示的内容，点击可进入查看

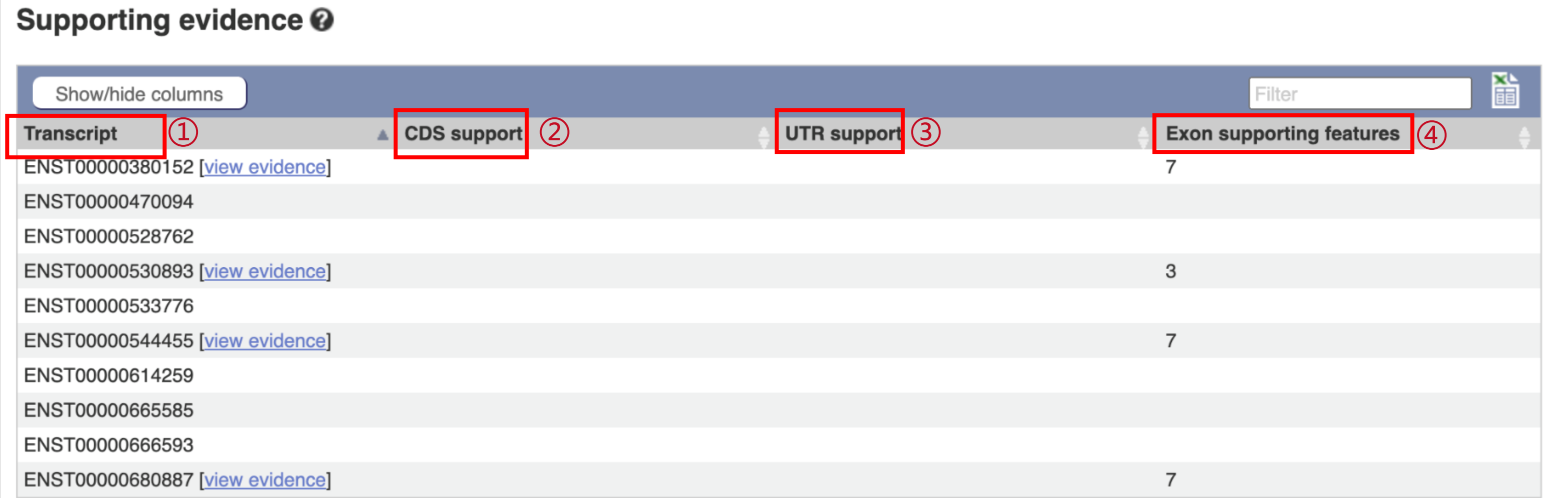
Ensembl检索方法---关键词检索

■ BRCA2的基因信息结果界面

➢ 在左侧导航栏中点击Supporting evidence，其右侧内容展示区将如下图所示。



Gene-based displays
- Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history



Supporting evidence

Transcript	CDS support	UTR support	Exon supporting features
ENST00000380152 [view evidence]			7
ENST00000470094			
ENST00000528762			
ENST00000530893 [view evidence]			3
ENST00000533776			
ENST00000544455 [view evidence]			7
ENST00000614259			
ENST00000665585			
ENST00000666593			
ENST00000680887 [view evidence]			7

①转录本信息

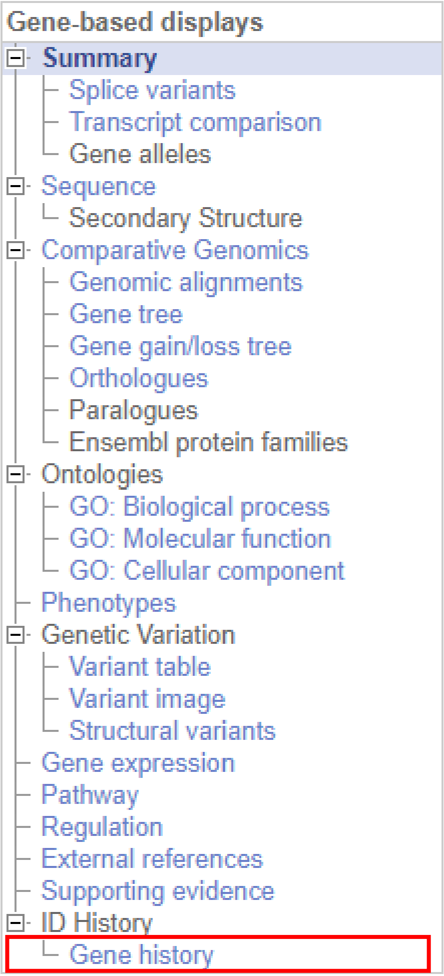
②CDS支持依据

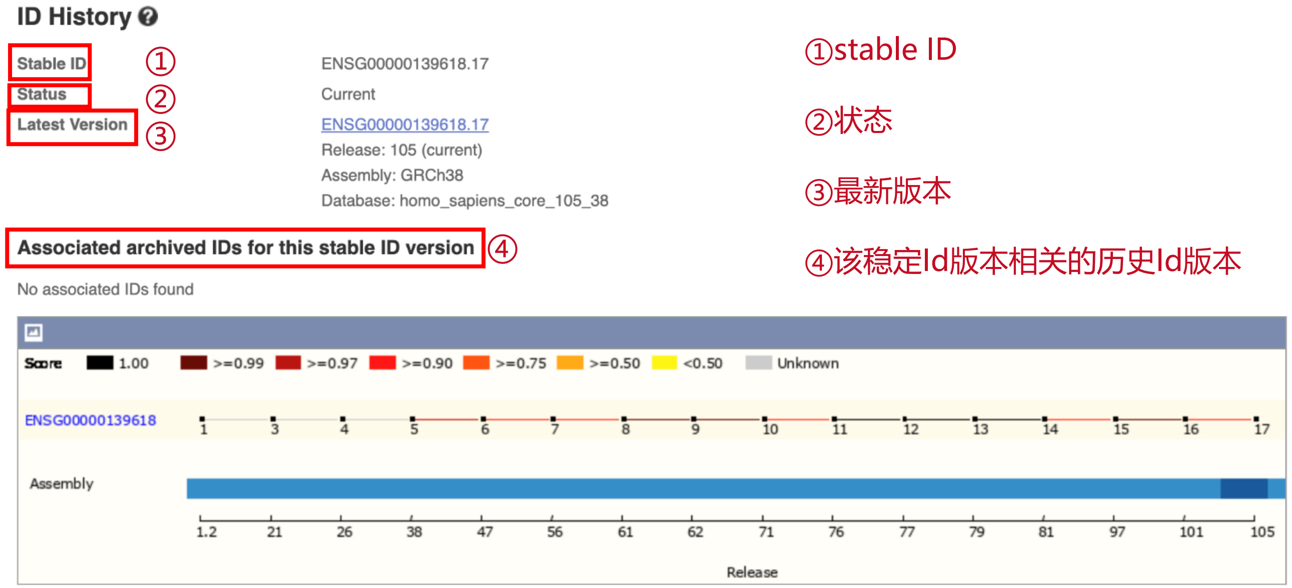
③UTR支持依据

④Exon支持依据

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面
  - ➢ 在左侧导航栏中点击Gene history，其右侧内容展示区将如下图所示。





ID History

Stable ID ① ENSG00000139618.17

Status ② Current

Latest Version ③ ENSG00000139618.17
Release: 105 (current)
Assembly: GRCh38
Database: homo_sapiens_core_105_38

Associated archived IDs for this stable ID version ④

No associated IDs found

①stable ID

②状态

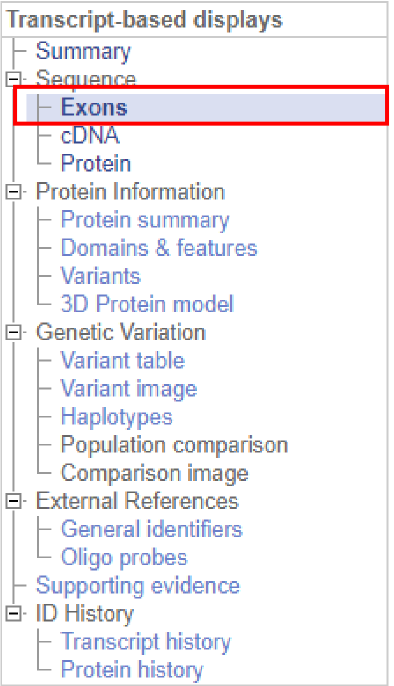
③最新版本

④该稳定Id版本相关的历史Id版本

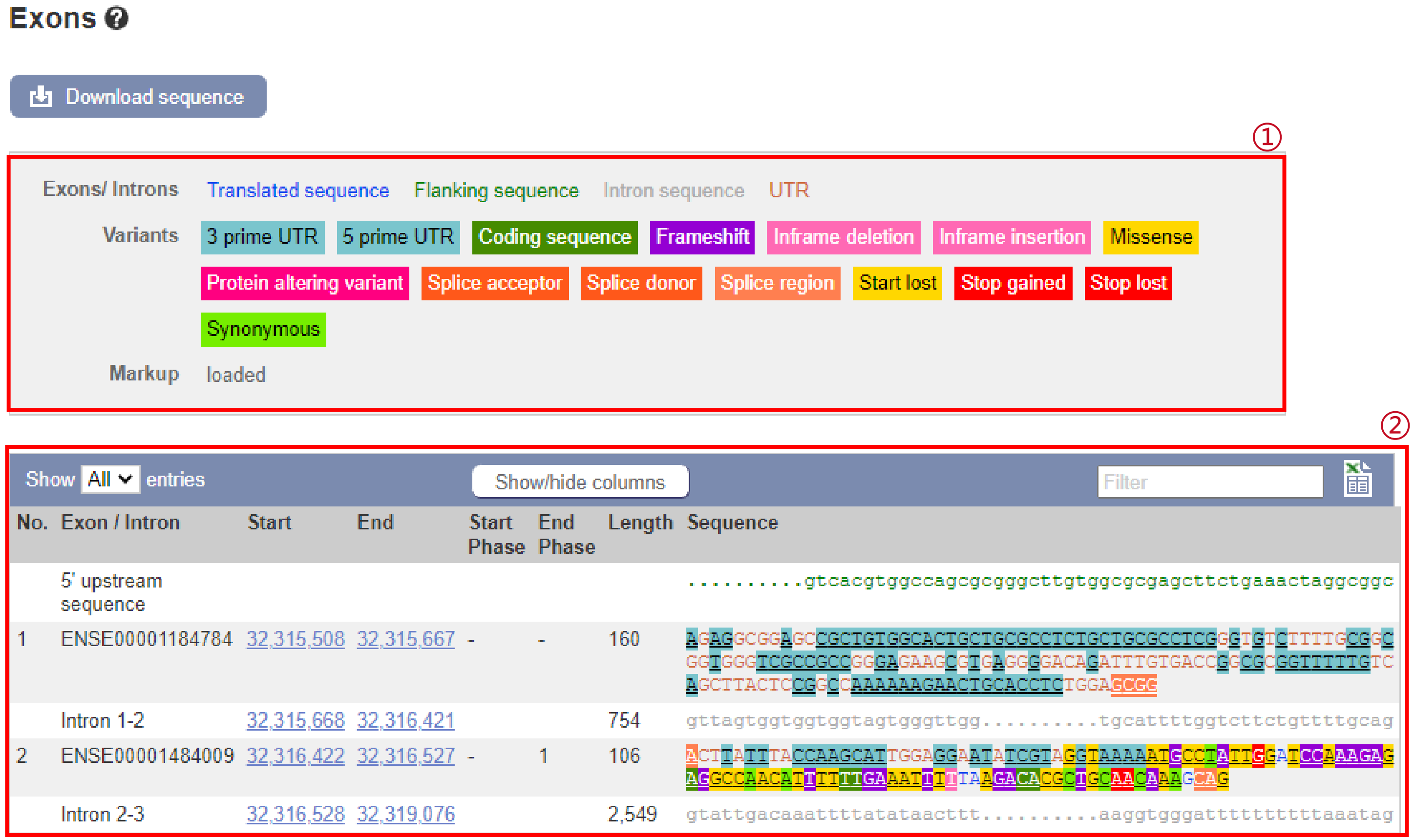
Ensembl检索方法---关键词检索

■ BRCA2的基因信息结果界面——转录本信息BRCA2-201

➢ 点击Transcript table中的第一个转录本信息BRCA2-201，在跳转的新界面左侧导航栏中点击Exons，其右侧内容展示区将如下图所示。



Transcript-based displays
- Summary
- Sequence
  - Exons
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
  - 3D Protein model
- Genetic Variation
  - Variant table
  - Variant image
  - Haplotypes
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history



Exons

①

Exons/ Introns: Translated sequence, Flanking sequence, Intron sequence, UTR

Variants: 3 prime UTR, 5 prime UTR, Coding sequence, Frameshift, Inframe deletion, Inframe insertion, Missense, Protein altering variant, Splice acceptor, Splice donor, Splice region, Start lost, Stop gained, Stop lost, Synonymous

Markup: loaded

②

No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
	5' upstream sequence						..........gtcacgtggccagcgcgggcttgtggcgcgagcttctgaaactaggcggc
1	ENSE00001184784	32,315,508	32,315,667	-	-	160	AGAGGCGGAGCCGCTGTGGCACTGCTGCGCCTCTGCTGCGCCTCGGGTGTCTTTTGCGGC GGTGGGTCGCCGCCGGGAGAAGCGTGAGGGGACAGATTTGTGACCGGCGCGGTTTTTGTC AGCTTACTCCGGCCAAAAAAGAACTGCACCTCTGGAGCGG
	Intron 1-2	32,315,668	32,316,421			754	gttagtggtggtggtagtgggttgg..........tgcattttggtcttctgttttgcag
2	ENSE00001484009	32,316,422	32,316,527	-	1	106	ACTTATTTACCAAGCATTGGAGGAATATCGTAGGTAAAAATGCCTATTGGATCCAAAGAG AGGCCAACATTTTTGAAATTTTTAAGACACGCTGCAACAAAGCAG
	Intron 2-3	32,316,528	32,319,076			2,549	gtattgacaaattttatataacttt..........aaggtgggattttttttttaaatag

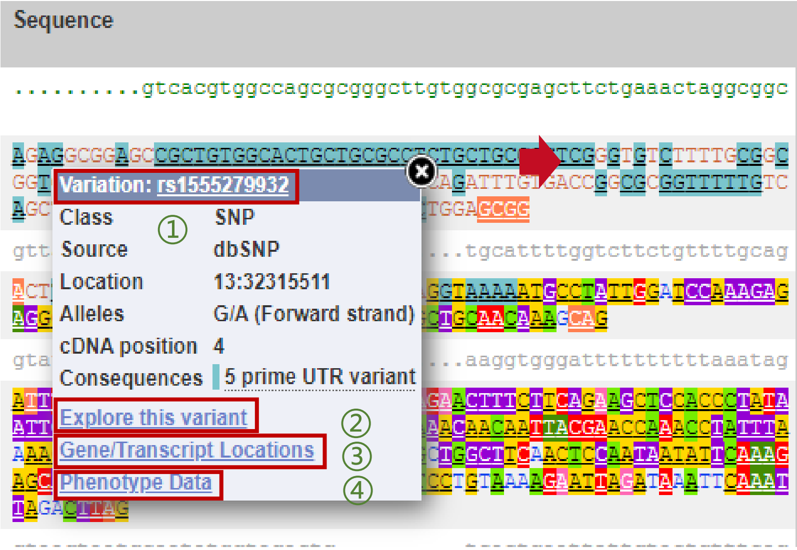
①外显子相关信息

②转录本详情

- 外显子/内含子
- 起始位置
- 结束位置
- 起始时相
- 结束时相
- 突变长度
- 序列信息，点击可查看具体变异，见下页PPT

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - 在跳转的新界面左侧导航栏中点击Exons，其右侧内容展示区将如下图所示——接上页PPT。点击序列中上页任意变异，可查看详细信息，其具体内容展示区将如下图所示。

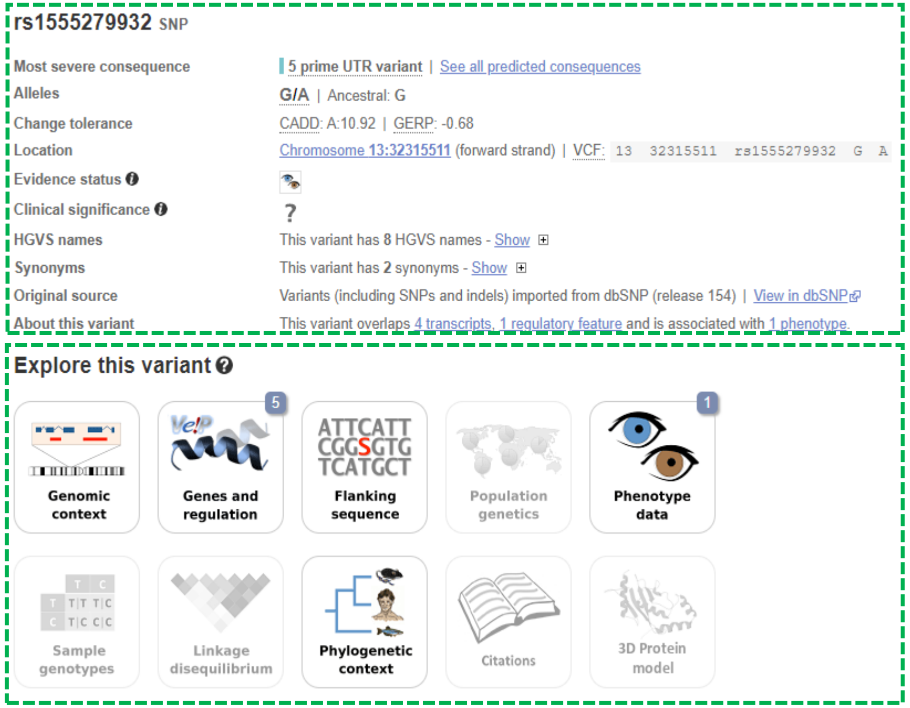


Sequence

.........gtcacgtggccagcgcgggcttgtggcgcgagcttctgaaactaggcggc

Variation: rs1555279932	
Class	SNP
Source	dbSNP
Location	13:32315511
Alleles	G/A (Forward strand)
cDNA position	4
Consequences	5 prime UTR variant

① Explore this variant ②
Gene/Transcript Locations ③
Phenotype Data ④



rs1555279932 SNP

Most severe consequence	5 prime UTR variant | See all predicted consequences
Alleles	G/A | Ancestral: G
Change tolerance	CADD: A:10.92 | GERP: -0.68
Location	Chromosome 13:32315511 (forward strand) | VCF: 13 32315511 rs1555279932 G A
Evidence status	
Clinical significance	?
HGVS names	This variant has 8 HGVS names - Show ⊞
Synonyms	This variant has 2 synonyms - Show ⊞
Original source	Variants (including SNPs and indels) imported from dbSNP (release 154) | View in dbSNP
About this variant	This variant overlaps 4 transcripts, 1 regulatory feature and is associated with 1 phenotype.

Explore this variant

①点击Variation中的变异体序号，可查看详细信息

- Most severe consequence（大部分序列）
- Alleles（等位基因）
- Change tolerance（容差）
- Location（定位）
- Evidence status（证据状态）
- Clinical significance（临床意义）
- HGVS names (HGVS名)
- Synonyms (别名)
- Original source (初始来源)
- About this variant (关于该转录本)

②点击Explore this variant即可跳转至此

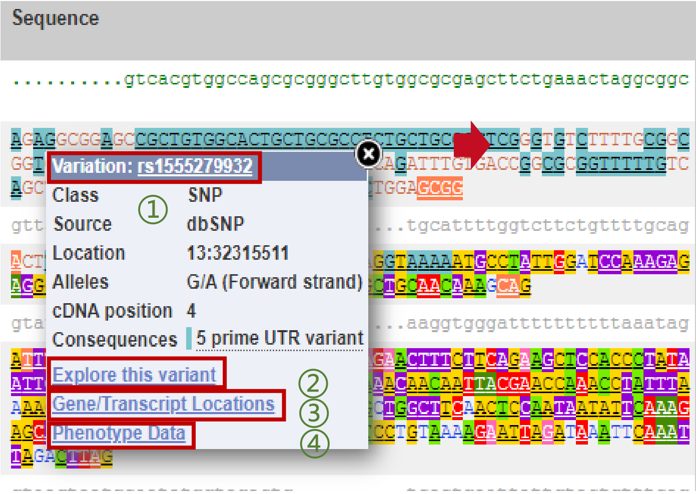
- Genomic context（遗传背景）
- Gene and regulation（基因调控）
- Flanking sequence（侧翼序列）
- Phenotype data（表型数据）
- Sample genotypes（样本基因型）
- Linkage disequilibrium (连锁不平衡)
- Phylogenetic context (种系背景)

③点击Gene/Transcript variant跳转至此页

④点击Phenotype data跳转至此页

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - ➢ 在跳转的新界面左侧导航栏中点击Exons，其右侧内容展示区将如下图所示——接上页PPT。点击序列中上页任意变异位点，可查看详细信息，其具体内容展示区将如下图所示。



Sequence

..........gtcacgtggccagcgcgggcttgtggcgcgagcttctgaaactaggcggc

...tgcattttggtcttctgttttgcag

...aaggtgggatttttttttaaatag

Variation: rs1555279932

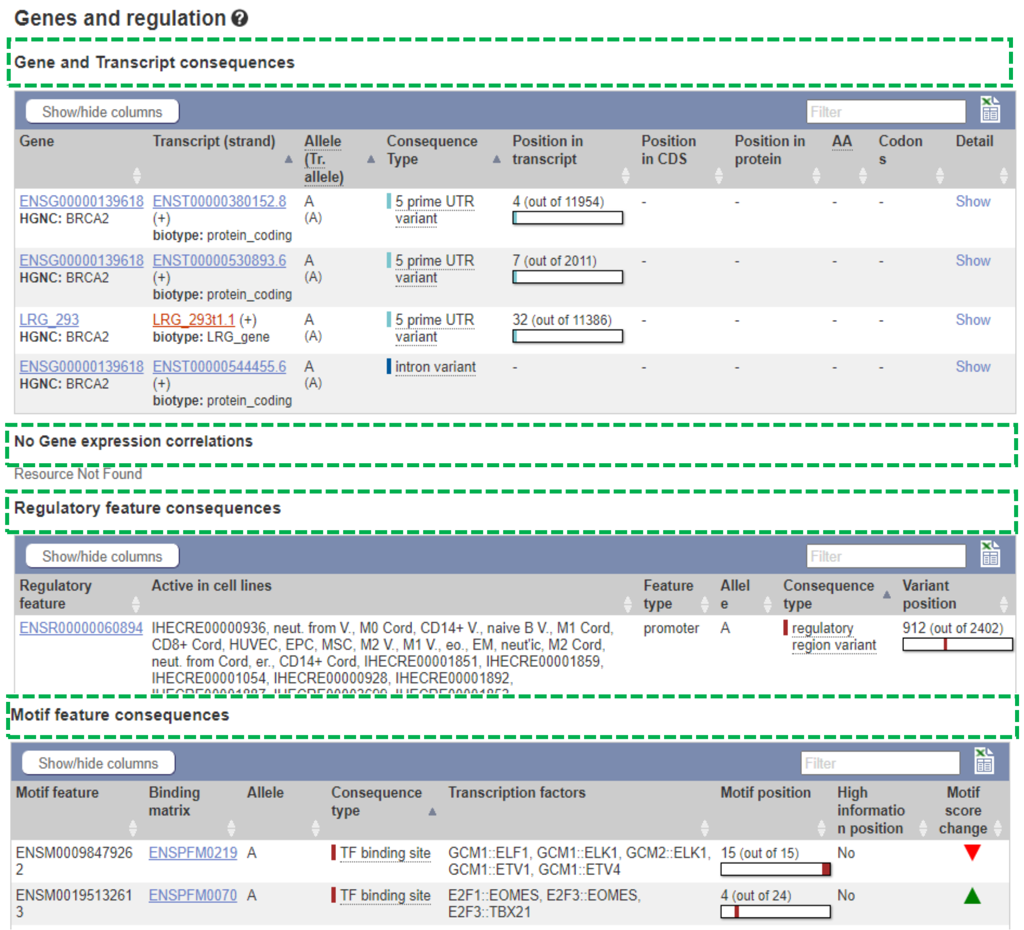
Class	SNP
Source	dbSNP
Location	13:32315511
Alleles	G/A (Forward strand)
cDNA position	4
Consequences	5 prime UTR variant

①

Explore this variant ②

Gene/Transcript Locations ③

Phenotype Data ④



Genes and regulation

Gene and Transcript consequences

Gene	Transcript (strand)	Allele (Tr. allele)	Consequence Type	Position in transcript	Position in CDS	Position in protein	AA	Codons	Detail
ENSG00000139618 HGNC: BRCA2	ENST00000380152.8 (+) biotype: protein_coding	A (A)	5 prime UTR variant	4 (out of 11954)	-	-	-	-	Show
ENSG00000139618 HGNC: BRCA2	ENST00000530893.6 (+) biotype: protein_coding	A (A)	5 prime UTR variant	7 (out of 2011)	-	-	-	-	Show
LRG_293 HGNC: BRCA2	LRG_293t1.1 (+) biotype: LRG_gene	A (A)	5 prime UTR variant	32 (out of 11386)	-	-	-	-	Show
ENSG00000139618 HGNC: BRCA2	ENST00000544455.6 (+) biotype: protein_coding	A (A)	intron variant	-	-	-	-	-	Show

No Gene expression correlations

Resource Not Found

Regulatory feature consequences

Regulatory feature	Active in cell lines	Feature type	Allele	Consequence type	Variant position
ENSR00000060894	IHECRE00000936, neut. from V., M0 Cord, CD14+ V., naive B V., M1 Cord, CD8+ Cord, HUVEC, EPC, MSC, M2 V., M1 V., eo., EM, neut'ic, M2 Cord, neut. from Cord, er., CD14+ Cord, IHECRE00001851, IHECRE00001859, IHECRE00001054, IHECRE00000928, IHECRE00001892,	promoter	A	regulatory region variant	912 (out of 2402)

Motif feature consequences

Motif feature	Binding matrix	Allele	Consequence type	Transcription factors	Motif position	High information position	Motif score change
ENSM00098479262	ENSPFM0219	A	TF binding site	GCM1::ELF1, GCM1::ELK1, GCM2::ELK1, GCM1::ETV1, GCM1::ETV4	15 (out of 15)	No	▼
ENSM00195132613	ENSPFM0070	A	TF binding site	E2F1::EOMES, E2F3::EOMES, E2F3::TBX21	4 (out of 24)	No	▲

③点击Gene/Transcript Location跳转至此页

Gene and Transcript consequences 基因和转录本结果

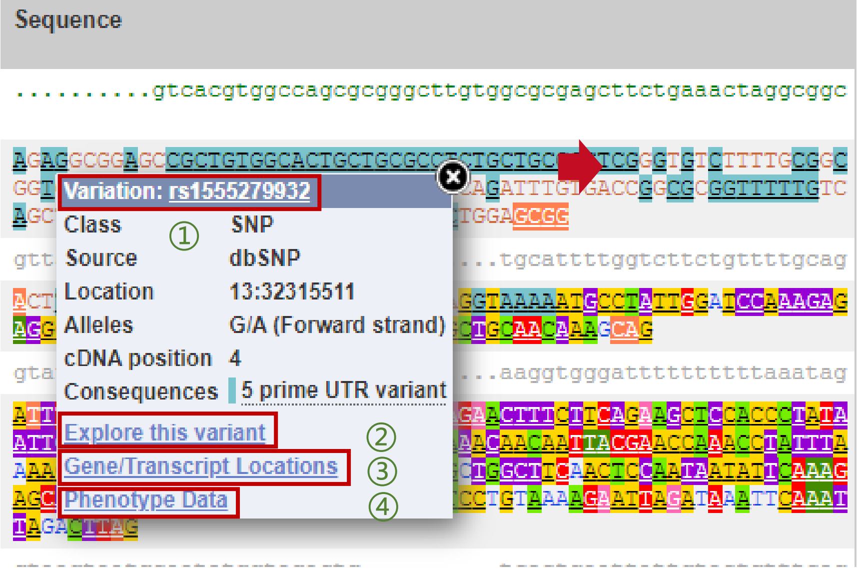
Gene expression correlations: 基因表达相关性分析

Regulatory feature consequences: 调节特征结果

Motif feature consequences：motif特征结果

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - 在跳转的新界面左侧导航栏中点击Exons，其右侧内容展示区将如下图所示——接上页PPT。点击序列中上页任意变异位点，可查看详细信息，其具体内容展示区将如下图所示。



Sequence

..........gtcacgtggccagcgcgggcttgtggcgcgagcttctgaaactaggcggc

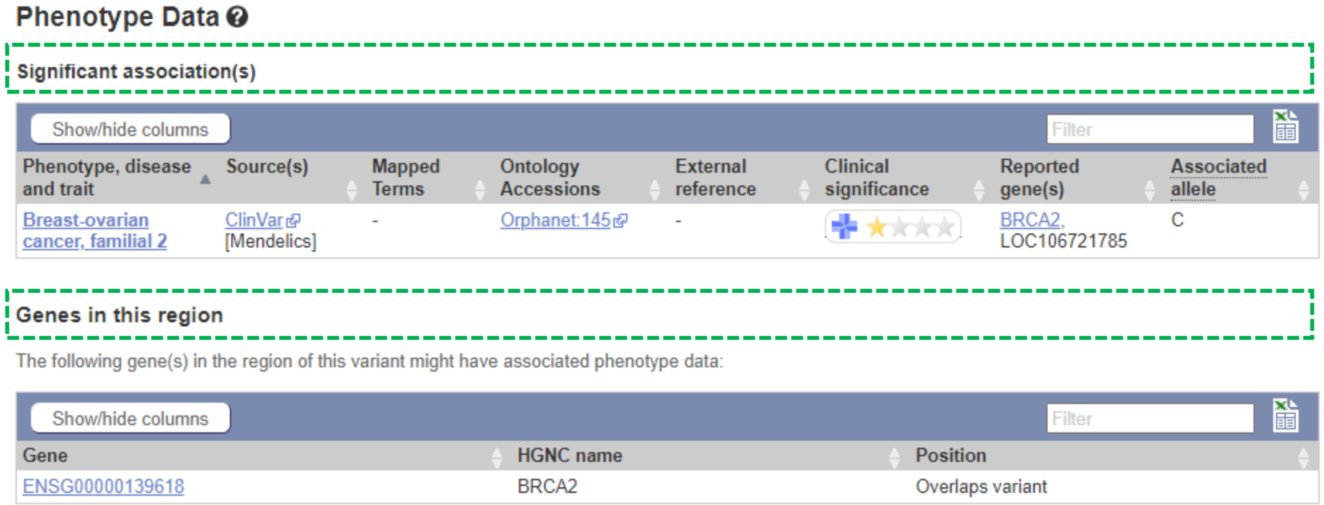
Variation: rs1555279932

Class ①	SNP
Source	dbSNP
Location	13:32315511
Alleles	G/A (Forward strand)
cDNA position	4
Consequences	5 prime UTR variant

Explore this variant ②

Gene/Transcript Locations ③

Phenotype Data ④



Phenotype Data

Significant association(s)

Phenotype, disease and trait	Source(s)	Mapped Terms	Ontology Accessions	External reference	Clinical significance	Reported gene(s)	Associated allele
Breast-ovarian cancer, familial 2	ClinVar [Mendelics]	-	Orphanet:145	-	★	BRCA2, LOC106721785	C

Genes in this region

The following gene(s) in the region of this variant might have associated phenotype data:

Gene	HGNC name	Position
ENSG00000139618	BRCA2	Overlaps variant

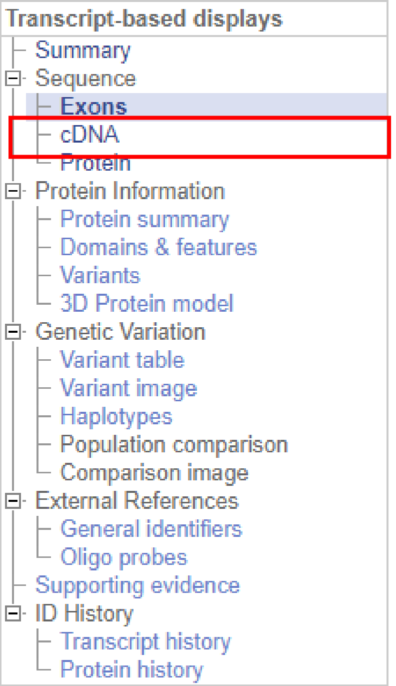
④点击Phenotype data跳转至此页

Significant associations: 显著相关的表型信息

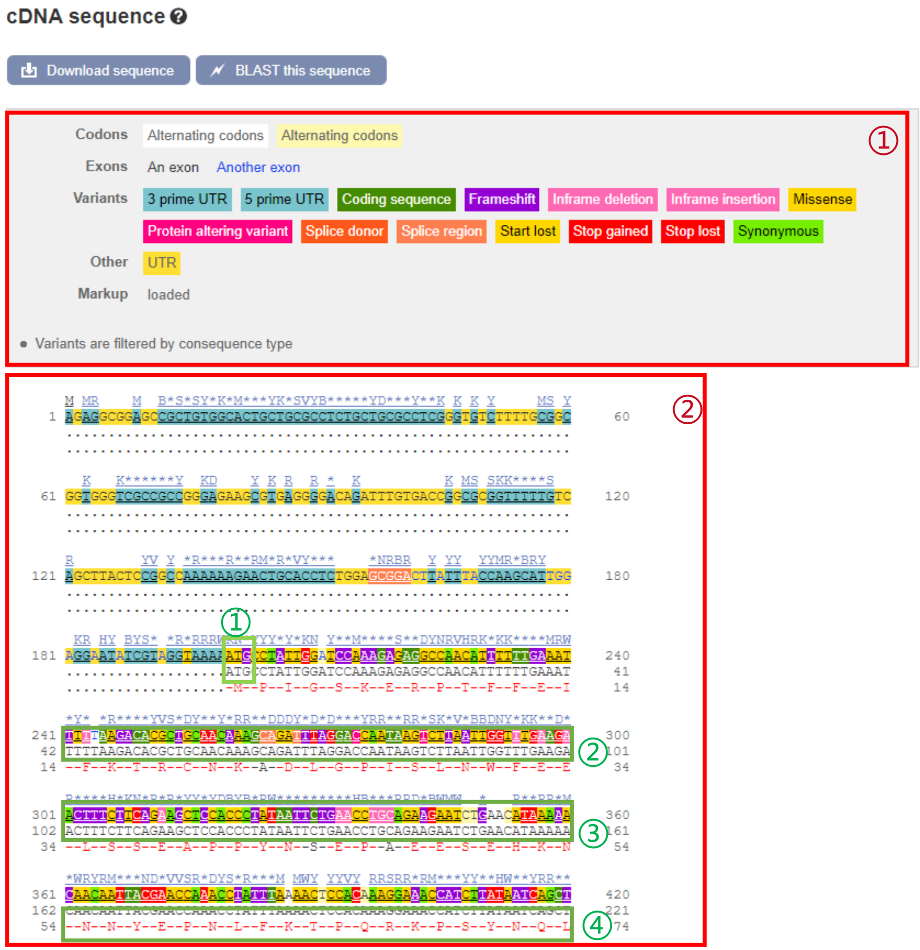
Genes in this region: 该变异相关的基因

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - 点击Transcript table中的第一个转录本信息BRCA2-201，在跳转的新界面（P25）左侧导航栏中点击cDNA，其右侧内容展示区将如下图所示。



Transcript-based displays
- Summary
- Sequence
  - Exons
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
  - 3D Protein model
- Genetic Variation
  - Variant table
  - Variant image
  - Haplotypes
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history



cDNA sequence

Codons	Alternating codons Alternating codons
Exons	An exon Another exon
Variants	3 prime UTR 5 prime UTR Coding sequence Frameshift Inframe deletion Inframe insertion Missense Protein altering variant Splice donor Splice region Start lost Stop gained Stop lost Synonymous
Other	UTR
Markup	loaded

- Variants are filtered by consequence type

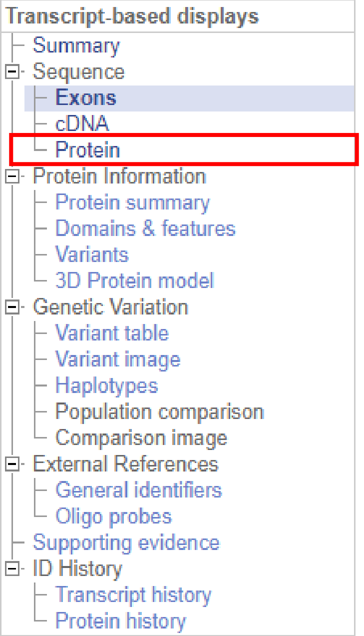
①不同颜色代表不同含义
- 密码子
- 外显子
- 变异
- 不同颜色分别代表3‘UTR，5’UTR区、编码区、移码突变、框内缺失突变，框内插入，错义突变，蛋白可变剪切，剪切供体、剪切区域、start-lost突变，stop-gain突变，stop-lost突变，同义突变
- 其他：黄色代表UTR区
- 标记

②展示详细的序列信息

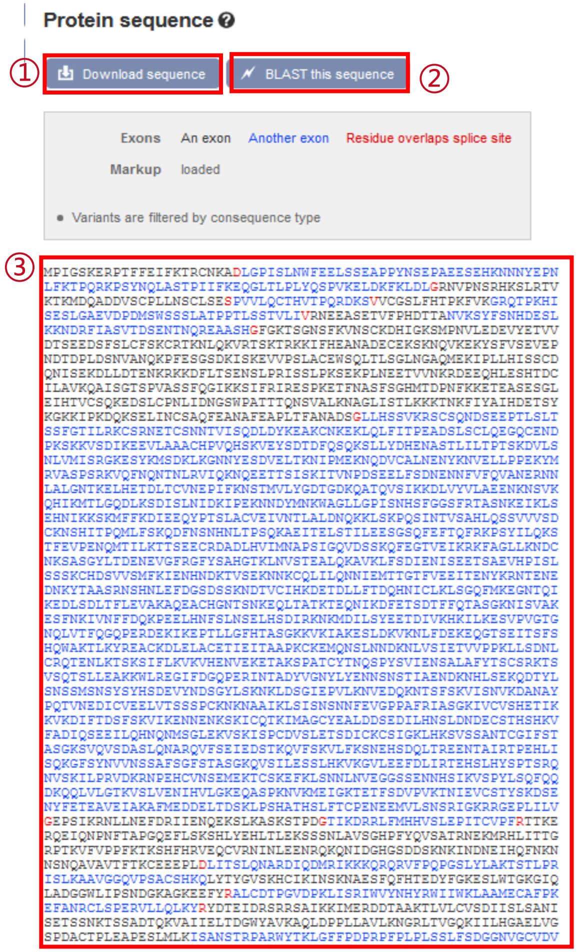
① 起始密码子
② 包含UTR区cDNA序列
③ 不包含UTR区cDNA序列
④ 对应的氨基酸序列

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - 新跳转界面的左侧导航栏中点击Protein，其右侧内容展示区将如下图所示。



Transcript-based displays
- Summary
- Sequence
  - Exons
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
  - 3D Protein model
- Genetic Variation
  - Variant table
  - Variant image
  - Haplotypes
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history



Protein sequence

① Download sequence

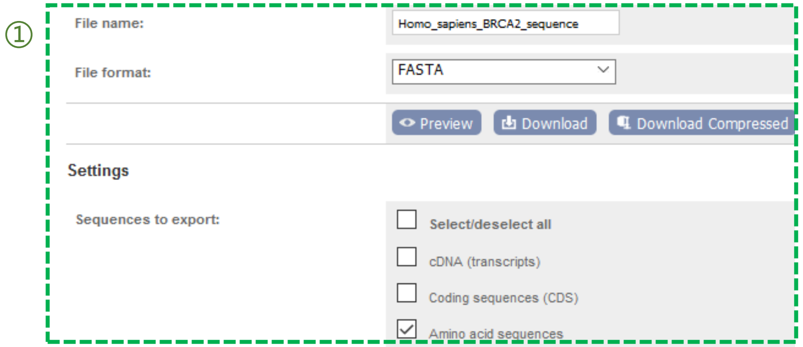
BLAST this sequence ②

Exons	An exon	Another exon	Residue overlaps splice site
Markup	loaded		

- Variants are filtered by consequence type

③

```
MPIGSKERPTFFEIFKTRCNKADLGPISLNWFEELSSEAPPYNSEPAEESEHKNNNYEPN
LFKTPQRKPSYNQLASTPIIFKEQGLTLPLYQSPVKELDKFKLDLGRNVPNSRHKSLRTV
KTKMDQADDVSCPLLNSCLSESPVVLQCTHVTPQRDKSVVCGSLFHTPKFVKGRQTPKHI
SESLGAEVDPDMSWSSSLATPPTLSSTVLIVRNEEASETVFPHDTTANVKSYFSNHDESL
KKNDRFIASVTDSENTNQREAASHGFGKTSGNSFKVNSCKDHIGKSMPNVLEDEVYETVV
DTSEEDSFSLCFSKCRTKNLQKVRTSKTRKKIFHEANADECEKSKNQVKEKYSFVSEVEP
NDTDPLDSNVANQKPFESGSDKISKEVVPSLACEWSQLTLSGLNGAQMEKIPLLHISSCD
QNISEKDLLDTENKRKKDFLTSENSLPRISSLPKSEKPLNEETVVNKRDEEQHLESHTDC
ILAVKQAISGTSPVASSFQGIKKSIFRIRESPKETFNASFSGHMTDPNFKKETEASESGL
EIHTVCSQKEDSLCPNLIDNGSWPATTTQNSVALKNAGLISTLKKKTNKFIYAIHDETSY
KGKKIPKDQKSELINCSAQFEANAFEAPLTFANADSGLLHSSVKRSCSQNDSEEPTLSLT
SSFGTILRKCSRNETCSNNTVISQDLDYKEAKCNKEKLQLFITPEADSLSCLQEGQCEND
PKSKKVSDIKEEVLAAACHPVQHSKVEYSDTDFQSQKSLLYDHENASTLILTPTSKDVLS
NLVMISRGKESYKMSDKLKGNNYESDVELTKNIPMEKNQDVCALNENYKNVELLPPEKYM
RVASPSRKVQFNQNTNLRVIQKNQEETTSISKITVNPDSEELFSDNENNFVFQVANERNN
LALGNTKELHETDLTCVNEPIFKNSTMVLYGDTGDKQATQVSIKKDLVYVLAEENKNSVK
QHIKMTLGQDLKSDISLNIDKIPEKNNDYMNKWAGLLGPISNHSFGGSFRTASNKEIKLS
EHNIKKSKMFFKDIEEQYPTSLACVEIVNTLALDNQKKLSKPQSINTVSAHLQSSVVVSD
CKNSHITPQMLFSKQDFNSNHNLTPSQKAEITELSTILEESGSQFEFTQFRKPSYILQKS
TFEVPENQMTILKTTSEECRDADLHVIMNAPSIGQVDSSKQFEGTVEIKRKFAGLLKNDC
NKSASGYLTDENEVGFRGFYSAHGTKLNVSTEALQKAVKLFSDIENISEETSAEVHPISL
SSSKCHDSVVSMFKIENHNDKTVSEKNNKCQLILQNNIEMTTGTFVEEITENYKRNTENE
DNKYTAASRNSHNLEFDGSDSSKNDTVCIHKDETDLLFTDQHNICLKLSGQFMKEGNTQI
KEDLSDLTFLEVAKAQEACHGNTSNKEQLTATKTEQNIKDFETSDTFFQTASGKNISVAK
ESFNKIVNFFDQKPEELHNFSLNSELHSDIRKNKMDILSYEETDIVKHKILKESVPVGTG
NQLVTFQGQPERDEKIKEPTLLGFHTASGKKVKIAKESLDKVKNLFDEKEQGTSEITSFS
HQWAKTLKYREACKDLELACETIEITAAPKCKEMQNSLNNDKNLVSIETVVPPKLLSDNL
CRQTENLKTSKSIFLKVKVHENVEKETAKSPATCYTNQSPYSVIENSALAFYTSCSRKTS
VSQTSLLEAKKWLREGIFDGQPERINTADYVGNYLYENNSNSTIAENDKNHLSEKQDTYL
SNSSMSNSYSYHSDEVYNDSGYLSKNKLDSGIEPVLKNVEDQKNTSFSKVISNVKDANAY
PQTVNEDICVEELVTSSSPCKNKNAAIKLSISNSNNFEVGPPAFRIASGKIVCVSHETIK
KVKDIFTDSFSKVIKENNENKSKICQTKIMAGCYEALDDSEDILHNSLDNDECSTHSHKV
FADIQSEEILQHNQNMSGLEKVSKISPCDVSLETSDICKCSIGKLHKSVSSANTCGIFST
ASGKSVQVSDASLQNARQVFSEIEDSTKQVFSKVLFKSNEHSDQLTREENTAIRTPEHLI
SQKGFSYNVVNSSAFSGFSTASGKQVSILESSLHKVKGVLEEFDLIRTEHSLHYSPTSRQ
NVSKILPRVDKRNPEHCVNSEMEKTCSKEFKLSNNLNVEGGSSENNHSIKVSPYLSQFQQ
DKQQLVLGTKVSLVENIHVLGKEQASPKNVKMEIGKTETFSDVPVKTNIEVCSTYSKDSE
NYFETEAVEIAKAFMEDDELTDSKLPSHATHSLFTCPENEEMVLSNSRIGKRRGEPLILV
GEPSIKRNLLNEFDRIIENQEKSLKASKSTPDGTIKDRRLFMHHVSLEPITCVPFRTTKE
RQEIQNPNFTAPGQEFLSKSHLYEHLTLEKSSSNLAVSGHPFYQVSATRNEKMRHLITTG
RPTKVFVPPFKTKSHFHRVEQCVRNINLEENRQKQNIDGHGSDDSKNKINDNEIHQFNKN
NSNQAVAVTFTKCEEEPLDLITSLQNARDIQDMRIKKKQRQRVFPQPGSLYLAKTSTLPR
ISLKAAVGGQVPSACSHKQLYTYGVSKHCIKINSKNAESFQFHTEDYFGKESLWTGKGIQ
LADGGWLIPSNDGKAGKEEFYRALCDTPGVDPKLISRIWVYNHYRWIIWKLAAMECAFPK
EFANRCLSPERVLLQLKYRYDTEIDRSRRSAIKKIMERDDTAAKTLVLCVSDIISLSANI
SETSSNKTSSADTQKVAIIELTDGWYAVKAQLDPPLLAVLKNGRLTVGQKIILHGAELVG
SPDACTPLEAPESLMLKISANSTRPARWYTKLGFFPDPRPFPLPLSSLFSDGGNVGCVDV
```



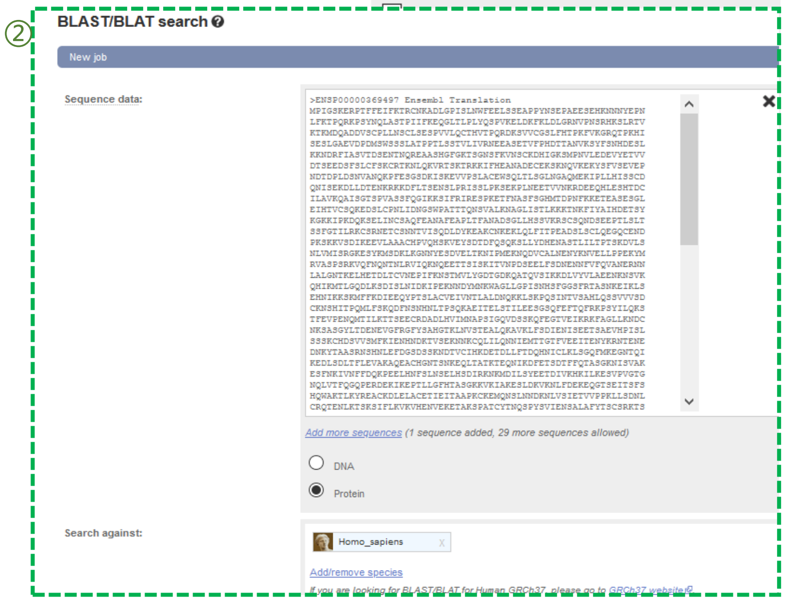
①

File name: Homo_sapiens_BRCA2_sequence

File format: FASTA

Settings

Sequences to export:
- ☐ Select/deselect all
- ☐ cDNA (transcripts)
- ☐ Coding sequences (CDS)
- ☑ Amino acid sequences



② BLAST/BLAT search

Sequence data:

```
>ENSP00000369497 Ensembl Translation
MPIGSKERPTFFEIFKTRCNKADLGPISLNWFEELSSEAPPYNSEPAEESEHKNNNYEPN
LFKTPQRKPSYNQLASTPIIFKEQGLTLPLYQSPVKELDKFKLDLGRNVPNSRHKSLRTV
KTKMDQADDVSCPLLNSCLSESPVVLQCTHVTPQRDKSVVCGSLFHTPKFVKGRQTPKHI
SESLGAEVDPDMSWSSSLATPPTLSSTVLIVRNEEASETVFPHDTTANVKSYFSNHDESL
KKNDRFIASVTDSENTNQREAASHGFGKTSGNSFKVNSCKDHIGKSMPNVLEDEVYETVV
DTSEEDSFSLCFSKCRTKNLQKVRTSKTRKKIFHEANADECEKSKNQVKEKYSFVSEVEP
NDTDPLDSNVANQKPFESGSDKISKEVVPSLACEWSQLTLSGLNGAQMEKIPLLHISSCD
QNISEKDLLDTENKRKKDFLTSENSLPRISSLPKSEKPLNEETVVNKRDEEQHLESHTDC
ILAVKQAISGTSPVASSFQGIKKSIFRIRESPKETFNASFSGHMTDPNFKKETEASESGL
EIHTVCSQKEDSLCPNLIDNGSWPATTTQNSVALKNAGLISTLKKKTNKFIYAIHDETSY
KGKKIPKDQKSELINCSAQFEANAFEAPLTFANADSGLLHSSVKRSCSQNDSEEPTLSLT
SSFGTILRKCSRNETCSNNTVISQDLDYKEAKCNKEKLQLFITPEADSLSCLQEGQCEND
PKSKKVSDIKEEVLAAACHPVQHSKVEYSDTDFQSQKSLLYDHENASTLILTPTSKDVLS
NLVMISRGKESYKMSDKLKGNNYESDVELTKNIPMEKNQDVCALNENYKNVELLPPEKYM
RVASPSRKVQFNQNTNLRVIQKNQEETTSISKITVNPDSEELFSDNENNFVFQVANERNN
LALGNTKELHETDLTCVNEPIFKNSTMVLYGDTGDKQATQVSIKKDLVYVLAEENKNSVK
QHIKMTLGQDLKSDISLNIDKIPEKNNDYMNKWAGLLGPISNHSFGGSFRTASNKEIKLS
EHNIKKSKMFFKDIEEQYPTSLACVEIVNTLALDNQKKLSKPQSINTVSAHLQSSVVVSD
CKNSHITPQMLFSKQDFNSNHNLTPSQKAEITELSTILEESGSQFEFTQFRKPSYILQKS
TFEVPENQMTILKTTSEECRDADLHVIMNAPSIGQVDSSKQFEGTVEIKRKFAGLLKNDC
NKSASGYLTDENEVGFRGFYSAHGTKLNVSTEALQKAVKLFSDIENISEETSAEVHPISL
SSSKCHDSVVSMFKIENHNDKTVSEKNNKCQLILQNNIEMTTGTFVEEITENYKRNTENE
DNKYTAASRNSHNLEFDGSDSSKNDTVCIHKDETDLLFTDQHNICLKLSGQFMKEGNTQI
KEDLSDLTFLEVAKAQEACHGNTSNKEQLTATKTEQNIKDFETSDTFFQTASGKNISVAK
ESFNKIVNFFDQKPEELHNFSLNSELHSDIRKNKMDILSYEETDIVKHKILKESVPVGTG
NQLVTFQGQPERDEKIKEPTLLGFHTASGKKVKIAKESLDKVKNLFDEKEQGTSEITSFS
HQWAKTLKYREACKDLELACETIEITAAPKCKEMQNSLNNDKNLVSIETVVPPKLLSDNL
CRQTENLKTSKSIFLKVKVHENVEKETAKSPATCYTNQSPYSVIENSALAFYTSCSRKTS
```

Add more sequences (1 sequence added, 29 more sequences allowed)

- ○ DNA
- ● Protein

Search against: Homo_sapiens

Add/remove species

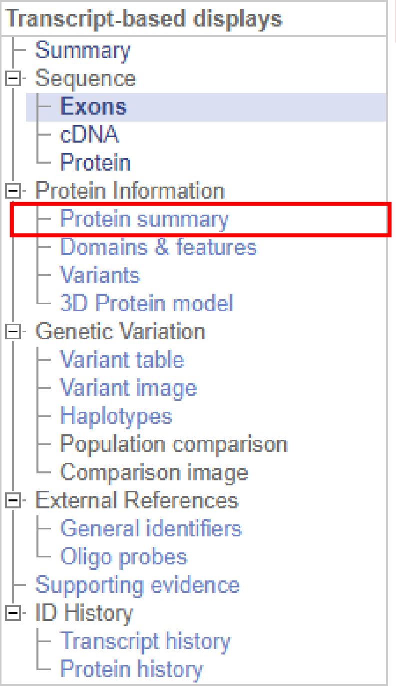
If you are looking for BLAST/BLAT for Human GRCh37, please go to GRCh37 website

①点击download sequence出现右侧①页面，可进行fasta等格式序列下载
① File name （文件名）
② File format （文件格式）
③ Settings (设置)
- 可根据需求设置下载内容

②点击BLAST this sequence进入②blast界面。 此界面具体内容见后述

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - 在新跳转界面的左侧导航栏中点击protein summary，其右侧内容展示区将如下图所示。



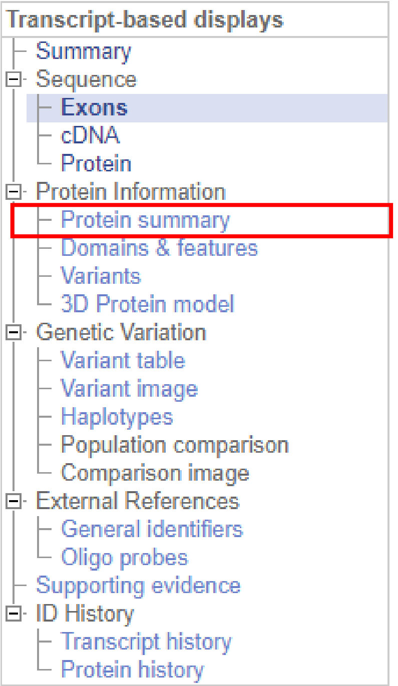
①该转录本对应的蛋白序列号

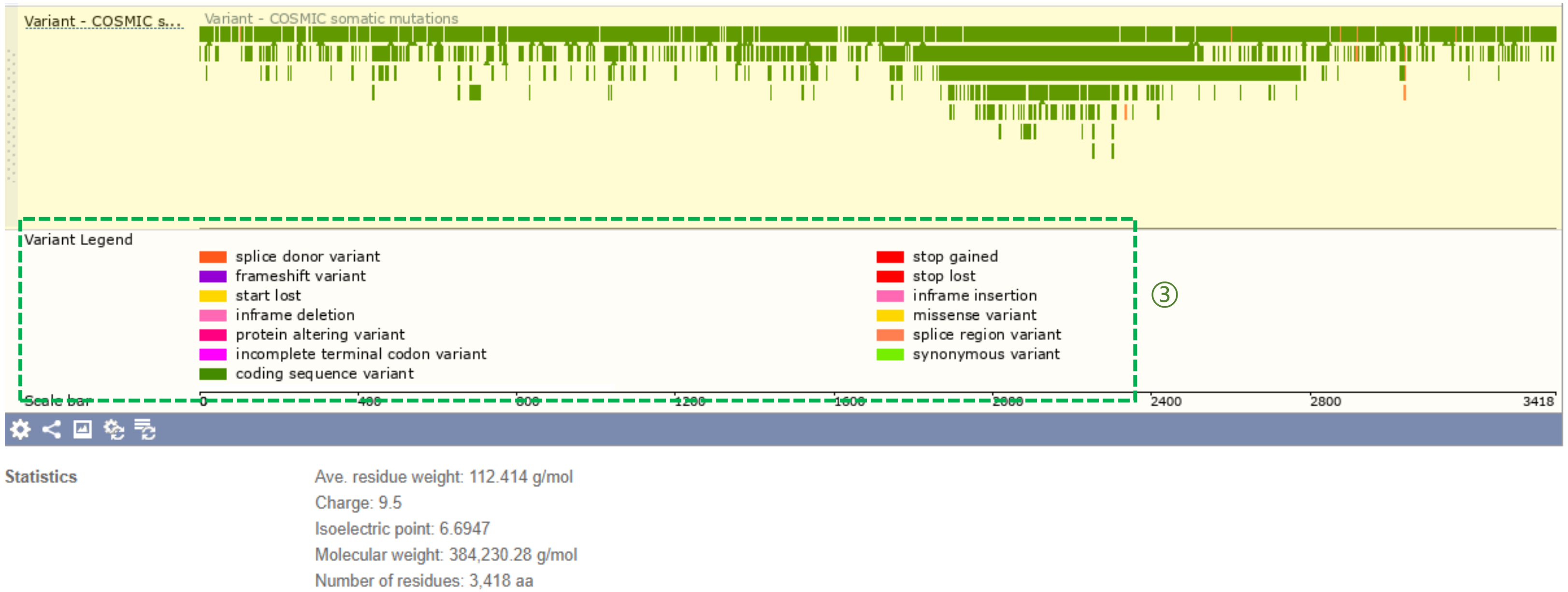
②不同数据库展现的该蛋白的结构信息

点击每个条目的具体信息，即可显示具体信息

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - ➢ 在新跳转界面的左侧导航栏中点击protein summary，其右侧内容展示区将如下图所示——接上页。





Statistics

Ave. residue weight: 112.414 g/mol
Charge: 9.5
Isoelectric point: 6.6947
Molecular weight: 384,230.28 g/mol
Number of residues: 3,418 aa

③变异标识

Splice donor variant剪切供体变异 stop gained 获得终止密码子

Frameshift variant 移码变异 stop lost终止密码子丢失

Start lost起始密码丢失 inframe insertion框内插入

Inframe deletion 框内删失 missense variant 错义突变

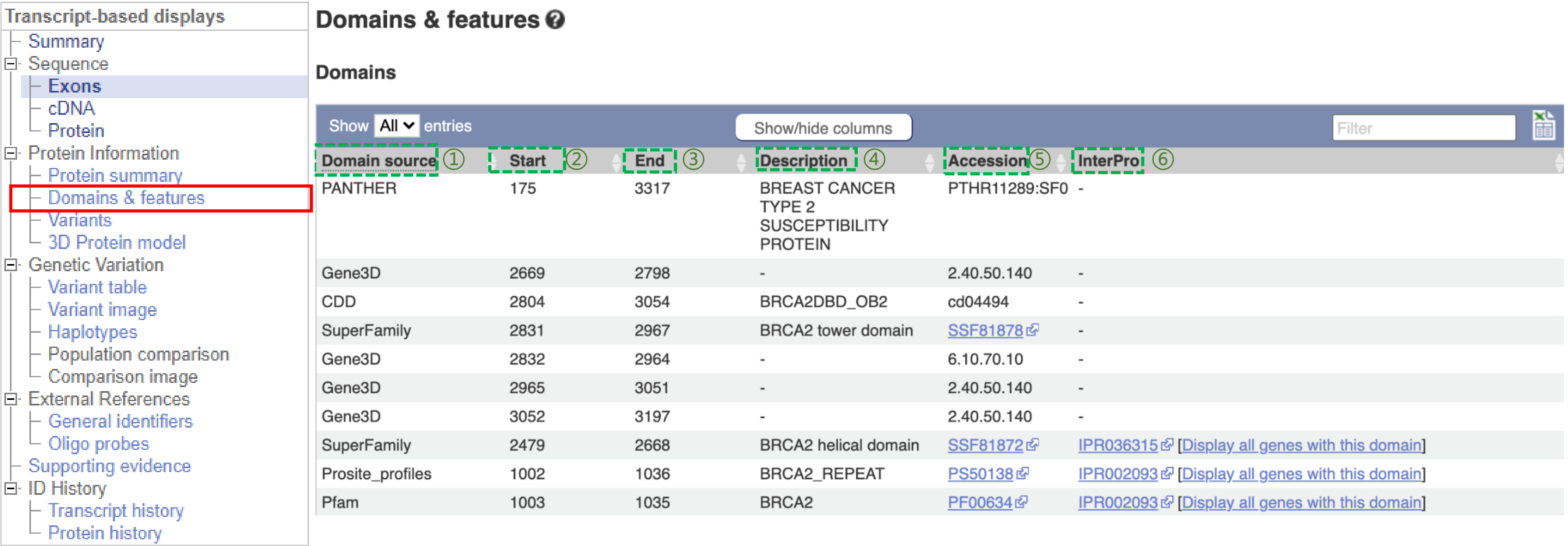
Protein altering variant蛋白改变变异 splice region variant剪切区域变异

Incomplete terminal codon variant不完全终止密码子变异 synonymous variant同义变体

Coding sequence variant 编码序列变异

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - ➢ 在新跳转界面的左侧导航栏中点击Domains features，其右侧内容展示区将如下图所示。



Domains & features

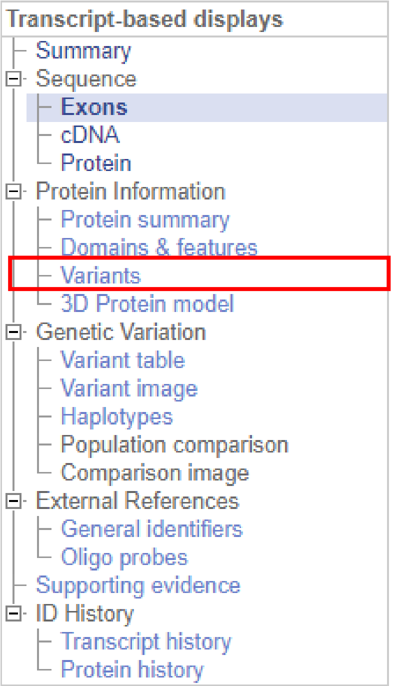
Domains

Domain source ①	Start ②	End ③	Description ④	Accession ⑤	InterPro ⑥
PANTHER	175	3317	BREAST CANCER TYPE 2 SUSCEPTIBILITY PROTEIN	PTHR11289:SF0	-
Gene3D	2669	2798	-	2.40.50.140	-
CDD	2804	3054	BRCA2DBD_OB2	cd04494	-
SuperFamily	2831	2967	BRCA2 tower domain	SSF81878	-
Gene3D	2832	2964	-	6.10.70.10	-
Gene3D	2965	3051	-	2.40.50.140	-
Gene3D	3052	3197	-	2.40.50.140	-
SuperFamily	2479	2668	BRCA2 helical domain	SSF81872	IPR036315 [Display all genes with this domain]
Prosite_profiles	1002	1036	BRCA2_REPEAT	PS50138	IPR002093 [Display all genes with this domain]
Pfam	1003	1035	BRCA2	PF00634	IPR002093 [Display all genes with this domain]

① Domain source (结构域来源)
② Start（起始位置)
③ End (结束位置)
④ Description (描述)
⑤ Accession (序列号)
⑥ InterPro数据库

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - 在新跳转界面的左侧导航栏中点击Variants，其右侧内容展示区将如下图所示。



Transcript-based displays
- Summary
- Sequence
  - Exons
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
  - 3D Protein model
- Genetic Variation
  - Variant table
  - Variant image
  - Haplotypes
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history

Residue	Variant ID	Conseq. Type	Source	Evidence	Alleles	Ambig. code	Residues	Codons	SIFT	PolyPhen	CADD	REVEL	MetaLR	Mutation Assessor
1	rs80359418	frameshift variant start lost	dbSNP		G/-	-	M, X	ATG, AT	-	-	-	-	-	-
1	rs80358650	start lost	dbSNP		G/A/T	D	M, I	ATG, ATA	0	0.511	24	0.396	0.014	-
1	rs80358650	start lost	dbSNP		G/A/T	D	M, I	ATG, ATT	0	0.511	24	0.396	0.014	-
1	rs80358547	start lost	dbSNP	ExAC gnomAD	T/A/C/G	N	M, K	ATG, AAG	0	0.898	24	0.287	0.013	-
1	rs80358547	start lost	dbSNP	ExAC gnomAD	T/A/C/G	N	M, T	ATG, ACG	0	0.863	24	0.346	0.013	-
1	rs80358547	start lost	dbSNP	ExAC gnomAD	T/A/C/G	N	M, R	ATG, AGG	0	0.935	24	0.325	0.013	-

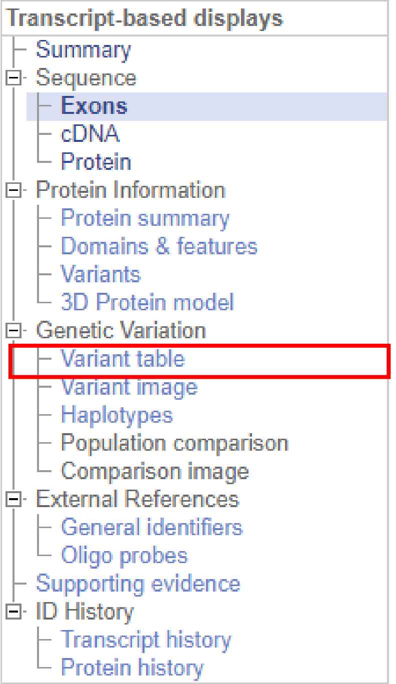
- 氨基酸残基数
- 变异ID号
- 序列类型
- 来源

- 证据等级
  - 突变频率
  - 被引用次数
  - 疾病表型发现
  - Topmed发现
  - gnomAD报道
- 等位基因

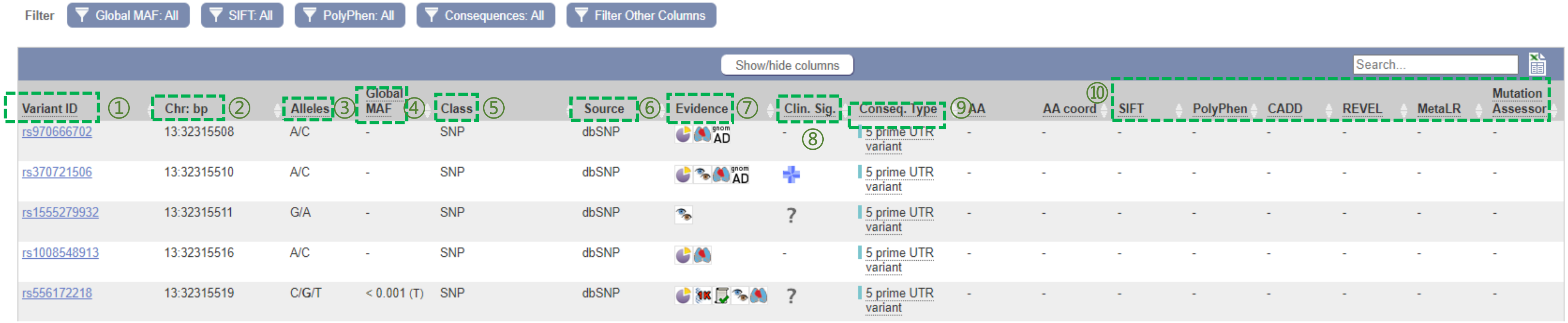
- 突变后氨基酸残基
- 突变后密码子序列
- 五种软件蛋白损伤预测结果

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - ➢ 在新跳转界面的左侧导航栏中点击Variant table， 其右侧内容展示区将如右图所示。



Transcript-based displays
- Summary
- Sequence
  - Exons
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
  - 3D Protein model
- Genetic Variation
  - Variant table
  - Variant image
  - Haplotypes
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history

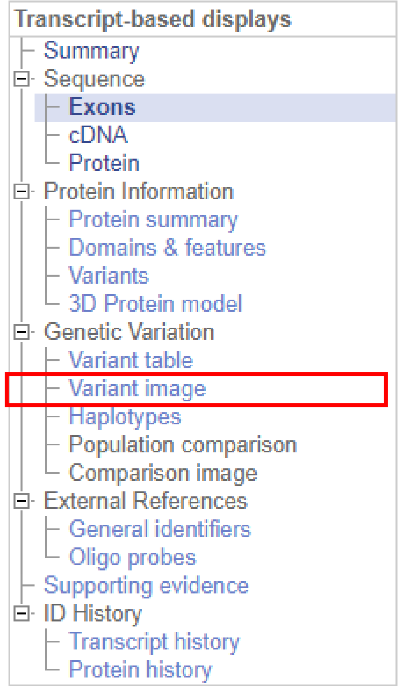


Variant ID	Chr: bp	Alleles	Global MAF	Class	Source	Evidence	Clin. Sig.	Conseq. Type	AA	AA coord	SIFT	PolyPhen	CADD	REVEL	MetaLR	Mutation Assessor
rs970666702	13:32315508	A/C	-	SNP	dbSNP	gnomAD	-	5 prime UTR variant	-	-	-	-	-	-	-	-
rs370721506	13:32315510	A/C	-	SNP	dbSNP	gnomAD	✚	5 prime UTR variant	-	-	-	-	-	-	-	-
rs1555279932	13:32315511	G/A	-	SNP	dbSNP		?	5 prime UTR variant	-	-	-	-	-	-	-	-
rs1008548913	13:32315516	A/C	-	SNP	dbSNP		-	5 prime UTR variant	-	-	-	-	-	-	-	-
rs556172218	13:32315519	C/G/T	< 0.001 (T)	SNP	dbSNP	1K	?	5 prime UTR variant	-	-	-	-	-	-	-	-

- ①变异ID号
- ②染色体定位
- ③等位基因
- ④第二位等位基因在全球突变频率
- ⑤分类
- ⑥来源
- ⑦证据等级
- ⑧临床贡献
- ⑨类型
- ⑩五种软件蛋白损伤预测结果

Ensembl检索方法---关键词检索

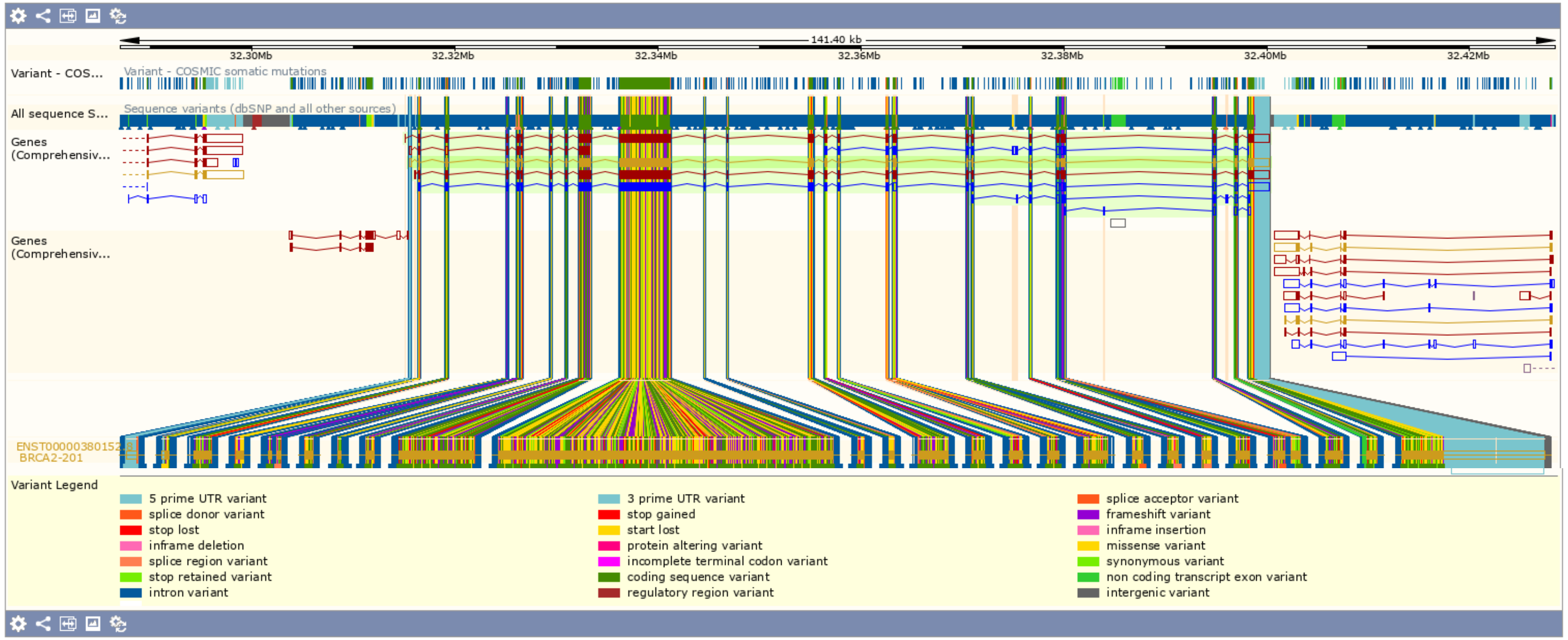
- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - 在新跳转界面的左侧导航栏中点击Variant image，其右侧内容展示区将如下图所示。



Transcript-based displays

- Summary
- Sequence
  - Exons
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
  - 3D Protein model
- Genetic Variation
  - Variant table
  - Variant image
  - Haplotypes
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history

Variant image



- 每个转录本的基因结构，外显子由框表示，连接框的线为内含子，填充框反应编码序列，未填充框反应UTRs（未翻译区）

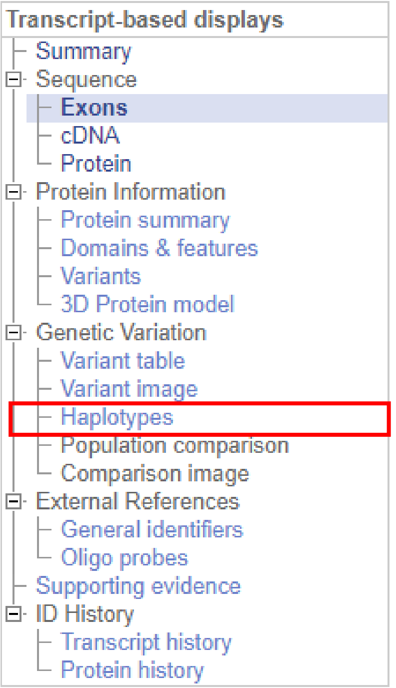
- 5 'UTR变异
- 剪切供体变异
- 终止密码子丢失
- 框内缺失突变
- 剪切区域变异
- 终止密码子保留变异
- 内含子变异

- 3' -UTR变异
- 获得终止密码子
- 起始密码子缺失
- 蛋白改变变异
- 不全终止密码子变异
- 编码序列变异
- 调节区域变异

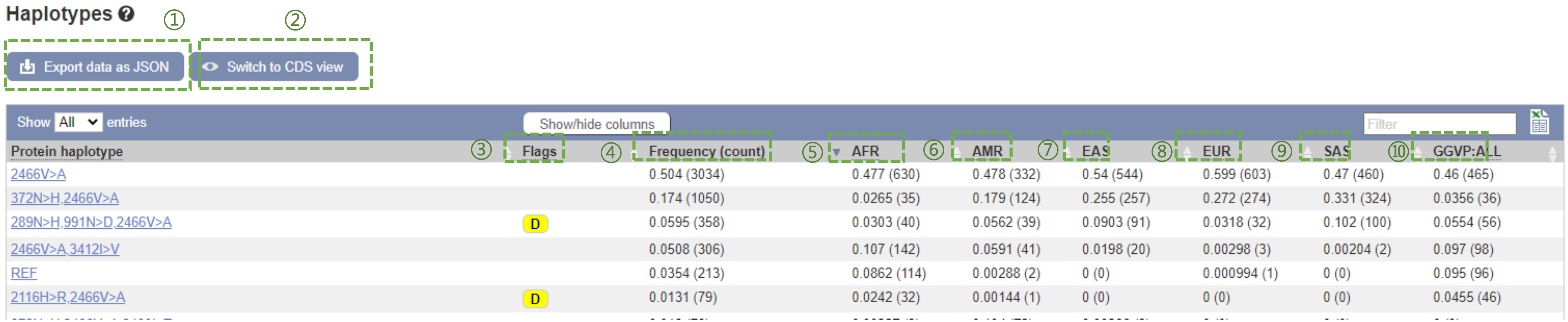
- 剪切受体变异
- 移码突变
- 框内插入突变
- 错义突变
- 同义突变
- 非编码转录外显子变异
- 基因间变异

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - 在新跳转界面的左侧导航栏中点击Haplotypes，其右侧内容展示区将如下图所示。



Transcript-based displays
- Summary
- Sequence
  - Exons
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
  - 3D Protein model
- Genetic Variation
  - Variant table
  - Variant image
  - Haplotypes
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history



Protein haplotype	Flags	Frequency (count)	AFR	AMR	EAS	EUR	SAS	GGVP:ALL
2466V>A		0.504 (3034)	0.477 (630)	0.478 (332)	0.54 (544)	0.599 (603)	0.47 (460)	0.46 (465)
372N>H,2466V>A		0.174 (1050)	0.0265 (35)	0.179 (124)	0.255 (257)	0.272 (274)	0.331 (324)	0.0356 (36)
289N>H,991N>D,2466V>A	D	0.0595 (358)	0.0303 (40)	0.0562 (39)	0.0903 (91)	0.0318 (32)	0.102 (100)	0.0554 (56)
2466V>A,3412I>V		0.0508 (306)	0.107 (142)	0.0591 (41)	0.0198 (20)	0.00298 (3)	0.00204 (2)	0.097 (98)
REF		0.0354 (213)	0.0862 (114)	0.00288 (2)	0 (0)	0.000994 (1)	0 (0)	0.095 (96)
2116H>R,2466V>A	D	0.0131 (79)	0.0242 (32)	0.00144 (1)	0 (0)	0 (0)	0 (0)	0.0455 (46)

①下载数据

②切换蛋白与CDS视图

③表示终止密码子改变或错义突变

④在1000个人群基因组中突变频率

⑤在非洲人群中突变频率

⑥在美国人群中突变频率

⑦东亚人群中突变频率

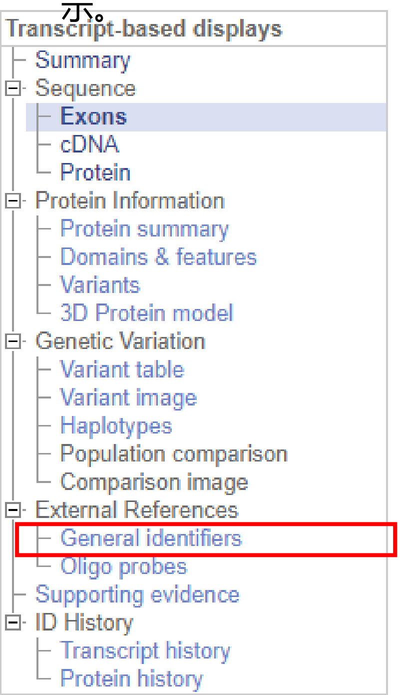
⑧欧洲人群中突变频率

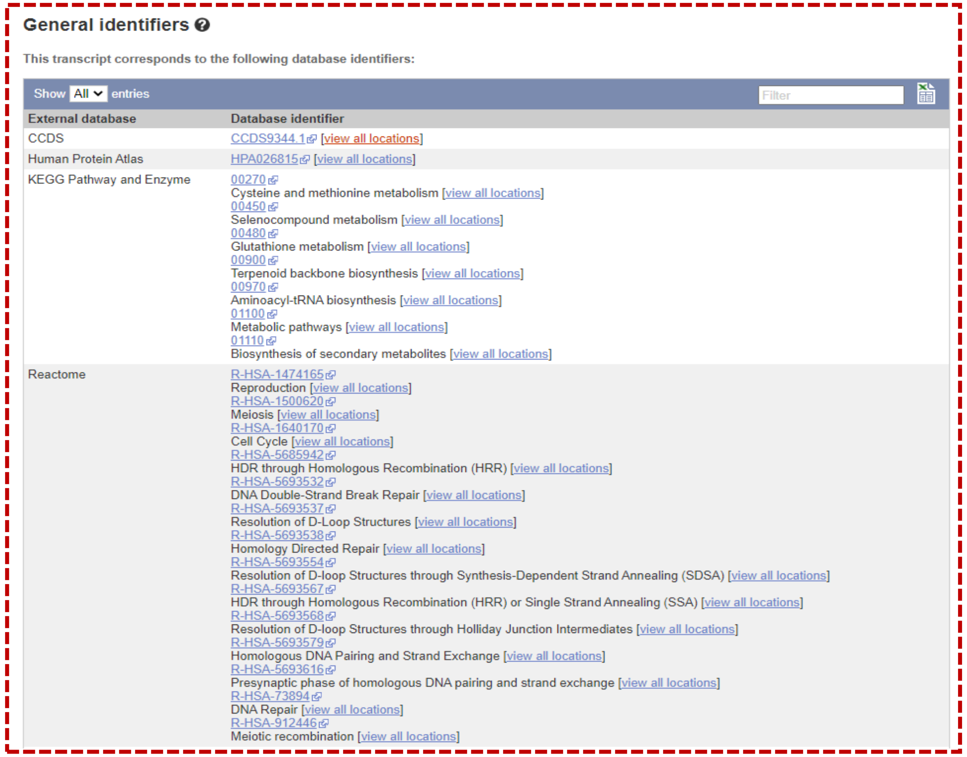
⑨南亚人群中突变频率

⑩冈比亚人种中突变频率

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - 在新跳转界面的左侧导航栏中分别点击General identifiers，其右侧内容展示区将如下图所示。

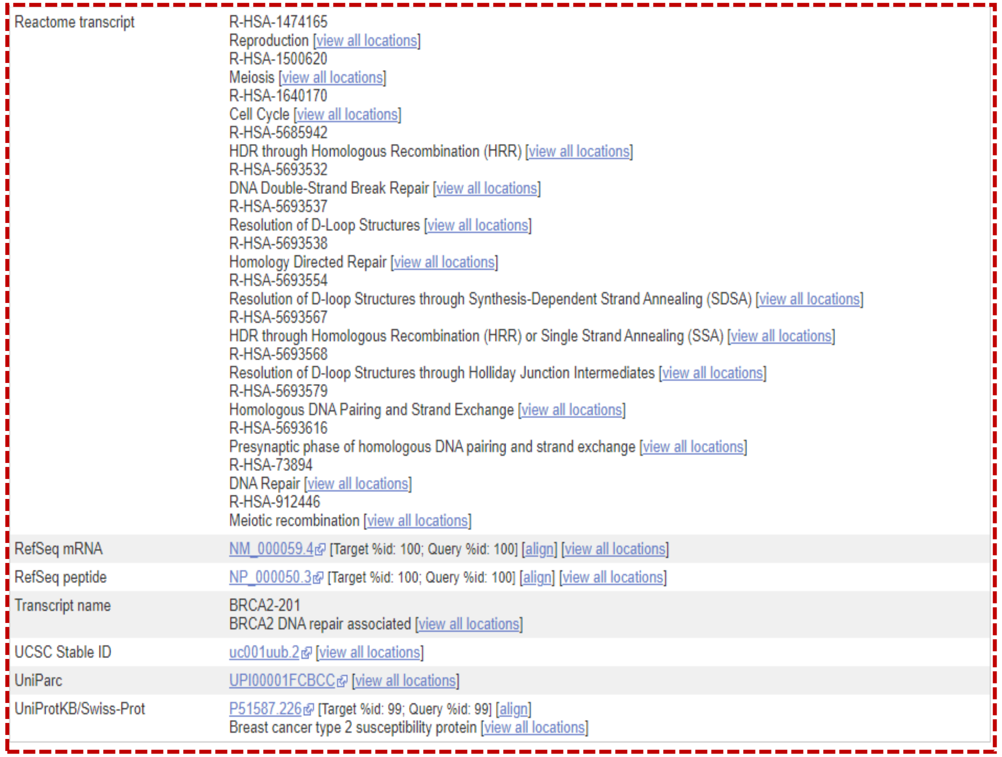




General identifiers

This transcript corresponds to the following database identifiers:

External database	Database identifier
CCDS	CCDS9344.1 [view all locations]
Human Protein Atlas	HPA026815 [view all locations]
KEGG Pathway and Enzyme	00270 Cysteine and methionine metabolism [view all locations] 00450 Selenocompound metabolism [view all locations] 00480 Glutathione metabolism [view all locations] 00900 Terpenoid backbone biosynthesis [view all locations] 00970 Aminoacyl-tRNA biosynthesis [view all locations] 01100 Metabolic pathways [view all locations] 01110 Biosynthesis of secondary metabolites [view all locations]
Reactome	R-HSA-1474165 Reproduction [view all locations] R-HSA-1500620 Meiosis [view all locations] R-HSA-1640170 Cell Cycle [view all locations] R-HSA-5685942 HDR through Homologous Recombination (HRR) [view all locations] R-HSA-5693532 DNA Double-Strand Break Repair [view all locations] R-HSA-5693537 Resolution of D-Loop Structures [view all locations] R-HSA-5693538 Homology Directed Repair [view all locations] R-HSA-5693554 Resolution of D-loop Structures through Synthesis-Dependent Strand Annealing (SDSA) [view all locations] R-HSA-5693567 HDR through Homologous Recombination (HRR) or Single Strand Annealing (SSA) [view all locations] R-HSA-5693568 Resolution of D-loop Structures through Holliday Junction Intermediates [view all locations] R-HSA-5693579 Homologous DNA Pairing and Strand Exchange [view all locations] R-HSA-5693616 Presynaptic phase of homologous DNA pairing and strand exchange [view all locations] R-HSA-73894 DNA Repair [view all locations] R-HSA-912446 Meiotic recombination [view all locations]



Reactome transcript	R-HSA-1474165 Reproduction [view all locations] R-HSA-1500620 Meiosis [view all locations] R-HSA-1640170 Cell Cycle [view all locations] R-HSA-5685942 HDR through Homologous Recombination (HRR) [view all locations] R-HSA-5693532 DNA Double-Strand Break Repair [view all locations] R-HSA-5693537 Resolution of D-Loop Structures [view all locations] R-HSA-5693538 Homology Directed Repair [view all locations] R-HSA-5693554 Resolution of D-loop Structures through Synthesis-Dependent Strand Annealing (SDSA) [view all locations] R-HSA-5693567 HDR through Homologous Recombination (HRR) or Single Strand Annealing (SSA) [view all locations] R-HSA-5693568 Resolution of D-loop Structures through Holliday Junction Intermediates [view all locations] R-HSA-5693579 Homologous DNA Pairing and Strand Exchange [view all locations] R-HSA-5693616 Presynaptic phase of homologous DNA pairing and strand exchange [view all locations] R-HSA-73894 DNA Repair [view all locations] R-HSA-912446 Meiotic recombination [view all locations]
RefSeq mRNA	NM_000059.4 [Target %id: 100; Query %id: 100] [align] [view all locations]
RefSeq peptide	NP_000050.3 [Target %id: 100; Query %id: 100] [align] [view all locations]
Transcript name	BRCA2-201 BRCA2 DNA repair associated [view all locations]
UCSC Stable ID	uc001uub.2 [view all locations]
UniParc	UPI00001FCBCC [view all locations]
UniProtKB/Swiss-Prot	P51587.226 [Target %id: 99; Query %id: 99] [align] Breast cancer type 2 susceptibility protein [view all locations]

- 此导航连接到其他外部数据库，点击每个数据库对应的identifier能够跳转到相应数据库查看详细内容

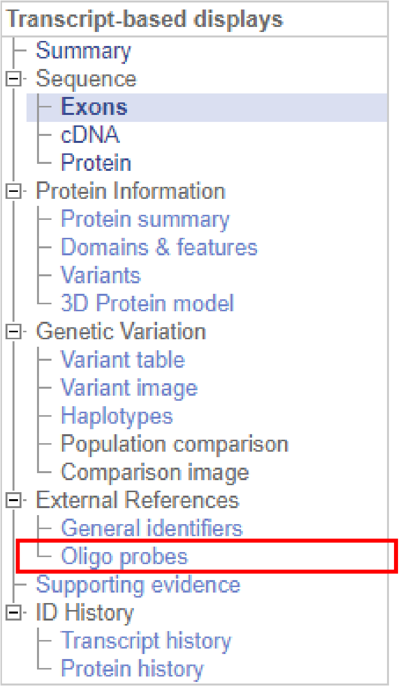
- CCDS数据库
- Human protein Atlas数据库
- KEGG数据库
- Reactome数据库

- Reactome transcript数据库
- Refseq mRNA数据库
- Refseq peptide数据库

- 转录本名称
- USCS固定ID
- Uniparc数据库信息
- UniProtKB/Swiss-Prot信息

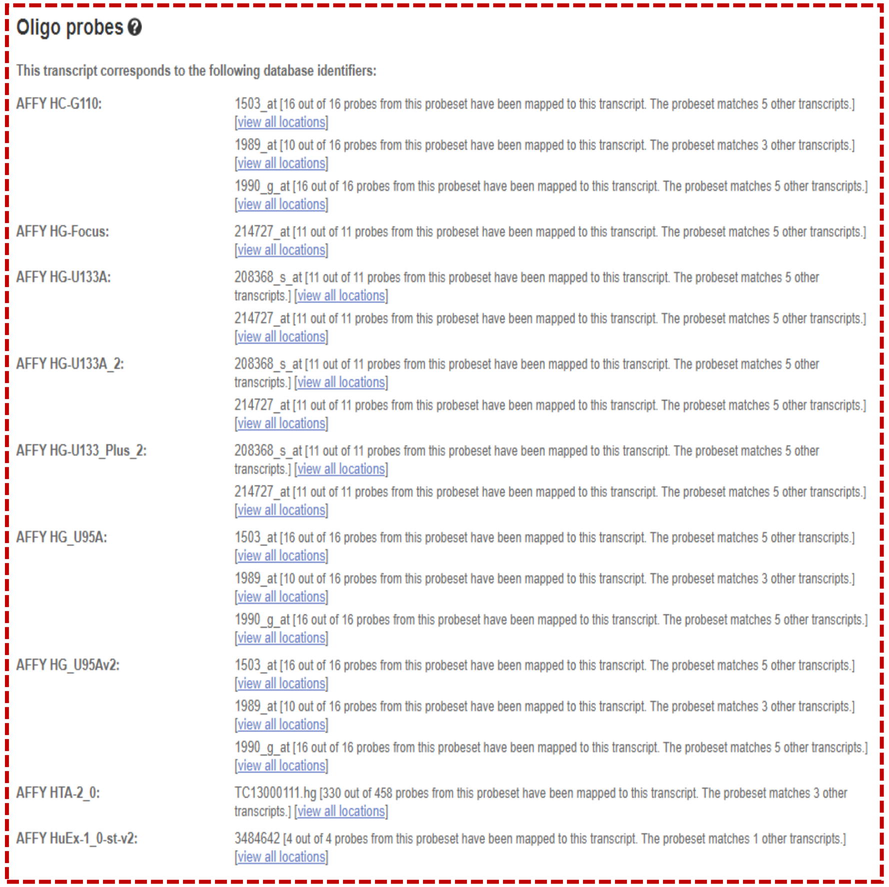
Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - 在新跳转界面的左侧导航栏中分别点击Oligo probes，其右侧内容展示区将如下图所示。



Transcript-based displays

- Summary
- Sequence
  - Exons
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
  - 3D Protein model
- Genetic Variation
  - Variant table
  - Variant image
  - Haplotypes
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history



Oligo probes

This transcript corresponds to the following database identifiers:

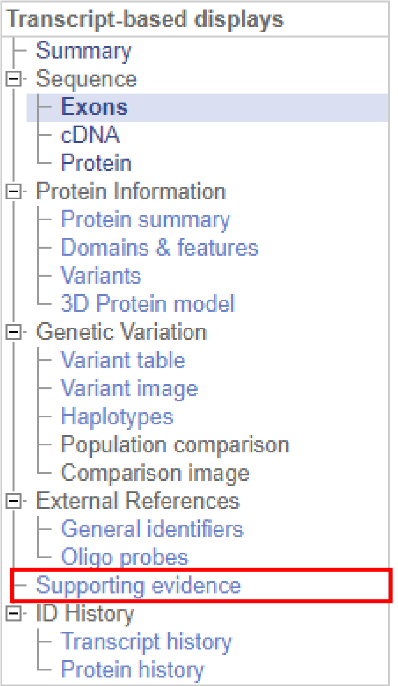
AFFY HC-G110:	1503_at [16 out of 16 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
	1989_at [10 out of 16 probes from this probeset have been mapped to this transcript. The probeset matches 3 other transcripts.] [view all locations]
	1990_g_at [16 out of 16 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
AFFY HG-Focus:	214727_at [11 out of 11 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
AFFY HG-U133A:	208368_s_at [11 out of 11 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
	214727_at [11 out of 11 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
AFFY HG-U133A_2:	208368_s_at [11 out of 11 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
	214727_at [11 out of 11 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
AFFY HG-U133_Plus_2:	208368_s_at [11 out of 11 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
	214727_at [11 out of 11 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
AFFY HG_U95A:	1503_at [16 out of 16 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
	1989_at [10 out of 16 probes from this probeset have been mapped to this transcript. The probeset matches 3 other transcripts.] [view all locations]
	1990_g_at [16 out of 16 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
AFFY HG_U95Av2:	1503_at [16 out of 16 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
	1989_at [10 out of 16 probes from this probeset have been mapped to this transcript. The probeset matches 3 other transcripts.] [view all locations]
	1990_g_at [16 out of 16 probes from this probeset have been mapped to this transcript. The probeset matches 5 other transcripts.] [view all locations]
AFFY HTA-2_0:	TC13000111.hg [330 out of 458 probes from this probeset have been mapped to this transcript. The probeset matches 3 other transcripts.] [view all locations]
AFFY HuEx-1_0-st-v2:	3484642 [4 out of 4 probes from this probeset have been mapped to this transcript. The probeset matches 1 other transcripts.] [view all locations]

Oligo探针信息

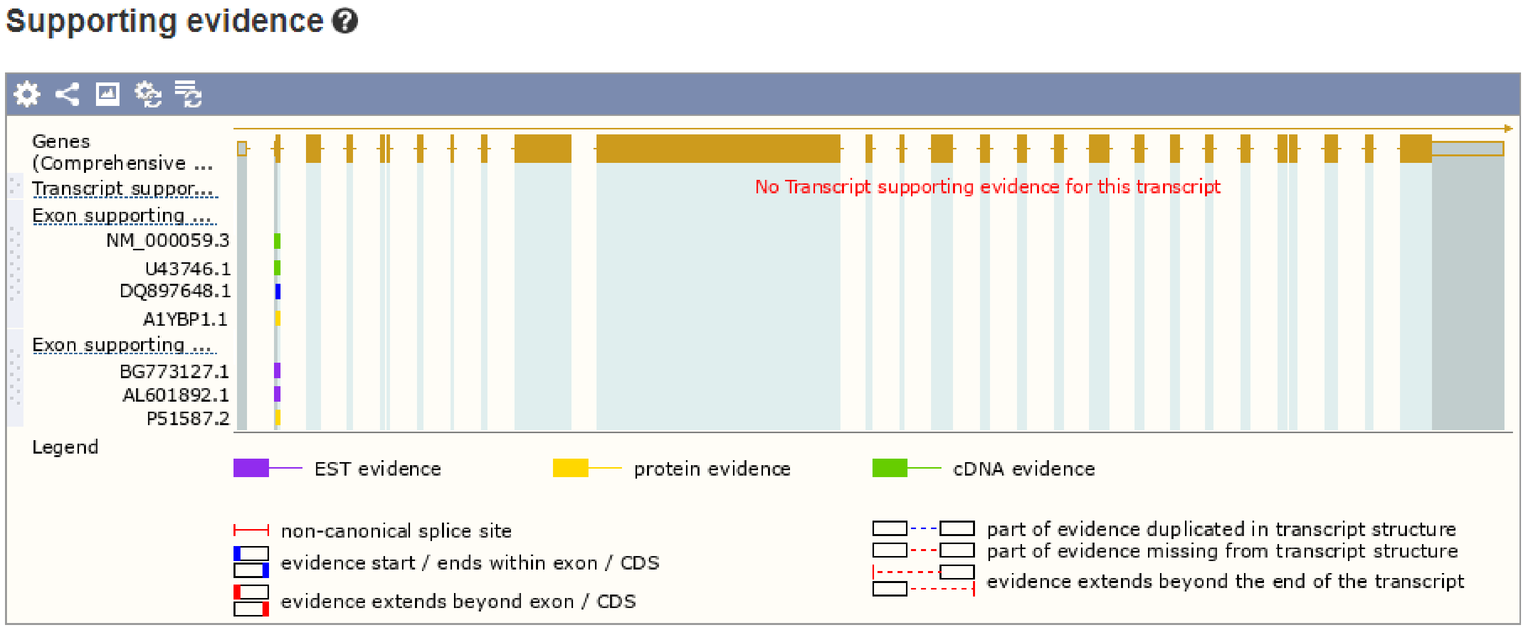
展示基于测序平台的探针位点信息

Ensembl检索方法---关键词检索

- BRCA2的基因信息结果界面——转录本信息BRCA2-201
  - ➢ 在新跳转界面的左侧导航栏中点击Supporting evidence，其右侧内容展示区将如下图所示。



Transcript-based displays
- Summary
- Sequence
  - **Exons**
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
  - 3D Protein model
- Genetic Variation
  - Variant table
  - Variant image
  - Haplotypes
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history



Supporting evidence

Genes (Comprehensive ...
Transcript suppor...
Exon supporting ...
NM_000059.3
U43746.1
DQ897648.1
A1YBP1.1
Exon supporting ...
BG773127.1
AL601892.1
P51587.2

No Transcript supporting evidence for this transcript

Legend

EST evidence	protein evidence	cDNA evidence

non-canonical splice site
evidence start / ends within exon / CDS
evidence extends beyond exon / CDS
part of evidence duplicated in transcript structure
part of evidence missing from transcript structure
evidence extends beyond the end of the transcript

该转录本没有支持依据

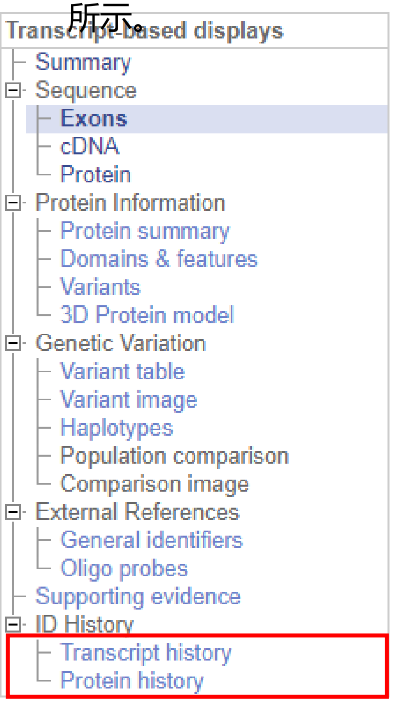
- 基因
- 转录本支持证据
- 外显子支持证据

图标代表含义：

- 黄色代表蛋白支持证据
- 红线代表：非经典剪切位点
- 蓝色开始或结尾方框代表：含外显子/CDS起点或终止位点
- 红色开始或结尾方框代表：外显子/CDS之上扩展区
- 绿色：cDNA证据
- 紫色：EST证据

Ensembl检索方法---关键词检索

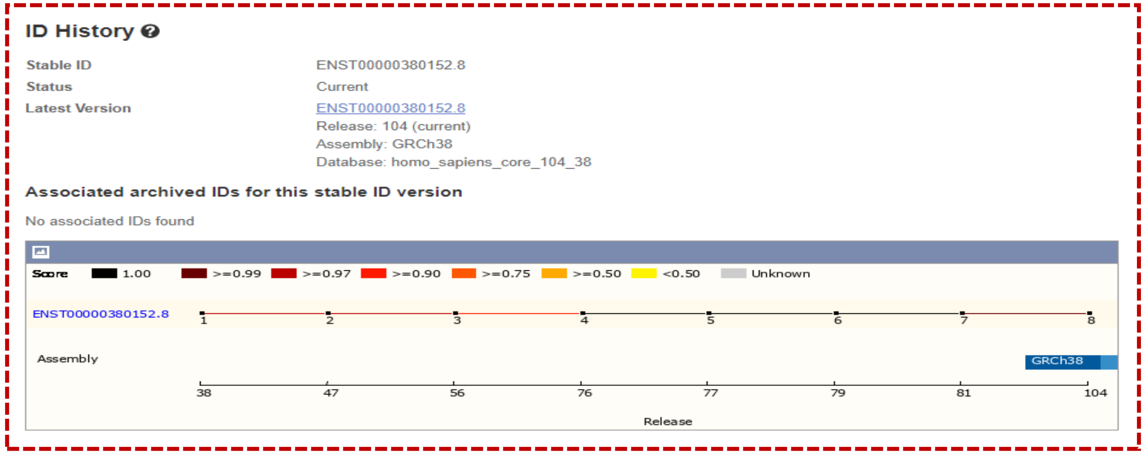
- BRCA2的基因信息结果界面
  - ➢ 点击Transcript table中的第一个转录本信息BRCA2-201，在新跳转的界面（P25）左侧导航栏中点击Transcript history/Protein history，其右侧内容展示区将如下图所示。



Transcript-based displays

- Summary
- Sequence
  - Exons
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
  - 3D Protein model
- Genetic Variation
  - Variant table
  - Variant image
  - Haplotypes
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history

①



ID History

Stable ID ENST00000380152.8
Status Current
Latest Version ENST00000380152.8
Release: 104 (current)
Assembly: GRCh38
Database: homo_sapiens_core_104_38

Associated archived IDs for this stable ID version

No associated IDs found

②

ID History

Stable ID ENSP00000369497.3
Status Current
Latest Version ENSP00000369497.3
Release: 104 (current)
Assembly: GRCh38
Database: homo_sapiens_core_104_38

Associated archived IDs for this stable ID version

Release	Gene	Transcript	Protein
75	ENSG00000139618	ENST00000380152.3	ENSP00000369497 MPIGSKERPTFFEIFKTRCNKADLGPISLNWFEELSSEAPPYNSEPAEESEHKNNNYEPN LFKTPQRKPSYNQLASTPIIFKEQGLTLPLYQSPVKELDKFKLDLGRNVPNSRHKSLRTV KTKMDQADDVSCPLLNSCLSESPVVLQCTHVTPQRDKSVVCGSLFHTPKFVKGRQTPKHI SESLGAEVDPDMSWSSSLATPPTLSSTVLIVRNEEASETVFPHDTTANVKSYFSNHDESL KKNDRFIASVTDSENTNQREAASHGFGKTSGNSFKVNSCKDHIGKSMPNVLEDEVYETVV DTSEEDSFSLCFSKCRTKNLQKVRTSKTRKKIFHEANADECEKSKNQVKEKYSFVSEVEP NDTDPLDSNVANQKPFESGSDKISKEVVPSLACEWSQLTLSGLNGAQMEKIPLLHISSCD QNISEKDLLDTENKRKKDFLTSENSLPRISSLPKSEKPLNEETVVNKRDEEQHLESHTDC ILAVKQAISGTSPVASSFQGIKKSIFRIRESPKETFNASFSGHMTDPNFKKETEASESGL EIHTVCSQKEDSLCPNLIDNGSWPATTTQNSVALKNAGLISTLKKKTNKFIYAIHDETSY KGKKIPKDQKSELINCSAQFEANAFEAPLTFANADSGLLHSSVKRSCSQNDSEEPTLSLT SSFGTILRKCSRNETCSNNTVISQDLDYKEAKCNKEKLQLFITPEADSLSCLQEGQCEND

① Transcript history 转录本历史

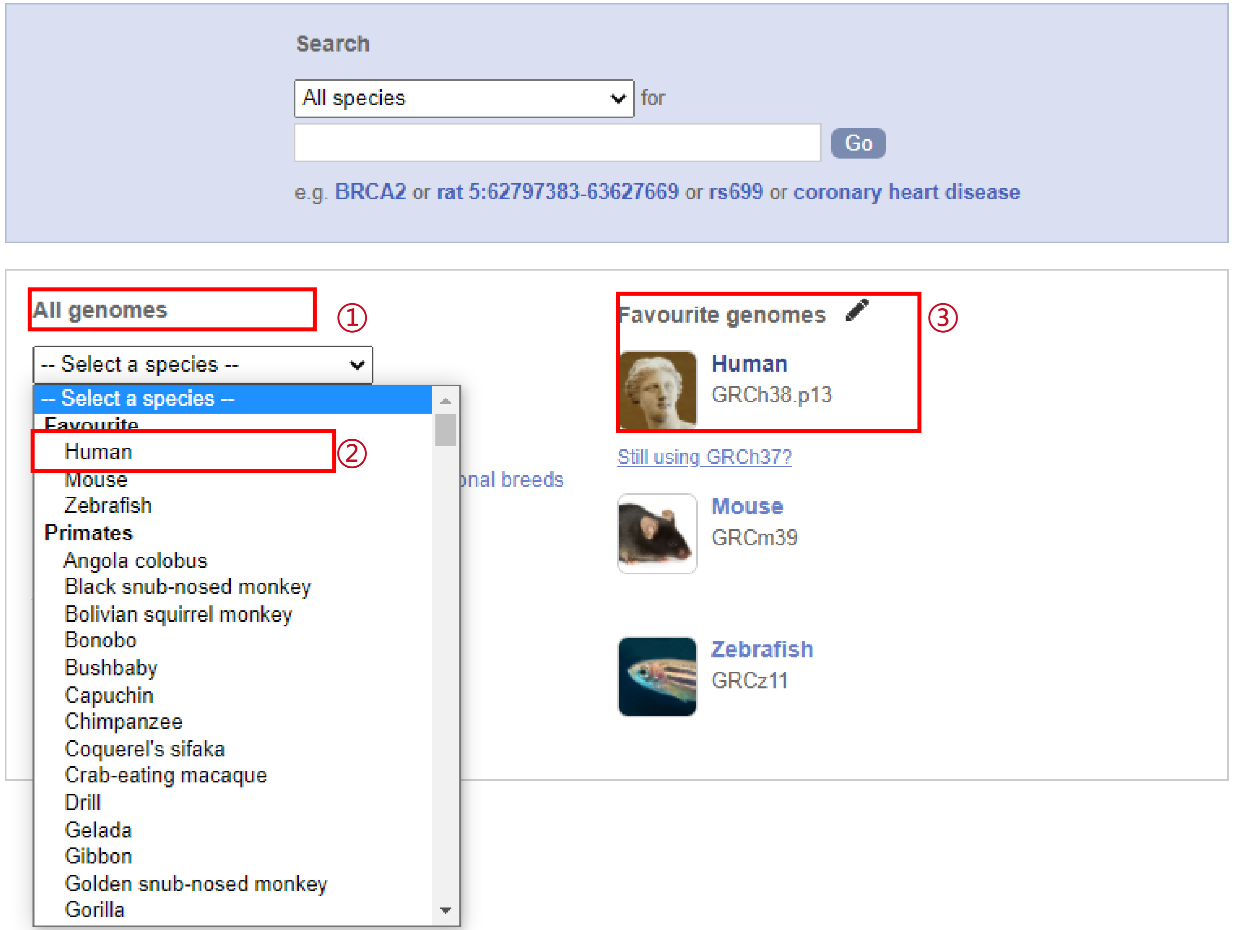
转录本固定ID
状态
最新版本
该固定ID版本相关的IDs

②Protein history 蛋白历史

蛋白固定ID
状态
最新版本
该固定ID版本相关的IDs

Ensembl检索方法---物种检索

- 以种属人为例，在Ensembl数据库主页面的all genomes下拉框中选择所需要的物种②，或者点击Favourite genomes中点击Human③，即可进行相关检索。



Search

All species for

Go

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes ①

-- Select a species --

-- Select a species --

Favourite

Human ②

Mouse

Zebrafish

Primates

Angola colobus

Black snub-nosed monkey

Bolivian squirrel monkey

Bonobo

Bushbaby

Capuchin

Chimpanzee

Coquerel's sifaka

Crab-eating macaque

Drill

Gelada

Gibbon

Golden snub-nosed monkey

Gorilla

onal breeds

Favourite genomes ③

Human
GRCh38.p13

Still using GRCh37?

Mouse
GRCm39

Zebrafish
GRCz11

Ensembl检索方法---物种检索

- 物种检索的结果界面如下图所示。

Human (GRCh38.p13) ▾

Search Human (Homo sapiens)

e.g. BRCA2 or 17:63992802-64038237 or rs699 or osteoarthritis

Genome assembly: GRCh38.p13 (GCA_000001405.28) • 版本信息

More information and statistics • 更多信息
Download DNA sequence (FASTA) • 下载DNA序列
Convert your data to GRCh38 coordinates • 版本转换
Display your data in Ensembl • 在Ensembl中显示个人数据

Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart

Comparative genomics • 比较基因组学

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

More about comparative analysis
Download alignments (EMF)

Regulation • 调控信息

What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and microarray annotations.

• 包括DNA甲基化、转录因子结合位点、组蛋白修饰、增强子等

More about the Ensembl regulatory build and microarray annotation
Experimental data sources
Download all regulatory features (GFF)

Gene annotation • 对人的基因多态性位点进行注释

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

More about this genebuild • 更多信息
Download FASTA files for genes, cDNAs, ncRNA, proteins
• 下载不同格式的基因注释信息
Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins
Update your old Ensembl IDs • 升级旧版本

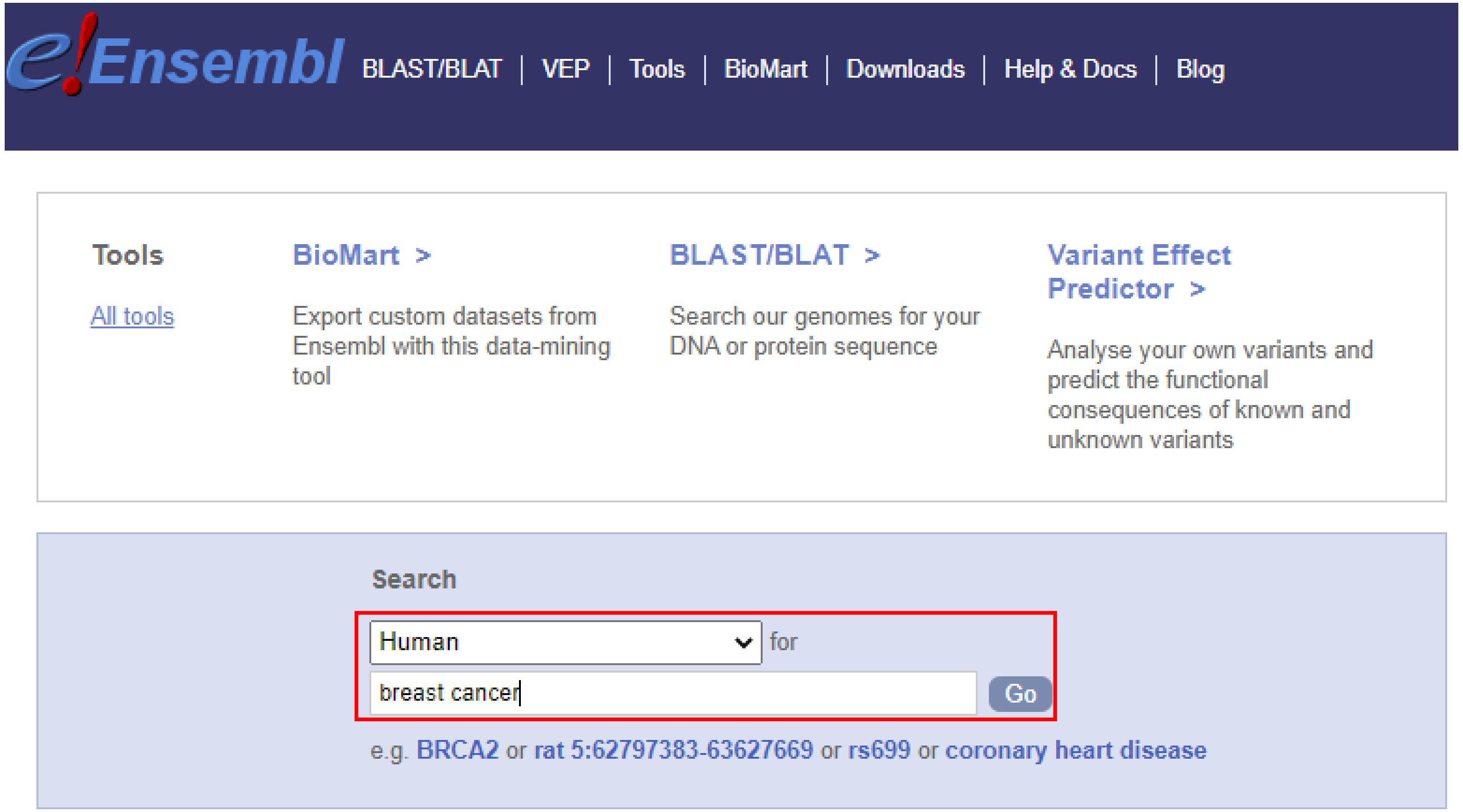
Variation • 变异

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes

More about variation in Ensembl • 更多变异相关信息
Download all variants (GVF) • 下载所有变异
Variant Effect Predictor • 变异效应预测

Ensembl检索方法---疾病名检索

- 以人乳腺癌为例，在Ensembl数据库主页面的search框中输入疾病名称breast cancer，点击GO，即可进行相关检索。



Export custom datasets from Ensembl with this data-mining tool

Search our genomes for your DNA or protein sequence

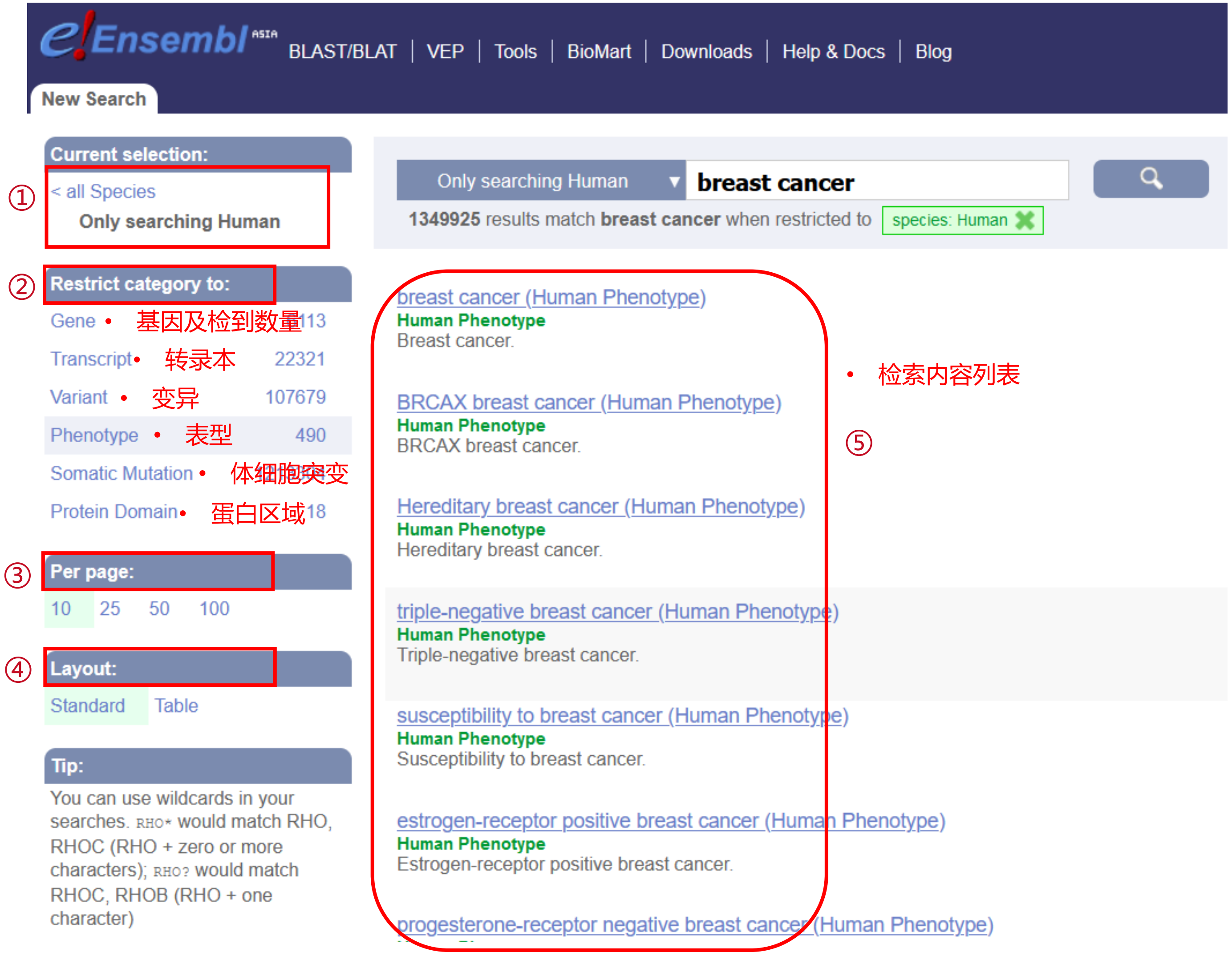
Analyse your own variants and predict the functional consequences of known and unknown variants

breast cancer

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

Ensembl检索方法---疾病名检索

■ 人类乳腺癌的表型结果界面



Gene •	基因及检到数量	113
Transcript•	转录本	22321
Variant •	变异	107679
Phenotype •	表型	490
Somatic Mutation •	体细胞突变	
Protein Domain•	蛋白区域	18

1349925 results match **breast cancer** when restricted to species: Human

Tip:
You can use wildcards in your searches. `RHO*` would match RHO, RHOC (RHO + zero or more characters); `RHO?` would match RHOC, RHOB (RHO + one character)

breast cancer (Human Phenotype)
Human Phenotype
Breast cancer.

BRCAX breast cancer (Human Phenotype)
Human Phenotype
BRCAX breast cancer.

Hereditary breast cancer (Human Phenotype)
Human Phenotype
Hereditary breast cancer.

triple-negative breast cancer (Human Phenotype)
Human Phenotype
Triple-negative breast cancer.

susceptibility to breast cancer (Human Phenotype)
Human Phenotype
Susceptibility to breast cancer.

estrogen-receptor positive breast cancer (Human Phenotype)
Human Phenotype
Estrogen-receptor positive breast cancer.

progesterone-receptor negative breast cancer (Human Phenotype)

①当前结果：显示人类中信息

②结果分类

③调整每页显示结果数

④页面展示切换

⑤检索结果，点击每一条可跳转到详细信息

Ensembl检索方法---疾病名检索

- 人类乳腺癌的表型结果界面
  - 在检索结果列表中点击第一行breast cancer (Human phenotype)后结果如下

Loci associated with breast cancer (MONDO:0007254)

① Name(s)	② Type	③ Genomic location (strand)	④ Reported gene(s)	⑤ Phenotype/Disease/Trait	⑥ Annotation source	Submitter	External reference	Supporting evidence
rs60226654	Variant	**14**:103585480-103585494 (+)	-	Breast Cancer	NHGRI-EBI GWAS catalog	-	PMID:29059683	-
rs113577745	Variant	**2**:9995553 (+)	GRHL1	Breast Cancer	NHGRI-EBI GWAS catalog	-	PMID:29059683	-
rs9262142	Variant	**CHR_HSCHR6_MHC_COX_CTG1**:30672112 (+)	-	Breast Cancer	NHGRI-EBI GWAS catalog	-	PMID:29059683	-
rs7430456	Variant	**3**:177768082 (+)	-	Breast Cancer	NHGRI-EBI GWAS catalog	-	PMID:29059683	-
rs6001982	Variant	**22**:40623944 (+)	-	Breast Cancer	NHGRI-EBI GWAS catalog	-	PMID:29059683	-
rs746598992	Variant	**11**:108281133-108281135 (+)	ATM	Breast Cancer	ClinVar	-	-	-
nsv4683113	Structural Variant	**16**:23629194-23641167 (+)	-	Breast Cancer	dbVar	-	-	-
rs35419901	Variant	**10**:62303692 (+)	-	Breast Cancer	NHGRI-EBI GWAS catalog	-	PMID:29059683	-
rs458685	Variant	**21**:29805194 (+)	GRIK1	Breast Cancer	NHGRI-EBI GWAS catalog	-	PMID:17903305	-
rs11814448	Variant	**10**:22026914 (+)	DNAJC1	Breast Cancer	NHGRI-EBI GWAS catalog	-	PMID:23535729	-
rs62079673	Variant	**17**:82597601 (+)	-	Breast Cancer	NHGRI-EBI GWAS catalog	-	PMID:29059683	-
rs56196003	Variant	**CHR_HSCHR7_1_CTG4_4**:103143072 (-)	-	Breast Cancer	NHGRI-EBI GWAS catalog	-	PMID:29059683	-
rs62070652	Variant	**17**:30894259 (+)	-	Breast Cancer	NHGRI-EBI GWAS	-	PMID:3288788	-

①位点名
②类型
③基因定位
④已报道基因
⑤疾病表型
⑥注释信息来源

Ensembl检索方法---BLAST/BLAT序列相似性查找

在浏览器打开Ensembl数据库的主页面，点击上方BLAST/BLAT，跳转至以下界面，点击New job即可进入Blast界面。这里以一段人源性序列为例进行实际操作。

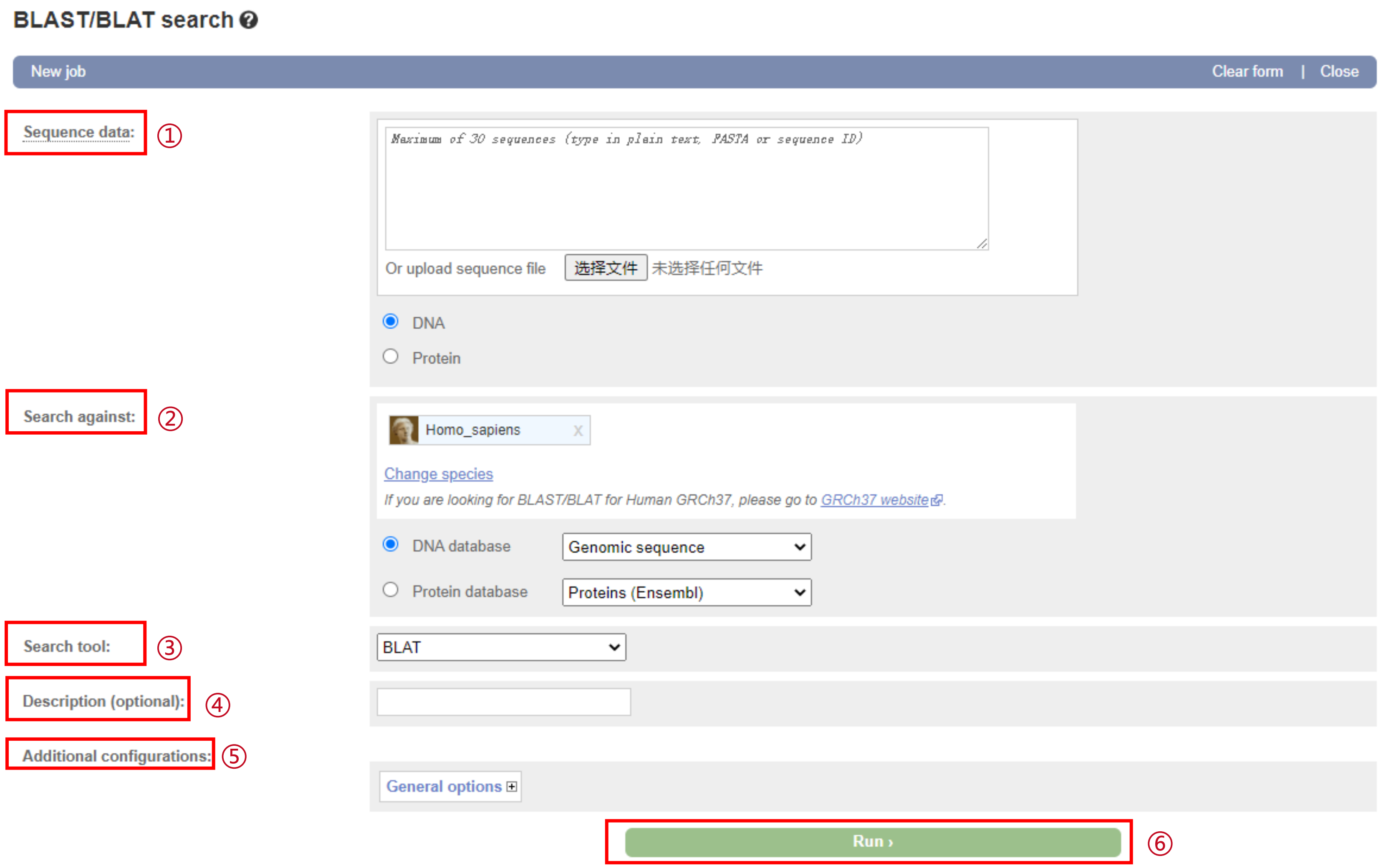
BLAST/BLAT search

Analysis	Jobs			Submitted at
BLAT	ENSG00000139618 Ensembl Gene chromosome:GRCh38:13:1:114364328:1	Done: 100 hits found	[View results]	08/12/2021, 09:58 (GMT)

Ensembl release 105 - Dec 2021 © EMBL-EBI

Ensembl检索方法---BLAST/BLAT序列相似性查找

Ensembl的Blast界面如下图所示。



BLAST/BLAT search

New job — Clear form | Close

Sequence data: ①
Maximum of 30 sequences (type in plain text, FASTA or sequence ID)
Or upload sequence file 选择文件 未选择任何文件
● DNA
○ Protein

Search against: ②
Homo_sapiens
Change species
If you are looking for BLAST/BLAT for Human GRCh37, please go to GRCh37 website.
● DNA database: Genomic sequence
○ Protein database: Proteins (Ensembl)

Search tool: ③ BLAT

Description (optional): ④

Additional configurations: ⑤ General options

Run › ⑥

BLAST页面

①序列信息输入

在右侧方框中输入要求的序列或文件，选择序列类型（DNA或蛋白序列）

后续会以人源序列（ENST00000672860.3）进行实操演示

②搜索范围

③搜索工具

- BLAT/BLASTN: 核酸序列与核酸数据库比对
- TBLASTX:核酸序列与翻译成蛋白的核苷酸数据库比对
- BLASTX:核苷酸序列翻译成蛋白质后与蛋白质数据库比对
- BLASTP: 氨基酸序列与蛋白数据库比对
- TBLASTN：氨基酸序列与核苷酸数据库中核酸序列翻译后的蛋白序列比对

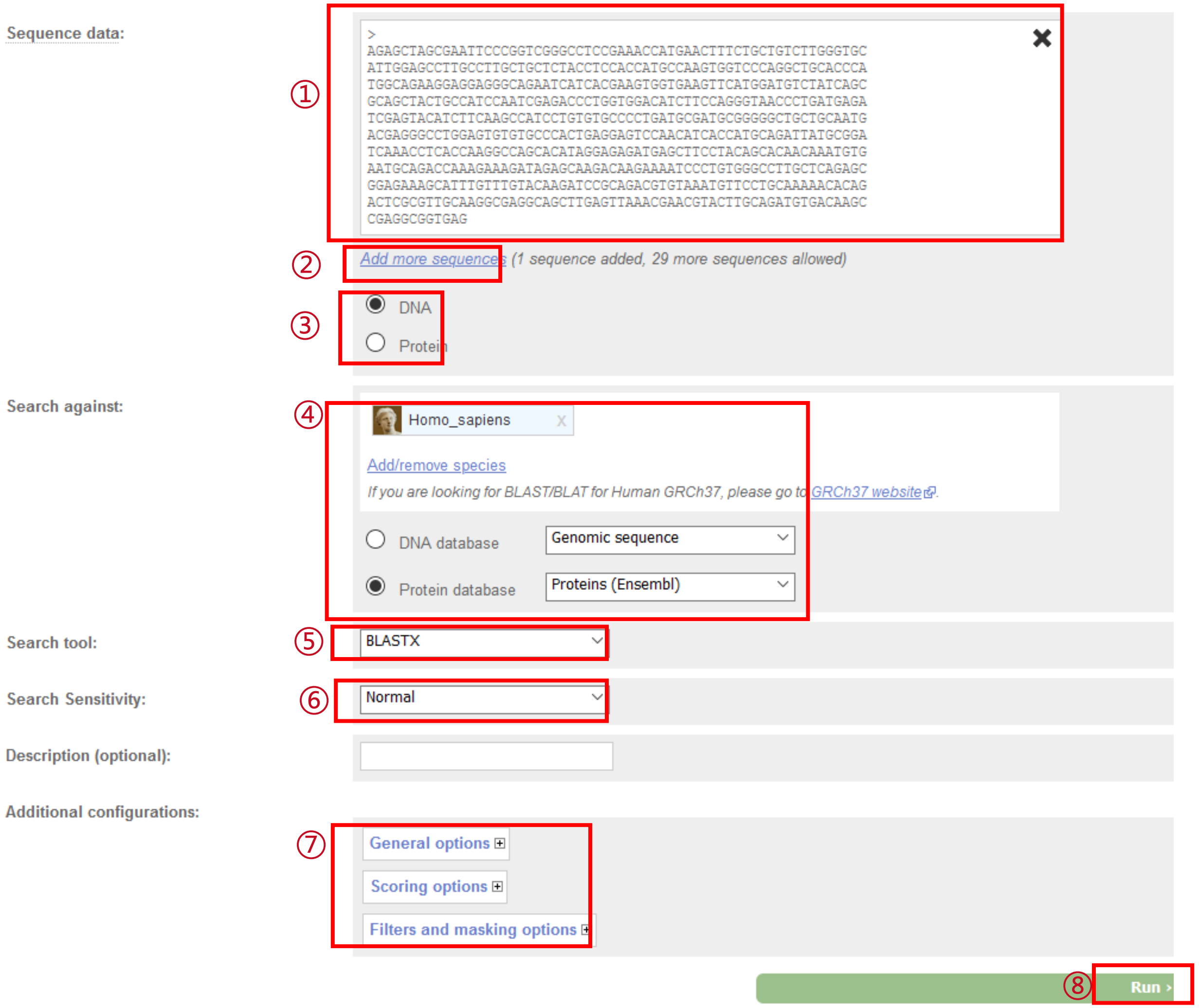
④描述

⑤其他设置（详见下页PPT）

⑥运行按钮

Ensembl检索方法---BLAST/BLAT序列相似性查找

以一段人源性序列为例进行实际操作。



Sequence data:

```
>
AGAGCTAGCGAATTCCCGGTCGGGCCTCCGAAACCATGAACTTTCTGCTGTCTTGGGTGC
ATTGGAGCCTTGCCTTGCTGCTCTACCTCCACCATGCCAAGTGGTCCCAGGCTGCACCCA
TGGCAGAAGGAGGAGGGCAGAATCATCACGAAGTGGTGAAGTTCATGGATGTCTATCAGC
GCAGCTACTGCCATCCAATCGAGACCCTGGTGGACATCTTCCAGGGTAACCCTGATGAGA
TCGAGTACATCTTCAAGCCATCCTGTGTGCCCCTGATGCGATGCGGGGGCTGCTGCAATG
ACGAGGGCCTGGAGTGTGTGCCCACTGAGGAGTCCAACATCACCATGCAGATTATGCGGA
TCAAACCTCACCAAGGCCAGCACATAGGAGAGATGAGCTTCCTACAGCACAACAAATGTG
AATGCAGACCAAAGAAAGATAGAGCAAGACAAGAAAATCCCTGTGGGCCTTGCTCAGAGC
GGAGAAAGCATTTGTTTGTACAAGATCCGCAGACGTGTAAATGTTCCTGCAAAAACACAG
ACTCGCGTTGCAAGGCGAGGCAGCTTGAGTTAAACGAACGTACTTGCAGATGTGACAAGC
CGAGGCGGTGAG
```

Add more sequences (1 sequence added, 29 more sequences allowed)

● DNA
○ Protei

Search against:

Homo_sapiens x

Add/remove species

If you are looking for BLAST/BLAT for Human GRCh37, please go to GRCh37 website.

○ DNA database: Genomic sequence
● Protein database: Proteins (Ensembl)

Search tool: BLASTX

Search Sensitivity: Normal

Description (optional):

Additional configurations:

General options ⊞
Scoring options ⊞
Filters and masking options ⊞

Run ›

①粘贴序列信息：可以为FASTA格式、plain格式等,此次我们利用一段人源性序列进行比对

②可以添加多条序列信息，最多30条

③选择以上序列为DNA序列还是蛋白信息

④选择物种及DNA/蛋白数据库

DNA数据库：Genomic sequence、Genomic sequence (hard masked)、Genomic sequence (soft masked)、cDNAs (transcripts/splice variant)、Ab-initio cDNAs (Genscan/SNAP)、Ensembl Non-coding RNA genes

蛋白数据库：Proteins (GENCODE/Ensembl)、Ab-initio peptides (Genscan/SNAP)

⑤检索工具

BLAT/BLASTN: 核酸序列与核酸数据库比对

TBLASTX:核酸序列与翻译成蛋白的核苷酸数据库比对

BLASTX:核苷酸序列翻译成蛋白质后与蛋白质数据库比对

BLASTP: 氨基酸序列与蛋白数据库比对

TBLASTN：氨基酸序列与核苷酸数据库中核酸序列翻译后的蛋白序列比对

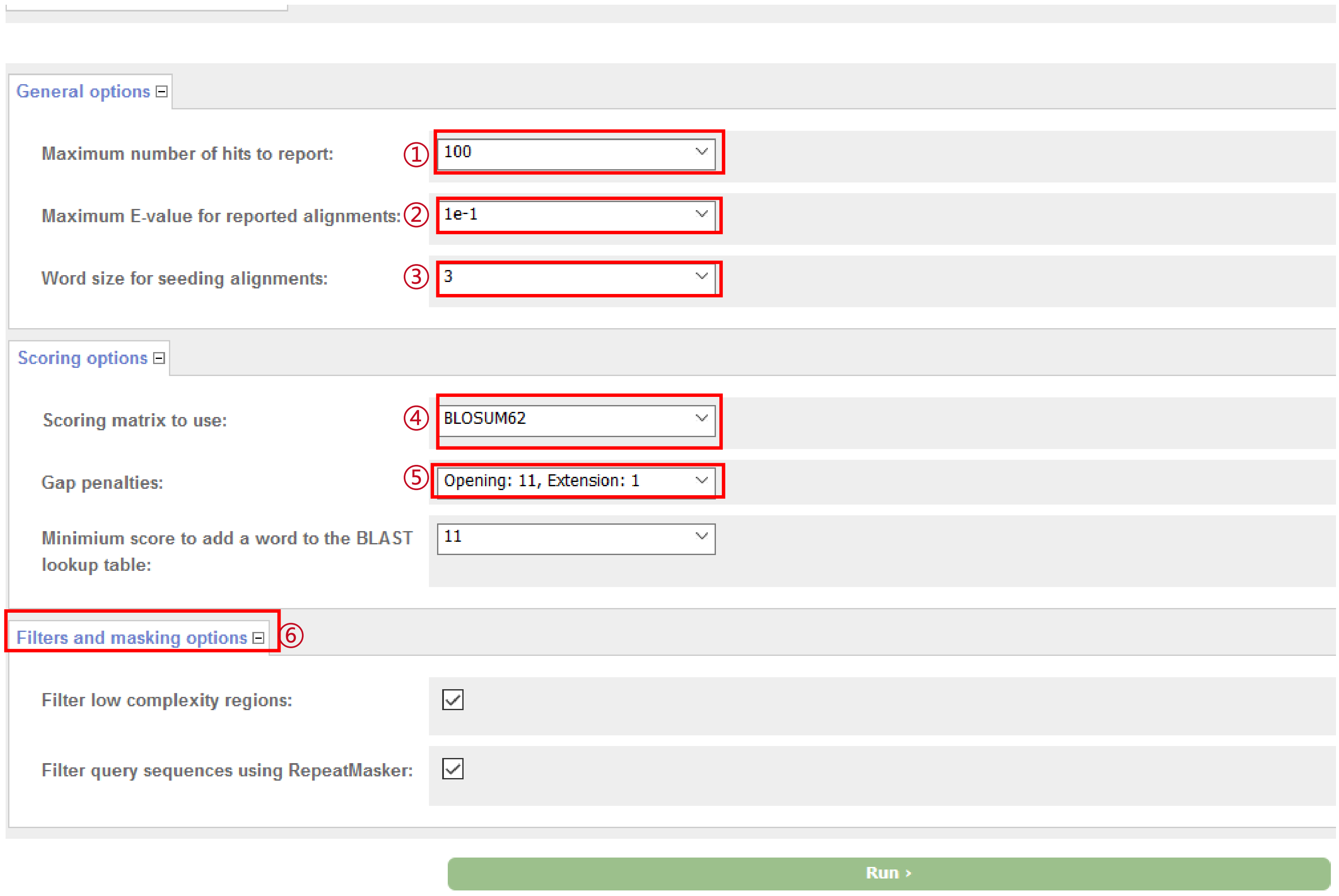
⑥检索敏感度设置

⑦更多的设置可以在additional configurations中设置，见下页PPT

⑧运行

Ensembl检索方法---BLAST/BLAT序列相似性查找

Ensembl的Blast界面如下图所示——接上页PPT。



General options ⊟

Maximum number of hits to report: ① 100

Maximum E-value for reported alignments: ② 1e-1

Word size for seeding alignments: ③ 3

Scoring options ⊟

Scoring matrix to use: ④ BLOSUM62

Gap penalties: ⑤ Opening: 11, Extension: 1

Minimium score to add a word to the BLAST lookup table: 11

Filters and masking options ⊟ ⑥

Filter low complexity regions: ☑

Filter query sequences using RepeatMasker: ☑

①显示的最大的序列数

②E-value：Blast程序在搜索空间中可随机找到获得这样高分的序列的可能性，因此E-value越高，代表结果越有可能是随机获得的，也就越不可信，一般设定为1即可。

③Blast算法将查询序列分割成一系列具有字段长度的小的序列段进行数据库搜索，因此当此值越小得到的搜索结果越多，但假阳性也越多。蛋白可设置2或3；核酸可设置为7、11或15。

④matrix 评价成对序列比对质量的一个关键因素，它为对齐任何可能的残基对分配一个分数。 Blast搜索中使用的矩阵可以根据搜索的序列类型进行更改。

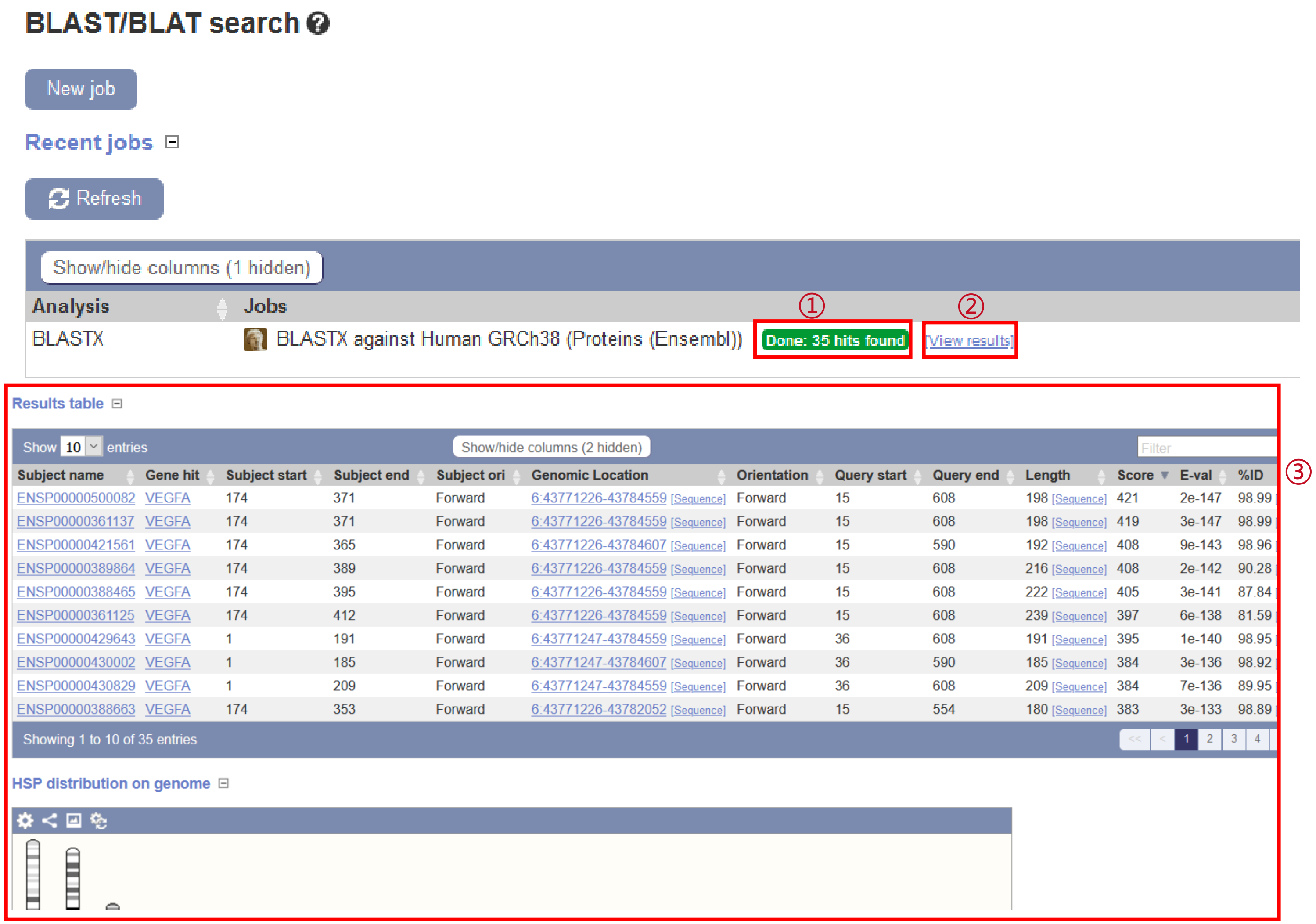
⑤Gap penalties 在比对中创建和扩大间距成本。下拉菜单显示所选矩阵的差距成本，这些参数只能有有限数量的选项， 增加差距成本将导致对齐方式的减少，从而减少引入差距数量

⑥过滤屏蔽

低成分复杂度的遮盖区域可能导致虚假或误导性的结果

Ensembl检索方法---BLAST/BLAT序列相似性查找

- 以人源序列（ENST00000672860.3） 为例进行蛋白比对，以确定该查询序列是哪种功能基因的同源基因，点击上页BLAST界面中的Run，结果页面见下



BLAST/BLAT search

Analysis	Jobs		
BLASTX	BLASTX against Human GRCh38 (Proteins (Ensembl))	① Done: 35 hits found	② [View results]

Results table

③

Subject name	Gene hit	Subject start	Subject end	Subject ori	Genomic Location	Orientation	Query start	Query end	Length	Score	E-val	%ID
ENSP00000500082	VEGFA	174	371	Forward	6:43771226-43784559 [Sequence]	Forward	15	608	198 [Sequence]	421	2e-147	98.99
ENSP00000361137	VEGFA	174	371	Forward	6:43771226-43784559 [Sequence]	Forward	15	608	198 [Sequence]	419	3e-147	98.99
ENSP00000421561	VEGFA	174	365	Forward	6:43771226-43784607 [Sequence]	Forward	15	590	192 [Sequence]	408	9e-143	98.96
ENSP00000389864	VEGFA	174	389	Forward	6:43771226-43784559 [Sequence]	Forward	15	608	216 [Sequence]	408	2e-142	90.28
ENSP00000388465	VEGFA	174	395	Forward	6:43771226-43784559 [Sequence]	Forward	15	608	222 [Sequence]	405	3e-141	87.84
ENSP00000361125	VEGFA	174	412	Forward	6:43771226-43784559 [Sequence]	Forward	15	608	239 [Sequence]	397	6e-138	81.59
ENSP00000429643	VEGFA	1	191	Forward	6:43771247-43784559 [Sequence]	Forward	36	608	191 [Sequence]	395	1e-140	98.95
ENSP00000430002	VEGFA	1	185	Forward	6:43771247-43784607 [Sequence]	Forward	36	590	185 [Sequence]	384	3e-136	98.92
ENSP00000430829	VEGFA	1	209	Forward	6:43771247-43784559 [Sequence]	Forward	36	608	209 [Sequence]	384	7e-136	89.95
ENSP00000388663	VEGFA	174	353	Forward	6:43771226-43782052 [Sequence]	Forward	15	554	180 [Sequence]	383	3e-133	98.89

Showing 1 to 10 of 35 entries

HSP distribution on genome

①Done：提示已完成

②View results:点击查看结果

③结果列表：

Subject name: 蛋白版本号

Gene hit:基因名

Subject start: 匹配起始位置

Subject end: 匹配结束位置

Genomic Location: 基因定位

Orientation: 匹配方向

Query start: 查询序列起始位置

Query end: 查询序列结束位置

Length：长度 (点击sequence可查看详细信息）

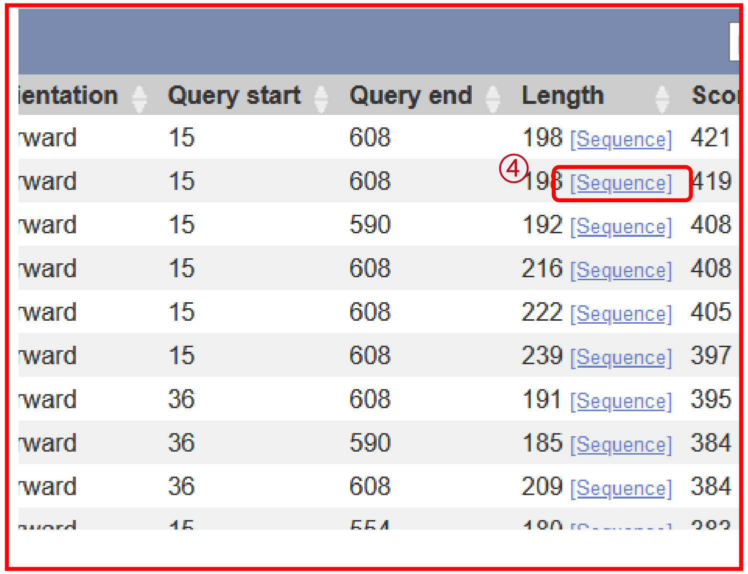
Score: 得分

E-val值

%ID：匹配比例

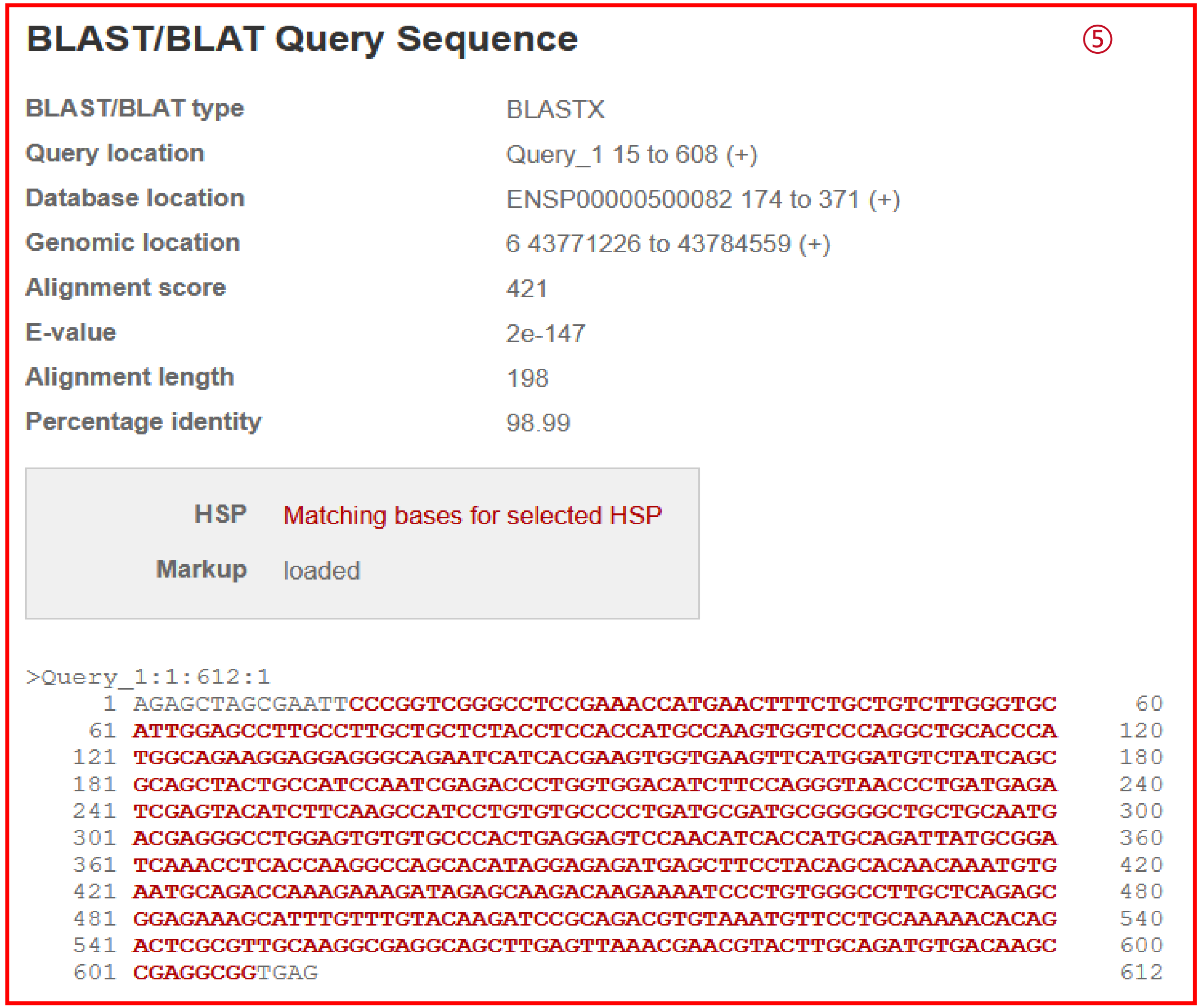
Ensembl检索方法---BLAST/BLAT序列相似性查找

- 以人源序列（ ENST00000672860.3 ） 为例进行蛋白比对，以确定该查询序列是哪种功能基因的同源基因，点击上页BLAST界面中的Run，结果页面见下



ientation	Query start	Query end	Length	Sco
ward	15	608	198 [Sequence]	421
ward	15	608	198 [Sequence]	419
ward	15	590	192 [Sequence]	408
ward	15	608	216 [Sequence]	408
ward	15	608	222 [Sequence]	405
ward	15	608	239 [Sequence]	397
ward	36	608	191 [Sequence]	395
ward	36	590	185 [Sequence]	384
ward	36	608	209 [Sequence]	384

④



BLAST/BLAT Query Sequence

⑤

BLAST/BLAT type	BLASTX
Query location	Query_1 15 to 608 (+)
Database location	ENSP00000500082 174 to 371 (+)
Genomic location	6 43771226 to 43784559 (+)
Alignment score	421
E-value	2e-147
Alignment length	198
Percentage identity	98.99

HSP	Matching bases for selected HSP
Markup	loaded

```
>Query_1:1:612:1
     1 AGAGCTAGCGAATTCCCGGTCGGGCCTCCGAAACCATGAACTTTCTGCTGTCTTGGGTGC      60
    61 ATTGGAGCCTTGCCTTGCTGCTCTACCTCCACCATGCCAAGTGGTCCCAGGCTGCACCCA     120
   121 TGGCAGAAGGAGGAGGGCAGAATCATCACGAAGTGGTGAAGTTCATGGATGTCTATCAGC     180
   181 GCAGCTACTGCCATCCAATCGAGACCCTGGTGGACATCTTCCAGGGTAACCCTGATGAGA     240
   241 TCGAGTACATCTTCAAGCCATCCTGTGTGCCCCTGATGCGATGCGGGGGCTGCTGCAATG     300
   301 ACGAGGGCCTGGAGTGTGTGCCCACTGAGGAGTCCAACATCACCATGCAGATTATGCGGA     360
   361 TCAAACCTCACCAAGGCCAGCACATAGGAGAGATGAGCTTCCTACAGCACAACAAATGTG     420
   421 AATGCAGACCAAAGAAAGATAGAGCAAGACAAGAAAATCCCTGTGGGCCTTGCTCAGAGC     480
   481 GGAGAAAGCATTTGTTTGTACAAGATCCGCAGACGTGTAAATGTTCCTGCAAAAACACAG     540
   541 ACTCGCGTTGCAAGGCGAGGCAGCTTGAGTTAAACGAACGTACTTGCAGATGTGACAAGC     600
   601 CGAGGCGGTGAG                                                     612
```

点击④sequence出现具体信息页面，见图⑤

课程目录

1 数据库介绍

2 检索

3 数据获取

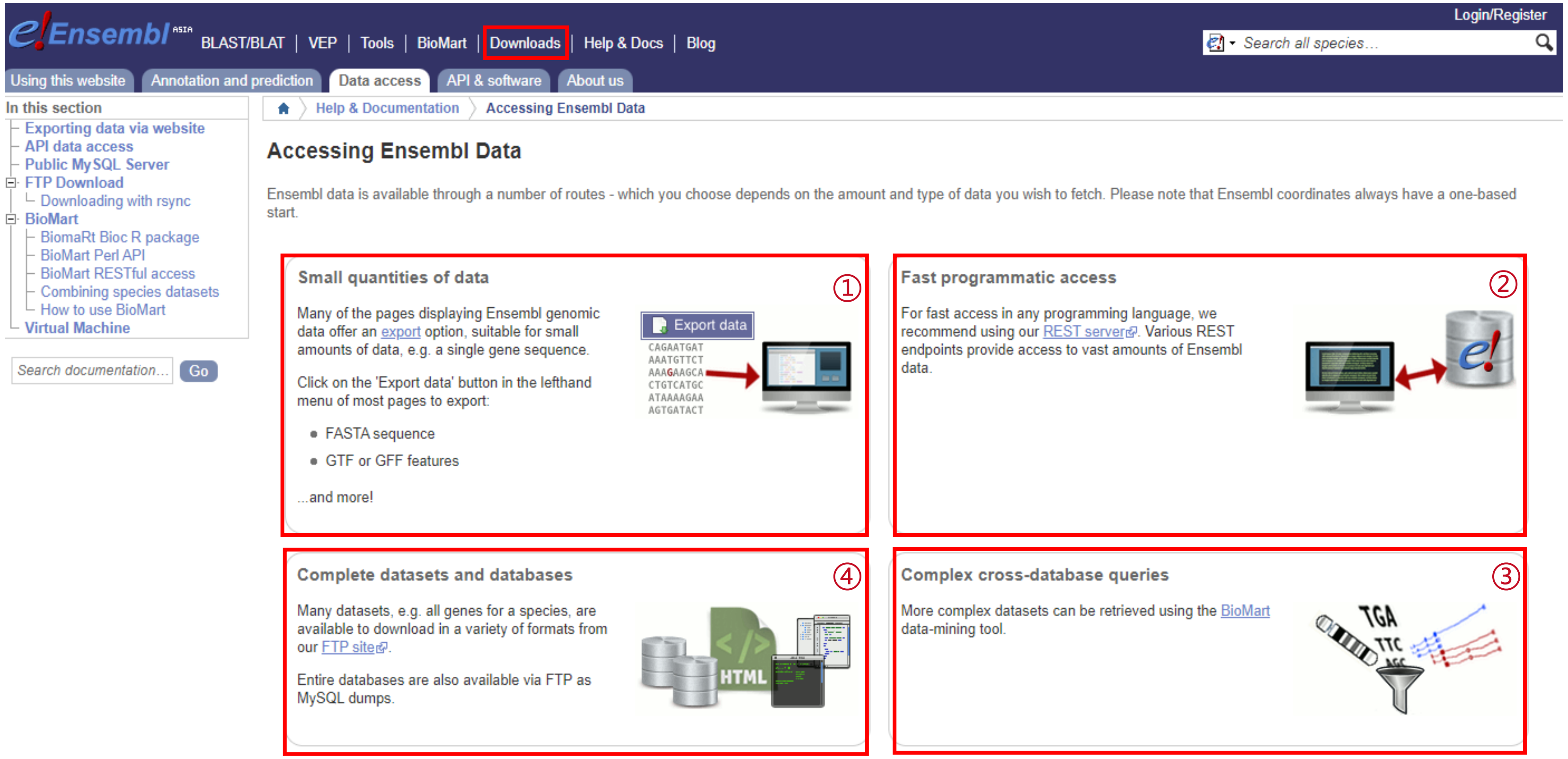
4 Biomart

5 版本转换

- 3.1 数据获取方法
- 3.2 Export下载
- 3.3 API or REST API下载
- 3.4 biomart批量下载
- 3.5 FIP下载

Ensembl数据获取方法

- 点击Ensembl主页上方Download，即可进入以下Download结果页面



Accessing Ensembl Data

Ensembl data is available through a number of routes - which you choose depends on the amount and type of data you wish to fetch. Please note that Ensembl coordinates always have a one-based start.

Small quantities of data ①

Many of the pages displaying Ensembl genomic data offer an export option, suitable for small amounts of data, e.g. a single gene sequence.

Click on the 'Export data' button in the lefthand menu of most pages to export:

- FASTA sequence
- GTF or GFF features

...and more!

Fast programmatic access ②

For fast access in any programming language, we recommend using our REST server. Various REST endpoints provide access to vast amounts of Ensembl data.

Complete datasets and databases ④

Many datasets, e.g. all genes for a species, are available to download in a variety of formats from our FTP site.

Entire databases are also available via FTP as MySQL dumps.

Complex cross-database queries ③

More complex datasets can be retrieved using the BioMart data-mining tool.

①少量的数据

✓ 大多数Ensembl 基因组数据的描述页有 “export” 功能，可以直接导出这一页的内容。

②大的数据集：

✓ PERL API http://www.ensembl.org/info/docs/api/index.html

✓ 如果不熟悉Perl语言，可以通过Ensembl REST API http://rest.ensembl.org/

③复杂的交叉数据库：

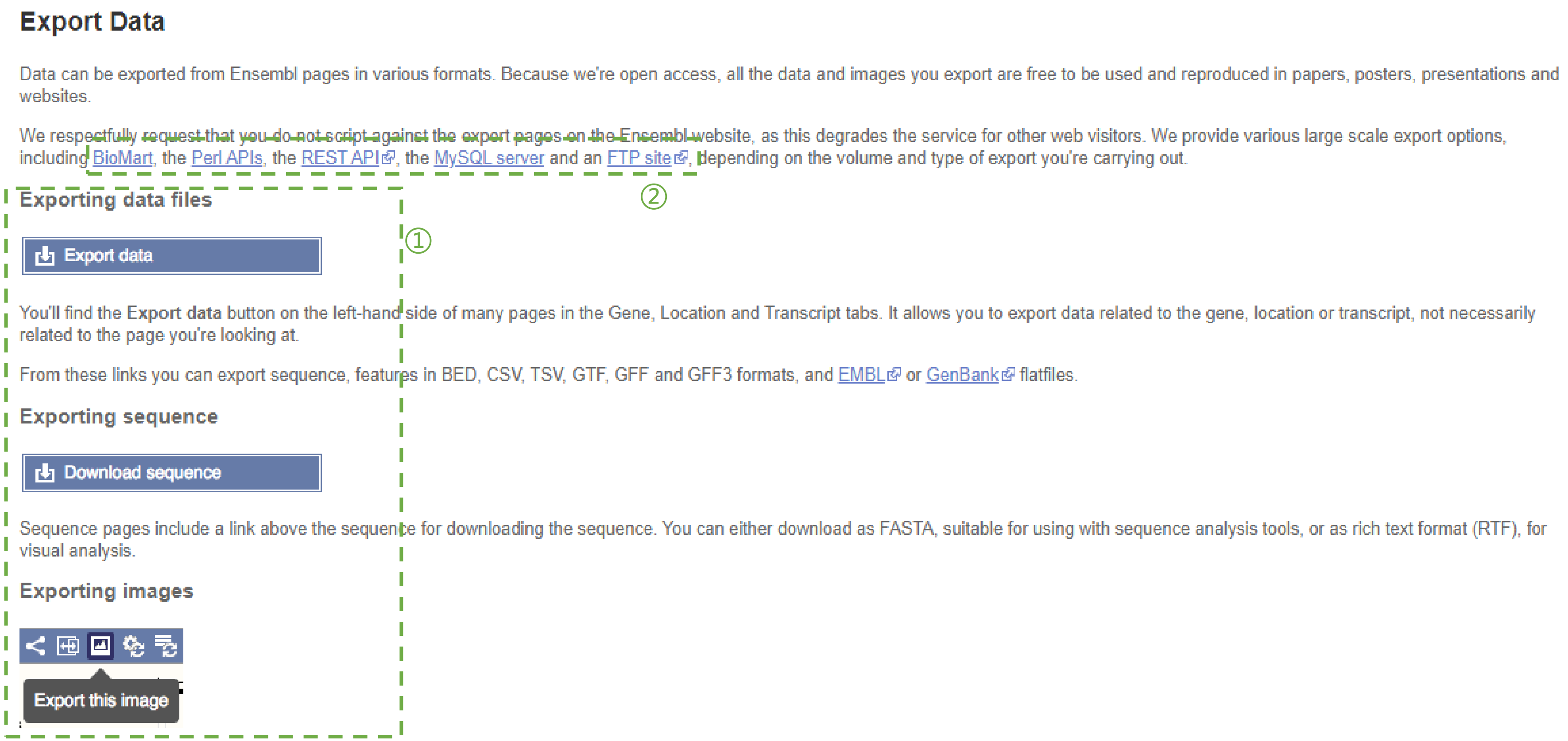
✓ BioMart http://www.ensembl.org/info/data/biomart/index.html

④全部的数据集

✓ FTP site：http://www.ensembl.org/info/data/ftp/index.html

Ensembl数据获取方法--Export下载

- 点击Download结果页面的中①Small quantities of data，即可进入以下页面



Export Data

Data can be exported from Ensembl pages in various formats. Because we're open access, all the data and images you export are free to be used and reproduced in papers, posters, presentations and websites.

We respectfully request that you do not script against the export pages on the Ensembl website, as this degrades the service for other web visitors. We provide various large scale export options, including BioMart, the Perl APIs, the REST API, the MySQL server and an FTP site, depending on the volume and type of export you're carrying out.

Exporting data files

You'll find the **Export data** button on the left-hand side of many pages in the Gene, Location and Transcript tabs. It allows you to export data related to the gene, location or transcript, not necessarily related to the page you're looking at.

From these links you can export sequence, features in BED, CSV, TSV, GTF, GFF and GFF3 formats, and EMBL or GenBank flatfiles.

Exporting sequence

Sequence pages include a link above the sequence for downloading the sequence. You can either download as FASTA, suitable for using with sequence analysis tools, or as rich text format (RTF), for visual analysis.

Exporting images

①少量的数据下载界面

Exporting data:导出数据

Exporting sequence:导出序列

Exporting images:导出图片

✓ 大多数Ensembl 基因组数据的描述页有“export”功能，可以直接导出这一页的内容。在前边功能介绍中结果展示部分均有说明如何下载

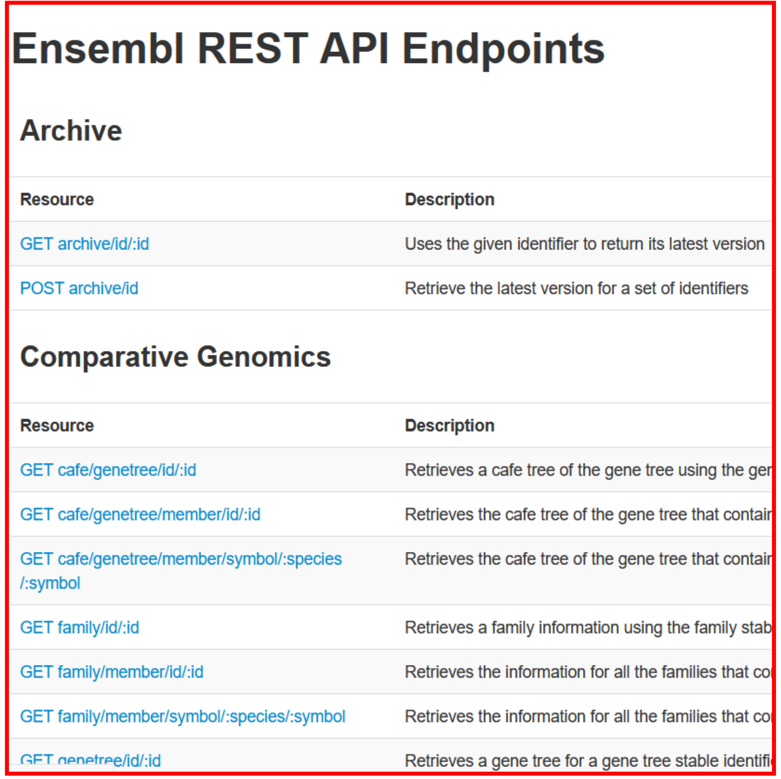
②提供其他宗数据下载的链接，可直接点击链接到其他方式的下载页面，见后述

Ensembl数据获取方法--API下载

■ 点击Download结果页面中②Fast programmatic access，即可进入API下载界面，页面如下

API是基于Perl的下载方式，需要有一定的语言基础，REST API则提供不同的语言代码实例，简单的方法就是按照提供的语言代码修改为自己需要的内容进行的下载，API既可以下载某个基因序列，也可以的下载大宗的基因组序列。

①



Ensembl REST API Endpoints

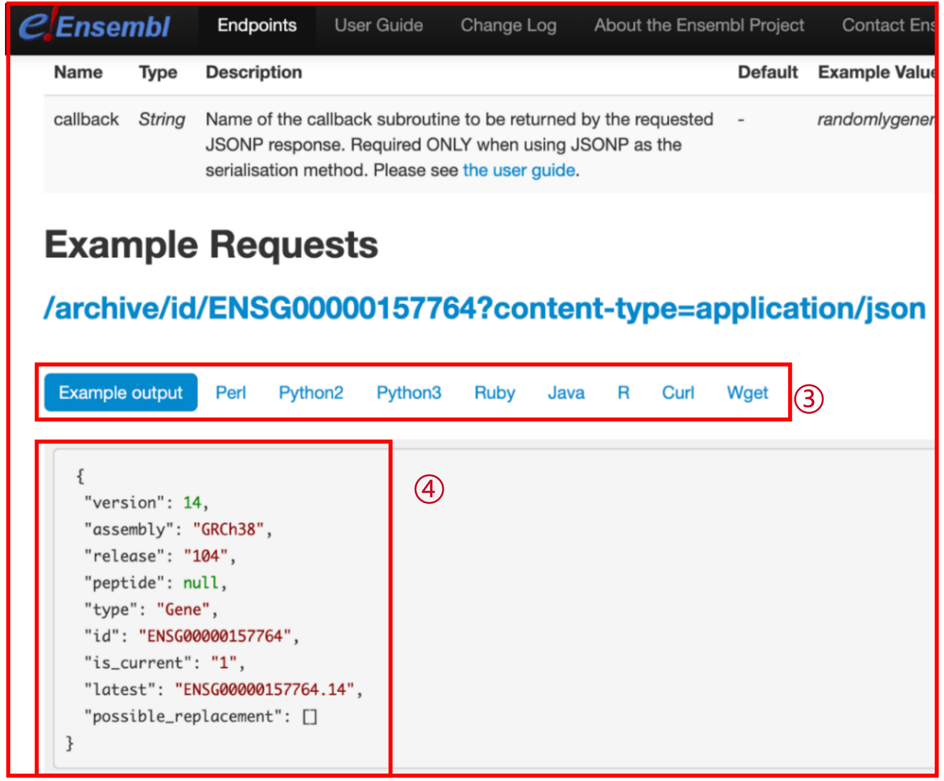
Archive

Resource	Description
GET archive/id/:id	Uses the given identifier to return its latest version
POST archive/id	Retrieve the latest version for a set of identifiers

Comparative Genomics

Resource	Description
GET cafe/genetree/id/:id	Retrieves a cafe tree of the gene tree using the ge
GET cafe/genetree/member/id/:id	Retrieves the cafe tree of the gene tree that contai
GET cafe/genetree/member/symbol/:species/:symbol	Retrieves the cafe tree of the gene tree that contai
GET family/id/:id	Retrieves a family information using the family stab
GET family/member/id/:id	Retrieves the information for all the families that co
GET family/member/symbol/:species/:symbol	Retrieves the information for all the families that co
GET genetree/id/:id	Retrieves a gene tree for a gene tree stable identifi

②



Name	Type	Description	Default	Example Value
callback	*String*	Name of the callback subroutine to be returned by the requested JSONP response. Required ONLY when using JSONP as the serialisation method. Please see the user guide.	-	*randomlygener*

Example Requests

/archive/id/ENSG00000157764?content-type=application/json

```
{
  "version": 14,
  "assembly": "GRCh38",
  "release": "104",
  "peptide": null,
  "type": "Gene",
  "id": "ENSG00000157764",
  "is_current": "1",
  "latest": "ENSG00000157764.14",
  "possible_replacement": []
}
```

①API下载界面

- 17个大目录
- 数据档案文件（Archive)
- 比较基因组学(Comparative Genomics)
- 交叉引用(Cross references)
- EQTL（表达数量性状基因座)
- 信息(information)、连锁不平衡(Linkage disequilibrium)
- 查找、图谱、本体和分类法(ontologies and taxonomy)、共有区域搜索(overlap)、表型注释调控（phenomotype annotations)、序列(sequence)、转录本单倍体、突变结果、变异GA4GH格式

点击每一个链接可以进入相应的脚本界面，个人可根据个人需求进行脚本的修改

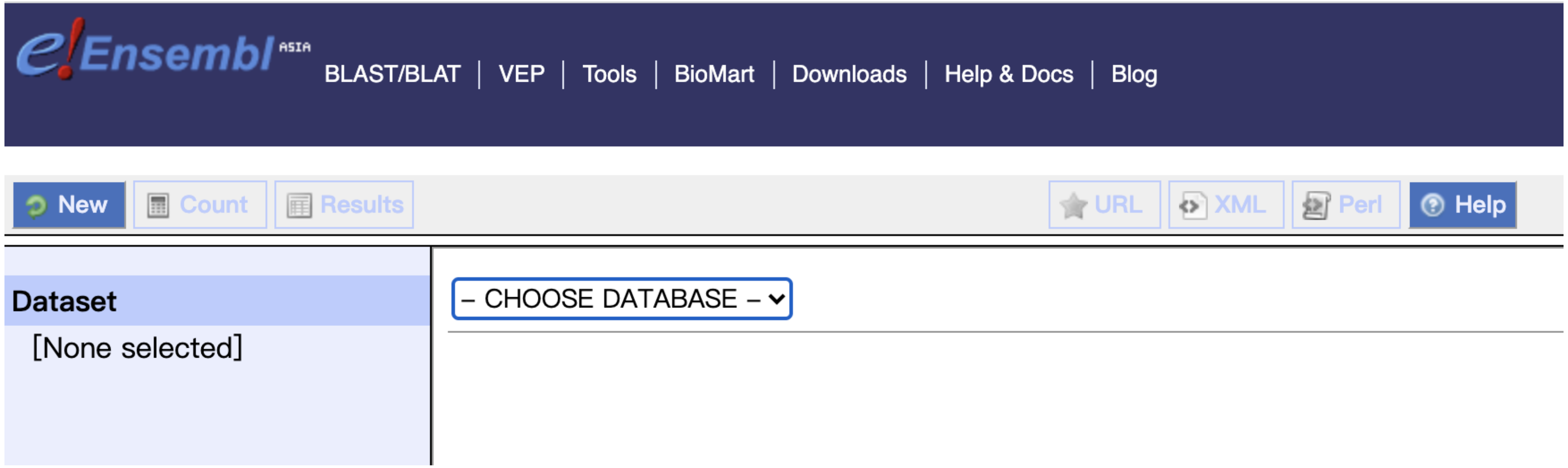
点击GET archive/id/:id即可进入界面②

③根据个人所熟悉的语言选择脚本

④根据需求，只需要修改代码中标红的部分即可

Ensembl数据获取方法--API下载Biomart下载

- 点击Download结果页面中③complex cross-database quries，即可进入Biomart下载界面，页面如下



后续有针对Biomart使用的详细介绍，此处不做进一步详解

Ensembl数据获取方法--FTP下载一

- 点击Download结果页面中④complete datasets and databases中FTP sites，即可进入FTP下载界面，页面如下图所示

Index of /pub/

```
../
IPI/                                        05-Dec-2016 11:53
assembly/                                   05-Dec-2016 11:53
assembly_mapping/                           05-Dec-2016 11:53
current_assembly_chain/                     15-Mar-2021 00:19
current_bamcov/                             04-May-2021 09:38
current_bed/                                11-Mar-2021 13:55
current_compara/                            11-Mar-2021 13:58
current_data_files/                         04-May-2021 09:38
current_embl/                               23-Mar-2021 15:10
current_emf/                                11-Mar-2021 13:55
current_fasta/                              12-Mar-2021 18:29
current_genbank/                            23-Mar-2021 15:10
current_gff3/                               21-Mar-2021 00:15
current_gtf/                                23-Mar-2021 15:06
current_json/                               21-Mar-2021 07:29
current_maf/                                11-Mar-2021 13:55
current_mysql/                              23-Apr-2021 17:20
current_ncbi_blast/                         12-Mar-2021 17:25
current_regulation/                         08-Feb-2021 12:22
current_solr_srch/                          29-Apr-2021 12:20
current_tsv/                                15-Mar-2021 07:29
current_variation/                          29-Mar-2021 14:08
current_virtual_machine/                    08-Mar-2021 15:50
current_xml/                                11-Mar-2021 13:55
data_files/①                                01-Feb-2021 10:26
functional_genomics/                        05-Dec-2016 11:53
gff3_so_terms_update/                       21-Jun-2017 14:49
grch37/                                     05-May-2021 13:40
misc/                                       18-Aug-2021 09:39
misc-scripts/                               05-Dec-2016 11:53
papers/                                     23-Oct-2018 08:58
pre/                                        05-Dec-2016 11:53
rapid-release/                              27-Jul-2021 20:19
release-100/                                28-Apr-2020 17:13
release-101/                                19-Aug-2020 17:08
release-102/                                26-Oct-2020 17:14
release-103/                                16-Feb-2021 17:52
release-104/                                30-Mar-2021 12:08
release-19/                                 05-Dec-2016 11:53
release-21/                                 05-Dec-2016 11:53
release-22/                                 05-Dec-2016 11:53
```

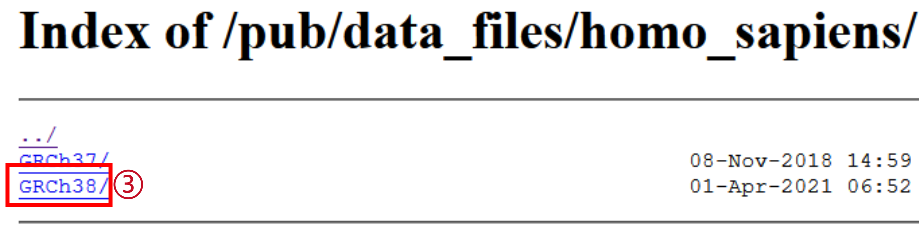
```
denticeps_clupeoides/                       31-Jul-2019 11:02
dicentrarchus_labrax/                       22-Jan-2020 11:48
dromaius_novaehollandiae/                   07-Feb-2019 13:57
echeneis_naucrates/                         29-Oct-2019 14:10
electrophorus_electricus/                   13-May-2019 09:50
eptatretus_burgeri/                         25-May-2018 09:12
equus_asinus_asinus/                        07-Nov-2018 16:07
equus_caballus/                             30-Oct-2018 17:03
erpetoichthys_calabaricus/                  31-Jul-2019 12:21
erythrura_gouldiae/                         29-Oct-2019 11:00
esox_lucius/                                30-Jan-2020 13:58
falco_tinnunculus/                          24-Apr-2020 06:56
felis_catus/                                18-May-2018 07:29
ficedula_albicollis/                        28-Jan-2021 08:51
fukomys_damarensis/                         22-Jun-2017 11:11
fundulus_heteroclitus/                      14-Aug-2018 11:34
gadus_morhua/                               09-Nov-2020 10:26
gallus_gallus/                              25-Oct-2018 08:15
gambusia_affinis/                           13-Aug-2018 15:57
geospiza_fortis/                            30-Jan-2020 11:07
gopherus_agassizii/                         07-Nov-2018 16:16
gopherus_evgoodei/                          23-Jan-2020 14:39
gorilla_gorilla/                            04-Oct-2017 13:04
gouania_willdenowi/                         31-Jul-2019 09:32
haplochromis_burtoni/                       14-Aug-2018 11:39
heterocephalus_glaber_female/               27-Jun-2017 11:29
heterocephalus_glaber_male/                 27-Jun-2017 11:29
hippocampus_comes/                          14-Aug-2018 11:43
homo_sapiens/②                              08-Dec-2016 15:16
hucho_hucho/                                13-May-2019 14:29
ictalurus_punctatus/                        13-Aug-2018 13:19
ictidomys_tridecemlineatus/                 27-Jun-2017 11:29
junco_hyemalis/                             14-Feb-2019 11:12
kryptolebias_marmoratus/                    13-Aug-2018 13:30
labrus_bergylta/                            15-Aug-2018 07:44
larimichthys_crocea/                        12-Jul-2019 13:59
lates_calcarifer/                           13-May-2019 13:16
laticauda_laticaudata/                      30-Jan-2020 11:15
latimeria_chalumnae/                        28-Aug-2012 16:28
lepidothrix_coronata/                       08-Feb-2019 11:24
lepisosteus_oculatus/                       14-Oct-2013 13:37
leptobrachium_leishanense/                  28-Apr-2020 11:46
```

①FTP下载界面索引，点击data files进入②

②data file中详细的信息，此处以人为例，点击homo sapiens，进入③

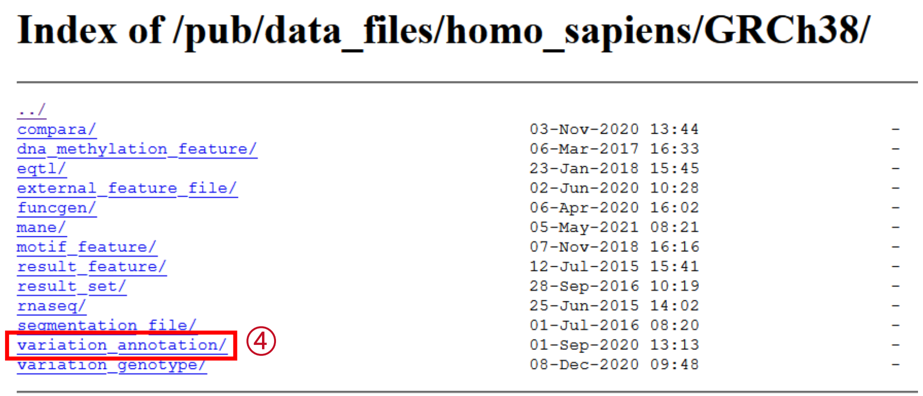
Ensembl数据获取方法--FTP下载一

- 点击Download结果页面中④complete datasets and databases中FTP sites，即可进入FTP下载界面，页面如下图所示—接上页PPT



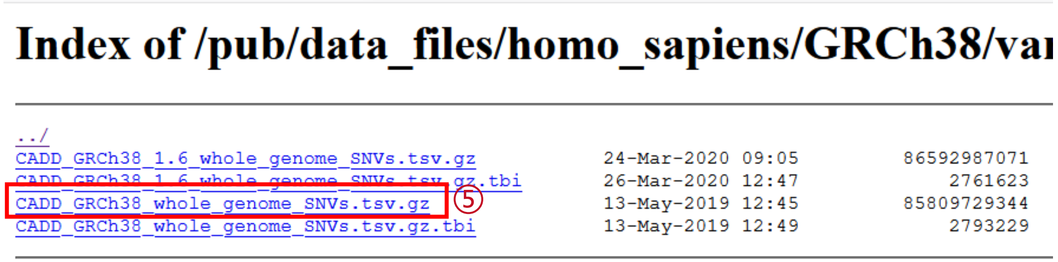
Index of /pub/data_files/homo_sapiens/

```
../
GRCh37/                                            08-Nov-2018 14:59
GRCh38/ ③                                          01-Apr-2021 06:52
```



Index of /pub/data_files/homo_sapiens/GRCh38/

```
../
compara/                       03-Nov-2020 13:44                   -
dna_methylation_feature/       06-Mar-2017 16:33                   -
eqtl/                          23-Jan-2018 15:45                   -
external_feature_file/         02-Jun-2020 10:28                   -
funcgen/                       06-Apr-2020 16:02                   -
mane/                          05-May-2021 08:21                   -
motif_feature/                 07-Nov-2018 16:16                   -
result_feature/                12-Jul-2015 15:41                   -
result_set/                    28-Sep-2016 10:19                   -
rnaseq/                        25-Jun-2015 14:02                   -
segmentation_file/             01-Jul-2016 08:20                   -
variation_annotation/ ④        01-Sep-2020 13:13                   -
variation_genotype/            08-Dec-2020 09:48                   -
```



Index of /pub/data_files/homo_sapiens/GRCh38/va

```
../
CADD_GRCh38_1.6_whole_genome_SNVs.tsv.gz          24-Mar-2020 09:05    86592987071
CADD_GRCh38_1.6_whole_genome_SNVs.tsv.gz.tbi      26-Mar-2020 12:47        2761623
CADD_GRCh38_whole_genome_SNVs.tsv.gz ⑤            13-May-2019 12:45    85809729344
CADD_GRCh38_whole_genome_SNVs.tsv.gz.tbi          13-May-2019 12:49        2793229
```

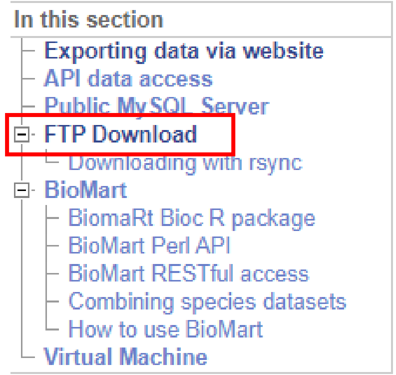
③选择版本，建议选择最新版本，点击GRCh38进入界面④

④该版本包含的信息有比较基因组学、DNA甲基化数据、motif特征数据，变异注释等。此处选择变异注释（variation annoation),点击进入界面⑤

⑤根据需求选择需要下载的数据，建议选择最新版本数据，此处选在最新版本全基因组数据（whole genome) tsv.gz格式文件。点击后即可弹出储存路径界面

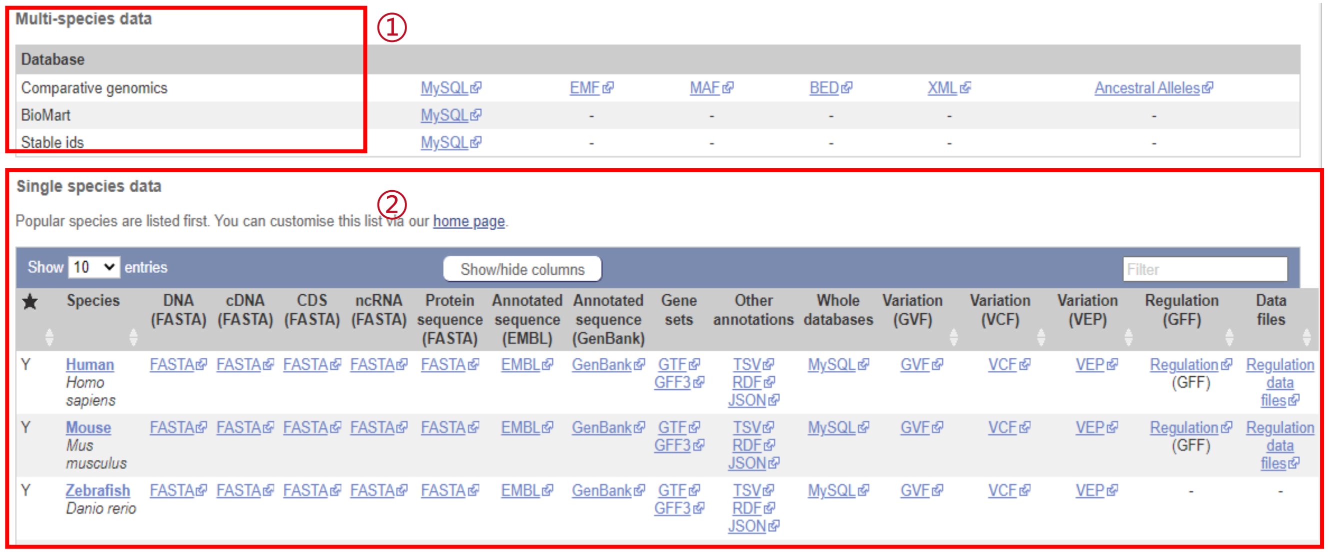
Ensembl数据获取方法--FTP下载二

- 在Download结果界面左侧导航栏中分别点击FTP Download，其右侧内容展示区将如下图所示。



In this section

- Exporting data via website
- API data access
- Public MySQL Server
- ⊟ FTP Download
  - Downloading with rsync
- ⊟ BioMart
  - BiomaRt Bioc R package
  - BioMart Perl API
  - BioMart RESTful access
  - Combining species datasets
  - How to use BioMart
- Virtual Machine



Multi-species data ①

Database						
Comparative genomics	MySQL	EMF	MAF	BED	XML	Ancestral Alleles
BioMart	MySQL	-	-	-	-	-
Stable ids	MySQL	-	-	-	-	-

Single species data ②

Popular species are listed first. You can customise this list via our home page.

★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Other annotations	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files
Y	Human *Homo sapiens*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files
Y	Mouse *Mus musculus*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files
Y	Zebrafish *Danio rerio*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	GVF	VCF	VEP	-	-

①多个物种数据

- Comparative genomics (比较基因组学)
- BioMart
- Stable ids （固定Id版本）

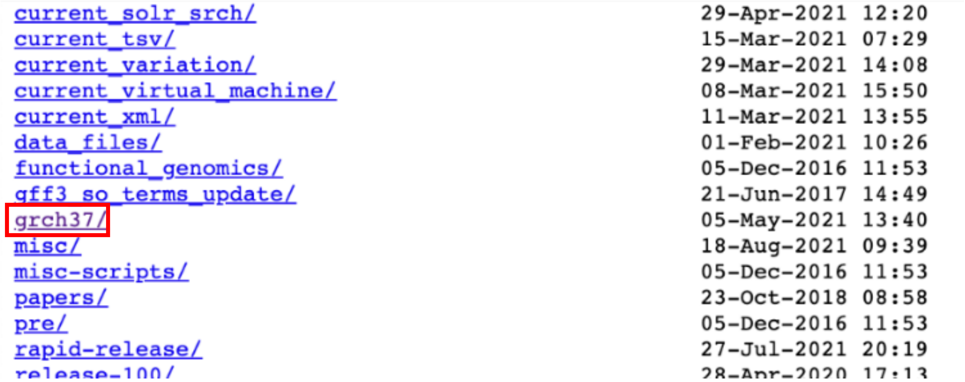
②单个物种数据

- Species：物种
- DNA（FASTA）：DNA序列（FASTA格式）
- cDNA序列
- CDS序列
- ncRNA：非编码RNA序列
- Protein sequence：蛋白序列
- Annotated sequence（EMBL）：注释序列（来自EMBL）
- Annotated sequence（GenBank）注释序列（来自GenBank）
- Gene sets:基因集
- Other annotations：其他注释信息

Ensembl数据获取方法--FTP下载三

- 点击Download结果界面的④complete datasets and databases中FTP sites，即可进入FTP下载界面，结果页面如下图所示

①



```
current_solr_srch/              29-Apr-2021 12:20
current_tsv/                    15-Mar-2021 07:29
current_variation/              29-Mar-2021 14:08
current_virtual_machine/        08-Mar-2021 15:50
current_xml/                    11-Mar-2021 13:55
data_files/                     01-Feb-2021 10:26
functional_genomics/            05-Dec-2016 11:53
gff3_so_terms_update/           21-Jun-2017 14:49
grch37/                         05-May-2021 13:40
misc/                           18-Aug-2021 09:39
misc-scripts/                   05-Dec-2016 11:53
papers/                         23-Oct-2018 08:58
pre/                            05-Dec-2016 11:53
rapid-release/                  27-Jul-2021 20:19
release-100/                    28-Apr-2020 17:13
```

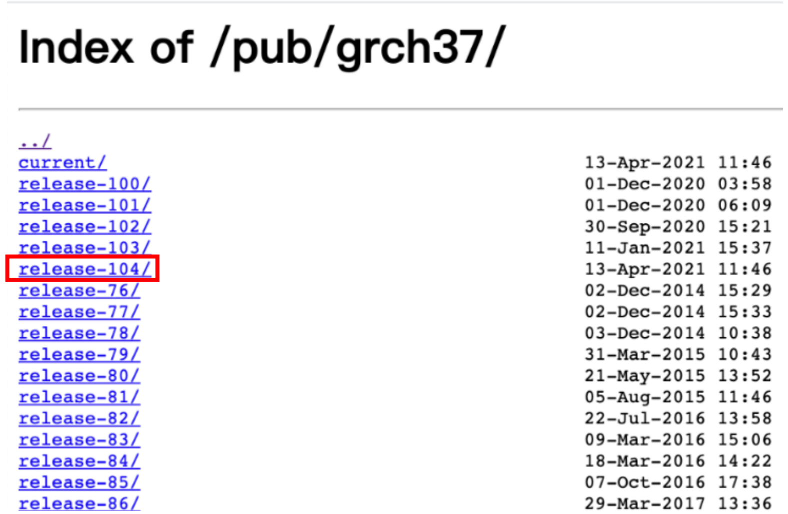
如果想下载人类的基因组序列，还可以按照最新日期选择文件、选择下载格式、人类、根据个人需求点击下载即可

①FTP下载界面索引，点击grch37进入②

②grch37中详细的信息，选择最新版本进入界面③

③选择数据格式，此处选择FASTA，点击进入界面④

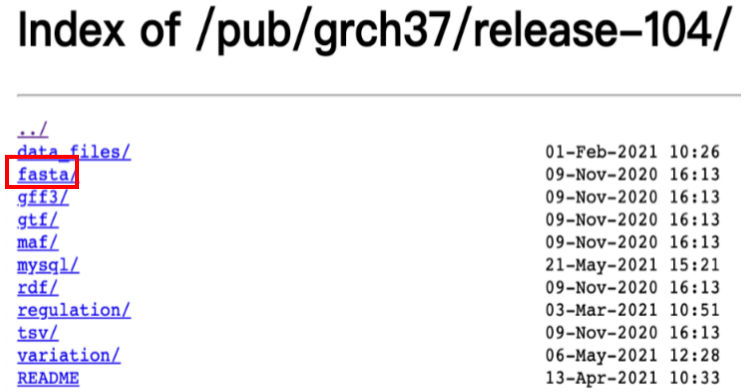
②



Index of /pub/grch37/

```
../
current/                        13-Apr-2021 11:46
release-100/                    01-Dec-2020 03:58
release-101/                    01-Dec-2020 06:09
release-102/                    30-Sep-2020 15:21
release-103/                    11-Jan-2021 15:37
release-104/                    13-Apr-2021 11:46
release-76/                     02-Dec-2014 15:29
release-77/                     02-Dec-2014 15:33
release-78/                     03-Dec-2014 10:38
release-79/                     31-Mar-2015 10:43
release-80/                     21-May-2015 13:52
release-81/                     05-Aug-2015 11:46
release-82/                     22-Jul-2016 13:58
release-83/                     09-Mar-2016 15:06
release-84/                     18-Mar-2016 14:22
release-85/                     07-Oct-2016 17:38
release-86/                     29-Mar-2017 13:36
```

③



Index of /pub/grch37/release-104/

```
../
data_files/                     01-Feb-2021 10:26
fasta/                          09-Nov-2020 16:13
gff3/                           09-Nov-2020 16:13
gtf/                            09-Nov-2020 16:13
maf/                            09-Nov-2020 16:13
mysql/                          21-May-2021 15:21
rdf/                            09-Nov-2020 16:13
regulation/                     03-Mar-2021 10:51
tsv/                            09-Nov-2020 16:13
variation/                      06-May-2021 12:28
README                          13-Apr-2021 10:33
```

Ensembl数据获取方法--FTP下载三

- 点击Download结果界面的④complete datasets and databases中FTP sites，即可进入FTP下载界面，结果页面如下图所示—接上页PPT

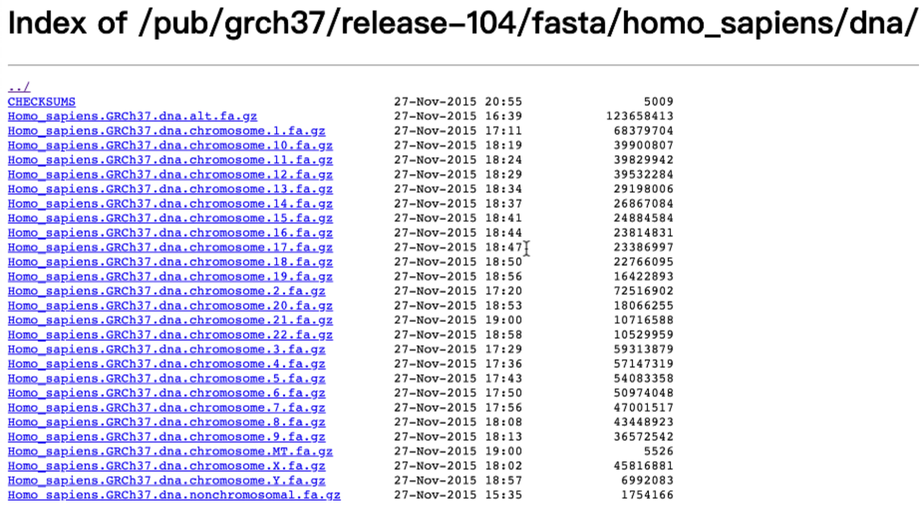
④

```
Index of /pub/grch37/release-104/fasta/

../
homo_sapiens/                                      27-Nov-2015 20:37                   -
fasta                                              09-Nov-2020 16:13                  21
```

⑤

```
Index of /pub/grch37/release-104/fasta/homo_sapiens/

../
cdna/                                              27-Nov-2015 20:54                   -
cds/                                               27-Nov-2015 20:54                   -
dna/                                               27-Nov-2015 20:55                   -
ncrna/                                             27-Nov-2015 20:54                   -
pep/                                               27-Nov-2015 20:54                   -
```

⑥



```
Index of /pub/grch37/release-104/fasta/homo_sapiens/dna/

../
CHECKSUMS                                          27-Nov-2015 20:55                5009
Homo_sapiens.GRCh37.dna.alt.fa.gz                  27-Nov-2015 16:39           123658413
Homo_sapiens.GRCh37.dna.chromosome.1.fa.gz         27-Nov-2015 17:11            68379704
Homo_sapiens.GRCh37.dna.chromosome.10.fa.gz        27-Nov-2015 18:19            39900807
Homo_sapiens.GRCh37.dna.chromosome.11.fa.gz        27-Nov-2015 18:24            39829942
Homo_sapiens.GRCh37.dna.chromosome.12.fa.gz        27-Nov-2015 18:29            39532284
Homo_sapiens.GRCh37.dna.chromosome.13.fa.gz        27-Nov-2015 18:34            29198006
Homo_sapiens.GRCh37.dna.chromosome.14.fa.gz        27-Nov-2015 18:37            26867084
Homo_sapiens.GRCh37.dna.chromosome.15.fa.gz        27-Nov-2015 18:41            24884584
Homo_sapiens.GRCh37.dna.chromosome.16.fa.gz        27-Nov-2015 18:44            23814831
Homo_sapiens.GRCh37.dna.chromosome.17.fa.gz        27-Nov-2015 18:47            23386997
Homo_sapiens.GRCh37.dna.chromosome.18.fa.gz        27-Nov-2015 18:50            22766095
Homo_sapiens.GRCh37.dna.chromosome.19.fa.gz        27-Nov-2015 18:56            16422893
Homo_sapiens.GRCh37.dna.chromosome.2.fa.gz         27-Nov-2015 17:20            72516902
Homo_sapiens.GRCh37.dna.chromosome.20.fa.gz        27-Nov-2015 18:53            18066255
Homo_sapiens.GRCh37.dna.chromosome.21.fa.gz        27-Nov-2015 19:00            10716588
Homo_sapiens.GRCh37.dna.chromosome.22.fa.gz        27-Nov-2015 18:58            10529959
Homo_sapiens.GRCh37.dna.chromosome.3.fa.gz         27-Nov-2015 17:29            59313879
Homo_sapiens.GRCh37.dna.chromosome.4.fa.gz         27-Nov-2015 17:36            57147319
Homo_sapiens.GRCh37.dna.chromosome.5.fa.gz         27-Nov-2015 17:43            54083358
Homo_sapiens.GRCh37.dna.chromosome.6.fa.gz         27-Nov-2015 17:50            50974048
Homo_sapiens.GRCh37.dna.chromosome.7.fa.gz         27-Nov-2015 17:56            47001517
Homo_sapiens.GRCh37.dna.chromosome.8.fa.gz         27-Nov-2015 18:08            43448923
Homo_sapiens.GRCh37.dna.chromosome.9.fa.gz         27-Nov-2015 18:13            36572542
Homo_sapiens.GRCh37.dna.chromosome.MT.fa.gz        27-Nov-2015 19:00                5526
Homo_sapiens.GRCh37.dna.chromosome.X.fa.gz         27-Nov-2015 18:02            45816881
Homo_sapiens.GRCh37.dna.chromosome.Y.fa.gz         27-Nov-2015 18:57             6992083
Homo_sapiens.GRCh37.dna.nonchromosomal.fa.gz       27-Nov-2015 15:35             1754166
```

④选择人类homo sapiens点击进入界面⑤

⑤根据需求选择所需序列

Cdna：cdna序列

Cds序列

Dna：DNA序列

Ncrna：非编码RNA序列

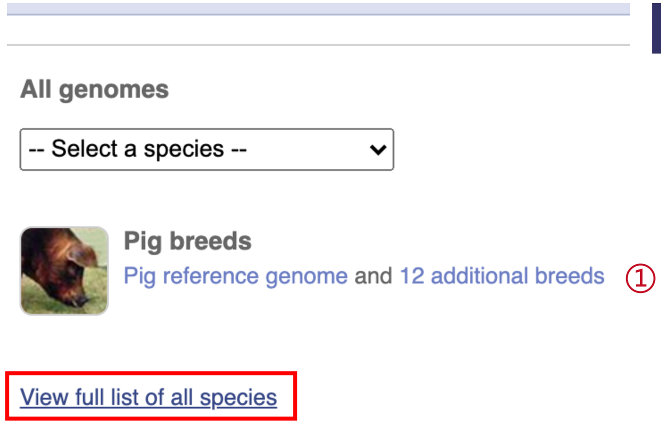
Pep：蛋白序列

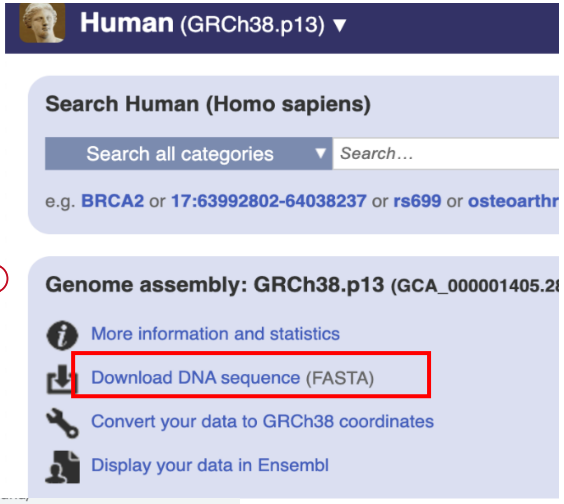
此处选择DNA序列，点击进入界面⑥

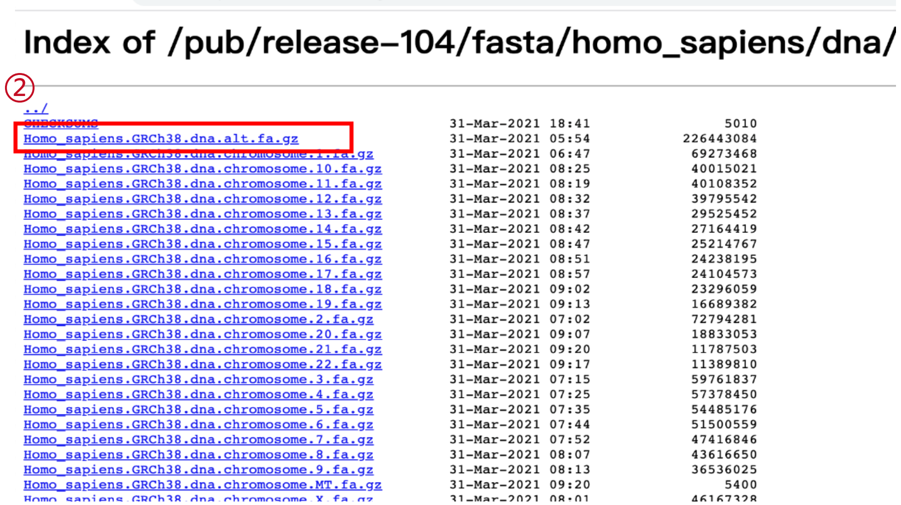
⑥点击最新的版本进行下载即可

Ensembl数据获取方法--FTP下载四

- Ensembl主页中view full list of all species，选择所研究物种，进入下载DNA序列界面，具体见以下结果页面







Index of /pub/release–104/fasta/homo_sapiens/dna/

②

```
../
CHECKSUMS                                          31-Mar-2021 18:41                5010
Homo_sapiens.GRCh38.dna.alt.fa.gz                  31-Mar-2021 05:54           226443084
Homo_sapiens.GRCh38.dna.chromosome.1.fa.gz         31-Mar-2021 06:47            69273468
Homo_sapiens.GRCh38.dna.chromosome.10.fa.gz        31-Mar-2021 08:25            40015021
Homo_sapiens.GRCh38.dna.chromosome.11.fa.gz        31-Mar-2021 08:19            40108352
Homo_sapiens.GRCh38.dna.chromosome.12.fa.gz        31-Mar-2021 08:32            39795542
Homo_sapiens.GRCh38.dna.chromosome.13.fa.gz        31-Mar-2021 08:37            29525452
Homo_sapiens.GRCh38.dna.chromosome.14.fa.gz        31-Mar-2021 08:42            27164419
Homo_sapiens.GRCh38.dna.chromosome.15.fa.gz        31-Mar-2021 08:47            25214767
Homo_sapiens.GRCh38.dna.chromosome.16.fa.gz        31-Mar-2021 08:51            24238195
Homo_sapiens.GRCh38.dna.chromosome.17.fa.gz        31-Mar-2021 08:57            24104573
Homo_sapiens.GRCh38.dna.chromosome.18.fa.gz        31-Mar-2021 09:02            23296059
Homo_sapiens.GRCh38.dna.chromosome.19.fa.gz        31-Mar-2021 09:13            16689382
Homo_sapiens.GRCh38.dna.chromosome.2.fa.gz         31-Mar-2021 07:02            72794281
Homo_sapiens.GRCh38.dna.chromosome.20.fa.gz        31-Mar-2021 09:07            18833053
Homo_sapiens.GRCh38.dna.chromosome.21.fa.gz        31-Mar-2021 09:20            11787503
Homo_sapiens.GRCh38.dna.chromosome.22.fa.gz        31-Mar-2021 09:17            11389810
Homo_sapiens.GRCh38.dna.chromosome.3.fa.gz         31-Mar-2021 07:15            59761837
Homo_sapiens.GRCh38.dna.chromosome.4.fa.gz         31-Mar-2021 07:25            57378450
Homo_sapiens.GRCh38.dna.chromosome.5.fa.gz         31-Mar-2021 07:35            54485176
Homo_sapiens.GRCh38.dna.chromosome.6.fa.gz         31-Mar-2021 07:44            51500559
Homo_sapiens.GRCh38.dna.chromosome.7.fa.gz         31-Mar-2021 07:52            47416846
Homo_sapiens.GRCh38.dna.chromosome.8.fa.gz         31-Mar-2021 08:07            43616650
Homo_sapiens.GRCh38.dna.chromosome.9.fa.gz         31-Mar-2021 08:13            36536025
Homo_sapiens.GRCh38.dna.chromosome.MT.fa.gz        31-Mar-2021 09:20                5400
Homo_sapiens.GRCh38.dna.chromosome.X.fa.gz         31-Mar-2021 08:01            46167328
```

①选择Download DNA sequence,进入下载界面②

②根据需求选择最新版本下载

数据既包含某个染色体的DNA数据，也包括全基因组的数据，根据个人需求选择就可

条条大路通罗马，还有其他多种方式

链接到全基因组序列下载，大家可以自行尝试

课程目录

1 数据库介绍

2 检索

3 数据获取

4 Biomart

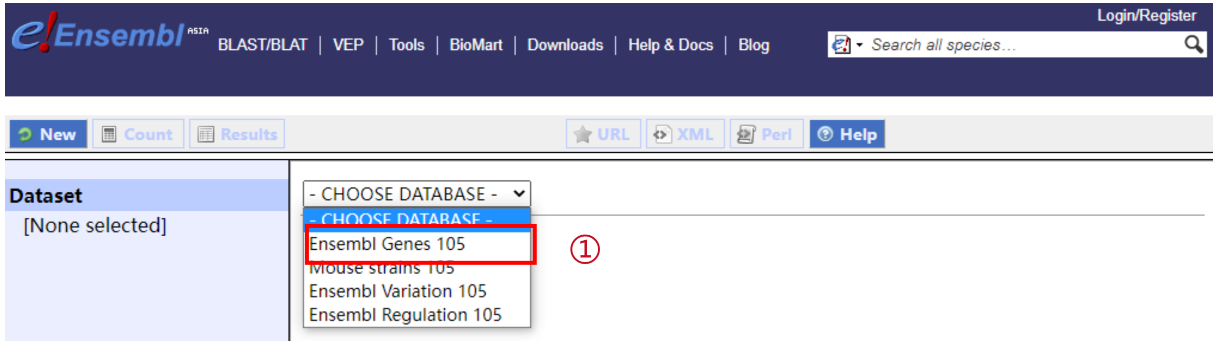
5 版本转换

- 4.1 Biomart入口
- 4.2 设置筛选条件
- 4.3 设置检索内容
- 4.4 下载设置及下载

Ensembl---BioMart

BioMart是一个集成的数据接口，可以通过其实现批量的数据检索

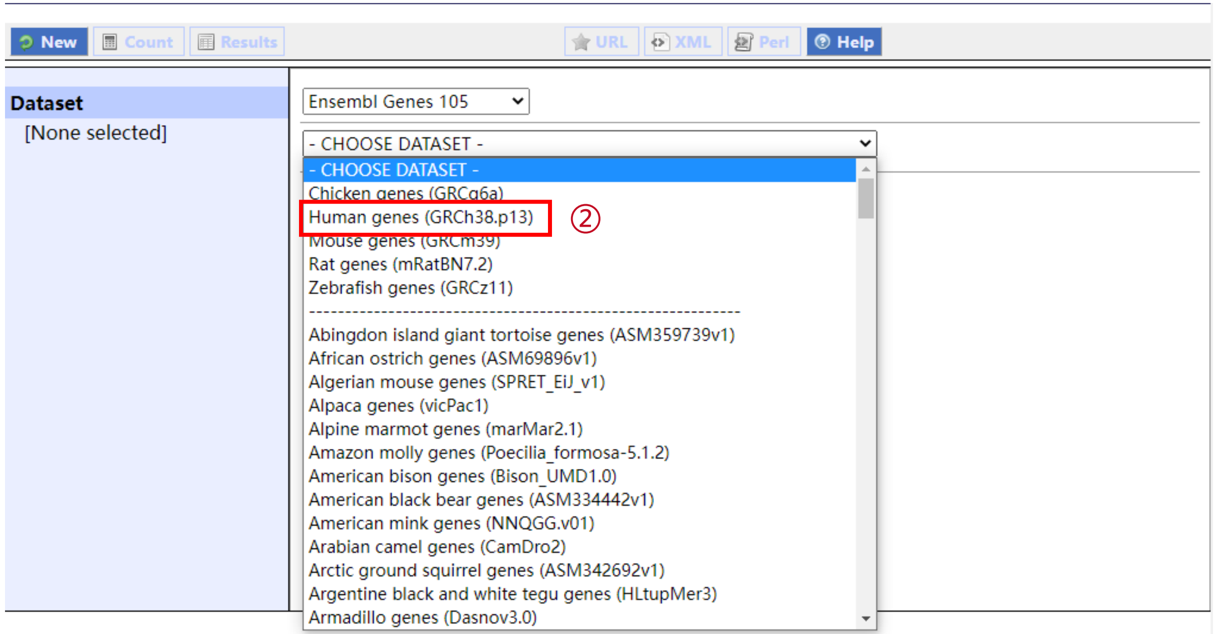
- 点击Ensembl主页中BioMart，进入BioMart界面，结果页面如图所示。



Dataset
[None selected]

- CHOOSE DATABASE -
- CHOOSE DATABASE -
Ensembl Genes 105
Mouse strains 105
Ensembl Variation 105
Ensembl Regulation 105

①

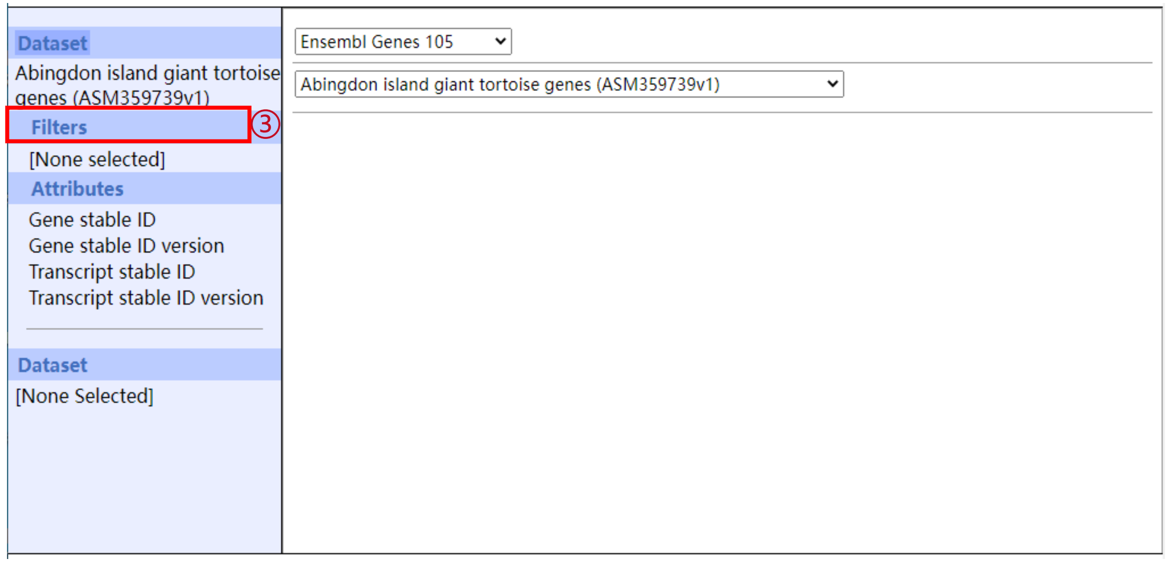


Dataset
[None selected]

Ensembl Genes 105
- CHOOSE DATASET -
- CHOOSE DATASET -
Chicken genes (GRCg6a)
Human genes (GRCh38.p13)
Mouse genes (GRCm39)
Rat genes (mRatBN7.2)
Zebrafish genes (GRCz11)

Abingdon island giant tortoise genes (ASM359739v1)
African ostrich genes (ASM69896v1)
Algerian mouse genes (SPRET_EiJ_v1)
Alpaca genes (vicPac1)
Alpine marmot genes (marMar2.1)
Amazon molly genes (Poecilia_formosa-5.1.2)
American bison genes (Bison_UMD1.0)
American black bear genes (ASM334442v1)
American mink genes (NNQGG.v01)
Arabian camel genes (CamDro2)
Arctic ground squirrel genes (ASM342692v1)
Argentine black and white tegu genes (HLtupMer3)
Armadillo genes (Dasnov3.0)

②


Dataset
Abingdon island giant tortoise genes (ASM359739v1)
Filters ③
[None selected]
Attributes
Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version

Dataset
[None Selected]

Ensembl Genes 105
Abingdon island giant tortoise genes (ASM359739v1)

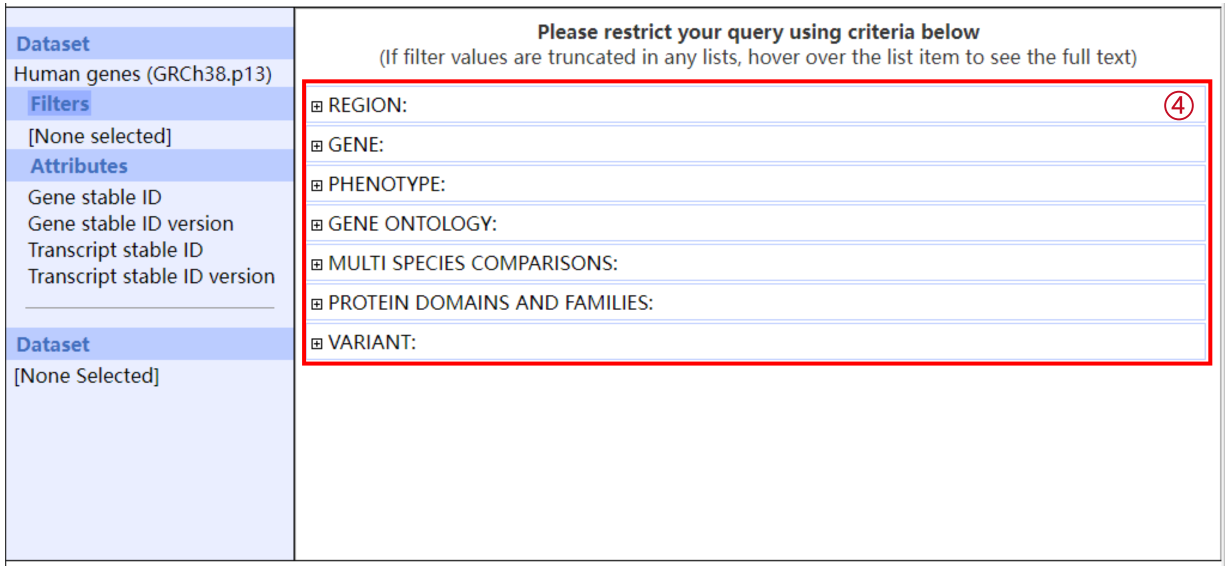
①选择数据库：建议最新版本

②选择种属：根据需求选择

③点击filter进入过滤页面，具体见下页PPT

Ensembl---BioMart

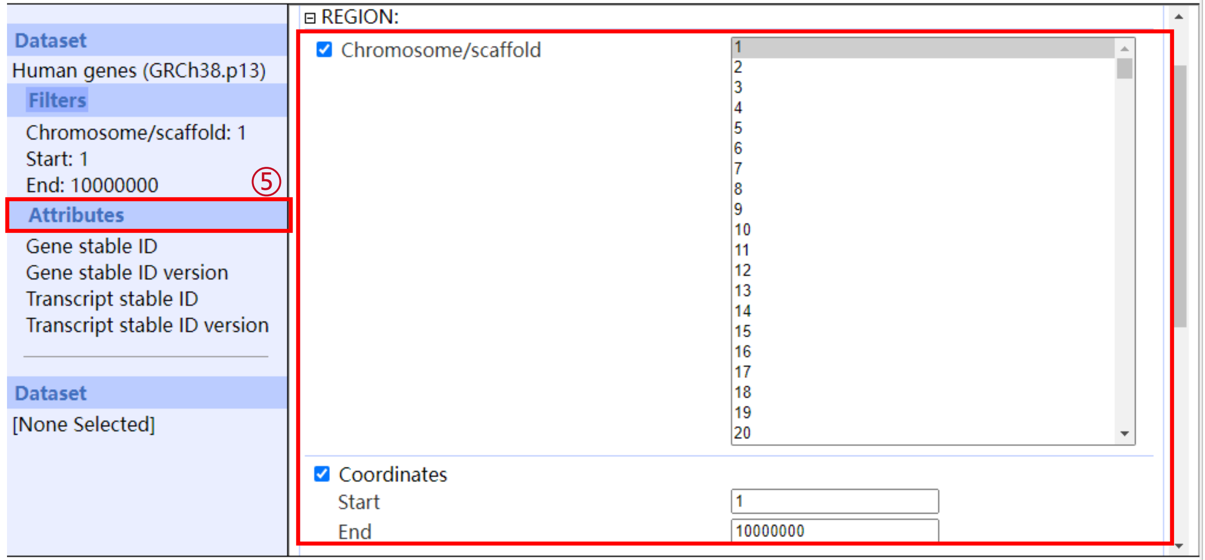
- 点击Ensembl主页中BioMart，进入BioMart界面，结果页面如图所示——接上页PPT。



Dataset
Human genes (GRCh38.p13)
Filters
[None selected]
Attributes
Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version
Dataset
[None Selected]

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

⊞ REGION: ④
⊞ GENE:
⊞ PHENOTYPE:
⊞ GENE ONTOLOGY:
⊞ MULTI SPECIES COMPARISONS:
⊞ PROTEIN DOMAINS AND FAMILIES:
⊞ VARIANT:



Dataset
Human genes (GRCh38.p13)
Filters
Chromosome/scaffold: 1
Start: 1
End: 10000000 ⑤
Attributes
Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version
Dataset
[None Selected]

⊟ REGION:
☑ Chromosome/scaffold

☑ Coordinates
Start 1
End 10000000

④filter过滤页面如左图所示

Region：选择染色体定位，见⑤

Gene:选择基因相关信息

Gene ontology:选择基因本体注释信息

Multi species comparisons: 选择多物种比较信息

Protein domains and families: 蛋白domain及家族

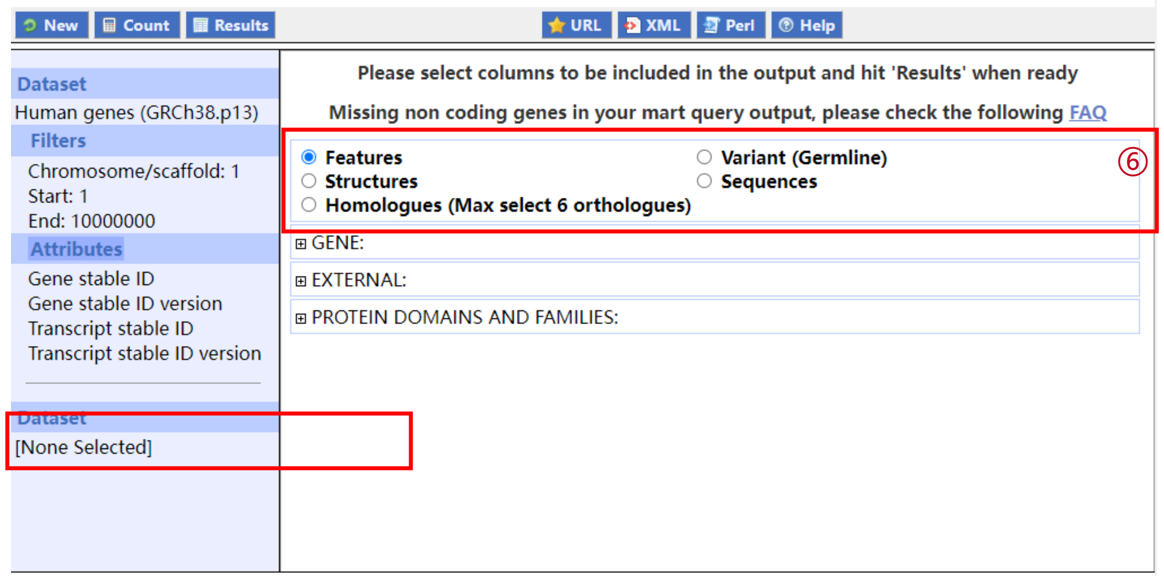
Variant：变异相关信息

⑤过滤条件筛选完毕后，点击左侧栏目中的attributes，进入界面⑥

见下页PPT

Ensembl---BioMart

- 点击Ensembl主页中BioMart，进入BioMart界面，结果页面如图所示——接上页PPT。



Dataset
Human genes (GRCh38.p13)
Filters
Chromosome/scaffold: 1
Start: 1
End: 10000000
Attributes
Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version
Dataset
[None Selected]

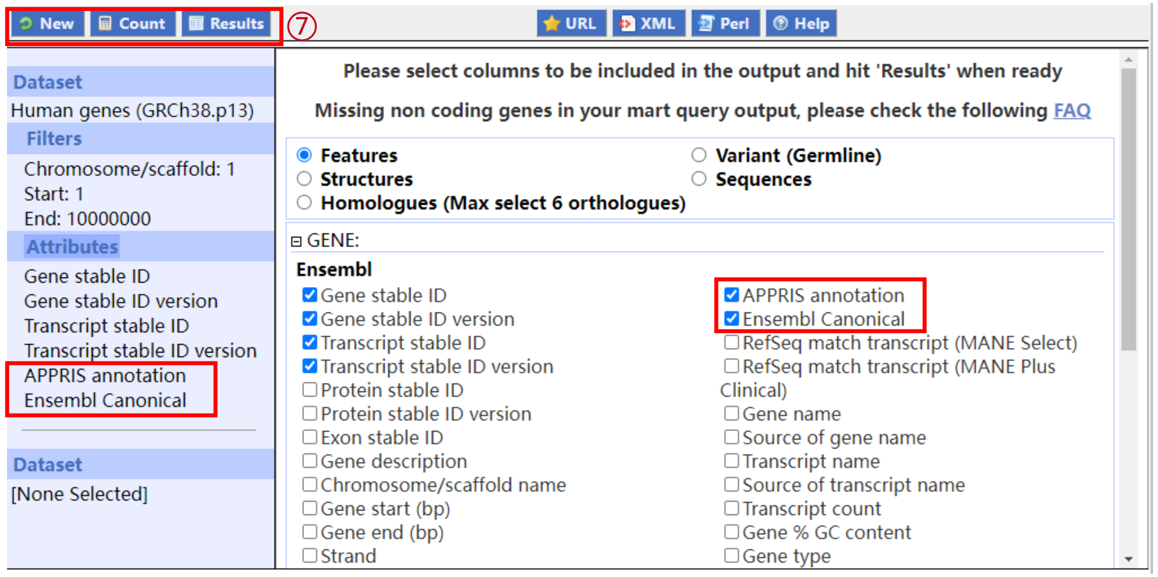
Please select columns to be included in the output and hit 'Results' when ready

Missing non coding genes in your mart query output, please check the following FAQ

● **Features** ○ **Variant (Germline)**
○ **Structures** ○ **Sequences**
○ **Homologues (Max select 6 orthologues)**

⑥

⊞ GENE:
⊞ EXTERNAL:
⊞ PROTEIN DOMAINS AND FAMILIES:



⑦

Dataset
Human genes (GRCh38.p13)
Filters
Chromosome/scaffold: 1
Start: 1
End: 10000000
Attributes
Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version
APPRIS annotation
Ensembl Canonical
Dataset
[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

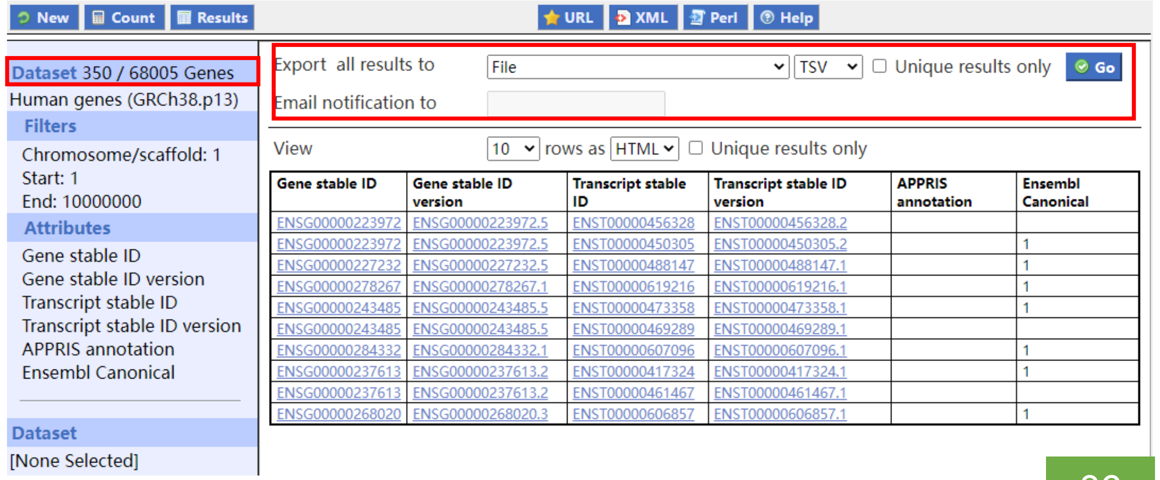
Missing non coding genes in your mart query output, please check the following FAQ

● **Features** ○ **Variant (Germline)**
○ **Structures** ○ **Sequences**
○ **Homologues (Max select 6 orthologues)**

⊟ GENE:

Ensembl

☑ Gene stable ID
☑ Gene stable ID version
☑ Transcript stable ID
☑ Transcript stable ID version
☐ Protein stable ID
☐ Protein stable ID version
☐ Exon stable ID
☐ Gene description
☐ Chromosome/scaffold name
☐ Gene start (bp)
☐ Gene end (bp)
☐ Strand
☑ APPRIS annotation
☑ Ensembl Canonical
☐ RefSeq match transcript (MANE Select)
☐ RefSeq match transcript (MANE Plus Clinical)
☐ Gene name
☐ Source of gene name
☐ Transcript name
☐ Source of transcript name
☐ Transcript count
☐ Gene % GC content
☐ Gene type



Dataset 350 / 68005 Genes
Human genes (GRCh38.p13)
Filters
Chromosome/scaffold: 1
Start: 1
End: 10000000
Attributes
Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version
APPRIS annotation
Ensembl Canonical
Dataset
[None Selected]

Export all results to File TSV ☐ Unique results only
Email notification to

View 10 rows as HTML ☐ Unique results only

Gene stable ID	Gene stable ID version	Transcript stable ID	Transcript stable ID version	APPRIS annotation	Ensembl Canonical
ENSG00000223972	ENSG00000223972.5	ENST00000456328	ENST00000456328.2		
ENSG00000223972	ENSG00000223972.5	ENST00000450305	ENST00000450305.2		1
ENSG00000227232	ENSG00000227232.5	ENST00000488147	ENST00000488147.1		1
ENSG00000278267	ENSG00000278267.1	ENST00000619216	ENST00000619216.1		1
ENSG00000243485	ENSG00000243485.5	ENST00000473358	ENST00000473358.1		1
ENSG00000243485	ENSG00000243485.5	ENST00000469289	ENST00000469289.1		
ENSG00000284332	ENSG00000284332.1	ENST00000607096	ENST00000607096.1		1
ENSG00000237613	ENSG00000237613.2	ENST00000417324	ENST00000417324.1		1
ENSG00000237613	ENSG00000237613.2	ENST00000461467	ENST00000461467.1		
ENSG00000268020	ENSG00000268020.3	ENST00000606857	ENST00000606857.1		1

⑥Attributes界面设置搜索显示内容，

- Features (特征)
- Structures (结构)
- Homologues（同源基因）
- Variant（Germline）（胚系变异）
- Variant(Somatic)(体系变异)
- Sequence（序列）

选择设置完成后点击左上角⑦框中的count按钮，可看Dataset具体情况。

⑦点击result，显示结果页面，点击下拉框选择输出文件格式，点击go即可批量下载

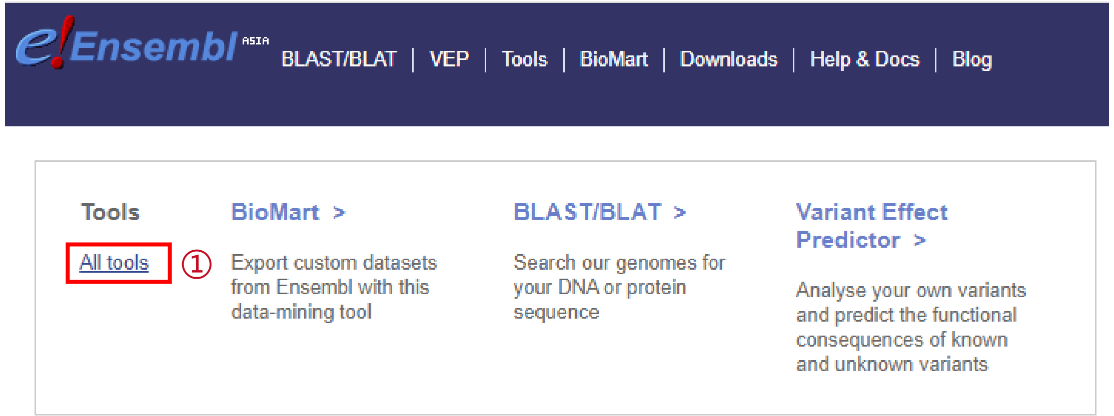
课程目录

1 数据库介绍
2 检索
3 数据获取
4 Biomart
5 版本转换
5.1 Assembly converter含义
5.2 操作步骤
5.3 结果展示

Ensembl版本转换

Assembly converter (版本转换）是Ensembl进行基因组坐标转换的工具，基于crossmap的，其他国际主流的生物信息学数据库的基因组坐标转换工具还包括 UCSC的liftover、NCBI的Remap

- 点击Ensembl主页中All tools，在界面中点击Assembly converter，会跳转至新界面，新界面的具体结果请看下张PPT。



Tools

All tools ①

BioMart >

Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >

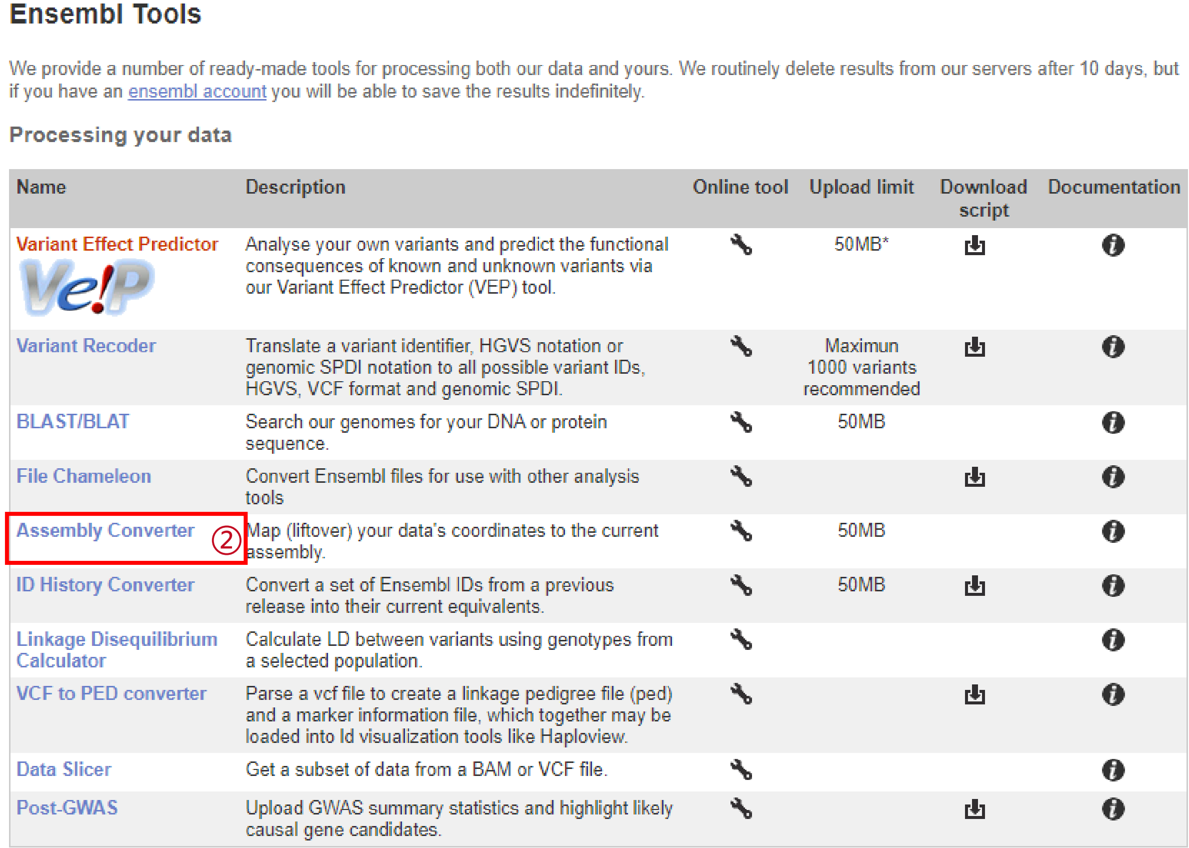
Search our genomes for your DNA or protein sequence

Variant Effect Predictor >

Analyse your own variants and predict the functional consequences of known and unknown variants

①All tools: 点击出现所有工具列表

②Assembly Converter版本转换



Ensembl Tools

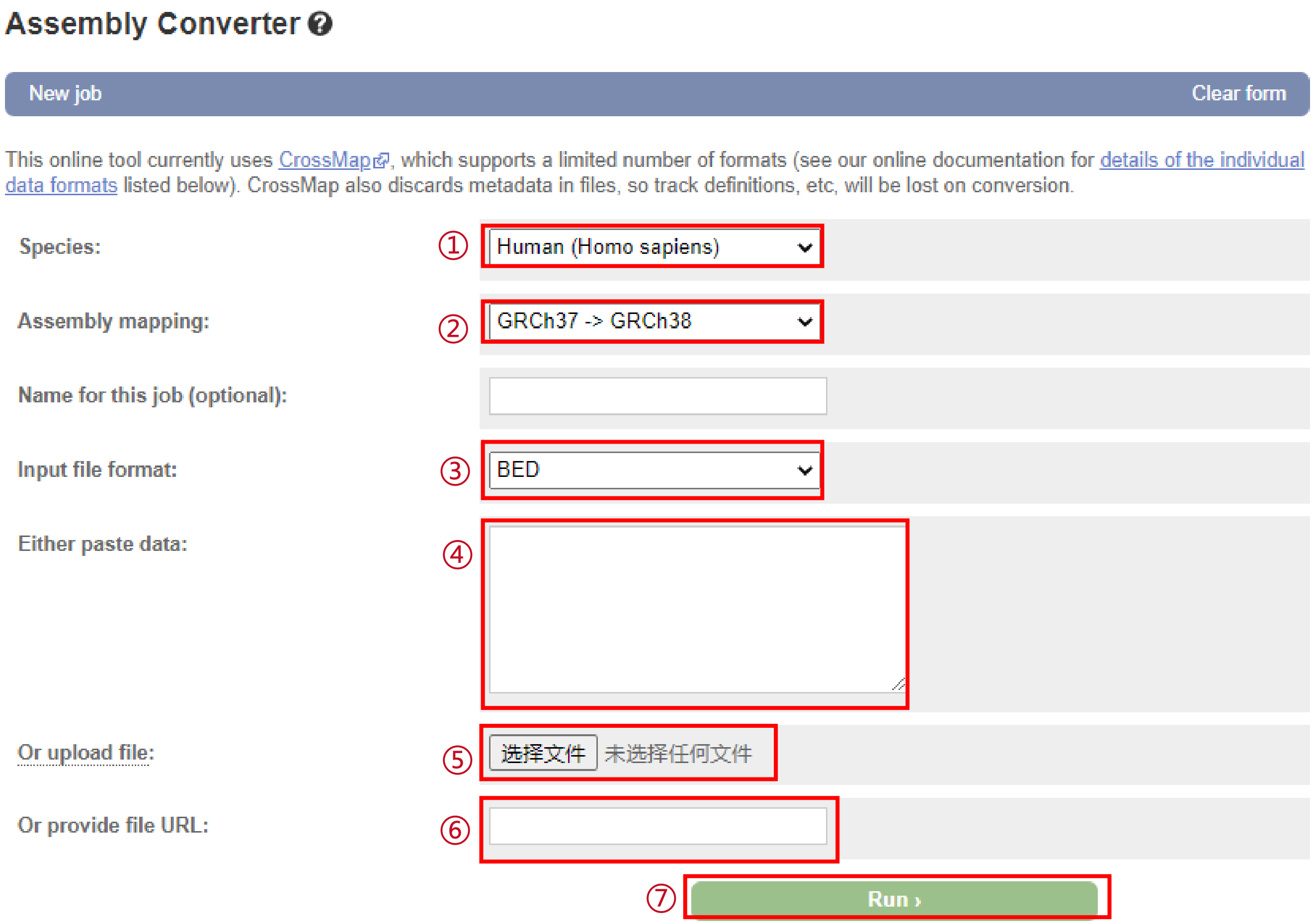
We provide a number of ready-made tools for processing both our data and yours. We routinely delete results from our servers after 10 days, but if you have an ensembl account you will be able to save the results indefinitely.

Processing your data

Name	Description	Online tool	Upload limit	Download script	Documentation
Variant Effect Predictor Ve!P	Analyse your own variants and predict the functional consequences of known and unknown variants via our Variant Effect Predictor (VEP) tool.	🔧	50MB*	⬇	ⓘ
Variant Recoder	Translate a variant identifier, HGVS notation or genomic SPDI notation to all possible variant IDs, HGVS, VCF format and genomic SPDI.	🔧	Maximun 1000 variants recommended	⬇	ⓘ
BLAST/BLAT	Search our genomes for your DNA or protein sequence.	🔧	50MB		ⓘ
File Chameleon	Convert Ensembl files for use with other analysis tools	🔧		⬇	ⓘ
Assembly Converter ②	Map (liftover) your data's coordinates to the current assembly.	🔧	50MB		ⓘ
ID History Converter	Convert a set of Ensembl IDs from a previous release into their current equivalents.	🔧	50MB	⬇	ⓘ
Linkage Disequilibrium Calculator	Calculate LD between variants using genotypes from a selected population.	🔧			ⓘ
VCF to PED converter	Parse a vcf file to create a linkage pedigree file (ped) and a marker information file, which together may be loaded into ld visualization tools like Haploview.	🔧		⬇	ⓘ
Data Slicer	Get a subset of data from a BAM or VCF file.	🔧			ⓘ
Post-GWAS	Upload GWAS summary statistics and highlight likely causal gene candidates.	🔧		⬇	ⓘ

Ensembl版本转换

- Assembly converter的跳转结果页面，如下图所示。



Assembly Converter

This online tool currently uses CrossMap, which supports a limited number of formats (see our online documentation for details of the individual data formats listed below). CrossMap also discards metadata in files, so track definitions, etc, will be lost on conversion.

Species: ① Human (Homo sapiens)

Assembly mapping: ② GRCh37 -> GRCh38

Name for this job (optional):

Input file format: ③ BED

Either paste data: ④

Or upload file: ⑤ 选择文件 未选择任何文件

Or provide file URL: ⑥

⑦ Run ›

①点击选择物种切换

②点击进行版本转换

③输入文件格式

④也可以直接将文件粘贴至此

⑤上传文件

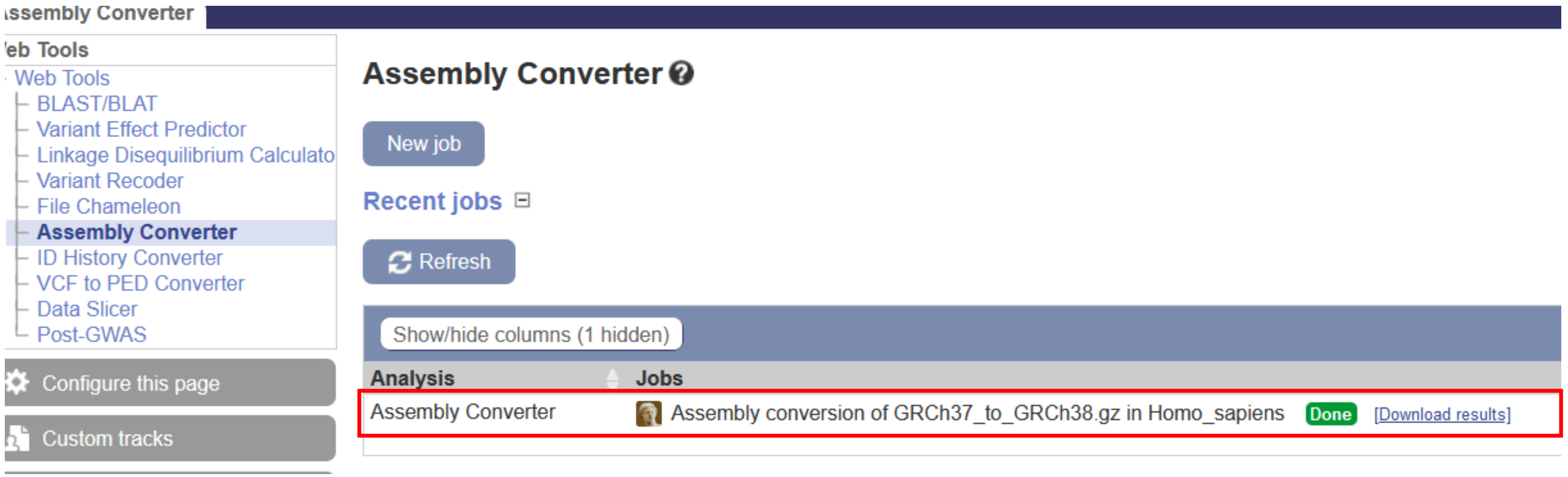
⑥通过文件URL链接提供文件

⑦运行

Ensembl版本转换

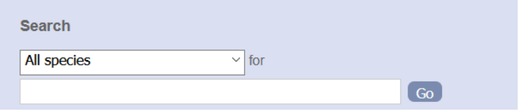
- 点击Assembly converter界面的run后，结果展示界面如下

利用网页Assembly converter工具进行转换需要利用chain file，才能知道两个版本的坐标对应关系。chain文件可以从Crossmap或UCSC下载。假设我们需要进行GRCh37_to_GRCh38的转换，我们需要的是GRCh37_to_GRCh38.chain.gz下载地址：
https://sourceforge.net/projects/crossmap/files/Ensembl_chain_files/homo_sapiens%28human%29/



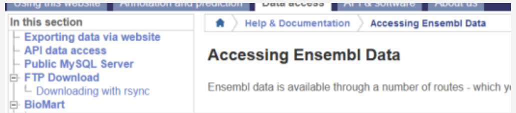
Analysis	Jobs
Assembly Converter	Assembly conversion of GRCh37_to_GRCh38.gz in Homo_sapiens Done [Download results]

小结：Ensembl可进行基因注释、变异、调控信息等检索、可下载


Search
All species for
Go

检索

检索基因信息


Data access
In this section
- Exporting data via website
- API data access
- Public MySQL Server
- FTP Download
  - Downloading with rsync
- BioMart

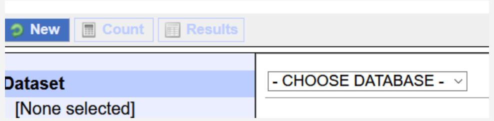
Help & Documentation > Accessing Ensembl Data

Accessing Ensembl Data

Ensembl data is available through a number of routes - which y

数据下载

下载基因或基因组信息



Dataset
[None selected]
- CHOOSE DATABASE -

BioMart

批量数据检索

[更多相关课程：](#)

《Origin软件使用教程》

《GDSC数据库使用教程》

《引物设计使用教程》

更多精彩课程，将陆续推出

酸谈学社

领悟科研，优人一步

36策 24型 套路课 单元课 一点通 文献精 SCI写作

- **请电脑使用谷歌浏览器访问**

www.helixlife.cn

THE END

Ensembl数据库使用教程

科研技能 单元课 01

- 针对医生做基础科研过程中需要的特定科研技能，解螺旋制作了一系列专项学习突破的图文或视频教程，谓之“单元课”，此为第 01 期。

后续课程

敬请关注

意见反馈

搜索公众号：解螺旋 | 陪伴医生科研成长