

# Finding functional relevant genes

Yolande F. M. Ramos  
[y.f.m.ramos@lumc.nl](mailto:y.f.m.ramos@lumc.nl)  
Molecular Epidemiology  
LEIDEN, THE NETHERLANDS



# AIMS of this lecture

- Understanding **genomic variation**, SNPs
- Functional **relevant variation**
- Use of online **databases**

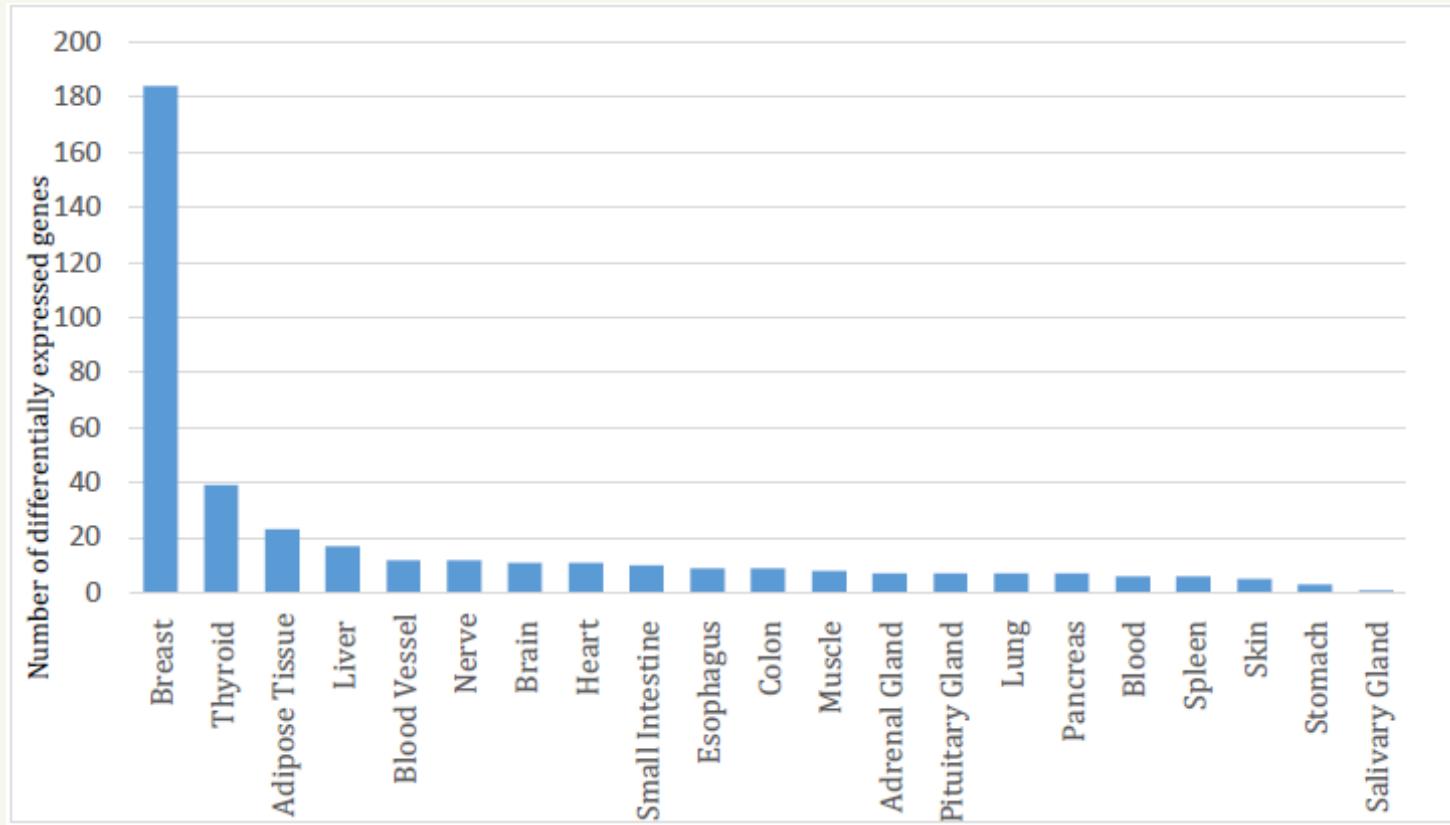
# AIMS of this lecture

- Understanding **genomic variation**, SNPs
- Functional relevant variation
- Use of online databases

# Human Genome

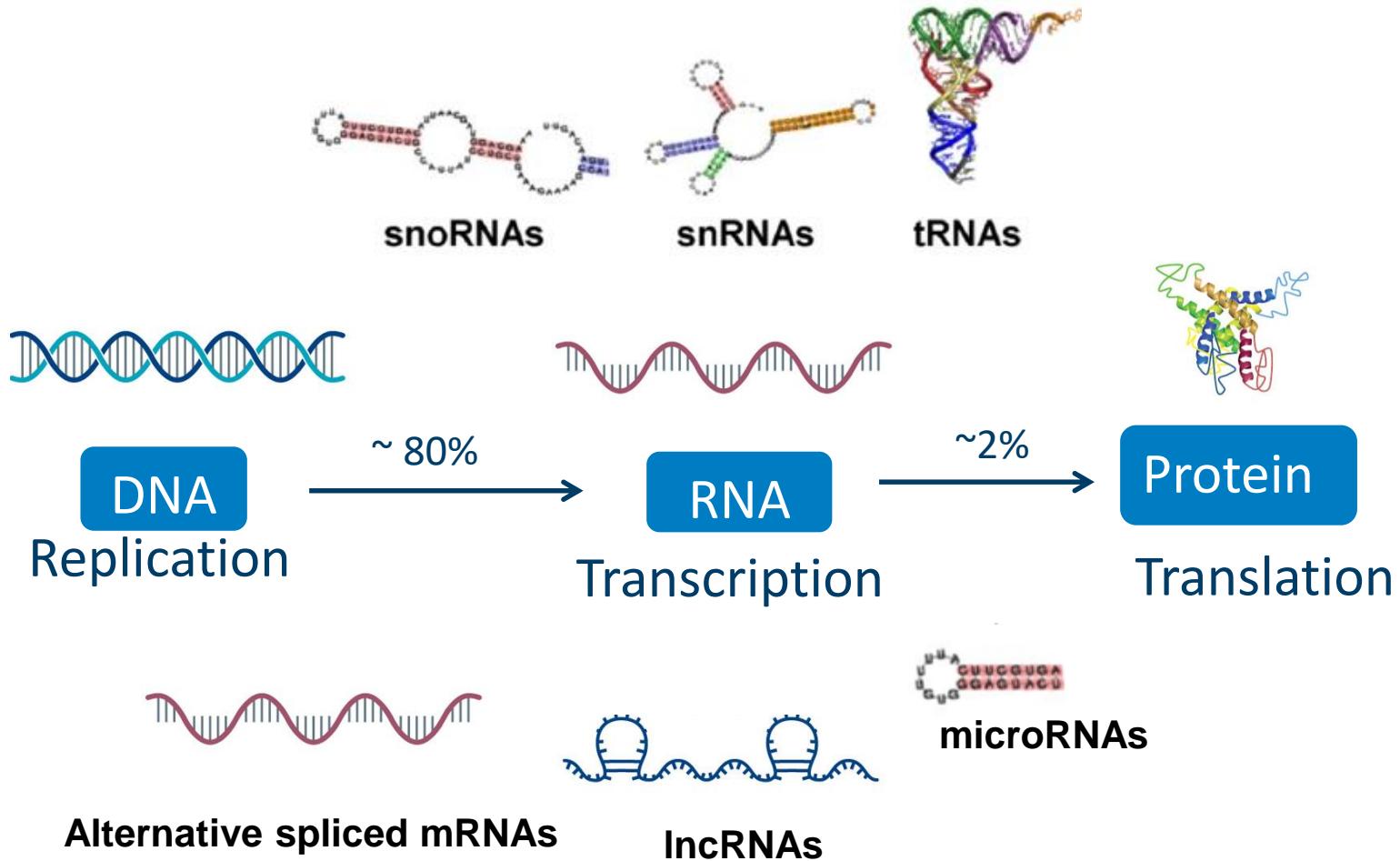
## GENE TALLY

Scientists still don't agree on how many protein-making genes the human genome holds, but the range of their estimates has narrowed in recent years.



©nature

# The Central Dogma of Molecular Biology



# Human Genome

- **Consists of ~3.3 billion basepairs**
- **Total ~20,000 protein-coding genes**
- **Whole-genome sequence equals ~825 Mb**
  - **Data storage requires Terabytes!**
- **Sequence variation: repeats/deletions, SNPs...**

**Single Nucleotide Polymorphisms:**  
*Specific nucleotides in the genome showing variation across the population in comparison to the reference sequence*

# Single Nucleotide Polymorphism

rs756599860

TGATGCCTTGTTATCTACTCAAG**A**CAGGATGAGGACTGGGTAAAGGAATG (**ref**)  
TGATGCCTTGTTATCTACTCAAG**C**CAGGATGAGGACTGGGTAAAGGAATG (**alt**)

# AIMS of this lecture

- Understanding **genomic variation**, SNPs
- Functional **relevant variation**
- Use of online **databases**

# Functional variants

- **Single Nucleotide Polymorphisms (SNPs)**
  - Intrinsic
  - Exonic
  - Promotor region
  - Regulatory elements

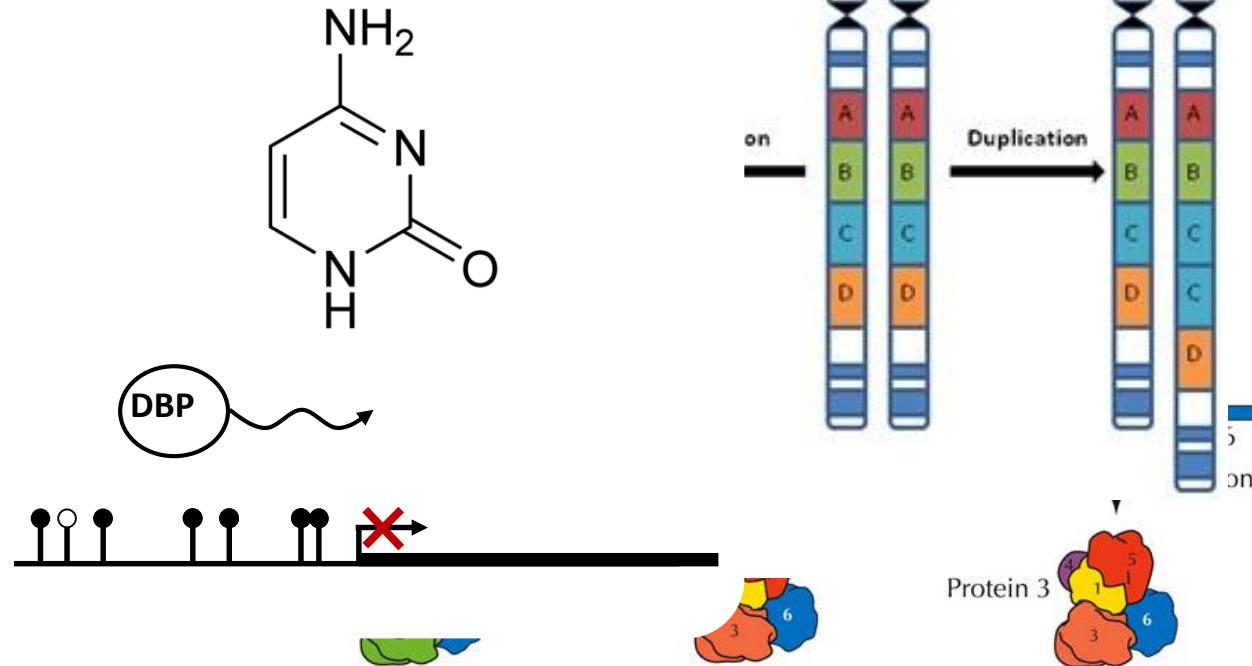
# Functional variants

- Single Nucleotide Polymorphisms (SNPs)
  - Intronic
  - **Exonic**
  - Promotor region
  - Regulatory elements

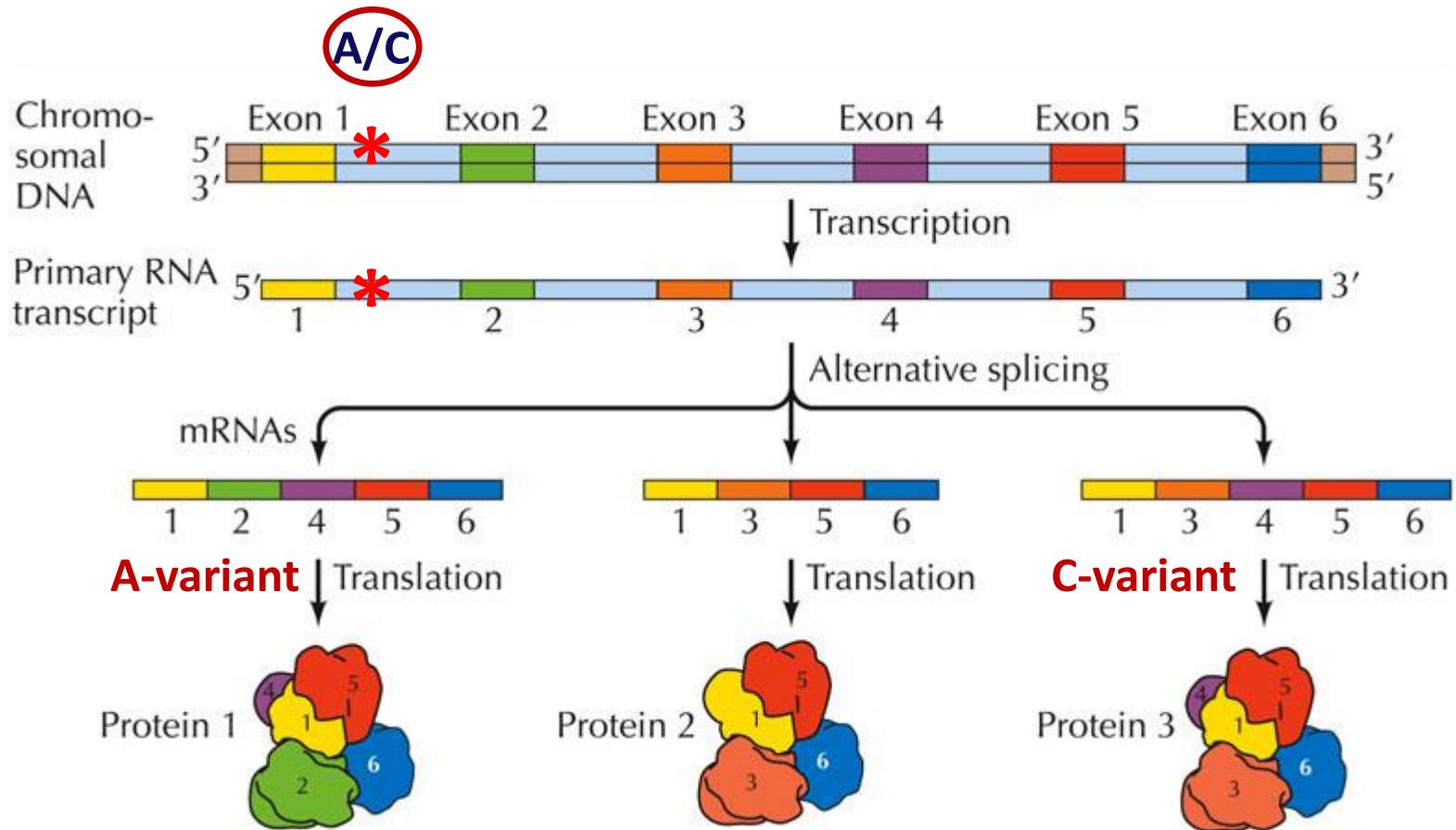
# Functional variants

- Single Nucleotide Polymorphisms (SNPs)

- Methylation
- Copy number variation (CNV)
- Alternative



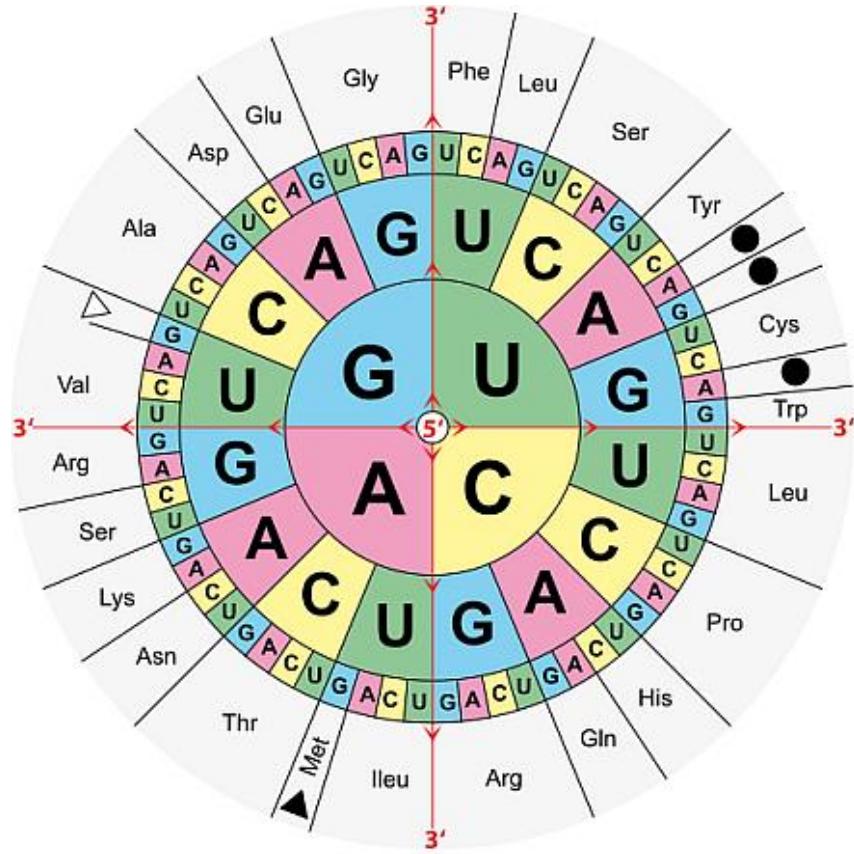
# Intronic SNPs



# Exonic SNPs

## Coding SNPs:

- Synonymous or non-synonymous



Codon: G A A Glu

Codon: G A G Glu

Codon: G G A Gly

Codon: T A A STOP

# Promotor SNPs

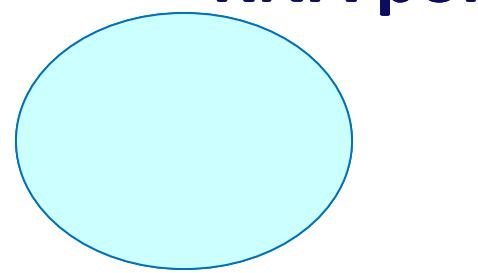
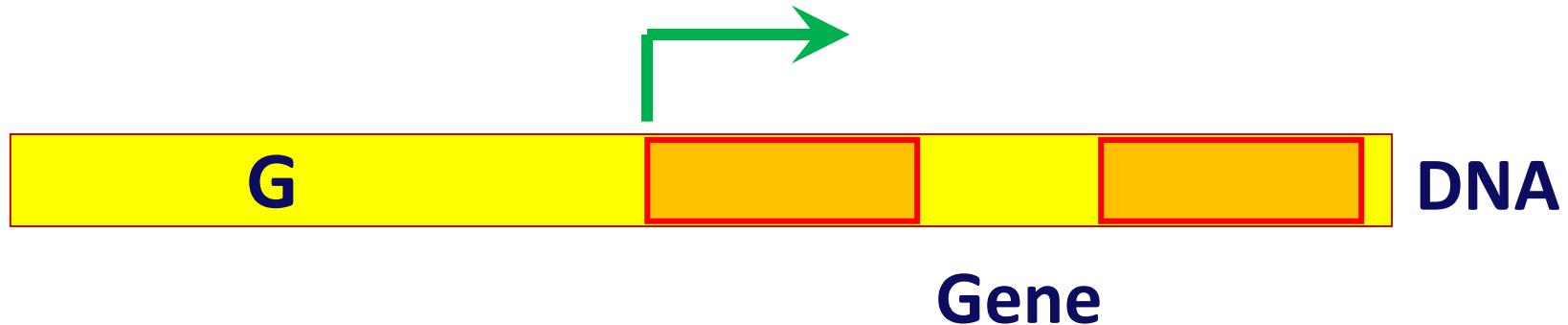
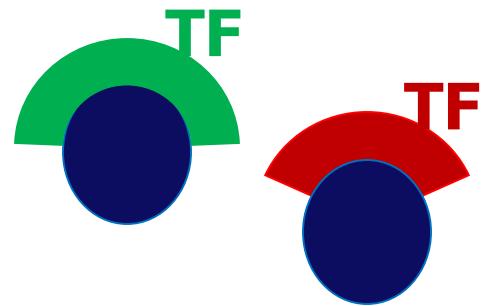
## Disruption of protein recognition sites



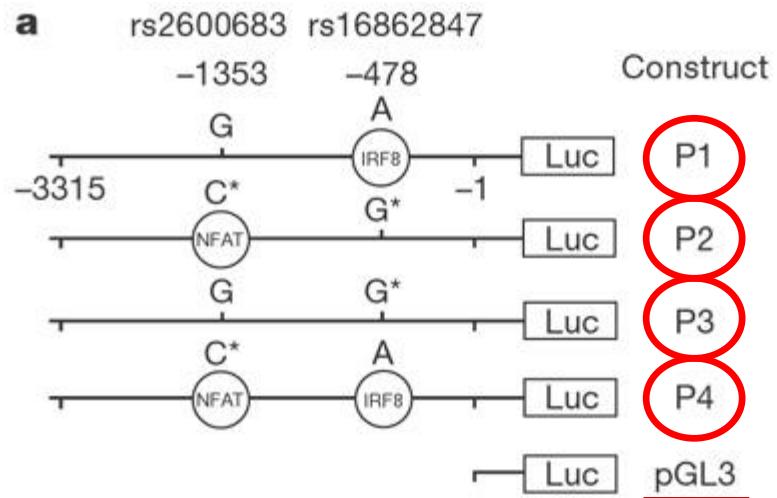
# Promotor SNPs

 *SNP G recognition*

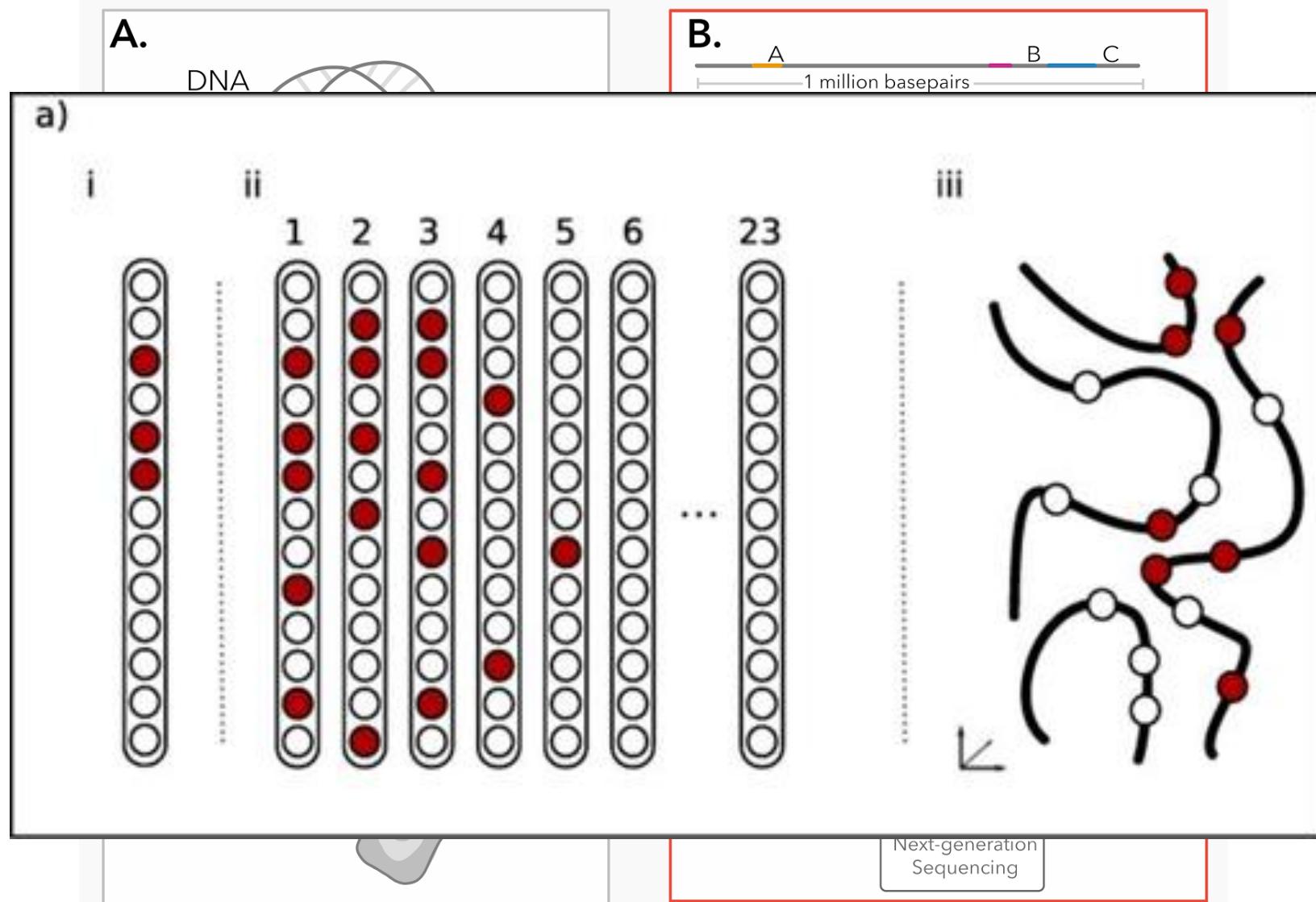
 *SNP A recognition*



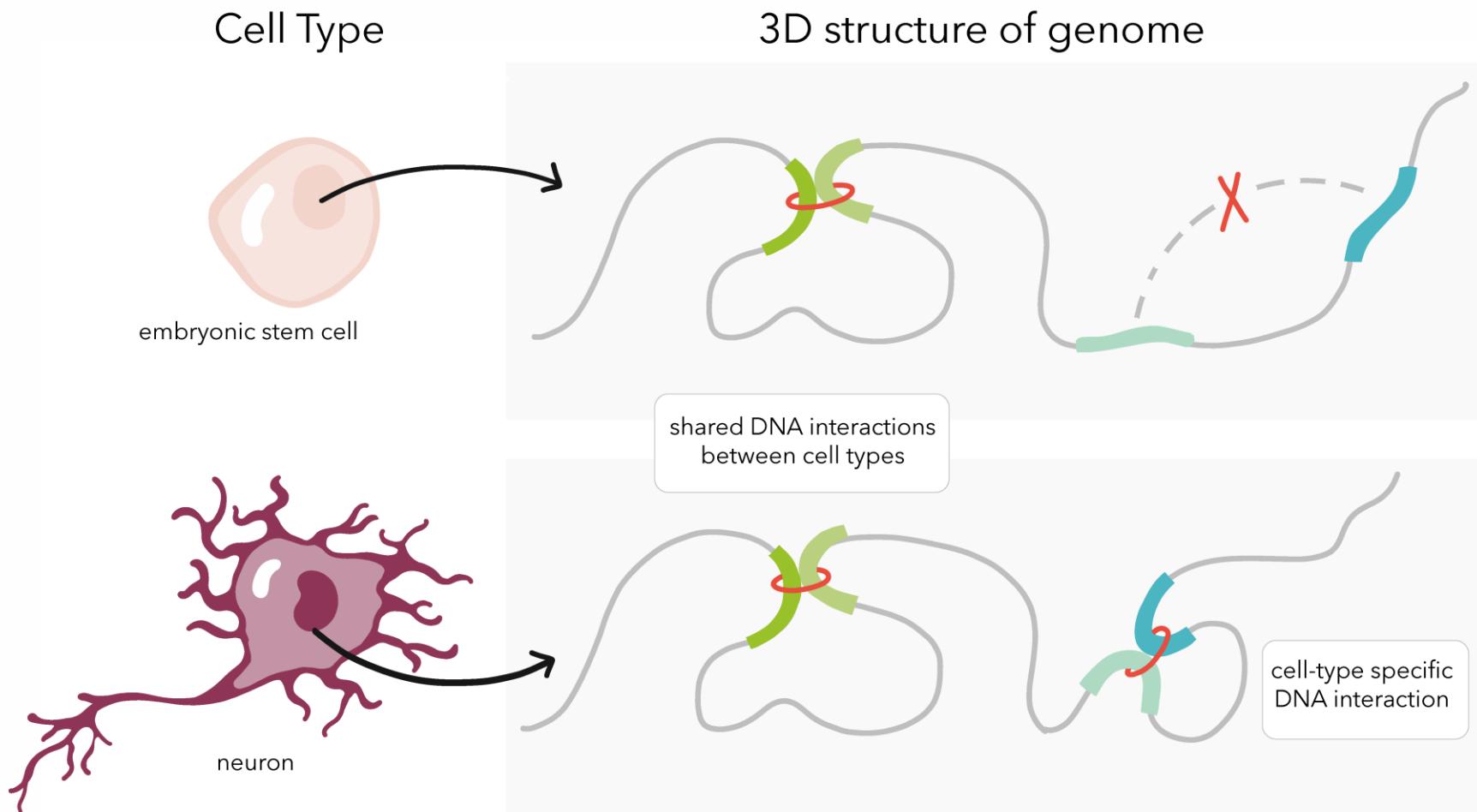
# Promotor SNPs



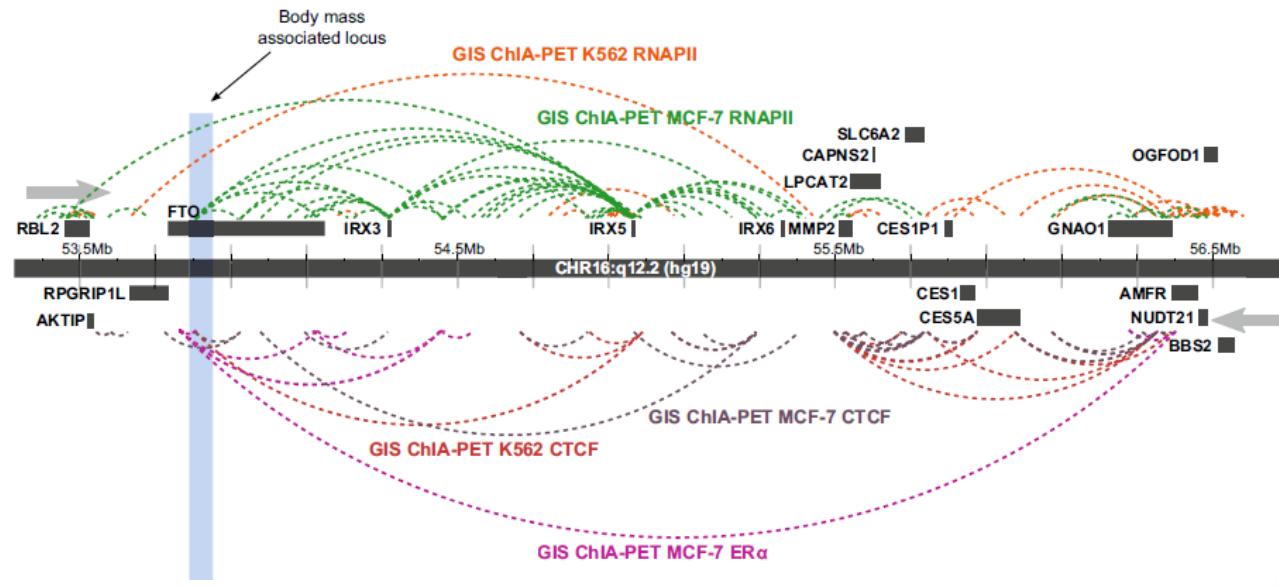
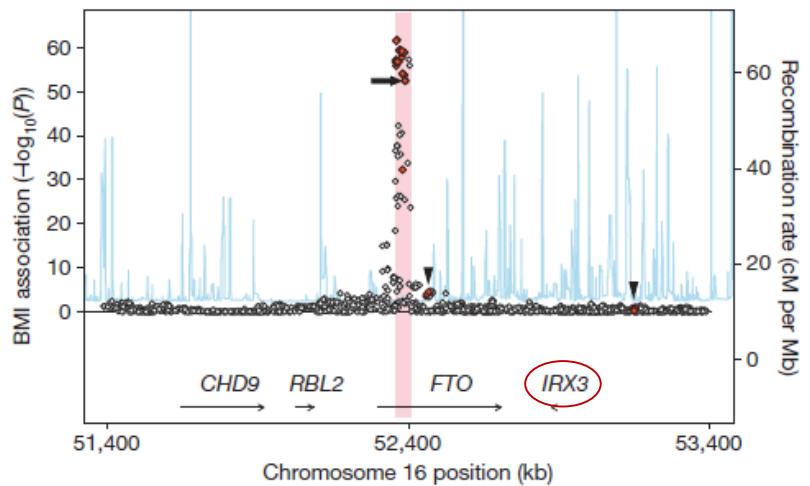
# Distal regulatory elements



# Distal regulatory elements



# Distal regulatory elements



## Functional SNPs

- Codon change
- Alternative splicing
- Variation in regulatory elements
  - Promotor
  - Distal regulatory element

# Example case – Genome Graphs



# Example case - Linkage Datafile

## LOD Score

▶ Pronunciation

LOD stands for "logarithm of the odds." In genetics, the LOD score is a statistical estimate of whether two genes, or a gene and a disease gene, are likely to be located near each other on a chromosome and are therefore likely to be inherited. A LOD score of 3 or higher is generally understood to mean that two genes are located close to each other on the chromosome. In terms of significance, a LOD score of 3 means the odds are a thousand to one that the two genes are linked, and therefore inherited together.

▶ Listen

Lawrence C. Brody,  
Ph.D. defines LOD  
Score



Profile

Illustration

3-D  
Animation



Lawrence C. Brody, Ph.D.

Chief & Senior Investigator, Genome Technology Branch; Head, Molecular Pathogenesis Section

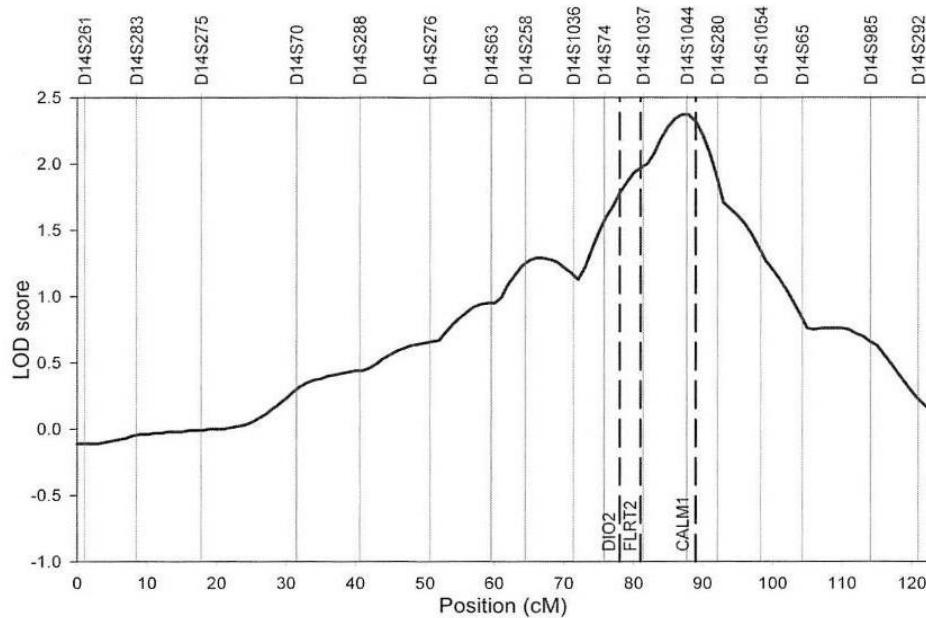
Dr. Brody investigates the genetics of breast cancer and neural tube defects. As chief of the NHGRI Genome Technology Branch's Molecular Pathogenesis section, he is interested in studying genetic mutations that lead to perturbations in normal metabolic pathways and cause disorders such as cancer and birth defects. His laboratory investigates mutations in two breast cancer-linked genes, breast cancer gene 1 (BRCA1) and breast cancer gene 2 (BRCA2). Dr. Brody's laboratory was among the first to report that women carrying BRCA1 or BRCA2 mutations have a higher risk of developing both breast and ovarian cancer than women without such mutations.

## LOD score:

logarithm ( $\log_{10}$ ) of the odds;  
for LOD=3 the odds is  $10^3$   
(1000) to 1 that the genomic  
region is linked with trait  
tested.

# Example case - Linkage Datafile

- Linkage at chromosome **14q32.11**
- 3 genes within 1-LOD drop interval
  - DIO2*
  - FLTR2*
  - CALM1*

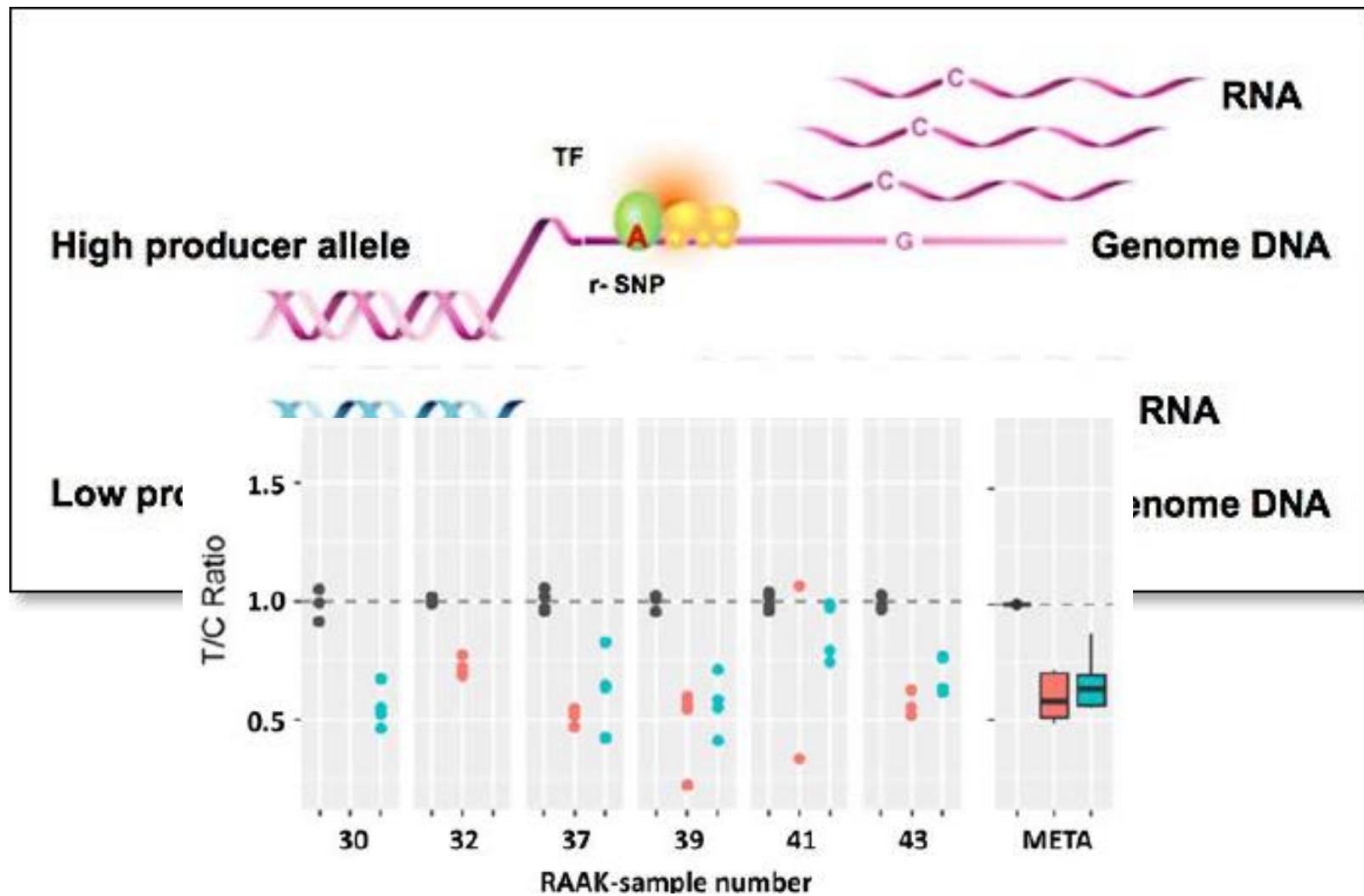


| Marker  | LOD   |
|---------|-------|
| D1S214  | 0     |
| D1S450  | 0     |
| D1S2667 | 0     |
| D1S2697 | 0     |
| D1S199  | 0.05  |
| D1S234  | 0.104 |
| D1S255  | 0     |
| D1S2797 | 0.006 |
| D1S2890 | 0.049 |
| D1S230  | 0.014 |
| D1S2841 | 0.326 |
| D1S207  | 0.298 |
| D1S2868 | 0.73  |
| D1S2793 | 0.599 |
| D1S206  | 0.876 |
| D1S495  | 0.733 |
| D1S2626 | 0.667 |
| D1S2778 | 0.218 |

## Functional SNPs

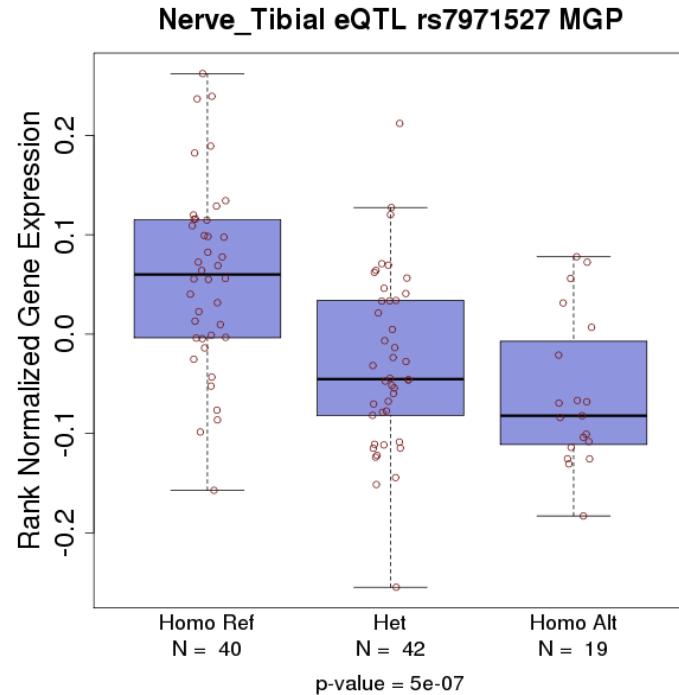
- Alternative splicing
- Codon change
- Variation in regulatory elements
  - Promotor
  - Distal regulatory element

# Allelic Imbalanced Expression

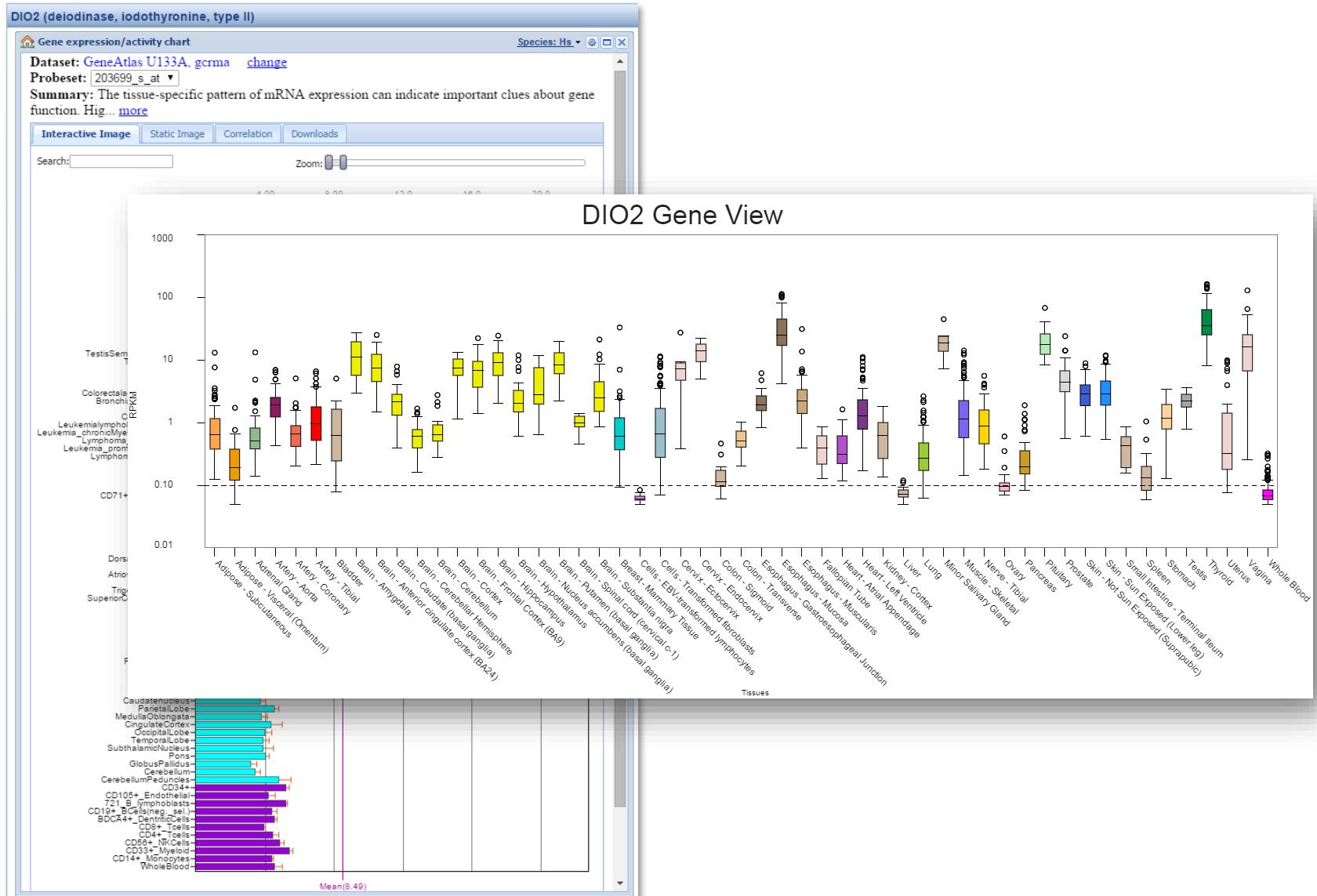


## Expression Quantitative Trait Locus

- SNP affects gene expression
- Either *in cis* or *in trans*



# Expression – Tissue of Interest (BioGPS & GTEx)

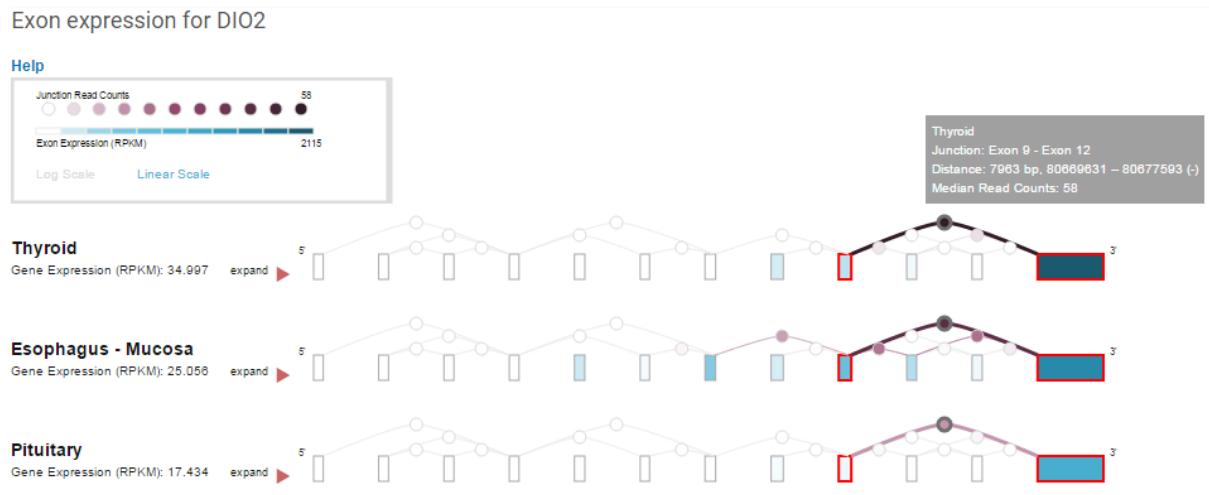


# Expression – Alternative splicing

**Ensembl** gives an  
overview of reported  
splice variants

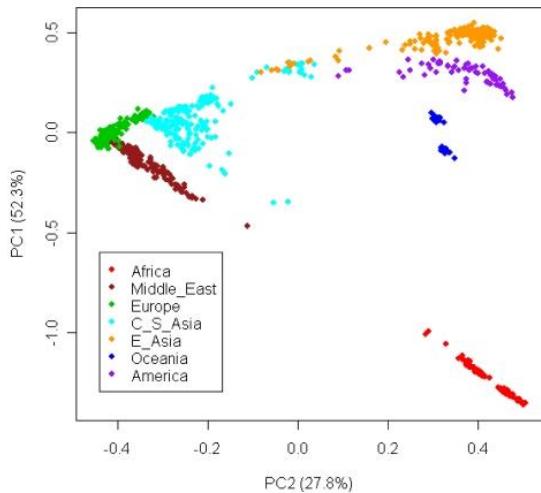
| Name     | Transcript ID                   | bp   | Protein    | Biotype                 | CCDS                      | UniProt                    | RefSeq   | Flags                         |
|----------|---------------------------------|------|------------|-------------------------|---------------------------|----------------------------|--|-------------------------------|
| DIO2-001 | <a href="#">ENST00000557010</a> | 6367 | 273aa      | Protein coding          | <a href="#">CCDS45146</a> | <a href="#">Q92813</a>     | <a href="#">NM_000793</a><br><a href="#">NP_000784</a>   | TSL:2 Gencode basic APPRIS P1 |
| DIO2-002 | <a href="#">ENST00000438257</a> | 6136 | 273aa      | Protein coding          | <a href="#">CCDS45146</a> | <a href="#">Q92813</a>     | <a href="#">NM_013989</a><br><a href="#">NP_054644</a>   | TSL:1 Gencode basic APPRIS P1 |
| DIO2-003 | <a href="#">ENST00000555750</a> | 1049 | 309aa      | Protein coding          | <a href="#">CCDS55934</a> | <a href="#">A0A0A0MTQ2</a> | <a href="#">NM_001007023</a><br><a href="#">NP_001007024</a>   | TSL:1 Gencode basic           |
| DIO2-201 | <a href="#">ENST00000422005</a> | 6272 | 145aa      | Protein coding          | -                         | <a href="#">J3KQY5</a>     | <a href="#">NM_001242502</a><br><a href="#">NM_001242503</a><br><a href="#">NP_001229431</a><br><a href="#">NP_001229432</a> | TSL:5 Gencode basic           |
| DIO2-004 | <a href="#">ENST00000556811</a> | 984  | 58aa       | Protein coding          | -                         | <a href="#">H0YJQ8</a>     | -  | CDS 5' incomplete TSL:1       |
| DIO2-006 | <a href="#">ENST00000554188</a> | 569  | 55aa       | Protein coding          | -                         | <a href="#">G3V3A8</a>     | -  | CDS 3' incomplete TSL:4       |
| DIO2-010 | <a href="#">ENST00000557125</a> | 517  | 56aa       | Protein coding          | -                         | <a href="#">G3V2A7</a>     | -  | TSL:3 Gencode basic           |
| DIO2-007 | <a href="#">ENST00000553594</a> | 491  | 6aa        | Protein coding          | -                         | -                          | -  | CDS 3' incomplete TSL:4       |
| DIO2-009 | <a href="#">ENST00000553968</a> | 417  | 6aa        | Protein coding          | -                         | -                          | -  | CDS 3' incomplete TSL:3       |
| DIO2-005 | <a href="#">ENST00000555844</a> | 776  | 37aa       | Nonsense mediated decay | -                         | <a href="#">H0YJ42</a>     | -  | CDS 5' incomplete TSL:1       |
| DIO2-008 | <a href="#">ENST00000556384</a> | 267  | No protein | Processed transcript    | -                         | -                          | -  | TSL:3                         |

# GTEx shows tissue specific splice information



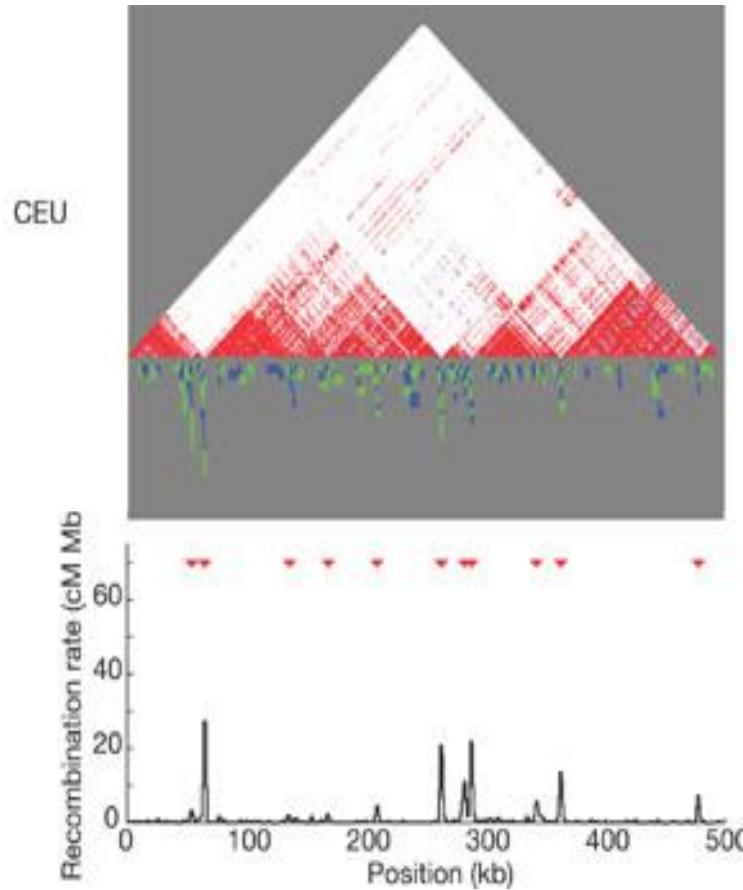
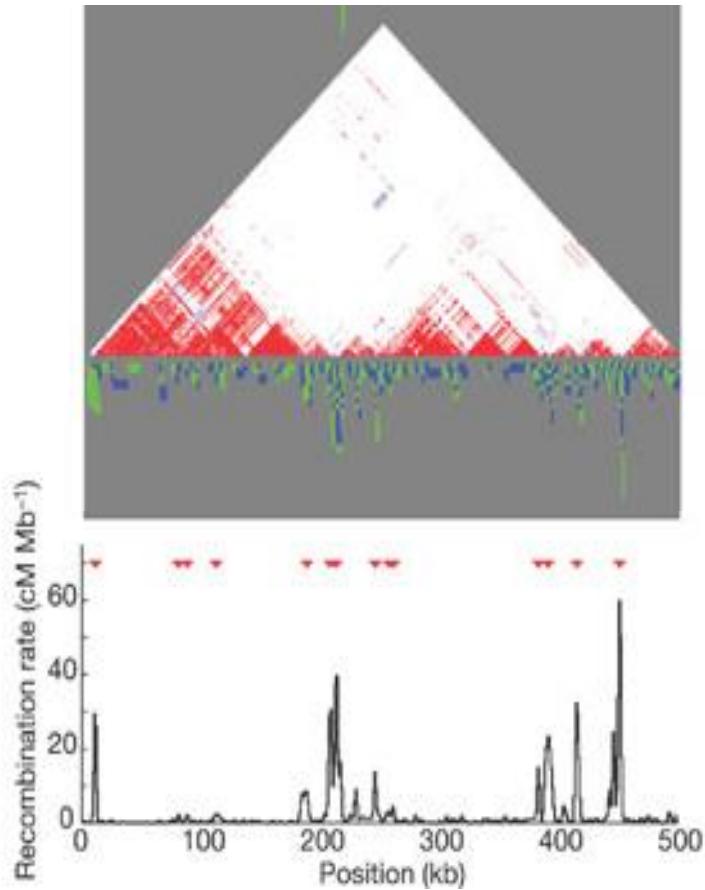
# Criteria relevant SNPs

- **Functional relevant variant**
- **Polymorphic in specific population**
- **Tagging SNP (possibly: tagging functional SNP)**
- **Assay design possible (repeat, GC-rich)**

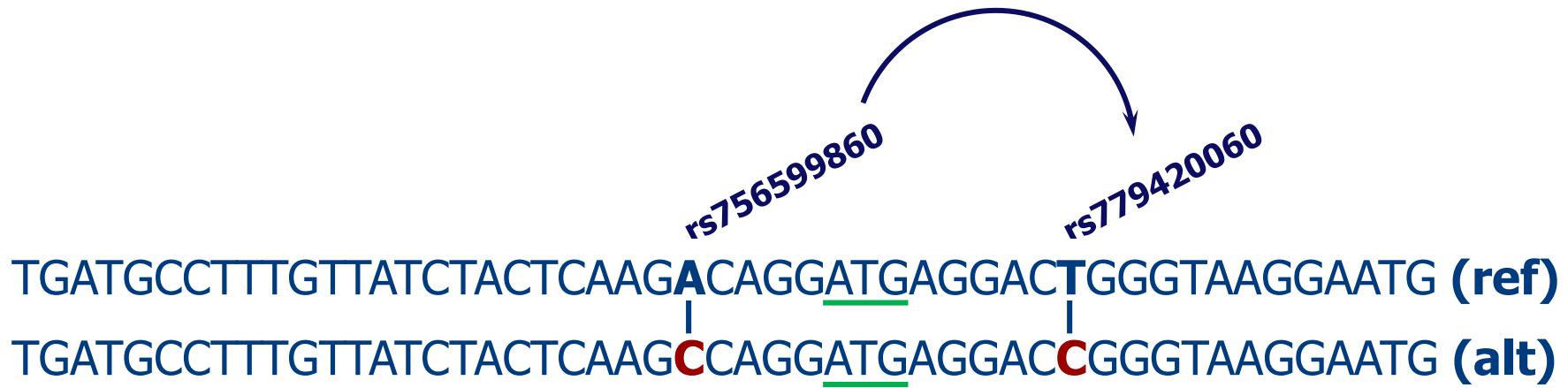


# Tagging SNP

Linkage across ‘blocks’: genetic variation is limited  
(several SNPs carry same information)



# Tagging SNP



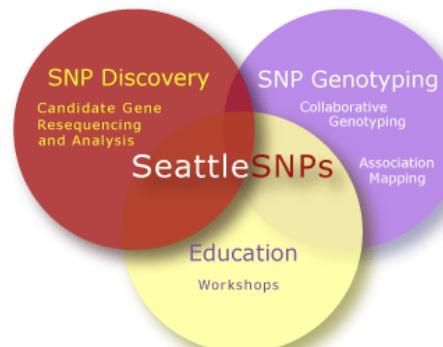
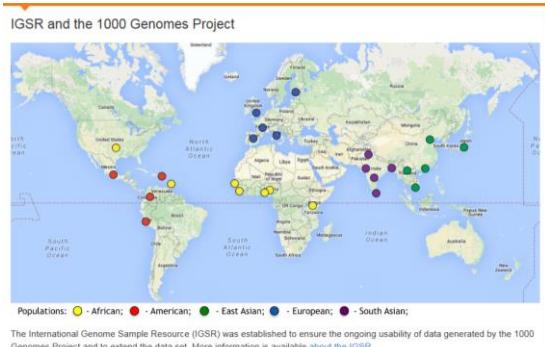
# AIMS of this lecture

- Understanding **genomic variation**, SNPs
- Functional **relevant variation**
- Use of online **databases**

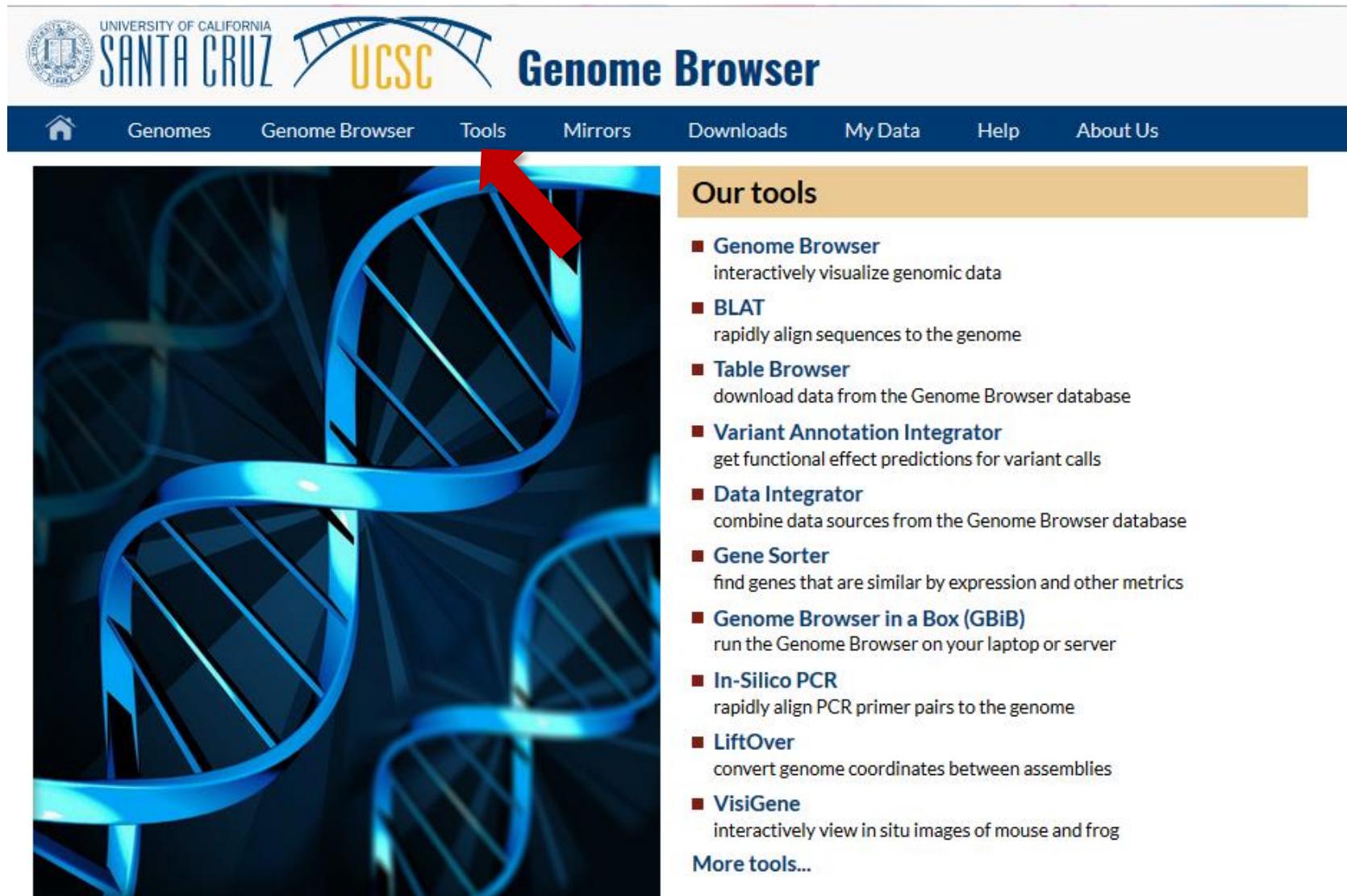
# Human Genome

- Consists of ~3.3 billion basepairs
- Whole-genome sequence equals ~825 Mb
  - Data storage requires Terabytes!
- Online databases contain information on thousands of individuals and millions of polymorphisms

Entrez Gene



# Example case – UCSC Genome Browser



The image shows the homepage of the UCSC Genome Browser. At the top left is the University of California Santa Cruz logo. Next to it is the text "UNIVERSITY OF CALIFORNIA SANTA CRUZ". To the right is the "UCSC" logo with a blue arch graphic. To the right of the logo is the text "Genome Browser". Below the header is a dark blue navigation bar with white text. From left to right, the menu items are: a house icon (Home), "Genomes", "Genome Browser", "Tools" (which has a red arrow pointing to it from the left), "Mirrors", "Downloads", "My Data", "Help", and "About Us". The main content area features a large, stylized blue DNA double helix on the left. To the right of the DNA is a yellow box containing the heading "Our tools" and a list of ten tools with descriptions. A "More tools..." link is at the bottom of the list.

**Our tools**

- **Genome Browser**  
interactively visualize genomic data
- **BLAT**  
rapidly align sequences to the genome
- **Table Browser**  
download data from the Genome Browser database
- **Variant Annotation Integrator**  
get functional effect predictions for variant calls
- **Data Integrator**  
combine data sources from the Genome Browser database
- **Gene Sorter**  
find genes that are similar by expression and other metrics
- **Genome Browser in a Box (GBiB)**  
run the Genome Browser on your laptop or server
- **In-Silico PCR**  
rapidly align PCR primer pairs to the genome
- **LiftOver**  
convert genome coordinates between assemblies
- **VisiGene**  
interactively view *in situ* images of mouse and frog

[More tools...](#)

# Surplus of information

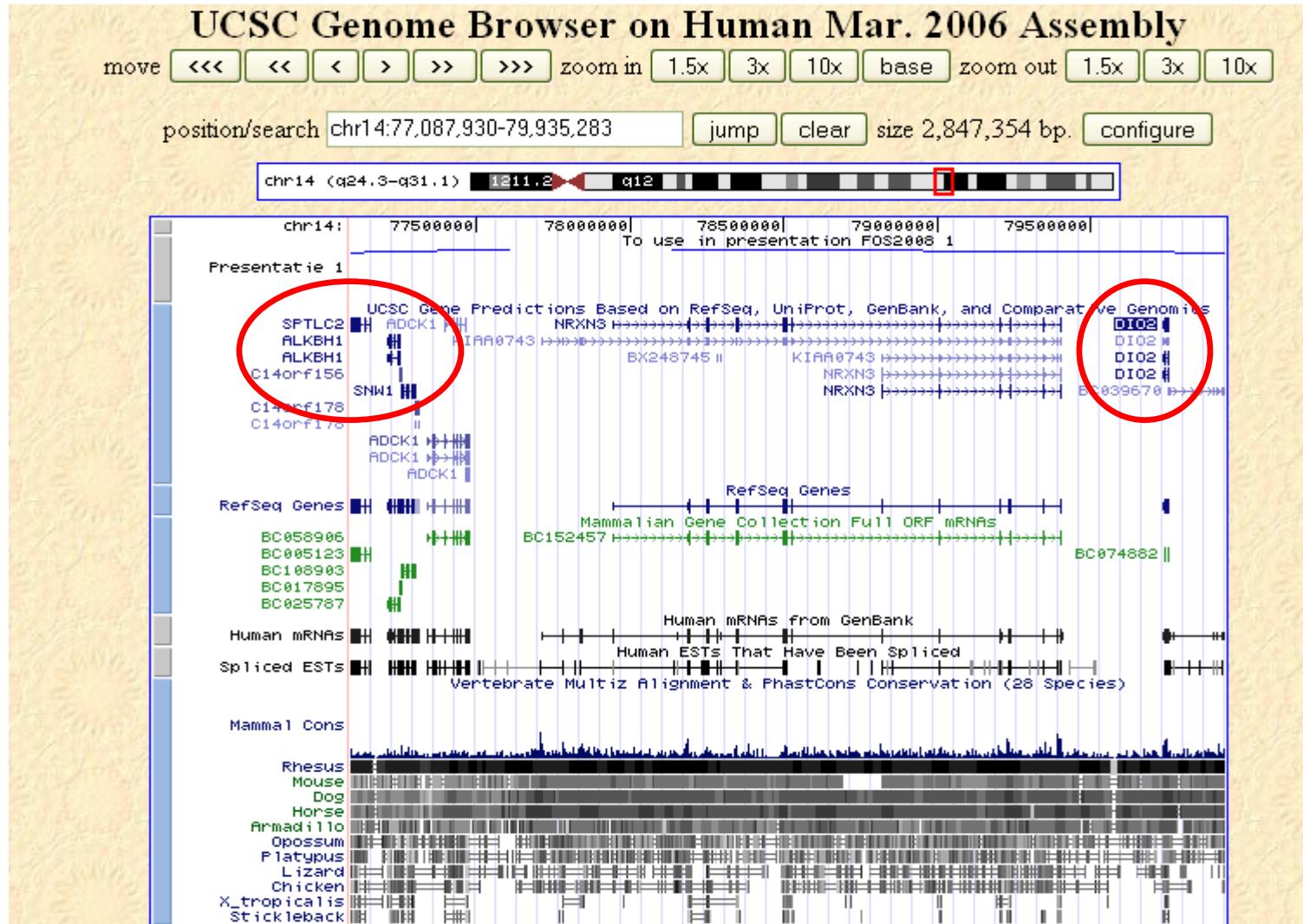
## Filter information by:

- Population of interest (CEU, CHB+JPT, etc.)
- Chromosome / locus
- Gene
- Intronic / Exonic / Promotor / 5' UTR/...
- Basepair

## (Some) criteria for disease relevant SNPs:

- If exonic: non-synonymous; damaging?
- If not: in regulatory element?

# Example case - Genome Browser



# Example case - Genome Graphs



# Example - Gene sorter

Home Genomes Blat Tables Gene Sorter PCR PDF/PS Session FAQ Help

## Finding Candidate Genes for Gene Sorter

Thresholding *Presentatie 1* at 3.4. There are 1 regions covering 4246031 bases.  
Installed a Gene Sorter filter that selects only genes in these regions.

[go to gene sorter](#)

Home UCSC Human Gene Sorter

genome Human assembly Mar. 2006 search uc001aaa.1 Go!  
sort by Expression (GNF Atlas2) configure filter (now on) display 50 output sequence text

| # | Name      | VisiGene | fetal brain | whole brain | amygdala | thymus | bone marrow | PB-CD4+ Tcells | skin | pancreatic islets | adipocyte | heart | lung | kidney | liver | ovary | testis | BLASTP E-Value | Genome | Position   | Description                              |
|---|-----------|----------|-------------|-------------|----------|--------|-------------|----------------|------|-------------------|-----------|-------|------|--------|-------|-------|--------|----------------|--------|------------|--|
| 1 | TMEM63C   | 172819   |             |             |          |        |             |                |      |                   |           |       |      |        |       |       |        | n/a            | chr14  | 76,756,757 | transmembrane protein 63C                |
| 2 | NGB       | 179188   |             |             |          |        |             |                |      |                   |           |       |      |        |       |       |        | n/a            | chr14  | 76,804,497 | neuroglobin                              |
| 3 | POMT2     | 179894   |             |             |          |        |             |                |      |                   |           |       |      |        |       |       |        | n/a            | chr14  | 76,834,011 | putative protein O-mannosyltransferase   |
| 4 | GSTZ1     | 176852   |             |             |          |        |             |                |      |                   |           |       |      |        |       |       |        | n/a            | chr14  | 76,862,400 | glutathione transferase zeta 1 isoform 1 |
| 5 | TMED8     | n/a      |             |             |          |        |             |                |      |                   |           |       |      |        |       |       |        | n/a            | chr14  | 76,895,098 | transmembrane emp24 domain containing 8  |
| 6 | C14orf174 | 177840   | n/a         |             |          |        |             | n/a            |      |                   |           |       |      |        |       |       |        | n/a            | chr14  | 76,920,427 | hypothetical protein LOC161394           |
| 7 | C14orf148 | 172443   | n/a         |             |          |        |             | n/a            |      |                   |           |       |      |        |       |       |        | n/a            | chr14  | 76,950,867 | hypothetical protein LOC122945           |
| 8 | C14orf133 | 168163   |             |             |          |        |             |                |      |                   |           |       |      |        |       |       |        | n/a            | chr14  | 76,978,214 | hypothetical protein LOC63894            |

# Description and known literature

- |  |   |
|--|---|
| <ul style="list-style-type: none"><li>• Gene function</li><li>• Literature</li><li>• Expression<ul style="list-style-type: none"><li>• eQTLs</li><li>• Tissue of interest</li><li>• Alternative splicing</li></ul></li><li>• Known genetic variation</li></ul> | <p><b>NCBI / OMIM</b></p> <p><b>PubMed</b></p> <p><b>Genevar / GTEx</b></p> <p><b>BioGPS / GTEx</b></p> <p><b>Ensembl / GTEx</b></p> <p><b>UCSC / HaploReg / SNPper</b></p> |
|--|---|

# NCBI – Entrez Gene

The screenshot shows the NCBI Entrez Gene search interface. The search bar at the top has 'Gene' selected and contains the query 'DIO2'. The results page for 'DIO2 deiodinase, iodothyronine, type II [Homo sapiens (human)]' (Gene ID: 1734) is displayed. On the left, a summary panel provides details like the official symbol (DIO2), full name (deiodinase, iodothyronine, type II), primary source (HGNC:HGNC\_2884), and gene type (protein coding). It also lists related genes (Ensembl:ENSG00000211448, HPRD:09027, MIM:601413, Vega:OTTHUMG00000171443), RefSeq status (REVIEWED), organism (Homo sapiens), and lineage (Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo). The 'Also known as' section includes D2, 5DII, SelY, DIOII, TXDII, and a detailed summary of the gene's function. Orthologs for mouse and all other species are listed. On the right, a sidebar titled 'Table of contents' lists various sections such as Summary, Genomic context, Genomic regions, transcripts, and products, Bibliography, Variation, Pathways from BioSystems, Interactions, General gene information, General protein information, NCBI Reference Sequences (RefSeq), Related sequences, and Additional links. A second sidebar titled 'Related information' lists Order cDNA clone, BioAssay, BioAssay by Target (List), BioAssay by Target (Summary), and BioAssay by Gene target.

GeneRIFs!

Gene Reference into Function

# Literature - PubMed

NCBI Resources How To

wdenhollander1 My NCBI Sign Out

PubMed DIO2 Search Help

Article types Clinical Trial Review Customize ...

Text availability Abstract Free full text Full text

Publication dates 5 years 10 years Custom range...

Species Humans Other Animals

[Clear all](#) [Show additional filters](#)

Summary ▾ 20 per page ▾ Sort by Most Recent ▾ Send to: ▾ Filters: [Manage Filters](#)

See 158 articles about **DIO2 gene function**  
See also: [DIO2 deiodinase, iodothyronine, type II](#) in the Gene database  
[dio2](#) in [Homo sapiens](#) [Mus musculus](#) [Rattus norvegicus](#) All 148 Gene records

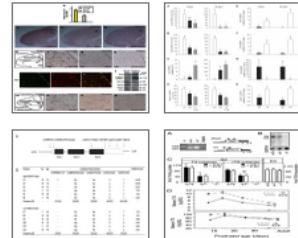
**Results: 1 to 20 of 267** << First < Prev Page  of 14 Next > Last >>

- [Disruption of Type 2 Iodothyronine Deiodinase Activity in Cultured Human Glial Cells by Polybrominated Diphenyl Ethers.](#)  
Roberts SC, Bianco AC, Stapleton HM.  
Chem Res Toxicol. 2015 Jun 2. [Epub ahead of print]  
PMID: 26004626  
[Similar articles](#)
- [An improved non-radioactive screening method identifies genistein and xanthohumol as potent inhibitors of iodothyronine deiodinases.](#)  
Renko K, Schäche S, Hoefig CS, Welsink T, Schwiebert C, Braun D, Becker NP, Koehrlé J, Schomburg L.  
Thyroid. 2015 May 12. [Epub ahead of print]  
PMID: 25962824  
[Similar articles](#)
- [Circadian synchronization determines critical day length for seasonal responses.](#)  
Majumdar G, Trivedi AK, Gupta NJ, Kumar V.  
Physiol Behav. 2015 Aug 1;147:282-90. doi: 10.1016/j.physbeh.2015.05.005. Epub 2015 May 7.  
PMID: 25957913  
[Similar articles](#)
- [Microcystin-RR exposure results in growth impairment by disrupting thyroid endocrine in zebrafish larvae.](#)  
Xie L, Yan W, Li J, Yu L, Wang J, Li G, Chen N, Steinman AD.  
Aquat Toxicol. 2015 Jul;164:16-22. doi: 10.1016/j.aquatox.2015.04.014. Epub 2015 Apr 13.  
PMID: 25897773  
[Similar articles](#)
- [Dual signal transduction pathways activated by TSH receptors in rat primary tanycyte cultures.](#)  
Bolborea M, Helfer G, Ebding FJ, Barrett P.  
Mol Endocrinol. 2015 Jan;29(1):15-26. doi: 10.1210/me.2014-1182. Epub 2014 Dec 15.

**New feature**  
Try the new Display Settings option - [Sort by Relevance](#)

**Related searches**  
[dio2 osteoarthritis](#)  
[dio2 brown adipose](#)  
[dio2 knockout](#)  
[dio2 polymorphism](#)  
[dio2 promoter](#)

**PMC Images search for DIO2**



[See more \(245\)...](#)

**Titles with your search terms**  
Common variation in the DIO2 gene predicts baseline psyche [J Clin Endocrinol Metab. 2009]  
Diet-induced obesity mediated by the JNK/DIO2 signal transduction pathway. [Genes Dev. 2013]  
Mice with targeted disruption of the Dio2 gene

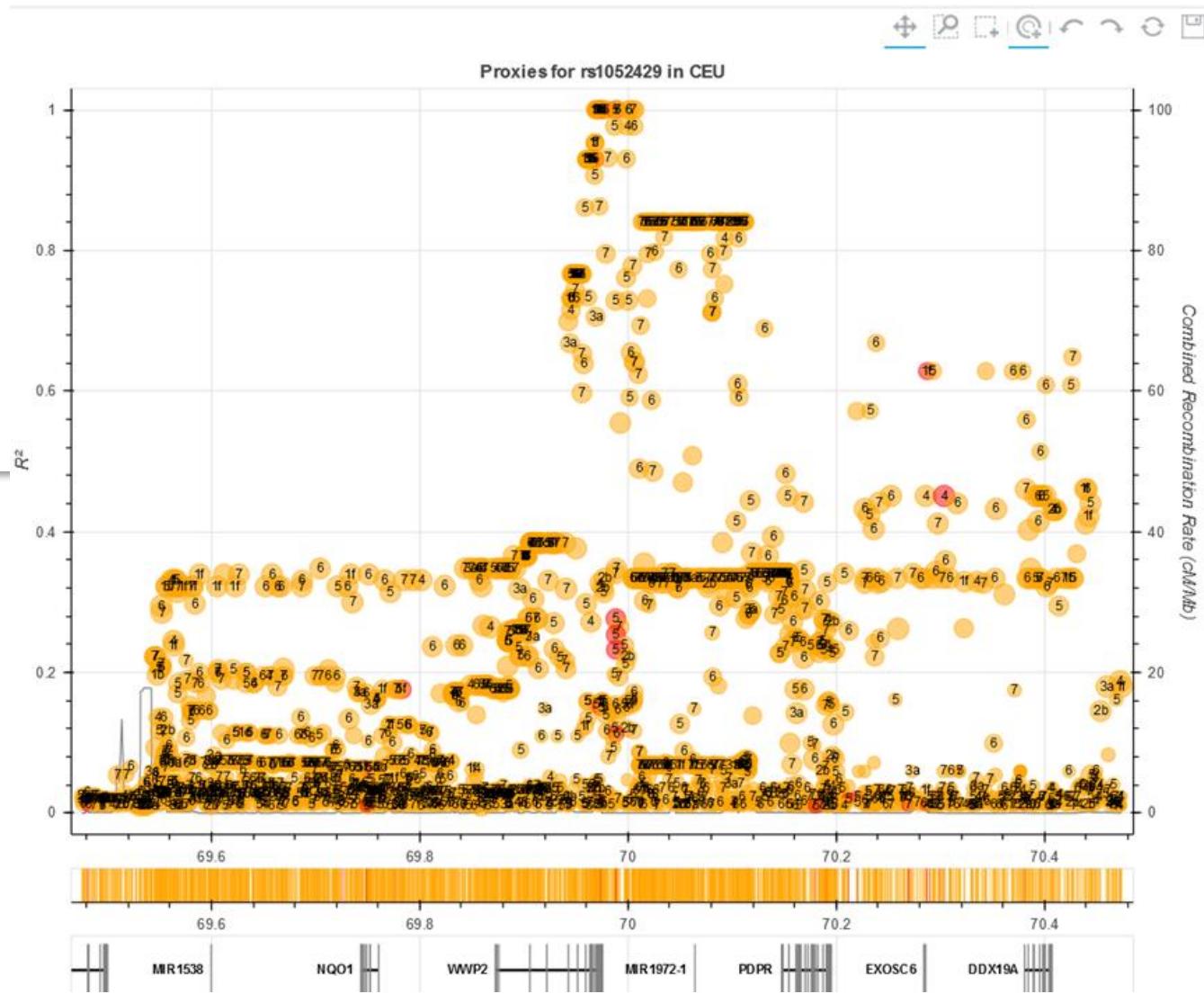
# Known genetic variation - HaploREG

- SNPs in LD
- Alleles
- Population frequencies
- Protein binding
- DNase sensitivity
- ...

Query SNP: rs225014 and variants with  $r^2 >= 0.8$

| pos (hg19)      | pos (hg38)     | LD (r <sup>2</sup> ) | LD (D') | variant    | Ref | Alt | AFR freq | AMR freq | ASN freq | EUR freq | SiPhy cons | Promoter histone marks | Enhancer histone marks | DNase        | Proteins bound   | eQTL tissues           | Motifs changed | Drivers disrupted | GENCODE genes | dbSNP func annot |
|-----------------|----------------|----------------------|---------|------------|-----|-----|----------|----------|----------|----------|------------|------------------------|------------------------|--------------|------------------|------------------------|----------------|-------------------|---------------|------------------|
| chr14:80634365  | chr14:80168022 | 0.84                 | 0.96    | rs4903903  | T   | C   | 0.22     | 0.41     | 0.35     | 0.32     | purple     | FAT                    | MUS,PANC,MUS           |              |                  | STAT                   |                | 30kb 3' of DIO2   |               |                  |
| chr14:80635104  | chr14:80168761 | 0.84                 | 0.96    | rs12588985 | A   | G   | 0.22     | 0.41     | 0.35     | 0.32     |            |                        |                        |              |                  | STAT                   |                | 29kb 3' of DIO2   |               |                  |
| chr14:80638440  | chr14:80172097 | 0.84                 | 0.96    | rs1491504  | G   | T   | 0.22     | 0.41     | 0.35     | 0.32     |            |                        |                        |              |                  | 10 altered motifs      |                | 25kb 3' of DIO2   |               |                  |
| chr14:806411302 | chr14:80174959 | 0.84                 | 0.96    | rs4899763  | C   | T   | 0.22     | 0.41     | 0.35     | 0.32     |            |                        |                        |              |                  | SRF                    |                | 23kb 3' of DIO2   |               |                  |
| chr14:80641672  | chr14:80175329 | 0.84                 | 0.96    | rs74064450 | T   | C   | 0.22     | 0.41     | 0.35     | 0.32     |            |                        |                        |              |                  | Arid5b,CHOP::CEBPalpha |                | 22kb 3' of DIO2   |               |                  |
| chr14:80642320  | chr14:80175977 | 0.84                 | 0.96    | rs2216086  | C   | T   | 0.35     | 0.43     | 0.35     | 0.32     |            |                        |                        |              |                  | GR                     |                | 22kb 3' of DIO2   |               |                  |
| chr14:80642939  | chr14:80176596 | 0.84                 | 0.96    | rs4899764  | G   | T   | 0.22     | 0.41     | 0.35     | 0.32     |            |                        |                        |              |                  | 5 altered motifs       |                | 21kb 3' of DIO2   |               |                  |
| chr14:80643110  | chr14:80176767 | 0.84                 | 0.96    | rs4899765  | C   | T   | 0.22     | 0.41     | 0.35     | 0.32     |            |                        |                        |              |                  | 11 altered motifs      |                | 21kb 3' of DIO2   |               |                  |
| chr14:80648603  | chr14:80182260 | 0.84                 | 0.96    | rs2005885  | C   | G   | 0.22     | 0.42     | 0.41     | 0.32     |            |                        |                        |              |                  | E2A,Pitx2,SETDB1       |                | 15kb 3' of DIO2   |               |                  |
| chr14:80649239  | chr14:80182896 | 0.83                 | 0.96    | rs759441   | T   | A   | 0.21     | 0.42     | 0.41     | 0.31     |            |                        |                        |              |                  | Hand1,Pax-6,Zbtb12     |                | 15kb 3' of DIO2   |               |                  |
| chr14:80649565  | chr14:80183222 | 0.84                 | 0.96    | rs4903904  | G   | A   | 0.22     | 0.42     | 0.41     | 0.32     |            |                        |                        |              |                  | Eif3,STAT              |                | 14kb 3' of DIO2   |               |                  |
| chr14:80655946  | chr14:80189603 | 0.82                 | 0.95    | rs74064456 | A   | T   | 0.23     | 0.42     | 0.44     | 0.32     |            |                        |                        |              |                  |                        |                | 7.9kb 3' of DIO2  |               |                  |
| chr14:80658261  | chr14:80191918 | 0.86                 | 0.96    | rs56017760 | T   | C   | 0.23     | 0.43     | 0.44     | 0.32     |            |                        |                        |              |                  | COMP1,CTCF             |                | 5.6kb 3' of DIO2  |               |                  |
| chr14:80660670  | chr14:80194327 | 0.84                 | 0.96    | rs74064457 | G   | A   | 0.24     | 0.42     | 0.44     | 0.32     |            |                        |                        |              |                  | Sox                    |                | 3.2kb 3' of DIO2  |               |                  |
| chr14:80661203  | chr14:80194860 | 0.86                 | 0.95    | rs35191251 | A   | C   | 0.23     | 0.43     | 0.44     | 0.33     |            |                        |                        |              |                  |                        |                | 2.7kb 3' of DIO2  |               |                  |
| chr14:80662335  | chr14:80195992 | 0.85                 | 0.99    | rs56025506 | C   | T   | 0.23     | 0.41     | 0.43     | 0.30     |            |                        |                        |              |                  | BAF155,Nkx3,Pou5f1     |                | 1.5kb 3' of DIO2  |               |                  |
| chr14:80667579  | chr14:80201236 | 0.87                 | 0.99    | rs225015   | G   | A   | 0.48     | 0.45     | 0.43     | 0.31     |            |                        |                        |              |                  | BCL,STAT               |                | DIO2              | 3'-UTR        |                  |
| chr14:80669580  | chr14:80203237 | 1                    | 1       | rs225014   | T   | C   | 0.50     | 0.46     | 0.44     | 0.34     |            |                        |                        |              |                  | RXRA                   |                | DIO2              | missense      |                  |
| chr14:80673242  | chr14:80206899 | 0.94                 | 1       | rs12437279 | C   | T   | 0.23     | 0.42     | 0.43     | 0.32     | purple     |                        | 4 organs               | ESDR,BRN,BRN | 6 bound proteins |                        |                |                   | DIO2          | intronic         |

# Known genetic variation - LDlink



# Subsequent steps

- Check for redundancy (LD):
  - Ldlink, HaploReg
  - Design assay
    - If unsuccessful, use alternative SNPs (based on LD)
  - Measure SNPs
- Determine functionality: wet lab!

QUESTIONS?

## Finding genes in practice

Yolande F. M. Ramos  
[y.f.m.ramos@lumc.nl](mailto:y.f.m.ramos@lumc.nl)  
Molecular Epidemiology  
LEIDEN, THE NETHERLANDS



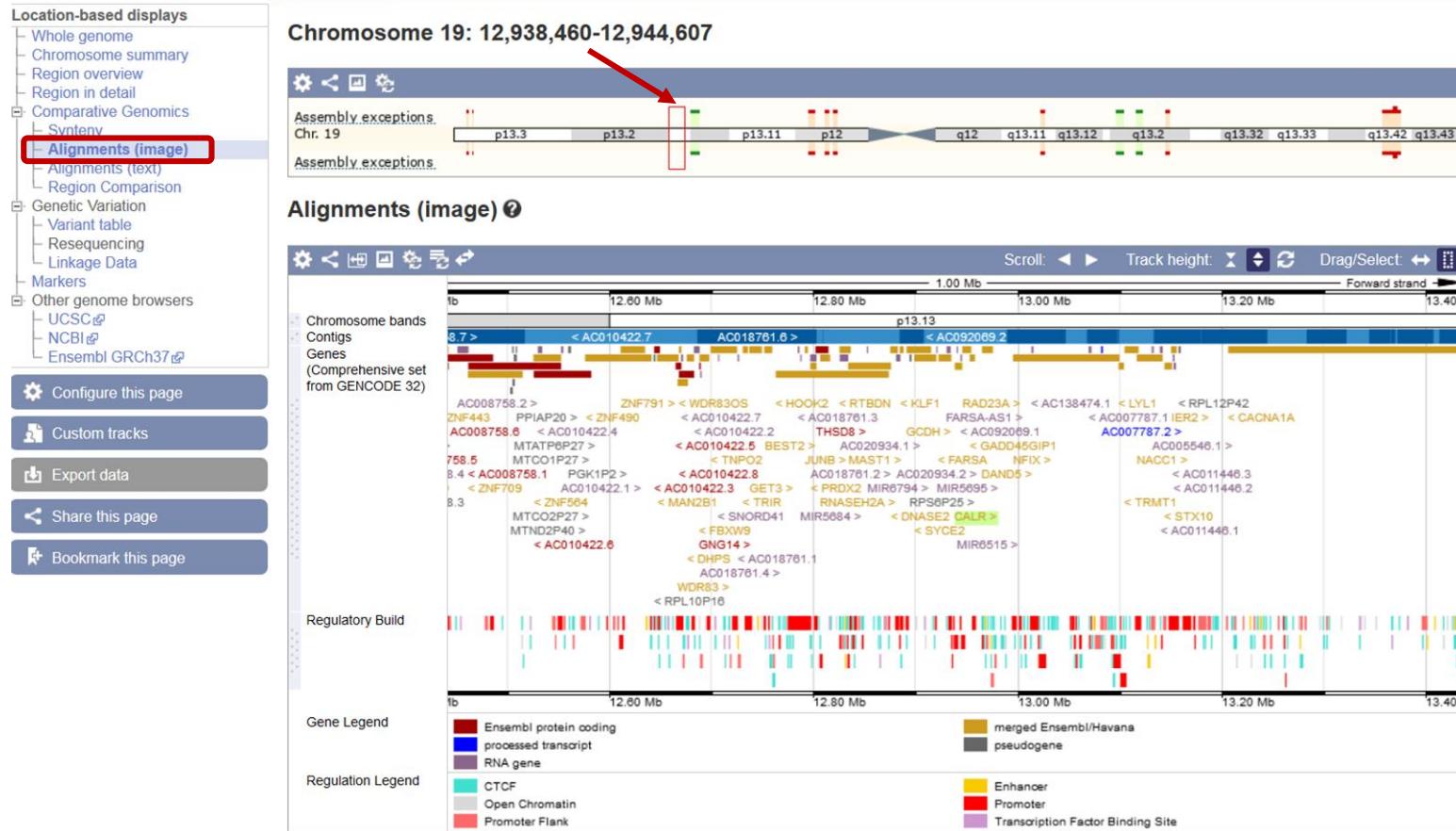
# AIMS of this practical session

- Understanding **genomic variation**, SNPs
- Functional **relevant variation**
- Use of online **databases**

# Questions & Exercise

- **Databases; hands on experience**
- **Select your own dataset (Data\_x) and follow manual**
- **Share tips & tricks**
- **Sometimes the web-based tools are a bit messy**
- **Ask questions!**

# Ensemble – an example



# Ensemble – an example

## TEXT

**Alignment**

Currently showing full alignment. Please click the button below to show the alignment for first 120 columns.

[Hide full alignment](#)

Human CCTAACATAGTGAACCGACGAAGCTCCAATGGAAAAAGACGCCATGGCATGACCRATGACAAGTGGCAGGGCCGGGCCAAGGCCTGGGTCAAGGTTGGTTGAGAGGCGGGTGGGT  
Mouse CCTAACTTGCTGAGCAAATGGAAAGGCATGGAAAGGCACAGTCAGGTCTAACCAAGCTCAAAGGACCGAGGGCCGGCTCAGCGCTGTGCAAGGTTGGGTGAGAGGTAGGTGAAT

Human ATAAAAGTCAAGGGCGCCGGCGCTCCGTACTGCAGAGCCGCTGCCGAGGGCTGTTAAAGGGCCCGCCTGGCCGCCCTCGGCCGCCATGCTATCCGTGCCGT  
Mouse ATAAA-TTGAACGCCGCTGGCCCGCTCCGTCAATACCGCACAGGCCGCTGCCGAGATCGTTAAAGGCCTGTGTCGCCGCCCTCGGCCGCCATGCTCTTGGGTGCCGT

Human GCTGCTCGGCTCCCTGCCCTGCCGCTGCCGAGCCGCTACTCAAGGAGCAGTTCTGGACGGAGGTAACGCTGGTCCCCTCGAGGCCGCCGACGACGCCGGCCCC  
Mouse CCTGCTTGGCTCCCTGCCCTGCCGAGGCCGACCCCTGCCATCTATTAAAGGACAGTCTTGGACGGAGGTAAGGCTGGGCCGCTCGAGGCCGCTTAG-CGACT--GCTGGGCC

Human CGATCCTGGATCTGGCTGGCGC---CGTAAATTACCGTTAGAGGTCCACACGGTGGCTCC-CGGGACTAGAGGCCGGGCGATTCTCTCTCGCTGCCCTGGGAGGCCGGAGG  
Mouse CGAGACCTCGGCTGCCCTCGCTGTGTAATGAAATTACCGTTAGAGGCCACACTGTGCCCTAGGGGACTAGGGCCGGGCTCATGTGCGCTTGTGGACTAGGAAG

Human GCGTAGCCCTCCGGCGGAGTTAGGGCTGAGGATCTGAAGGCACCCAGCTGAAACTAGAGGTGGAATGGGAGTGGGAGTCTGGGGATCTCTTCTGCCCCACAG  
Mouse GCTAGCCCTCCGGCGGAGTTAGGAATTCTGCAAGGACT--GAAGCCATCGAGATGTCAGTTGGAGTGGGATGGAA-----TCCCTTCCCCACCCACCTG

Human CTTGTCCTCGCAGATGTTGGTGGGGGG-GGGATTAGCACAGCGCTCTGACCTACCCCTTAATCCCCACT-AGACGGGTTGACTCCCGCTGGATCGAATCCAACACAAG  
Mouse CTTGTCAGCTCGCAGACGTTGGTTATGGGGCGGGGCTGAGGCCGCGCTTAATCTGACTCTATGCCCTTCTGAGCTGGACCAACCGCTGGTGAATCCAACATAAG

Human TCAGATTTGGCAATTCGTTCTCAGTCCGGCAAGTTCTACGGTGACGGAGAAAGATAAGGTAAGAGCCTAGGAGTGGGTCTCAGATCAGGGAGGACTTCTGGAGAAGCTCTT  
Mouse TCCGATTTGGCAATTTGTCCTCAGTCTGGCAATTTACGGGGACTGGAGAAGGATAAGGATACAAG----GGAGTGGGTGCAAGGACAACCTCTGGAGGAAGCTGTA

Human GTCTGTACACACACA-GCCGGGACAGTCCCC-----TTGGAG--GAGGACAGGTGGAGGAAGTGGGGAG-TCTTCTCTATTC-TCTAAGTCGAGGGTCTCGCGAGTCAGGGCCA  
Mouse GTCAAGCACATACATTTGTTGGGCTTTCAGGAAACCGGGAGCAGGACAGTGGAGGAAGGAAGGGAGTCAGGCCGCCCCCTCTAGGTCTAGGGT

Human ACGGTGACCTCACTACCGCTCCGCTCAGGTTTGCAAGRCAGGCCAGGATGCAAGCTTTATGCTGTCGGCAGTTGAGCCCTTCAGCAACAAAGGCCAGACGGCTGGGGAGCTT  
Mouse CC-TGATGACAGTCCATTTCATCACAGGGCTGCAAGCAAGCCAGATGCCGAAATTGCAACCCCTTCAGCAATTAAGGGCCAGACACTGGTGGTACAGTT

Human CACGGTGAACATGAGCAGAACATCGACTGTGGGGGCGGGTATGTAAGGCTTCTTAATAGTTGGACCAGACAGACATGCACGGAGACTCAGAAATACAACATCATGTTGGTGAAGGG  
Mouse CACGGTGAAGCATGAGCAGAACATCGACTGTGGGGGCGGGTACGTTGAAGGCTGTTGGAGGAGATCGACAGGAGACTCAGAAATACAACATCATGTTGGTGAAGGG

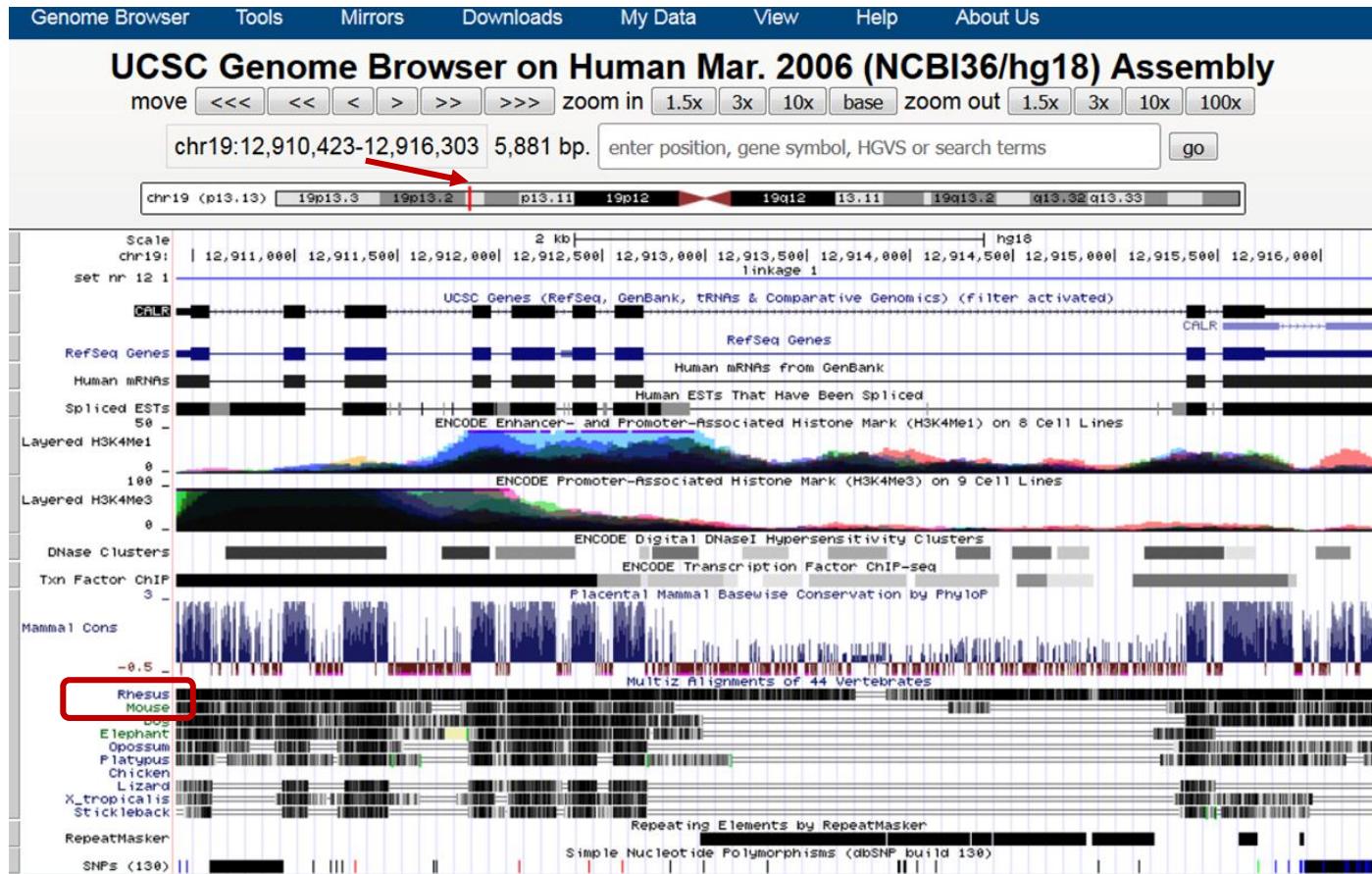
Human CCTCTTCTGGTGTCTGATCTGTCCTCATTAGTTAGAGGG---AGACCCAGACCCATTGACTTTCTTAATAATGATTTTGGAGGGAGCTAAAGAATAAGTCCCAGCAACAA  
Mouse GCCCCCAGCCAGTGTGACCTCTG-CGGTTGGACAGTGGGTGAGACTGAACCC

Human TTTATTGCAATTGATCGCAGATCTAGGTGTTGATATAGTTATTCCAATCTTAATGAGGATTGAGTTCTAGAGCACTGATTTTTCT  
Mouse .....

Human CCTTAAACTTAAGCTCCACCCACAGCCATTAGGACAGAACATCAGGTCTGAGTTCTCTCTCAGCCTTGACAGACCCAGGTTGAAGAACAGGCTTCTTATAAGAGGGGTG  
Mouse .....

Human AGAGCCTGAGATGATGGGAGTCTGACTCTTAACCTGAGGACTTCAACCTAGTCCCAGACATCTGTGGCCCTGGACCAAGAAGGTCATGTCATCTCAACTACAAGGGCAAGA  
Mouse .....

# UCSC – an example



# Ensemble – an example

Genes and regulation  
**Flanking sequence**  
Population genetics  
Phenotype data  
Sample genotypes  
Linkage disequilibrium  
Phylogenetic context  
Citations  
3D Protein model

Most severe consequence  
Alleles  
Change tolerance  
Location  
Evidence status ⓘ  
HGVS name  
Original source  
About this variant

TF binding site | See all predicted consequences  
**GIA** | Ancestral: G | Highest population MAF: < 0.01 →  
CADD: A:6.251 | GERP: -2.97  
Chromosome 19:12938008 (forward strand) | VCF: 19 12938008 rs1180072170 G A  
NC\_00019.10:g.12938008G>A  
Variants (including SNPs and indels) imported from dbSNP (release 152) | View in dbSNP ⓘ  
This variant overlaps 1 regulatory feature.

**Flanking sequence ?**

Download sequence | BLAST this sequence

**Flanking sequence**

The sequence below is from the **reference genome** flanking the variant location. The variant is shown in **red** text. Neighbouring variants are shown with highlighted letters and ambiguity codes.  
To change the display of the flanking sequence (e.g. hide the other variants, change the length of the flanking sequence), use the "Configure this page" link on the left.

Variants Focus variant Regulatory region Tf binding site

Markup loaded

rs1180072170 SNP

CTAYCAYCTSKGTCKTTTYCCTYCCCARACGGAAYMCCCCTGCCCTCTCCYCTGCCTRT  
AACCGTTTAATTSCARAAGCCAGGCTTTGTGGGAGRCCACAGACAGYGAGACCTTCA  
TTTACCRGTTGAGASGAGGGTARWGGGCYGGCKKCAATCKSKRGTAATACCCATAIMSACA  
CTCCASRAGTCBAGTCMVATCCCKRGGCYTVTCCCNFTTGTGYMAGGACRGCCTTTA  
AAAACRTTAAALSCATTTCGTCGRGTRGORTYGGGSAGGGTGCCTCCMYTCTGTCRMT  
CAGGAACRCYHRYCASTHYAGAGAGCYTAARTSACIYGCGCTCACACAGCWSCARPC  
Y-SAYAGGCTGCCARCRYISTASEGGCAGAGGGMRGAR-RGC-RCATTYTCCCCACYRG  
CCCRBERTGACCTAGCAACGGGGTGGCCTTACRCCCAWACCCCAAYSRGTGACCTST  
CTCCCATBMYAVGYWGGGTS-RGRAGCACRGTTGGGGTCTGTTCTCARMTGGRASAT  
AAGAGCCKRCTAAGAAAACCTTGCCCAAGCCCCCTCCACCTAGAGGGAAATGDGAGGGARAGA  
AKCTGAGGGYAGSGTSCYRGTCYMSKGGSACARCTGCRCTCYRCCRWWTCTTAAAG  
YCCCCAGRYRRGCAACRAYGCRCGCGGAMRGGSCGGDGTTGGGBTCNGRTCTRSTYRCA

# Ensemble – an example

Show/hide columns (1 hidden)

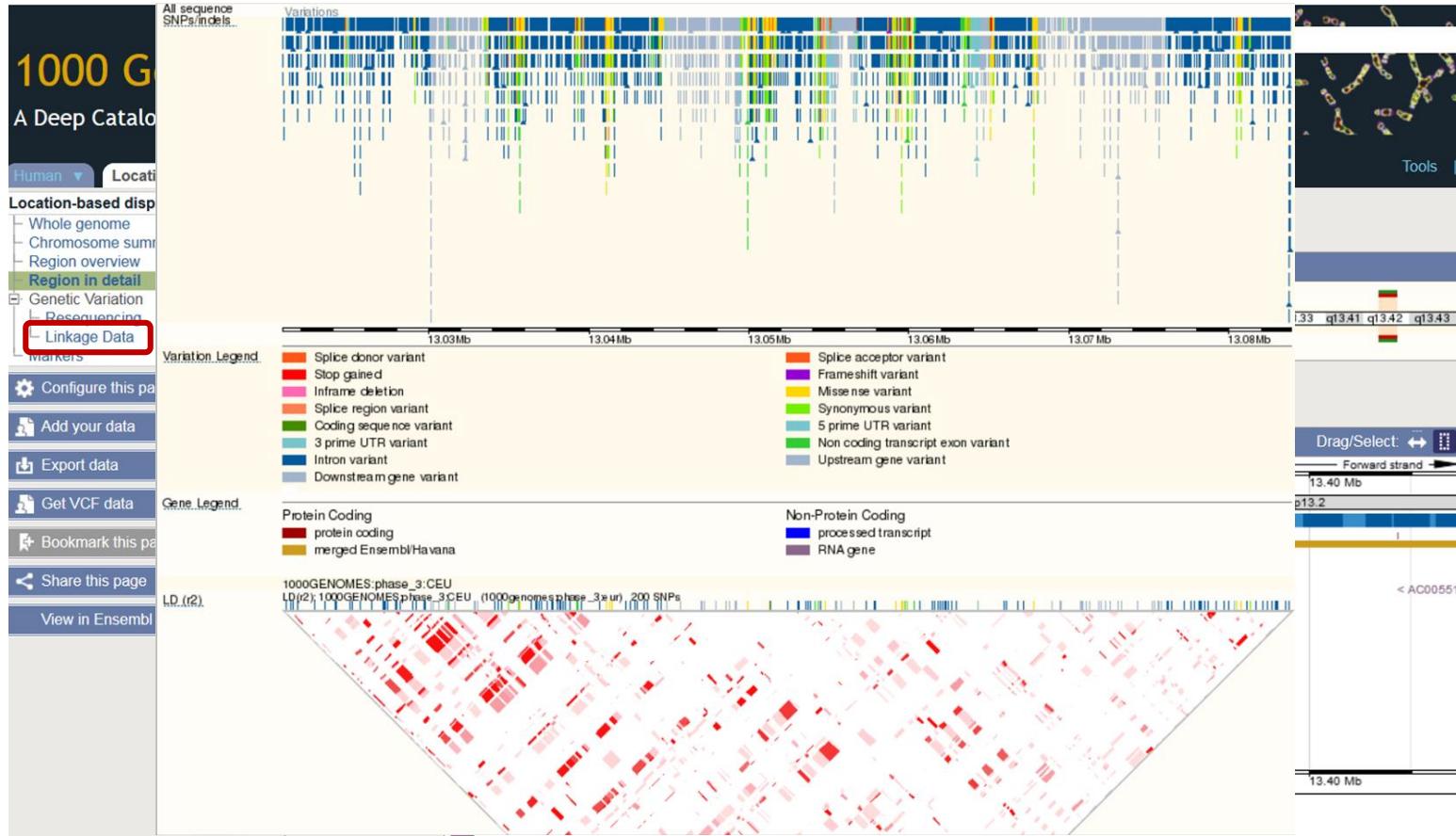
| Name     | Transcript ID     | bp   | Protein    | Biotype                 | CCDS      | UniProt       | RefSeq Match | Flags                        |
|----------|-------------------|------|------------|-------------------------|-----------|---------------|--------------|------------------------------|
| CALR-201 | ENST00000316448.9 | 1903 | 41aa       | Protein coding          | CCDS12288 | P27797 V9HW88 | -            | TSL:1 GENCODE basic APPRIS F |
| CALR-202 | ENST00000586760.1 | 913  | 163aa      | Protein coding          | -         | K7EL50        | -            | CDS 5' incomplete TSL:2      |
| CALR-206 | ENST00000588454.5 | 776  | 247aa      | Protein coding          | -         | K7EJB9        | -            | CDS 3' incomplete TSL:3      |
| CALR-204 | ENST00000586967.1 | 394  | 13aa       | Nonsense mediated decay | -         | K7ELE2        | -            | CDS 5' incomplete TSL:2      |
| CALR-205 | ENST00000587486.6 | 957  | No protein | Retained intron         | -         | -             | -            | TSL:2                        |
| CALR-207 | ENST00000590325.1 | 781  | No protein | Retained intron         | -         | -             | -            | TSL:2                        |
| CALR-203 | ENST00000586803.1 | 667  | No protein | Retained intron         | -         | -             | -            | TSL:2                        |

Splice variants ?

The figure displays a genomic track for the CALR gene region. At the top, a table lists transcripts with their lengths and protein products. The main panel shows a genomic scale from 12.938 Mb to 12.946 Mb. Multiple exons are shown as colored boxes (blue, yellow, red, orange) with arrows indicating their direction. Splicing is indicated by lines connecting exons. Below the genome track, several databases provide additional information:

- Genes (Comprehensive...):** Shows multiple gene models across the region.
- ENST00000316448.9 CALR-201:** A specific transcript model.
- PIRSF:** Annotations include PIRSF002356 Calreticulin.
- PROSITE patterns:** Annotations include PS00803 Calreticulin/calnexin, conserved site, PS00804 Calreticulin/calnexin, conserved site, and PS00805 Calreticulin/calnexin, conserved site.
- Pfam:** Annotation includes PF00262.

# 1000 Genome – an example



# Ensembl – an example

- Supporting evidence
- Gene alleles
- Sequence**
  - └ Secondary Structure
  - External references
  - Regulation
  - Phenotype
- Genetic Variation**
  - Variation table**
    - └ Structural variation
    - └ Variation image
  - External data
    - └ Gene expression
  - ID History
    - └ Gene history

**Synonyms** cC1qR, CRT, FLJ26680, RO, SSA, cC1qR, CRT, FLJ26680, RO, SSA

**Location** Chromosome 19: 13,049,392-13,055,303 forward strand.

**INSDC coordinates** chromosome:GRCh37:CM000681.1:13049392:13055303:1

**About this gene** This gene has 7 transcripts ([splice variants](#)) and is associated with [150 phenotypes](#).

**Transcripts** [Show transcript table](#)

### Variation table

**Configuring the page**

The full intronic sequence around this Gene is used. To extend or reduce the intronic sequence, use the "Configure this page - Intron Context" link on the left.

Note: From release 68, Ensembl uses Sequence Ontology (SO) terms to describe consequences. [More information about this table](#).

#### Summary of variation consequences in ENSG00000179218

[Switch to tree view](#) 

| Show                           | All  entries | Filter                   |
|--------------------------------|---|--------------------------|
| Number of variant consequences | Type  | Description              |
| 0                              | -   | Transcript ablation      |
| 28                             | <a href="#">Hide</a>  | Splice donor variant     |
| 1                              | <a href="#">Hide</a>  | Splice acceptor variant  |
| 7                              | <a href="#">Show</a>  | Stop gained              |
| 70                             | <a href="#">Show</a>  | Frameshift variant       |
| 0                              | -   | Stop lost                |
| 0                              | -   | Start lost               |
| 0                              | -   | Transcript amplification |
| 0                              | -   | Inframe insertion        |
| 5                              | <a href="#">Show</a>  | Inframe deletion         |
| 0                              | -   | protein altering variant |
| 164                            | <a href="#">Show</a>  | Missense variant         |

55

23-Oct-20

# HaploReg – an example

Query SNP: **rs1052429** and variants with  $r^2 \geq 0.8$

| chr | pos<br>(hg38) | LD<br>(r <sup>2</sup> ) | LD<br>(D') | variant                    | Ref | Alt | AFR freq | AMR freq | ASN freq | EUR freq | SiPhy cons | Promoter histone marks | Enhancer histone marks | DNase           | Proteins bound   | Motifs changed         | NHGRI/EBI GWAS hits | GRASP QTL hits | Selected eQTL hits | GENCODE genes | dbSNP func annot |
|-----|---------------|-------------------------|------------|----------------------------|-----|-----|----------|----------|----------|----------|------------|------------------------|------------------------|-----------------|------------------|------------------------|---------------------|----------------|--------------------|---------------|------------------|
| 16  | 69924592      | 0.85                    | 0.96       | <a href="#">rs904808</a>   | T   | C   | 0.93     | 0.78     | 0.97     | 0.75     |            | 13 tissues             | 4 tissues              |                 |                  | 23 altered motifs      |                     | 72 hits        | WWP2               | intronic      |                  |
| 16  | 69926085      | 0.9                     | 0.97       | <a href="#">rs904807</a>   | G   | C   | 0.92     | 0.79     | 0.97     | 0.74     |            | 6 tissues              | 13 tissues             | IPSC            |                  | MZF1::1-4              | 1 hit               | 75 hits        | WWP2               | intronic      |                  |
| 16  | 69926361      | 0.9                     | 0.97       | <a href="#">rs904806</a>   | A   | G   | 0.92     | 0.79     | 0.97     | 0.74     |            | 5 tissues              | 10 tissues             | BRN, BRN        |                  | 5 altered motifs       |                     | 72 hits        | WWP2               | intronic      |                  |
| 16  | 69926589      | 0.9                     | 0.97       | <a href="#">rs904805</a>   | G   | A   | 0.92     | 0.79     | 0.97     | 0.74     |            | 6 tissues              | 10 tissues             |                 |                  | CTCF, En-1             |                     | 72 hits        | WWP2               | intronic      |                  |
| 16  | 69926702      | 0.9                     | 0.97       | <a href="#">rs904804</a>   | T   | C   | 0.94     | 0.79     | 0.97     | 0.74     |            | 9 tissues              | 10 tissues             |                 |                  |                        |                     | 72 hits        | WWP2               | intronic      |                  |
| 16  | 69927089      | 0.9                     | 0.97       | <a href="#">rs1983015</a>  | A   | G   | 0.94     | 0.79     | 0.97     | 0.74     |            | 9 tissues              | 10 tissues             | MUS, BRN, LNG   |                  | AFP1, RhoX11           |                     | 72 hits        | WWP2               | intronic      |                  |
| 16  | 69928578      | 0.91                    | 0.97       | <a href="#">rs11864678</a> | C   | T   | 0.94     | 0.79     | 0.97     | 0.74     |            | VAS                    | 10 tissues             | IPSC, ADRL      |                  | Pou1f1, Pou2f2, Pou5f1 |                     | 69 hits        | WWP2               | intronic      |                  |
| 16  | 69929452      | 0.91                    | 0.97       | <a href="#">rs1566452</a>  | G   | A   | 0.92     | 0.79     | 0.97     | 0.74     |            |                        | 8 tissues              |                 | Evi-1            |                        | 2 hits              | 71 hits        | WWP2               | synonymous    |                  |
| 16  | 69931121      | 0.91                    | 0.97       | <a href="#">rs2270840</a>  | A   | T   | 0.94     | 0.80     | 0.97     | 0.74     |            |                        | BRN                    |                 | Hsf              |                        |                     | 69 hits        | WWP2               | intronic      |                  |
| 16  | 69931378      | 0.91                    | 0.97       | <a href="#">rs8052727</a>  | T   | G   | 0.92     | 0.79     | 0.97     | 0.74     |            |                        | BRN                    |                 | NERF1a           |                        |                     | 69 hits        | WWP2               | intronic      |                  |
| 16  | 69931689      | 0.91                    | 0.97       | <a href="#">rs2291961</a>  | A   | G   | 0.94     | 0.79     | 0.97     | 0.74     |            |                        | BRN                    |                 | 6 altered motifs |                        |                     | 71 hits        | WWP2               | intronic      |                  |
| 16  | 69932072      | 0.92                    | 0.98       | <a href="#">rs4985381</a>  | A   | G   | 0.94     | 0.79     | 0.97     | 0.74     |            |                        | STRM, BRN              |                 | Crx, Pitx2       |                        |                     | 70 hits        | WWP2               | intronic      |                  |
| 16  | 69932247      | 0.83                    | 0.97       | <a href="#">rs10048129</a> | T   | C   | 0.86     | 0.78     | 0.93     | 0.71     |            |                        |                        |                 |                  | 8 altered motifs       |                     | 67 hits        | WWP2               | intronic      |                  |
| 16  | 69932414      | 0.92                    | 0.98       | <a href="#">rs10048088</a> | A   | G   | 0.94     | 0.79     | 0.97     | 0.74     |            |                        |                        |                 |                  | BDP1                   |                     | 70 hits        | WWP2               | intronic      |                  |
| 16  | 69932459      | 0.92                    | 0.98       | <a href="#">rs10048090</a> | G   | A   | 0.92     | 0.79     | 0.97     | 0.74     |            |                        |                        |                 |                  | ATF3                   |                     | 70 hits        | WWP2               | intronic      |                  |
| 16  | 69933190      | 0.92                    | 0.98       | <a href="#">rs2102066</a>  | G   | A   | 0.89     | 0.79     | 0.97     | 0.74     |            |                        | BRN                    | BLD, BLD        |                  |                        | CHOP::CEBPalpha     |                | 64 hits            | WWP2          | intronic         |
| 16  | 69933413      | 0.92                    | 0.98       | <a href="#">rs7206222</a>  | G   | A   | 0.92     | 0.79     | 0.97     | 0.74     |            |                        | STRM, BRN              |                 |                  |                        | 5 altered motifs    |                | 70 hits            | WWP2          | intronic         |
| 16  | 69933829      | 0.9                     | 0.95       | <a href="#">rs8047818</a>  | G   | A   | 0.76     | 0.78     | 0.97     | 0.73     |            |                        | STRM, BRN              | ESDR, IPSC, MUS |                  |                        | CHD2, CTCF          |                | 56 hits            | WWP2          | intronic         |
| 16  | 69933994      | 0.93                    | 0.98       | <a href="#">rs2270841</a>  | T   | C   | 0.92     | 0.79     | 0.97     | 0.74     |            |                        | STRM                   |                 |                  |                        | 5 altered motifs    |                | 73 hits            | WWP2          | synonymous       |
| 16  | 69934298      | 0.91                    | 0.98       | <a href="#">rs8048678</a>  | G   | A   | 0.88     | 0.78     | 0.97     | 0.74     |            |                        | MUS                    |                 |                  |                        | 2 hits              | 71 hits        | WWP2               | intronic      |                  |
| 16  | 69934464      | 0.91                    | 0.98       | <a href="#">rs8048590</a>  | C   | T   | 0.88     | 0.79     | 0.97     | 0.74     |            |                        | MUS                    |                 |                  |                        | 2 hits              | 65 hits        | WWP2               | intronic      |                  |
| 16  | 69934493      | 0.91                    | 0.98       | <a href="#">rs8049004</a>  | G   | T   | 0.93     | 0.79     | 0.97     | 0.74     |            |                        | MUS                    | LNG             |                  |                        | ERalpha-a           |                | 68 hits            | WWP2          | intronic         |
| 16  | 69934708      | 0.88                    | 0.95       | <a href="#">rs12932286</a> | G   | C   | 0.90     | 0.77     | 0.95     | 0.73     |            |                        | MUS                    |                 |                  |                        | ERalpha-a, HNF4     |                | 64 hits            | WWP2          | intronic         |
| 16  | 69934824      | 0.91                    | 0.97       | <a href="#">rs12932078</a> | A   | G   | 0.89     | 0.85     | 0.97     | 0.74     |            |                        | MUS                    |                 |                  |                        | DMRT7, Pou2f2       |                | 68 hits            | WWP2          | intronic         |
| 16  | 69936097      | 0.97                    | 0.99       | <a href="#">rs2270842</a>  | T   | G   | 0.94     | 0.80     | 0.96     | 0.73     |            |                        | STRM, VAS, MUS         |                 |                  |                        | NF-AT1              | 4 hits         | 81 hits            | WWP2          | intronic         |
| 16  | 69936426      | 0.97                    | 0.99       | <a href="#">rs1983016</a>  | A   | G   | 0.89     | 0.79     | 0.96     | 0.73     |            |                        | STRM, MUS              |                 |                  |                        | 4 altered motifs    | 6 hits         | 74 hits            | WWP2          | synonymous       |
| 16  | 69936626      | 0.97                    | 0.99       | <a href="#">rs904803</a>   | A   | G   | 0.89     | 0.79     | 0.96     | 0.73     |            |                        | STRM                   |                 |                  |                        | USF1                | 3 hits         | 74 hits            | WWP2          | intronic         |
| 16  | 69938062      | 0.97                    | 0.99       | <a href="#">rs3762177</a>  | A   | G   | 0.93     | 0.80     | 0.96     | 0.73     |            |                        | BRN                    |                 |                  |                        | RhoX11, p300        | 4 hits         | 82 hits            | WWP2          | intronic         |
| 16  | 69938583      | 0.83                    | 0.99       | <a href="#">rs9302605</a>  | T   | A   | 0.95     | 0.81     | 0.97     | 0.76     |            |                        |                        |                 |                  | HDAC2, LXR, Zbtb3      |                     | 63 hits        | WWP2               | intronic      |                  |
| 16  | 69939719      | 0.96                    | 0.99       | <a href="#">rs10712484</a> | GC  | G   | 0.89     | 0.79     | 0.97     | 0.74     |            |                        |                        | ESDR            |                  | 6 altered motifs       |                     | 75 hits        | WWP2               | intronic      |                  |
| 16  | 69939752      | 0.99                    | 1          | <a href="#">rs4985461</a>  | C   | G   | 0.53     | 0.78     | 0.97     | 0.73     |            |                        |                        |                 |                  | 5 altered motifs       |                     | 62 hits        | WWP2               | intronic      |                  |
| 16  | 69940545      | 1                       | 1          | <a href="#">rs3748388</a>  | C   | A   | 0.94     | 0.79     | 0.97     | 0.73     |            |                        | LIV                    | SKIN            |                  |                        | 8 altered motifs    | 1 hit          | 82 hits            | WWP2          | 3'-UTR           |
| 16  | 69941457      | 1                       | 1          | <a href="#">rs1052429</a>  | G   | A   | 0.94     | 0.79     | 0.97     | 0.73     |            |                        |                        |                 |                  | HEY1, LXR              | 5 hits              | 81 hits        | WWP2               | 3'-UTR        |                  |
| 16  | 69942186      | 0.99                    | 1          | <a href="#">rs7184994</a>  | T   | C   | 0.90     | 0.80     | 0.97     | 0.73     |            | FAT                    | 9 tissues              | THYM            |                  |                        |                     | 71 hits        | 444bp 3' of WWP2   |               |                  |

# Explore

- **Gene:** **WWP2**
- **SNP:** **rs1052429**

<https://pt.coursera.org/lecture/disease-genes/identifying-causative-genes-integration-of-the-genomic-data-6jOLF>

<https://www.labroots.com/trending/cell-and-molecular-biology/5499/impact-neanderthal-dna-human-gene-expression>





