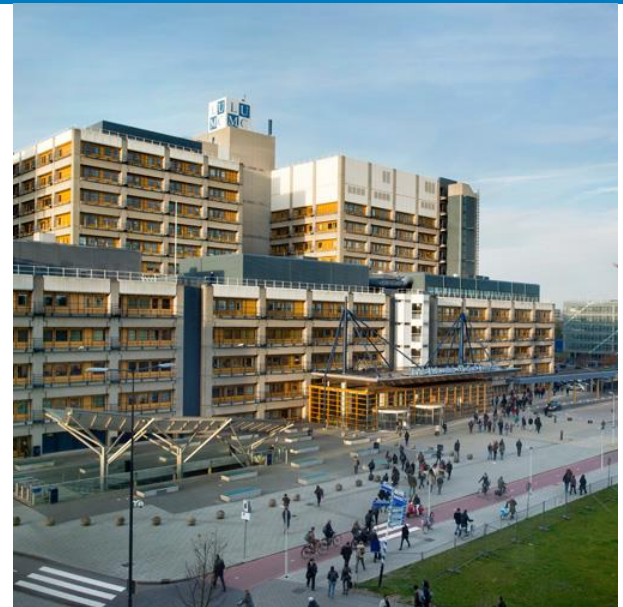




Leids Universitair  
Medisch Centrum

# 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**

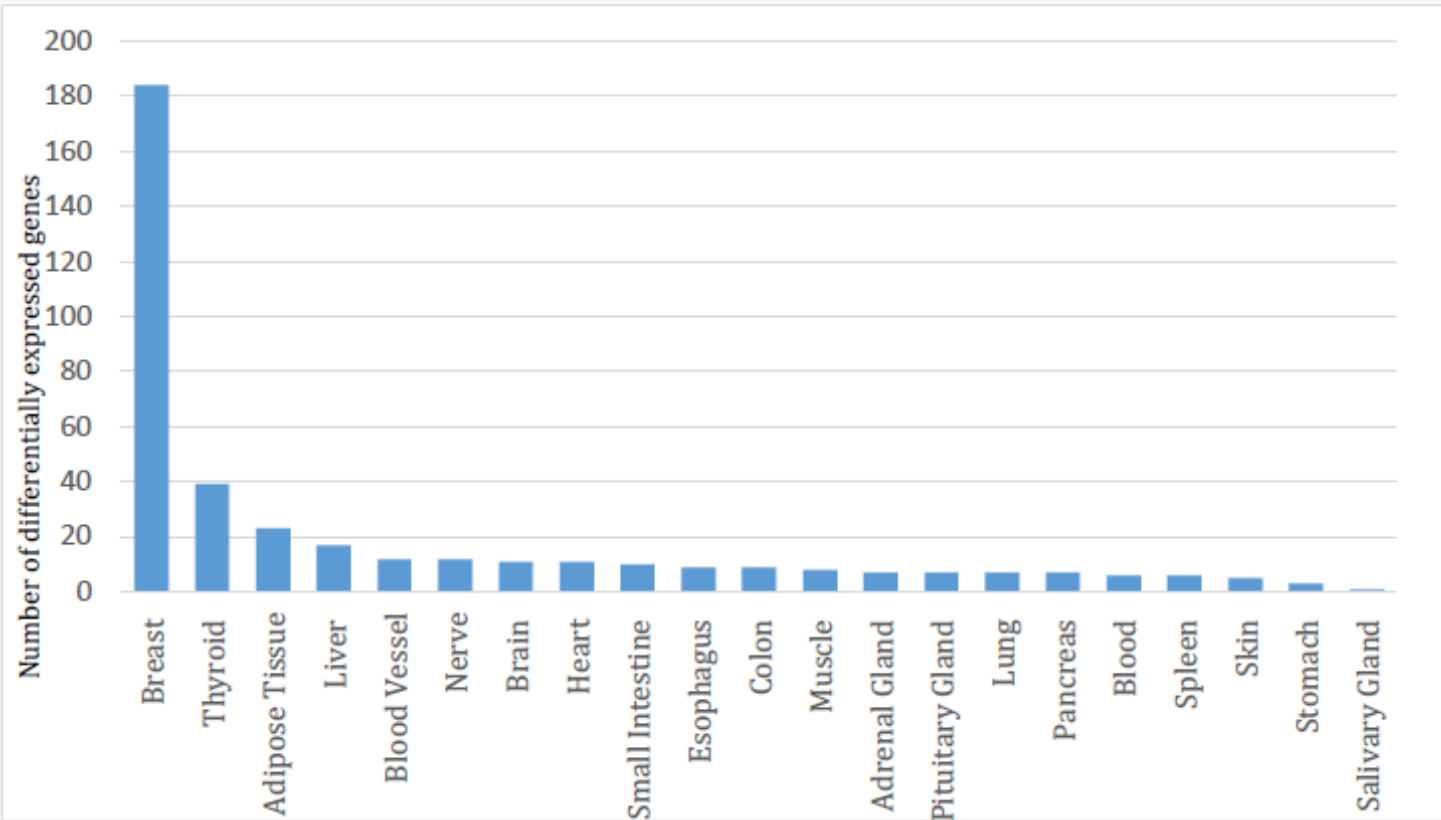
# AIMS of this lecture

- 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

# 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

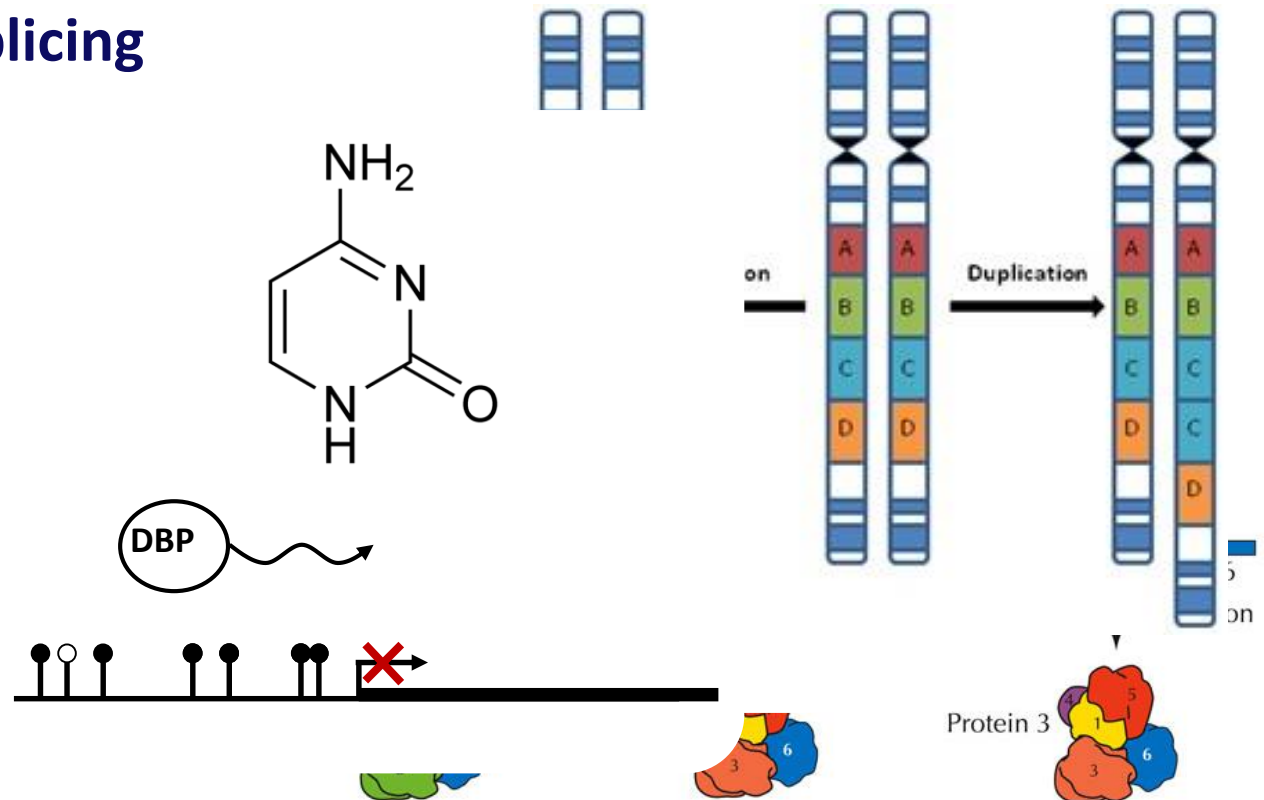
TGATGCCTTTGTTATCTACTCAAG**A**CAGGATGAGGACTGGGTAAGGAATG (**ref**)  
TGATGCCTTTGTTATCTACTCAAG**C**CAGGATGAGGACTGGGTAAGGAATG (**alt**)

# AIMS of this lecture

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

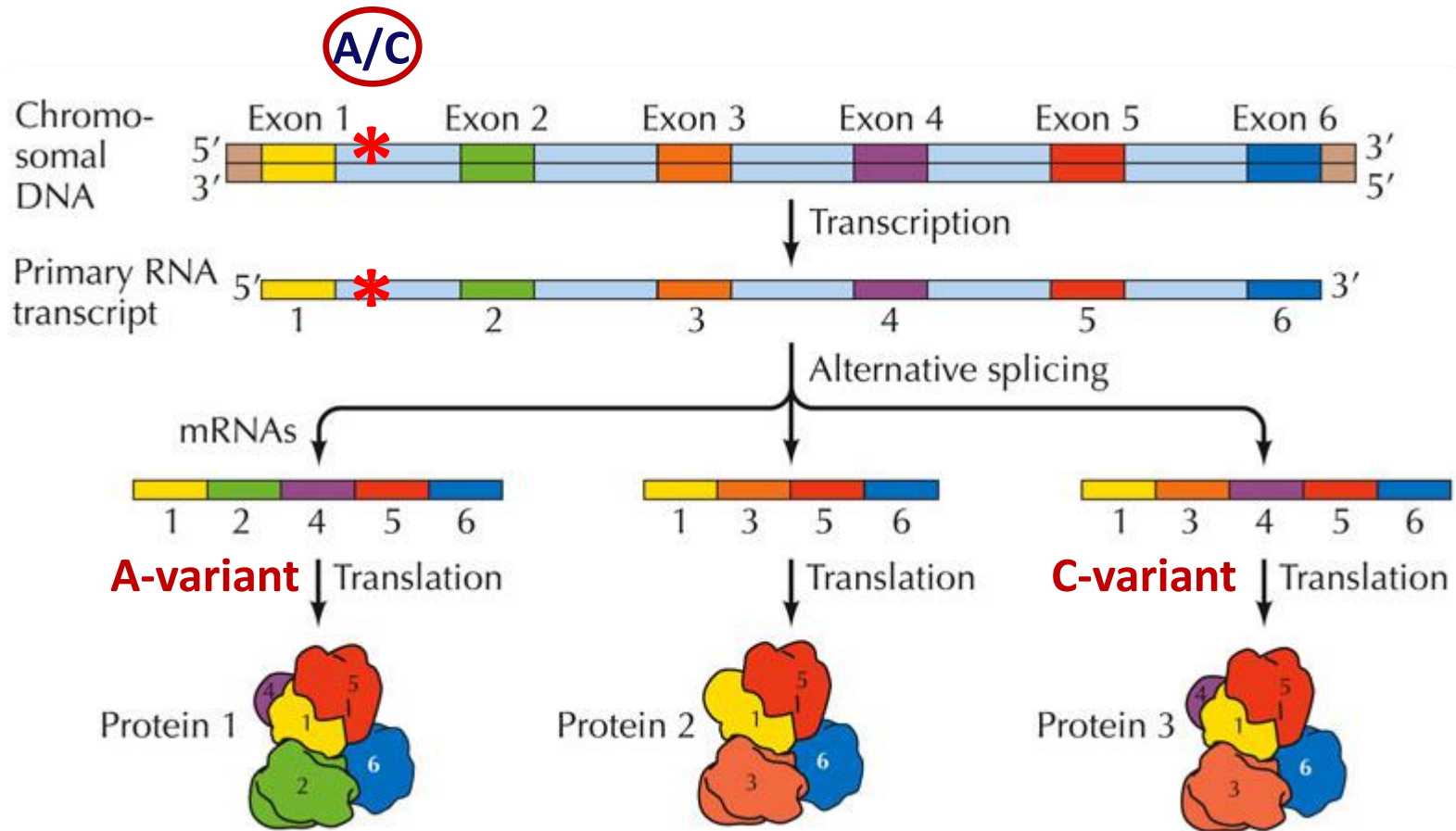
# Functional variants

- Methylation
- Copy number variation (CNV)
- Alternative splicing
- Single Nucleotide  
  - Intronic
  - Exonic
  - Promotor
  - Regulator





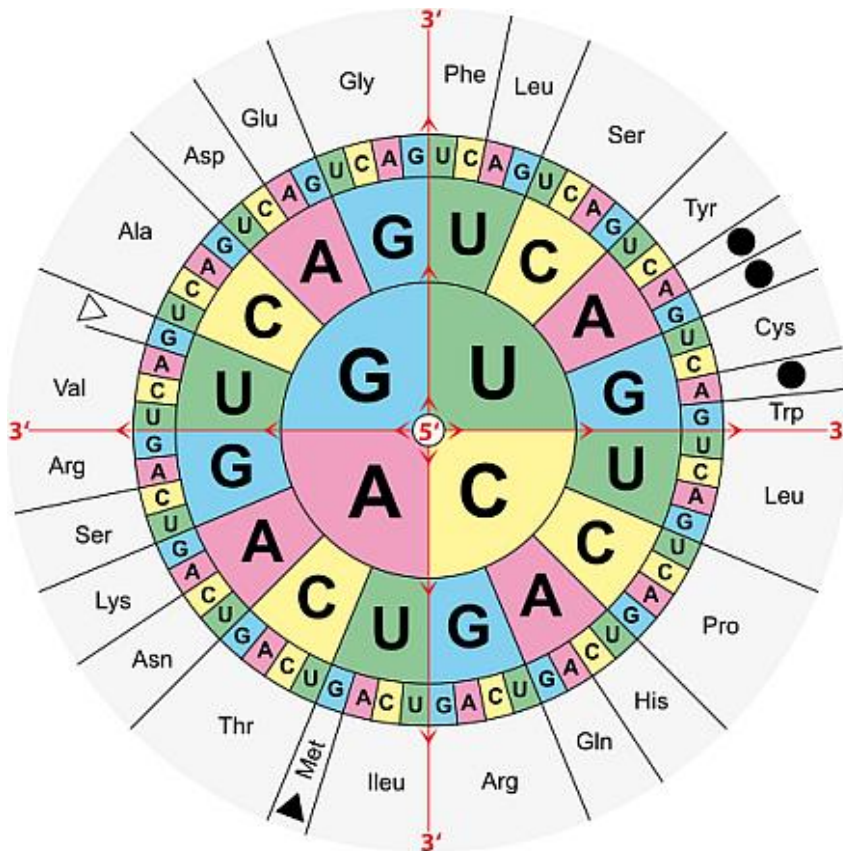
# Intronic SNPs



# 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**

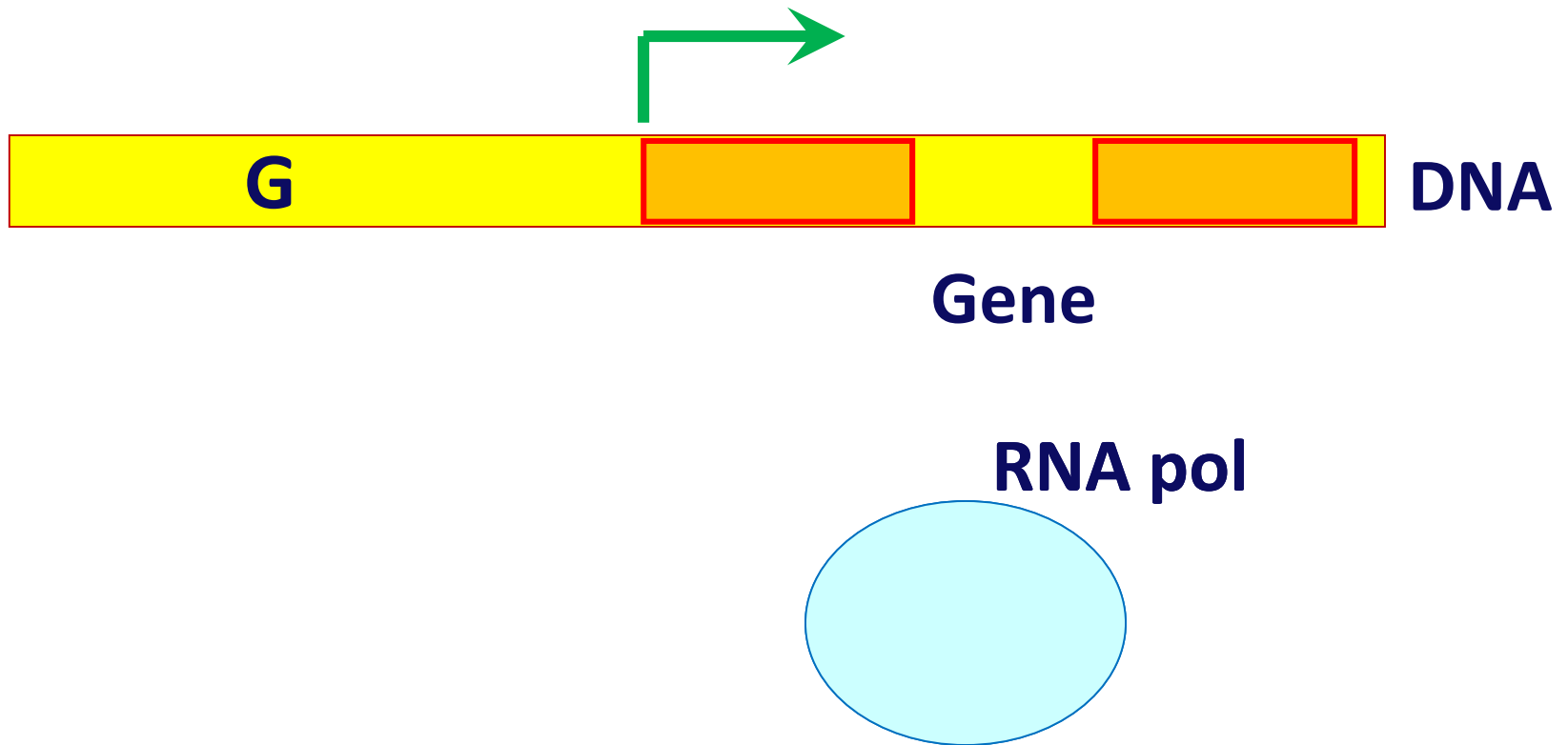
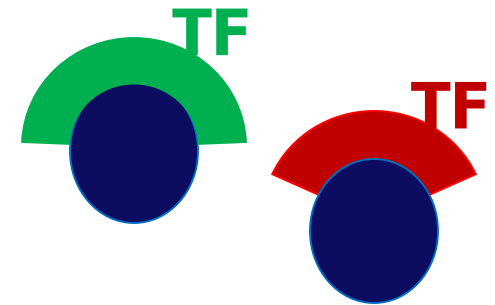
## Disruption of protein recognition sites



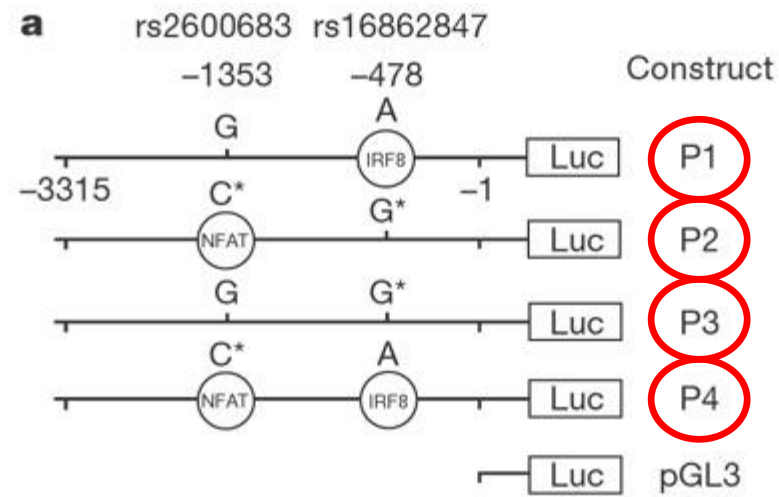
# Promotor SNPs

 *SNP G recognition*

 *SNP A recognition*



# Promotor SNPs



# Distal regulatory elements

A.

DNA



B.



a)

i



ii

1

2

3

4

5

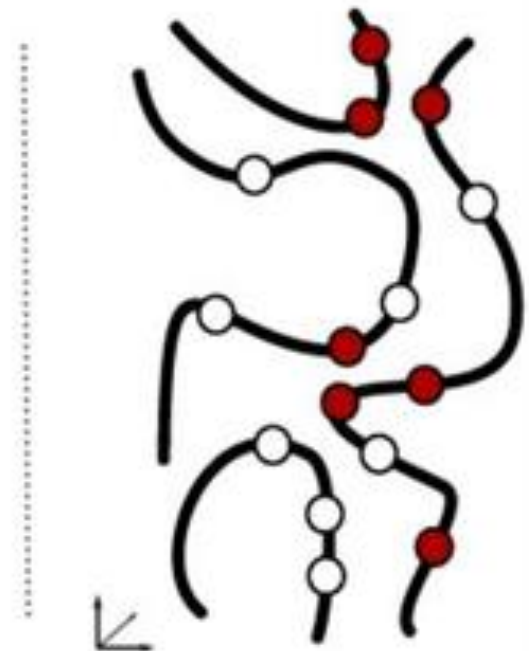
6

23

...

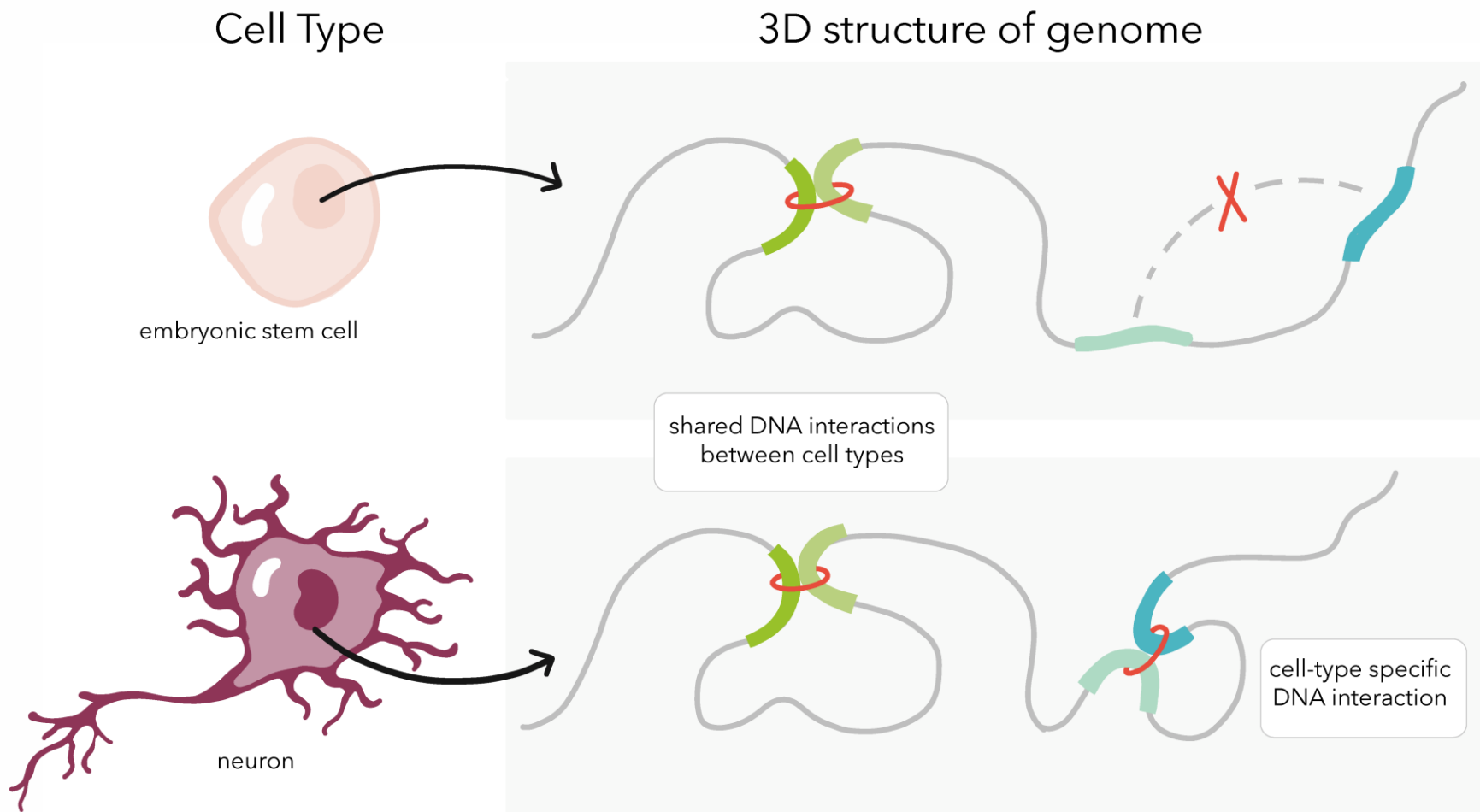


iii



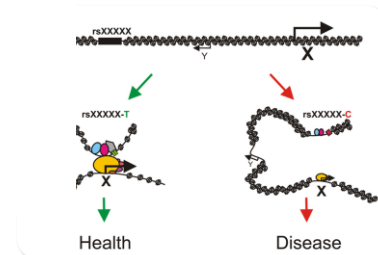
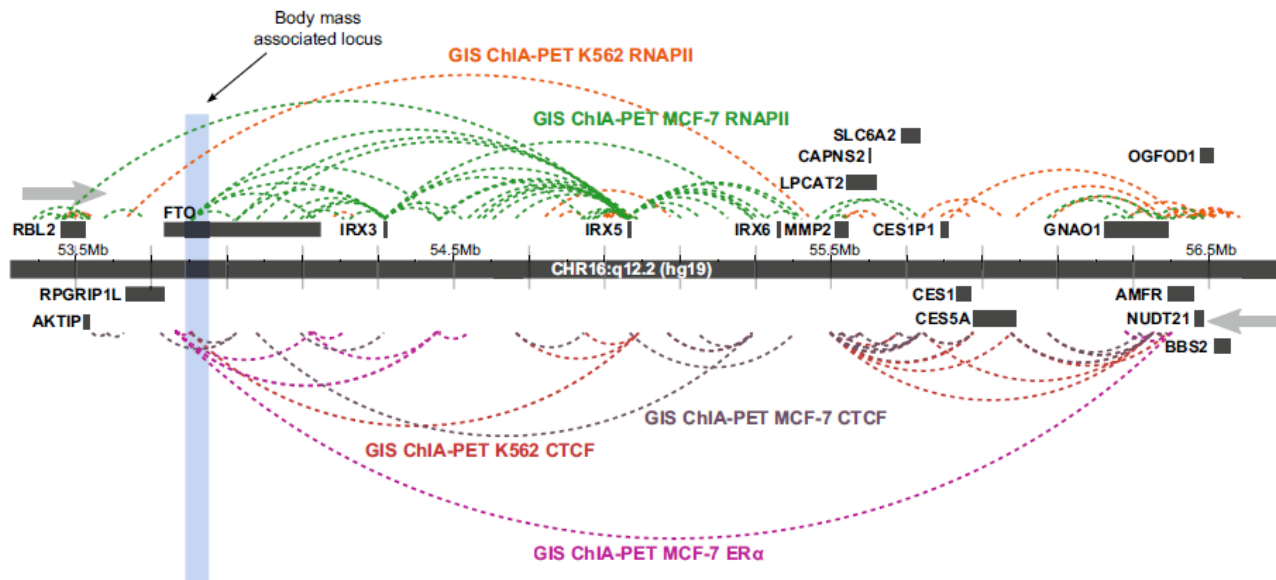
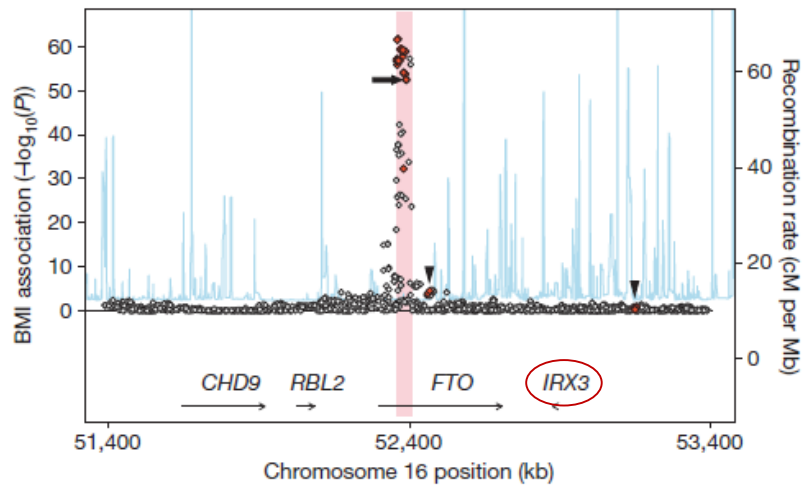
Next-generation  
Sequencing

# Distal regulatory elements





# Distal regulatory elements



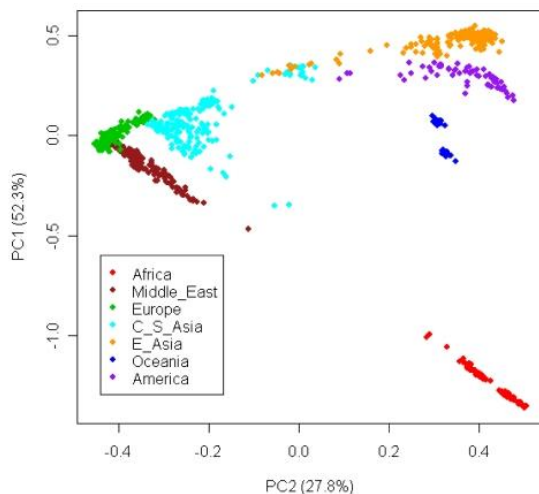


## Functional SNPs

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

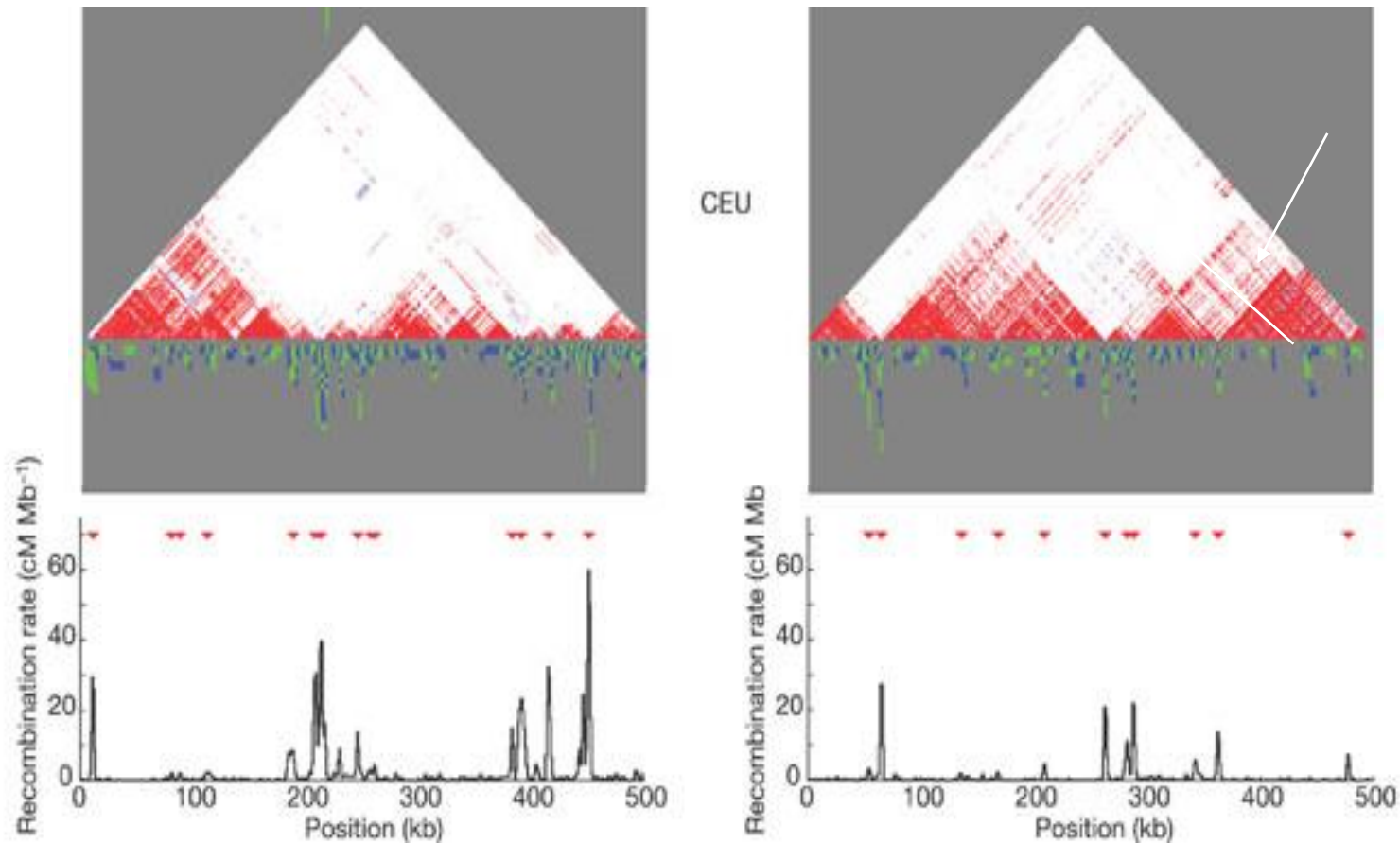
# Criteria relevant SNPs

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

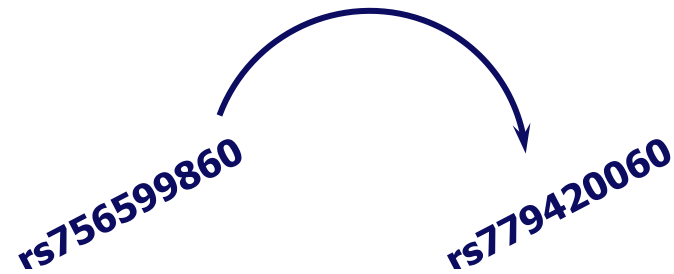


# Tagging SNP

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



# Tagging SNP



TGATGCCTTTGTTATCTACTCAAG**A**CAGGATGAGGACT**T**GGGGTAAGGAATG (**ref**)  
TGATGCCTTTGTTATCTACTCAAG**C**CAGGATGAGGAC**C**GGGGTAAGGAATG (**alt**)

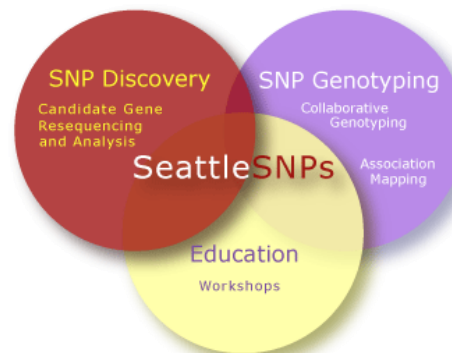
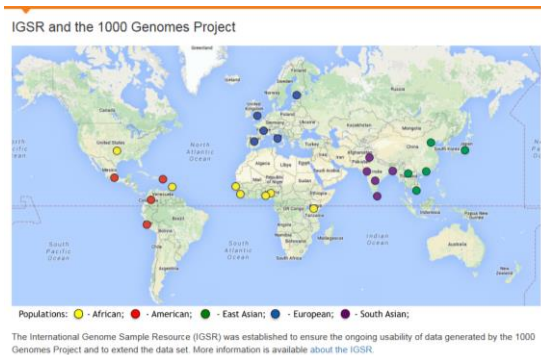
# AIMS of this lecture

- 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




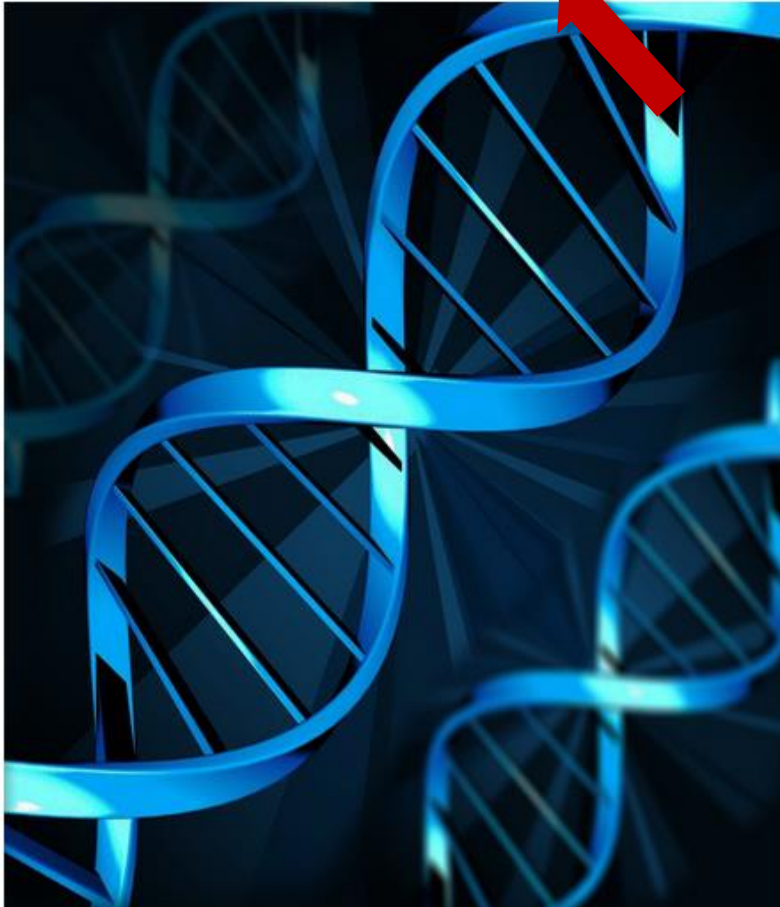
UNIVERSITY OF CALIFORNIA  
**SANTA CRUZ**



**UCSC**

## Genome Browser

 Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us



### 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...



## 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

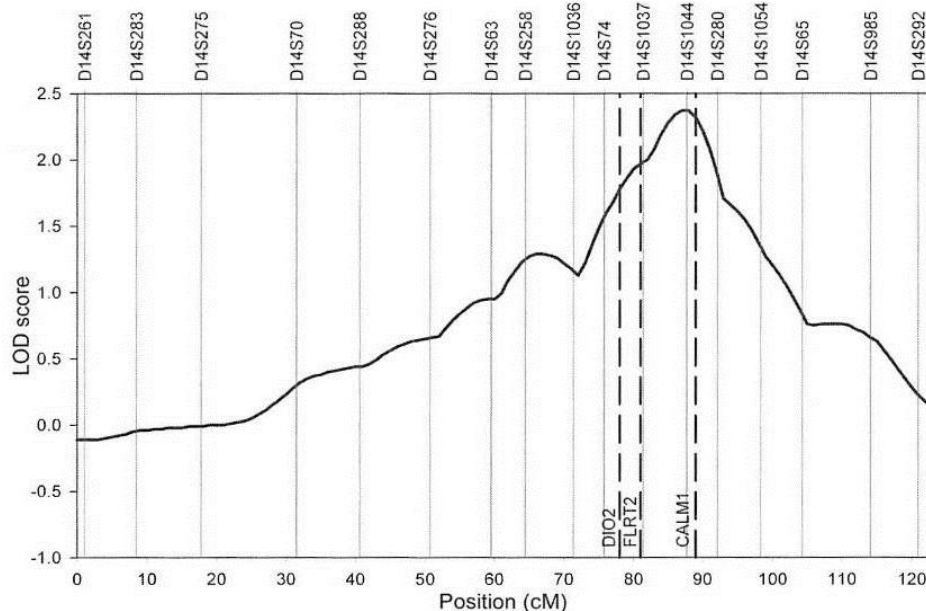
Dr. Brody investigates the genetics of breast cancer and neural tube defects. As chief of the NHGRI Genome Technology Branch's Molecular Pathogenesis section, he is interested in studying genetic mutations that lead to perturbations in normal metabolic pathways and cause disorders such as cancer and 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  $\text{LOD}=3$  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

# Surplus of information

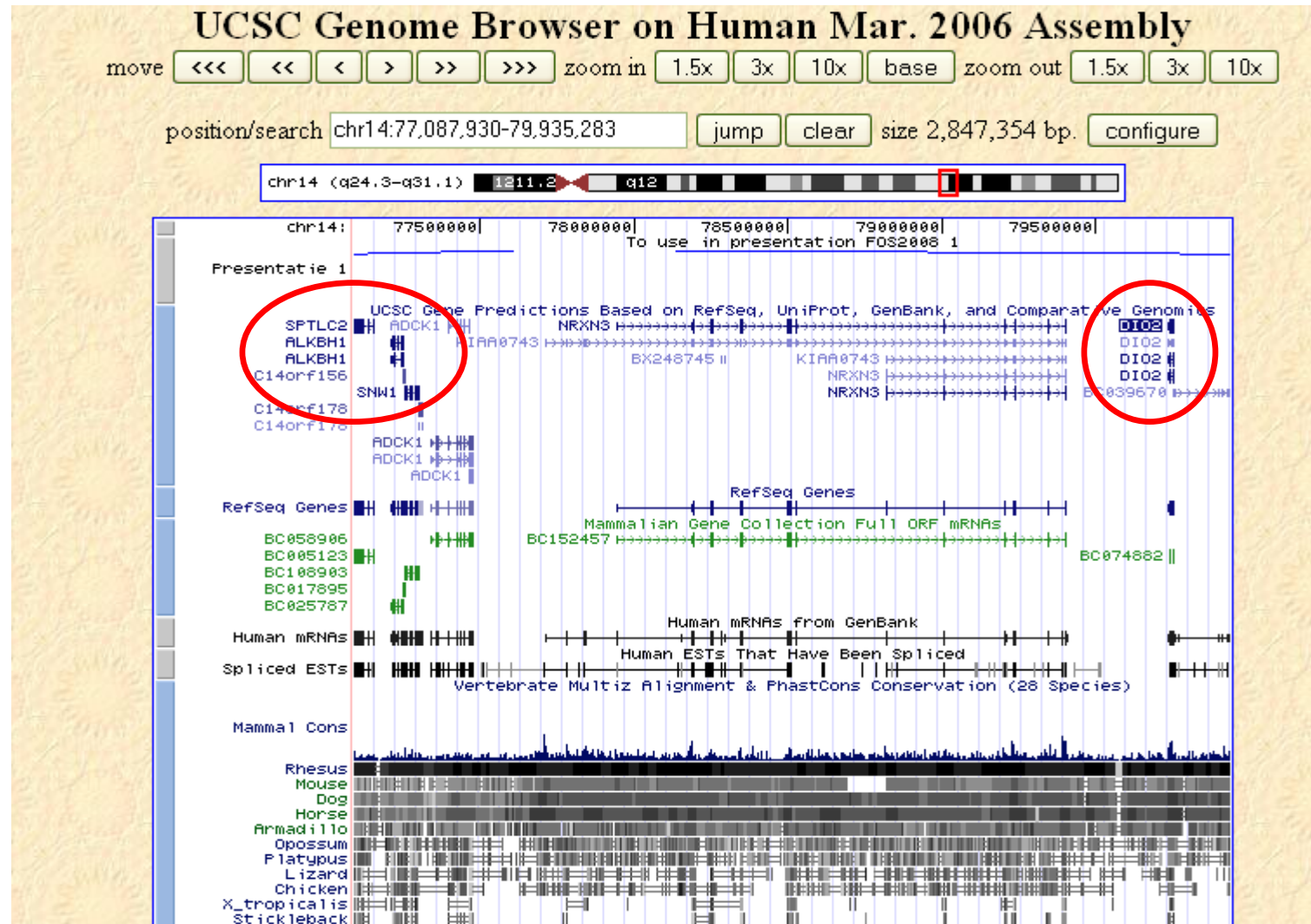
## Filter information by:

- 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  assembly  search

sort by   filter (now on) display  output

#	Name	VisiGene	fetal brain	whole brain	amygdala	thymus	bone marrow	PB-CD4+ Tcells	skin	adipocyte	pancreatic islets	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					n/a	chr14 76,920,427	hypothetical protein LOC161394
7	C14orf148	172443	n/a					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

- |                           |                                 |
|---------------------------|---------------------------------|
| • Gene function           | <b>NCBI / OMIM</b>              |
| • Literature              | <b>PubMed</b>                   |
| • Expression              |                                 |
| • eQTLs                   | <b>Genevar / GTEx</b>           |
| • Tissue of interest      | <b>BioGPS / GTEx</b>            |
| • Alternative splicing    | <b>Ensembl / GTEx</b>           |
| • Known genetic variation | <b>UCSC / HaploReg / SNPper</b> |

The screenshot shows the NCBI Entrez Gene interface. At the top, there's a navigation bar with 'NCBI', 'Resources', 'How To', and user links. Below is a search bar with 'Gene' selected. The main content area displays the gene 'DIO2 deiodinase, iodothyronine, type II [Homo sapiens (human)]' with its Gene ID (1734) and update date (7-Jun-2015). A 'Summary' tab is active, showing details like Official Symbol, Full Name, Primary source, and a detailed description of the protein's function in thyroid hormone metabolism. A right-hand sidebar contains a 'Table of contents' with links to various sections like Summary, Genomic context, and Bibliography, and a 'Related information' section with links to cDNA clones and assays.

**DIO2 deiodinase, iodothyronine, type II [Homo sapiens (human)]**  
Gene ID: 1734, updated on 7-Jun-2015

**Summary**

**Official Symbol** DIO2 provided by HGNC  
**Official Full Name** deiodinase, iodothyronine, type II provided by HGNC  
**Primary source** HGNC:HGNC:2884  
**See related** Ensembl: ENSG00000211448; HPRD:09027; MIM:601413; Vega:OTTHUMG00000171443  
**Gene type** protein coding  
**RefSeq status** REVIEWED  
**Organism** Homo sapiens  
**Lineage** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo  
**Also known as** D2; 5DII; SeLY; DIOII; TXDI2  
**Summary** The protein encoded by this gene belongs to the iodothyronine deiodinase family. It activates thyroid hormone by converting the prohormone thyroxine (T4) by outer ring deiodination (ORD) to bioactive 3,3',5-triiodothyronine (T3). It is highly expressed in the thyroid, and may contribute significantly to the relative increase in thyroidal T3 production in patients with Graves disease and thyroid adenomas. This protein contains selenocysteine (Sec) residues encoded by the UGA codon, which normally signals translation termination. The 3' UTR of Sec-containing genes have a common stem-loop structure, the sec insertion sequence (SECIS), which is necessary for the recognition of UGA as a Sec codon rather than as a stop signal. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, Jul 2008]  
**Orthologs** mouse all

**Table of contents**

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- Variation
- Pathways from BioSystems
- Interactions
- General gene information
  - Markers, Homology, Gene Ontology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences
- Additional links

**Related information**

- Order cDNA clone
- BioAssay
- BioAssay by Target (List)
- BioAssay by Target (Summary)
- BioAssay, by Gene target

## GeneRIFs!

Gene Reference into Function



# Literature - PubMed

NCBI Resources ▾ How To ▾ wdenhollander1 My NCBI Sign Out

PubMed.gov  
US National Library of Medicine  
National Institutes of Health

PubMed  Search

Create RSS Create alert Advanced 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 1 of 14 Next > Last >>

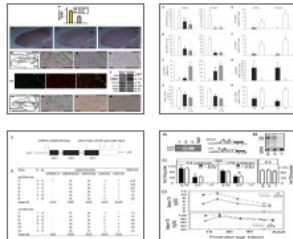
- ☐ [Disruption of Type 2 Iodothyronine Deiodinase Activity in Cultured Human Glial Cells by Polybrominated Diphenyl Ethers.](#)  
Roberts SC, Bianco AC, Stapleton HM.  
Chem Res Toxicol. 2015 Jun 2. [Epub ahead of print]  
PMID: 26004626  
[Similar articles](#)
- ☐ [An improved non-radioactive screening 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.](#)  
Borborea M, Helfer G, Ebling FJ, Barrett P.  
Mol Cell Neurosci. 2015 Jun 1;59:1-10. doi: 10.1016/j.mcn.2015.05.005. Epub 2015 Jun 1.

**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**



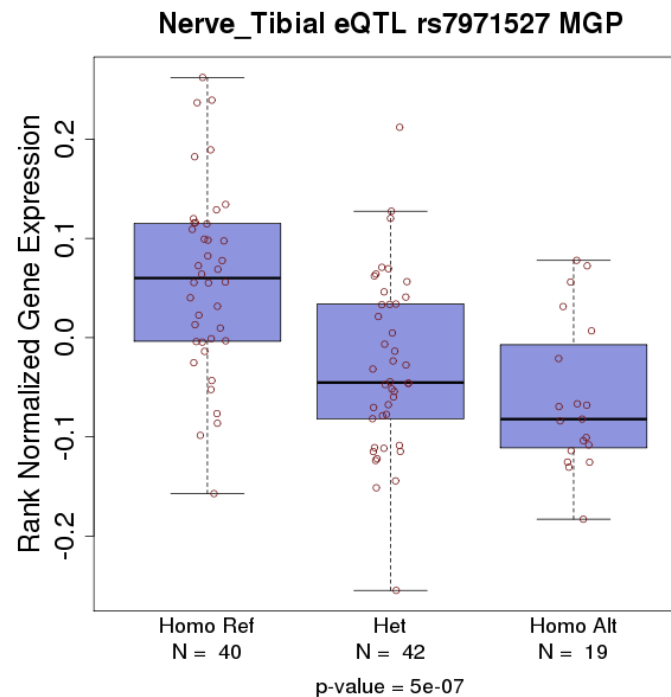
See more (245)...

**Titles with your search terms**

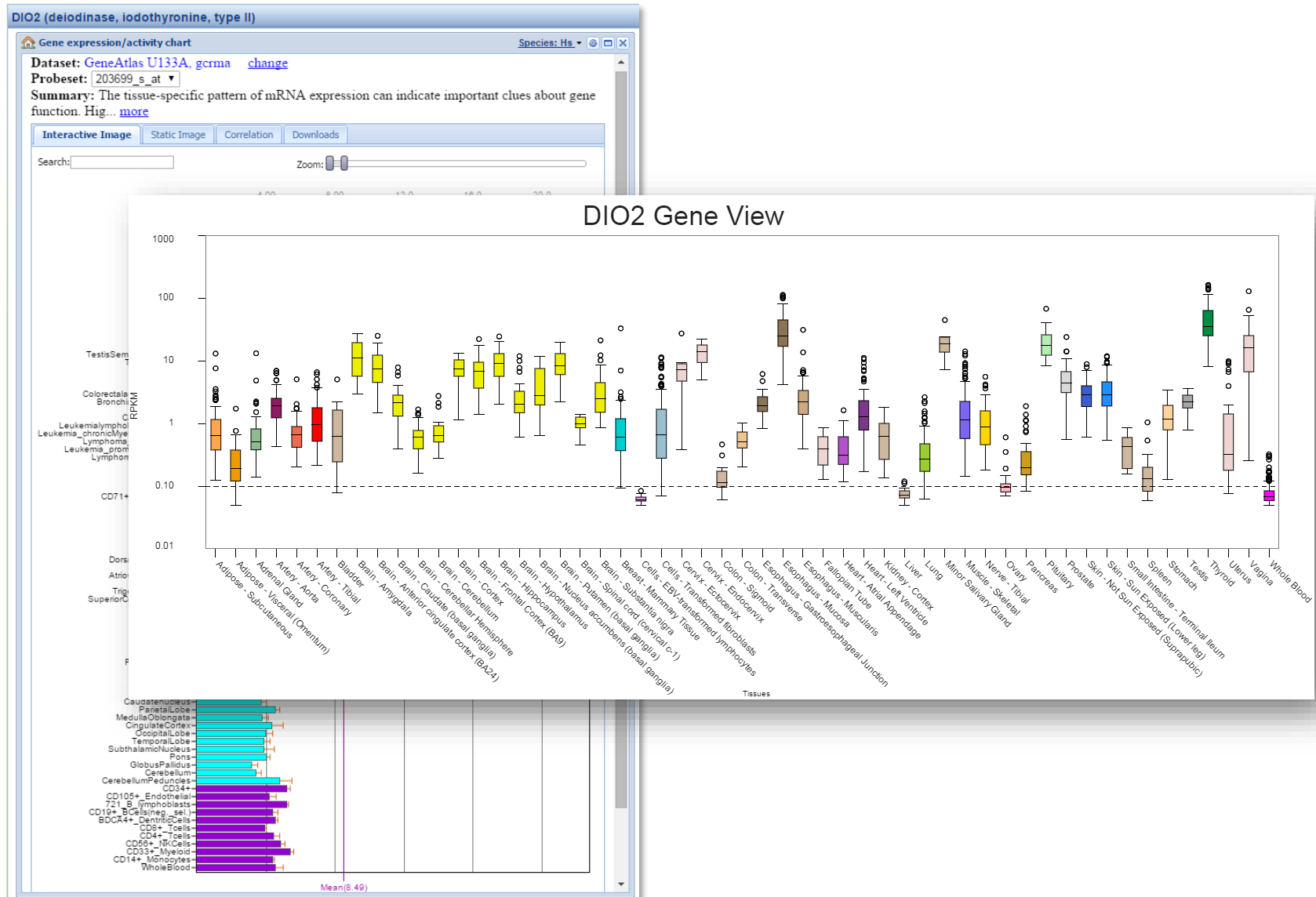
Common variation in the **DIO2** gene predicts baseline psycho [J Clin Endocrinol Metab. 2009]  
Diet-induced obesity mediated by the JNK/**DIO2** signal transduction pathway. [Genes Dev. 2013]  
Mice with targeted disruption of the **Dio2** gene

## Expression Quantitative Trait Locus

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



# Expression – Tissue of Interest (BioGPS & GTEx)

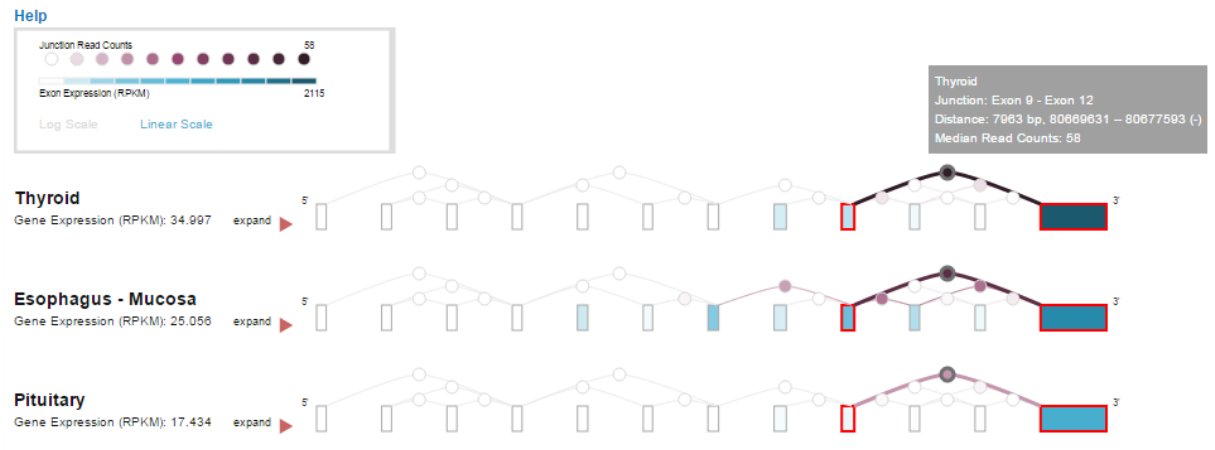


# Expression – Alternative splicing

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

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

## Exon expression for DIO2



**GTEx** shows tissue specific splice information

# Known genetic variation - HaploREG

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

Query SNP: **rs225014** and variants with  $r^2 \geq 0.8$

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	<a href="#">rs4903903</a>	T	C	0.22	0.41	0.35	0.32			<b>FAT</b>	MUS,PANC,MUS			STAT		30kb 3' of DIO2	
chr14:80635104	chr14:80168761	0.84	0.96	<a href="#">rs12588985</a>	A	G	0.22	0.41	0.35	0.32							STAT		29kb 3' of DIO2	
chr14:80638440	chr14:80172097	0.84	0.96	<a href="#">rs1491504</a>	G	T	0.22	0.41	0.35	0.32							10 altered motifs		25kb 3' of DIO2	
chr14:80641302	chr14:80174959	0.84	0.96	<a href="#">rs4899763</a>	C	T	0.22	0.41	0.35	0.32							SRF		23kb 3' of DIO2	
chr14:80641672	chr14:80175329	0.84	0.96	<a href="#">rs74064450</a>	T	C	0.22	0.41	0.35	0.32							Arid5b,CHOP:CEBPaIalpha		22kb 3' of DIO2	
chr14:80642320	chr14:80175977	0.84	0.96	<a href="#">rs2216086</a>	C	T	0.35	0.43	0.35	0.32				PANC			GR		22kb 3' of DIO2	
chr14:80642939	chr14:80176596	0.84	0.96	<a href="#">rs4999764</a>	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	<a href="#">rs4899765</a>	C	T	0.22	0.41	0.35	0.32				SKIN			11 altered motifs		21kb 3' of DIO2	
chr14:80648603	chr14:80182260	0.84	0.96	<a href="#">rs2005885</a>	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	<a href="#">rs759441</a>	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	<a href="#">rs4903904</a>	G	A	0.22	0.42	0.41	0.32							Elf3,STAT		14kb 3' of DIO2	
chr14:80655946	chr14:80189603	0.82	0.95	<a href="#">rs74064456</a>	A	T	0.23	0.42	0.44	0.32				ESDR					7.9kb 3' of DIO2	
chr14:80658261	chr14:80191918	0.86	0.96	<a href="#">rs56017760</a>	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	<a href="#">rs74064457</a>	G	A	0.24	0.42	0.44	0.32							Sox		3.2kb 3' of DIO2	
chr14:80661203	chr14:80194860	0.86	0.95	<a href="#">rs35191251</a>	A	C	0.23	0.43	0.44	0.33									2.7kb 3' of DIO2	
chr14:80662335	chr14:80195992	0.85	0.99	<a href="#">rs56025506</a>	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	<a href="#">rs225015</a>	G	A	0.48	0.45	0.43	0.31							BCL,STAT		DIO2	3'-UTR
chr14:80669580	chr14:80203237	1	1	<b>rs225014</b>	T	C	0.50	0.46	0.44	0.34				4 organs	6 bound proteins		RXRA		DIO2	missense
chr14:80673242	chr14:80206899	0.94	1	<a href="#">rs12437279</a>	C	T	0.23	0.42	0.43	0.32			<b>4 organs</b>	ESDR,BRN,BRN					DIO2	intronic

# Known genetic variation - SNPper


- No direct information mRNA or protein consequences
- Focussed on regional variation
- Submit lists of SNPs

**Find a gene**  
Enter a gene symbol (e.g. SRPR), part of a gene product name (e.g. 'liver'), or an identifier (GenBank mRNA accession number, Entrez GeneID, Ensembl ID, OMIM, Unigene, or Swissprot). Alternatively, choose from a [list of genes in alphabetical order](#).  
Symbol:  Find  
Product:   
Identifier:

**Find genes from a list**  
Enter a list of gene symbols or identifiers (see above), one per line. You can also upload a text file containing the gene identifiers.  
 Find  
Bestand kiezen Geen bestand gekozen

**Find genes by position**  
Select a chromosome and enter the start and end position of the interval you are interested in. You can also use STS marker names in place of the start and end positions.  
Chromosome: chr1 Find  
Start:   
End:

**Find genes by cytogenetic band**  
Select the chromosome and enter the cytogenetic band you are interested in (e.g., p34.1). Leave empty to see all bands. Band: chr1  Find



## Subsequent steps

- **Check for redundancy (LD):**
  - Haploview with HapMap data
  - Design assay
    - If unsuccessful, use alternative SNPs (based on LD)
  - Measure SNPs
- **Determine functionality: wet lab!**

*QUESTIONS?*



# 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!**



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