Module 6: Genomic Variation

Aims

- Introduction to genomic variation
- Introduction to various SNP resources on the web
- Integration of information from various databases to identify SNPs in your favourite gene or chromosomal region
- Choosing SNPs to genotype
- Introduction to genotypes and haplotypes

Introduction

Genetic variation is at the basis of heritable phenotype. Together with the environment, genetic variation makes each one of us different. A key goal of the human genome project is the compilation of a catalogue of common human sequence polymorphisms. With the exception of identical twins, who have identical genomes, differences between two genomes occur on average between 0.3 and 1 kb, equating to 5 - 10 million differences in a genome of 3.2 billion base pairs. Two types of genetic mutation event give rise to all genetic variants. The simplest type of variant is the substitution of a single nucleotide for another, a so-called single nucleotide polymorphism (SNP). SNPs are the commonest form of variation and when comparing 2 genomes, SNPs with a frequency > 1% typically occur every 1000 bp. Insertions or deletions of a section of DNA, so-called INDELs, account for many other types of variation. Variable number tandem repeats (VNTRs) are the commonest type of INDEL and occur where nucleotide patterns are repeated. The difference in size of VNTRs is used to divide them into minisatellites (10 -100s bp) and simple tandem repeats (STRs or microsatellites) which are 2-6bp in length. It is these SNPs and INDELs that account for most inherited phenotypes, including disease susceptibility. Analysis of sequence variation provides a powerful tool for understanding susceptibility to disease.

A single nucleotide polymorphism (SNP) is defined as a single base change occurring in a population at a frequency >1%. Single base changes that occur at <1% are often referred to as mutations or rare SNPs. However, there is a

lack of agreement between databases on this terminology and some diseasecausing mutations occur with quite a high frequency in some populations. For example, the carrier frequencies of mutations in the CFTR gene that cause cystic fibrosis are around 2% in European populations.

SNPs are highly abundant and are thought to be more stable than STRs due to low mutation rates. Nucleotide diversity is lower in exons and approximately half of the exonic SNPs are non-synonymous. SNPs can act as surrogate markers for an adjacent functional variant or can have direct functional consequences if they occur in coding or regulatory regions. The development of high-throughput genotyping platforms makes SNPs well suited to the identification of factors involved in multi-gene diseases as large sample sizes can be analysed quickly.

It is important to remember that the current SNP maps are not exhaustive and rare nucleotide substitutions, that may be critical for disease, may not be represented in the SNP maps. Re-sequencing of genomic DNA from a large number of individuals can be used to identify sequence variation. One such project is the **1000 Genomes Project**, an international research consortium formed to create the most detailed and medically useful picture to date of human genetic variation. The project involves sequencing the genomes of approximately about 2500 unidentified people from about 25 populations around the world. It's being supported by the Wellcome Trust Sanger Institute in England, the Beijing Genomics Institute Shenzhen in China and the National Human Genome Research Institute (NHGRI), part of the National Institutes of Health (NIH) in the US. The project draws on the expertise of multidisciplinary research teams and will develop a new map of the human genome that will provide "a view of biomedically relevant DNA variations at a resolution unmatched by current resources". More information about the 1000 Genome Project can be found at the project website: www.1000genomes.org

Table 2 •	Detection	rate for SN	Ps with a gi	ven minima	al allele fre	quency
n	1%	5%	10%	20%	30%	40%
2	.21	.30	.36	.43	.47	.49
3	.32	.46	.55	.65	.71	.74
4	.39	.56	.66	.77	.83	.86
5	.44	.62	.73	.84	.90	.93
6	.48	.68	.78	.89	.94	.96
7	.52	.72	.83	.92	.96	.98
8	.55	.75	.86	.94	.98	.99
9	.57	.78	.88	.96	.98	.99
10	.59	.80	.90	.97	.99	.997
16	.69	.89	.96	.99	.999	>.999
24	.76	.95	.99	.999	>.999	>.999
48	.87	.99	.999	>.999	>.999	>.999
96	.95	.999	>.999	>.999	>.999	>.999
192	.99	>.999	>.999	>.999	>.999	>.999

Kruglvak and Nickerson

Table adapted from Kruglyak and Nickerson, Nature Genetics vol 27, page 234, showing the detection rate for SNPs with a given minimal allele frequency in *n* chromosomes.

The Evolution of SNPs

The appearance of mutations and their evolution to SNPs has been defined in four phases (Miller and Kwok, 2001):

- 1. Appearance of new variant allele by mutation
- 2. Survival of allele through early generations against the odds
- 3. Increase of the allele to a substantial population frequency
- 4. Fixation of allele in populations

Survival of mutations is limited and a lot are lost in early generations. A heterozygous individual having 2 offspring has a 0.75 probability of passing on the mutation to at least one child. If the mutation is neutral, there is a 94 % probability of loss in 10 generations (approximately 200 years). Deleterious mutations disappear more quickly.

Ambiguity codes:

M => a/c	V => a/c/g	N => a/c/g/t
H => a/c/t	R => a/g	D => a/g/t
W => a/t	S => c/g	B => c/g/t
Y => c/t	K => g/t	

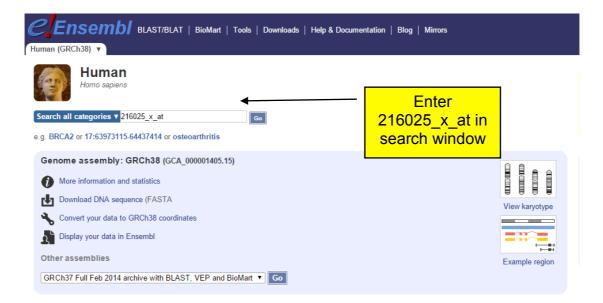
Worked Example:

You have been using Affymetrix GeneChips to identify genes that are differentially regulated in patients and controls. Following analysis, the gene corresponding to the probe set 216025_x_at is found to be up-regulated in the patient samples from an Affymetrix gene expression experiment. You need to find out information about this gene, in particular whether there are any polymorphisms in the gene and if these could affect its activity. *To determine a complete SNP map for a gene, information from several databases may need to be combined.*

Questions:

- 1. What is the function of the gene that probe set 216025_x_at corresponds to?
- 2. How many single nucleotide polymorphisms are there in this gene?
- 3. How many of the SNPs are coding and how many alter the amino acid sequence?

To start, access Ensembl, <u>http://www.ensembl.org</u> and select the Human homepage to begin the database search



Colour

Human (GRCh38) 🔻	/BLAT BioMart Tools Downloads Help & Documentation Blog Mirrors
Current selection:	Only searching Human V 216025 x at
< all Species Only searching Human	1 results match 216025_x_at when restricted to species: Human 💥
Restrict category to: ProbeFeature 1	216025_x_at (Human AFFY Probe) 216025_x_at AFFY probeset 216025_x_at has probes which hit the genome in 9 locations. They hit transcripts in the following gene: CYP2C9 (ENST00000260632).
Per page: 10 25 50 100 Layout:	<< < 1 > >> Ensembl release 78 - December 2014 © WTSI / OligoProbe
Standard Table	EBI Permanent link - View in archive site
	Discidiner Conductos
near the top of your search by updating your favourite species.	Disclaimer Contact Us
You can choose which results appear near the top of your search by updating your favourite species.	Check for multiple locations

Key

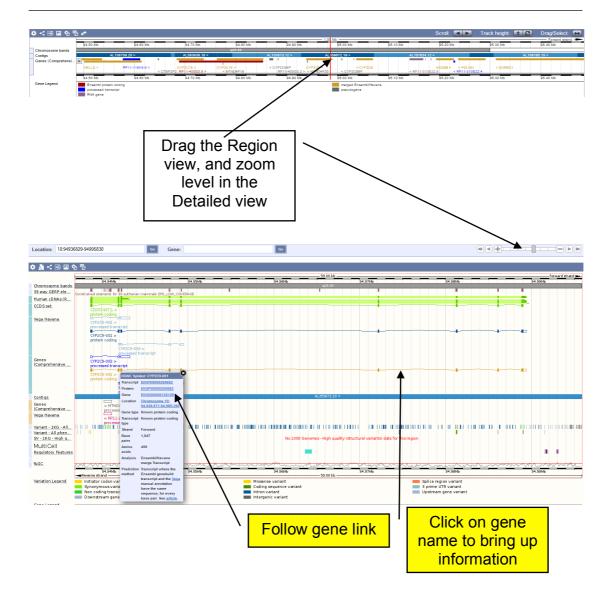
Feature type Oligoprobe

Transcript

Oligoprobe Information

Show/hide columns		
Genomic location (strand)	🕂 Length	Ensembl ID
10:94986052-94986076(1)	25	HG-U133A:216025_x_at:323:257 HG-U133A_2:216025_x_at:533:249 HG-U133_Plus_2:
10:94986097-94986121(1)	25	HG-U133A:216025_x_at:274:63 HG-U133A_2:216025_x_at:529:61 HG-U133_Plus_2:21 Follow Genomic
10:94986148-94986172(1)	25	HG-U133A-216025_x_at:318:455 HG-U133A_2:216025_x_at:317:441 HG-U133_Plus_2 HG_U132A_216025_x_at:318:455 HG-U133A_2:216025_x_at:317:441 HG-U133_Plus_2
10:94988884-94988908(1)	25	HG-U133A:216025_x_at:310:525 HG-U133A_2:216025_x_at:34:507 HG-U133_Plus_2:2
10:94988970-94988994(1)	25	HG-U133A:216025_x_at:504:707 HG-U133A_2:216025_x_at:339:685 HG-U133_Plus_2:210025_x_at:100:1157
10:94989078-94989102(1)	25	HG-U133A:216025_x_at:190:607 HG-U133A_2:216025_x_at:686:587 HG-U133_Plus_2:216025_x_at:274:983
10:94989102-94989126(1)	25	HG-U133A:216025_x_at:129:179 HG-U133A_2:216025_x_at:628:173 HG-U133_Plus_2:216025_x_at:1009:311
10:94989137-94989161(1)	25	HG-U133A:216025_x_at:529:103 HG-U133A_2:216025_x_at:303:101 HG-U133_Plus_2:216025_x_at:20:181
10:94989187-94989211(1)	25	HG-U133A:216025_x_at:475:401 HG-U133A_2:216025_x_at:525:389 HG-U133_Plus_2:216025_x_at:732:647

183



an (GRCh38) 🔻 \Upsilon Location: 10								OIIOW	HGNC		
b-based displays Immary	Gene: CYP20	C9 ENSG0000138109									
lice variants (3)											
anscript comparison	Description		ome P450, family 2, subfam	ily C, polypep	tide 9 [Source	e:HGNC Symbol	I;Acc:HGNC	:2623]			
pporting evidence ene alleles	Synonyms	CYP2C	10, P450IIC9			/					
quence	Location	Chromo	some 10: 94,938,658-94,98	390 forward	strand.						
Secondary Structure ternal references	INSDC coordinates	s chromo	some:GRCh38:CM000672.2	94938658:94	989390:1						
gulation	Transcripts	This as	ne has 3 transcripts (splice	ariante) Hid	r transcript ta	abla					
mparative Genomics		rins ge	ne nas 5 transcripts (spilce		transcript ta						
Genomic alignments Gene tree								XL			
Gene gain/loss tree	Show/hide colum	ns (1 hidden)				Filter					
Orthologues (168) Paralogues (17) Ensembl protein families (1)		Inscript ID bp Prote T00000260682 1847 490		CCDS <u>CCDS7437</u>			Flags CODE basic	APPRIS PI			
enotype netic Variation	CYP2C9-002 ENS		otein Processed ranscrip		<u>NP_000762</u> -		TSL:1				
/ariation table /ariation image Structural variation	CYP2C9-003 ENS		otein Processed transcrip	-	-		TSL:2				
ernal data Personal annotation History	Summary										
Gene history	Name		9 (HGNC Symbol)								
onfigure this page	CCDS	-	ne is a member of the Huma								
oningare uno page	UniprotKB	This ge	ne has proteins that corresp	and to the follo	owing Uniprot	identifiers: P117	712				
dd your data	RefSeq		ping RefSeq Gene ID <u>1559</u> r								
	Ensembl version		0000138109.9				-				
xport data	GRCh37 assembly			19 1/7 in OD4	h37 coerdia-	atos					
ookmark this page	Groupr assembly		ne maps to <u>96,698,415-96,7</u> ID ENSG00000138109 not p			ncə.					
oonanana ana paye	Gene type		protein coding	eaent iff GRU	mut.						
hare this page				ale and a set				and and the			
	Prediction Method		tion for this gene includes bo				i <u>Havana</u> ma	nual curation, se	e <u>article</u> .		
	Alternative genes		ene corresponds to the fol		oase identifie	ers:					
		Havan	a gene: OTTHUMG0000001	<u>3805</u>							
IGO Gene Nomenclature Committee											
			1 11 1					and a state			
tome Search Genes Dor Gene Symbol F		milies HCOP Useful Li	nks About Contact	US		Re	equest Sy				
Gene Symbol F CYP2C9		CYP2C9		US		Re	equest sy				
Gene Symbol F YP2C9 hpproved Symbol + hpproved Name +		CYP2C9 cytochrome P450, family 2, sub		US		Re	equest sy				
Gene Symbol F TYP2C9 (Approved Symbol + (GNC ID +		CYP2C9 cytochrome P450, family 2, sub HGINC:2623	family C, polypeptide 9	<u></u>	<u>.</u>	Re			Paragona		
sene Symbol F YP2C9 pproved Symbol + pproved Name + GIGKE DP + revious Symbols & Names +		CYP2C9 cytochrome P450, family 2, sul HGNC:2623 CYP2C10, "cytochrome P450, r	family C, polypeptide 9 ubfam				ברע	פכוז וייצמי ד	19 2	with	W. HEP NAVE
iene Symbol F YP2C9 pproved Symbol + pproved Name + KGKC ID + revious Symbols & Names + ynonyms +		CYP2C9 cytochrome P450, family 2, sub HGNC:2623 CYP2C10, "cytochrome P450, r P45011C9	family C, polypeptide 9 ubfam		-		ברע		SCIENCE	with	XENT
iene Symbol F YP2C9 pproved Symbol + upproved Name + GMC TD + revious Symbols & Names + ynonyms + ocus Type +		CYP2C9 cytochrome P450, family 2, sul HGNC:2623 CYP2C10, "cytochrome P450, r	family C, polypeptide 9 ubfam	rds®		WE	ברע IZMANN I	מכוז ויצמז נ אstitute of			
pproved Symbol + pproved Symbol + pproved Name + GAC ID + revious Symbols & Names + ynonyms + ocus Type +		CYP2C9 cytochrome P450, family 2, sub HGNC:2623 CYP2C10, "cytochrome P450, r P45011C9	family C, polypeptide 9 ubfam	rds®	sne Compendi	WH: Free for aca	ברע IZMANN I ademic non pro	פכון ויצפן נ אנו אנען אנען ofit institutions. All	L other users need a <u>CO</u>	mmercial licen	
eene Symbol F YP2C9 pproved Symbol + pproved Name + GNC ID + revious Symbols & Names + ynonyms + ocus Type + hromosomal Location +		CYP2C9 cytochrome P450, family 2, sub HGNC:2623 CYP2C10, "cytochrome P450, s P45011C9 gene with protein product 10q24.1	family C, polypeptide 9 ubfam	rds®		WH: Free for aca	ברע IZMANN I	פכון ויצפן נ אנו אנען אנען ofit institutions. All	L other users need a <u>CO</u>		
ene Symbol F YP2C9 pproved Symbol + pproved Name + GNC ID + revious Symbols & Names + ynonyms + scus Type + hromosomal Location + ENE FAMILY +		CYP2C9 cytochrome P450, family 2, sut HGNC:2623 CYP2C10, "cytochrome P450, f P4501C9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D	rfamily C, polypaptide 9	rds® The Human Gen GeneCards Gu		WH: Free for aca	ברע IZMANN I ademic non pro	פכון ויצפן נ אנו אנען אנען ofit institutions. All	L other users need a <u>CO</u>	er Feedback	se from Xennex, 1
eene Symbol F YP2C9 pproved Symbol + pproved Name + GNC ID + revious Symbols & Names + ynonyms + ocus Type + hromosomal Location + ENE FAMILY + OMOLOGS +		CYP2C9 cytochrome P450, family 2, sub HGNC:2623 CYP2C10, "cytochrome P450, s P45011C9 gene with protein product 10q24.1 Cytochrome P450 family C	rfamily C, polypeptide 9 ubfam Genecia Home Set Analyses:	rds® The Human Ge GeneCards Gi Gene	uide Si neDecks (Previous n	WEL Tree for sicility uite Terms cytochrome I names: cytochrom	ערצי IZMANN I ademic non pro is and Conditi P450, fam me P450, sub	עכון ויצבין ר אזדער פינטין ויצבין אזדער איזער איזער איזער אניער איזער איזער איזער אניער איזער איזער איזער איזער אניער איזער איזער איזער איזער איזער איזער אניער איזער איזער אניער איזער איזע	L other users need a <u>co</u> Us Use Illy C, polypepti nytoin 4-hydroxylase	inter Seedback	se from Xennex, 1 Mirro Search Ad
eene Symbol F YP2C9 pproved Symbol + pproved Name + GNC ID + revious Symbols & Names + ynonyms + ocus Type + hromosomal Location + ENE FAMILY + OMOLOGS +		CYP2C9 cytochrome P450, family 2, sult HGNC:2623 CYP2C10, "cytochrome P450, r P45011C9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DDB3 RefSectIM. 00772 D CCOS:cCOSY273.1 C	family C, polypeptide 9 ubfam Genecia Home Set Analyses:	rds® The Human Ge GeneCards Gi Gene GIFtS: 63	uide Si neDecks (Previous n	WEL Tree for sicility uite Terms cytochrome I names: cytochrom	בירע IZMANN I is and Condition P450, stat me P450, stat the Huido G	עכון ויצבין ר אזדער פינטין ויצבין אזדער איזער איזער איזער אניער איזער איזער איזער אניער איזער איזער איזער איזער אניער איזער איזער איזער איזער איזער איזער אניער איזער איזער אניער איזער איזע	L other users need a <u>co</u> Us Use Us Use illy C , polypepti mytoin 4-hydroxylase; <i>c</i> committee (HGNC) (inter Seedback	se from Xennex, 1 Mirro Search Ad
eene Symbol F YP2C9 pproved Symbol + pproved Name + GNC ID + revious Symbols & Names + ynonyms + ocus Type + hromosomal Location + ENE FAMILY + OMOLOGS +		CYP2C9 cytochrome P450, family 2, sub HGNC:2623 CYP2C10, cytochrome P450, r P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DDB3 RefSeatIML 000721 D CCD5::CCD5:373.1 C Yeas:0TTHUMG0000018805	family C, polypaptide 9 ubfam Geneca Home Set Analyses: C C CYP2C9 protein-coding	rds® The Human Ge GeneCards Gi Gene GIFtS: 63	uide Si neDecks (Previous n	WEL Tree for sicility uite Terms cytochrome I names: cytochrom	لاعتری IZMANN I addemic non pr addemic non pr addemic non pr s and Conditii s and Conditii P450, sub P450, fam the <u>HUGO &</u> (Previo	ניבלא ויצבא בי ארארי לה וווווווווווווווווווווווווווווווווווו	L other users need a <u>co</u> Us Use Us Use illy C , polypepti mytoin 4-hydroxylase; <i>c</i> committee (HGNC) (inter Seedback	se from Xennex, 1 Mirro Search Ad
eene Symbol F YP2C9 pproved Symbol + pproved Name + GNC ID + revious Symbols & Names + ynonyms + ocus Type + hromosomal Location + ENE FAMILY + OMOLOGS +		CYP2C9 cytochrome P450, family 2, sul HGNC:2623 CYP2C10, "cytochrome P450, r P450IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCO2P D TreeFam D GenBank:M51855 EMBL DDB3 RefSea:1ML 000721 D CCD5:CCD57437.1 C Yea:OTHUMS0000018805 Entre: Gene:1555 C	family C, polypeptide 9 ubfam Genecia Home Set Analyses: C C C C C C	rds® The Human Ge GeneCards Gr Gene Genes Genes Gares Gares	uide Si neDecks (Previous n Syrr	WE! Im Preeforsta uite Terms cytochrome I names: cytochron bol approved by	العلام IZMANN I Is and Conditions Is and Conditions Is and Conditions Is and Conditions Is and Conditions Is a conditions (Previo Previo	ככז ויצביז רבי NSTITUTE OF Mematutions. A bout keyword(c) Nily 2, subfanity (C (memacularity) us symbol: CYP2(d) us symbol: CYP2(d) dicts	L other users need a <u>co</u> Us Use illy C, polypepti nytoin 4.hydroxylase) <i>Committee (HGNC) o</i> C10)	mmercial licen: r Feedback ide 9), polypeptide 9 database	Search Annex, I
ene Symbol F YP2C9 pproved Symbol + pproved Name + GKC ID + revious Symbols & Names + yonoyms + ocuts Type + hromosomal Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES +		CYP2C9 cytochrome P450, family 2, sub HGNC:2623 CYP2C10, cytochrome P450, r P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DDB3 RefSeatIML 000721 D CCD5::CCD5:373.1 C Yeas:0TTHUMG0000018805	Afamily C, polypeptide 9 ubfam Genecia Home Set Analyses: C C C C C C C C C C C C C	rds® The Human Ge GeneCards Gi Gene <u>CiFtS</u> : 63 688	uide Si ne Decks (Previous r Syrr	WEL Tree for sicility uite Terms cytochrome I names: cytochrom	IZMANN I ademic non pro- ta ademic non pro- ta and Condition (Previous) (Prev	ניבלא ויצבא בי ארארי לה וווווווווווווווווווווווווווווווווווו	L other users need a <u>co</u> Us Use Us Use illy C , polypepti mytoin 4-hydroxylase; <i>c</i> committee (HGNC) (interestal licens ir Feedback ide 9), polypeptide 9 database Antibodies	se from Xennex, 1 Mirro Search Ad
ene Symbol F YP2C9 pproved Symbol + pproved Name + GKC ID + revious Symbols & Names + ynonyms + pocus Type + hromosoma Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES +		CYP2C9 cytochrome P450, family 2, sul HGNC:2623 CYP2C10, "cytochrome P450, r P450IC5 gene with protein product 10q24.1 Cytochrome P450 family C Adochrome P450 family C BenBank:M61855 EMBL DDB3 RefSeat:IM 000721 D CCD5:CCD5X137.1 C Yea::OTTUNK0000018805 Entre: Gene:1552 C Enter: Gene:1552 C Enter: Gene:1552 C	Mamily C, polypaptide 9 ubfam Genediat None Set Analyses: C C C C C C C C C C C C C	rds ® The Human Ge GeneCards Ge Genes Gen	uide Si ne Decks (Previous r Syrr	WEI Terre for sca uite Terrer cytochrome I hames: cytochrom bol approved by ss/cDNA/RNAi ss/cDNA/RNAi ss/cDNA/RNAi	IZMANN I ademic non pro- ta ademic non pro- ta and Condition (Previous) (Prev	אלט איז	L other users need a <u>co</u> Us Use illy C , polypepti inty C , polypepti inty committee (<i>HGNC</i>) (c) i Committee (<i>HGNC</i>) (c) C ONTIGENE	interestal licens ir Feedback ide 9), polypeptide 9 database Antibodies	Search Ad
eene Symbol F YP2C9 pproved Symbol + pproved Name + GKC ID + revious Symbols & Names + ynonyms + ocus Type + hromosoma Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES +		CYP2C9 cytochrome P450, family 2, aul HGNC:2623 CYP2C10, "cytochrome P450, r P4501C9 gene with protein product 10q24.1 Cxtochrome P450 family C HCOP D TreeFam D GenBank:M61855 EMBL DDBJ RefSeat/M 000721 D CCD5:CCD574372.4 C Yeas:0THUHG0000018805 Entres Gene1552 C Entere Science1552 C Entere Science1552 C	Afamily C, polypeptide 9 ubfam Genecia Home Set Analyses: C C C C C C C C C C C C C	rtds ® The Human Ge GeneCards Gi Gene CIFES: 63 688	uide Sr neDecks (Previous n Syn M Antibodie Proteins Assays & s & Description rome P450, fan	WEI Terre for sca uite Terrer cytochrome I hames: cytochrom bol approved by ss/cDNA/RNAi ss/cDNA/RNAi ss/cDNA/RNAi	IZMANN I ademic non pro- ss and Condition P450, fan Pro- (Previo Pro- stand Pro- Pro- stand Pro- Pro- Pro- Pro- Pro- Pro- Pro- Pro-	אלי איז איז איז איז איז איז איז איז איז אי	L other usery need a co Us Use Us Us Us Use Use Use Use Use Use Use	in Feedback	se from Xennex, I Mirro Search Ard))) () () () () () () () () () () () (
eene Symbol F YP2C9 pproved Symbol + pproved Name + GKC ID + revious Symbols & Names + ynonyms + ocus Type + hromosoma Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES +		CYP2C9 Cytochrome P450, family 2, sult HGNC:2623 CYP2C10, "cytochrome P450, r P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GanBank:M61855 EMEL DDB1 RefSeat1M 00771 D CCDS:CCDS7371 C Yeaa:OTTHUMG0000018805 Entre: Gene:1552 C Entre: Gene:1552 Entre: Gene:1552 Entre: Gene:1552 Entre: Gene:1552 CUCS:CLUCOLIKka.3 D Veca:OTTHUMG0000018805 UICSCLUCOLIKka.3 D Veca:OTTHUMG0000018805	family C, polypaptide 9 ubfam Genecica Nome Set Analyses: C C C C C C C Set Analyses: C C C C C C Set Analyses: C C C C C C C C C C C C C	rtds ® The Human Get GeneCards Go Constant Galfris: 63 Galfris: 63 Galfris: 63 Galfris: 63 Calfris: 63	uide St mcDecks (Previous n Syn Proteins Assays F s & Description rome P450, fan C9 ¹²	WEI Terres for sca uite Terms cytochrome I names: cytochron hol approved by rs/cDNA/RNAi s Enzymss kts, # athways mily 2, subfamily C	IZMANN I ademic non pro- ss and Condition P450, fan Pro- (Previo Pro- stand Pro- Pro- stand Pro- Pro- Pro- Pro- Pro- Pro- Pro- Pro-	אלי איז איז איז איז איז איז איז איז איז אי	L other users need a CO Us Us Use Thily C, polypepti http://committee/HGNC/c C Committee/HGNC/c C Committee/HGNC/c C C C C C C C C C C C C C C C C C C	mmercial licen rr Feedback ide 9), potypeptide 9 database Protaias Antibades NA/Primers	se from Xennex, I Mirrow Search Add
Gene Symbol F CYP2C9 pproved Symbol + pproved Name + GIGK ID + revious Symbols & Names + ynonyms + oorsona Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES +		CYP2C9 Cytochrome P450, family 2, sult HGNC:2623 CY2C10, "cytochrome P450, r P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeFam D GanBank:M61855 EMEL DDB1 RefSealMM 000771 D CCDS:CCDS:CAST31 C Yeaa:OTTHUMS0000018805 Entre: Gane:1552 C Entre: Gane:1552 C Entre: Gane:1552 C Entre: Gane:1552 C UCSC:LucOXIkka.3 D Veca:OTTHUMS0000018805 Uniprot:P11712 D Interco D OMIM D GaneTets D	Hamily C, polypeptide 9 wubfam Genecia Nome Set Analyses: C C C C C C Jump to Section Aliases &	The Human Ge GeneCards Gi Gene GiffS: 63 668 CiffS: 63 CiffS: 63 C	uide Si nrcDccks (Previous f Syrr Proteins Syrr s & Description come P450, fan Cgi ²	WEL More for act utte Terms cytochrome I names: cytochrom bol aproved by sc/DDNA/RNAi B Etsynetways mily 2, subfamily C s21	IZMANN I ademic non pro- ss and Condition P450, fan Pro- (Previo Pro- stand Pro- Pro- stand Pro- Pro- Pro- Pro- Pro- Pro- Pro- Pro-	אלי איז איז איז איז איז איז איז איז איז אי	L other usery need a co Us Use Use Use Use Use Use Use Use Use Use	mmercial ficen rr Feedback ide 9), polypeptide 9 database Protains Antibades NA/Primers Linked monooxy	Search Address J Addres J Address J Address J Address J Address J Address J
eene Symbol F YP2C9 pproved Symbol + pproved Name + GNC ID + revious Symbols & Names + ynonyms + cous Type + hromosomal Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES +		CYP2C9 cytochrome P450, family 2, sub HGNC:2623 CYP2C10, "cytochrome P450, r P450IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeFam D GenBant:M61855 EMBL DDBJ RefSeat:MN 00072 D CCD5:CCD57437.1 C Year:07THUMS0000018805 Entres Gene1552 C Enters:Gene1552 C Enters:Charles D Veca:07THUMS0000018805 Uniprot:P1172 D Interror D OMIM D GeneTiests D Orbans D	family C, polypeptide 9 ubfam Genecical Nome Set Analyses: Control Genecical Nome Set Analyses: Control C C CYP2C9 protein-coding GC 10098 C C CYP2C9 Set Analyses: Control C C CYP2C9 Set Analyses: Control C C CYP2C9 Set Analyses: Control C C C CYP2C9 Set Analyses: Control C C C C C C C C C C C C C C C C C C C	The Human Get GeneCards Gi GeneCards Gi Gene Genes Genes Genes Genes Genes Genes Genes Genes Genes Cifeis: 63 688	uide Si ncDzctSS (Previous r Syn Parteins s & Description rome P450, fan Cgi ² rome P450 PB	WEL More for act utte Terms cytochrome I names: cytochrom bol aproved by sc/DDNA/RNAi B Etsynetways mily 2, subfamily C s21	IZMANN I ademic non pro- ss and Condition P450, fan Pro- (Previo Pro- stand Pro- Pro- stand Pro- Pro- Pro- Pro- Pro- Pro- Pro- Pro-	אלי איז איז איז איז איז איז איז איז איז אי	L other usery need a co Us Us Use illy C, polypepti Inity C, polypepti Committee (HONC) co Committee (HONC) co To) Committee (HONC) co To) To) To) To) To) To) To) To) To)	Interest fices In Feedback ide 9), potypeptide 9 Proteins Autibides SNA/Primers linked monooxy monooxygenase 6-monooxygenase	se from Xennex, I Mirre Search Add)) () () () () () () () () () () () ()
ene Symbol F YP2C9 pproved Symbol + pproved Name + GNC ID + revious Symbols & Names + ynonyms + cous Type + hromosomal Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ROTEIN RESOURCES +		CYP2C9 Cytochrome P450, family 2, sul HGNC:2623 CYP2C10, "cytochrome P450, s P45011C9 gene with protein product 10q24.1 Cytochrome P450 family C HO2P D GenBank:M61855 EMBL DDBJ BefSau:MM 000721 D CCDS:CCD57437.1 C Veaa:OTTHURG000018805 Ehres Gene1553 C Ehres Gene1553 C Ehres Gene1553 C UICSCLudOtIkka.3 D Veaa:OTTHURG000018805 UICSCLUDOTHURG000018805 DCSCLUDOTHURG000018805 UICSCLUDOTHURG0000000000 UICSCLUDOTHURG000000000000000000000000000000000000	family C, polypeptide 9 ubfam Genecical Nome Set Analyses: Control Genecical C CYP2C9 protein-coding GC CYP2C9 protein-coding GC CYP2C9 C Jump to Settor and For CYP2C9 gene (According to FCNC-3 Genecical C CYP2C9 gene (According to FCNC-3 C CYP2C9 gene	The Human Get GeneCards Ge GeneCards Ge GeneCards Ge Genes G	uide Su neDecks (Previous n Syn Parteine s & Description Cel ² rome P450, fan Cel ² rome P450, fan Cel ² rome P450 PB	WEL More for act utte Terms cytochrome I names: cytochrom bol aproved by sc/DDNA/RNAi B Etsynetways mily 2, subfamily C s21	IZMANN I ademic non pro- ss and Condition P450, fan Pro- (Previo Pro- stand Pro- Pro- stand Pro- Pro- Pro- Pro- Pro- Pro- Pro- Pro-	אלי איז איז איז איז איז איז איז איז איז אי	L other usery need a co Us Use Use Use Use Use Use Use Use Use Use	mmereial ficen r Feedback ide 9), potypeptide 9 databace Pretains Antibades NA: Primers linked monooxy e 6-monooxygenasi	se from Xennex, I Mirri Search Ad))) () () () () () () () () () () ()
ene Symbol F YP2C9 pproved Symbol + pproved Name + GNC ID + revious Symbols & Names + ynonyms + cous Type + hromosomal Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ROTEIN RESOURCES +		CYP2C9 CYP2C9 Cytochrome P450, family 2, sul HGNC:2623 CYP2C10, "cytochrome P450, 1 P450IC5 gane with protein product 10q24.1 Cytochrome P450 family C BanBank:161855 EMBL DDB1 RefSata1M 000771 D CCDS:CCDS7437.1 C Veaa:07THUNG000018805 Entra: Gane:1559 C Entra: Gane:1559 C Entra: Gane:1559 C Entra: Gane:1559 C Uniprot:P1172 D Veaa:07THUNG000018805 Uniprot:P1172 D Constants D Gombanst D DECIPHER D COSMIC D	Mamily C, polypeptide 9 wubfam Genetical Bone Set Analyses: C C C C C C C C C C C C C C C C C C C	The Human Ge GeneCards Gi GeneCards Gi GeneCards Gi Gene Cife(5): 63 888	uide Si netbecks (Previous r Syrr Syrr s & Description come P-450, fan Cgl ² rome P-450 PB rome P-4	WEL IIII Pres for aca uite Terms cytochrome I annes: cytochrom bol approved by s:/dpNA/INNAI s: Ets.rymtwsys Is: AppNA/INNAI s: Ap	IZMANN I IZMANN I ademic non pro- se and Condition P450, stub (Previo Pro- se assence PCR Arrays	e دیک ۱۲۷۵ ۲۰۰۰ NSTITUTE OF Af unstructors AL About Keyword(ع) Hily 2, subfan family IIC (mophe ene Nomenclaus Suppoli CYP24 Subfan	L other usery need a co Us Us Use Villy C, polypepti Nilly C, polypepti Villy C, polypepti Villy C, polypepti Villy C, polypepti Villy C, polypepti Villy C, polypepti Navery Ceases / MR Raveprotein- microsomal Xenobiolic m (R)-limonen (S)-limonen	Innered Ficehock In Feedback In Feedback In Proteins Proteins Antibides Anti	se from Xennex, I Mirri Search Ad))) () () () () () () () () () () ()
eene Symbol F YP2C9 pproved Symbol + pproved Name + GRC ID + revious Symbols & Names + ynonyms + ocus Type + hnomosomal Location + ENE FAMILY + OHOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ENE RESOURCES + LINICAL RESOURCES +		CYP2C9 cytochrome P450, family 2, sul HGNC:2623 CYP2C10, "cytochrome P450, r P450IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCO2 D TreeFam D GenBank:M51855 EMBL DDB3 RefSeat/M 00072 D CCD5:CCD57437.1 C Yeas:OTTHUMG0000018805 Entre: Gene:1552 C Ensemble:M50000018805 Uniprot:D12122 D Interror D OMIM D GenEriests D Cytochant D DECIDHER D COSMIC D LSDB;Human Cytochrome P455	Annily C, polypeptide 9 Wofam Genetical Home Set Analyses: C C C C C C C C C C C C C C C C C C C	The Human Get GeneCards Ge GeneCards Ge GeneCards Ge Genes G	uide St neDecks (Previous r Synt Proteins rome P450 Markov 10 ²¹ rome P450 Markov 10 ²¹ rome P450 Markov 10 ²¹ rg2	WEL More for act utte Terms cytochrome I names: cytochrom bol aproved by sc/DDNA/RNAi B Etsynetways mily 2, subfamily C s21	IZMANN I IZMANN I ademic non pro- se and Condition P450, stub (Previo Pro- se assence PCR Arrays	e دیک ۱۲۷۵ ۲۰۰۰ NSTITUTE OF Af unstructors AL About Keyword(ع) Hily 2, subfan family IIC (mophe ene Nomenclaus Suppoli CYP24 Subfan	L other usery need a co Us Us Use Villy C, polypepti Nilly C, polypepti Villy C, polypepti Villy C, polypepti Villy C, polypepti Villy C, polypepti Villy C, polypepti Navery Ceases / MR Raveprotein- microsomal Xenobiolic m (R)-limonen (S)-limonen	Interest ficen rr Feedback ide 9), potypeptide 9 database Pretains Antibédies NA/Primers linked monooxy e 6-monooxygenasé e 6-monooxygenasé e 7-monooxygenasé	se from Xennex, I Mirri Search Ad))) () () () () () () () () () () ()
eene Symbol F YP2C9 pproved Symbol + pproved Name + GRC ID + revious Symbols & Names + ynonyms + ocus Type + hnomosomal Location + ENE FAMILY + OHOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ENE RESOURCES + LINICAL RESOURCES +		CYP2C9 cytochrome P450, family 2, sul HGNC:2623 CYP2C10, "cytochrome P450, s P4501C9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeFam D GenEant.M61855 EMBL DDBJ RefSeat.1M 0.00721 D CCD51:CCD57437.1 C Yeas:OTTHUHG0000018805 Entres Gene1553 C EnsembleNSG000018805 Uniport.P14712 D Interpro D OMIM D GenETests D Orabanet D DECIDHER D CCOSIC D LOSPILLING CAChrome P45 PMID:202263, PMID:784144	family C, polypeptide 9 ubfam Genecca Bome Set Analyses: Construction Genecca Genecca Jump to Section Aliases & Descriptions for CYP2C9 gene (According to Section Aliases & Descriptions for CYP2C9 gene (According to Section Aliases & Descriptions for CYP2C9 gene (According to Section	The Human Get GeneCards Ge GeneCards GeneCards Ge GeneCards GeneCards GeneCards GeneCards GeneCards GeneCards GeneCards GeneCards	uide Si ncDccks (Previous F Syrr Proteina Protei	WEL IIII Pres for aca uite Terms cytochrome I annes: cytochrom bol approved by s:/dpNA/INNAI s: Ets.rymtwsys Is: AppNA/INNAI s: Ap	IZMANN I IZMANN I ademic non pro- se and Condition P450, stub (Previo Pro- se assence PCR Arrays	e دیک ۱۲۷۵ ۲۰۰۰ NSTITUTE OF Af unstructors AL About Keyword(ع) Hily 2, subfan family IIC (mophe ene Nomenclaus Suppoli CYP24 Subfan	L other usery need a CO Us Us Use Inity C, polypepti Inity C, polypepti Committee (HONC) (Committee (HONC)	Interced Recon r Feedback ide 9), polypeptide 9 Pretains Autholice Pretains Autholice CNA/Primers Hinked monooxy e 6-monooxy e 6-monooxy e 7-monooxy e 7-monooxy	se from Xennex, I Mirri Search Ad))) () () () () () () () () () () ()
ene Symbol F YP2C9 pproved Symbol + pproved Name + GKC ID + revious Symbols & Names + ynonyms + pocus Type + hromosomal Location + ENE FAMILY + OHOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ENE RESOURCES + LINICAL RESOURCES +		CYP2C9 CYP2C9 CYDchrome P450, family 2, sull HGNC:2623 CYP2C10, "cytochrome P450, s P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DD23 RefSea:10M 000721 D CCDS:CCDS7437.1 C CGESCCDS7437.1 C	Annily C, polypeptide 9 Wofam Genecica Nome Set Analyses: C C C Jump to Section Aliases & Descriptions for CYP209 Centry Control Co	The Human Get GeneCards Ge GeneCards Cotoch Cot	uide Si ncDzc2s5 (Previous C (Previous C Protiona Assays 8 1 s & Description Si & Descripti	WEL IIII Pres for aca uite Terms cytochrome I annes: cytochrom bol approved by s:/dpNA/INNAI s: Ets.rymtwsys Is: AppNA/INNAI s: Ap	IZMANN I IZMANN I ademic non pro- se and Condition P450, stub (Previo Pro- se assence PCR Arrays	e دیک ۱۲۷۵ ۲۰۰۰ NSTITUTE OF Af unstructors AL About Keyword(ع) Hily 2, subfan family IIC (mophe ene Nomenclaus Suppoli CYP24 Subfan	L other users need a co Us Use Us Use Us Use Use Use Use Use Use Use Use	Innered Rech r Feedback ide 9 , potypeptic database Pretains Arathadies NA./Primers linked monooxygenasei e 6-monooxyge e 7-monooxyge e 7-monooxyge gal potypeptic gal potypeptic poty	se from Xennex, I Mirri Search Ad))) () () () () () () () () () () ()
eene Symbol F YP2C9 pproved Symbol + pproved Name + GRC ID + revious Symbols & Names + ynonyms + ocus Type + hnomosomal Location + ENE FAMILY + OHOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ENE RESOURCES + LINICAL RESOURCES +		CYP2C9 cytochrome P450, family 2, sul HGNC:2623 CYP2C10, "cytochrome P450, s P4501C9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeFam D GenEant.M61855 EMBL DDBJ RefSeat.1M 0.00721 D CCD51:CCD57437.1 C Yeas:OTTHUHG0000018805 Entres Gene1553 C EnsembleNSG000018805 Uniport.P14712 D Interpro D OMIM D GenETests D Orabanet D DECIDHER D CCOSIC D LOSPILLING CAChrome P45 PMID:202263, PMID:784144	family C, polypeptide 9 ubfam Genecca Bome Set Analyses: Control Genecca C C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C C CYP2C9 protein-coding GC C C C C C C C C C C C C C C C C C C	The Human Ge GeneCards Ge GeneCards Ge GeneCards Ge Genes Ge	uide Sk mcDzc2SS (Previous F Syrr Proteins Assays & 1 s & Description Gall Coll 2007 1007 1007 1007 1007 1007 1007 1007	WEI Terrer for sca uite Terrer cytochrome hames: cytochrom hol approved by scaptochrome is & Enzymes is & Enzymes is & Enzymes is & Enzymes is & Enzymes hilly 2, subfamily C 22 j 12 bfamily IIC (meph- mephenytoin 4-hy)	IZMANN I IZMANN I ademic non pro- se and Condition P450, state (Previo Previo	e دیک ۱۲۷۵ ۲۰۰۰ NSTITUTE OF Af unstructors AL About Keyword(ع) Hily 2, subfan family IIC (mophe ene Nomenclaus Suppoli CYP24 Subfan	L other usery nod 4 c0 Us Us Us illy C, polypepti inity C, polypepti vitin 4-hydroxylase Committee (HSNC) c Committee (H	Interest ficen r Feedback ide 9), potypetide 9 diabase Proteins Antibades BNA/Primers Linked monooxy monooxygenasie 6 E-monooxygenasie 6 E-monooxygenasie 6 E-monooxygenasie 6 E-monooxygenasie 8 E-monooxygenasie 8 E-monooxygenasie 8 E-monooxygenasie 8 E-monooxygenasie 9 P-monooxygenasie 8 E-monooxygenasie 9 P-monooxygenasie 9 P-monooxygenasie 8 P-monooxygenasie 9 P-monoxygenasie 9 P-monooxygenasie 9 P-monooxygenasie	Sectron Xonnex L Mirre Search Ad
eene Symbol F YP2C9 pproved Symbol + pproved Name + GRC ID + revious Symbols & Names + ynonyms + ocus Type + hnomosomal Location + ENE FAMILY + OHOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ENE RESOURCES + LINICAL RESOURCES +		CYP2C9 CYP2C9 CYDchrome P450, family 2, sull HGNC:2623 CYP2C10, "cytochrome P450, s P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DD23 RefSea:10M 000721 D CCDS:CCDS7437.1 C CGESCCDS7437.1 C	family C, polypeptide 9 ubfam Genecca Bome Set Analyses: Control Genecca C C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C C CYP2C9 protein-coding GC C C C C C C C C C C C C C C C C C C	The Human Ge GeneCards Ge GeneCards Ge GeneCards Ge Genes Ge	uide St ncbbcks (Previous F (Previous F Syn Parbeins Parbeins St Description Cgi² Pato ins Cgi² Pato ins Sg² Pato ins Sig² Pato ins Sig² Sig² Sig² Sig² Sig² Sig² Sig² Sig² Sig² Sig² Sig² Sig²	WEI Terrer for sca uite Terrer cytochrome hames: cytochrom hol approved by scaptochrome is & Enzymes is & Enzymes is & Enzymes is & Enzymes is & Enzymes hilly 2, subfamily C 22 j 12 bfamily IIC (meph- mephenytoin 4-hy)	IZMANN I IZMANN I ademic non pro- se and Condition P450, state (Previo Previo	e دیک ۱۲۷۵ ۲۰۰۰ NSTITUTE OF Af unstructors AL About Keyword(ع) Hily 2, subfan family IIC (mophe ene Nomenclaus Suppoli CYP24 Subfan	L other usery need a co Us Us Use Vis Use Vis C, polypepti Vis	Interest ficen r Feedback ide 9), potypetide 9 diabase Proteins Antibades BNA/Primers Linked monooxy monooxygenasie 6 E-monooxygenasie 6 E-monooxygenasie 6 E-monooxygenasie 6 E-monooxygenasie 8 E-monooxygenasie 8 E-monooxygenasie 8 E-monooxygenasie 8 E-monooxygenasie 9 P-monooxygenasie 8 E-monooxygenasie 9 P-monooxygenasie 9 P-monooxygenasie 8 P-monooxygenasie 9 P-monoxygenasie 9 P-monooxygenasie 9 P-monooxygenasie	Sectron Xonnex L Mirre Search Ad
Gene Symbol F YP2C9 upproved Symbol + upproved Name + (GIC ID + revious Symbols & Names + ynonyms + ocus Type + hromosomal Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ENE RESOURCES + LINICAL RESOURCES +		CYP2C9 CYP2C9 CYDchrome P450, family 2, sull HGNC:2623 CYP2C10, "cytochrome P450, s P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DD23 RefSea:10M 000721 D CCDS:CCDS7437.1 C CGESCCDS7437.1 C	family C, polypeptide 9 ubfam Genecca Bome Set Analyses: Control Genecca C C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C C CYP2C9 protein-coding GC C C C C C C C C C C C C C C C C C C	The Human Ge GeneCards Ge GeneCards Ge Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Crock Cro	uide Ski mcDzc2SS (Previous F Syrr Proteins Assays & 1 s & Description Gall Coll 10 ⁶¹ sgl a rome P450, sul sgl a socome P450, sul sgl a sgl a sg	WEI Terme for acc uite Terme cytochrome tames: cytochrom bol approved by ss/cDNA/RNAi s & Enzymes title, Pathways mily 2, subfamily C 23 bfamily IIC (meph- mephenytoin 4-hy- gi	عدر بر IZMANN I adomic non pro se and Condition (Previo (Previo PCB Arrays) C, polypeptide henytoin 4-hyd	ercori الانتخاب المعالي About the official and the offi	L other usery need a co Us Us Use Us Use Us Use Us C, polypepti ntily C, polypepti ntily C, polypepti commitee (FRVC) commitee (FRVC) committee (FRVC) commit	Inthereial Ficehover r Feedback ide 9), polypeptide 9 dialabase Proteins Antibades e 6-monocxygena nonocxygenas nonocxygenas e 6-monocxyge e 6-monocxyge e 6-monocxyge e 6-monocxyge e 7-monocxygena polygena e 7-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena e 6-monocxygena polygena e 6-monocxygena polygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena e 6-m	Sectron Xonnex L Mirre Search Ad
Gene Symbol F YP2C9 upproved Symbol + upproved Name + (GIC ID + revious Symbols & Names + ynonyms + ocus Type + hromosomal Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ENE RESOURCES + LINICAL RESOURCES +		CYP2C9 CYP2C9 CYDchrome P450, family 2, sull HGNC:2623 CYP2C10, "cytochrome P450, s P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DD23 RefSea:10M 000721 D CCDS:CCDS7437.1 C CGESCCDS7437.1 C	family C, polypeptide 9 ubfam Genecca Bome Set Analyses: Control Genecca C C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C C CYP2C9 protein-coding GC C C C C C C C C C C C C C C C C C C	The Human Ge GeneCards Ge GeneCards Ge GeneS Giffs: 633 688 VV27: Citoth Cytoth	uide Ski mcDzc2SS (Previous F Syrr Previous Syrr Syrr s & Description Gall Syll Trome P450, Fan Gall Syll Syll Syll Syll Syll Syll Syll S	WEI Terrer for acc uite Term cytochrome hames: cytochrom hames: cytochrom hames: cytochrom hames: cytochrom hames: cytochrom hames: cytochrom bit England tit England tit England hames: cytochrom hames: cytochrom h	الكترين IZMANN I ademic non pro se and Condition Processor (Prevol Processor Proceso	ercori التخديم fit institutions من fit institutions من fit institutions at hereine bisenesses family IIC (mephanismi to symbol: CYP2 ducts us Symbol: CYP2 ducts g12 roxylase), polypep Ensembl: ENSCO	L other usery nod 4 c0 Us Us Us illy C, polypepti inity C, polypepti vitin 4-hydroxylase Committee (HSNC) c Committee (H	Inthereial Ficehover r Feedback ide 9), polypeptide 9 dialabase Proteins Antibades e 6-monocxygena nonocxygenas nonocxygenas e 6-monocxyge e 6-monocxyge e 6-monocxyge e 6-monocxyge e 7-monocxygena polygena e 7-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena e 6-monocxygena polygena e 6-monocxygena polygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena polygena e 6-monocxygena e 6-m	Sectron Xonnex L Mirre Search Ad
Gene Symbol F YP2C9 upproved Symbol + upproved Name + (GIC ID + revious Symbols & Names + ynonyms + ocus Type + hromosomal Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ENE RESOURCES + LINICAL RESOURCES +		CYP2C9 CYP2C9 CYDchrome P450, family 2, sull HGNC:2623 CYP2C10, "cytochrome P450, s P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DD23 RefSea:10M 000721 D CCDS:CCDS7437.1 C CGESCCDS7437.1 C	family C, polypeptide 9 ubfam Genecca Bome Set Analyses: Control Genecca C C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C C CYP2C9 protein-coding GC C C C C C C C C C C C C C C C C C C	The Human Ge GeneCards Gi GeneCards Gi GeneCards Gi Gene Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Cres Cres Cres Cres Cres Cres Cres Cr	uide Si ncDccRS (Previous F Syrr Autho die (Previous F Syrr Autho die Syrr Autho die Syrr Autho die Syrr Autho die Syrr Autho die Autho die Syrr	WEL Im Pree for accurate uite Termination cytochrome Image: cytochrome hames: cytochrome Image: cytochrome initial approved by Image: cytochrome initiapproved by Image: cytochrome	IZMANN I adomic non pro se and Condition Part and Condition (Previo Proce Arrays) C, polypeptide nenytoin 4-hyd ydroxylase ² Sene: 1559 ⁴ side databasy	ecori (الالالالم المحمد الالالالم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد ال	L other usery need a co Us Us Use Us Use Us Use Us C, polypepti mytion 4 hydroxylase Committee (FRNC) co 10) Committee (FRNC) co 10) Co	Inthereal Ficehover r Feedback ide 9), polypeptide 9 didatabase Proteins Antibides Proteins Antibides Proteins Proteins Antibides Proteins Antibides Proteins Antibides Proteins Proteins Antibides Proteins Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Antibides Proteins Antibides Proteins Antibides Proteins Antibides Antibides Proteins Antibides Anti	Sectron Xonnex L Mirre Search Ad
iene Symbol F YP2C9 pproved Symbol + .pproved Name +		CYP2C9 CYP2C9 CYDchrome P450, family 2, sull HGNC:2623 CYP2C10, "cytochrome P450, s P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DD23 RefSea:10M 000721 D CCDS:CCDS7437.1 C CGESCCDS7437.1 C	family C, polypeptide 9 ubfam Genecca Bome Set Analyses: Control Genecca C C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C C CYP2C9 protein-coding GC C C C C C C C C C C C C C C C C C C	The Human Ge GeneCards Gi GeneCards Gi GeneCards Gi Gene Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Genes Cres Cres Cres Cres Cres Cres Cres Cr	uide Si ncDccRS (Previous F Syrr Autho die (Previous F Syrr Autho die Syrr Autho die Syrr Autho die Syrr Autho die Syrr Autho die Autho die Syrr	WEL Im Pree for accurate uite Termination cytochrome Image: cytochrome hames: cytochrome Image: cytochrome initial approved by Image: cytochrome initiapproved by Image: cytochrome	IZMANN I adomic non pro se and Condition Part and Condition (Previo Proce Arrays) C, polypeptide nenytoin 4-hyd ydroxylase ² Sene: 1559 ⁴ side databasy	ecori (الالالالم المحمد الالالالم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد ال	L other usery need a co Us Us Use Us Use Us Use Us C, polypepti ntily C, polypepti ntily C, polypepti commitee (FRVC) commitee (FRVC) committee (FRVC) commit	Inthereal Ficehover r Feedback ide 9), polypeptide 9 didatabase Proteins Antibides Proteins Antibides Proteins Proteins Antibides Proteins Antibides Proteins Antibides Proteins Proteins Antibides Proteins Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Antibides Proteins Antibides Proteins Antibides Proteins Antibides Antibides Proteins Antibides Anti	Sectron Xonnex L Mirre Search Ad
Gene Symbol F YP2C9 upproved Symbol + upproved Name + (GIC ID + revious Symbols & Names + ynonyms + ocus Type + hromosomal Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ENE RESOURCES + LINICAL RESOURCES +		CYP2C9 CYP2C9 CYDchrome P450, family 2, sull HGNC:2623 CYP2C10, "cytochrome P450, s P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DD23 RefSea:10M 000721 D CCDS:CCDS7437.1 C CGESCCDS7437.1 C	family C, polypeptide 9 ubfam Genecca Bome Set Analyses: Control Genecca C C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C CYP2C9 protein-coding GC C C CYP2C9 protein-coding GC C C C C C C C C C C C C C C C C C C	The Human Ge GeneCards Gi GeneCards Gi GeneC	uide St ncDzcks (Previous ''''''''''''''''''''''''''''''''''''	WEL WEL WEL WEL WE Cytochrome Cytochrome sytochrome sytochrom sytochrom state state s	IZMANN I adomic non pro se and Condition Part and Condition (Previo Proce Arrays) C, polypeptide nenytoin 4-hyd ydroxylase ² Sene: 1559 ⁴ side databasy	ecori (الالالالم المحمد الالالالم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحم المحمد المحم المحمد المحمد المحمد المحمد المحمد المحمد المحمد ال	L other usery need a co Us Us Use Us Use Us Use Us C, polypepti mytion 4 hydroxylase Committee (FRNC) co 10) Committee (FRNC) co 10) Co	Inthereal Ficehover r Feedback ide 9), polypeptide 9 didatabase Proteins Antibides Proteins Antibides Proteins Proteins Antibides Proteins Antibides Proteins Antibides Proteins Proteins Antibides Proteins Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Proteins Antibides Antibides Proteins Antibides Proteins Antibides Proteins Antibides Antibides Proteins Antibides Anti	Sectron Xonnex L Mirre Search Ad
Gene Symbol F YP2C9 upproved Symbol + upproved Name + (GIC ID + revious Symbols & Names + ynonyms + ocus Type + hromosomal Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ENE RESOURCES + LINICAL RESOURCES +		CYP2C9 CYP2C9 CYDchrome P450, family 2, sull HGNC:2623 CYP2C10, "cytochrome P450, s P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DD23 RefSea:10M 000721 D CCDS:CCDS7437.1 C CGESCCDS7437.1 C	Hamily C, polypeptide 9 Wolfam Genediat Set Analyses: Construction C C C C C C C C C C C C C C C C C C C	The Human Ger GeneCards Gi GeneCards Gi Gene	uide Si ncDzeks (Previous F Syrr Proteine	WEL Terms Ute Terms Ute Terms Ute Terms Ute Cytochrome I annes: cytochrom bol approved by s: dDNA/INNAI E Expanse s: cytochrom bol approved by s: dDNA/INNAI E Expanse s: cytochrom pol approved by cytochrom	2722 274 274	درما الالاديم مان الالاديم مان المانيم مان المان مان المان مان المان المان مان المان مان المان المان مان المان مان المان المان مان المان المان مان المان المان المان مان المان المان المان المان مان المان الما	L other usery need a co Us Us Use Us Use Us Use Us Construction Use Construction Committee (netWold) Committee (netWold) Commit	Interced Recon r Feedback ide 9), polypeptide 9 defaultate Pretaine Pretaine Pretaine Pretaine Pretaine Pretaine Pretaine Pretaine e F-monoxygenas nonooxygenas nonooxygenas nonooxygenas e F-monoxyge e 7-monooxygenas e F-monoxyge e 7-monooxygenas interce Pretaine	Section Xonnex I Mirrow Search Add () () () () () () () () () () () () ()
Gene Symbol F YP2C9 upproved Symbol + upproved Name + (GIC ID + revious Symbols & Names + ynonyms + ocus Type + hromosomal Location + ENE FAMILY + OMOLOGS + UCLEOTIDE SEQUENCES + ENE RESOURCES + ENE RESOURCES + LINICAL RESOURCES +		CYP2C9 CYP2C9 CYDchrome P450, family 2, sull HGNC:2623 CYP2C10, "cytochrome P450, s P4501IC9 gene with protein product 10q24.1 Cytochrome P450 family C HCOP D TreeEam D GenBank:M61855 EMBL DD23 RefSea:10M 000721 D CCDS:CCDS7437.1 C CGESCCDS7437.1 C	family C, polypeptide 9 wbfam Genetical Bome Set Analyses: C C C C C C C C C C C C C	The Human Ge GeneCards GeneC	idde Sk mcDxc3S (Previous f	WEI IIII Free for sca uite Term cytochrome hames: cytochrom bol approved by scape of the scape scape of the scape scap	IZMANN I IZMANN I ademic non pro se and Condition P450, stat me P450, stat P26, stat (Pfred) P26 (Pfr	ecologiase), polyper roxylase), polyper Ensembl: ENSCO	L other users need a co Us Us Use V V V V V V V V V V V V V	Innarceal faceh r Feedback ide 9 proteins (all book of the second (all book of the second	se from Xonex, I Mirri Search Add)) () () () () () () () () () () () ()

An

Go back to Ensembl GeneView

Human (GRCh38) Location:									
Gene-based displays	Gene: CV	P2C9 ENSG0000	0138109						
 Summary Splice variants (3) 	oche. or	1 203 EN30000	0130103						
 Transcript comparison 	Description		cytoch	rome P450, family 2, subfan	ily C, polypep	tide 9 [Source	HGNC Symbol;	Acc:HGNC:2623]	
 Supporting evidence 	Synonyms		CYP20	10, P450IIC9					
– Gene alleles ⊟ Sequence	Location	*							
Secondary Structure	INSDC coordi	nator	chromosome:GRCh38:CM000672.2:94938658:94989390:1						
 External references 		nates							
 Regulation Comparative Genomics 	Transcripts		This ge	ne has 3 transcripts (splice	variants) Hid	e transcript ta	ble		
Genomic alignments									
 Gene tree 	Show/hide	columns (1 hidden)					Filter		
 Gene gain/loss tree Orthologues (168) 			bp 🕴 Prote	in 🕴 Biotype	CCDS 🔶	RefSeg 🝦		Flags	
- Paralogues (17)			bp Prote 1847 490		CCDS7437			-	v
 Ensembl protein families (1) 	CTP2C9-001	EINS10000200002	1047 490	aa Protein coding	<u>CCD31431</u>	NP 000762	TSL:1 GENCO	ODE basic APPRIS	S PI
 Phenotype Genetic Variation 	CYP2C9-002	ENST00000461906	1431 No pi	otein Processed transcrip	t -			TSL:1	
Series variation table		ENST00000473496		otein Processed transcrip		-			
- Variation image	CTF2C3-003	<u>EN310000413430</u>	041 140 pi	otein Processed transcrip		-		TSL:2	
- Structurar variation - External data	Summary								
Personal annotation	Statingary								
ID History	Name		CYP20	9 (HGNC Symbol)					
Gene history	CCDS			ne is a member of the Hum	n CCDS eat:	00087437			
🔅 Configure this page	UniprotKB			ne has proteins that corresp			identifiere: D4474	10	
Nodel serve data									
Add your data	RefSeq		Overla		matches and h	ias similar bio	type of protein_c	oding	
🛃 Export data	Ensembl vers			00000158109.5					
	GRCh37 asser	mbly		ne maps to <u>56,698,415,96,7</u>			tes.		
🕈 Bookmark this page				ID ENSG00000138109 not p	resent in GRO	h37.			
< Share this page	Gene type			protein coding					
- enale and page	Prediction Me	ethod	Annota	tion for this gene includes b	oth automatic	ann otation from	m Ensembl and <u>F</u>	Havana manual cura	ation, see <u>article</u> .
	Alternative ge	enes		ene corresponds to the fol		ase identifie	rs:		
			Havan	a gene: OTTHUMG000001	8805				
¢<⊞⊠%									
COSMIC 94.93	4b	94.94Mb							
			94.95Mb	94.96Mb	94.97Mb		94.98Mb	94.99Mb	95.00
			1	94.22 kb 94.96Mb			94.98Mb	94.99Mb	95.00
Genes			1	94.56Mb			94.98Mb	94.99Mb	
	•		1	<u>84 22 kb -</u> 94 96Mb			94.98Mb	94.99Mb	95.00 95.00
Senes Comprehensiv Genes			1	<u>84 22 kb -</u> 94 96Mb			94.56Mb	54.99%6	95.00 95.00
Senes Comprehensiv Genes			1	91.2215 - 91.966				54 5946	55.00
Genes Comprehensw Genes Comprehensw			1	34 5946				54 5940	55.00
Grees Gromprhensv Gromprhensv			1					54.5916	
Grees Gregorbensv Gregorbensv ENSTRO000260682 CP2C5401 KW IR. UK			1						
Groep Annay Comprehensy Extra construction of the second of the			1						
Groes Gorgenhensv Comprehensv Evertuopop2eose2 KV Tru, La KV			1						VVV VA AA NS SS GS IV IV AAT
Existing Comprehensiv Existing Crigo20002 Existing Crigo20001 Existing Crigo20001 Existing			1			SS ULU NS ULU INS ULU INS ULU INS ULU INS ULU INT INT INT INT			
Groep hensy Comprehensy Protocological and a second			1						MVV VAA ULU AVVV AAA NVV ULU AVVV ILU AVVV ILU AVVV ILU AVVV ILU AVVV ILU AVVV
Groep hensy Comprehensy Exerctocop2ce622 Exerctocop			1						MANINA SEE
Comprehensy Comprehensy Exercised and the second sec			1						
Groep hensy Comprehensy Exerctocop2ce622 Exerctocop			1						
Groep hensy Comprehensy Cropose hensy Cropose he			1						
Grand Stand			1						
Groep hensy Comprehensy Cropose hensy Cropose he			1						
Groep hensy Comprehensy Cropose hensy Cropose he			1						
Groep hensy Comprehensy Cropose hensy Cropose he			1						
Groep hensy Comprehensy Cropose hensy Crocose he			1						
Groep hensy Comprehensy Cropose hensy Crocose he			1						
Groep hensy Comprehensy Cropose hensy Crocose he			1						
Grand Stand	TI IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII		1						
Grand Stand			1						We way a set of a set
Grand Stand		Type	1		the deleted region	Control of the second sec			We way a set of a set
Grand Stand	A Construction of the second o	Type Transcript ablation Spice accord/or variant	1	Description A spice variant that chasp	the deleted region es the 2 base regions es the 2 base regions				With our and a second s
Grand Stand	Received and a second and a sec	Type Transcript ablation Spice donor variant	1	A second seco	the deleted region es the 2 base regions es the 2 base regions				With our and a second s
Grand Stand	A Construction of the second o	Type Transcript ablation Spice accord/or variant	1	Description A spice variant that chang A spice va	the deleted region es the 2 base regis	Control of the second sec	pt feature (sc.cocoust) mintone (sc.cocoust) minton	We to an	The shortened transcription
Grand Stand	B M	Type Transcript addation Spice acceptor variant Spice acceptor variant	1	Description A squence variant which main care and car	the deleted region es the 2 base region	A Constant of the translational	pt feature (sc.coc)(sc) minton (sc) minton (sc) minton (sc) minton (sc) minton (sc) minton (sc) minton (sc)	We the second se	Avail Net 0 available of the second
Grand Stands Comprehense	B M	Type Francuit altation Spice donor variant Spice acceptor variant Spice ac	1	Description A fasture ablation whereby A spice waint that chang A spice waint that cha	the deleted region es the 2 base regions as the 2 bas the 2 base regions as the 2 base r	includes a transcription of a costn state of a costn stat	Pf feiture (sc.2002122) in intron (sc.2002122)	mature stop codor, leading an elongated	Book and a second
Grand Stands Comprehense	Restartion of the second secon	Type Transcript addation Spice acceptor variant Spice acceptor variant	1	Description A squence variant which main care and car	the deleted region as the 2 base regions as the 2 base regions at least one base of a stat one base of a sta	Includes a transcerio an at the 3 end of a	Pf fature (c_coccustor) in inform (c_coccustor)	mature stop codor, leading an elongated	Book and a second
Grand Stands Comprehense	Image: Section of the sectio	Type Transcript ablation Splice acceptor variant Splice acceptor variant Stop Joal Hillator codon variant Transcript aplification	1	Description A faster abalance variant whech A spice variant that charg A spi	the deleted region of the set of	Includes a transcri and the Sea of a set and the Sea of a of the translations of the tran	Pf feature (c), c),c),c),c),c),c),c),c),c),c),c),c),c),c	mature stop codon, leading to the number of nucleotic transmission of the stop codon	Book and a second
Grand Stands Comprehense	A Constant of the second of th	Type Transcipt ablation Spice acceptor variant Stop gained Frameshift variant Transcipt amplification Inframe interaction Infr	1	Description A fasture ablation whereby A sequence variant where A sequence variant whe	the deleted regions end of the second	Includes a transcription of the translation of the	pt feature (sc. cost into the cost into the cost into the cost into the cost into the provide the cost into the cost into the cost into the cost into the provide the cost into the cost into the cost into the cost into the provide the cost into the cost into the cost into the provide the cost into the cost into the cost into the provide the cost into the cost into the cost into the provide the cost into the cost into the cost into the provide the cost into the cost into the cost into the cost into the provide the cost into the cost into the cost into the cost into the provide the cost into the cost into the cost into the cost into the provide the cost into the cost into the cost into the cost into the provide the cost into	mature stop coden, leading an elemented of the context of the code	Alexandree de la alexandre de la alexandr
Grand Stands Comprehense	Image: Section of the sectio	Fype Transcript ablation Splice donor variant Splice acceptor variant Splice a	1	Description A fastare addation whereby A spice wrainst that chang A fastare amplification of A inframe non synonym A spice wrainst that chang A inframe non synonym	the deleted region es the 2 base region es the 2 base region es the 2 base region at least one base of es at least one base of so at least one base of	includes a transcript of a code is characteristic interview of a code is characteristic is the code is the code is characteristic is the code is characteristic is the is the code is the code is characteristic is the code is the code is the code is the is the code is the code is the code is the is the code is the code is the code is the is the code is the code is the code is the is the code is the code is the code is the is the code is the code is the code is the code is the is the code is the code is the code is the code is the is the code is the code is the code is the code is the is the code is the code is the code is the code is the is the code is the code is the code is the code is the is the code is the is the code is the c	pt feature (c).20011800 minimum citizeness provide and the second provide a	mature stop codon, leading and inclusion of the second sec	Piler
Grand Stands Comprehense	A Constant of the second of th	Type Transcipt ablation Spice acceptor variant Stop gained Frameshift variant Transcipt amplification Inframe interaction Infr	1	Description A fasture ablation whereby A sequence variant where A sequence variant whe	the deleted region es the 2 base region es the 2 base region es the 2 base region at least one base of es at least one base of so at least one base of	includes a transcript of a code is characteristic interview of a code is characteristic is the code is the code is characteristic is the code is characteristic is the is the code is the code is characteristic is the code is the code is the code is the is the code is the code is the code is the is the code is the code is the code is the is the code is the code is the code is the is the code is the code is the code is the is the code is the code is the code is the code is the is the code is the code is the code is the code is the is the code is the code is the code is the code is the is the code is the code is the code is the code is the is the code is the code is the code is the code is the is the code is the is the code is the c	pt feature (c).2001150) mintron (c).2001150	mature stop codon, leading and inclusion of the second sec	Piler
Grand Stands Comprehense	A Carl A and	Type Transcript and tion Splice across variant Transcript anglification Inframe insection In		Description A fasture ablation of three anguards A sequence variant the three	the deleted region of the second seco	Includes a transcription of the spaces from the space f	P fature (c.cocosta) n intron (c.cocosta)	mature stop codor, leading at the number of nucleotic resulting an elongated to the stop codor. The stop codor is an elongated to the stop cod	Pier Pier
Grand Stands Comprehense	8 0 - 1 9 0 - 1 1000 Horse 9 0 - 1 1000 Horse 1000 Horse 9 0 - 1 1000 Horse 1000 Horse 1000 Horse 9 0 - 1000 Horse	Type Transcript addation Splice donor variant Splice acceptor variant Splice acceptor variant Splice acceptor variant Splice acceptor variant Stop last Indiatro codon variant Transcript addation Infarme deletion Infarme fueletion Infarme deletion Infarme deletion Infa		Description A feature ablation whereby A spice variant that chan	the deleted region es the 2 base region es the 2 base region es the 2 base region as the 2 base region as the 2 base region as the 2 base region at least one base of a least one base of the sin an estillation as variant that dires has of a least one base of the is an estillation of the sin and the is an estillation of the sin and the sin and the sin and the is an estillation of the sin and the s	includes a transcript of a coden school of a cod	Pf feiture (cc.constant) re constant) re define the feiture (cc.constant) re constant) re define the feiture (cc.constant) re re r	mature stop coden, leading an elongated the sumber of nucleotic resulting in an elongated the store within 13 bases of adequarts pictures pictur	The second secon
Grand Stands Comprehense	1 Show Show 1 Show Show	Type Transcript ablation Spice acceptor variant Stop Ig1 Initiator codon variant Spice region variant Spice region variant Spice region variant Spice terminal codor variant Spice terminal codor variant Spice terminal codor variant		Description A fasture abalance variant where A sequence varia	the deleted region of the second seco	Includes a transcer Includes	Pf flature (c), coordination of the solution of the solution	mature stop coden, leading an elongated the sumber of nucleotic resulting in an elongated the store within 13 bases of adequarts pictures pictur	The second secon
Grand Stands Comprehense	8 0 - 1 9 0 - 1 1000 Horse 9 0 - 1 1000 Horse 1000 Horse 9 0 - 1 1000 Horse 1000 Horse 1000 Horse 9 0 - 1000 Horse	Type Transcript addation Splice donor variant Splice acceptor variant Splice acceptor variant Splice acceptor variant Splice acceptor variant Stop last Indiatro codon variant Transcript addation Infarme deletion Infarme fueletion Infarme deletion Infarme deletion Infa		Description A fasture ablation whereby A spice variant that charg A spice variant that charge A spice variant that t	the deleted region es the 2 base regio at least one base is at least one base at a stanta that lines at least one base	includes a transcript events of a coon is char of a coon is char of a coon is char of a coon is char of the translations the transcript (coop) to base for the translations of a coon is char of the translations of the translations of the translations of the transcript (coop) to base for the house, is reality in the bases, resulting in curved within the re-	pt feature (2, 002120) metric and service (2, 002120) metric	mature stop coden, leading an elongated the sumber of nucleotic resulting in an elongated the store within 13 bases of adequarts pictures pictur	The second secon
Grand Stands Comprehense	A Constant of the second of th	Type Type <t< td=""><td></td><td>Description A fasture abalance variant where A sequence varia</td><td>the deleted region as the 2 base region of the second second second second second set at least one base the second second</td><td>includes a transcript events of a coon is char of a coon is char of a coon is char of a coon is char of the translations the transcript (coop) to base for the translations of a coon is char of the translations of the translations of the translations of the transcript (coop) to base for the house, is reality in the bases, resulting in curved within the re-</td><td>pt feature (2, 002120) metric and service (2, 002120) metric</td><td>mature stop coden, leading an elongated the sumber of nucleotic resulting in an elongated the store within 13 bases of adequarts pictures pictur</td><td>The second secon</td></t<>		Description A fasture abalance variant where A sequence varia	the deleted region as the 2 base region of the second second second second second set at least one base the second	includes a transcript events of a coon is char of a coon is char of a coon is char of a coon is char of the translations the transcript (coop) to base for the translations of a coon is char of the translations of the translations of the translations of the transcript (coop) to base for the house, is reality in the bases, resulting in curved within the re-	pt feature (2, 002120) metric and service (2, 002120) metric	mature stop coden, leading an elongated the sumber of nucleotic resulting in an elongated the store within 13 bases of adequarts pictures pictur	The second secon
Grand Stands Comprehense	Results of the second s	Type Transcript ablation Splee donor variant Splee donor variant Splee donor variant Splee donor variant Stop lost Initiator codon variant Transcript ablation Splee donor variant Stop lost Initiator codon variant Transcript aplication Inframe deletion Missense variant Splice region variant Stop retained variant Stope retained variant Stope retained variant Stop retained variant Sprime UTR variant Sprime UTR variant	variant	Description A fasture addation whereby A spice wrant that chan A spice variant which A spice variant	the deleted region estimates and the second	includes a transcript of the translation of a code site of a code site of a site of a code site of a code site of the site of a code site of a code site of the site of a code site of a code site of the site of a code site of a code site of the site of a code site of a code site of the site of a code site of a code site of the site of a code site of the site of a code site of the site of a code site of a code site of the site of the site of a code site of the site of a code site of the site of a code site of the site of a code site of the site of a code site of the site of a code site of the site of a code site of the site of the site of a code site of the si	pf feature (c_cocol1202) monormal with the feature of the feature (c_cocol1202) monormal feature of the feature of the feature (c_cocol1202) monormal feature monormal feature of the feature (c_cocol1202) monormal feature monormal	mature stop codon, leading are shown in the store of the	The second secon
Grand Stands Comprehense	Image: Section of the sectio	Type Transcript ablation Spice acceptor variant Stop lost Initiator codon variant Spice region variant Spine UTR variant	variant	Description A sequence variant that charge A sequence variant that that A sequence variant that that A sequence variant that that A sequence variant that	the deleted region of the second seco	includes a transceri to the second se	pf feature (c_cocol1202) monormal with the feature of the feature (c_cocol1202) monormal feature of the feature of the feature (c_cocol1202) monormal feature monormal feature of the feature (c_cocol1202) monormal feature monormal	mature stop codon, leading are shown in the store of the	The second secon
Grand Stands Comprehense	8 0	Type Transcript addaton Spice acceptor variant Spire UTR variant Spirme UTR variant	variant	Description A fasture ablation where by Region and the constraint of the constraint the constraint of the constraint the constraint of the constraint the constra	the deleted region es the 2 base region es the 2 ba	includes a transcript of a coden is characteristic and the 5 ⁻ end of a or a coden is characteristic to base in the the second of the translations of the mature mile of the mature mile approximation of the translations of the translations of the translations	pf feature (c_cocol1202) monormal with the feature of the feature (c_cocol1202) monormal feature of the feature of the feature (c_cocol1202) monormal feature monormal feature of the feature (c_cocol1202) monormal feature monormal	mature stop codon, leading are shown in the store of the	The second secon
Grand Stands Comprehense	Image: Section of the sectio	Transcript availant Spice donor variant Spice acceptor variant <td>variant</td> <td>Description A faster abalance synaptice A spice variant that char A spice variant that char</td> <td>the deleted region of the deleted region as the 2 base region as the 2 b</td> <td>Includes a transcri where the second second</td> <td>pf feature (c_cocol1202) monormal with the feature of the feature (c_cocol1202) monormal feature of the feature of the feature (c_cocol1202) monormal feature monormal feature of the feature (c_cocol1202) monormal feature monormal feature monormal</td> <td>mature stop codon, leading are shown in the store of the</td> <td>The second secon</td>	variant	Description A faster abalance synaptice A spice variant that char	the deleted region of the deleted region as the 2 base region as the 2 b	Includes a transcri where the second	pf feature (c_cocol1202) monormal with the feature of the feature (c_cocol1202) monormal feature of the feature of the feature (c_cocol1202) monormal feature monormal feature of the feature (c_cocol1202) monormal feature monormal	mature stop codon, leading are shown in the store of the	The second secon
Grand Stands Comprehense	Image: Section of the sectio	Type Transcript addaton Spice acceptor variant Spire UTR variant Spirme UTR variant	variant	Description A fasture ablation where by Region and the constraint of the constraint the constraint of the constraint the constraint of the constraint the constra	the deleted region of the second seco	Includes a transcription of the second secon	pf feature (c_cocol1202) monormal with the feature of the feature (c_cocol1202) monormal feature of the feature of the feature (c_cocol1202) monormal feature monormal feature of the feature (c_cocol1202) monormal feature monormal	mature stop codon, leading are shown in the store of the	The second secon

The Open Door Workshop

Module 6: Genomic Variation

Select "Transcript" Ensembl Blast/Blat | Biom 🚮 • Se Transcript: CY 2C9-001 me P450, family 2, Syn Sh ***** CCDS Flags NCODE b Protein + Biotype 490 aa Protein coding NM_000771 NP_000762 CYP2C9-002 CYP2C9-003 TSL:1 TSL:2 841 No protein Summary 0 E ID History Transcript history Configure this page ł Follow "Variations"

Variations 0

Show All	entries	Sho	w/hide columns					Filter	
Residue	Variation ID	О Туре	Evidence	Alleles	Ambig. code	Residues	Codons	SIFT	PolyPhen
1	rs114071557	Initiator codon variant	💧 🕷 🕒	A/G	R	M, V	ATG, GTG	0	0
1	COSM106812	Initiator codon variant		G/A	R	M, I	ATG, ATA	0.02	0
1	rs150891702	Initiator codon variant		G/A	R	M, I	ATG, ATA	0.02	0
2	rs139414138	Missense variant		G/A	R	D, N	GAT, AAT	0.08	0
2	COSM107499	Missense variant		G/A	R	D, N	GAT, AAT	0.08	0
5	rs138957855	Missense variant	📥 3K 🕒	T/C	Y	V, A	GTG, GCG	0.1	0
5	rs267602638	Synonymous variant		G/A	R	v	GTG, GTA		
359	<u>rs1057<mark>910</mark></u>	Missense variant	a 6 9 % I 🖓	A/C	м	I, L	ATT, CTT	0.04	0.45
359 3 59	rs1057 <mark>910</mark> 0.0005032	Missense variant Toosing Sequence variant PT earch on mattern	.	A/C HOME_INSTATION	м	l, L	ΑΤΤ, ΟΤΤ	0.04	0.45
359 3 55 359			C C 31 (C)		M -	I, L	ATT, CTT	0.04	0.45
399 — — —	<u></u>	- Cooling-Sequence variant Pressure-etonyalism		HOMO_NOTATION					
3 59 — — —	<u>CM960481</u> CM960481	Coding sequence variant Feature elongation		HGMD_MUTATION					
359 360	CM960481 rs28371686	Coding sequence variant Feature elongation Coding sequence variant Feature elongation Missense variant		HGMD_NUTATION C/G	- S	- D, E	GAC, GAG		- 0.928
359 — — — 359 360 360	CM909962 CM960481 rs28371686 CM014176	Coding Supported variant [Feature elongation [Coding sequence variant] Feature elongation [Missense variant [Coding sequence variant] Feature elongation		HGMD_MUTATION C/G HGMD_MUTATION	- S -	- D, E -	- GAC, GAG -	- - -	- 0.928 -
359 359 360 360 360	CM905452 CM960481 rs28371686 CM014176 CM090527	Coding sequence variant Feature elongation Missense variant Coding sequence variant	.	HGMD_INDIANON HGMD_MUTATION C/G HGMD_MUTATION HGMD_MUTATION	- S -	- D, E -	- GAC, GAG -		- 0.928 - -

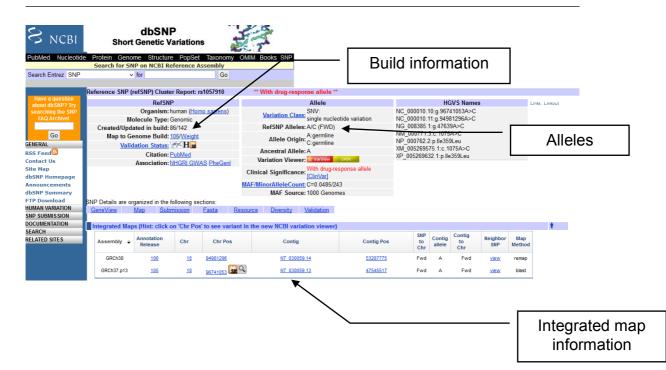
Follow rs1057910

rs1057910 SNP

Original source Alleles Location Co-located Validation status Clinical significance Synonyms I HGVS names I Genotyping chips I

Variants	(including SNPs and indels) i	mported from dbSNP (rele	ease 137) <u>View in</u>
Referen	e/Alternative: A/C Ancestral:	A Ambiguity code: M M/	AF: 0.04 (C)
Chromo	some 10:96741053 (forward s	trand) <u>View in location ta</u>	ab
with HGI	ID-PUBLIC CM960481		/
This vari	ation is validated by 1000 Gen	omes, HapMap and also	frequency
pathoge	nic (from dbSNP) <u>View explar</u>	nation	/
This fea	ure has <mark>6</mark> synonyms - click the	e plus to show	/
This fea	ure has <mark>3</mark> HGVS names - click	the plus to show	/
	ation has assays on 8 chips -	click the plus to show	1

Follow link to dbSNP



Scrolling down the page reveals information about the allele frequency and a link through to genotypes.

					/		со	Seq ntex	uen t of	ice SN	۱P	
NCBI Assay ID	Handle Submitter ID	Validation Status	ss to rs Orientation /Strand	Alleles	5' Near Seq 30 bp	3' Near Seq 30 bp	Entry Date	Update Bui Date Add		Freq Warning		ISuccess Rate
ss1538933	LEE 741019		fwd/T	A/C	gatgetgtggtgcacgaggtccagagatac	ttgacettetecceaceageetgeecea	tg 09/13/00	0 10/10/0386	cDNA			unknown
ss2419886	HGBASE SNP000000187		fwd/T	A/C	tgtggtgcacgaggtccagagatac	ttgaccttctccccaccagcctgcc	11/07/00	0 10/10/0389	Genomic			unknown
ss4426472	LEE e741019		fwd/T	A/C	gatgctgtggtgcacgaggtccagagatac	ttgaccttctccccaccagcctgcccca	t.g 04/26/02	210/10/03106	cDNA			unknown
ss5586419	SNP500CANCERICYP2C9-01	K	fwd/T	A/C	gatgctgtggtgcacgaggtccagagatac	ttgaccttctccccaccagcctgcccca	t.g 09/26/02	2 04/07/04 113	Genomic			unknown
ss12588583	EGP_SNPSICYP2C9-045324	\varkappa	fwd/T	A/C	gatgctgtggtgcacgaggtccagagatac	ttgaccttctccccaccagcctgcccca	tg08/20/03	3 04/07/04 119	Genomic			unknown
ss28501344	IMCJ-GDTJMCJ-CYP2C9_9-AC		fwd/T	A/C	gatgctgtggtgcacgaggtccagagatac	ttgaccttctccccaccagcctgcccca	tg 08/20/0 4	4 08/20/04 126	Genomic			unknown

		Genotype a			Genot	ype Det	ail NEW	All	leles
ss#	Population	allele informa		rom. Die Cnt. ^{Sou}	urce A/A	A/C	HWP	Α	С
<u>ss105107895</u>	PA152209538		184	AF				0.918	0.082
ss105108091	PA152211301		696	AF			\backslash	0.953	0.047
<u>ss105109763</u>	PA154394460		584	AF		1	×	0.938	0.062
ss12588583	PDR90	Global	176	IG	0.920	0.080	0.752	0.960	0.040

Scroll up and go to Gene (ID)

		RefSeqGene Mapping		
RefSeqGene	Gene (ID)	SNP to RefSeqGene	Position	Allele
<u>NG_008385.1</u>	CYP2C9 (1559)	Fwd	<u>47639</u>	A

CYP2C9 cytochrome P450, family 2, subfamily C, polypeptide 9 [Homo sapiens (human)]

Gene ID: 1559, updated o	n 7-Dec-2014		
Summary			٦) (٩)
Official Full Name Primary source Locus tag See related Gene type Ref Seq status Organism Lineage Also known as	CYP2C9 provided by <u>HONC</u> cytochrome P450, family 2, subfamily C, polypeptide 9 provided by <u>HONC</u> <u>HONC-HONC-2623</u> RP11-208C176 7 Reviewed by <u>HONC 2003</u> MIM: 601130, teal 2000 Protein coding REVIEWED <u>Homo sagatems</u> Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi, Mammalia, Eutheni CPC9; CYP2C; CYP2C10; CYP10C9; P450IC9 This gene encodes a member of the cytochrome P450 superfamily of enzymes. The cy of cholesterol, steroids and other lipids. This protein localizes to the endoplasmic refectu phenytoin, tolbutamide, ibuprofen and Swarfain. Studies identifying individuals who are cluster of cytochrome P450 genes on chromosome 10q24. [provided by RefSeq, Jul 201	a; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Ho tochrome P450 proteins are monooxygenases which catalyze many lum and its expression is induced by rifampin. The enzyme is know poor metabolizers of phenytoin and tolbutamide suggest that this g	reactions involved in drug metabolism and synthesis to metabolize many xenobiotics, including
*601130 CYTOCHROME	P450, SUBFAMILY IIC, POLYPEPTIDE 9; CYP2C9		
HGNC Approved Gene S	iymbol: CYP2C9		

Cytogenetic location: 10q23.33 Genomic coordinates (GRCh37): 10:96,698,349-96,749,485 (mm NCB)

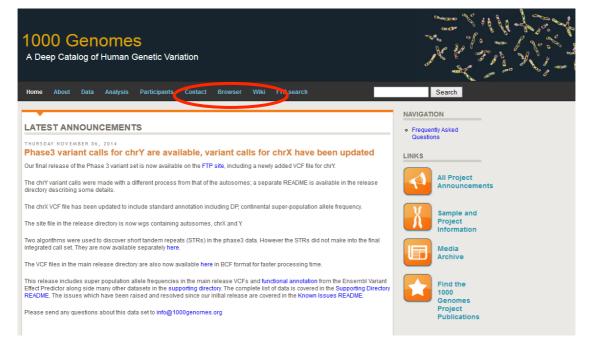
Gene-Phenotype Relationships

Location	Phenotype	Phenotype MIM number	Phenotype mapping key
10q23.33	Tolbutamide poor metabolizer		3
	Warfarin sensitivity	122700	3

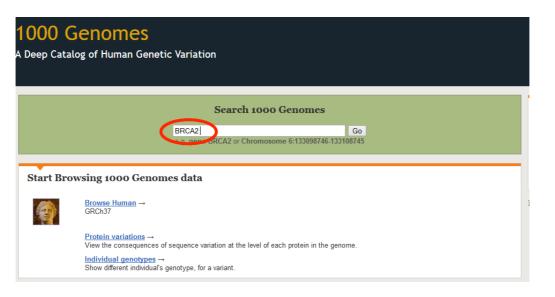
The 1000 Genomes Project

Data from increasing numbers of human genome sequences is currently available through the public website (www.1000genomes.org).

These can be searched through the familiar Ensembl browser format for displaying data (browser.1000genomes.org)



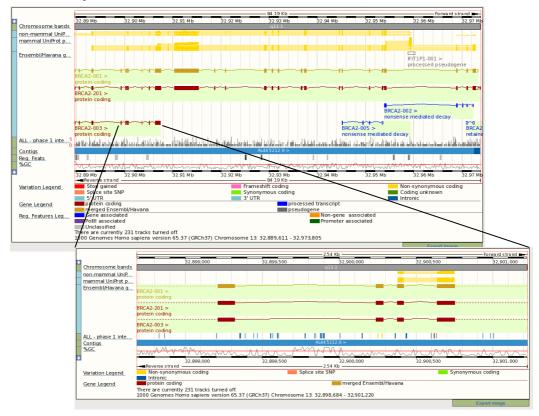
1000 Genomes A Deep Catalog of Human Genetic Variation	A Contraction
Home About Data Analysis Participants Contact Browser Wiki FTP search	Search
Home >	NAVIGATION
1000 GENOMES BROWSERS	 Frequently Asked Questions
The novo genomes project uses a project specific version of the Ensembl Browser to visualise its variants. This browser can be found at http://browser.1000genomes.org .	
The main browser currently displays the snps and indels from the Integrated Phase 1 Release. This runs Ensembl version 63.	
There is also a version of the browser which holds the pilot data from A map of human genome variation from population-scale sequencing , Nature 467 ,1061.1073 at http://pilotbrowser.1000genomes.org. This uses Ensembl version 60.	
You can find instructions for how to use our browser here in doc format. There is also a tutorial for the pilot browser here	
You can also access the databases which sit behind our browser, for more details about this please look at the Public Mysql Instance page.	
There are also more tips for ensembl both on the main ensembl site and their blog	
The NCBI also provide a 1000 genomes browser hosted on their site	



The project uses the familiar Ensembl browser format:

Search 1000 Genomes	Results Summary
- Results Summary	You searched for 'BRCA2 '
	Gene or Gene Product
Add your data	18 entrie(s) matched your search strings.
Export data	1. Gene: ENSG00000170037 Region in detail [17:7835419-7853236] centrobin, centrosomal BRCA2 interacting protein [Source:HGNC Symbol;Acc:29616]
👌 Get VCF data	2. Variations in gene ENSG00000170037: Variations in gene [17:7835419-7853236]
Bookmark this page	3 Gene: ENSG00000107949 Region in detail [10:127512115-127542264] BRCA2 and CDKN1A interacting protein [Source:HGNC Symbol;Acc:978]
Share this page	4. Variations in gene ENSG00000107949: Variations in gene [10:127512115-127542264]
View in Ensembl	5 Gene: ENSG00000185515 Region in detail [X:154299695-154351349] BRCA1/BRCA2-containing complex, subunit 3 [Source:HGNC Symbol;Acc:24185]
	6. Variations in gene ENSG00000185515: Variations in gene [X:154299695-154351349]
	7. Gene: ENSG0000083093 Region in detail [16:23614488-23652631] partner and localizer of BRCA2 [Source:HGNC Symbol;Acc:26144]
	8. Variations in gene ENS600000083093: Variations in gene [16:23614488-23652631]
	9. Gene (from Patch 2013-07-15 14:43:48): ENSG0000269884 Region in detail [HG1497_PATCH:154239888-154291542] BRCA1/BRCA2-containing complex, subunit 3 [Source:HGNC Symbol;Acc:24185]
	10. Variations in gene ENS600000269884 (from Patch 2013-07-15 14:43:48): Variations in gene [HG1497_PATCH:154239888-154291542]
	11. Gene: LRG 308 Region in detail [LRG 308:5001.43196] partner and localizer of BRCA2 [Source:HGNC Symbol;Acc:26144]
	12. Variations in gene LRG_308: Variations in gener [1.13] 200-5001-43196]
	13. Gene: ENSG00000139618 Region in detail 13:32889611-32973805) BRCA2 - breast cancer 2, early onset [Source HGNC Symbol:Accent01]

Make sure you select BRCA2 - breast cancer 2, early onset ENSG00000139618



Choosing SNPs to Genotype

Once you have identified the SNPs associated with your gene(s) of interest it is time to get some data. However, as we have seen there are typically dozens of SNPs in or close to any given gene. So how do you decide which ones to look at? There are no set rules for this, but you should take into consideration:

- 1) Previously published data
- 2) Validation status
- 3) Population frequency
- 4) In silico predictions of potential phenotypic effect
- 5) Haplotypes

Previously published data

The first place to look for SNPs is a literature database such as PubMed (<u>http://www.ncbi.nlm.nih.gov/pubmed/</u>) using the appropriate search terms (eg 'polymorphism' AND 'CYP17A1'). If somebody has already found a functional SNP, or linked a SNP to a disease then it is an excellent candidate for your study.

The advent of microarrays means that some groups are investigating associations between SNPs and gene expression levels on a large scale, so called expression Quantitative Trait Loci (eQTL). For example the GENe Expression VARiation (GENEVAR) project at the Sanger has looked at the expression of 48,000 genes in all 270 HapMap lymphoblastoid cell lines. A searchable database is under construction to allow you to see what SNPs (if any) are associated with expression of your gene(s) of interest. Data is currently available at: http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/

Genome-wide association studies (GWAS) are becoming increasingly common. The US National Human Genome Research Institute maintains a searchable catalogue of GWAS (<u>http://www.genome.gov/gwastudies/</u>). If you

know what region or disease you are interested in you can see if any SNPs have already been linked to it:

Division of Genomic Medicine		🛨 Share 📇 Prin
A Catalog of Published Genome-Wi	de Association Studies	
	iomic Medicine Activities GWAS Catalog Meetings & Workshops cing Programs Publications Trans-NIH Sequencing Inventory	
Additional information has been added to th	e HTML catalog columns below. For a description of column heading	js for the HTML catalog, go to: Catalog Heading Descriptions 🐲 🦚
	is of genome-wide association loci for human diseases and traits is Academy of Science <mark>s (PNAS) article on catalog methods and an</mark> alys	s .
View the Full Catalog and Download	the Catalog and Search the Catalog and	
Search By:		
Journal:	Select Journal	
First Author: (last name)		
Disease/Trait:	(string search)	
	or	
	Attention deficit hyperactivity disorder symptoms (interaction) Autism Basal cell carcinoma (cutaneous) Behcat's disease Beta thalassemia/hemoglobin E disease Bilirubin levels Biochemical measures Biochemical quantitative traits Bipolar disorder Black vs. blond hair color (hold CtT-key when selecting multiple entries)	
Chromosomal Region: (e.g., "13q21.31")	(noid car-key when selecting induple entries)	Type in OXTR
Gene: (e.g., "LRP5")	LRP5	
SNP: (e.g., "rs20755555")		
OR greater than:		
p-Value threshold: Enter the exponent. For example, enter "5" for p<10 ⁻⁵		
Se	Clear Query	

The search returns details of the study and SNPs identified:

Date Added to Catalog (since 11/25/08)	First Author/Date/ Journal/Study	Disease/Trait	Initial Sample Size	Replication Sample Size	Region	Reported Gene(s)	Mapped Gene (s)	Strongest SNP- Risk Allele	Context	Risk Allele Frequency in Controls	P-value	OR or beta- coefficient and [95% CI]	Platform [SNPs passing QC]	CNV
12/14/11	Edwards AC November 05, 2011 Psychiatr Genet Genome-wide association study of comorbid depressive syndrome and alcohol dependence.	Depression and alcohol dependence	467 European ancestry cases, 407 European ancestry controls		3p25.3	OXTR	OXTR	<u>r#237899-?</u>	intron	NR	<u>2 x 10⁻⁶</u>	1.69 [1.36-2.10]	Illumina [876,476]	N

Validation status

It is important to remember that not all SNPs in dbSNP are useful or even real. Some are sequencing errors and others may be unique to the individual they were found in. Looking at the validation status of a SNP gives an idea of how reliable it is. Unvalidated SNPs should be treated with caution. SNPs with frequency information available are best.

	Validation status description
8	Validated by multiple, independent submissions to the refSNP cluster
K	Validated by frequency or genotype data: minor alleles observed in at least two chromosomes.
	Validated by submitter confirmation
	All alleles have been observed in at least two chromosomes apiece
Η	Genotyped by HapMap project
<u>ik</u>	SNP has been sequenced in 1000Genome project.

Population frequency

The population frequency of a SNP should match the study that you are conducting. How many samples do you have? Do you have the power to detect associations in SNPs with a frequency of 5%, 10% etc? If not then there's little point genotyping SNPs that have such low frequencies.

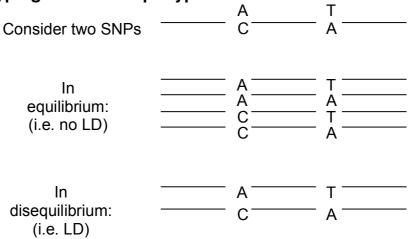
In silico predictions of potential phenotypic effect

SNPs are more likely to have phenotypic effects if they are:

- 1) Frameshift
- 2) Non-synonymous
- 3) Synonymous (exonic splice enhancer/suppressor)
- 4) Splice site
- 5) Untranslated region (UTR)
- 6) In regulatory regions

Expressed SNPs are easy to identify as shown previously in this module, but determining which of these to prioritize is not always obvious.

Genotyping data and haplotypes



Linkage and Linkage Disequilibrium (LD)

Linkage and linkage-disequilibrium (LD) measure a correlation, cosegregation or association between a genetic marker and disease. They can be distinguished a number of ways:

1. Linkage is focused on a locus whilst LD is focused on an allele

2. Linkage results from recombination events in the last 2-3 generations. LD on the other hand results from much earlier, ancestral recombination events

3. Linkage measures co-segregation in a pedigree. LD measures cosegregation in a population (essentially a very large pedigree)

4. From the dynamical system point of view, Linkage is the "dynamical equation", LD is the "initial condition"

5. In a pedigree likelihood calculation (LOD score), the result tells you whether you have Linkage or not. Conversely, LD is provided by the user prior to performing the calculation.

 Linkage is usually detected for markers reasonable close to the disease gene (one centiMorgan). LD is detected for markers even closer (0.01-0.02 cM). Several metrics have been devised to measure linkage disequilibrium (LD). The two most commonly used of these are D' and r^2 . Both are related to the basic unit of LD, D.

D

D measures the deviation of haplotype frequencies from the equilibrium state. LD occurs when *D* is significantly greater than zero. Consider two linked SNPs with alleles (*A*, *a*) and (*B*, *b*), resulting in four possible haplotypes: *AB*, *Ab*, *aB* and *ab*. *D* can be calculated as in equation 1, where f(X) represents the frequency of the X allele or haplotype.

$$D=f(AB)-f(A)f(B) \tag{1}$$

D'

D' is the absolute ratio of *D* compared with its maximum value, D_{max} , when $D \ge 0$, or compared with its minimal value, D_{min} , when D < 0. *D'*=1 denotes complete LD, and historical recombination results in the decay of *D'* towards zero.

r²

 r^2 is the statistical coefficient of determination – a measurement of correlation between a pair of variables (see equation 2).

$$r^{2} = \frac{D^{2}}{f(A) f(a)f(B) f(b)}$$
(2)

 r^2 is of particular importance in genetic mapping as it is inversely related to the required sample size for association mapping, given a fixed genetic effect. For example, if only one pair of SNPs was genotyped and r^2 between the SNPs was found to be 0.5, then to provide the same statistical power for ungenotyped SNP compared with the case where $r^2=1$, knowing the genotypes of alleles of one SNP is directly predictive of the genotypes of another SNP. The alternative notation R^2 is used when individual variables are predicted using the multiple regression of a constellation of other variables.

Relationship between D' and r^2

D' and r^2 can be written in terms of each other and allele frequencies. Without losing generality, the four alleles can be chosen such that $D \ge 0$ and $f(A) \ge f(B)$. So *D'* and D_{max} have the relations in equations 3 and 4.

$$D'= \underbrace{D}_{D_{max}}$$
(3)

$$D_{max} = f(a)f(B)$$
(4)
and

$$r^{2} = (D')^{2} \times \underbrace{f(a)f(B)}_{f(A)f(b)}$$
(5)

Equation 5 shows the relationship between D', r^2 and allele frequencies. As $f(A) \ge f(B)$, r^2 has the upper bound of $(D')^2$, and reaches it only when f(A) = f(B). The implication of this is that D', a commonly used measure of historical recombination, provides information on the physical extent of useful LD (in terms of association mapping and statistical power) by providing the upper limit of r^2 . Dense LD maps that are based on high frequency SNPs (MAF>0.1) can reveal regions of historical recombination. Knowing the level of D' decay in these maps directly provides the maximum potential level of useful LD in association mapping (based on r^2) for high-frequency SNPs even if a significant proportion of common SNPs remains undiscovered. For example, if a recombination point resulted in a D' of 0.7 for SNPs on either side of it, the maximum possible r^2 for these SNPs would be 0.49, and sample sizes would need to be more than doubled to maintain the same statistical power for association mapping. It should be noted that both D' and r^2 suffer from sampling biases given a small number of individuals and for rare variants. Confidence intervals for D' have been used by some investigators. LD varies throughout the human genome. Near complete LD has been observed over >800 Kb on chromosome 22, however, these regions of near complete LD are interspersed with regions of little or no LD.

There are a number of web based programs available for determining linkage disequilibrium such as Arlequin (<u>http://cmpg.unibe.ch/software/arlequin3/</u>)

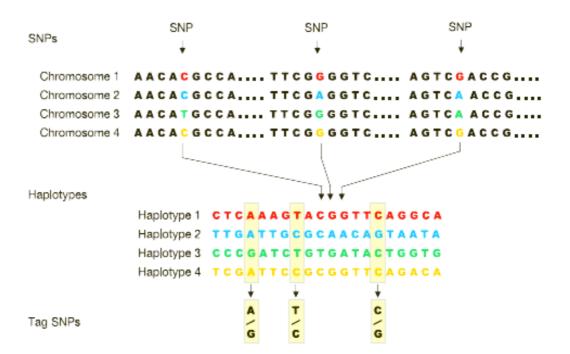
Genepop (<u>http://genepop.curtin.edu.au/</u>) and Haploview (www.broad.mit.edu/personal/jcbarret/**haploview**/).

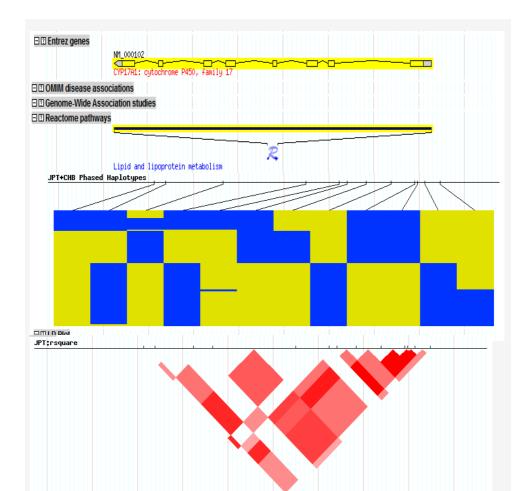
Haplotypes

While individual SNPs can have an effect on gene expression or protein structure they do not exist in isolation, but as part of haplotypes. Haplotypes are blocks of sequence that derive from the same ancestral chromosome and have not been disrupted by recombination. Haplotypes are defined by groups of closely linked alleles that tend to be inherited together. Thus each SNP investigated is linked to and interacts with a number of other SNPs. As high throughput genotyping platforms become more widely used, attention will inevitably move from individual SNPs to haplotypes.

Taking Haplotype block structure into account

- Discrete chromosome region of high LD and low haplotype diversity
- All pairs of polymorphisms within a block are in strong LD, whereas other pairs show weaker association
- Blocks hypothesized to be regions of low recombination flanked by recombination hotspots: In other words, SNPs in blocks have had little ancestral recombination happen between them.





_	-10002702	-17115100		27/0207	1004467	=755442	4010607		202/755	rs10786712	rs6163	rs6162	rs762563	742572	rs2486758	17115140
										104586386						
NA18545 c		G	A	G	A	G	A	G	G	T	A	A	G	G	T	T
NA18608 c		G	A	G	A	G	A	G	G	T	A	A	G	G	T	Ť
NA18611 c		G	A	G	A	G	A	G	G	T	A	A	G	G	T	T
NA18537 c		G	A	G	A	G	A	G	G	T	A	A	G	G	T	T
NA18526 c:	A	G	A	G	Α	G	A	G	G	Т	А	A	G	G	Т	Т
NA18532 c:	A	G	Α	G	Α	G	A	G	G	Т	А	A	G	G	Т	Т
NA18974_c:	1 A	G	A	G	A	G	A	G	G	Т	А	A	G	G	Т	Т
NA18953_c	2 A	G	Α	G	A	G	A	G	G	Т	А	Α	G	G	Т	Т
NA18945_c	2 A	G	Α	G	Α	G	Α	G	G	Т	А	Α	G	G	Т	Т
NA18943_c	2 A	G	Α	G	Α	G	Α	G	G	Т	Α	Α	G	G	Т	Т
NA18635_c	2 A	G	A	G	A	G	A	G	G	Т	A	A	G	G	Т	Т
NA18632_c	2 A	G	Α	G	A	G	A	G	G	Т	Α	A	G	G	Т	T
NA18540_c	2 A	G	С	С	A	G	A	G	G	Т	Α	A	G	G	Т	G
NA18555_c:	1 A	G	С	С	A	G	A	G	G	Т	Α	A	G	G	Т	G
NA18570_c:	1 A	G	С	С	A	G	A	G	G	Т	A	A	G	G	Т	G
NA18571_c:	1 A	G	С	С	A	G	A	G	G	Т	Α	A	G	G	Т	G
NA18572_c	2 A	G	С	С	A	G	A	G	G	Т	Α	A	G	G	Т	G
NA18593_c	2 A	G	С	С	A	G	A	G	G	Т	Α	A	G	G	T	G
NA18594_c:	1 A	G	С	С	A	G	A	G	G	Т	Α	Α	G	G	T	G
NA18603_c	2 A	G	С	С	A	G	Α	G	G	Т	A	A	G	G	Т	G
NA18609_c:	1 A	G	С	С	A	G	A	G	G	Т	A	A	G	G	Т	G
NA18947_c:	1 A	G	С	С	A	G	A	G	G	Т	A	A	G	G	Т	G
NA18960_c:		G	С	С	A	G	A	G	G	Т	A	A	G	G	Т	G
NA18526_c		G	С	С	A	G	A	G	G	Т	A	A	G	G	Т	G
NA18995_c:		G	С	С	A	G	A	G	G	Т	A	A	G	G	Т	G
NA18994_c:		G	С	С	A	G	A	G	G	Т	A	A	G	G	Т	G
NA18976_c		G	С	С	A	G	A	G	G	Т	A	A	G	G	Т	G
NA18965_c		G	С	С	A	G	A	G	G	Т	A	A	G	G	Т	G
NA18577_c:		G	A	С	A	A	A	G	G	Т	A	A	G	G	Т	G
NA18973_c		G	A	С	A	A	A	G	G	Т	A	A	G	G	Т	G
NA18637_c		G	A	G	G	G	G	G	G	Т	A	A	G	G	Т	Т
NA18940_c:		G	A	G	G	G	G	G	G	Т	A	A	G	G	Т	Т
NA18582_c		Т	A	С	G	G	G	G	С	Т	A	A	С	G	T	G
NA18542_c		T	A	С	G	G	G	G	С	Т	A	A	G	G	T	G
NA18545_c:	1 T	Т	A	С	G	G	G	G	С	Т	A	A	G	G	Т	G

The International HapMap Project

This is a multi-country project to identify and catalogue genetic similarities and differences in human beings http://www.hapmap.org. In phase I & II 270 DNAs haplotyped were from the CEPH, Han Chinese, Japanese and Yoruba populations 6 million SNPs. This means that there will be a common, genotyped SNP every 600bp on average. The Generic Genome Browser enables one to look at the genotyped SNPs associated with a particular region or landmark and provides links to frequency and genotype data.

- Goal: Determine common patterns of DNA sequence variation in human genome in samples from different populations
- Attempts to capture most of the variation due to existing ~10 million SNPs by genotyping 200,000-1,000,000 tag SNPs
- BUT by focusing on common variants, may miss rare, diseaseassociated variants.

		•	ASW
•	African ancestry in Southwest USA		CEU
•	Utah residents with Northern and Western European ancestry from the collection	CEF	
	Han Chinese in Reijing, China	•	СНВ
•	Han Chinese in Beijing, China	•	CHD
•	Chinese in Metropolitan Denver, Colorado		GIH
•	Gujarati Indians in Houston, Texas	•	GIH
		•	JPT
•	Japanese in Tokyo, Japan	•	LWK
•	Luhya in Webuye, Kenya		
•	Mexican ancestry in Los Angeles, California	•	MXL
		•	МКК
•	Maasai in Kinyawa, Kenya		TSI
•	Toscani in Italia		
•	Yoruba in Ibadan, Nigeria	•	YRI

A HapMap tutorial section (<u>http://hapmap.ncbi.nlm.nih.gov/tutorials.html</u>) includes presentations from the the HapMap tutorials at American Society of Human Genetics Annual Convention on the 27th of October 2005 and a 'Users Guide to the web site'. The Generic Genome Browser enables one to look at the genotyped SNPs associated with a particular region or landmark and provides links to frequency and genotype data.

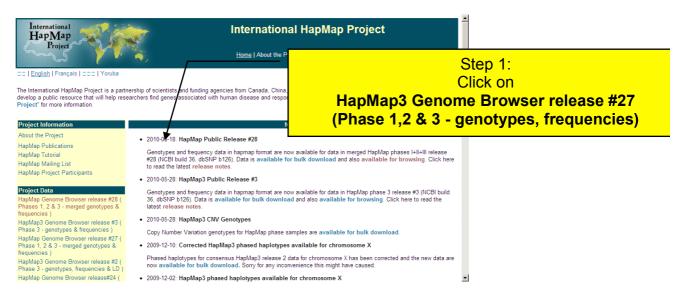
Worked Example:

You have identified IL7R as being differentially expressed in patients and controls. You want to see what SNPs have been genotyped for IL7R as part of the HapMap project and to find out their frequencies in the Caucasian population.

Questions:

- 1. How many SNPs have been genotyped within 10kb of IL7R?
- 2. How many of these have minor allele frequencies >0.1 in the Caucasian HapMap population?
- 3. Are any of the SNPs in linkage disequilibrium?
- 4. What are the tag SNPs?

To start, access HapMap, http://hapmap.ncbi.nlm.nih.gov/



NB. Since so much has to be precomputed, the visualisation features (especially plots) lag behind the latest data releases. Below we use release 27 containing data from phases 1, 2 & 3, where these tools are available at the time of writing.

Instructions Searching: Search using a sequence name, ge Navigation: Click one of the rulers to center on Examples : Chr20, Chr9:660,000760,000, SN [Help] [Reset] Search	Choose Annotate LD	to change magnification and position.
Help links:		
LD tagSNPs	Phased Haplotype - Genotype data -	- Frequency data - Symbols and colours used -
Landmark or Region :		Reports & Analysis :
IL7R Search		Annotate LD Plot Configure Go
Data Source HapMap Data Rel 27 Phasell+III, Feb09, on N	BI B36 assembly, dbSNP b126	

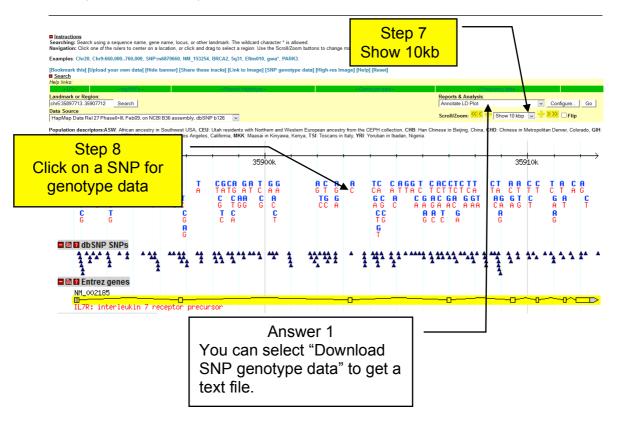
Population descriptors:ASW: African ancestry in Southwest USA, CEU: Utah residents with Northern and Western European ancestry from the CEPH collection, CHB: Han Chinese in Beijing, China, CHD: Chinese in Metropoltan Denver, Colorado, GH: Gujarati Indians in Houston, Texas, JPT: Japanese in Tokyo, Japan, LWK: Luhya in Webuye, Kenya, MEX: Mexican ancestry in Los Angeles, California, MKK: Maasai in Kinyawa, Kenya, TSI: Toccansi Intaly, YRI: Youhan in Ibadan, Nigeria.

Make sure 'LD Plot' track is switched on (bottom of page) also 'Phased

Haplotype Display' if available for that dataset.

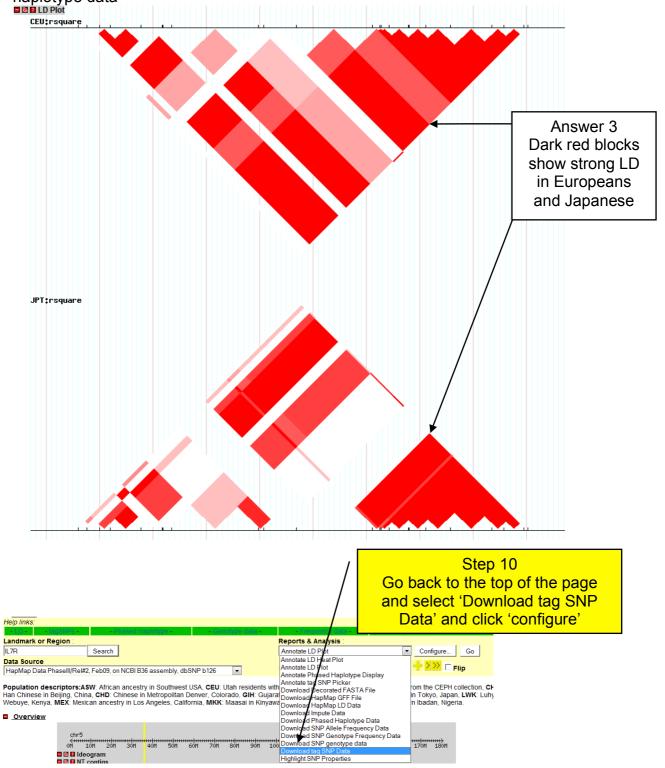
Analysis Advon Advor Analysis Advon Advor Analysis Advon Advort Analysis Advon Advort Analysis Advon Advort Analysis Advon Advort Advontation advort Display Settings Display Setti	<u>Variation</u>	All on 🔲 All off						
Image: Display Sattings Image: Bisplay Sattings Image: Bisplay Sattings Image: Bisplay Sattings Image: Display Sattings Step 3: Chooses r ² for LD properties > 0.3 and <10			🗹 Genotyped	SNPs	Recombination ra	te (cM/Mb)	Sequence Tagged	Sites
Step 3: Choose r ² for LD properties >0.3 and <1.0	<mark>■ <u>Analysis</u> □</mark> A	\ll on 🔲 All off						
Lisplay Settings Step 3: Choose r ² for LD properties >0.3 and <1.0 Step 4: Turn on the European (CEU) and Japanese (JPT) populations, inverting the latter. Population: of a a data dest and a data dest and a data dest and a data dest and a data dest dest dest dest dest dest and a data dest dest dest dest dest dest dest dest	LD Heat Plot		LD Plot					
<pre>properties >0.3 and <1.0</pre>			_	—	Sten 3		ifigure tracks	Jpdate Image
<pre>properties >0.3 and <1.0</pre>	Display Settings				$O_{\rm L}$			
ConfigureLD Plot Carcel Onfigure Max. Segnent Size Max. # grd 5 Wr 200 Droperties: requare greater than and less than 1 greater than 200 Coior: Paining pdt Coior: Pdt Coio								
ConfigureLD Plot Carcel Onfigure Max. Segnent Size Max. # grd 5 W Box Size [250k] Droperties: require greater than releas than LD Properties: require greater than releas than Color: Painties pld Color: Painties pld Color: AddW CEU Off Born Off on off				gorg	erties >0.3 a	and <1.0)	
Cancel Configure Step 4: Nax. Segment Size You we we have a set of the		-4						
Max. Segnent Sire Max. figt 4 Stresses LD Properties: reader than 1.0 reader than 1.1 mailess than 1.1 reader than 2500 w reader than 1.1 mailess than 1.1 reader than 2500 w reader reader than <td></td> <td>ot</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>		ot						
250Kb 20 Proportionate LD Properties: Isquare Image: Stand St	Cancel Configure	」,	/					
250Kb 20 Proportionale LD Properties: reader than and less than Color: Periodic particle Proportionate Color: Periodic particle Periodic particle Populations: ASW CEU CHB CHD Off On Off <on< td=""> Off On Off<on< td=""> O</on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<></on<>	Max. Segment Size	Max. # gt'd SNPs	Box Size		Г		Step 4:	
LD Properties: require greater than and less than 250 greater than 250 gre					/	Turn		nean
Color: Pairwise plot red		×			/			
Color: Pairwise plot red Populations: ASW CEU CHB CHD GH JFT LWK MEX MKK TSI YRI off on off on	LD Properties:	rsquare 💌	<u> </u>			•	· · ·	
Color: Pairwise pot red ASW off off off off off off off of			·		/	(JP	T) population	ons,
Populations: ASW CEU CHB CHD GH OF OF<	Color:					•	· · · ·	
Populations: e off on	COIDT.	red 💌			4	inve		
Populations: e off on		ASW	CEU	СНВ СНО	GIH JPT	IWK MEX	MKK TSI	YRI
Cancel Configure Cancel Configure Step 5: Click configure Later configure Click configu	Populations:							
Step 5: Click configure Step 5: Click configure Search sing a sequence name, gene name, locus, or other landmark. The widdcard character * is allowed. Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the ScrollZoom buttons to change magnification and position. Examples: Chr20, Chr2:660,000, SNP:rs6870660, NM_153254, BRCA2, 5q31, Elm010, gwa [*] , PARK3. Search Telpi IResc Search Telpi IResc Search Telpi Integer Search Telpi Integer Search S	Orientation:	invert 💌	normal 💌	invert 💌 normal 💌	invert 💌 invert 💌	invert 💌 norma	al 💌 invert 💌 norma	al 💌 invert 💌
Instructions Search sing a sequence name, gene name, locus, or other landmark. The widdcard character* is allowed. Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position. Examples: Chr20, Chr2:660,000, SNP:rs6870660, NM_153254, BRCA2, 5q31, EHm010, gwa*, PARK3. Fielp IResc: Search Search Search Search Instructions Reports & Analysis: Innotate LD Plot Annotate LD Plot Monotate LD Plot Annotate Ann	Cancel Configure	1						
		· •						
								on E.
Instructions Search using a sequence name, gene name, locus, or other landmark. The wildcard character* is allowed. Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position. Examples : Chr20, Chr3:660,000, SNP:re6870660, NM_153254, BRCA2, 5q31, ENm010, gwa*, PARK3. File[n] Resot Search File[n] Resot Landmark of Region HapMap Data Rel 27 Phasell-III. Feb09, on NCBI B36 assembly, dbSNP b126								
Searching: Search using a sequence name, locus, or other landmark. The wildcard character "is allowed. Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position. Examples : Chr20, Chr3:660,000, SNP:rs6870660, NM_153254, BRCA2, 5q31, ENm010, gwa [*] , PARK3. [Help] Resot] Search Help links: Communication of the rulers of the							Click of	configure
Searching: Search using a sequence name, locus, or other landmark. The wildcard character "is allowed. Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position. Examples : Chr20, Chr9:660,000, SNP:rs6870660, NM_153254, BRCA2, 5q31, ENm010, gwa", PARK3. [Help] [Reset] Search Help links: Command or Region : Reports & Analysis : IL/R Configure Go Data Source HapMap Data Rel 27 Phaseli-III. Feb09, on NCBI B36 assembly, dbSNP b126								
Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position. Examples : Chr20, Chr9:660,000, SNP:rs6870660, NM_153254, BRCA2, 5q31, ENm010, gwa [*] , PARK3. [Help] [Reset] Search Help inks:	Instructions							
Examples : Chr20, Chr2660,000,.760,000, SNP:rs6870660, NM_153254, BRCA2, 5q31, ENm010, gwa*, PARK3. [Help] [Reset] Search Help links: Landmark or Region : Landmark or Regio	Searching: Search using a s					ion and position.		
Integritives of search Search for search search search Search for search search search search Search for search searc								Cto
Landmark or Region: Search Configure Go Data Source HapMap Data Rel 27 Phaseli-III, Feb09, on NCBI B36 assembly, dbSNP b126 V	[Help] [Reset]						-	
Landmark or Region: Landmark or Region: Landmark or Region: Landmark or Region: Landmark or Region: Configure Go Data Source HapMap Data Rel 27 Phasell+IIL Feb09, on NCBI B36 assembly, dbSNP b126 v	Search							Search
IL/R Annotate LD Plot Configure Go Data Source HapMap Data Rel 27 PhaseII+III, Feb09, on NCBI B36 assembly, dbSNP b126	LD tagSNPs		^o hasedi Hapletyoe -	- Genotype dat	a - Frequency o	lata.	- Symbols and colt	
HapMap Data Rel 27 Phasell+III, Feb09, on NCBI B36 assembly, dbSNP b126 💌		Search				F	Configure Go	
		all+III Feb09 on NCBLB	36 assembly dbSNP b126					
Population descriptors:ASW: African ancestry in Southwest USA, CEU: Utah residents with Northern and Western European ancestry from the CEPH collection, CHB: Han Chinese in Beijing, China, CHD: Chinese in			-		Jestern European ancestry from the	CEPH collection CHR	Han Chinese in Beijing, China (Chinese in

Population descriptors:ASW: African ancestry in Southwest USA, CEU: Utah residents with Northern and Western European ancestry from the CEPH collection, CHB: Han Chinese in Beijing, China, CHD: Chinese in Metropolitan Derwer, Colorado, GH: Gujarati Indians in Houston, Texas, JPT: Japanese in Tokyo, Japan, LWK: Luhya in Webuye, Kenya, MEX: Mexican ancestry in Los Angeles, California, MKK: Maasai in Kinyawa, Kenya, TSI: Toscans in Italy, YRI: Yoruban in Ibadan, Nigeria.



chr5:35910332..35910332, (+) strand relative to the human reference sequence

Population					ype frec							Ref-allel	е	ċ	uencies)ther-alle			
	genotype	freq	countg	enotype	e freq o	countg	enoty	pe freq c	ount	Total	allele	e freq (count	allel	e freq c	ount	Total	
ASW (A)	C/C	0.830	44	C/T	0.132	7	T/T	0.038	2	53	с	0.896	95	т	0.104	11	106	retrieve genotypes
CEU (C)	C/C	0.593	67	C/T	0.327	37	T/T	0.080	9	113	С	0.757	171	Т	0,243	55	226	retrieve genotypes
CHB (H)	C/C	0.726	61	C/T	0.202	17	T/T	0.071	6	84	С	0.827	139	Т	0.173	29	168	retrieve genotypes
CHD (D)	C/C	0.624	53	C/T	0.341	29	T/T	0.035	3	85	С	0.794	135	Т	0.206	35	170	retrieve genotypes
GIH (G)	C/C	0.739	65	C/T	0.261	23	T/T	0	0	88	С	0.869	153	7	0.131	23	176	retrieve genotypes
JPT (J)	C/C	0.663	57	C/T	0.314	27	T/T	0.023	2	86	С	0.820	141	t	0.180	31	172	retrieve genotypes
LWK (L)	C/C	0.899	80	C/T	0.101	9	T/T	0	0	89	С	0.949	169	/Τ	0.051	9	178	retrieve genotypes
MEX (M)	C/C	0.600	30	C/T	0.340	17	T/T	0.060	3	50	С	0.770	77	Т	0.230	23	100	retrieve genotypes
MKK (K)	C/C	0.797	114	C/T	0.189	27	T/T	0.014		143	С	0.892	255	Т	0.108	31	286	retrieve genotypes
TSI (T)			Δnc	wer	2			0.034	3	88	С	0.778	137	Т	0.222	39		retrieve genotypes
YRI (Y)			-	-				-0		113		0.951	215	Т	0.049	11	226	retrieve genotypes
Note: the 're	All	ele	freq	uenc	y in	pop	C	e genome	e seq	uence	at this	s locatio	n					
						• •												
	OR select "Download																	
	SNP genotype <u>frequency</u>																	
	data"																	
			u	ald														



Step 9 - Go back one page to the chromosome view page and scroll down for haplotype data

Configur	e tag SNP Data	4			Step 11 Choose your options
Population	<u> </u>	CEU -			and click 'Go'
Pairwise M		Tagger Pairwise*	▼ [?]		
RSquare c		0.7 • [?]	L.1		
MAF cut o		0.05 - [?]			
include SN			Brows	se [?]	
Exclude Si			Brows		
			Brows		
Design sc Output for		I ● text ○ Save to		ie	
<pre>#tag SNPs CC rs139930 rs7737000 rs7737005 rs10461559 rs6451229 rs11867751 rs11867751</pre>	iromosome Pos r shr5 35899776 shr5 358907030 shr5 35898598 shr5 35804109 shr5 35801975 shr5 35898742	maf 0.258 0.152 0.243 0.183 0.450 0.075 0.347	out for population CEU chr5:3	589771335907712 using the algor	ithm-Tagger-pairwiseTagging
	cent of alleles with me s in 7 tests.				
	37717955 1.0	r			
rs6451226 r:	11567714 1.0 • 1389830 1.0	\sim -			
rs1389830 r:	1389830 1.0 1389830 1.0		Δ	nswer 4	
	1389830 1.0 6451229 1.0				.
rs6891095 r:	\$7737000 1.0		SNPs identifie	d as tag SNPs us	sina l
	1389830 1.0 10461959 1.0				
	1389830 1.0		the above c	riteria in selecteo	
	1389830 1.0				
	1389830 1.0		genor	ne window.	
	1389830 1.0 1389830 1.0		gener	ne window.	
	1389830 1.0				
	1389830 1.0				
	1389830 1.0				
	1389830 1.0	_			
	37737000 1.0 311567751 1.0				
Test Alleles C rs1389830 r: rs7737000 r: rs10461959 r: rs6451229 r: rs64512714 r:	aptured	11567737,rs1389830,rs	10074095,r#6893142,r#6451226,r#	7711202,rs10063445,rs10941267,rs1	494555, rs1494556, rs10044838, rs9292616

Haploview - haplotyping software

Haploview is a user friendly piece of freeware that has been designed to generate haplotypes directly from HapMap or from your own data. It can be downloaded at http://www.broadinstitute.org/haploview/haploview-downloads Comprehensive documentation is also available at this web page.

Worked Example:

Import your data	a files into	Haploview.			Choos	Step 1 se the sample.ped file for the
Ор	oen new data					file and sample.info for the
	Linkage Format					
	Haps Format	Data File:	gram Files\HaploView\sam	ple.ped		Locus information file
	HapMap Format	Locus Information File:	gram Files\HaploView\sam	ple.info	Browse	
	HapMap PHASE	□ X Ch	romosome 🔲 Do associa	tion test		-
H	apMap Download			lion cosc		-
	PLINK Format	 Fami 	ily trio data 🛛 🔿 Case/Conl	rol data		
		۲	Standard TDT 🔿 ParenT	DT		
		Test list file (optional):			Browse	
	Ignor	e pairwise comparisons of	f markers > 500 kb	apart.		- - -
	E	\times clude individuals with >	50 % missing genotype	s.		-
		OK	Cancel	Pro	xy Settings	

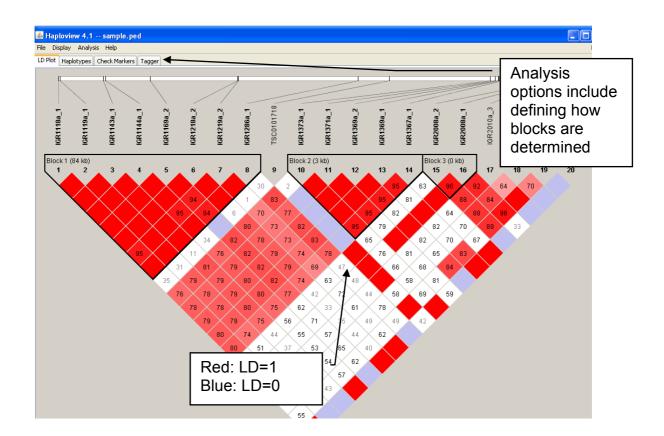
Г

The sample files contain data on 40 trios (father, mother and child), so simply click ok.

Haploview 3.0 File Display An													L 🗆 🗙 Kev	I		
LD Plot Haplotyc	· ·]										1107	-		
			Using 0 singletons	and 40 tric	os from 40	families.	Show E	xcluded ir	ndividual	: ▲	<u> </u>					Step 2
	#	Name	Position		PredH	HWpval		FamTric	Mend		Rating	-		<u>Che</u>	ck	the data looks ok
	1	IGR1118a_1	274044	0.282	0.269	0.762	97.5	39	0	0.16						
	2	IGR1119a_1	274541	0.267	0.257	0.938	96.7	37	0	0.151						
	3	IGR1143a_1	286593	0.3	0.289	0.516	100.0	40	0	0.175		\mathbf{k}				
	4	IGR1144a_1	287261	0.283	0.272	0.696	100.0	40	0	0.162	V					
	5	IGR1169a_2	299755	0.268	0.241	0.392	93.3	33	0	0.14	<u> </u>					r
	5	IGR1218a_2 IGR1219a_2	324341 324379	0.301	0.284	0.63	94.2 90.8	33 31	0	0.171	ঘ					
	6	IGR1215a_2	358048	0.275	0.278	1.0	95.0	35	0	0.167	<u>v</u>					Minor Allele
	<u>°</u>	TSC0101718	366811	0.265		1.0	95.0	34	0	0.145	- VI					
	10	IGR1373a 1	395079	0.283		0.176	100.0	40	0	0.162	- T					Frequency
	11	IGR1371a 1	396353	0.277		0.215	93.3	33	0	0.162	<u> </u>					
	12	IGR1369a 2	397334	0.311		0.139	88.3	31	0	0.181	V					
	13	IGR1369a_1	397381	0.275	0.264	0.216	100.0	40	0	0.156	V					
	14	IGR1367a_1	398352	0.283	0.264	0.216	100.0	40	0	0.156	V					
	15	IGR2008a_2	411823	0.393	0.441	0.695	93.3	34	0	0.329	V					
	16	IGR2008a_1	411873	0.294	0.403	0.04	85.0	29	0	0.28	V					
	17	IGR2010a_3	412456	0.336	0.403	0.143	96.7	38	0	0.279	V					
	18	IGR2011b_1	413233	0.489	0.499	0.84	75.0	27	0	0.483	- V	-				
	19	IGR2016a_1	415579	0.351	0.422	0.151	95.0	37	0	0.303	- 1	·				
			H	∦ p•value	cutoff: 0.	0010									_	
				Min geno	type %: 🛛	75								-		Are the SNPs in
				Max # mer	ndel errors:	1		Select All								
			Minin	num minor a	allele freq.	0.0010										HW equilibrium?
				_	Rescore	Markers										

There are three haplotype blocks, defined by confidence intervals (Gabriel et al, Science, 2002), which can be genotyped using four tag SNPs.

🕌 Haploview 4.1 sample.	ped		
File Display Analysis Help			Кеу
LD Plot Haplotypes Check Marke	rs Tagger		
Block 1	Block 2	Block 3	
001000000000000000000000000000000000000	11 11 14 14	16	
GGACAACC.82	5 TTACG .83	1 CC .664	
AATTCGTG.14			
	.78	TC .054	
		.79	
Europies basistance	above 1.0 % ^{Dir}	splay alleles as:	
Examine haplotypes	abuve 1.0 70	letters	
Connect with thin lir	nes if > 1.0 % (🔵 numbers	
Connect with thick li	nes if > 10.0 % (🔵 colored squares	
	Go		



LD plot shows how the haplotype blocks are composed.

This data set is available online as part of the Haploview tutorial, along with descriptions of file formats and how Haploview analyses data.

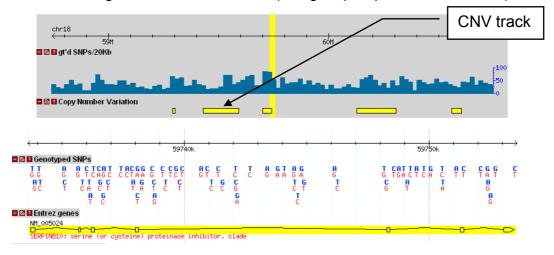
Copy Number Variation

Another frequent source of polymorphisms is Copy Number Variation (CNVs), which are a type of "structural variation" in the genome. These include anything from small insertions and deletions (≥1kb) and duplications to large scale duplications (≥50kb). Such CNVs may be on different chromosomes through duplications followed by translocation events, or are segmental duplications arisen through non-allelic homologous recombination. Many of these CNVs have been found through analyses of the HapMap data, and more are being found with the 1000 genome project. For more see Conrad et al. Nature 2009 and the CNV discovery project at WTSI

(http://www.sanger.ac.uk/humgen/cnv/42mio/)

While our knowledge of CNVs is far from as complete as with SNPs, they are starting to be implicated in human diseases, often due to their effects via gene dosage. So it's definitely worth checking if any are known in your area of interest.

The HapMap site (<u>http://hapmap.ncbi.nlm.nih.gov</u>) we looked at earlier has a track for CNVs e.g. here for SERPINB10 (using HapMap data Release 24).



The Database of Genomic Variance (http://dgv.tcag.ca/dgv/app/home) is a

more specialised site for finding CNVs.

			.			Search for
	Datal	base of	f <mark>G</mark> eno	mic ${\cal V}$ a	ariants	SERPINB10
			-		ural variation	
	A curatea ca	tulogue oj	numun gei		urur variation	
				/		
	About the Project Genome Browser	Downloads Query Tool	Links Submissions	Statistics Contact Us	FAQ Training Resources	
Keyword,	Landmark o	r Region S	Search:	•	Search NC	BI36/hg18 💌
	Example	es: RP11-34P13	; CFTR, 7q11.21	; chr7:71890181	-72690180	
		Find D	GV Varian	ts		
		by Study	by Sample			
		by Method	by Variant			
		by Platform	n by Chromoso	me		
		Summa	ary Statist	ics		
		Stat	Merged-level	Sample-level		
		CNVs: Inversions:	109863	2304349 3380		
			238 f Studies: 55	3300		
		11011001 0				

News: July 2013 Update and Newsletter has been issued

Opening the results in the genome browser, shows that Database of Genomic Variance has three CNV encompassing at least part of SERPINB10 (Red = gain, Blue = loss). Clicking on the CNV (here the red) gives more data:



It may also be useful to check CNVD <u>http://202.97.205.78/CNVD/</u> which text mines CNVs from publications in addition to use large studies.

Exercises:

1. Analysis of sequence variation at the RUNX1 locus

Using the Ensembl database determine the number of coding SNPs within the longest transcript of transcription factor RUNX1. How many stop gain, frame shift coding and non-synonymous SNPs are there? What are their ambiguity codes and do they encode amino acid substitutions. How many non-synonymous SNPs have Validation information and which single SNP would you type first?

2. Sequence variation in SERPINB10

Microarray data suggests that differential SERPINB10 expression is involved in prostate cancer. You have a medium sized, European, case-control population with 500 cases and 500 controls. You can only afford to genotype a couple of SNPs to see if any are associated with prostate cancer risk in your cohort. Which SNPs do you pick and why?

3. What are the significant SNPs and MYH11 haplotype that predisposes to disease?

Your microarray results showed that MYH11 shows differential expression in diseased compared to non-diseased aorta. You decide to genotype 28 SNPs in all of your case and control coronary artery disease samples to identify genetic association of SNPs within the gene. Using Haploview [two files "Genotype_data.ped" and SNP_Locations.info] determine how many haplotypes are there? What is the name of the SNP significantly out of HWE? Are any of the SNPs or haplotypes significant or trending in cases than controls (or vice versa)? What are the tag SNPs and how many of them are there?

Hint: Check "Do association test" and "Case/Control data" from the first window

Hint: Find statistical significance under the 'Association' tab, then look at 'Single Marker' and 'Haplotypes' for significant SNPs'

Hint: Tag SNP data can be found under the 'Tagger' tab, alter the r^2 to 0.7 then "Run Tagger" at the bottom of the page

4. Using dbGaP for identifying heart disease loci.

You have a heterogeneous Type-1 Diabetes cohort, and you decide to subset your population to help define the multiple genetic components of the disorder.

You segregate a T1D sub-population with myocardial infarction within your cohort. A small, seemingly underpowered screen that you carry out shows some significance on chromosome 16p13.13. Are there any studies within dbGaP that might help you decide if your finding is real or novel? *Hint:* Type 1 Diabetes Genetics Consortium (T1DGC): Genome-Wide Association Study in Type 1 Diabetes, 2008 (pha002862.1) *Hint:* Try changing the filter of the GWAS to <10e-6 for visualisation *Hint:* 16p is the petit (shorter) arm.

Answers

Task 1: Variation in the RUNX1 gene

Search Human Ensembl for RUNX1 to determine the gene ID and link to the gene view. You can then link to SNP information in Variation Table for the gene. *Currently*, a total of 153 Stops, 789 Frame shift and 995 non-synonymous (missense). 8 have validation data (multiple observations), possibly rs74315451 because of PolyPhen2 and SIFT predictions, but there are lots of other potentially pathogenic variants!

Task 2: Sequence variation in SERPINB10

There are no 'right' answers to this question!

Searching PubMed with the terms 'SERPINB10' and 'polymorphism' identifies a paper by Shioji et al (J Hum Genet (2005) 50: 507-515). Two cSNPs, rs8097425 and rs963075, are shown to have significant associations with prostate cancer in a Japanese cohort. These SNPs and three SNPs in SERPINB2 form a haplotype block which can be defined by genotyping just two SNPs (one in SERPINB2 and one in SERPINB10).

Task 3: Does a MYH11 haplotype predispose to disease?

There are 7 haplotype blocks as first defined in Haploview The SNP significantly out of HWE is rs7203040 – do you think that this could be important? The significant SNPs are rs215571 (trending – almost significant) and rs2306860 (p-val 0.02). There are no significant haplotypes, though 1 haplotype in block 5 and block 7 are trending.

	1 2 3	R51050163 R51050162	т	0.488, 0.484	0.011	0.916
	2	R51050162				0.910
	3		C	0.519, 0.518	0.0020	0.9609
		RS2075511	G	0.515, 0.500	0.28	0.5965
	4	R511130	A	0.522, 0.500	0.544	0.4606
	5	R512907	A	0.021, 0.020	0.02	0.8881
	6	R516967494	T	0.279, 0.258	0.7	0.4028
	7	R51050113	A	0.337, 0.310	1.005	0.3161
	8	R52272554	G	0.425, 0.393	1.331	0.2486
	9	R54781689	A	0.095, 0.092	0.04	0.8417
	10	RS6498574	G	0.368, 0.349	0.474	0.491
	11	R58044595	G	0.334, 0.306	1.097	0.2948
	12	R51050111	A	0.136, 0.115	1.233	0.2669
	13	R57184472	c	0.817, 0.815	0.0050	0.9415
	14	RS215590	A	0.416, 0.379	1.646	0.1995
	15	R5215581	c	0.786, 0.768	0.529	0.4668
	16	RS215579	Т	0.700, 0.669	1.335	0.248
	17	RS215573	c	0.696, 0.671	0.805	0.2606
	18	RS215571	Т	0.632, 0.581	3.361	0.0668
	19	RS215570	Т	0.633, 0.587	2.563	0.1094
	20	R59935015	G	0.795, 0.774	0.822	0.3646
	21	R53851706	G	0.737, 0.701	1.835	0.1700
	22	R52306860	G	0.592, 0.527	5.214	0.0224
	23	R58057023	G	0.943, 0.926	1.546	0.2138
	24	R512446688	G	0.760, 0.753	0.07	0.7912
	25	R53826056	G	0.941, 0.927	1.005	0.3161
	27	RS12597051	Т	0.944, 0.928	1.368	0.2422
	28	R53213476	A	0.841, 0.821	0.916	0.3384

What happens to the results when you change the LD blocks?

Haplotype	Freq.	Case, Control Ratios	Chi Square	p value	LD Plot Haplotypes Check Markers Tagger Association							
Maplotype Associations					Configuration Results							
Block 1					Sec. 2	1	Name	Position	Design Score	Force Include	Force Euclude	Capture this Allele
-cc	0.515	0.515, 0.513	0.0070	0.9337		1	R51050163	15718524	0			
TT	0.485	0.485, 0.487	0.0070	0.9337		2	R51050162	15718563	0		Ĭ	
Block 2						3	RS2075511	15725642	0			
GA	0.506	0.513, 0.493	0.471	0.4924		4	R511130	15725811	0			
TG	0.476	0.474, 0.480	0.037	0.8479		2	R512907 R516967494	15726154	0		<u> </u>	<u> </u>
TA		0.010, 0.020	2.122	0.1452		7	R51050113	15726304	n	H	H	
Block 3		· · ·				8	R52272554	15757705	0	n	H H	Ň
CGAG	0.492	0.481, 0.516	1.556	0.2123		9	R54781689	15772973	0			
TAGG		0.274, 0.254	0.666	0.4143		10	R56498574	15795766	0			
CGAA		0.093, 0.088	0.08	0.7776		11	RS8044595	15813631	0			<u> </u>
CGGG		0.084, 0.075	0.323	0.5698		12	R51050111 R57184472	15824698	0			
CAGG		0.061, 0.056	0.126	0.7224		14	R5215590	15835092	0	H	H	
Block 4	0.000	0.001/ 0.000	01200	011001		15	RS215581	15840675	0	ň	Η	
AG	0.676	0.666, 0.694	1.097	0.2948		16	RS215579	15843113	0			
GG		0.197, 0.192	0.046	0.8306		17	RS215573	15849085	0			
GA		0.137, 0.114	1.464	0.2263		18	RS215571 RS215570	15851834 15852530	0		<u> </u>	
Block 5	0.125	0.137, 0.114	1.404	0.2203		20	R5215570 R59935015	15856402	'n	H	- H-	
ACTCTT	0.392	0.408, 0.359	3.067	0.0799		21	RS3851706	15858148	0	n	h h	
		0.205, 0.203		0.0799		22	RS2306860	15858260	0			
CCTCTT		0.195, 0.216	0.014	0.3556		23	R58057023	15861671	0			
CTCACC						24	R512446688	15862631	0			
CCCACC		0.094, 0.104	0.351	0.5537		25	R53826056 R512597051	15867043 15881106	0		<u> </u>	
CCTCCC		0.053, 0.072	1.956	0.1619		28	R53213476	15889759	n	H	H	
CTTCTT	0.015	0.017, 0.010	1.011	0.3145		-						
Block 6							Include Al	Exclude Al	Uncapture Al	Exclude A/T 8	cican.	Reset Table
GG		0.737, 0.701	2.017	0.1555			10006 MI	EXCOUPE HE		E1000 APT 0		RESEL TALIE
AA		0.205, 0.225	0.792	0.3735								
GA	0.062	0.056, 0.073	1.632	0.2015		💽 pa	rwise tagging only				7 Ma design	score
Block 7						0.40	pressive tagging: u	se 2-marker hanin		100 threshold	for multi-marker b	sets 3.0
GGGGTA		0.560, 0.513	2.716	0.0994						200 0100000	rer maio marrier e	
CGAGTA		0.211, 0.233	01000	0.3627		() ag	pressive tagging: u	se 2- and 3-marks	r haplotypes Max	tags 🛛 🕅	fin distance betwe	en tags 0 b
CGGGTG		0.096, 0.106	0.287	0.5923								
CAGCCG	0.062	0.057, 0.072	1.274	0.259		Run Tage	er 🚺 Load In	dudes Load	Excludes Al	eles to Capture	Design Score	s Reset Thres
CGGGTA	0.051	0.045, 0.062	1.756	0.1851								
GGAGTA	0.023	0.028, 0.014	2.67	0.1022								

There are 15 tag SNPs that capture 27 SNPs typed in the region – notice that the unchecked HWE SNP is not included. Check the "Results" Tab to see the tag SNPs

Task 4: T1D, MI endophenotype in dbGaP

C16, 11.15Mb, CLEC16A - check OMIM!