SOLVED BIOLOGICAL AND CLINICAL DATABASES EXERCISES. GDA2016

Exercise 1. Search information for specific SNVs in different databases.

Questions:

- A) dbSNP database: what can you say about dbSNP id rs158691 from dbSNP database? has it been validated? how?
- Type the dbSNP URL on your browser (<u>http://www.ncbi.nlm.nih.gov/SNP/</u>). There are two fields for searching using dbSNP id: the first one at the upper part of the web page and the second one at the "Search by IDs on All Assemblies" section.
- Search for dbSNP id rs158691 through the first option,

S NCBI	dbSNP Short Genetic Variations
dbVar	ClinVar GaP PubMed Nucleotide Protein
Search Entrez d	bSNP V for rs158691 Go
Have a question about dbSNP? Try searching the SNP FAQ Archive! Go	dbSNP dbVar
RSS FeedW Contact Us Organism Data dbSNP Homepage NCBI Variation Resources Announcements dbSNP Summary ETR Download	Search by IDs on All Assemblies Note: p= and set must be prefixed with "rs" or "ss", respectively (i.e. rs25, ss25) ID: Search Reset Reset
SNP SUBMISSION DOCUMENTATION SEARCH RELATED SITES	Submission Information - By Submitter - New Submitter Batches - Method - Population - Publication
	Batch
	Enter List NCBL Assay: JD(xs) Reference: SNP: ID(xs) Local SNP ID Upload List NCBL Assay: JD(xs) Reference: SNP: ID(xs) Local SNP ID Batch Query Help

- The result of the first field search gives information about how rs158691 has been validated:

SNP	 rs158691 	
	Save search	Advanced
Dis	splay Settings: 🕑 Summar	y, Sorted by SNP_ID
Re	esults: 2	
	rs158691 [Homo sapiens]	
1.		
	tgagggtggcaattcaaac	tgttgg[C/T]taggtgtgtataggagagtcacaat
	Chromosome:	19:23017251
	Gene:	LOC101929164 (GeneView)
	Functional Consequence:	intron variant
- 1	Validated:	by 1000G,by 2hit 2allele,by cluster,by frequency,by hapmap
	Global MAF:	C=0.3135/1570
	HGVS:	NC_000019.10:g.23017251T>C, NC_000019.9:g.23200053T>C,
		NR_110746.1:n.544+467A>G, XR_244111.1:n.544+467A>G
	rs60992747 has merged i	nto rs158691 [Homo sapiens]
2.		
	tgagggtggcaattcaaac	tgttgg <mark>[C/T]</mark> taggtgtgtataggagagtcacaat
	Chromosome:	19:23017251
	Gene:	LOC101929164 (GeneView)
	Functional Consequence:	intron variant
	Validated:	by 1000G,by 2hit 2allele,by cluster,by frequency,by hapmap
	Global MAF:	C=0.3135/1570
	HGVS:	NC_000019.10:g.23017251T>C, NC_000019.9:g.23200053T>C,
		NR_110746.1:n.544+467A>G, XR_244111.1:n.544+467A>G

- Then, search for dbSNP id rs158691 through the second option,



- With this search field, you go straightforward to the report page where you can find more information about this SNP. Specifically the information about validation can be found in several parts of the web page (most relevant are highlighted in red),

S ncbi	Shor	dbSN t Genetic	IP Variations	1	T						
dbVar Clin	Var GaP	PubMe	d	Nucleotide	Protein						
Search Entrez dbSN	P for	in dbSNP or lar	ge structural	variations in dbVa	r						
Jearen Enrez (abora				2							
	Reference SNP	(refSNP) Clus	ster Report: r	s158691							
Have a question		RefSt	NP		Allele		1	HG	/S Name	s	
searching the SNP	Organism: human (<u>Homo sapiens</u>)				Variation Class: SNV:	ucleotide variation	NC_000019.10:g.23017251T>C NC_000019.9:g.23200053T>C				
The function of	Created/Updated in build: 70/146				RefSNP Alleles: C/T (FV	NR_110746.1:n.544+467A>G					
Go	Map to Genome Build: 107/Weight				Allele Origin:			n.544+4	67A>G		
ENERAL	Vali	dation Status:			Ancestral Allele: C						
SS Enge			-	_	Variation Viewer:	w					
antact Lis					Clinical Significance: NA						
maniom Data					MAF/MinorAlleleCount: C=0.31	35/1570					
SND Homonogo					MAF Source: 1000 G	enomes					
ODI Verletler											
Conversion											
esources											
SND Summary	SNP Details are o	organized man	e tonowing Se	ctions:							
Download	GeneView	Map Sul	bmission	Fasta Reso	urce Diversity Validation						
IP Download											
OCUMENTATION	integrated Ma	ps (Hint: click	on Chr Pos	to see variant i	n the new NCBI variation viewer						
EARCH	Assembly 🜩	Annotation Release	Chr	Chr Pos	Contig	Contig Pos	SNP to Chr	Contig	Contig to Chr	N	
ELATED SITES	GRCh38.p2	107	19	23017251	NT 011295.12	22957251	Fwd	т	Fwd		
	GRCh37.p13	105	19	23200053	NT 011295.11	14462855	Fwd	т	Fwd		

 When looking at the Validation status in the dbSNP report, the field is empty and it differs from what we found when searching with the first option. Looking at other sections does not clarify the question. For example, in the Validation section of the report we find the following information,

Validation Sum	imary:		
Validation status	Marker displays Mendelian segregation	PCR results confirmed in multiple reactions	Homozygotes detected in individual genotype data
	UNKNOWN	UNKNOWN	UNKNOWN

- So, can we consider this information reliable? One option is to search for the rs158691 in other databases such as Ensembl or check its population frequency in different human variation catalogs.
- More information about validation status in dbSNP: <u>http://www.ncbi.nlm.nih.gov/books/NBK44476/#Reports.what_exactly_does_it_mean</u> <u>when_a</u> More information about validation status in Ensembl:

http://www.ensembl.org/info/genome/variation/data_description.html#evidence_statu

- B) COSMIC database: which is the KRAS gene position with highest substitution rate found in cancers? which is the most common substitution in this position? Is there any specific tissue distribution for this mutation?
- Type the COSMIC URL on your browser (<u>http://cancer.sanger.ac.uk/cosmic</u>) and search for KRAS gene in the "Search" field.



- Select the first Gene ID ("KRAS") in the results page.

About v	Resources v	Curation 🔻	Tools	<u>د</u>	Data 🔻	News 🔻	Help 🔻	Enter search	h here				
SMIC search results													
r keyword "KRAS" return	ned following results in	n the sections,											
	Show 10 T er	ntries	Se	earch:									
			Туре 🕒 🔺		All Hits	\$			Keyword se	arch:		Go	
		Disease Cla	assification		0								
			Pubmed		1028								
			Samples		0								
			Study		0								
		1	'umour site		0								
		Unique	Mutations		342								
			Gene		3								
	Showing 1 to 7 of	7 entries			Previous 1	Next							
10 Ventries	ied												
Gene		Alt Id	s ¢		Tested sample:	s ¢	Sim	ple Mutations	φ	Fusions	φ	Coding M	utations
S		KRAS		189621		3	37953		1		37953		
S ENST00000256078		KRAS ENSTOOD	00256078	25784		1	2806		0		2806		

In the next page you can find different bar plots with gene information. The first plot includes the counts of substitutions along the gene. Here, you can find that the position with the highest number of substitutions is position 12. Passing the mouse over the bars in the plot, some pop-up information appear. If you pass the mouse over the widest bar of the position 12, you can see that substitution p.G12D/c.35G>A has been observed 13350 times.



- Click on the region of the previous bar at position 12 (p.G12D/c.35G>A). There you can find information about the selected substitution. Click on the "Tissue Distribution" at the tab menu on the top to see its tissue frequency.



- C) humsaVar database: could you find the previous rs158691 SNP in this file? why?
- Type the humsaVar URL on your browser (<u>http://www.uniprot.org/docs/humsavar</u>).
 The information of this database is contained in a text file that you can download from its web page. You can search for rs158691 either within the text file or directly in the web page using the search option of the browser.

						rs158691	0 of 0	~ ~	~
Main gene name	Swiss-Prot AC	FTId	AA change	Type of variant	dbSNP	Disease name			
A1BG	P04217	VAR 018369	p.His52Arg	Polymorphism	rs893184	2 <u></u>			
A1BG	P04217	VAR 018370	p.His395Arg	Polymorphism	rs2241788				
A1CF	Q9NQ94	VAR 052201	p.Val555Met	Polymorphism	rs9073	-			
A1CF	Q9NQ94	VAR 059821	p.Ala558Ser	Polymorphism	rs11817448	-			
A2ML1	A8K2U0	VAR_055463	p.Gly207Arg	Polymorphism	rs11047499	ā.			
A2ML1	A8K2U0	VAR 055464	p.Cys970Tyr	Polymorphism	rs1558526	9 9			
A2ML1	A8K2U0	VAR_055465	p.Thr1131Met	Polymorphism	rs7959680	-			
A2ML1	A8K2U0	VAR_055466	p.Thr1412Ala	Polymorphism	rs7315591	ā.			
A2ML1	A8K2U0	VAR_059083	p.Asp850Glu	Polymorphism	rs1860926	12 C			
A2ML1	A8K2U0	VAR_059084	p.His1229Arg	Polymorphism	rs10219561	5			
A2ML1	A8K2U0	VAR_071854	p.Arg1122Trp	Polymorphism	rs1860967	-			
A2ML1	A8K2U0	VAR_071855	p.Met1257Val	Polymorphism	rs7308811	÷			
A2ML1	A8K2U0	VAR 071856	p.Thr1312Met	Polymorphism	rs201083574				

- Searching directly in the web page, you can't find any result for rs158691 because it is an intron variant. Note that humsaVar has been developed by UNIPROT, which is a well known and curated database for proteins (gene exons).
- D) ClinVar database: browse the clinical information reported for the conserved domain database (CDD) id NP_203524.1. Does it include the variant detected in B? which is its clinical significance? ant its review status? Note: CDS Mutation ID c.35G>A
- Type the ClinVar URL on your browser (<u>http://www.ncbi.nlm.nih.gov/clinvar/</u>) and search for NP_203524.1.

ClinVar ClinVar NP_20352	4.1	Search					
Home About Access Using the website	Submission Statistics FTP site						
ACTGATGGTATGGGGGCCAAGAGATATATC CAGGTACGGCTGTCATCACTTAGACCTCA	ClinVar						
CAGGGCTGGGCATAAAAGTCAGGGCAGAGC CCATGGTGCATCTGACTCCTGAGGAGAAGT GCAGGTTGGTATCAAGGTTACAAGACAGGT GCACGTTGGTATCAAGGTTACAAGACAGGT GGCACTGACTCTCTCTGCCTATTGGTCTAT							
Using ClinVar	Tools	Related Sites					
About ClinVar	ACMG Recommendations for Reporting of Incidental Findings	ClinGen					
Data Dictionary	Variation Submission Portal	GeneReviews®					
Downloads/FTP site	Submissions	GTR®					
FAQ	Variation Viewer	MedGen					
Contact Us	Clinical Remapping - Between assemblies and RefSeqGenes	<u>OMIM®</u>					
RSS feed/What's new?	RefSeqGene/LRG	Variation					
Factsheet	Variation Reporter						

- The results page reports 62 items for NP_203524.1. In this page, you can search for c.35G>A, which is the CDS mutation ID from Exercise 1B. Then, you can find that its clinical significance states that is pathogenic and no assertion criteria is provided.

	Variation Location	Gene(s)	Condition(s)	Frequency	Clinical significance (Last reviewed)	c.35G>A
4 9.	NM_03360.3(KRAS):c.38G>A (p.Gly13Asp) GRCh37: Chr12:25398281 GRCh38: Chr12:25245347	KRAS	Juvenile myelomonocytic leukemia, Non-small cell lung cancer, Breast cancer, somatic, RAS-associated autoimmune leukoproliferative disorder, Breast adenocarcinoma		Pathogenic (Ail 1, 2015)	criteria provided, single submitter
5 0.	<u>NM_033360.3(KRAS):c.37G>T (p.Gly13Cys)</u> GRCh37: Chr12:25398282 GRCh38: Chr12:25245348	KRAS	Non-small cell lung cancer, RAS- associated autoimmune leukoproliferative disorder		Pathogenic (Sep 17, 2012)	criteria provided, single submitter
51.	<u>NM_033360.3(KRAS):c.37G>C (p.Gly13Arg)</u> GRCh37: Chr12:25398282 GRCh38: Chr12:25245348	KRAS	Non-small cell lung cancer, Pilocytic astrocytoma, somatic, Pilocytic astrocytoma		Pathogenic (Apr 15, 2011)	criteria provided, single submitter
52.	NM_033360.3(KRAS):c.35G>C (p.Gly12Ala) GRCh37: Chr12:25398284 GRCh38: Chr12:25245350	KRAS	Non-small cell lung cancer		Pathogenic (Dec 7, 2007)	no assertion criteria provided
53.	<u>NM_004985.4(KRAS):c.35G>T (p.Gly12Val)</u> GRCh37: Chr12:25398284 GRCh38: Chr12:25245350	KRAS	Juvenile myelomonocytic leukemia, Carcinoma of pancreas, Non-small cell lung cancer, Nevus sebaceous, NEVUS SEBACEOUS, SOMATIC, Rasopathy		Pathogenic (Mar 25, 2013)	criteria provided, single submitter
5 4.	NM_033360.3(KRAS); c.35G>A (p.Gly12Asp) GRCh37: Chr12:25398284 GRCh38: Chr12:25245350	KRAS	Epidermal nevus syndrome, Juvenile myelomonocytic leukemia, Epidermal nevus,		Pathogenic (Jun 10, 2012)	no assertion criteria provided
			Neoplasm of ovary, Carcinoma of pancreas, Non-small cell lung cancer, RAS-associated autoimmune leukoproliferative disorder, Neoplasm of stomach, Nevus sebaceous, NEVUS SEBACEOUS, SOMATIC			
	NM 033360.3(KRAS):c.34G>A (p.Glv12Ser)	KRAS	Juvenile myelomonocytic leukemia,		Pathogenic	criteria provided, single submitter

E) OMIM database: search for the chromosome location of the B result. Is there any nearby clinical annotation that makes sense with the KRAS gene? (Note that OMIM mapping uses build GRCh38) - Type the OMIM URL on your browser (<u>http://www.omim.org/</u>) and click on "Gene Map" at "Advanced Search" section.



- Then, search for the location 12:25,245,350-25,245,350. Note the OMIM special format with commas.

12:25,245,350-25,245,350	Search
Entries per page: 10 🔻	
Search by genomic region (or cyto location range) to get a list of a	all OMIM Gene/Loci in that region, for example:
'1:0-124, 300, 000' or '1p36-p32'	all OMIM Gene/Loci in that region, for example:
Search by genomic region (or cyto location range) to get a list of a '1:0-124, 300, 000' or '1p36-p32' A message will be displayed indicating when a genomic region se	all OMIM Gene/Loci in that region, for example:
Search by genomic region (or cyto location range) to get a list of a '1:0-124,300,000' or '1p36-p32' A message will be displayed indicating when a genomic region se Search by genomic location (or cyto location band) to jump to tha	all OMIM Gene/Loci in that region, for example: arch is run. I location in the chromosome, for example:

- In the next results page, you can find 12:25,204,788 as the nearest KRAS position to the selected substitution in Exercise 1B.

12:25,245	350-25,245,350		Search							
iearch: Oh	IN Advanced Sea	rch + Display Options + Phenotype Only Entrie	All Entries							
Search: '1: Results: 7 Downlo	2.25,245,350-25,24 entries. and As +	5,350								
Senomic context able	Location (genomic start, cyto location)	GenelLocus	Gene/Locus name	Gene/Locus MIM number	Phenotype	Phenotype MIM number	(n progress)	Pheno map key	Comments	Mouse symbo
	12:0 12p	KAR	Aromatic alpha-keto acid reductase	107920					7same as MDH1	
	12:0 12p	PKS	Pallister-Killian syndrome	601803	Pallister-Killian syndrome	601803	SMo	4		
	12:10,000,000 12p13.2-p11.23	DFNB62	Deafness, autosomal recessive 62	630143	Deafrens, autosomal recessive 62	610143	AR	2	between D12S358 and D12S1042	
	12:10,000,000 12p13.2-q24.1	1802	Inflammatory bowel disease 2	601458	[inflammatory bowel disease 2]	601458		2	mainly alcerative colitis	
-	12:19,000,000 12p12.2-p12.1	HYT4	Hypertension, essential, susceptibility to, 4	606742	(Hypertension, essential, susceptibility to, 4)	145500	Mu	2		
	12:25,204,788	KRAS, KRAS2, RASK2, NS, CPC2, RALD	Kirsten rat sarcoma-2 viral (v-Ki-ras2) oncogene homolog	190070	Bladder cancer, somatic	109800		3	pseudogene KRAS1P on 6p12-p11	Kras
	represe				Cardiofaciocataneous synchrome 2	615278		3		
					Gastric cancer, sematic	137215		3		
					Leukemia, acute myeloid	601626	AD	3		
					Lung cancer, somatic	211980		3		
					Noonan syndrome 3	609942		3		
					Pancreatic carcinoma, somatic	266350		3		
					RAS-associated autoimmune leukoproliferative disorder	614470	AD	3		
					Schimmelpenning-Feuerstein-Mins syndrome, somatic mosaic	163200		3		
							_			

- F) HGMD database: register for the public version and try it at home.
- Type the HGMD URL on your web browser (<u>http://www.hgmd.cf.ac.uk/ac/index.php</u>). Click on "Register for public version" button.



- Then, fill the form to get access to the public version of HGMD.

С немо ^р	HGMD User Registration
PI	ease read the following before registering
Please note that the public version of our database is free only for regist from <u>QIAGEN®</u> , our commercial partner. The HGMD Professional lice database. HGMD mutation data are made available via the public site 3	ered users from academic institutions/non-profit organisations. Commercial users are required to purchase a license ense is available to both commercial and academic/non-profit users wishing to access the most up-to-date version of the years after initial inclusion. Please read the <u>DisCLAIMER</u> !
For legal reasons, only users who give an email address that can be CL1 hospital email address etc). Please DO NOT use your personal email ac access HGMD. Once registered, users will receive a password via the ar enabled, and will require their email address, country and password to l	EARLY assigned to an academic or non-profit organization will be allowed to register successfully (i.e a university or iddress from hotmail, yahoo, gmail, MSN or any other commercial web-based system. You WILL NOT be allowed to cademic/non-profit email address supplied when registering. To use the HGMD login, users must have session cookies og in.
	Registration data (*required)
First name*:	
Last name*:	
Background*:	Select background
Role/title*:	Select role/title
Company/Organisation*:	
Department:	
Address1*:	
Address2:	
City*:	
Post/Zip code*:	
Country*:	Select country
Telephone*:	
Fax:	
Email*:	
Privacy policy & disclaimer	Accept and register

- Once you have logged in, search for KRAS gene on the upper left.

KRAS	Gene symbol V Go!
The Human Gene Get HGMD Professional	Mutation Database (HGMD®) represents an attempt to collate kr *Please note that this less up-to-date public version of our database is freely availa up-to-date version of the database (visit QIAGEN® to request a free trial of HGM Cardiff University 2015. All rights reserved.
<u>Table:</u>	Description:
Cono cumbol	The gene description, gene symbol (as recommended by the H

- Select KRAS in the following table.

Gene symbol	
KRAS	V-ki-ras2 kirsten rat sarcoma viral oncogene homologu

- The next results page includes several information about KRAS, but some of it is only accessible from the professional version. Click on "Get mutations" of missense/nonsense type.

Gene Symbol	Chromosomal location		Gene name	cDNA sequence	Extended cDNA	Mutation viewer
KRAS (Allasse: available to <u>subscribers</u>)	12p12.1	V-ki-ras2 kirsten rat sarcoma viral oncogene (Aliane: avalable to salverbers)	homologue	NM_004985.4	Not available	BIOB SE Feature available to subscribers
	Mutation type		Number of mutations		Mutation data by type (regist	ter or los in)
Missense/nonsense			25		Get mutations	
Splicing			0		No mutations	
Regulatory			1		Get mutations	
Small deletions			1		Get mutations	
Small insertions			0		No mutations	
Small indels			0		No mutations	
Gross deletions			0		No mutations	
Gross insertions/duplications			0		No mutations	
Complex rearrangements			0		No mutations	
Repeat variations			0		No mutations	
Get all mutations by type				BIOB SE Feature swallable to subs	loibers	
Public total (HGMD Professional 2015.4 total)		27 (33)			
	Disease/nhenotyne		Number of mutations		Mutation data by dis	essembraature
Noonan candrome	and a second sec				BIOBA	SE
Noonan syndrome			14		Feature evaluable to	subscribers
Cardio-facio-cutaneous syndrome			6		Feature available to	SE subscribers
Costello syndrome			2		BIOB Feature evaluable to	SE subscribers
Cardio-facio-cutaneous syndrome ?			1		BIOB Feature available to	SE subscribers
Gallbladder carcinoma, increased risk, assoc w	ith		1		BIOB Feature available to	SE subscribers
Lung cancer, risk, association with			1		Feature evaluable to	SE subscribers
Multiple mole melanoma syndrome			1		BIOB Peature available to	SE subscribers
Myelodysplastic/myeloproliferative disease ?			1		BIOBA Feature available to	SE subscribers

- Here, you can find the codon and amino acid changes, as well as the phenotype it has been associated with.

Missense/no	isense	Splicing	Regulatory	Small deletions		Small insertions	Small indels	Gross deletions	Gross insertions	Complex	Repeats
29 mutations in HGMD pr	ofencional 2005.4	is mutations 2 m	tations in HGMD professional 2015.4	1 mutation in HGMD preferring	2015.4	No mutations	No mutations	1 mutation in HGMD prefermined 2015.4	No mutations	No mutations	No mutations
				Further o	ptions available in 🔣	CMD professional 2015.4					
Accession Number	Codon change	Amino acid chan	e Codon number	Genomic coordinates & HGVS nonsucciature		Phenotype		Refe	rence		Comments
CM070963	888-88T	Lys-Asn		BIOBASE Peature available to subscribers	Costello sync	drome		Zenker (2007) J Med Genet 44, 131 Factional characterisation report available to <u>subscriben</u> Additional report available to <u>subscriben</u>	1		
CM073168	ARR-GAR	Lys-6lu	:	BIOBASE Feature available to subscribers	Costello sync	drome		Bertola (2007) J Hum Genet 52, 521 Additional phenotype report available to <u>subscribers</u> Additional report available to <u>subscribers</u> Additional phenotype report available to <u>subscribers</u>			
CM076251	GGT-RGT	Gly-Ser	12	BIOBASE Feature available to subscribers	Cardio-facio-	-cutaneous syndrome		Nava (2007) J Med Genet 44, 263 Additional report available to <u>subscribers</u>			
CM087372	GGT-GRT	Gly-Asp	12	BIOBASE Feature available to subscribers	Multiple mol	le melanoma syndrome		Koorstra (2008) Am J Song Pathol 32, 3 Additional report available to <u>subserilien</u> Additional report available to <u>subserilien</u>	905		
CM125166	66C-69C	Gly-Rep	13	BIOBASE Feature available to subscribers	Myelodyspla	astic/myeloproliferative disear	ж?	Ismael (2012) Br J Haematol 158, 129 Additional report available to <u>subscribers</u> Additional report available to <u>subscribers</u> Additional phenotype report available to <u>subscribers</u>		5	Somatic mosa
CM061082	GTR-RTR	Val-Ile	14	BIOBASE Feature available to subscribers	Noonan synd	drome		Schubbert (2005) Nat Genet 38, 331 Functional characterisation report available to <u>subscribert</u> Additional report available to <u>subscribert</u>	1		
CM070966	CRG-CGG	Gin-Arg	22	BIOBASE Peature available to subscribers	Noonan synd	drome		Zenker (2007) J Med Genet 44, 131 Functional characterisation report available to <u>subscriben</u> Additional report available to <u>subscriben</u>	2		
CM070964	CRG-GRG	Gln=Glu	22	BIOBASE Feature available to subscribers	Cardio-facio-	-cutaneous syndrome		Zenker (2007) J Med Genet 44, 131 Factoral characterisation report available to <u>subscriber</u> Additional report available to <u>subscribers</u>	2		

Exercise 2. Retrieve genomic variation data from CellBase using its web services API. Note that the main host is <u>http://ws.bioinfo.cipf.es/</u> (GRCh37) but there is another mirror in <u>http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/rest</u> (GRCh38)

Some examples:

Get species included in CellBase:

http://ws.bioinfo.cipf.es/cellbase/rest/latest

Get all the mutations from BRCA2 gene:

http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/feature/gene/BRCA2/mutation

Get all the genes within a specific genomic region:

http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/genomic/region/1:3972105-12973105/gene Get the phenotype from rs3934834 SNP:

http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/feature/snp/rs3934834/phenotype

Questions:

A) We are interested in a particular region of the human genome chr12:25,350,000-25,245,000 (GRCh37), and we want to know if this region contains

mutations already catalogued. Help: latest (version), hsa (species), genomic (category), region (subcategory), 12:25350000-25450000 (id), mutation (resource).

- Query result: http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/genomic/region/12:25350000-25450000/mutation
- B) We want to know the allelic and genotypic frequencies for a SNP, rs158691, across populations. Help: latest (version), hsa (species), feature (category), snp (subcategory), rs158691 (id), population_frequency (resource).
 - Query result: http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/feature/snp/rs158691/populati on_frequency
- C) We have obtained a SNP of interest (rs28937313, location GRCh37 9:107584801) in our analysis and we want to know if it has been related with any disease.
 - Query result: http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/feature/snp/rs28937313/pheno type

Exercise 3. Browse different catalogs of human genetic variation.

Questions:

- A) Go to the latest release of the HapMap project and check the KRAS gene region (Note that HapMap uses NCBI build 36). Can you find the allele frequencies of genotyped SNPs in the HapMap populations?
- Type the HapMap URL on your web browser (<u>https://hapmap.ncbi.nlm.nih.gov/</u>). Select HapMap release #28 on the left menu (Project Data).

International Project Project Project Project Project I and Project I a	International HapMap Project Lace Alout the Project Data Publications Tutorial Lace Alout the Project Data Publications Tutorial Lace Alout the Project Data Publications Tutorial Lace stom Canada, China, Japan, Ngena, the United Kingdom and the United States to develop a public resource that will help researchers find pr
Project Information	News
Abox the Project Naphtap Publications Haphtap Tubicition Haphtap Abuing List Haphtap Project Panticipants Project Data	2013-05-14: Highlap data conversion tool There are several inquires for a conversion tool to convert Highlap data into the VCF format. Please take a look of The Genome Analys 2012-12-06. Downtime for hardware maintenance From December 15- 16, Hapmap site will be taken offline for an internal hardware maintenance. Sony for the inconvenience. 2011-0.01: Hughlap help deta kanouncement
Haphago Genome Brower release 2/2 (Plases 1, 2 & 3 - merged genotypes 4 Haphago Genome Brower release 2/2 (Plase 3 - genotypes 8 flequencies) Haphago Genome Brower release 2/2 (Plase 3 - genotypes 6 flequencies) Haphago Genome Brower release 2/2 (Plase 1, 2 & 3 - merged genotypes 6 Lo) Haphago Genome Brower release 2/2 (Plase 1, 2 - bit dataset) Haphago Genome Brower release2/2 (Plase 1, 2 - bit dataset) GWA Karyogram Haphag	There was a problem with the HapMap help desk system. In the past several weeks, emails sent to hapmap-help@rcbi.nlm.nlh.gov did r request if you serit emails to be HapMap help desk system. In the past several weeks. Sonry for the Inconventence. 2011-0-20: Mapma help desk suspont from 05002011 to 05023011. Sonry for the Inconventence. There will be no help desk suspont from 05002011 to 05023011. Sonry for the Inconventence. 2011-0-20: Maphoteve traces with rel 28 data. Recently, there are several questions about Haploview data formal errors when users theid to analyze HapMap release 28 data. The curr will general are nor Smith to "Happen data formal errors. VALIB/76" when typing to open the data.
HapMag FTP Buik Data Drezes for Publication ENCODE Project Guidelines For Data Use	Haptonews is developed and maintained by an organization different Ton HapMap. Please contact Happonew help desk (happonew)@bro 2 0110-01.9: HapMap phase I recombination rate on GRCh07 The ithore of the HapMap II genetic map from human genome build b35 to GRCh07 is available. Data is available for built download. 2 010-09-36: HapMap Public Release #28
Useful Linko TSC SNP Downlands HapMap Bayelier tass Cortell institute HapMap Project Proces Release NH-SRI HapMap Page NH-SRI HapMap Database (dSNP) Jaganese SNP Database (dSNP)	Centroppes and frequency data in hapmap format are now available for data in merged HapMap phases i Hi-III release #22 (NCBI build 5 to not the filese intesses notes: 2 2010 05-52: HapMap3 Dublic Release #3 Centroppes and Horgency data in hapmap format are now available for data in HapMap phase 3 release #3 (NCBI build 36, dbSNP b126 laterd release notes: 2 2010 05-32: HapMap3 CNV Genctypes Copy Number Variation genchpes for HapMap phase samples are available for build forwilead.

- Then, search for KRAS gene on the landmark or region search field.



- In the results page, there is a section named "Genotyped SNPs" where you can pass the mouse over the letters and check the frequency of the alleles. If you click directly on the letter, you can see a frequency report with population genotype and allele frequencies.

andmark or Region :	l e un l				Reports & Analysis :		1.0.1
CRAS	Search				Download Decorated FASTA File	• Conngure	G0
Jata Source LlanMan Data Rel 29 Dhasella	August10 on NCBI B26 assembly dbSNB b126 T				Scroll/Zoom: K Show 45.67 k	bp 🔻 🕂 🔀 🗉	Flip
rapinap baarner zorraben	In the second						
opulation descriptors:ASW: WK: Luhya in Webuye, Kenya,	African ancestry in Southwest USA, CEU: Utah residents MEX: Mexican ancestry in Los Angeles, California, MKR	with Northern and Western European Maasai in Kinyawa, Kenya, TSI: Tuso	ancestry from the CEP an in Italy, YRI: Yorub	PH collection, an in Ibadan,	CHB: Han Chinese in Beijing, China, CHD: Chinese in Metr Nigeria.	opolitan Deriver, Co	lorado, GIH
Overview							
		chr12				****>	
		0H 10H 20H	3011 4011 50	el eou	70H 80H 90H 100H 110H 120H	1304	
I 🛾 🖬 Ideogram			M				
		U		ſ			
E NT contigs				rs79671	20: Allele Frequencies in HapMap Popula	tions	
III III gt'd SNPs/500Kb						1	-
12 OMIM disease association	ns			Panel	Description	of T (ref)	Frequer of C
B GWA studies (NHGRI Cat	alog)			ASW(A)	African ancestry in Southwest USA	NA	NA
		• • •	•• •	CEU(C)	Utah residents with Northern and Western European ancestry from the CEPH collection	100%	0%
Region		cite12		CHB(H)	Han Chinese in Beijing, China	100%	0%
		÷ · · · · ·		CHD(D)	Chinese in Metropolitan Denver, Colorado	NA	NA
e at'd SNPs/20Kb				GIH(G)	Gujarati Indians in Houston, Texas	NA	NA
				JPT(J)	Japanese in Tokyo, Japan	100%	0%
				LWK(L)	Luhva in Webuve, Kenva	NA	NA
				MEX(M)	Mexican ancestry in Los Angeles, California	NA	NA
E Copy Number Variation				MKK(K)	Maasai in Kinyawa, Kenya	NA	NA
				TSI(T)	Tuscan in Italy	NA	NA
Details				YRI(Y)	Yoruban in Ibadan, Niperja	NA	NA
- Designation		353604 3850	24			1 1 1 2	10.000
Genotyped SNPs		111000 1100		1			
		TGT GTRR RG G G	ee ci i	f ç ç	T CRICI CORG R C T R TRGC TR	1	
		0 0 0	1 9	A 6	C A GA GG AC SCTTS	I	

- B) Now, go to the 1,000 Genomes browser and search for the KRAS genomic region (example: 12:25350000-25450000). Can you find the global MAFs of the SNPS in this region from the 1,000 Genome populations?
- Type the 1,000 Genomes URL in your browser (<u>http://browser.1000genomes.org/index.html</u>) and search for KRAS region.



- In the results page, there is a section named "1KG All SNPs/indels" where you can find the Global MAFs of variations by clicking on each position.

1KG AI SNPs/indels	vice AFAS-003 politic coltra		<kras-002 probin coding</kras-002 	0 1 0		
				5 features		U
MutiCel		Variation: rs529328197	Variation: rs369251072	Variation: rs550808613	Variation: rs61762434	Variation: rs10642508
Regulatory Features.		more about rs529328197	more about rs369251072	more about rs550808613	more about rs61762434	more about rs10842508
960		Class SNP	Class insertion	Class SNP	Class SNP	Class SNP
in the second se	Mar	Location 12:25371190	Location 12: between 25371276 & 25371277	Location 12:25371320	Location 12:25371340	Location 12:25371462
	25.35Mb 25.36Mb 25.3	Alleles T/C	Alleles -ITCTAAAATCAAT.	Alleles C/T	Alleles T/A	Alleles C/T
Maximizer I approved	College descentions	Ambiguity code Y	GIODAI MAF 0.0129695 (TCTAAAATCAATATGAATGTGTCTA)	Ambiguity code Y	Ambiguity W code	Ambiguity Y
vanaron segeno.	Spice color variant	Concession Lintropic	Consequence Intronic	Consequence Untropic	Global MAF 0.0123802 (A)	Globel MAF 0.196885 (T)
	Intron variant	Evidence 1000Genomes	Source dbSNP	Evidence 1000Genomes	Consequence Intronic	Consequence Intronic
	Intercentic variant	Source dushp	Evidence Multiple_observations,	Source dbSNP	Evidence Multiple_observations,	Evidence Multiple_observations,
0		Population genetics	1000Genomes	Population genetics	1000Genomes	1000Genomes, Cited
Gene Legend	Protein Coding		- Obreactori deserve		Source dbSNP	Source dbSNP
	protein cooling				Population genetics	Population genetics
	merged EnsembilHavana					

- Then, from the pop-up box, you can click on "Population genetics" and get more information about allele and genotype frequencies of each variant and population.

rs369251072 insertio	N N						
Original source	Variants (includ	ting SNPs and indel	s) imported from dbS	NP (release 142)	/iew in dbSNP]		
Alleles	-/TCTAAAA	TCAATATGAAT	GTGTCTA MAF: 0	.02 (TCTAAAATCA	ATATGAATGTGTCTA)		
Location	Chromosome 1	12: between 253712	76 and 25371277 (fo	rward strand) View	in location tab		
Most severe consequence	Intron variant	See all predicted of	consequences (Genes	and regulation]			
Evidence status ()	9K 🛕						
HGVS names 🗉	This variation h	nas 4 HGVS names	- click the plus to sho	w			
About this variant	This variant ov	erlaps <u>3 transc</u> ripts	and has 2545 individ	ual genotypes.			
ALL -: 98%	AFR -: 93%	AMR .: 99	EAS	• -: 100%	• -: 100%	SAS	 -: 100%
	Sub-populations	Sub-populations	E Sub-popu	ulations 🗉 🛛 S	iub-populations 🗉	Sub-popu	ulations E
Show All V entries	se 3 (32) ⊟			Show/hide column	3		
Population		Allele: frequenc	y (count)		Genotype:	frequency	(count)
1000GENOMES:phase_3:	ALL	-: 0.981 (4913)	TCTAAAATCA: 0.019 (95)		- -: 0.964 (2	2413)	- TCTAAAAT: 0. (87)
1000GENOMES:phase_3:	AFR	-: 0.935 (1236)	TCTAAAATCA: 0.065 (86)		- -: 0.876 (5	579)	-[TCTAAAAT: 0. (78)
1000GENOMES:phase 3	ACB	-: 0.938 (180)	ТСТААААТСА:		- -: 0.875 (8	34)	-ITCTAAAAT: 0.
			0.062 (12)				(12)

- C) Check the allele frequencies of same genomic region in the ESP 6,500 samples.
- Type the ExAC URL on your browser (<u>http://exac.broadinstitute.org/</u>) and search for 12:25350000-25450000 region.

ExAC Browser (Beta) Exome Aggrega	tion Consortium
12:25350000-25450000 Svampler Cene PCSK0, Transcrift: ENST00000407236, Variant: 22:46615880-T-C, Multi-allelic variant: rs1800234, Re	gion: 22:46615715-46615880
About ExAC	Recent News
The Exome Aggregation Consortium (ExAC) is a coalition of investigators seeking to aggregate and	March 14, 2016
harmonize exome sequencing data from a wide variety of large-scale sequencing projects, and to make summary data available for the wider scientific community.	- Version 0.3.1 ExAC data and browser (beta) is

 In the results page, you can find information about the variants located in the selected region. One of the columns of the table shown is called "Allele frequency".



- D) Finally, check the genetic variation of KRAS in ExAC browser. Which is the allele frequency of rs121913529 in the European (Non-Finnish) population?
- Select the third gene Ensembl ID (ENSG00000133703, KRAS gene) from the previous web page and search for "rs121913529" in the results web page.

12:25398264 A / G	12	25398264	p.Leu19Leu	PASS	synonymous	1	102040	0	0.000009800	
12:25398268 A / G	12	25398268	p.Ser17Ser	PASS	synonymous	1	102060	0	0.000009798	
12:25398279 C / T (rs104894365)	12	25398279	p.Val14lle	PASS	missense	1	101898	0	0.000009814	
12:25398284 C / T (ts121913529)	12	25398284	p.Gly12Asp	PASS	missense	2	101204	0	0.00001976	
12:25398285 C / A (rs121913530)	12	25398285	p.Gly12Cys	PASS	missense	2	101218	0	0.00001976	
12:25398295 T / C (rs147406419)	12	25398295	p.Val8Val	PASS	synonymous	36	98618	0	0.0003650	1011
12:25398321 T / C	12	25398321		PASS	5' UTR	3	83546	0	0.00003591	

- Click on the link and check the European (Non-Finnish) population frequency (1.873e-05).

Filter Status dbSNP	PASS rs121913529	Genotyp	e Quality Metr	ics		
Allele Frequency Allele Count UCSC	1.976e-05 2 / 101204 12-25398284-C-T ☑	Site Qua	ality Metrics			
ClinVar	Click to search for variant in Clinvar					
Annotations his variant falls on 4 trans hissense	cripts in 1 genes:	Population	Allele Count +	Allele Number	 Number of Homozygotes 	+ Allele Frequency
Transprinto				0004		0.0004440
KRAS Transcripts	•	African European (Non- Finnish)	1	8994 53382	0	0.0001112 1.873e-05
KRAS Transcripts	 additional transcripts in the same gene that the variant does not overlap. 	African European (Non- Finnish) East Asian	1 1 0	8994 53382 7960	0 0	0.0001112 1.873e-05 0
KRAS Transcripts	 additional transcripts in the same gene that the variant does not overlap. 	African European (Non- Finnish) East Asian European (Finnish)	1 1 0 0	8994 53382 7960 5844	0 0 0	0.0001112 1.873e-05 0 0
KRAS Transcripts	 additional transcripts in the same gene that the variant does not overlap. 	African European (Non- Finnish) East Aslan European (Finnish) Latino	1 1 0 0 0	8994 53382 7960 5844 10162	0 0 0 0 0	0.0001112 1.873e-05 0 0 0
KRAS Transcripts	 additional transcripts in the same gene that the variant does not overlap. 	African European (Non- Finnish) East Aslan European (Finnish) Latino Other	1 1 0 0 0 0	8994 53382 7960 5844 10162 772	0 0 0 0 0 0	0.0001112 1.873e-05 0 0 0 0 0
KRAS Transcripts	 additional transcripts in the same gene that the variant does not overlap. 	African European (Non- Finnish) East Asian European (Finnish) Latino Other South Asian	1 1 0 0 0 0 0	8994 53382 7960 5844 10162 772 14090	0 0 0 0 0 0 0	0.0001112 1.873e-05 0 0 0 0 0 0

Exercise 4. Browse genomic variation using the CIBERER Spanish Variant Server.

Questions:

- A) Search all the genomic variants of KRAS gene in the Spanish population. How many variants do you find? Now, try again but selecting only the IBS population from the 1,000 Genomes project. How many variants do you find?
- Type the CSVS URL (<u>http://csvs.babelomics.org/</u>) and search for KRAS gene. 51 variants are obtained.

Search									MAF					1000	C MAR (c)	(t and				1000G M	E Inhose 3			FS2 640	0				
	Chr	Position	Alleles	Gene	ы		Genol	types			Freq.		ALL	AME	ASI	AFR	EUR	ALL	AME	South ASI	East ASI	AFR	EUR	Eur. Ame. Al	Y. Ame.	SIFT	POLYPHEN	PhastCons	phylop
						00	0/1	1/1	L	Ofreq	1 Preq	MAP																	
mal Location:	12	2030/662	OI.	LTKMO, KKAS		5//	1	0	0	U.YYY	0.001	0001																0.553	0.655
860,2:1-108000	12	25357893	T≻C	LYRM5,KRAS	rs192368726	577	1	0	0	0.999	0.001	0.001	0.003	0.000	0.000	0.000	0.010	0.002	0.000	0.001	0.000	0.000	0.007					0.040	0.655
	12	25358033	T>A	LYRM5,KRAS	rs11047892	576	1	1	0	0.997	0.003	0.003	0.070	0.070	0.010	0.010	0.160	0.058	0.071	0.079	0.005	0.004	0.153					0.010	-0.353
	12	25358198	ATTAC>	LYRM5,KRAS		577	1	0	0	0.999	0.001	0.001																0.996	0.533
	12	25358301	C≻G	LYRM5,KRAS		513	1	0	64	0.999	0.001	0.001																0.996	0.533
	12	25358368	A>G	LYRM5,KRAS		577	1	0	0	0.999	0.001	0.001																0.996	0.655
ulations 🐨 🗆	12	25358418	T>C	LYRM5,KRAS	rs61764374	540	33	4	1	0.964	0.036	0.036	0.020	0.020	0.000	0.000	0.040	0.017	0.023	0.017	0.001	0.001	0.051					0.997	0.533
267 healthy	12	25358423	≻AA	LYRM5,KRAS		568	10	0	0	0.991	0.009	0.009																0.998	0.655
	12	25358650	A>T	LYRM5,KRAS	r\$12245	568	6	4	0	0.965	0.012	0.012	0.450	0.460	0.230	0.320	0.460	0.452	0.490	0.312	0.583	0.279	0.477					0.001	-0.450
07 Spanish	12	25358663	TD-	LYRM5,KRAS		571	3	4	0	0.99	0.01	0.01																0.372	0.585
ames) ain infectious and diseases plasms eases of the blood I-forming organs and	G	Page 1 enomic co X 毛 1	ontext	ot 6 > > Frequ	encies	Pheno 159	type	Effe 104529-10	ect 04687		So! «	< > »	*														ite	ms 1 -	10 0
sorders involving the nechanism focrine, nutritional	Regic 10039	n overview 73						_	_	_			-	Window	size: 12	71 nts	_	_											

- Select only the IBS subpopulation and click on "Search" button. 14 variants are obtained.

Clear	Search									MAF					10000	MAF (ph	ase 1)				1000G MA	F (phase 3)			ESPO	500				
		Chr	Position	Alleles	Gene	1d		Genot	types			Freq.		AU	AME	ACI	AER	0.10	A11	AME	South AS	East AG	AED	E1 12	Eur Area	Atr Ama	SIFT	POLYPHEN	PhastCons	phyloP
sition							0/0	0/1	1/1	L	OFreq	1.Freq	MAF	ALL	10.12	~~~	Park	Low .	76.5	Para	- South Party	CARLINGT	Part.	Lon	Lot. Parts	No. Port.				
omosomal Locati	ion	12	25358418	THC	LYRMS,KRAS	rs61764374	98	8	1	0	0.953	0.047	0.047	0.020	0.020	0.000	0.000	0.040	0.017	0.023	0.017	0.001	0.001	0.051					0.997	0.533
1-100000,2:1	-100000	12	25358423	>AA	LYRM5,KRAS		104	3	0	0	0.985	0.014	0.014																0.998	0.655
		12	25359046	>A	LYRMS,KRAS		104	3	0	0	0.986	0.014	0.014																0.036	-0.229
e:		12	25359227	A>T	LYRM3,KRAS	rs61764372	106	1	0	0	0.995	0.005	0.005	0.010	0.010	0.000	0.040	0.010	0.034	0.004	0.107	0.000	0.037	0.011					0.979	0.655
6		12	25359328	A>T	LVRM3,KRAS	rs1137189	30	55	22	0	0.537	0.463	0.463	0.450	0.460	0.230	0.310	0.460	0.452	0.403	0.313	0.104	0.279	0.477					0.019	-0.669
		12	25359352	G⊁A	LYRM3,KRAS	rs1137188	30	55	22	0	0.537	0.463	0.463	0.450	0.460	0.230	0.320	0.460	0.451	0.488	0.312	0.183	0.279	0.477					0.513	0.533
bpopulations	80	12	25361074	GFA	LYRM5,KRAS	rs7973623	62	42	3	0	0.776	0.224	0.224	0.150	0.150	0.110	0.120	0.210	0.158	0.187	0.197	0.091	0.120	0.219					0.013	0.528
MGP (267 healt	tw î	12	25361102	>A	LYRM5,KRAS		14	56	37	0	0.393	0.607	0.393																0.007	-0.344
trois)		12	25361142	A+G	LYRM5,KRAS	rs7973450	62	42	э	0	0.776	0.224	0.224	0.170	0.150	0.110	0.180	0.210	0.175	0.193	0.197	0.092	0.180	0.219					0.845	0.557
IBS (107 Spanis)	h	12	25362552	A>C	LYRMS,KRAS	rs712	30	56	21	0	0.542	0.458	0.458	0.470	0.460	0.230	0.210	0.460	0.479	0.494	0.313	0.182	0.177	0.478					0.954	0.383
Ngenomes)	ous and	e e	Page 1		of 2 > >																							it	ens 1 ·	10 of

- B) Which information can we obtain searching the 1:24536 position? (Effect, phenotype, etc.)
- There is no information retrieved by searching for the position 1:24536. Then, we try other genomic region and search for variants included in 12:25368400-25368500.

Clear	Soarch									MAF					1000	CHARGE			1000C MAE (-h 2)							
Great	Jearch	Chr	Position	Alleles	Gene	Id		Genotypes			Freq.				1000	C MAR (D	1406 1/				10000 1994	r (priase 3)				
Position							0/0	0/1	1/1	J.	0 Freq	1 Freq	MAF	ALL	AME	ASI	AFR	EUR	ALL	AME	South ASI	East ASI	AFR			
Chromosomal Location		12	25368410	C>T	KRAS	rs200970347	576	2	0	0	0.998	0.002	0.002	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0		
12:25368400-2536	8500	12	25368462	C>T	KRAS	rs4362222	110	1	467	0	0.191	0.809	0.191	0.000	0.000	0.000	0.000	0.000	0.002	0.003	0.000	0.000	0.008	0		
Gene:																										
BRCA2, PPL																										
Subpopulations	80																									
MGP (267 healthy	î																									
IBS (107 Spanish	- 1																									
1000genomes)		•																					-	-		
I Certain infectious parasitic diseases	sand	« (Page 1		or 1 > >																					
II Neoplasms		G	enomic co	ontext	Frequ	encies	Pheno	type	Ef	fect																
III Diseases of the I and blood-forming org	blood ans and	86	X 🗟 I	Min -		+ Max	159	1:1	004529-:	1004687		Go! «	< > »	, x ^t												
certain disorders invol	lving the	Region	overview												Window	v size: 12	71 nts									

- Select the first one (12:25368410) and explore the phenotype and effect information.

Clear Search								MAF				10000	MAF (shar	ur 1)			1000	G MAF (share	ESP	6500			
	Chr	Position	Alleles	Gene Id		Gen	otypes		F	Freq.	ALL	AME	ASI	AFR	EUR	ALL	AME Sou	h ASI East AS	AFR	EUR	Eur. Ame.	Afr. Ame.	SIFT POLYPHEN
Position					0/0	0/1	1/1	1 1	OFreq 1	Freq MA													
Chromosomal Location:	12	25368410	C>T	KRAS rs2009	0347 576	2	0	0	0.998 0	.002 0.00	0.001	0.000	0.000	0.000	0.001	0.000	0.000 0.	000.000	0.000	0.001	0.000	0.000	
12:25368400-25368500	12	25368462	C>T	KRAS rs436	110	1	467	0	0.191 0	809 0.15	0.000	0.000	0.000	0.000	0.000	0.002	0.003 0.	000.000	0.008	0.000	0.000	0.004	
Gene:																							
BRCA2, PPL																							
Subpopulations 🐨 🗆																							
MGP (267 healthy controls) IBS (107 Spanish individuals from 1000 genomes) C Contain inforctions and																							
individuals from 1000genomes) I Certain infectious and	e e e	Page 1	of	1 > >	_		_					-	-	-					-		_		
Individuals from 1000genomes) I Certain infectious and parasitic diseases I Neoplasms	< « «	Page 1	of	1 > > Frequencies	Phenol	type	Eff	ect													_		
individuals from 1000genomes)	< c Gen	Page 1 enomic co e Name E	of ontext Ensembl Gene I	1 > > Frequencies d Ensembl Transc	Phenot Conseq. type	type Relat	Eff.	ect Coc	don	Strand	В	iotype	cDna F	Position	cds P	osition	AAPosi	tion A	A Change	2	Sift	Pol	pher
Individuals from 1000genomes) I Certain Infectious and parasitic diseases I Neoplasms I Neoplasms I Noiseases of the blood and blood-forming organs and certain disorders involving the	< c Gen K	Page 1 enomic co e Name E RAS E	of ontext Ensembl Gene I	1 > > Frequencies d Ensembl Transc	Phenot Conseq. type intron_variant	type Relat	Eff.	ect Cor	don	Strand -	B	iotype in_coding	cDna F	osition	cds P	osition	AAPosi	tion A.	A Change	2	Sift	Pol	pher
individuals from 1000genomes) [™] I Certain infectious and parasitic diseases [™] III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism [™] IV Reference anticibanal	Gen K	Page 1 enomic co e Name E RAS E RAS E	of ontext Ensembl Gene I INSG0000013: INSG0000013:	1 > > Frequencies d Ensembl Transc. 3. ENST00000311.	Phenot Conseq. type intron_variant intron_variant	type Relat	Eff	ect Coo	don	Strand -	B prote	iotype in_coding in_coding	cDna F	Position	cds P	osition	AA Posi	tion A.	A Change	2	Sift	Pol	pher
Individuals from 1000genomes) I Certain infectious and parasitic diseases II I Neoplasms II Diseases of the blood and blood-forming organs and certain disorders moving the immune mechanism IV Endocrine, nutritional and metabolic diseases	Gen K	Page 1 enomic co e Name E RAS E RAS E	of ontext Insembl Gene I INSG0000013: INSG0000013: INSG0000013:	1 > > Frequencies d Ensembl Transc 3. ENST00000311. 3. ENST00000557 5. ENST00000256.	Phenol Conseq. type intron_variant intron_variant missense_varian	Relat	_Eff	ect Coo	don /Agc	Strand -	B prote prote	iotype in_coding in_coding	cDna F	Position	cds P	tosition	AA Posi 179	tion A	A Change	2	Sift 0.14	Pol	pher 002
Individuals from 1000pronnes) If Certain Infectious and paratitic diseases If IN Recolutions If IN Recolu	Gen K K	Page 1 enomic co e Name E IRAS E IRAS E	of ontext Ensembl Gene I INSG00000133 INSG00000133	1 > > Frequencies d Ensembl Transc 8.ENST00000311. 8.ENST00000256.	Phenol Conseq. type intron_variant intron_variant missense_varian regulatory_regio	type Relati	Eff.	ect Coo	don /Agc	Strand - -	B prote prote	iotype in_coding in_coding in_coding	cDna F	Position 99	cds P	losition	AA Posi 179	tion A	A Change GLY/SER		Sift 0.14	Pol	pher 102

- C) Now, search for BRCA2 gene only in the MGP population. Is there any variant that could be characteristic of the Spanish population?
- Good candidates to be characteristic of a population are those variants that can be found in that population and not in others.

Clear Search	chr	Baritian	Allalar	Gene	м		Gano	have	MAF	Free			1000G MAF (phase 1)							1000G M/	F (phase 3	1		ESP	6500	GET	DOLVENEN	Bhartford	chadaD	
Position	Cit	T GALLAT	PERMIT	Ocin	~	0/0	0/1	1/1	7	Ofreq	1.Freq	MAF	ALL	AME	ASI	AFR	EUR	ALL	AME	South ASI	East ASI	AFR	EUR	Eur. Ame.	Afr. Arrie.				prijest	
Chromosomal Location	13	32903685	C>T	BRCA2		181	59	5	22	0.859	0.141	0.141)		0.279	0.491	Breast-ova
1:1-100000,2:1-100000	13	32905220	T>	BRCA2		244	23	0	0	0.957	0.043	0.043								-		355						0.007	-1.232	
	13	32906480	A>C	BRCA2	rs766173	244	23	0	0	0.957	0.043	0.043	0.060	0.090	0.100	0.010	0.040	0.074	0.092	0.133	0.096	0.033	0.035	0.037	0.020	0.12 (tol.	0.052 (ben	0.361	0.533	not specify
Gene:	13	32906571	A>C	BRCA2	rs55939572	265	2	0	0	0.996	0.004	0.004												0.000	0.000	0.08 (tol.	0.713 (pos	0.008	0.533	Breast-ova
BRCA2	13	32906729	A>C	BRCA2	rs144848	144	96	27	0	0.719	0.281	0.281	0.240	0.310	0.260	0.100	0.290	0.249	0.300	0.354	0.285	0.084	0.295	0.286	0.129	0.14 (tol.	0.022 (ben.	0.297	-0.333	not specifie
	13	32906980	A>G	BRCA2	rs1801439	244	23	0	0	0.957	0.043	0.043	0.060	0.090	0.100	0.010	0.040	0.074	0.092	0.133	0.096	0.033	0.035	0.037	0.020			0.231	0.655	Breast-ova
Subpopulations 🛛 🖓 🗆	13	32907129	T+C	BRCA2	n28897708	265	1	0	0	0.998	0.002	0.002	0.001	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.000	0.15 (tol.	0.037 (ben	0.419	0.528	Breast can
MGP (267 healthy	13	32907401	G>C	BRCA2	rs56328701	258	1	0	8	0.998	0.002	0.002												0.000	0.000	0 (delete	0.864 (pos	0.920	0.528	Breast can
controls)	13	32907536	эT	BRCA2		210	55	2	0	0.89	0.11	0.11																0.002	-0.378	
BS (107 Spanish	13	32907536	TH	BRCA2		210	35	2	0	0.87	0.11	0.11																0.002	-0.378	
1000genomes)																														
I Certain infectious and	<	Page 2		of 7 > >																							iter	15 11 -	20 of	70 =