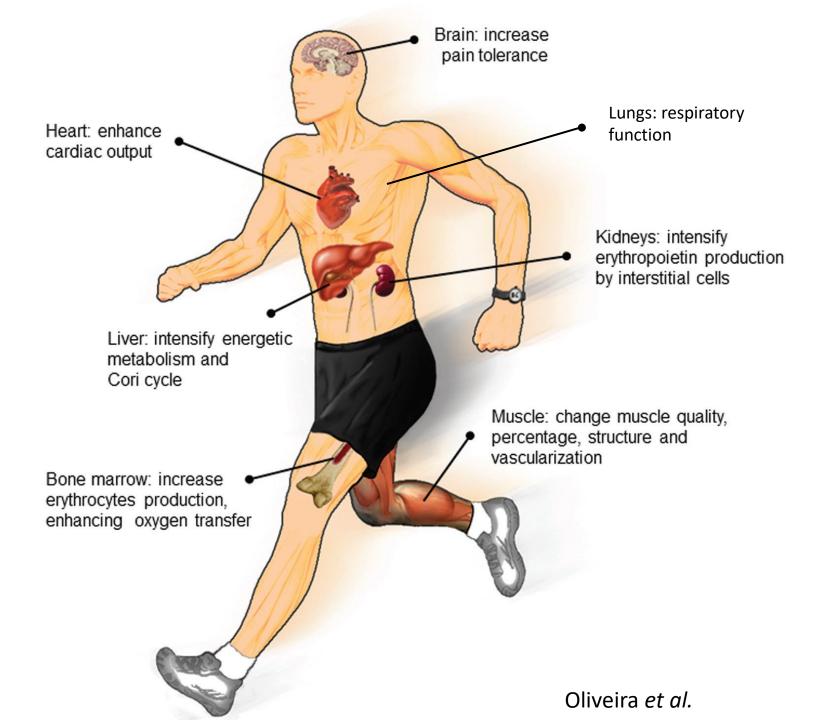


Image: http://www.bio.georgiasouthern.edu

Genetic variations and athletic performance?





IJPH - Year 9, Volume 8, Number 3, 2011

ITALIAN JOURNAL OF PUBLIC HEALTH

Athleticogenomics and elite athletes: a review of the state of the art and a possible relationship with inflammatory response

Alessandra Colombini¹, Giovanni Lombardi¹, Giuseppe Banfi^{1,2}, Marisa Arpesella³, Gabriele Pelissero^{3,4}

¹I.R.C.C.S. Istituto Ortopedico Galeazzi, Milano, Italy; ²School of Medicine, University of Milan, Milano, Italy; ³Dipartimento di Medicina Preventiva Occupazionale e di Comunità - Sezione di igiene, University of Pavia, Pavia, Italy; ⁴I.R.C.C.S. Policlinico San Donato, San Donato Milanese, Milano, Italy

Correspondence to: Giuseppe Banfi, I.R.C.C.S. Galeazzi and School of Medicine, University of Milan, via R. Galeazzi 4, 20161 Milano, Italy. E-mail: giuseppe.banfi1@unimi.it

Abstract

Background: Recent research in athleticogenomics has begun to reveal how particular genetic polymorphisms may influence athletic status and confer an individual predisposition for better sports performances. This is of particular interest for elite athletes because it could help to assess an athlete's potential, to enhance specific training protocols for selected performances, to monitor the individual response to training load and recovery and, finally, aid in the prevention of accidents.

Methods: Using a topics search in the PubMed database, the search strategy included studies examining the relationship between the presence of polymorphisms in genes influencing selected physiological parameters and the elite athletic status. English written case (elite athletes) -control (general population) studies were selected.

Results: 26 research articles concerning polymorphic genes involved in muscle physiology, cell respiratory

IJPH - Year 9, Volume 8, Number 3, 2011

ITALIAN JOURNAL OF PUBLIC HEALTH

Table 1. The main polymorphic genes in nuclear DNA analyzed in elite athletes and the more significant case-control studies of their association with sports performance. (Continue)

Gene	Location	Polymorphism	Number of subjects		P Value (genotype/ allele frequency)	Reference	Year
			Elite atbletes	Controls			
<i>ADRA2A</i> (alpha2a adrenergic receptor)	10q24-q26	Dra I RFLP	140	141	0.037	29	2000
ADRB2 (beta2 adrenergic receptor)	5q31-q32	Arg16Gly polymorphism	303	297	0.03	30	2007
GNB3 (guanine nucleotide binding protein β polypeptide 3)	12p13	exon 10 C825T polymorphism	155	234	0.046	31	2009



Braz J Med Biol Res, December 2011, Volume 44(12) 1194-1201

doi: 10.1590/S0100-879X2011007500145

The use of genes for performance enhancement: doping or therapy?

R.S. Oliveira, T.F. Collares, K.R. Smith, T.V. Collares and F.K. Seixas

"Gene doping"

ORIGINAL ARTICLE

Endocrine Care

Doping Test Results Dependent on Genotype of Uridine Diphospho-Glucuronosyl Transferase 2B17, the Major Enzyme for Testosterone Glucuronidation

Jenny Jakobsson Schulze, Jonas Lundmark, Mats Garle, Ilona Skilving, Lena Ekström, and Anders Rane

Department of Laboratory Medicine, Karolinska Institutet at Division of Clinical Pharmacology, Karolinska University Hospital, SE-141 86 Stockholm, Sweden

Context: Testosterone abuse is conventionally assessed by the urinary testosterone/epitestosterone (T/E) ratio, levels above 4.0 being considered suspicious. The large variation in testosterone glucuronide (TG) excretion and its strong association with a deletion polymorphism in the uridine diphospho-glucuronosyl transferase (UGT) 2B17 gene challenge the accuracy of the T/E ratio test.

Objective: Our objective was to investigate whether genotype-based cutoff values will improve the sensitivity and specificity of the test.

Design: This was an open three-armed comparative study.

Participants: A total of 55 healthy male volunteers with either two, one, or no allele [*insertion*/ *insertion*, *insertion*/*deletion*, or deletion/deletion (*del/del*)] of the *UGT2B17* gene was included in the study.

Intervention: A single im dose of 500 mg testosterone enanthate was administered.

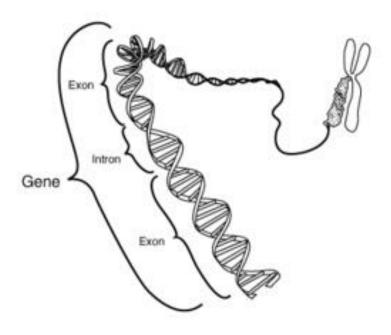
Contents

- 1. Basics of genetic variation
- 2. Technology to measure variation
- 3. Linking SNPs to traits
- 4. SNPs in genome browsers

1. Basics of genetic variation

Genetic variations

- In human beings, 99.9 percent of the bases are the same.
- Remaining 0.1 percent makes a person unique.
 - Different attributes / characteristics / traits
 - how a person looks
 - diseases he or she develops
- Most of those variations are in non-coding regions
 - This does not necessarily mean they have no effect!

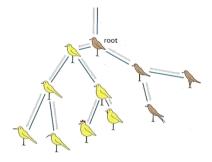


Genetic variations (II)

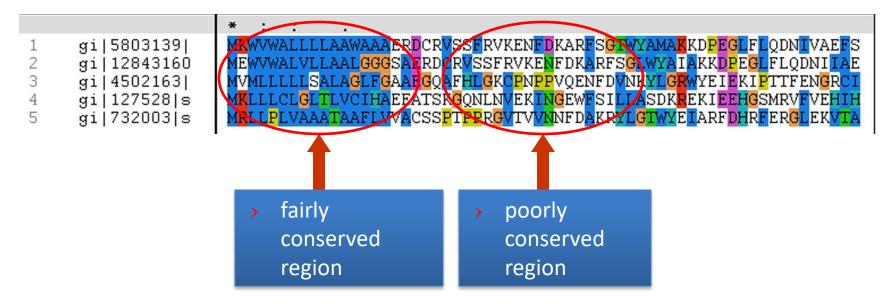
- Variations can be:
 - Harmless (change in phenotype)
 - Harmful (diabetes, cancer, heart disease, Huntington's disease, and hemophilia)
 - Latent (variations found in coding and regulatory regions, are not harmful on their own, and the change in each gene only becomes apparent under certain conditions e.g. susceptibility to lung cancer)
- A variation is called a mutation if a disadvantageous effect on disease is proven
 - This is far from easy

Genetic variations and evolution

- (Advantageous) variations drive evolution
- The less variation in a region or the more conserved the region is – the more important we can assume it to be

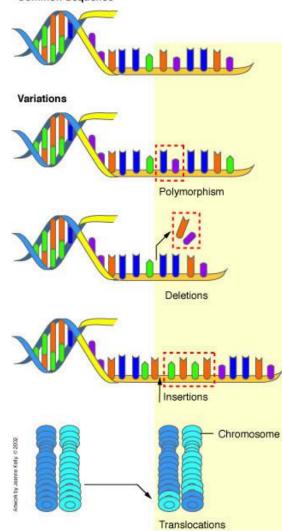


- Higher evolutionary pressure
- Functional domains of proteins, regulatory elements in promotors, ...



Types of genetic variations

Common Sequence



Other variations in genes

- Apart from variations in the sequence of the genes, other (inheritable) variations occur
 - These are called **epi**genetic variations

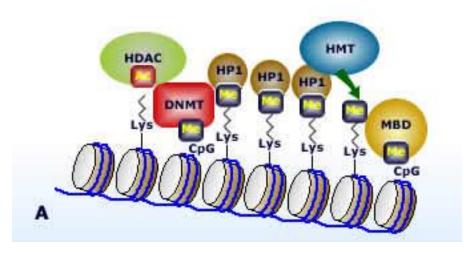


Image: http://www.avernes.fr

Single Nucleotide Polymorphisms (SNP)

• A SNP (single nucleotide polymorphism) is defined as a single base change in a DNA sequence *that occurs in a significant proportion* (more than 1 percent) of a large population

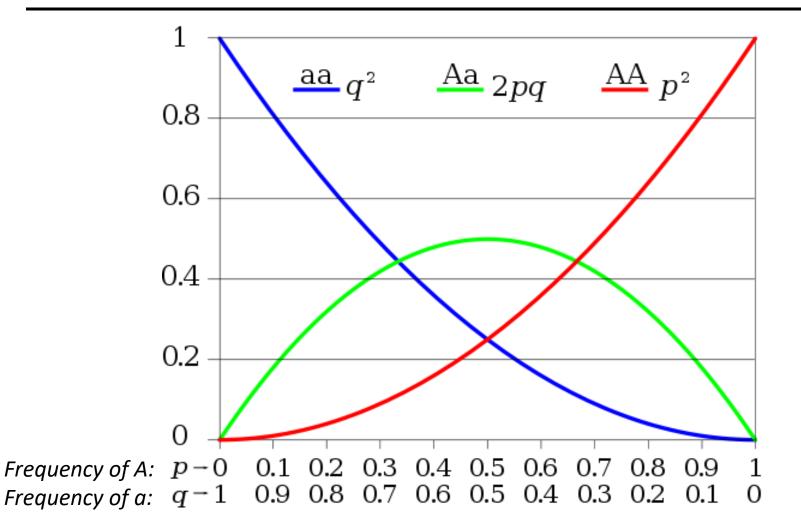


- Currently, dbSNP at NCBI (build 132) has about 7 million human SNPs (4.5 million validated refSNPs)
 - The minimal frequency criterion is <u>not</u> used

To refresh memory

- The two copies of a gene we have (each of both chromosomes) are called **alleles**
- In case of a SNP the nucleotide at a position can be different between the alleles of an individual
- The combination of alleles one has, is called the **genotype**
- The related trait one has is called the **phenotype**
- Allele frequency: the frequency of occurence of a specific allele
- **Penetrance**: the number of people with a certain genotype that also develop the associated phenotype

Hardy-Weinberg principle

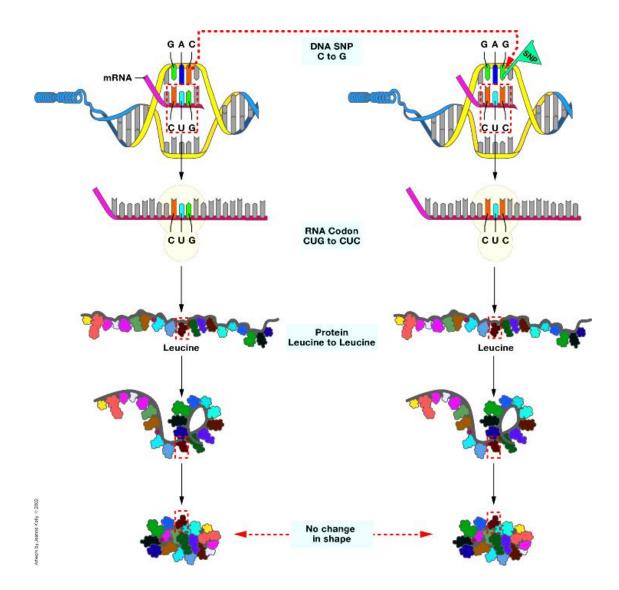


For SNPs we can check whether the observerd allele frequencies correspond to the expected frequencies in HW equilibrium; if not there may be evolutionary selection for certain genotypes



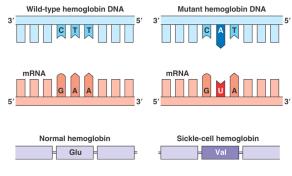
- SNPs are found in
 - coding and (mostly) non-coding regions.
- Occur with a very high frequency
 - about 1 in 1000 bases to 1 in 100 to 300 bases.
- The abundance of SNPs and the ease with which they can be measured make these genetic variations significant.
- SNPs close to a particular gene acts as a **genetic marker** for that gene.
- SNPs in coding regions alter the protein sequence made by that coding region:
 - **Synonymous** SNP: no protein sequence alteration
 - Non-synonymous SNP: protein sequence alteration -> aka as missense mutation.

SNPs may / may not alter protein structure



Functional effects of SNPs

- When amino acid (AA) changes, is the change relevant?
 - Type of AA
 - Site of the change
 - Functional domain
 - Conservation in other species
 - Truncating mutation



http://avonapbio.pbworks.com

- When is a variation a mutation?
 - In fact functional proof is needed (functional testing protein, enzymes, protein-protein binding, cellular localization, make KO cells, substitute defect)

Functional effects of SNPs

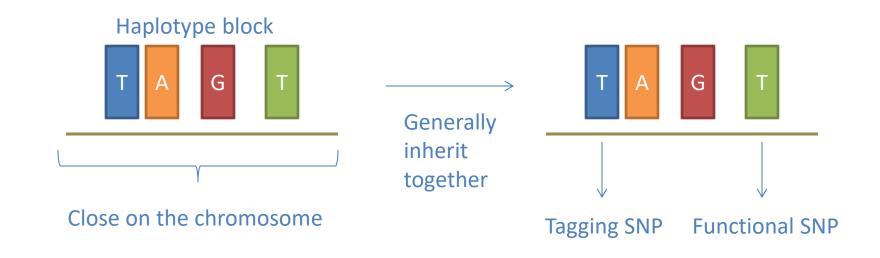
- Also non-coding SNPs may have an effect:
 - Effect on target sites of Transcription Factors (regulation of transcription)
 - Effect on target sites of miRNAs (regulation of transcript decay)
 - effect on splice donor or acceptor sites (regulation of alternative splicing)
 - Other...

Predicting functional effects of SNPs

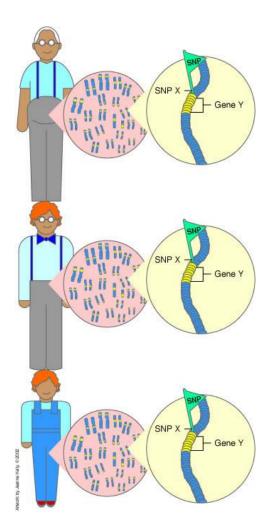
- **SIFT** is a tool that predicts whether an amino acid change affects protein function, based on the conservation of amino acid residues in closely related sequences
- Polyphen uses annotations (e.g. is the position in a known binding site), 3D-structural (if available) and sequence alignment information to predict the effect of an amino acid change
 - → Both only work for non-synonymous coding SNPs

Haplotypes

- The combination of a number of SNPs one has is called a **haplotype**
- SNPs that are close on the genome form haplotype blocks, as they inherit as a group due to the mechanism of cross-over and recombination
- A SNP that is informative for a haplotype block is called a tagging SNP
- A tagging SNP may have no direct function, but still be predictive as it may correlate with functional SNPs



SNPs act as gene markers



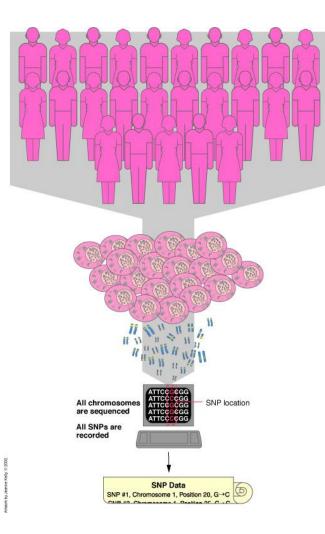
More on markers

- Several types of genetic polymorphic markers can be used to identify chromosomes, alleles, samples:
 - SNPs
 - SNPs with more than 2 alleles are of special interest as markers
 - Tandem repeats (microsatellites)
 - These can vary in number of repeats (repeat length)
 - Longer variabele sequences of DNA
 - Satellites



- Sequence genomes of a large number of people
- Compare the base sequences to discover SNPs
- Generate a single map of the human genome containing all possible SNPs \rightarrow SNP maps
- Examples:
 - Hapmap project
 - 1000 Genomes

SNP Maps (II)





- Millions of SNPs selected and measured in hundreds of people (minor allele frequency > 1%)
- From several populations all over the world
 - Nigeria (YRI)
 - Americans from European origin (CEU)
 - Tokyo residents (JPT)
 - Beijing residents (CHB)





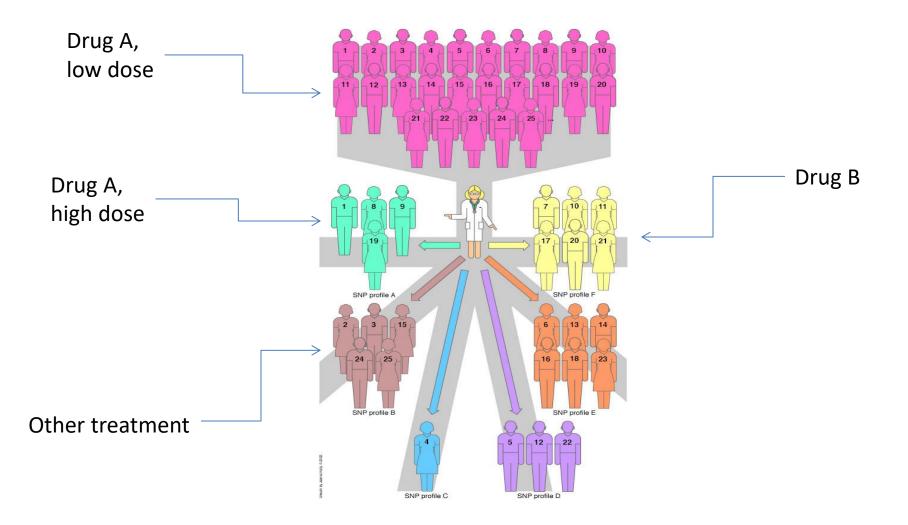
- Sequence the full DNA of many people
 - This also reveals all polymorphisms with frequency > 1%
- 2,500 people sequenced
- Haplotype block information used to complete predictions





- Genome of each individual contains distinct SNP pattern.
- People can be grouped based on the SNP profile.
- SNPs Profiles important for identifying response to Drug Therapy:
 - Correlations might emerge between certain SNP profiles and specific responses to treatment.
 - Personalized medicine

SNP Profiles (II)



SNP Profiles (II)

You can already get your own SNP profile (for selected SNPs)

– 23andme – deCODEme – Navigenics – ...

- Costs: few hundred Euros
- Personalised health risk estimates
- Ancestry
- Knome even determines DNA sequence
 - Costs: few thousand Euros

Collection of 23andme data



Data/Code

Genomes Unzipped is a collaborative online project aiming to provide genetic testing customers with the knowledge and tools they need to make the most of their own genetic data. As part of the project our members are taking commercial genetic tests and making the raw data publicly available for others to <u>download</u>, analyse and reuse. There are currently three ways for users to access our genotype data.

1: Our Genome Browser

You can download our genotype data for particular parts of the genome, using the "Download Genotypes" link on the Genome Browser. Download genotype data for a particular SNP by clicking on it.

2: Download Complete Data

You can download all genotypes for particular individuals from our Amazon S3 genotype bucket, from the following locations:

23andMe v2 data:

- Daniel MacArthur (DGM001)
- Luke Jostins (LXJ001)
- Dan Vorhaus (DBV001)
- Caroline Wright (CFW001)
- Kate Morley (KIM001)
- Vincent Plagnol (VXP001)

About

Genomes Unzipped is a group blog providing expert, independent commentary on the personal genomics industry.

About The Project About The Contributors

Subscribe

Recent Posts

f 100M for whole patient genomes - an implementation challenge f 100M for whole patient genomes - revolutionising genetic diagnostics or squandering NHS cash? Crowd-funding personalized bioscience Response to "Exaggerations and errors in the promotion of genetic ancestry testing" Exaggerations and errors in the promotion of genetic ancestry testing

Categories Admin (7)

Resources for analyzing your own genome

- 1. <u>SNPedia</u>: a community-annotated wiki containing <u>information</u> on traits associated with single nucleotide polymorphisms.
- 2. <u>Promethease</u>: a freely downloadable tool for <u>analysing</u> 23andMe, deCODEme or Navigenics raw data with information from SNPedia.
- 3. <u>Catalogue of genome-wide association studies</u>: produced by the National Human Genome Research Institute.
- 4. <u>DIYgenomics</u>: A website providing apps and tools to compare predicted disease risk from different personal genomics companies.
- 5. <u>Ensembl</u> and <u>UCSC</u>: genome browsers containing information on genes, variants and associated phenotypes. Can be somewhat complex to use.
- <u>SNPtips</u>: a FireFox extension that allows you to rapidly check your
 23andMe genotype for any SNP you see mentioned on a web page.

Personalized medicine

- **Concept**: use information about a patient's genome, epigenome, proteome, etc. to adapt medical care:
 - select between different medications
 - optimize drug dosage
 - provide a specific therapy
- This **omics-centered approach** is not yet in widespread use clinically.
- **Important unanswered question**: does personalized medicine really offer any significant advantages over the traditional combination of clinical approaches (medical history, family history, data from imaging, laboratory, and other tests)

Personalized food

- Same concept as personalized medicine: use omics information of an individual or specific group of individuals to "optimize" nutrition
- Optimization on product level:
 - "You should eat more kiwi's"
 - "You shouldn't eat kiwi's at all"
- Optimization on molecular level:
 - "You should take extra vitamin B12"
 - "Taking extra vitamin B12 is dangerous for you"
- Big business for food companies

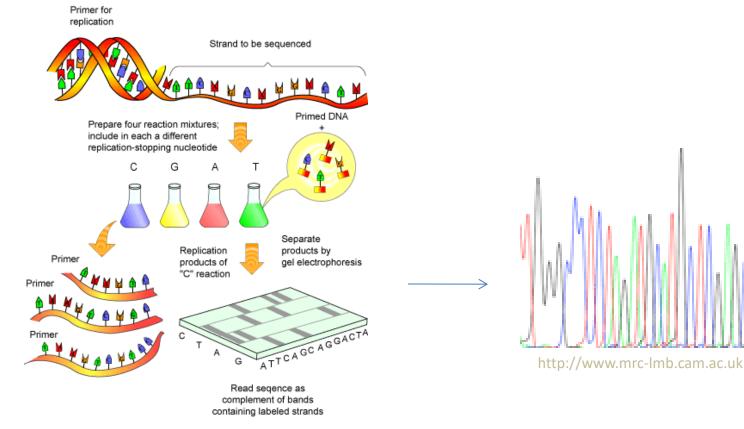


2. Technology to measure variation

Measuring SNPs

Sanger sequencing

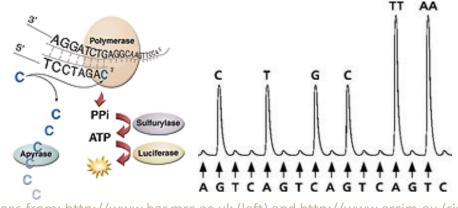
terminates the chain with incorporation of a ddNT



http://www.scq.ubc.ca

Measuring SNPs

- SNPs can be measured using several technologies
 - Pyrosequencing
 - detects pyrophosphate (light)



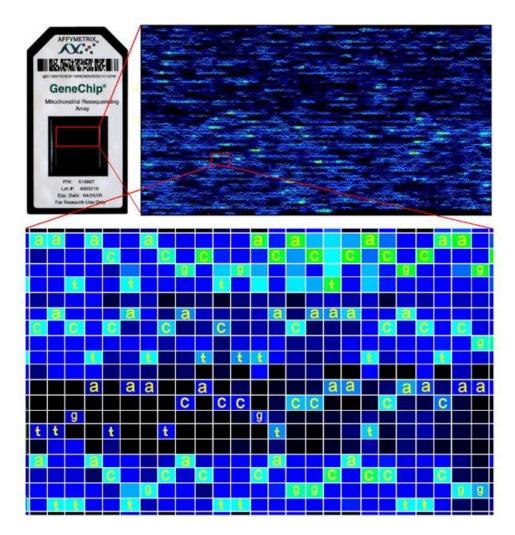
Images from: http://www.har.mrc.ac.uk (left) and http://www.ercim.eu (right)

Marker sets

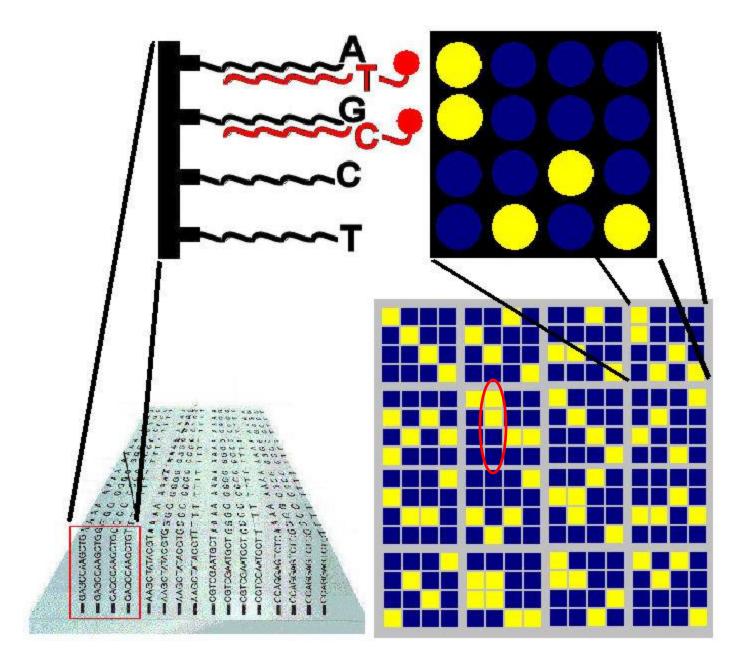
- A way to measure genetic variation on a more global scale is the construction of genetic markers sets
 - Covering the whole genome or part of it
 - Often, microsatellite markers are used for construction of these sets
 - Regions in which changes are observed can be explored further by sequencing or SNPs
- Marker sets are also used in forensics or in paternity determination
- Less used in other applications with the development of large scale technologies

Large scale measurement of SNPs

- Affymetrix SNP chip
- 500,000 or
 1M SNPs
- Genome wide studies
- Data analysis?



(SM Carr et al. 2008. Comp. Biochem. Physiol. D, 3:11)



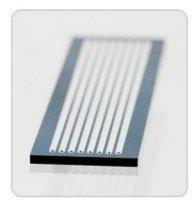
(after SM Carr *et al.* 2008. Comp. Biochem. Physiol. D, 3:11) http://www.mun.ca/biology/scarr/DNA_Chips.html

Resequencing chips

- Another type of chip allows sequencing genes or genomic regions of interest
 - One can design the chips depending on the genes of interest
 - As such one can measure all known mutations related to a disease, also the yet unknown SNPs (in these genes)

Sequencing the whole genome

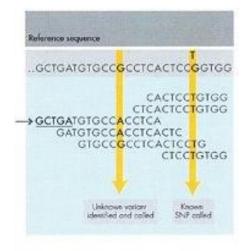
- Next Generation Sequecing (NGS) has made it possible to sequence the whole genome of an organism
 - In principle, all variations between individuals can be determined
 - In any case: massive amounts of data are generated (Gbs per sample)



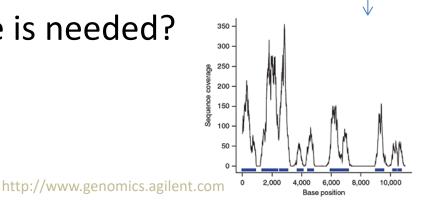
Up to eight samples can be loaded onto the flow cell for simultaneous analysis http://seganswers.com on the Illumina Genome Analyzer.

Sequencing the whole genome

- Data analysis is not that easy
 - Aligning
 - Calling ('peak' calling)
 - Real changes or sequencing errors
 - Error file
 - Same issue with 'regular' sequencing, but there one can evaluate by eyesight
 - How many fold coverage is needed?



http://seqanswers.com



3. Linking SNPs to traits

Traits

- A trait is just a characteristic
 Length, weights, eye color, sex, ...
- Traits can be discrete (sex, ...) or continuous (weight, ...)
- Discrete = 'quantitative'
- Continuous = 'qualitative'



http://phe.rockefeller.edu

Heritability

- This term is often MISinterpreted
- The heritability of a trait means how much of its variation can be explained by genetic variation
 - ... in the population in which it is measured
 - Thus high heritability does not mean that the trait is genetically determined in general

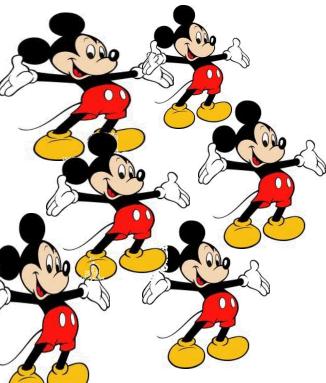
Heritability

 The more genetically uniform the population (e.g. inbred strains) the the heritability

of the trait

All kinds of mice -genetic variation contributes to differences in phenotype -environment does too

Mickey clones -no genetic variation -all differences in phenotype must be due to environment



Heritability

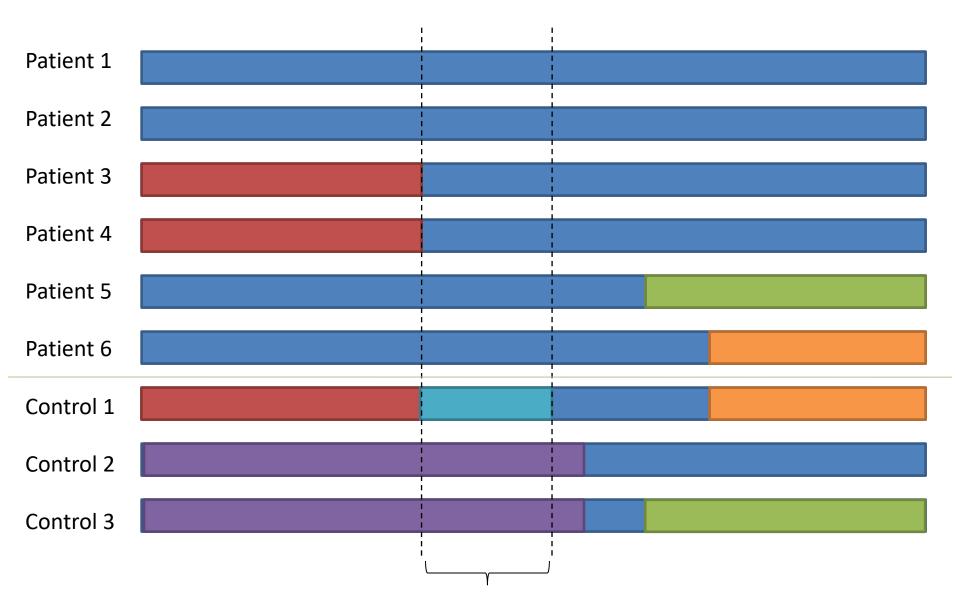
 Heritability estimates in your population can tell you whether you can find possible genetic factors contributing to a phenotype using the population of choice

Genome wide association studies

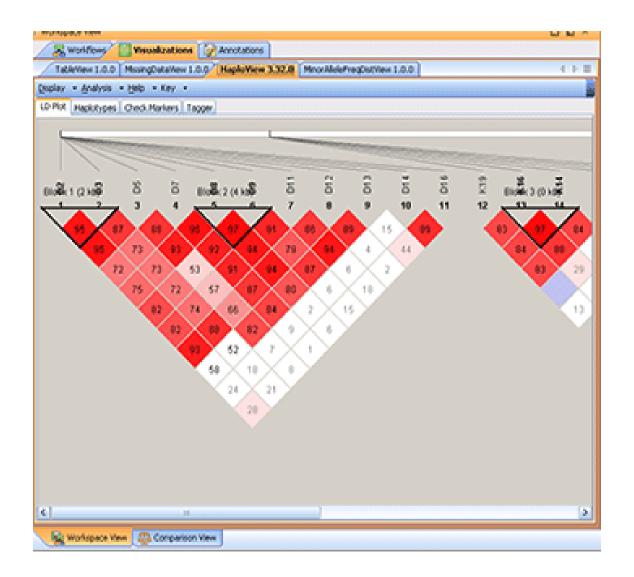
- **GWAS** ('association') tries to link SNPs to traits (diseases) in a genome wide way
- Makes use of *unrelated* individuals
 - So no family members
- Tries to find which allelic variants, correlate with the phenotype of interest
- If a complete haplotype goes together with the phenotype, this is considered association
 - Tools such as HaploView¹ can assist

¹ www.broad.mit.edu/mpg/haploview/

Each color indicates a different haplotype in the study population



Region of interest (determine in more detail, or check genes it contains)



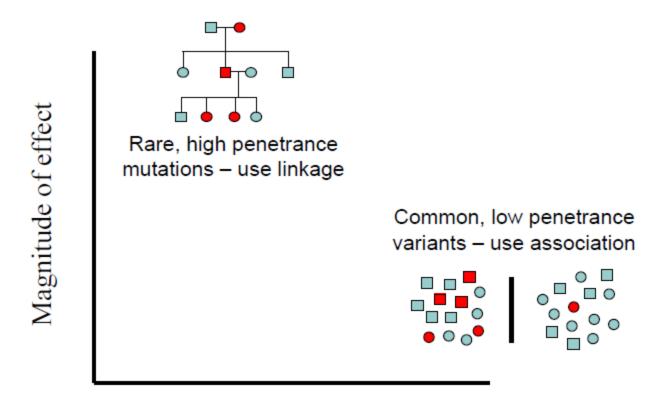
http://www.htbiology.com

Linkage studies

- Linkage makes use of related individuals

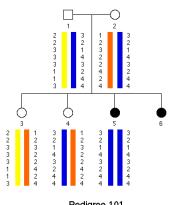
 family members
- Adantage is higher power as compared to GWAS
- But one needs large enough families with enough (informative) 'cross overs' and preferably several generations
- Principle is the same as with GWAS, using markers or SNPs

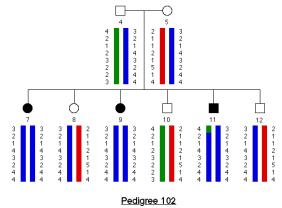
When linkage, when association?

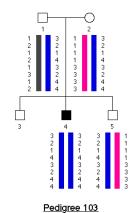


Frequency in population

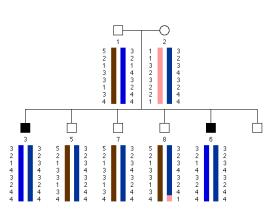
Image: University of Michigan School of Public Health



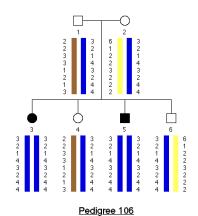


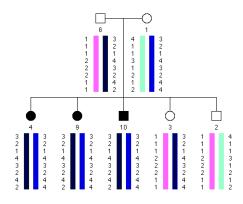


Pedigree 101

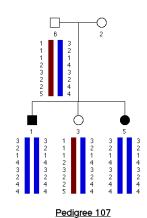








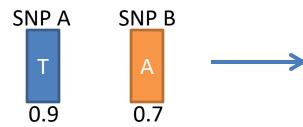
Pedigree 105



http://www.molvis.org

Computations

- The linkage disequilibrium (D or LD) indicates the deviation of a haplotype's frequency from its expected frequency
 - The expected frequency for each combination of SNP A and SNP B is the multiplication of the separate frequencies



Frequency of **T** for SNP A = 0.9 Frequency of **A** for SNP B = 0.7 Expected frequency for both = 0.9 x 0.7 = 0.63

 When this does **not** hold the SNPs are in a haplotype block, also called 'in linkage disequilibrium'

Computations (II)

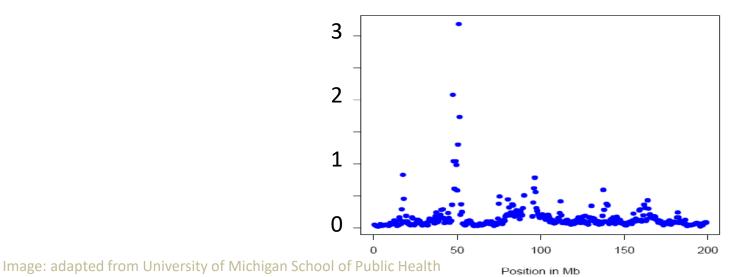
- Disadvantage of the LD: it has a range that depends on the allele frequencies
 Alifi cult to compare hotomorphic to be allocated.
 - \rightarrow difficult to compare between haplotype blocks
- **D'** is an adapted <u>scaled</u> score, to correct for the influence of allele frequences
 - It ranges between -1 and +1
 - Disadvantages: small sample size or low allele frequences tend to overestimate the score

Computations (III)

- An alternative way of correcting is computing the correlation coefficient (r) between loci, also sometimes given as its square (r²)
 - more stringent than D'
 - r^2 ranges between 0 and 1

Computations (IV)

- The LOD score (¹⁰log of the odds) indicates the likelihood of obtaining the data given that the loci are indeed linked, versus obtaining the data by chance
 - A score higher than 3 (which means a 1000:1 odds) is considered evidence of linkage



Limitations

- A very large sample size is needed but also population uniformity

 Trade-off
- To most common diseases, many SNPs/genes contribute for a few percent each
 - Difficult to detect
- Often many genes in haplotype blocks, which one is linked to the disease (contains the relevant SNP(s))
- Rare alleles make sampling even more difficult

 \rightarrow often discrepancy even between large studies

Strategies

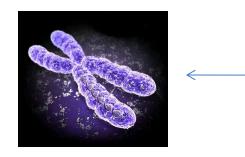
• Populations to use to increase power:

Families (linkage more power)

- Simplify by candidate gene approach
 - Rank the genes in a linked genomic region by their likelihood of involvement in the disease, based on knowledge on their biological functions and/or literature
- Scan globally, and scan regions of interest in more detail

What's more?

- Always realise that the SNPs linked to the phenotype are not (neccesarily) the functional or causing SNPs, they are just close enough to be markers
- Now we only discussed <u>genetic</u> contributions to a phenotype
- Other aspects to study:
 - Genes modifying the effects of other genes (epistatis)
 - Gene-environment interactions





- ADE models can help in estimating genetic and environmental contributions (here twins are of interest)
- Specific study of interactions is very difficult
 - Even more possibilities
 - Even smaller effects

Images: http://theosophical.wordpress.com (left) and http://www.foodfacts.info (right)

4. SNPs in genome browsers

SNP IDs

- A standard ID for SNPs is the dbSNP ID
 - also called "rs number"
 - example: rs4986852
 - Standardised, unique, stable
- An alternative for disease related SNPs is the OMIM variation ID
 - example: 113705.0011
 - Standardised, unique, stable
- A final possibility is the
 - For non-coding or coding SNPs: variation
 - Example: BRCA1, 2978G>A
 - For coding SNPs (also): mutation
 - Example: BRCA1, SER1040ASN
 - Easier to interpret, but not stable

Investigating SNPs in NCBI (Entrez SNP)

gene model	Contig Label	Cont	g	mrna		protein	mrna orientation	transcript s	snp count
(contig mRNA transcript):	GRCh37	NT_0107	<u>33.15 NN</u>	<u>vi_007294.3</u>	<u>NP</u>	009225.1	forward	plus strand {	53, coding

Region	Chr. position	mRNA pos		Hetero- zygosity	<u>Validation</u>	MAF	Allele origin	3D	Linkout	Function		Protein residue	Codon	Amino acid pos	PubMed
	<u>41197708</u>	<u>5811</u>	<u>rs116185952</u>	0.222	Μ 🖬					missense	С	Pro [P]	2	1860	
										contig reference	A	His [H]	2	<u>1860</u>	
	<u>41197716</u>	<u>5803</u>	rs28897699	0.013	% ×			Yes		missense	Т	His [H]	3	1857	
										contig reference	G	Gln [Q]	3	<u>1857</u>	
	<u>41197816</u>	<u>5703</u>	rs70953662	N.D.				Yes		missense	A	Asn [N]	2	1824	
										contig reference	т	lle [l]	2	<u>1824</u>	
	<u>41199693</u>	<u>5666</u>	<u>rs1800751</u>	N.D.	3			Yes		missense	G	Ala [A]	1	1812	E
										contig reference	с	Pro [P]	1	<u>1812</u>	
	<u>41199694</u>	<u>5665</u>	<u>rs4438367</u>	N.D.	3			Yes		synonymous	A	GIn [Q]	3	1811	
										contig reference	G	Gln [Q]	3	<u>1811</u>	
	<u>41203086</u>	<u>5558</u>	<u>rs1800757</u>	N.D.				Yes		missense	Т	Ser [S]	1	1776	
										contig reference	с	Pro [P]	1	<u>1776</u>	
	<u>41209082</u>	<u>5496</u>	<u>rs76171189</u>	N.D.						frame shift			2	1755	
										contig reference	с	Ser [S]	2	<u>1755</u>	
	<u>41215360</u>	<u>5414</u>	<u>rs34570933</u>	N.D.				Yes		frame shift		Met [M]	1	1728	
										frame shift	A	Asn [N]	1	<u> 1728 –</u>	

Investigating SNPs in Ensembl

<i>C</i> Ensembl	· · ·	<u> </u>	, ,			<u> </u>	Q
Home > Human	5.857 Gene: PTPN22	3				Login / Register E	BLAST/BLAT BioMart Docs & FAQs
Location: 1:114,157,963-114,21 Gene: PTPN22 Gene summary Splice variants (3) Supporting evidence Sequence External references (5) Regulation Genomic alignments (37) Gene Tree Gene Tree (text) Gene Tree (alignment) Orthologues (12) Protein families (1) Genetic Variation Variation Table	Gene: PTPN22 (E Tyrosine-protein phos Location Tra	sphatase non-receptor type 22 () Chromosome 1: 114.157.963-1 Click on PTPN22-202 I Gene level displays	(B/Swiss-Prot Q9Y2R2				
Variation image External Data D History Gene history Configure this page Add custom data to page Export data	This view is a gene information click on	nenu at the left hand side of the	page. To return to viewing gene level Splice variants »				
Bookmark this page	Synonyms CCDS Gene type Prediction Method Transcripts	This gene is a member of the H Known protein coding	ew all Ensembl genes linked to the name <u>o</u> Human CCDS set: <u>CCDS863, CCDS86</u> I genebuild transcripts and <u>Havana</u> m	<u>64</u>			
	Ensembl/Havana Contigs Ensembl/Havana			114.17 Mb AL13 < PTPN22-202 Known protein cod		114.20 Mb HIPK1-202 > Known protein	Forward st

Investigating SNPs in Ensembl

<i>C</i> Ensembl										2 -
Home > Human										Login / Register BLAST/BLAT BioMart Docs & F
Location: 1:114,157,963-114,21	5.857 Gener PTP	N22								Login / Register BLASI/BLAT Blowart Docs & P
Gene: PTPN22		2 (ENSG00000134242)								
- Gene summary			C 3.1.3.48)(Hematopoietic cell protein-tyro	sine phosp	hatase 70Z-PEP	(Lymphoid pl	hosphatase)(LvP) Sou	rce: UniProtK	B/Swiss-Prot Q9Y2R2	
 Splice variants (3) 	Location	Chromosome 1: 114,157,963-1								
 Supporting evidence Sequence 	Transcripts									
- External references (5)		Name	Transcript ID Protein I)	Description					
Regulation Comparative Genomics		PTPN22-001	ENST00000359785 ENSP0000035	2833	protein_codin	g		~ 1		
- Genomic alignments (37)		PTPN22-201	ENST00000307489 ENSP0000030		protein_codin	g		Sh	ows all SNPs in tl	hic gongl
E· Gene Tree ⊢ Gene Tree (text)		PTPN22-202	ENST00000354605 ENSP0000034	6621	protein_codin	g		JII		ins gene:
Gene Tree (alignment)		🕕 Transcript and Ge	ne level displays							
 Orthologues (41) Paralogues (12) 		In Ensembl a gene is ma	de up of one or more transcripts. We prov	ide displays	s at two levels:					
Protein families (1)		Transcript view	s which provide information specific to ar	individual t	ranscrint such a	s the cDNA a	and CDS sequences ar	nd protein d	omain apportation	
Genetic Variation Variation Table			ich provide displays for data associated							
Variation Image				-		-				
- External Data			I view. To access the transcript level disp Sene tab in the menu bar at the top of the		a Transcript ID II	n the table ab	ove and then navigate	e to the info	mation you want using the menu at the left hand side of the page sturn to	viewing gene level
ID History Gene history				age.						
 Configure this page 	« Protein	1 families					Variation Table	he!p		Variation Image »
 Add custom data to page 	Variation	00307489:								
 Export data Bookmark this page 	Variation	00001400.								
	ID		Chr: bp	Alleles	Ambiguity	AA change	AA co-ordinate	Class	Sou	Validation
	rs958008	3PRIME_UTR	1: 114158106	C/T	Y	-	-	snp	HGVbase, dbSNP, TSC	hapmap
	rs3811021	3PRIME_UTR		A/G	R	-	-	snp	Affy GeneChip 100K Array, HC ASEMBL:celera, dbSNP	cluster, freq, hapmap
	rs34857282	3PRIME_UTR 3PRIME_UTR	1: 114158464-114156463	G/T -/G	К	-	-	snp insertion	HGVbase, dbSNP, TSC dbSNP	hapmap
	1834037202	SPRIME_UTR	1.114130404-114130403	-/G A/T	w	-	-	snp	dbSNP	-
				G/A	R		-	snp	ENSEMBL:Watson, HGVpase, dbSNP, TSC, ENSEMBL:Venter	cluster, freg, doublehit
Solort	"Conf	igure nage"	to select for	с/т	Y	-	-	snp	dbSNP	-
Juicu	Com	iguic page		A/T	W	-	-	snp	dbSNP	-
				T/A	W	٧L	174 (1)	snp	dbSNP	-
exa	mple (only non-syi	nonymous	C/G	S	-	-	snp	ENSEMBL:Watson	-
0/10		e,e e,.		C/G	S	-	-	snp	dbSNP	-
	1557077024	INTRONIC	1. 1141/300/	A/-		-	-	deletion	dbSNP	-
	<u>rs34209542</u>	INTRONIC	1: 114174047	A/G	R	-	-	snp	dbSNP	-
	rs3761935	INTRONIC	1: 114174051	T/G	К	-	-	snp	HGVbase, dbSNP	cluster, freq
	rs56354629	INTRONIC	1: 114178608	G/A	R	-	-	snp	dbSNP	-
	rs1599971	INTRONIC	1: 114178616	A/G	R	-	-	snp	ENSEMBL:Watson, dbSNP, TSC, ENSEMBL:Venter	cluster, freq, hapmap
	rs1970559	INTRONIC	1: 114178671	T/C	Y R	-	-	snp	Affy GeneChip 100K Array, HGVbase, dbSNP, TSC	cluster, freq, doublehit, hapmap
	rs2476601 rs12759178	INTRONIC	1: 114179091 1: 114181670	A/G A/C	м		-	snp snp	ENSEMBL:Watson, HGVbase, dbSNP, ENSEMBL:Venter dbSNP	cluster, freq, hapmap hapmap
	rs35747917	INTRONIC	1: 114181698	A/C A/T	W		-	snp	dbSNP	партар
	1833747317	INTRONIC	1. 114101090	AV1	-		-	anh	005NP	-

Investigating SNPs in Ensembl

			Variation: rs2476601	Variation: rs2476601										
0 Frances	61		- Summary		0110 (
C Ensem	וס		- Gene/Transcript (3)											
Home > Human	114.215.857 Gene: PTPN22		 Population genetics (4) Individual genotypes (272) 	Synonyms	/nonyms ENSEMBL:Watson ENSSNP6762723 ENSEMBL:Venter ENSSNP19339									
	Gene: PTPN22 (EN	CC00000424242	- Context	Alleles										
Gene: PTPN22		,	Phenotype Data (4)	Arrieles A/G (Ambiguity code: R) Ancestral allele: G										
 Splice variants (3) 	Tyrosine-protein phospi	hatase non-receptor type	Configure this page	Location			1 genomic	location(s).	hide locations					
 Supporting eviden Sequence 			 Add custom data to page 				-	ard strand)	Jump to reg					
 External reference 			Export data			1.114173	001 (1014)	ara suana)	Sump to req					
Regulation Generative Generative	<i>E</i> Ense	embl	Bookmark this page	// Genel	Transcript						Population genotypes and alle	le frequencies		
- Genomic alignn				« Gene/	rransonpt						r opulation genetypes and ane	ie nequencies		
Gene Tree	lome > Human			Demulation		Allelee	All-1	Constant	Constance	Description				
Gene Tree (Location: 1:114,07	71 121-114 302		Population		Alleles A	G	Genotypes A G	s Genotypes GIG	Description				
 Orthologues (4 Paralogues (12 	20044011.1111,01	1,121-111,002		CSHL-HAPMAP:	HapMap-CEU	0.158	0.858	0.317	0.683	30 mother-father-child trios from th	e CEPH collection (Utah residents with ano	estry.from northern ar		
Protein families Genetic Variation	Variation: rs2476	6601		CSHL-HAPMAP:		0.057	1.000	0.114	0.886		ng, China, representing one of the population			
- Variation Tab	1			CSHL-HAPMAP:	HapMap-JPT	0.022	1.000	0.044	0.956	44 unrelated Japanese in Tokyo, J	lapan, representing one of the populations	tudied, in the Interna		
 Variation Image External Data 	 Summary 			CSHL-HAPMAP:	HapMap-YRI	0.033	1.000	0.067	0.933	30 Yoruba mother-father-child trios	s in Ibadan, Nigeria, representing,one of the	populations studied		
E ID History	 Gene/Transcri 	pt (3)												
Gene history	Population gen	etics (4)		Ensembl release 53 - Mar 2009 © WTSI / EBI										
 Configure this pag Add custom data t 	 Individual geno 	types (272)		Ensembl releas	ie 53 - Mar 200	9 © <u>WISI</u> / <u>EI</u>	31							
 Export data 	- Context			Permanent link	- <u>View in archi</u>	ve site								
 Bookmark this pag 	Phenotype Dat	a (4)		100 (7 000)	guny couc									
			Ancestral allele: G											
	 Configure this p 	page	Location	This featur	e maps to	1 genom	nic loca	ition(s).	hide locat	ions	ons			
	 Add custom da 	ita to page												
	 Export data 	in to page			1:114179	091 (for	ward s	strand)	Jump t	to region in detail				
	 Bookmark this p 											ıblehit		
	 BOOKINALK UNS (page												
	ENSSNP11610818	INTRONIC	1: 114163868	C/G	S	-	-	snp (ENSEMBL:Watson		-	-		
	rs56048322	INTRONIC	1: 114173737	C/G	S	-	-	snp (dbSNP		-			
	<u>rs57877024</u>	INTRONIC	1: 114173867	A/-		-	-	deletion	dbSNP		-			
	rs34209542	INTRO	1: 114174047	A/G	R	-	-		dbSNP		-			
	rs3761935	NT	1: 114174051	T/G	К	-	-		HGVbase, dbSNP		cluster, fre	p		
	rs56354629 rs1599971	INTRONIC	1: 114178608 1: 114178616	G/A A/G	R	-	-		dbSNP	SNP, TSC, ENSEMBL:Venter	cluster, freg, h	200000		
	rs1970559	INTRONIC	1: 114178671	T/C	Y	-	-			Array, HGVbase, dbSNP, TSC	cluster, freq, double			
	rs2476601	IN LONIC	1: 114179091	A/G	R	-	-			GVbase, dbSNP, ENSEMBL:Venter	cluster, freq, h			
	rs12759178	INTRONIC	1: 114181670	A/C	М	-	-	snp (dbSNP		hapmap			
	<u>rs35747917</u>	INTRONIC	1: 114181698	A/T	W	-	-	snp (dbSNP		-			
					-									

Afternoon session

- You will look into SNPs related to athletic performance
 - Look up information in NCBI (dbSNP) and Ensembl
 - Look into personal SNPs profiles, what does the data look like, how can we predict risks using the data
 - Look up information about SNPs and their characteristic in populations in the HapMap genome browser