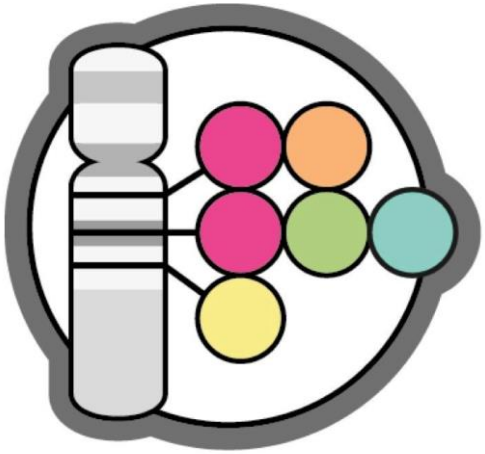
The GWAS Catalog



https://www.ebi.ac.uk/gwas/home

- Useful concepts
- The failure of linkage studies
- The common disease/common variant hypothesis
- What is a genome-wide association study (GWAS)?
  - The logic behind GWAS
  - GWAS ushered the personalized medicine era
  - Reasons for non-replication of GWAS
  - Has GWAS failed us or have we misunderstood GWAS?
- The GWAS catalog
- Practice

The pipeline so far - Tools

Obtaining DNA

Case/control cohort

Microarray Scanner

Quality control

Sequencer

Fastq

Quality control

Fastq

Alignment to reference genome

BAM

Visualization

SAM

Variant calling

VCF

Variant Annotation

Text file

Variant Prioritization

The Genome Wide Association Studies Era

FastQC

Bowtie2

BAM-TO-SAM, IGV

GATK,SAMtools

Accessible through Galaxy

Annovar, wAnnovar, Variant Effect Predictor, dbSNP, ExAC browser, annotation tools

A free ticket to a movie of your choice

1

Hacksaw Ridge

15

See more information about this film

Release date: 26/01/2017 **Running time:** 139 mins

Synopsis: Mel Gibson directs Andrew Garfield in the visceral true story of a war hero who didn't fire a single shot

Scarred by childhood experiences with his alcoholic WWI veteran father (Hugo Weaving), devout Seventh Day Adventist Desmond Doss (Andrew Garfield) resolves never to touch a gun. This proves something of a challenge when he enlists in the army to fight in WWII. As a conscientious objector, Desmond insists on his right to serve as a medic. But he's bullied by his unsympathetic comrades, who consider him to be a coward. His greatest test comes during the notoriously bloody Battle of Okinawa. Returning to the director's chair for the first time in a decade, Mel Gibson celebrates a very different kind of hero in this intense, blood-soaked true-life tale of faith, courage and patriotism. Andrew Garfield is on outstanding form as the pacifist whose single-handed acts of bravery saved 75 lives.

2

Rogue One: A Star Wars Story

12A

See more information about this film

Average user rating: ☆☆☆☆☆

Release date: 15/12/2016 **Running time:** 134 mins

Synopsis: Felicity Jones takes the lead in the first Star Wars spin-off movie.

It is a period of civil war. The Galactic Empire rules the galaxy, and are putting the finishing touches to their ultimate super-weapon, the Death Star. The Rebellion plan to steal the its plans in order to detect a weak spot in which to destroy it. They recruit Jyn Erso (Felicity Jones) to work with Rebel fighter Cassian Andor (Diego Luna) and a team to undertake the deadly mission.

3

Underworld: Blood Wars

15

See more information about this film

Average user rating: ☆☆☆☆☆

Release date: 13/01/2017 **Running time:** 91 mins

Synopsis: Kate Beckinsale is back in black.

In the latest Underworld movie, vampire death dealer Selene (Kate Beckinsale) continues her struggle against the Lycan clan and the vampire faction that betrayed her, with both sides trying to use the blood of her and her daughter to become Vampire-Corvinus hybrids. With her only allies, David (Theo James) and his father Thomas (Charles Dance), she must stop the eternal war between Lycans and Vampires, even if it means she has to make the ultimate sacrifice.

4

Passengers

12A

See more information about this film

Average user rating: ☆☆☆☆☆

Release date: 21/12/2016 **Running time:** 116 mins

Synopsis: When two people on a deep-space mission are unexpectedly awoken from their slumber, they are forced to depend on one another for survival.

Jim (Chris Pratt) and Aurora (Jennifer Lawrence) are destined for life on another planet, but their stasis is suddenly interrupted 90 years before they were due to wake up. Unable to return to their slumber, they're stuck together on a ship hurtling through the unknown. As they begin to investigate the mystery the two begin to fall for one another – but will they be able to survive the deadly truth about why they woke up in the first place?

Why???

69 upcoming comic book movies, and when to expect them

William Shatner will star alongside Adam West in the latest addition to our comic book movies list...

“Remembrance of things past is not necessarily the remembrance of things as they were”
Marcel Proust

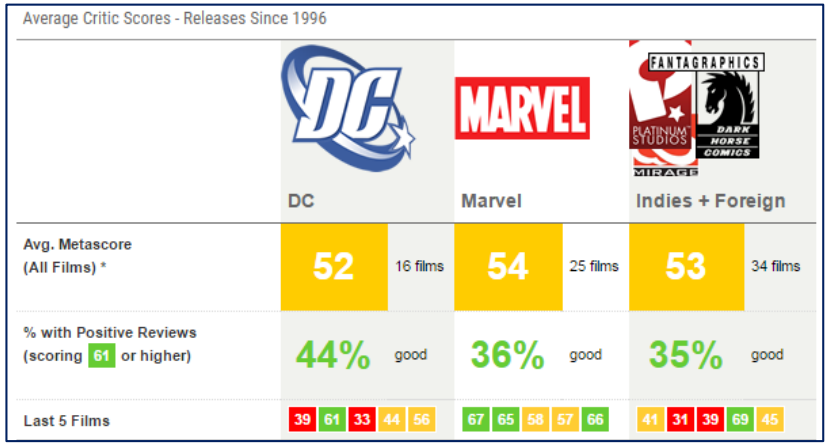
Highest-grossing films[13]

Rank	Peak	Title	Worldwide gross	Year
1	1	*Avatar*	$2,787,965,087	2009
2	1	*Titanic*	$2,186,772,302	1997
3	3	*Star Wars: The Force Awakens*	$2,068,223,624	2015
4	3	*Jurassic World*	$1,670,400,637	2015
5	3	*The Avengers*	$1,518,812,988	2012
6	4	*Furious 7*	$1,516,045,911	2015
7	5	*Avengers: Age of Ultron*	$1,405,403,694	2015
8	3	*Harry Potter and the Deathly Hallows – Part 2*	$1,341,511,219	2011
9[F]	5	*Frozen*	$1,287,000,000	2013
10	5	*Iron Man 3*	$1,214,811,252	2013

https://en.wikipedia.org/wiki/List_of_highest-grossing_films#Highest-grossing_films

	DC	Marvel	Indies + Foreign
Avg. Gross	$157.5m	$178.9m	$67.9m
Last 5 Films	$70.8m	$149.2m	$39.6m
Films > $100m	13 films (48%)	19 films (70%)	8 films (16%)
Biggest Hit (inflation-adjusted)	$583.9 million *The Dark Knight* (2008)	$546.2 million *Spider-Man* (2002)	$429.3 million *Men in Black* (1997)
Major Flop (not inflation-adjusted)	$10.4m ($47m budget) *Jonah Hex* (2010)	$6.9m ($15m budget) *Red Sonia* (1985)	$4.1m ($25m budget) *Tank Girl* (1995)

** Excludes a few films that do not have at least 7 critic reviews in our database.*
*** Box office grosses for all films are adjusted for inflation to 2011 dollars. Note that several films are still playing in theaters; all box office data is through July 24, 2011 (source: Boxofficemojo.com).*



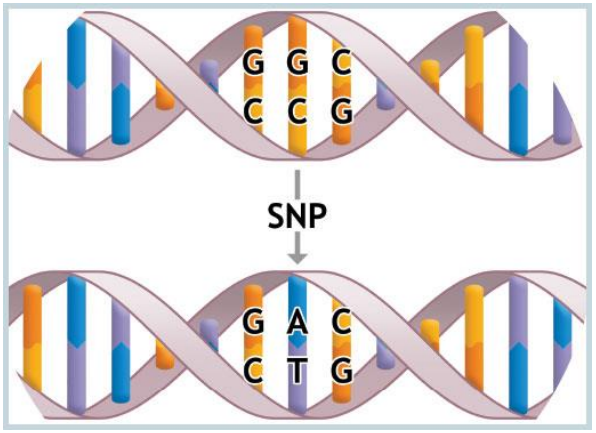
http://www.metacritic.com/feature/comic-book-movies-by-publisher

Concepts Underlying the Study Design

The modern unit of genetic variation is the single nucleotide polymorphism or SNP (single base-pair changes in the DNA sequence). In genetic studies, SNPs are used as markers of a genomic region and are the most abundant form of genetic variation in the human genome.

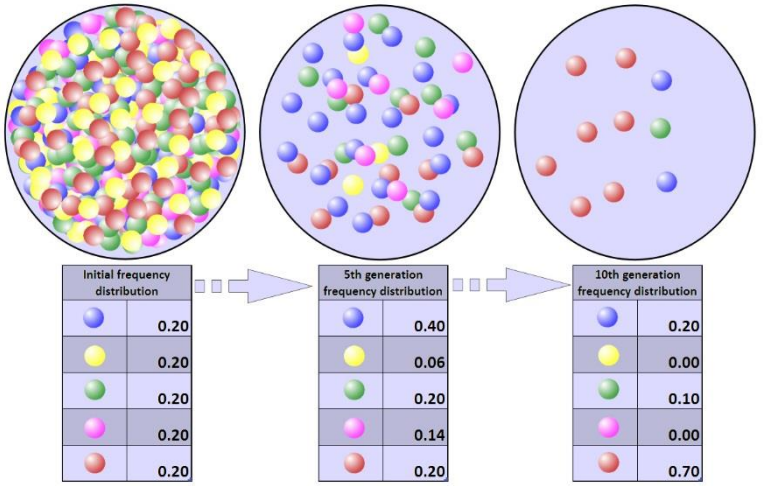
Most SNPs are non-functional or have a minimal impact on biological systems. **Some** SNPs have functional consequences, causing amino acid changes, changes to mRNA transcript stability, and changes to transcription factor binding affinity.

Most SNPs have two alleles, meaning within a population there are two commonly occurring base-pair possibilities for a SNP location.



The frequency of a SNP is given in terms of the minor allele frequency (MAF) or the frequency of the less common allele. For example, a SNP with a minor allele frequency of 0.40 implies that 40% of a population has the *G* allele versus *A* the more common allele (the major allele), which is found in 60% of the population.

Random genetic drift responsible to changes in allele frequencies of most SNPs creating a confusion between causal and neutral alleles



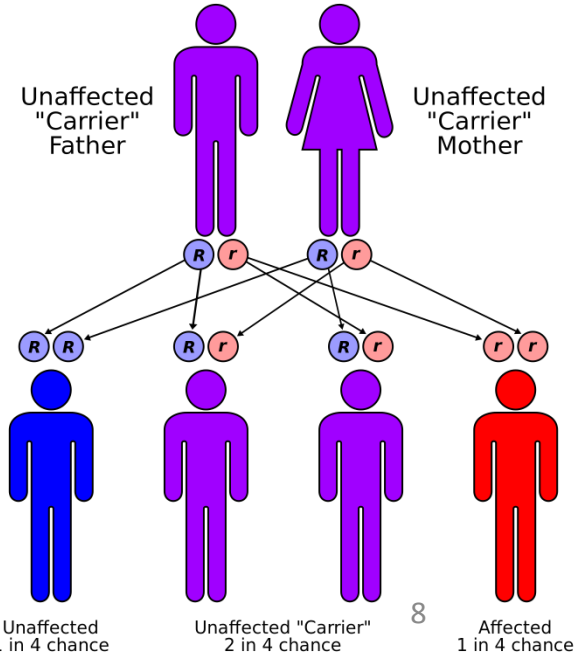
Rare genetic disorders (e.g., cystic fibrosis) are caused due to extremely rare SNPs (=*mutations*) that induce a detrimental change to protein function.

In the genetics literature, the term SNP is generally applied to *common* single base-pair changes, and the term mutation is applied to *rare* genetic variants.

The failure of linkage for common disorders

The identification of *CFTR* mutations was achieved by genotyping affected, and examining how those genetic markers segregate with the disease across families. *Linkage analysis*, was successfully applied to identify genetic variants in other rare disorders like Huntington disease. Applied to more common disorders, linkage analysis has **not** fared well.

The *common disease/common variant* (CD/CV) hypothesis



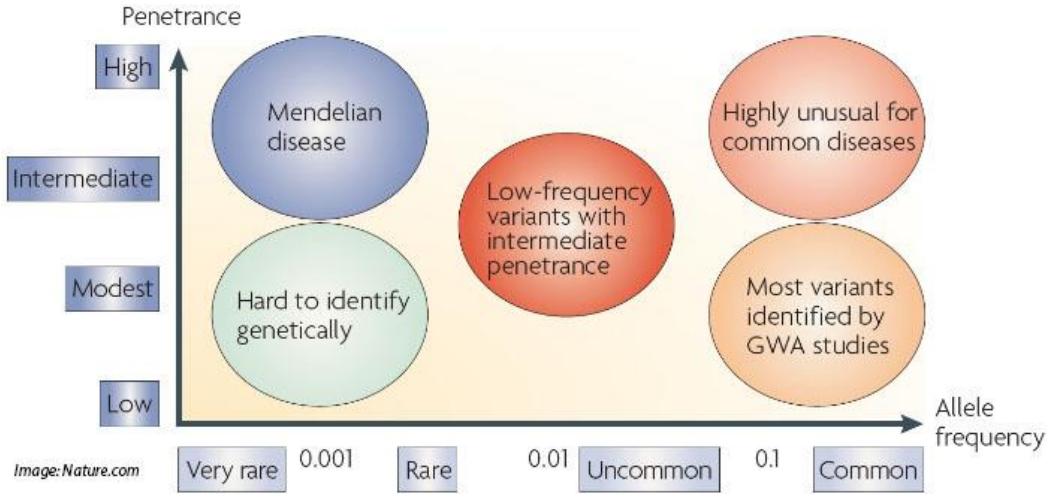
The common disease/common variant hypothesis

Common disorders are likely influenced by genetic variation common in the population. **This has ramifications of this for studying diseases.**

If common genetic variants influence a disease, the *effect size* (or *penetrance*) must be small, compare to rare disorders. **Here is why**:

- If a common a disease allele (MAF=0.4) inflicts the phenotype in 40% of the population then the population prevalence and MAF would be correlated.
- If that allele causes a small change in gene expression that alters risk for a disease by some small amount then the prevalence of the disease and the MAF would be slightly correlated.

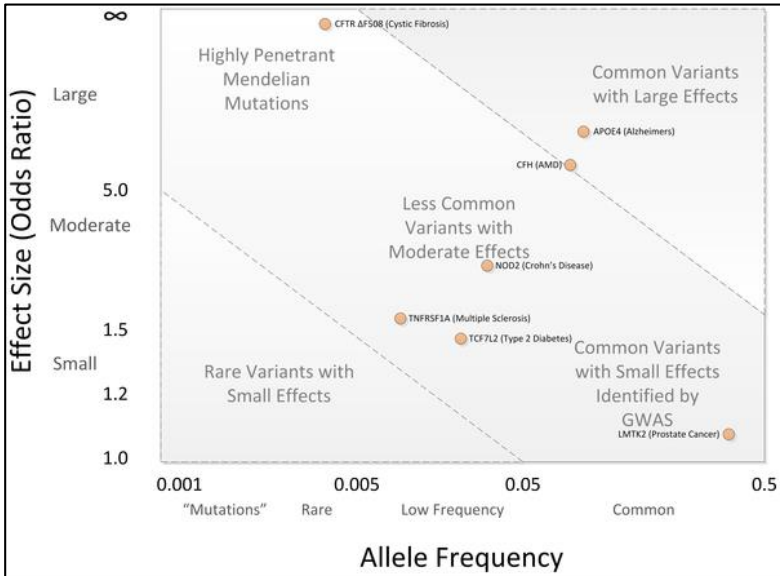
Therefore, common variants have low penetrance.



If common alleles have small genetic effects (=low penetrance), but common disorders show heritability (=inheritance in families), then multiple common alleles must influence disease susceptibility. **Here is why**:

- If a twin study estimates the heritability of a common disease as 40%, then 40% of the total variance in disease risk is due to genetic factors.
- If the allele of a SNP incurs only a small degree of disease risk, then that SNP explains only a small proportion of the total genetic variation. Therefore, the total genetic risk due to common genetic variation must be spread across genetic factors.

Therefore, family-based studies would be more successful in studying complex disorders than population-based studies.



Bush WS, Moore JH (2012) Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol 8(12): e1002822. doi:10.1371/journal.pcbi.1002822
http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1002822

What is a Genome-Wide Association Study (GWAS)?

A central goal of human genetics is to identify genetic risk factors for common, complex diseases (e.g., schizophrenia) and for rare Mendelian diseases (e.g., cystic fibrosis). Understanding the biological basis of genetic effects has an important role in developing new pharmacologic therapies.

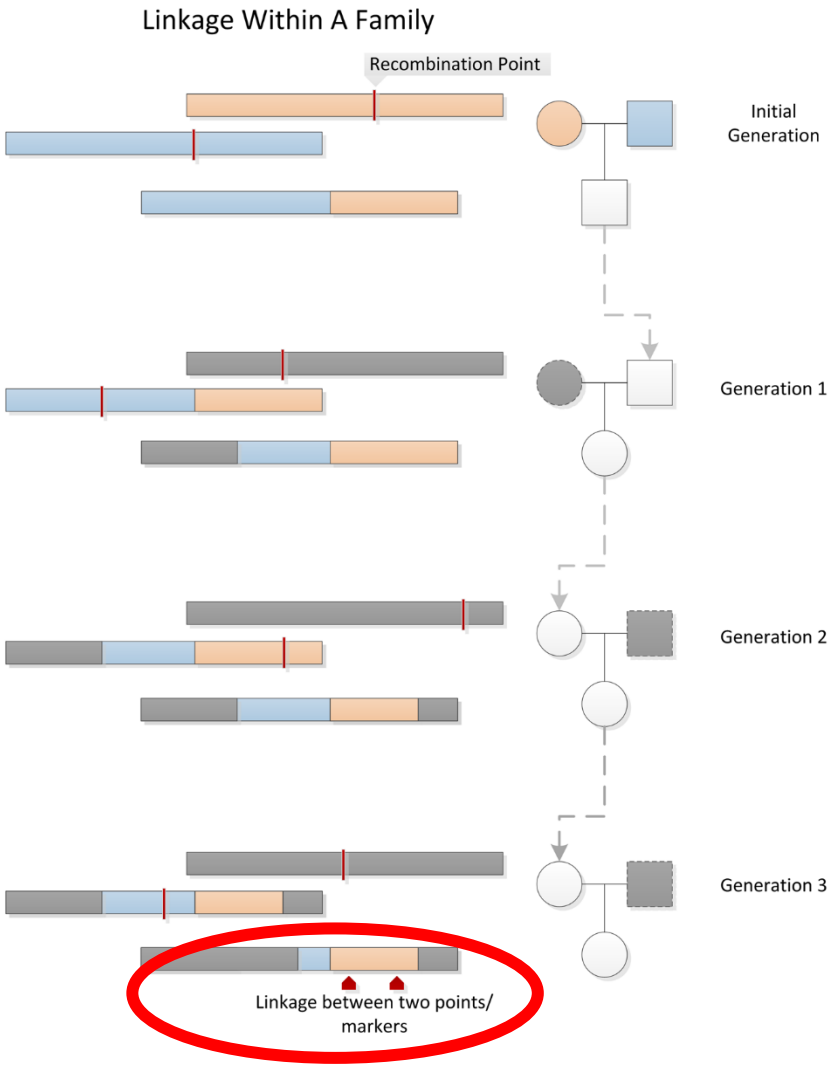
- There are many different technologies, study designs, and analytical tools for identifying genetic risk factors.

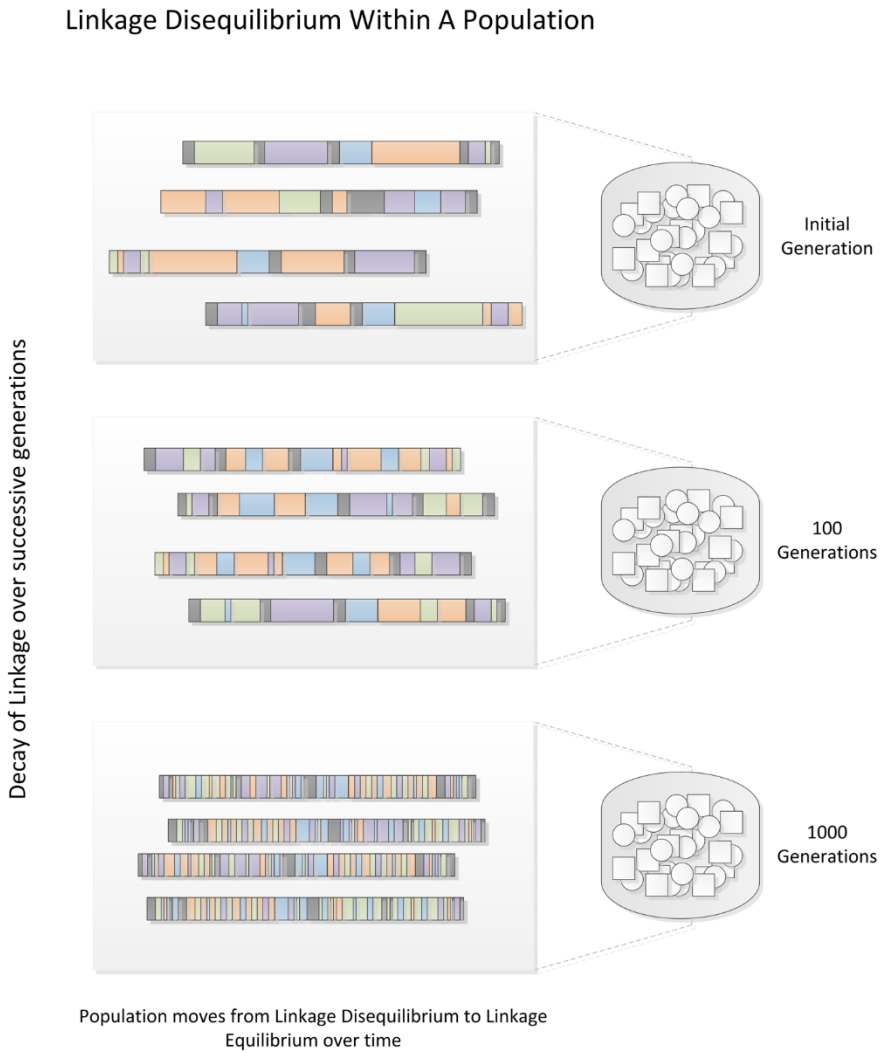
- GWAS analyzes DNA sequence variations throughout the genome to identify genetic risk factors for common diseases.

- GWAS's goals are to use genetic risk factors to make predictions about who is at risk and to identify the biological underpinnings of disease susceptibility for developing new prevention and treatment strategies.

- GWAS can be powerful tools for investigating the genetic architecture of **common** (single gene) diseases.

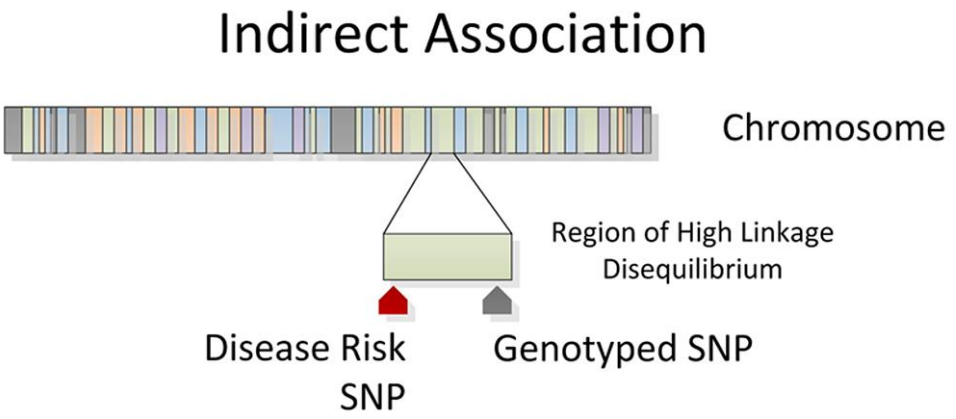
The logic behind GWAS



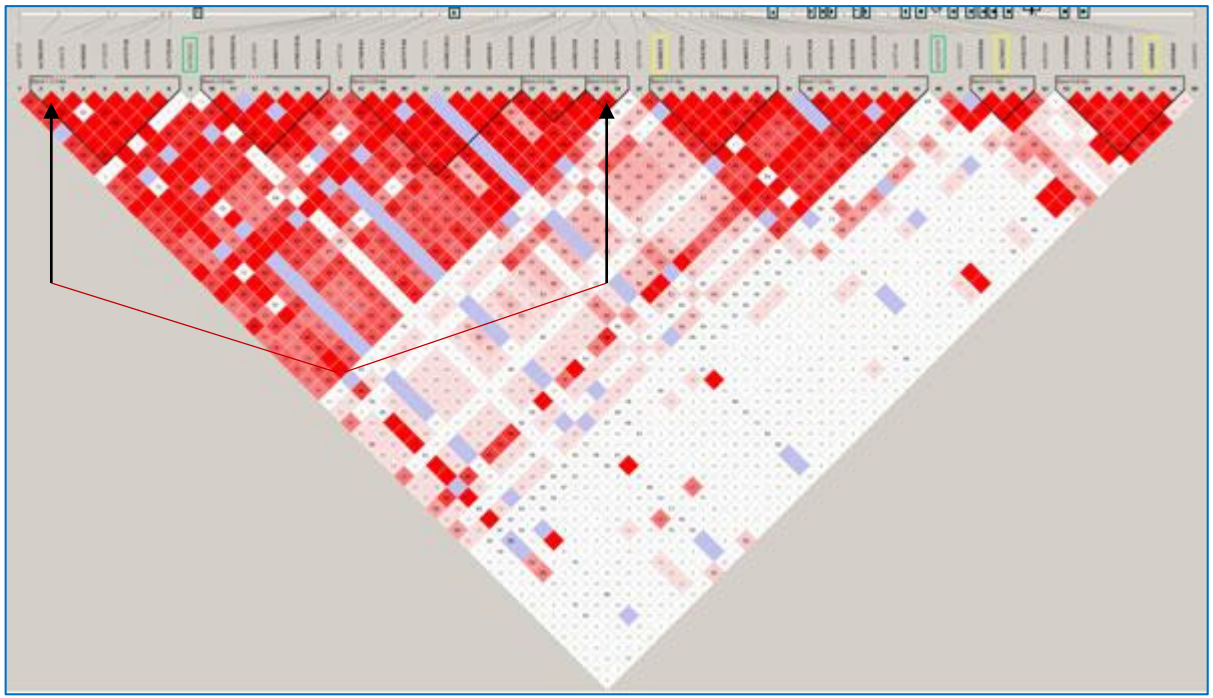


Within a family, linkage occurs when two genetic markers remain linked on a chromosome rather than being broken apart by recombination events during meiosis. **In a population**, contiguous stretches of founder chromosomes from the initial generation are sequentially reduced in size by recombination events.

Assumption. Genotyped SNPs often lie in a region of high linkage disequilibrium with an influential allele. The genotyped SNP will be statistically associated with disease as a surrogate for the disease SNP through an indirect association.



"Tag SNPs" allow predicting the value of the SNPs that they tag.

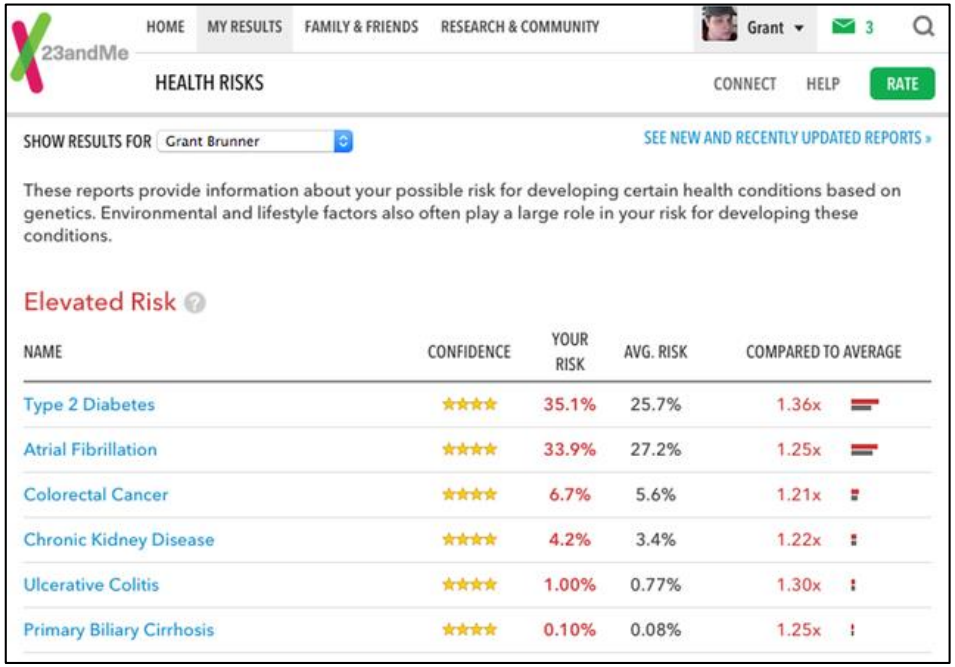


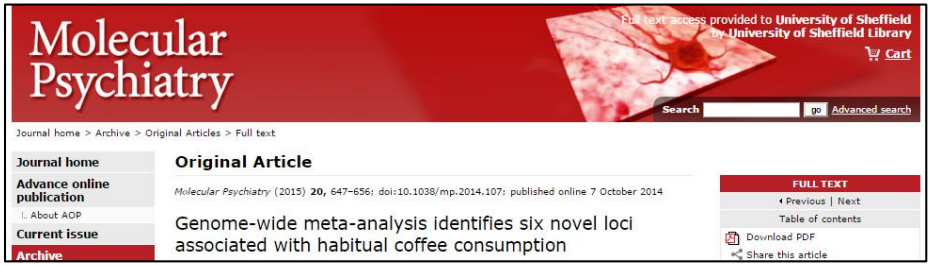
Estimating the correlation between every two adjacent SNPs. Red=highly correlated. White=not correlated.

Can you identify the recombination points?

GWAS ushered the personalized medicine era

The widespread availability of low-cost technology for measuring an individual's genetic background has been harnessed by businesses that market genetic testing directly to the consumer.





Replication?!?!?!?

For better or worse, GWAS have paved the way to the era of personalized medicine and personal genetic testing.

Reasons for the non-replicability of GWAS

1. The original observation was a false-positive due to sampling error.

2. The follow-up study had insufficient power.

3. The genotypic coding used in the initial study may not accurately reflect the true underlying association, leading to a loss of power (**correlation<>causation!**).

4. The variant may be a poor marker for the trait due to differences in linkage-disequilibrium structure between the studies.

5. Differences in design or trait definition may lead to inconsistencies.

6. The absence of an association in the subsequent studies may be due to true etiologic heterogeneity.

Has GWAS failed us or have we misunderstood GWAS?

Is the GWAS approach founded on a flawed assumption that genetics plays an important role in the risk for common diseases?

GWAS showed that 10%–50% of phenotypic variation is captured when all SNPs are considered simultaneously for a number of complex diseases and traits in support of pedigree studies suggest that a substantial proportion of variation in susceptibility for common disease is due to genetic factors.

Have GWASs been disappointing in not explaining more genetic variation in the population?

The aim of GWAS is to detect loci that are **associated** with complex traits. In some cases, the detection of such loci has led to the discovery of new biological knowledge about disease.

Have GWASs delivered meaningful biologically relevant knowledge or results of clinical or any other utility?

Yes, in some cases. No, in others cases.

Are GWAS results spurious?

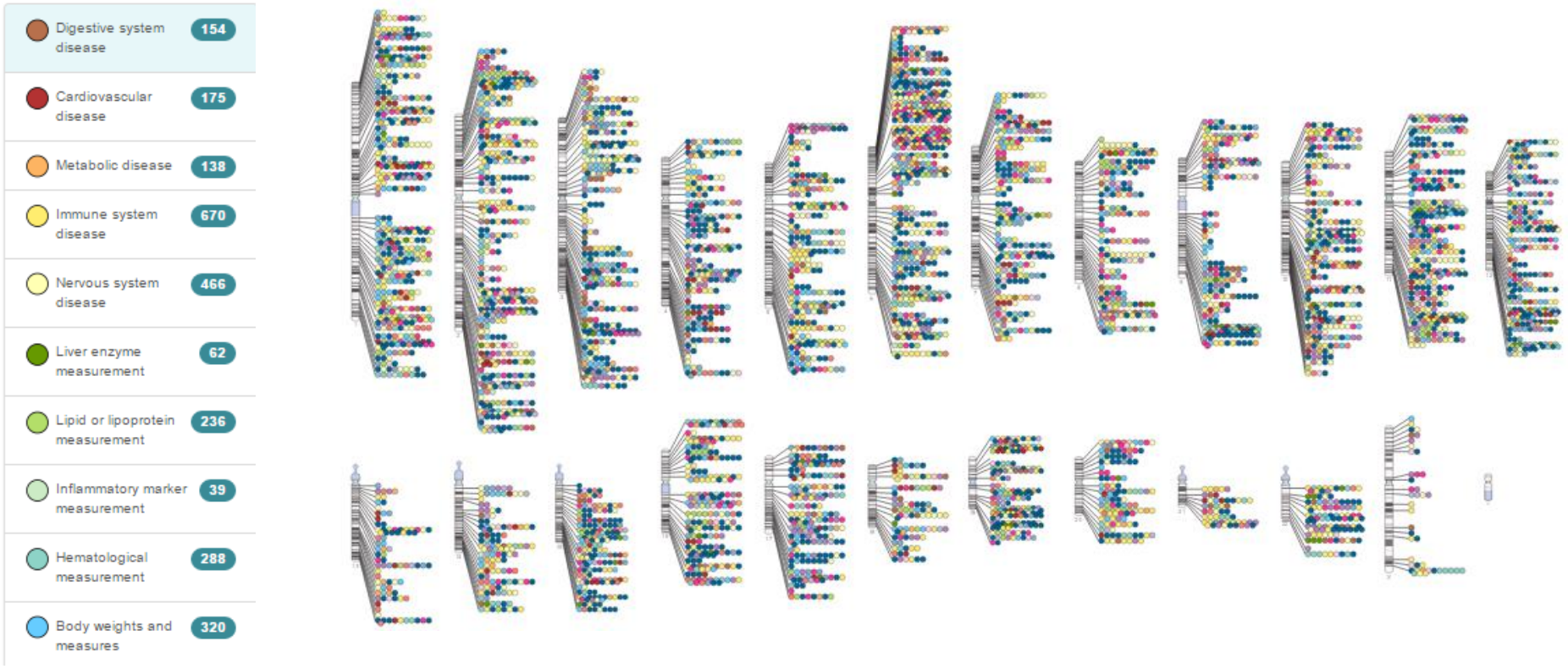
Yes, in some cases. No, in others cases.

Visscher et al. (2012) http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3257326/

The GWAS catalog

https://www.ebi.ac.uk/gwas/

The GWAS Catalog was founded by the NHGRI in 2008. The Catalog is a quality controlled, manually curated, literature-derived collection of all published genome-wide association studies **assaying** at least 100,000 SNPs (typical size of GWAS) and all SNP-trait associations with p-values < 1.0×10^{-5} *(Hindorff et al., 2009)*.



Exercise

https://www.ebi.ac.uk/gwas/

Search for GWAS studies on autism and download the association results.

- How many autism GWAS studies (publications) are in the catalog?
- Which SNP (and gene) has the highest OR for autism?
- How many SNPs were studied in this study (Under PLATFORM).
- The SFARI database ranks genes associated with autism. Search for the gene that you found https://gene.sfari.org/autdb/search. Does it appear inSFARI?

Search for GWAS studies on coffee and download the association results.

- Which gene is best associated with Coffee consumption?
- Where is it located?
- Is this association true for all worldwide individuals?
- Read the paper. How many SNPs were used in a typical study to infer association?

Download the entire GWAS catalog (https://www.ebi.ac.uk/gwas/docs/file-downloads 18.3Mb) and observe the data.

- How many SNPs are in the catalog?
- How many "stop" mutations are in the catalog?
- Rank the SNPs by OR. What is the most common functional classification of SNPs?