

What are good drug targets and how to find them?

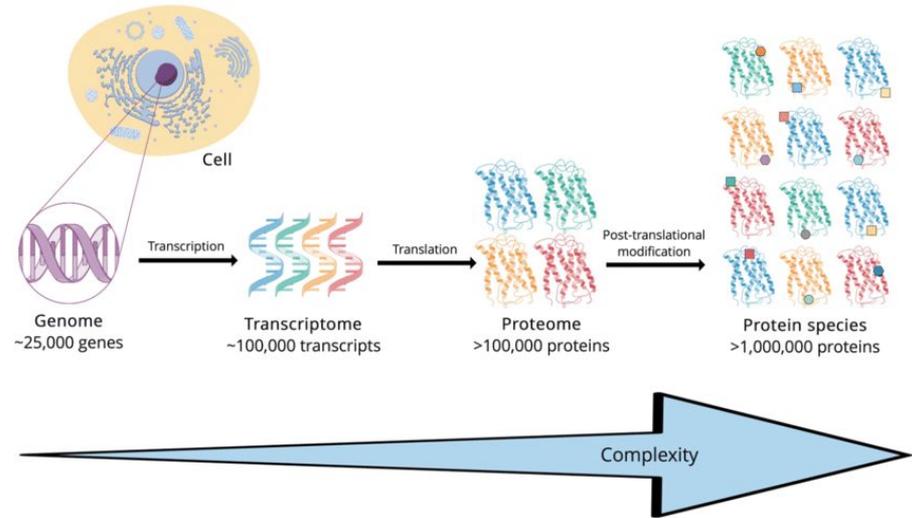
*Mathematical and Computational Biology in Drug
Discovery (MCBDD), Module I*

Dr. Jitao David Zhang, March 2026

Outline

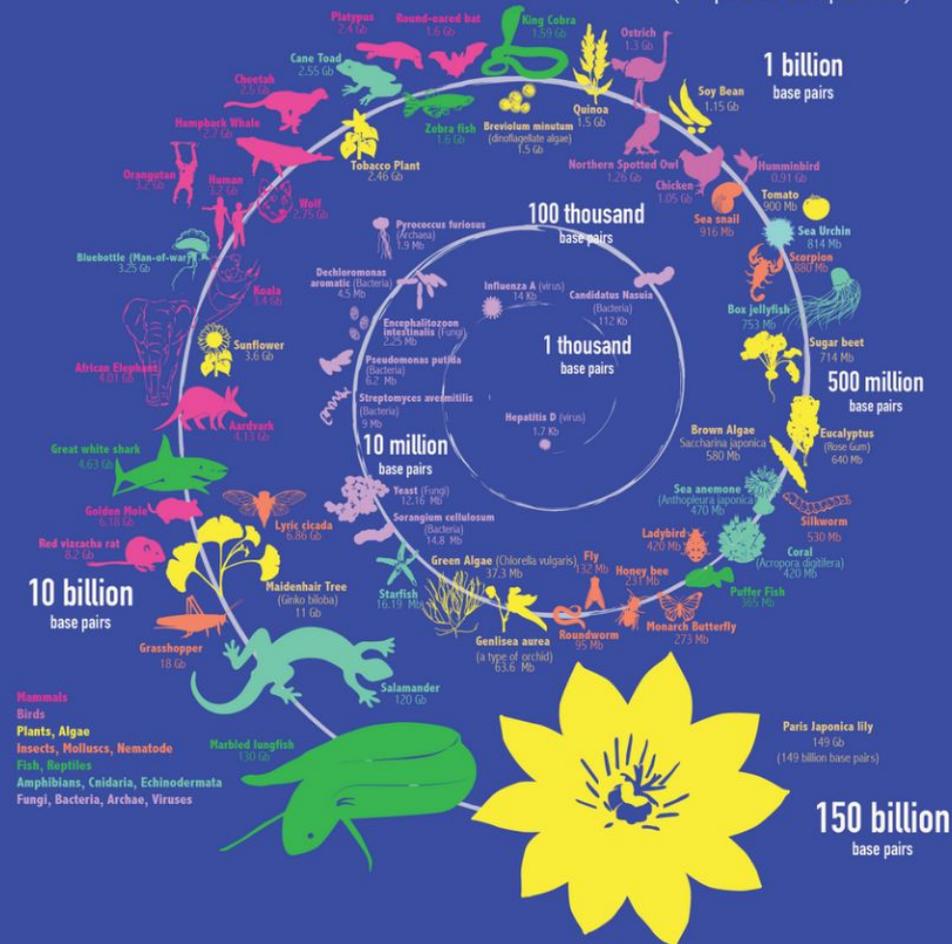
- A brief introduction to inference.
- Biological foundations of target identification.
- Drugs targeting targets with genetic evidence support are twice more likely to be successful.

Finding targets: needle in a haystack?



Left: [Kathy Lemmon-Warde, CC-NC-ND](#), Right: [Virag et al., 2020, CC-BY 4.0](#)

Genome size (A species comparison)



Exercise of *inference* (I)

The company *Fränzi and Friends* developed a new quick test at home for seasonal Influenza, which is pending regulatory agency's review. The test has been shown to have a sensitivity of 99% and a specificity of 99%.

Suppose that Fred uses the test by *Fränzi and Friends* and the test was positive. Assume that 5% of the population is in fact infected. Was is your guess about the probability that Fred is indeed infected?

		Ground truth	
		Positive	Negative
Prediction	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{FP + TN}$$

Example

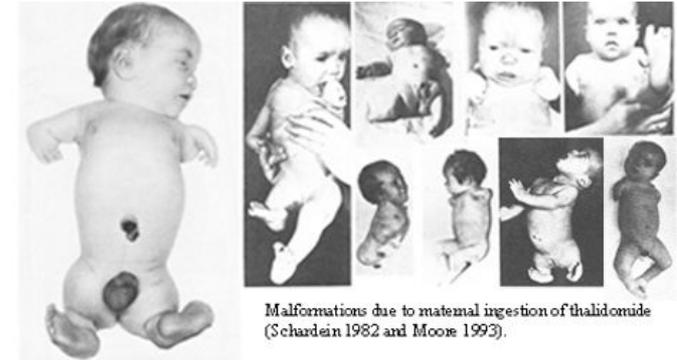
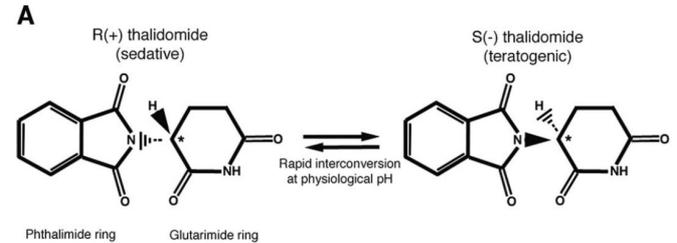
		Hidden truth	
		Healthy	Disease
Diagnosis	Healthy	45	5
	Disease	10	40

$$\text{Sensitivity} = \frac{45}{45 + 10} = 81.8\%$$

$$\text{Specificity} = \frac{40}{5 + 40} = 88.9\%$$

Exercise of *inference* (II)

- S-thalidomide is *teratogenic* and induces severe malformation .
- Among about 20,000 approved drugs, it is estimated that 30-140 have known teratogenic effects *in vivo*. Let's assume that 100 (i.e. 0.5%) are teratogens.
- We have developed a biological assay to predict teratogenicity before clinical trials in human. Applying the assay to 30 known teratogens and 50 non-teratogens, we report that the assay has 95% sensitivity and 98% specificity.
- **Question:** Is our assay good enough to be used for drug discovery screening?



Thalidomide can induce severe phocomelia
([source: Wikipedia](https://en.wikipedia.org/wiki/Thalidomide))

Exercise of *inference* (III)

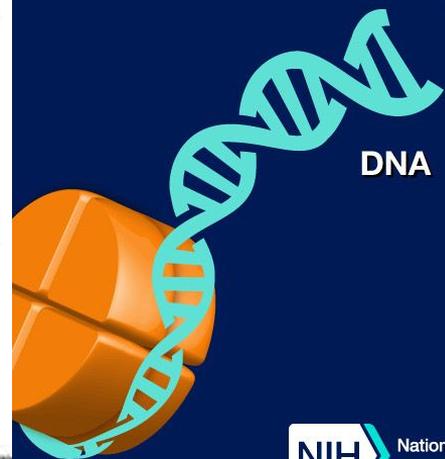
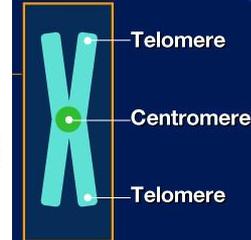
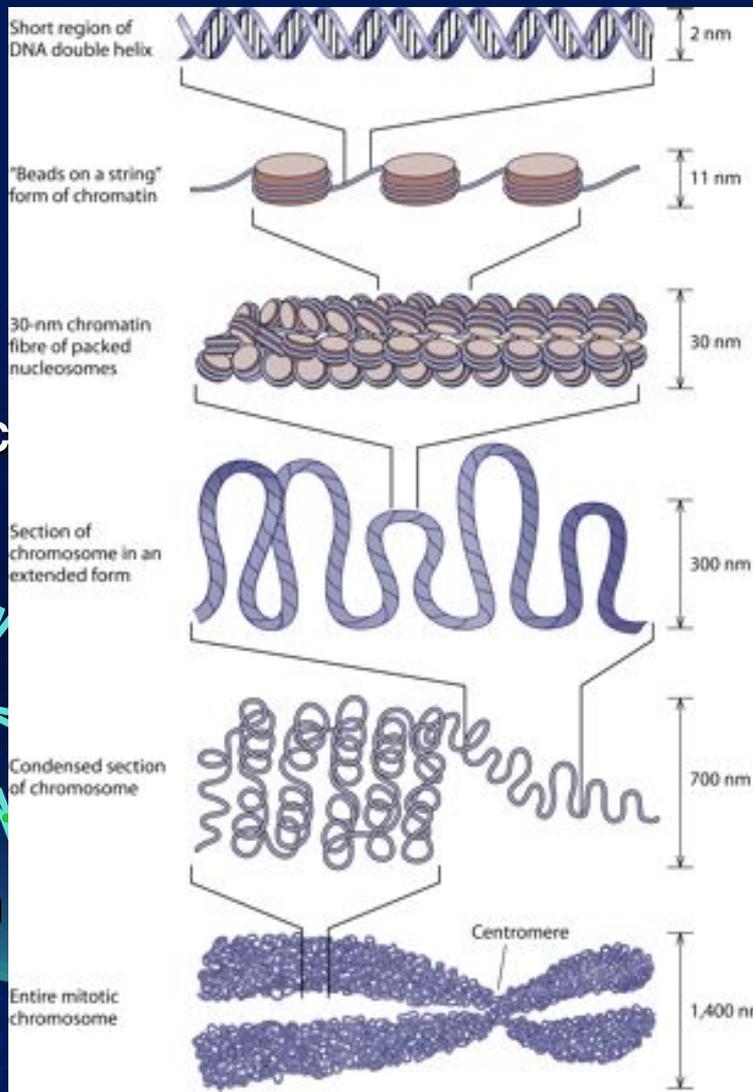
I have two pills and two hamsters. The pills are identical. The two hamsters are optically identical, too, so that I cannot tell the difference between them.

I know that both pills are one of the following three types:

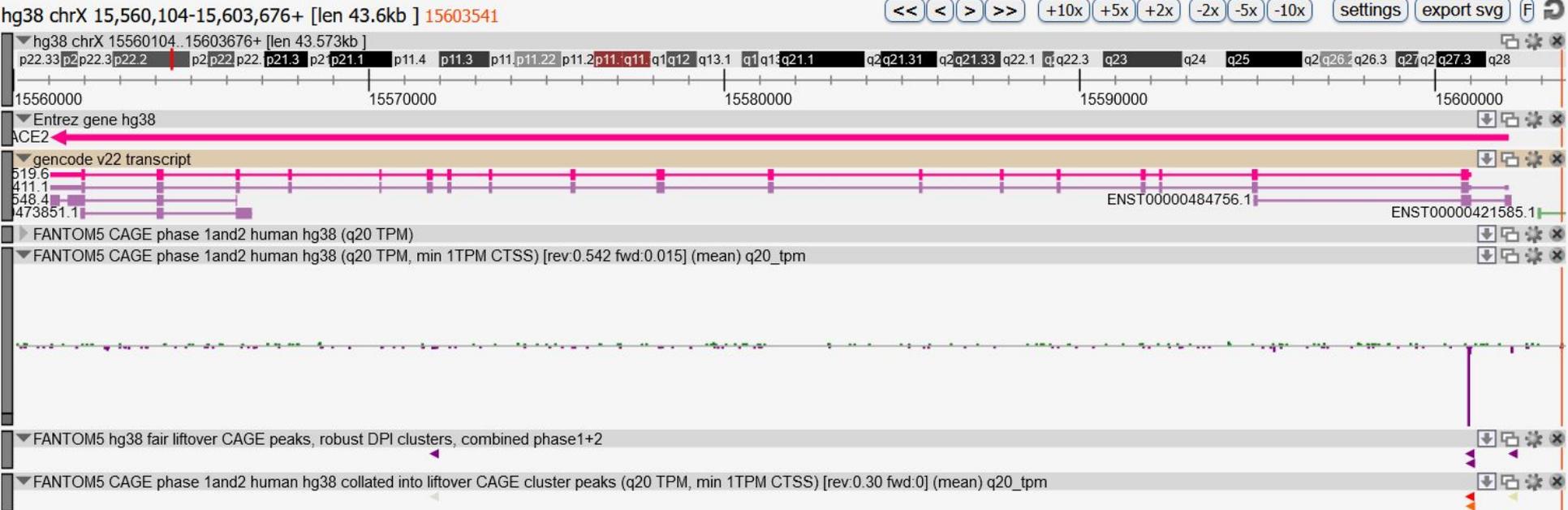
1. Pill A makes both hamsters sleep.
2. Pill B makes neither animal sleep.
3. Pill C makes one animal sleep but not the other.

Now I pick a pill, feed it to one hamster, and the hamster falls asleep. What's the probability that the second pill makes the other animal sleep, too?

Chromosome



Gene structure and gene expression



ACE2 viewed in [FANTOM5/ZENBU](#)

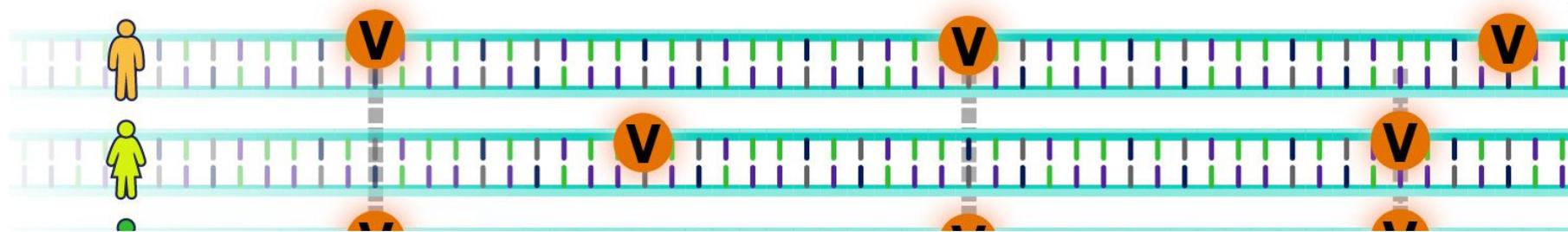


Person one

A G A C G C T

<u>Variant ID</u>	<u>Source</u>	<u>HGVS Consequence</u>	<u>VEP Annotation</u>	<u>LoF Curation</u>	<u>Clinical Significance</u> ^	<u>Flags</u>	<u>Allele Count</u>
17-7579617-C-T	E	c.74+22G>A	● intron		Likely benign		1
17-7579831-C-T	E	c.74+8G>A	● splice region		Likely benign		1
17-7579924-G-A	E G	c.-12C>T	● 5' UTR		Likely benign		7
17-7579932-G-C	E	c.-20C>G	● 5' UTR		Likely benign		2
17-7578142-C-A	E G	c.672+35G>T	● intron		not provided		9
17-7577142-C-A	E	p.Gly266Ter	● stop gained		Pathogenic		1
17-7578188-C-A	E	p.Glu221Ter	● stop gained		Pathogenic		1
17-7578263-G-A	E	p.Arg196Ter	● stop gained		Pathogenic		1
17-7576928-TAGGAA...	E	c.920-14_920-3delTGC...	● splice region		Uncertain significance		2
17-7578171-C-A	E G	c.672+6G>T	● splice region		Uncertain significance		2
17-7578171-C-T	E	c.672+6G>A	● splice region		Uncertain significance		1
17-7579934-C-T	G	c.-22G>A	● 5' UTR		Uncertain significance		1
17-7565206-T-A	G	c.*51A>T †	● 3' UTR				1
17-7565222-C-T	G	c.*35G>A †	● 3' UTR				1



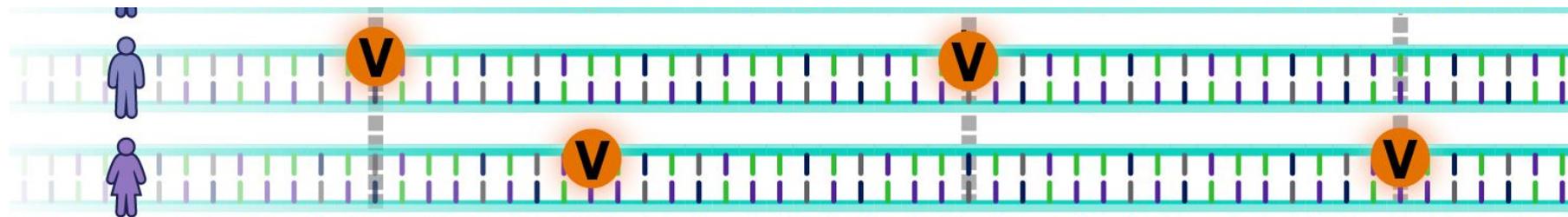

a

CNV

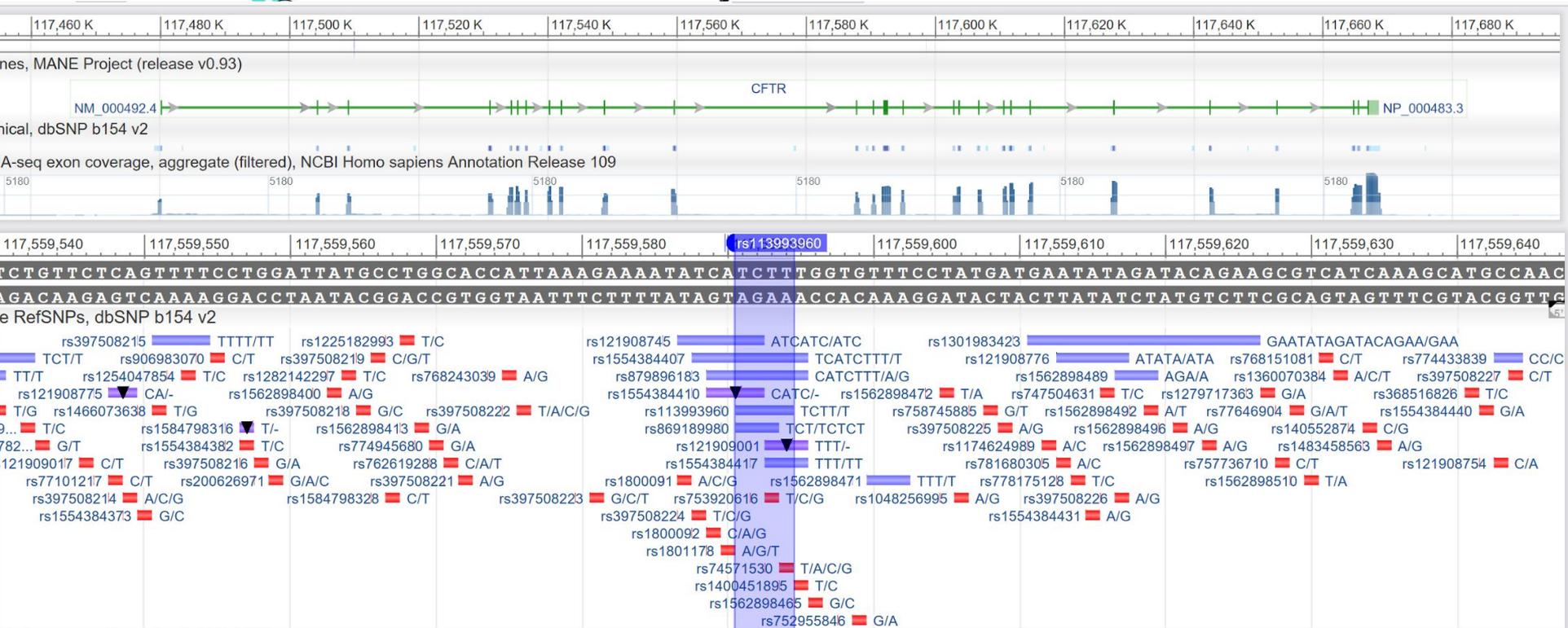
Other SV (non-CNV)

Unresolved

SV class	Deletion	Duplication	Multiallelic CNV	Insertion	Inversion	Translocation	Complex SV	Breakends
Abbrev.	-DEL	-DUP	-MCNV	-INS	-INV	-CTX	-CPX	-BND
Ref.								
Example alternatives								
						(See Fig. 2)	Discarded	



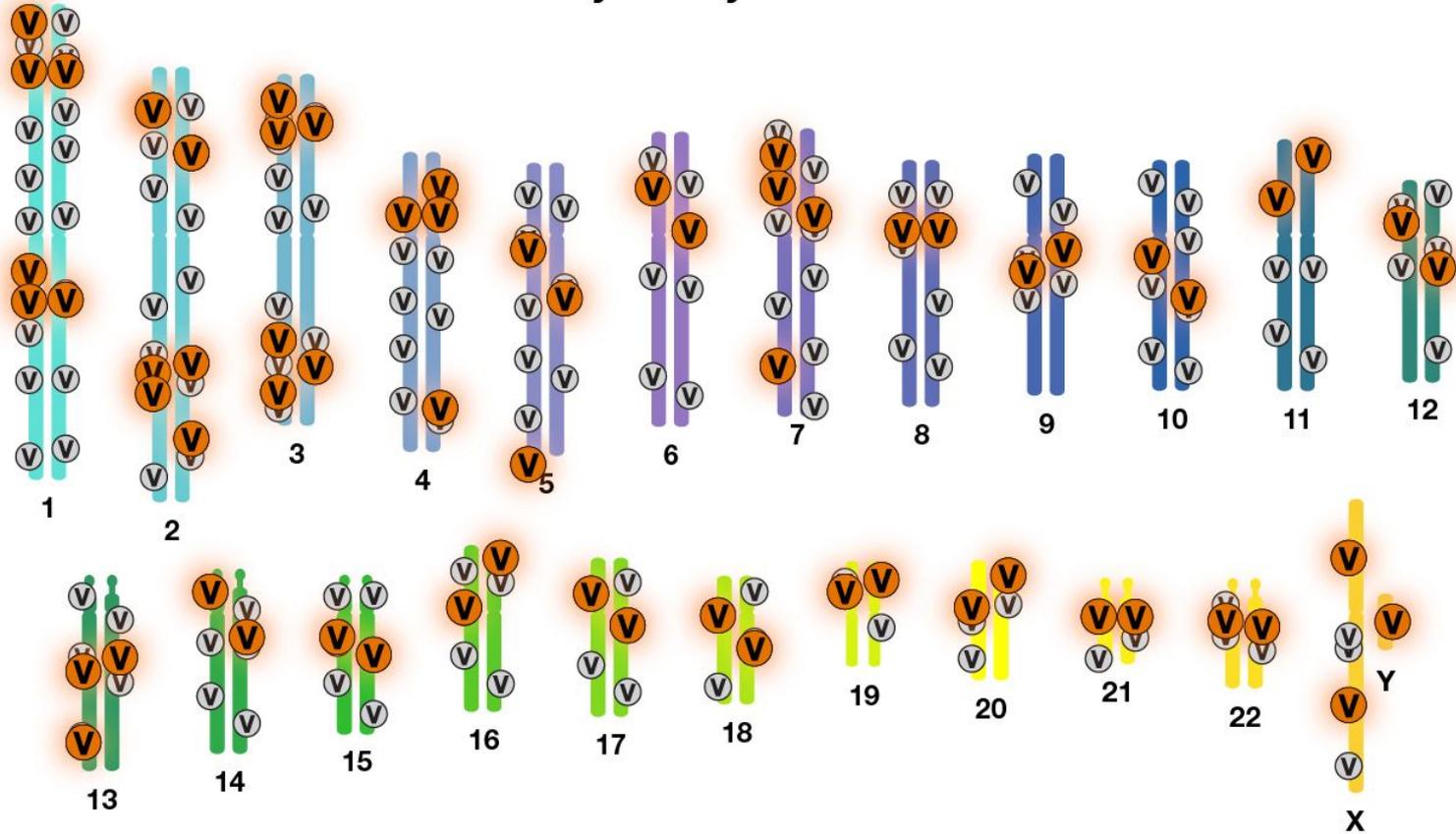
Cystic fibrosis



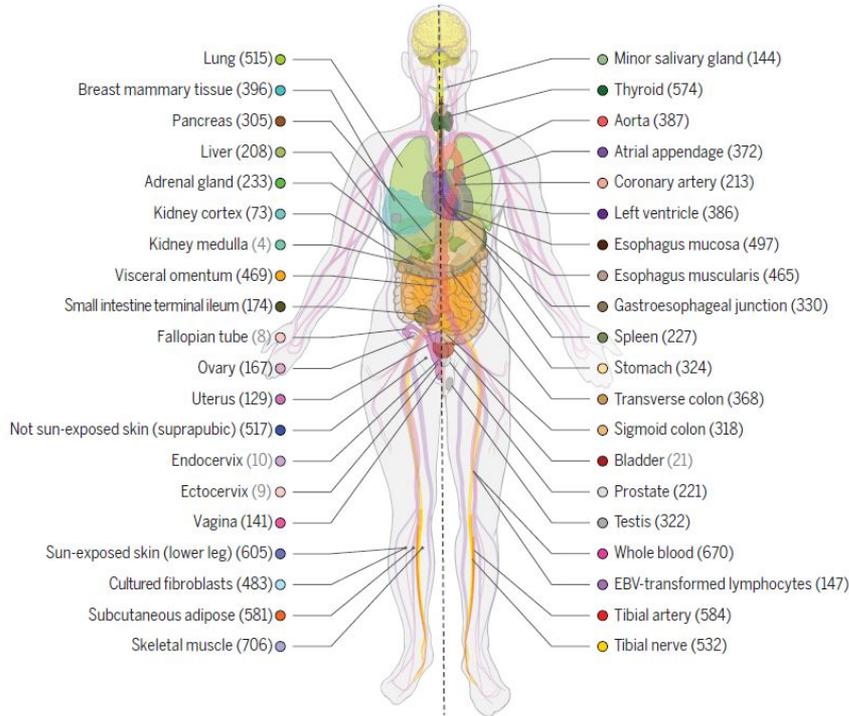
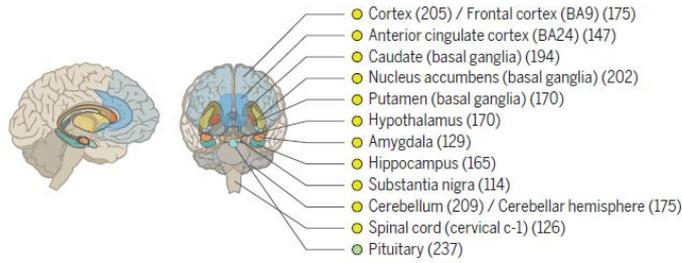
The [CFTR gene](#) (Chr 7), and [rs113993960](#) (F508del), the most common cause of cystic fibrosis (CF). [Read more about CFTR modulator therapies](#) (CF Foundation), and a [deep coverage by Sarah Zhang](#) (Atlantic).

Coronary artery disease

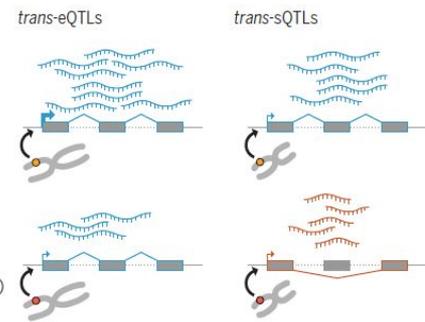
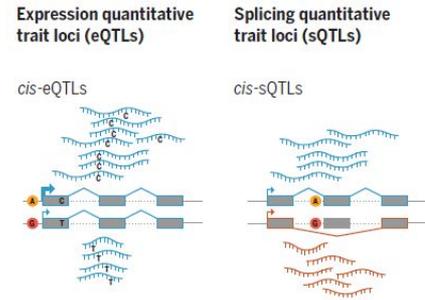
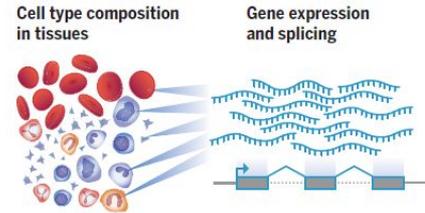
Person
two



A



B



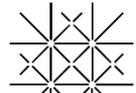
GTE_x (v8)

Cost per Human Genome

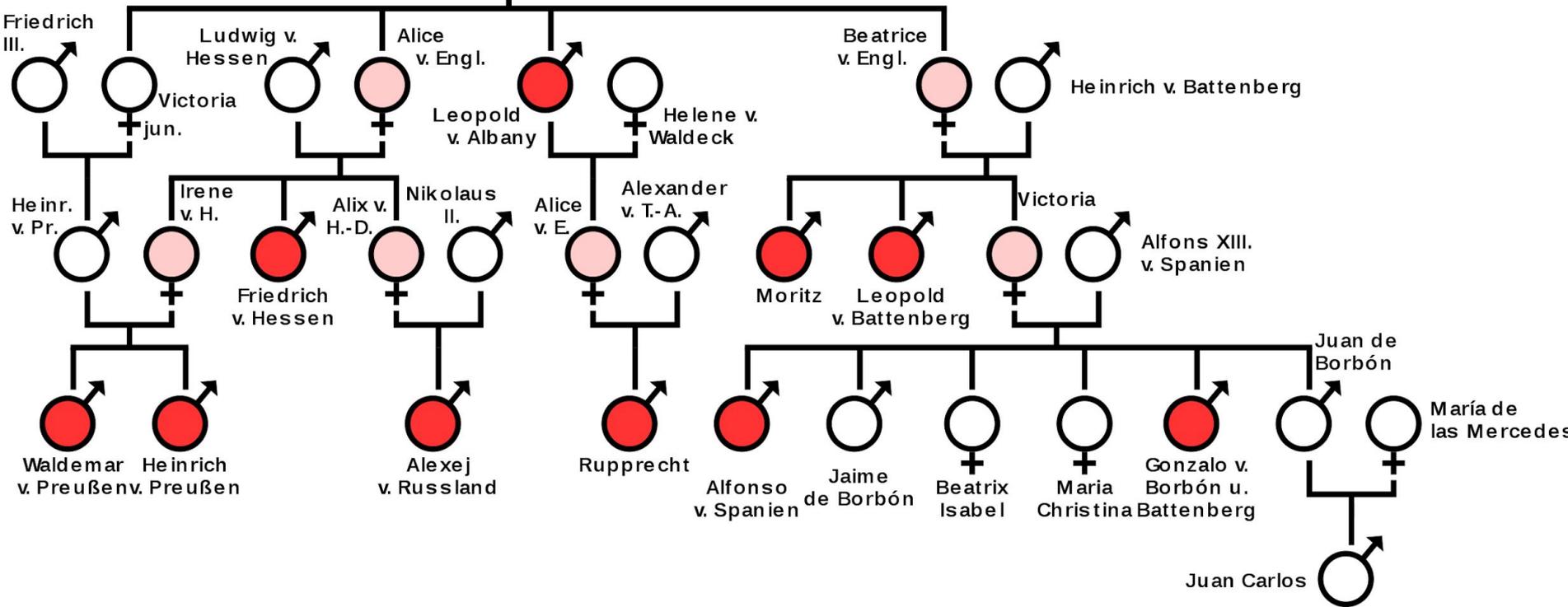




Haemophilia in the descendants of Queen Victoria

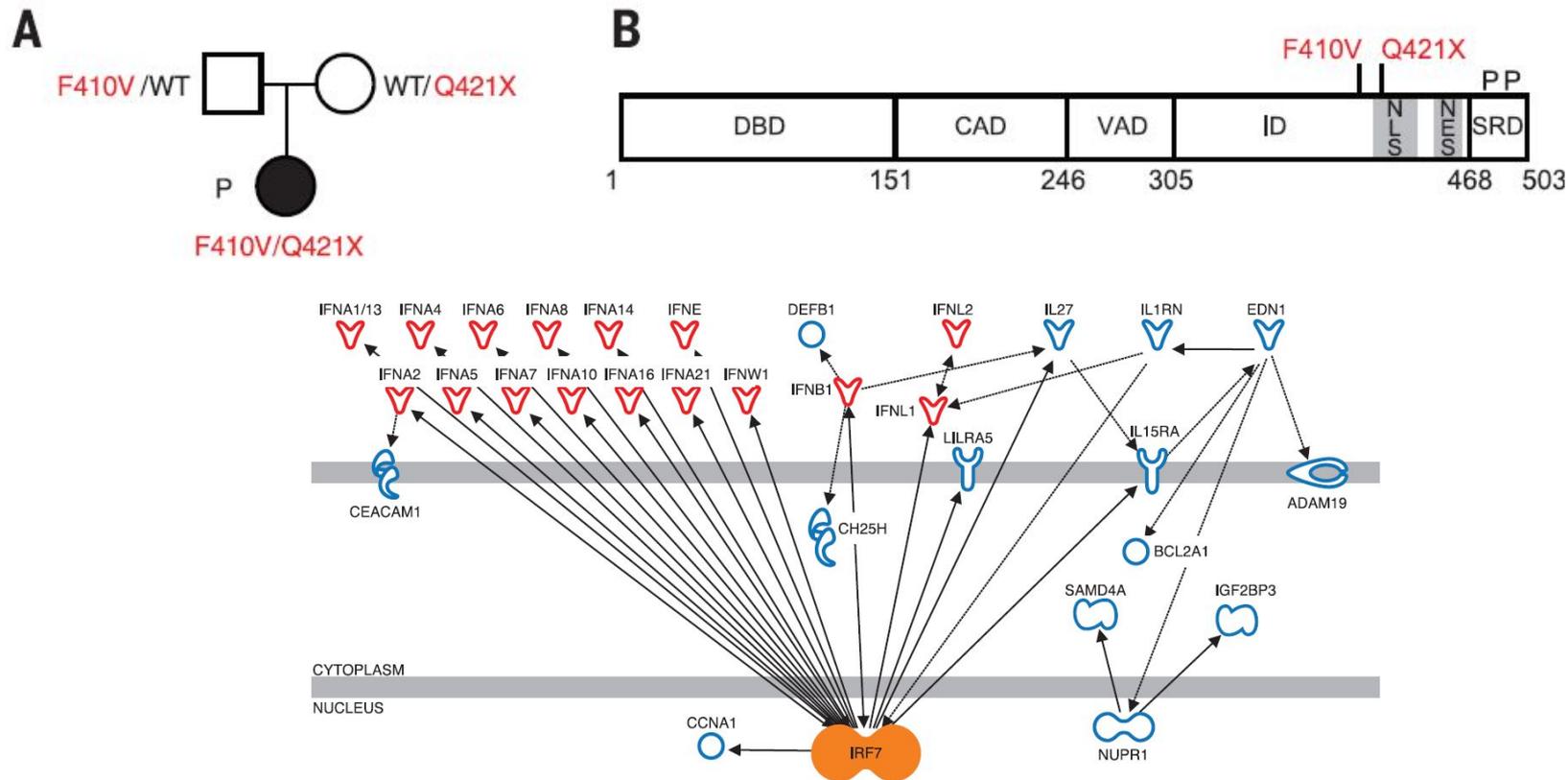


Prince Albert  Queen Victoria 



Prussia (1889-1945)	Prussia (1900-1904)	of Russia (1904-1918)	Teck (1907-1928)	Asturias (1907-1938)	Spain (1914-1934)
------------------------	------------------------	--------------------------	---------------------	-------------------------	----------------------

Life-threatening influenza infection in human IRF7 deficiency detected by trio sequencing



Genetics helps to find drug targets

The hypothesis: genes that are associated with disease-associated traits are more likely to be a valid drug target for an indication with similar phenotypes than genes that are not associated. The more causal the association, the more likely.

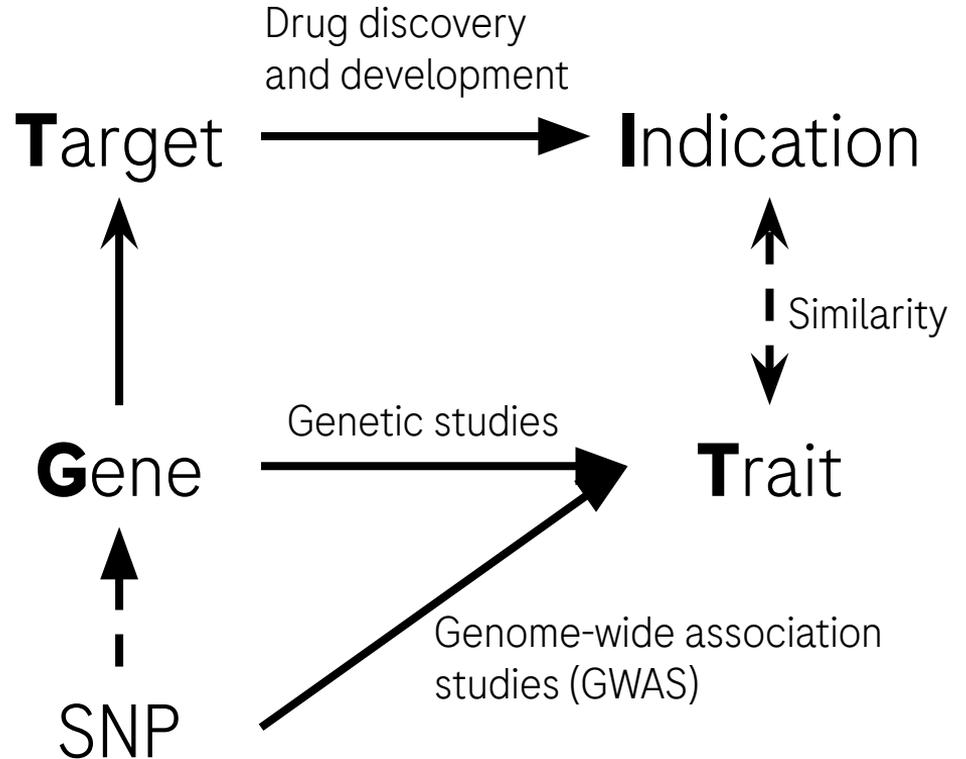


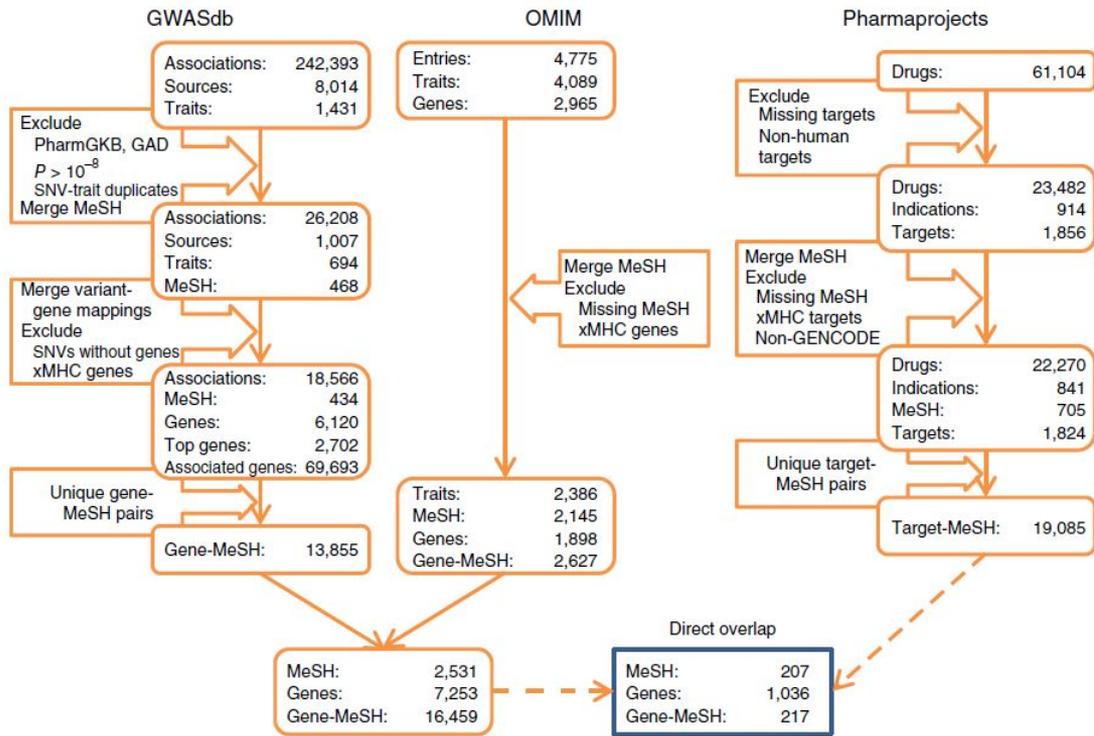
Illustration of the definition of genetic support for drug target

drug programs				human genetic associations			
gene	indication MeSH ID	indication MeSH term	phase	gene	association MeSH ID	association MeSH term	source
<i>ABCC8</i>	D000070642	Brain injury, traumatic	Phase II	<i>ABCC8</i>	D003924	Diabetes Mellitus, Type 2	OTG
<i>ABCC8</i>	D003924	Diabetes Mellitus, Type 2	Launched	<i>ABCC8</i>	D003924	Diabetes Mellitus, Type 2	OMIM
<i>FFAR1</i>	D003924	Diabetes Mellitus, Type 2	Phase III	<i>ABCC8</i>	D007003	Hypoglycemia	OMIM
<i>IL1R1</i>	D003924	Diabetes Mellitus, Type 2	Phase II	<i>ABCC8</i>	D000428	Alcohol Drinking	Genebass
				<i>IL1R1</i>	D015212	Inflammatory Bowel Diseases	OTG

similarity = 1.0

- MeSH: [Medical Subject Headings](#) (NIH)
- OTG: [Open Targets Genetics](#), with [an example of ABCC8](#)
- OMIM: [Online Mendelian Inheritance in Man](#)
- Genebass: [Gene\(exome\)-based association studies](#), with an example of [ABCC8](#)

Impact of genetics on target identification: a factor of ~2 estimated by Nelson *et al.*



GWASdb: a database for human genetic variants identified by genome-wide association studies (Li et al., 2012, now almost obsolete). Alternatives include the [GWAS Catalog](#) (NHGRI-EBI).

Pharmaprojects: commercial tool to track and analyse global drug R&D landscape ([the CITELINE company](#)).

Nelson, Matthew R. et al. 2015. "The Support of Human Genetic Evidence for Approved Drug Indications." *Nature Genetics* 47 (8): 856–60. <https://doi.org/10.1038/ng.3314>.

Disease ← Gene ← Drug

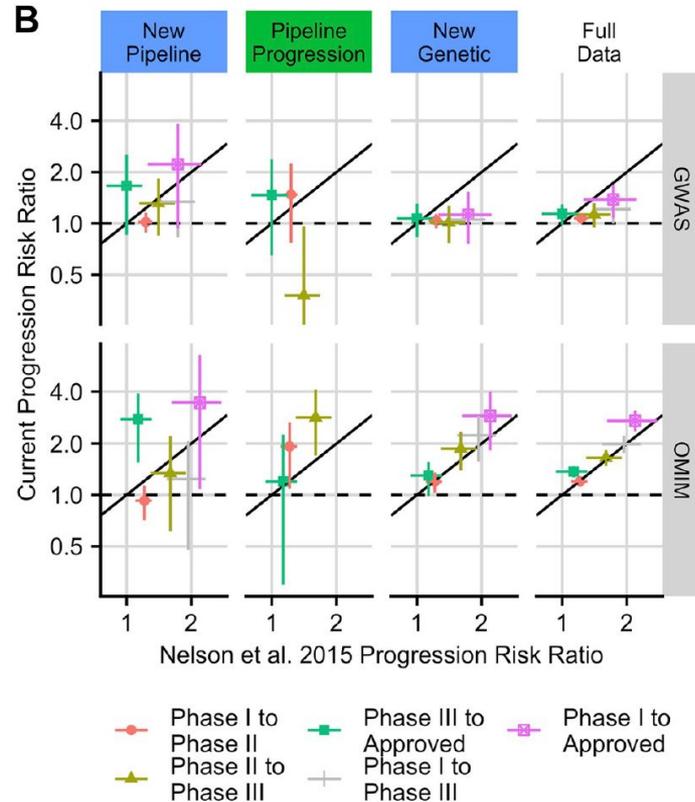
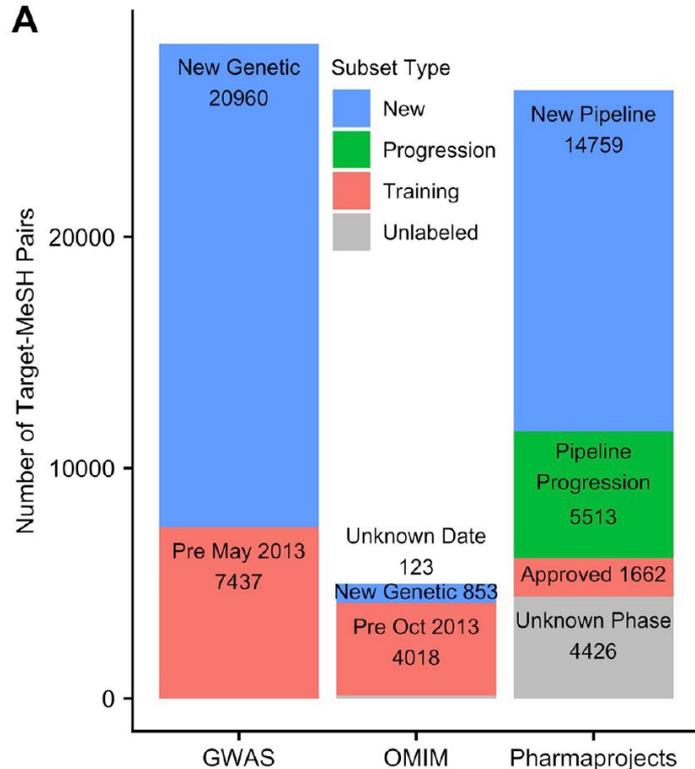
Nelson *et al.* inferred that genetic support offers an likelihood ratio ~2

Table 1 The relative value of genetic support for the probability that a target-indication pair progresses along the drug development pipeline, based on historical drug trial information

Progression	$p(\text{progress} \text{genetic support})/(\text{progress} \text{no genetic support})$		
	GWASdb and OMIM	GWASdb	OMIM
Phase I to phase II	1.2 (1.1–1.3)	1.2 (1.1–1.3)	1.2 (1.1–1.3)
Phase II to phase III	1.5 (1.3–1.7)	1.4 (1.2–1.7)	1.6 (1.3–1.9)
Phase III to approval	1.1 (1.0–1.2)	1.0 (0.8–1.2)	1.1 (0.9–1.3)
Phase I to phase III	1.8 (1.5–2.1)	1.8 (1.4–2.1)	1.9 (1.5–2.3)
Phase I to approval	2.0 (1.6–2.4)	1.8 (1.3–2.3)	2.2 (1.6–2.8)

Values give the ratio of the probability of a target-indication pair progressing given genetic support to the probability of progressing without genetic support; 95% confidence intervals are given in parentheses.

Follow-up study by King *et al.*, 2019: benefits more than 2 fold when causal evidences are strong



King, Emily A., J. Wade Davis, and Jacob F. Degner. 2019. "Are Drug Targets with Genetic Support Twice as Likely to Be Approved? Revised Estimates of the Impact of Genetic Support for Drug Mechanisms on the Probability of Drug Approval." *PLOS Genetics* 15 (12): e1008489. <https://doi.org/10.1371/journal.pgen.1008489>.

Genes with *biologically understandable* genetic association are more likely to be good targets

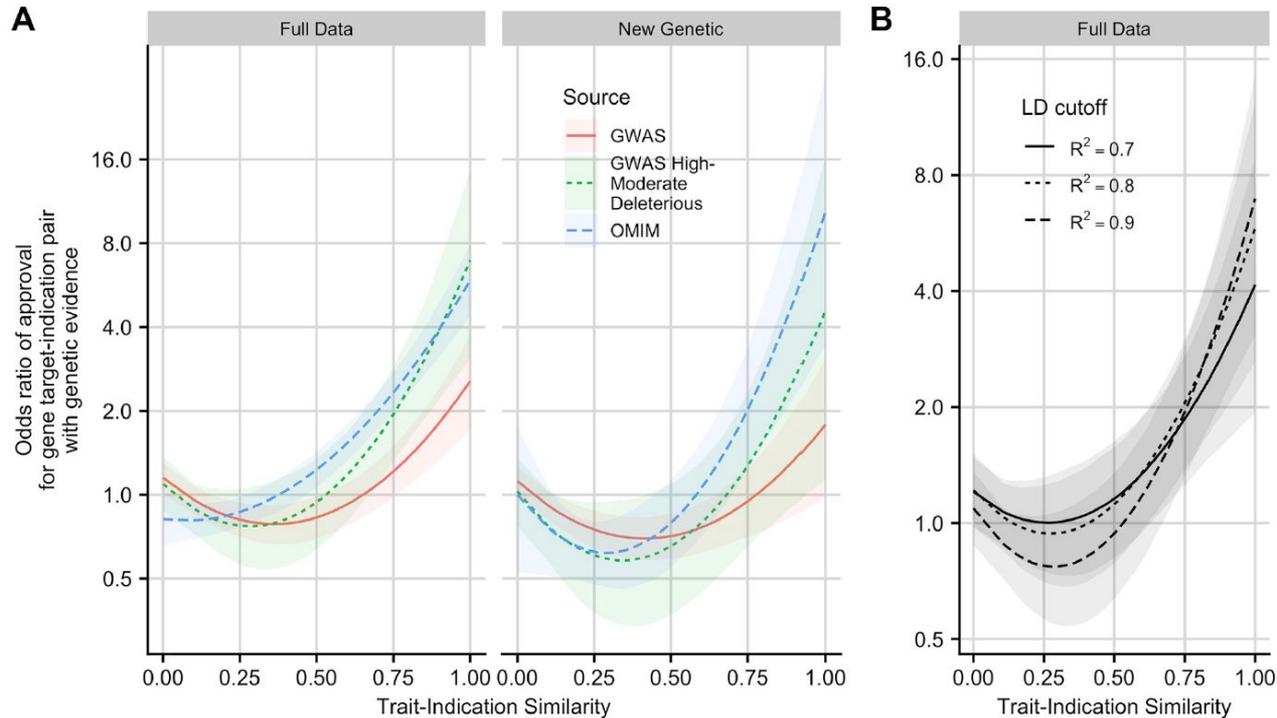
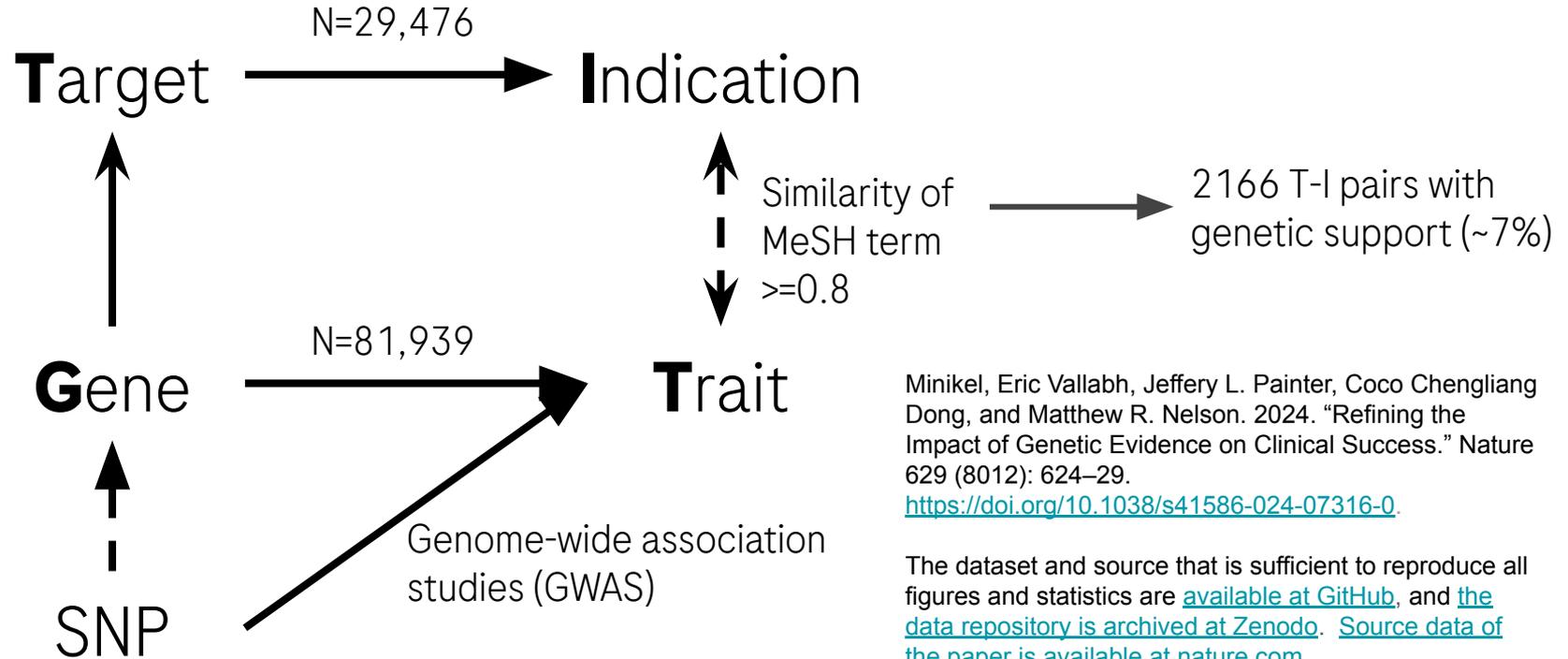


Fig 2. Estimated odds ratio of gene target-indication pair attaining approval, as a function of similarity between drug indication and the most similar trait associated with the target. A: Left: All genetic associations. Right: Only genetic associations reported after 2013 download. B: Effect of LD expansion threshold R^2 on the estimated approval odds ratio of a drug gene target-indication pair supported by a GWAS high-moderate deleterious variant. Posterior median and pointwise 95% credible interval from Bayesian logistic regression.

Follow-follow-up study Minikel *et al.* 2024



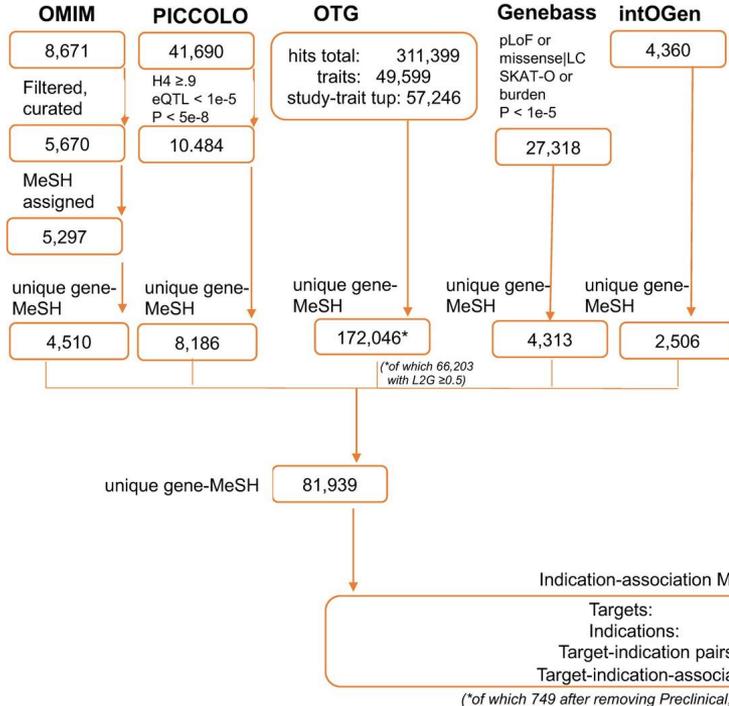
Minikel, Eric Vallabh, Jeffery L. Painter, Coco Chengliang Dong, and Matthew R. Nelson. 2024. "Refining the Impact of Genetic Evidence on Clinical Success." *Nature* 629 (8012): 624–29. <https://doi.org/10.1038/s41586-024-07316-0>.

The dataset and source that is sufficient to reproduce all figures and statistics are [available at GitHub](#), and [the data repository is archived at Zenodo](#). [Source data of the paper is available at nature.com](#).

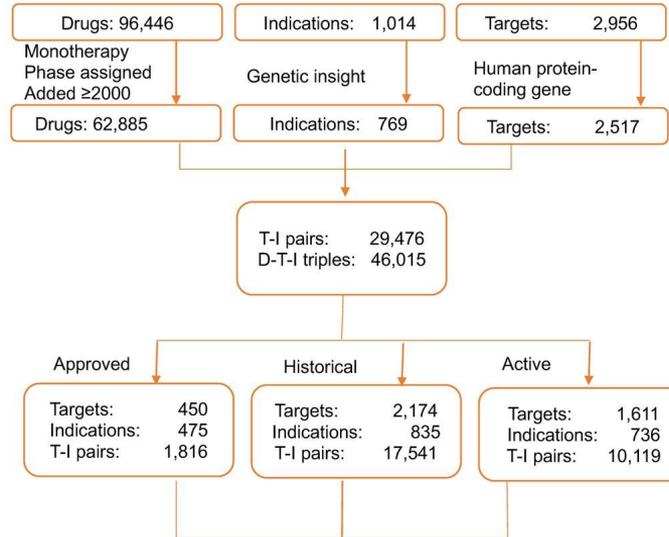
Data processing schematic

a

Human genetic associations
(all counts are gene-trait links unless otherwise specified)



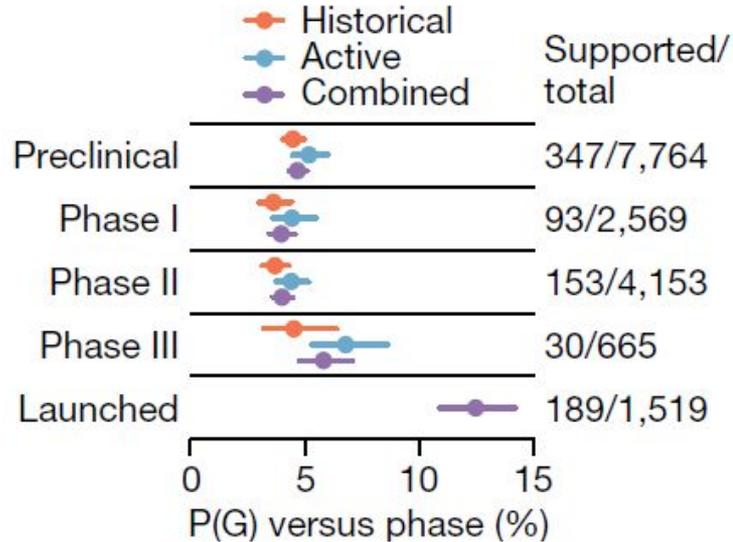
Pharmaprojects



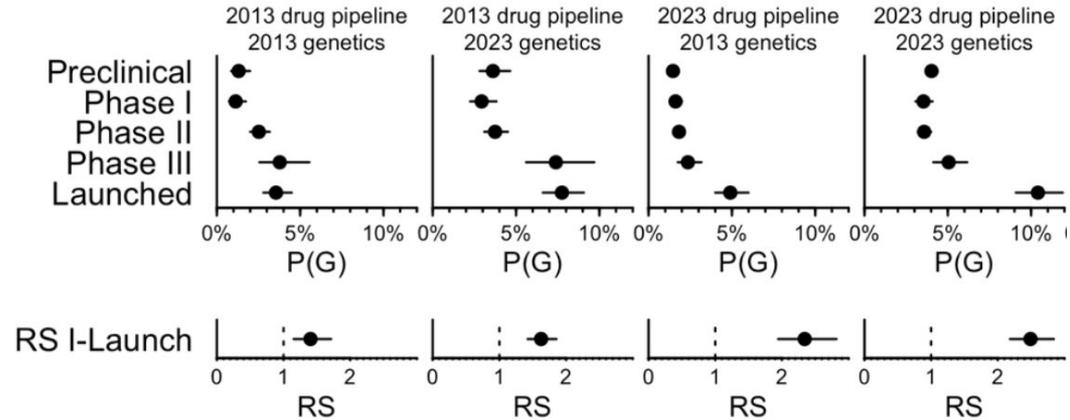
[PICCOLO](#): a GWAS-hits database with gene considering putative causality (GSK).

[intOGen](#): somatic mutations, genes and pathways involved in tumor formation (tumorigenesis). (Barcelona Biomedical Genomics Lab)

Follow-follow-up study Minikel *et al.* 2024

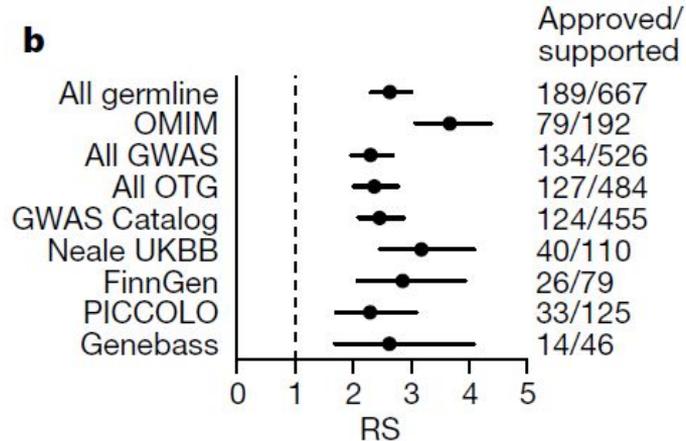


P(G): target-indication pairs with genetic support. Supported/Total: in the unit of target-indication pairs. Error bars represent 95% confidence interval.

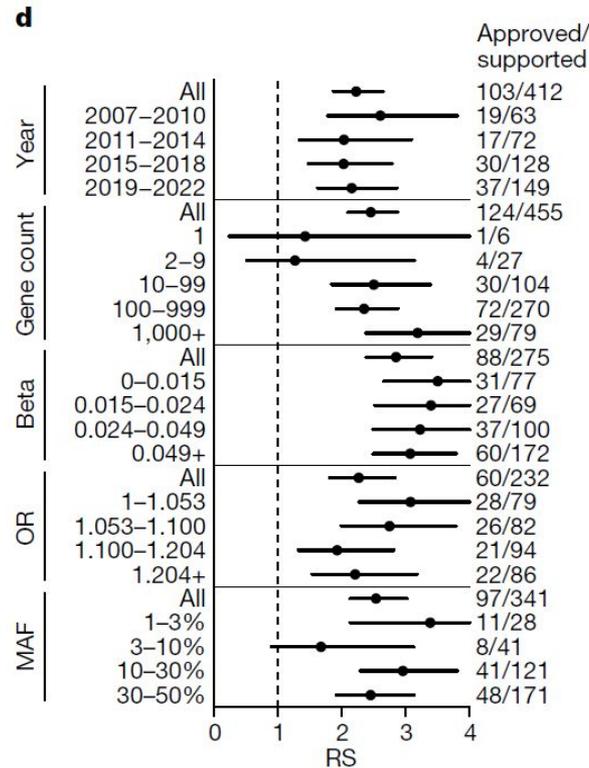


Accumulation of genetic data leads to more targets with genetic support, though only 5-10% target-indication pairs with genetic evidence are exploited. RS=relative success.

The probability of success for drug mechanisms with genetic support is estimated 2.6 times greater than those without



Sensitivity of phase I-launch relative success (RS) to source of human genetic information. *GWAS Catalog*, *Neale UKBB*, and *FinnGen* are subsets of OTG. Approved/supported: number of T-I pairs that were launched/number of those with support from each source.



Year: in which a target-indication pair got first support

Gene count: number of genes associated with the trait that is similar to an indication.

Beta: effect size of an quantitative trait.

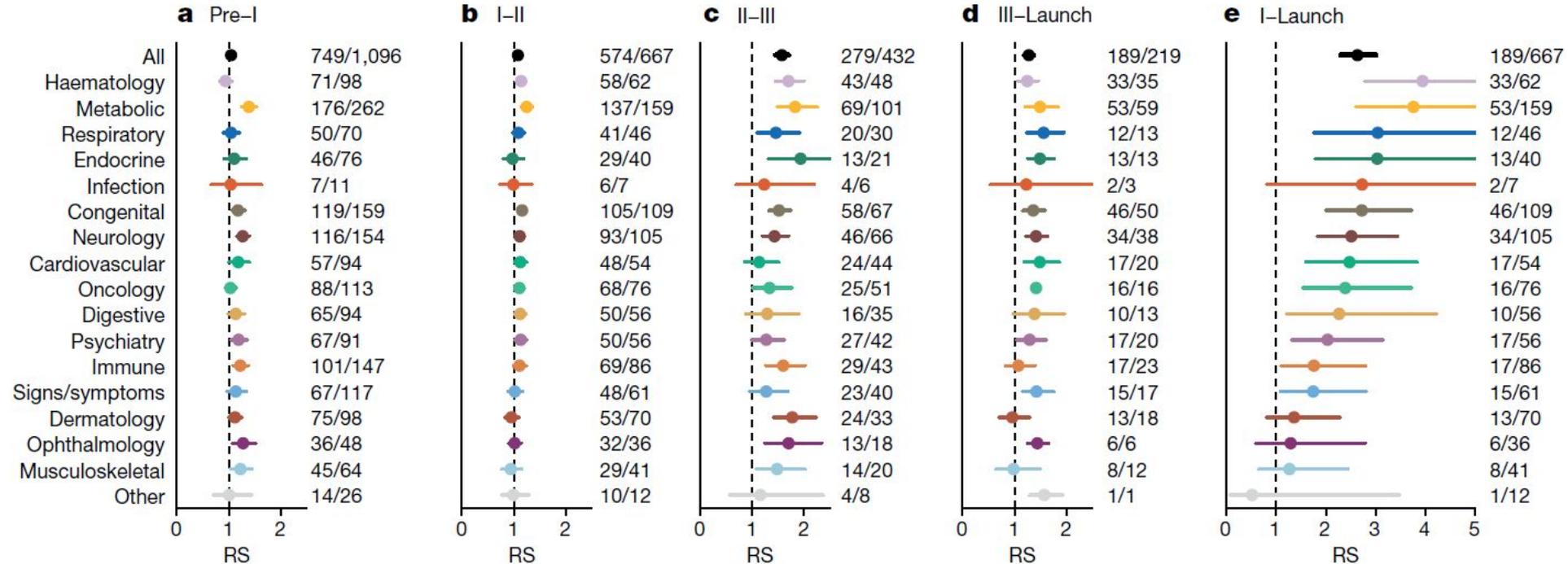
Odds ratio: effect size of a binary trait.

MAF: minor allele frequency

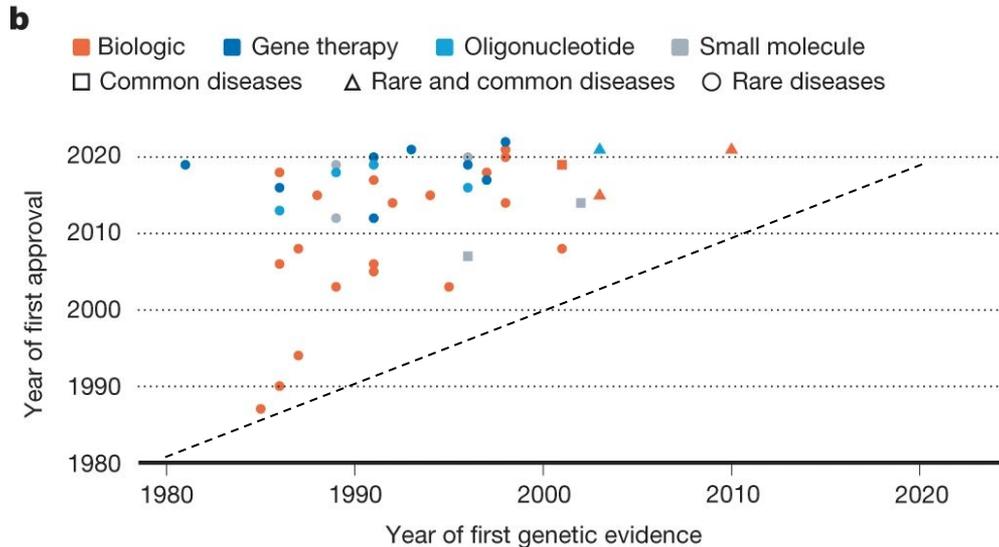
Genetically supported T-I pairs across diseases and target classes

	All genes	Predicted Ab tractable	Predicted SM tractable	Rhodopsin-like GPCRs	Nuclear receptors	Enzymes	Ion channels	Kinases
All areas	719/34,190 2.1%	588/17,811 3.3%	614/12,391 5%	46/508 9.1%	22/245 9%	116/2,278 5.1%	33/714 4.6%	135/1,134 12%
Haematology	42/4,794 0.88%	35/2,409 1.5%	37/1,698 2.2%	2/87 2.3%	0/30 0%	8/276 2.9%	0/58 0%	4/156 2.6%
Metabolic	116/6,819 1.7%	94/3,594 2.6%	95/2,487 3.8%	8/90 8.9%	10/51 20%	14/471 3%	2/98 2%	3/185 1.6%
Respiratory	24/884 2.7%	23/480 4.8%	12/318 3.8%	2/9 22%	0/5 0%	2/51 3.9%	1/26 3.8%	0/13 0%
Endocrine	35/2,676 1.3%	31/1,327 2.3%	31/922 3.4%	7/53 13%	2/32 6.2%	1/187 0.53%	2/53 3.8%	3/80 3.8%
Infection	3/234 1.3%	3/138 2.2%	1/88 1.1%	0/4 0%	0/2 0%	0/13 0%	0/7 0%	0/2 0%
Congenital	87/4,488 1.9%	65/2,444 2.7%	63/1,697 3.7%	2/46 4.3%	2/28 7.1%	15/296 5.1%	1/103 0.97%	2/114 1.8%
Neurology	81/3,481 2.3%	70/1,911 3.7%	65/1,212 5.4%	8/53 15%	0/13 0%	10/198 5.1%	14/133 11%	3/86 3.5%
Cardiovascular	39/2,681 1.5%	37/1,436 2.6%	36/952 3.8%	3/43 7%	1/17 5.9%	10/212 4.7%	10/97 10%	1/90 1.1%
Oncology	295/4,304 6.9%	227/2,088 11%	277/1,975 14%	6/37 16%	4/25 16%	62/419 15%	0/56 0%	109/250 44%
Digestive	38/2,309 1.6%	30/1,252 2.4%	26/802 3.2%	3/47 6.4%	3/21 14%	6/129 4.7%	1/24 4.2%	4/71 5.6%
Psychiatry	40/2,382 1.7%	38/1,300 2.9%	39/786 5%	14/53 26%	0/18 0%	4/131 3.1%	6/84 7.1%	0/74 0%
Immune	66/2,406 2.7%	59/1,301 4.5%	37/855 4.3%	4/40 10%	0/7 0%	10/125 8%	0/26 0%	5/79 6.3%
Signs/symptoms	47/3,629 1.3%	41/1,897 2.2%	41/1,200 3.4%	7/70 10%	4/30 13%	4/206 1.9%	12/124 9.7%	3/107 2.8%
Dermatology	49/2,217 2.2%	43/1,183 3.6%	32/731 4.4%	2/24 8.3%	0/6 0%	7/131 5.3%	0/28 0%	6/71 8.5%
Ophthalmology	23/2,044 1.1%	18/1,153 1.6%	15/626 2.4%	1/19 5.3%	1/18 5.6%	3/108 2.8%	2/54 3.7%	0/61 0%
Musculoskeletal	30/1,620 1.9%	27/887 3%	22/539 4.1%	1/15 6.7%	2/6 33%	2/91 2.2%	1/31 3.2%	4/54 7.4%
Other	7/1,513 0.46%	6/761 0.79%	5/487 1%	0/24 0%	1/11 9.1%	0/93 0%	0/35 0%	1/41 2.4%

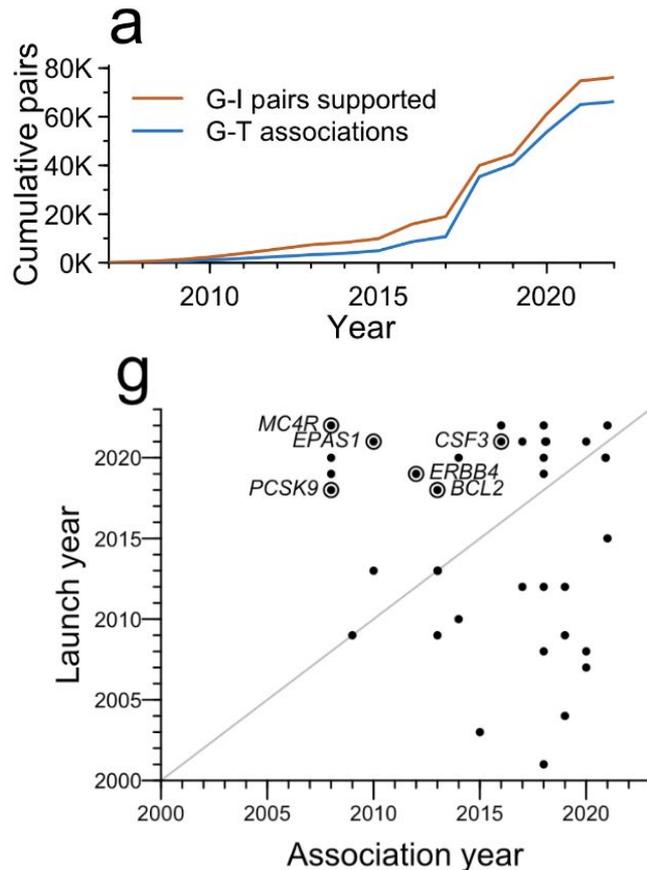
Benefits of genetics-supported targets vary by disease and clinical development phase



Much genetic support nowadays is found retrospectively



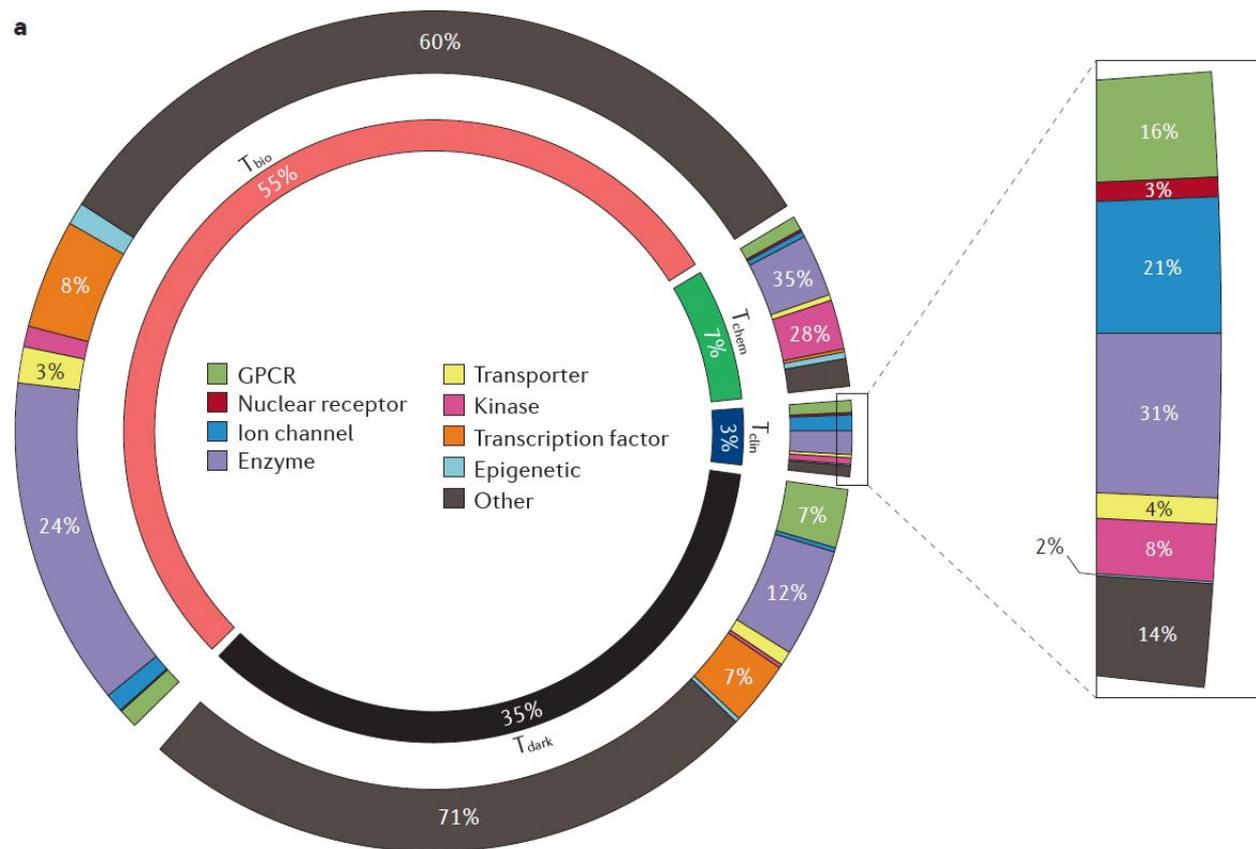
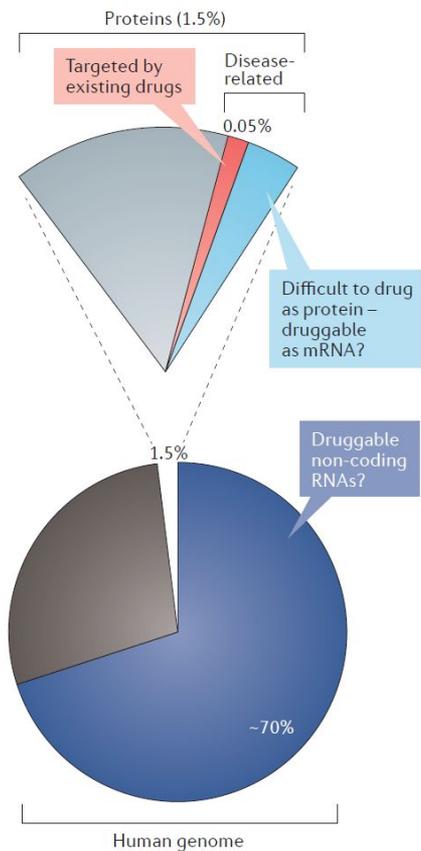
Trajanoska, K. et al. From target discovery to clinical drug development with human genetics. *Nature* 620, 737–745 (2023).



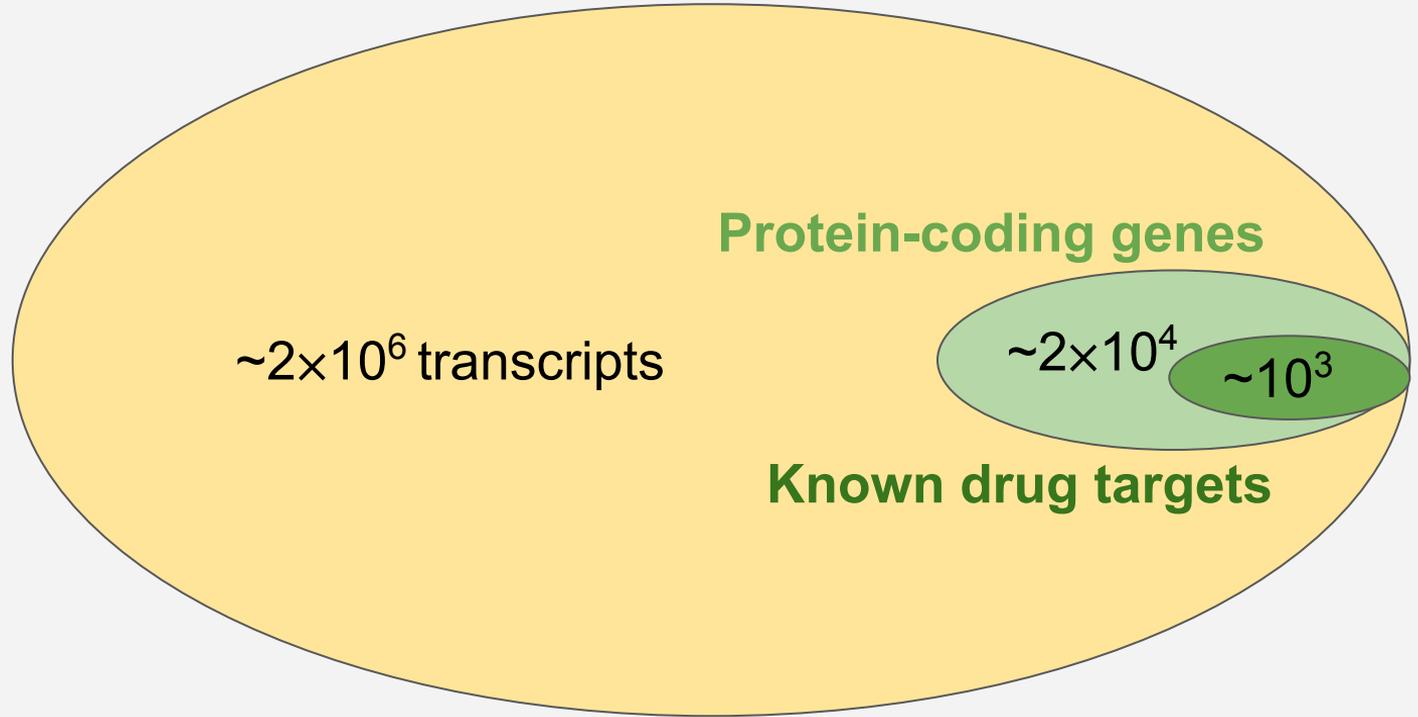
Discussions

- What practical consequences do the findings that genetics increase target-indication success rates by about two fold?
- What other evidences can we use to increase the likelihood that a gene is a good drug target?
- What are the challenges?

Challenge #1: little knowledge on much of the genome

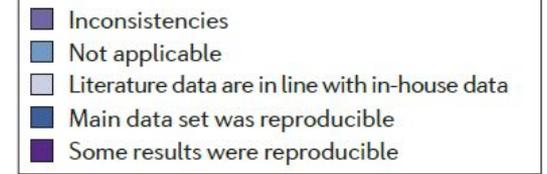
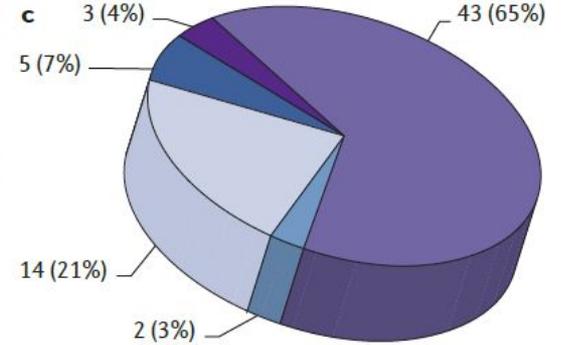
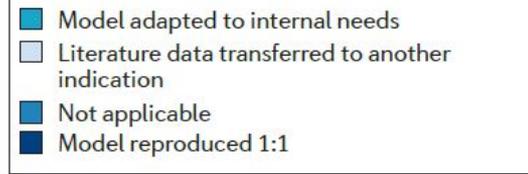
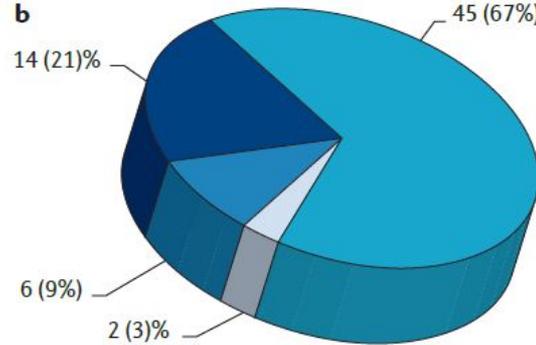
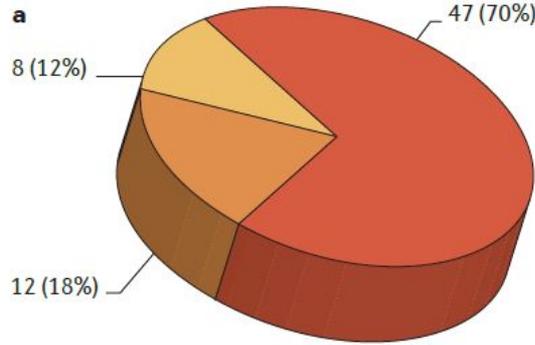


Protein, RNA, or DNA as target?



$\sim 3 \times 10^9$ DNA bases from maternal and paternal each

Challenge #2: Lack of reproducibility

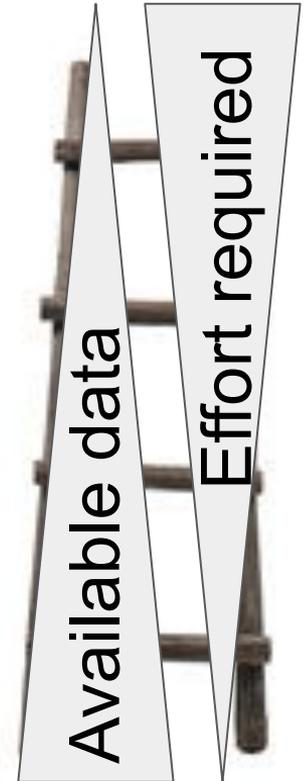


d

	Model reproduced 1:1	Model adapted to internal needs (cell line, assays)	Literature data transferred to another indication	Not applicable
In-house data in line with published results	1 (7%)	12 (86%)	0	1 (7%)
Inconsistencies that led to project termination	11 (26%)	26 (60%)	2 (5%)	4 (9%)

Challenge #3: The Target Ladder

3. [**Real-world test**] What would happen if we inhibit the activity of the kinase domain in the *WKN3* gene in patients with Alzheimer's Disease?
2. [**Intervention**] What would happen to human cells or to a rat model if we inhibit the activity of the kinase domain in the *WKN3* gene?
1. [**Association**] What are the evolutionary conservation, sequence, expression profile, expression regulation patterns, ... of the *WKN3* gene?



Conclusions

- Genomics and genetics offer unprecedented opportunities and challenges for target identification and assessment;
- Drugs targeting targets with genetic evidence support are twice more likely to be successful.
- Central tasks of mathematical and computational biology in drug discovery include (1) performing inference to reduce uncertainty, and (2) identifying causal factors that can be modulated pharmacologically.

**Offline activity of Module I: [submission link](#)
(submission deadline: March 19th, 2026)**

Offline activity of Module I (Part 1)

Read [Refining the Impact of Genetic Evidence on Clinical Success](#) by Minikel *et al.* Report what surprises you most, and submit any questions you may have about the analysis.

Offline activity of Module I (Part 2)

Task 1: The company Fränzi and Friends developed a 2nd-generation quick test at home for Influenza, which is pending regulatory agency's review. The test has been shown to have a sensitivity of 99.5% and a specificity of 99.5%. Suppose that Fred uses the test by Fränzi and Friends and the test was positive. Assume that 5% of the population is in fact infected. What is your guess about the probability that Fred is indeed infected?

Task 2: Please share a piece of code that visualizes the probability that Fred is indeed infected as the dependent variable, with the infection prevalence (5% in the example above, which takes any real-number value between 0.001% to 50%) and the specificity (99% in the example above, which takes values 99%, 99.9%, 99.99%, and 99.999%) as independent variables. For simplicity, we fix the sensitivity at 99%. Visualize the results if possible, and use integers to check and explain your results. Use any programming language that you prefer. Please put your code in GitHub or GitLab or other code-hosting service and paste the link below.

Task 3: What are your interpretations of the results?

Backup slides

References

1. Overington, John P., Bissan Al-Lazikani, and Andrew L. Hopkins. 2006. “How Many Drug Targets Are There?” *Nature Reviews Drug Discovery* 5 (12): 993–96. <https://doi.org/10.1038/nrd2199>.
2. Santos, Rita, Oleg Ursu, Anna Gaulton, A. Patrícia Bento, Ramesh S. Donadi, Cristian G. Bologa, Anneli Karlsson, et al. 2017. “A Comprehensive Map of Molecular Drug Targets.” *Nature Reviews Drug Discovery* 16 (1): 19–34. <https://doi.org/10.1038/nrd.2016.230>.
3. Oprea, Tudor I., Cristian G. Bologa, Søren Brunak, Allen Campbell, Gregory N. Gan, Anna Gaulton, Shawn M. Gomez, et al. 2018. “Unexplored Therapeutic Opportunities in the Human Genome.” *Nature Reviews Drug Discovery* 17 (February): 317–32. <https://doi.org/10.1038/nrd.2018.14>.
4. Finan, Chris, Anna Gaulton, Felix A. Kruger, R. Thomas Lumbers, Tina Shah, Jorgen Engmann, Luana Galver, et al. 2017. “The Druggable Genome and Support for Target Identification and Validation in Drug Development.” *Science Translational Medicine* 9 (383): eaag1166. <https://doi.org/10.1126/scitranslmed.aag1166>.
5. Plenge, Robert M., Edward M. Scolnick, and David Altshuler. 2013. “Validating Therapeutic Targets through Human Genetics.” *Nature Reviews Drug Discovery* 12 (8): 581–94. <https://doi.org/10.1038/nrd4051>.
6. Prinz, Florian, Thomas Schlange, and Khusru Asadullah. 2011. “Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets?” *Nature Reviews Drug Discovery* 10 (9): 712–712. <https://doi.org/10.1038/nrd3439-c1>.
7. Matsui, Masayuki, and David R. Corey. 2017. “Non-Coding RNAs as Drug Targets.” *Nature Reviews Drug Discovery* 16 (3): 167–79. <https://doi.org/10.1038/nrd.2016.117>.
8. Setten, Ryan L., John J. Rossi, and Si-ping Han. 2019. “The Current State and Future Directions of RNAi-Based Therapeutics.” *Nature Reviews Drug Discovery* 18 (6): 421–46. <https://doi.org/10.1038/s41573-019-0017-4>.
9. Warner, Katherine Deigan, Christine E. Hajdin, and Kevin M. Weeks. 2018. “Principles for Targeting RNA with Drug-like Small Molecules.” *Nature Reviews Drug Discovery* 17 (8): 547–58. <https://doi.org/10.1038/nrd.2018.93>.
10. Donovan, Katherine A., Fleur M. Ferguson, Jonathan W. Bushman, Nicholas A. Eleuteri, Debabrata Bhunia, SeongShick Ryu, Li Tan, et al. 2020. “Mapping the Degradable Kinome Provides a Resource for Expedited Degradation Development.” *Cell* 183 (6): 1714–1731.e10. <https://doi.org/10.1016/j.cell.2020.10.038>.

References (continued)

11. Shih, Hsin-Pei, Xiaodan Zhang, and Alex M. Aronov. 2017. “Drug Discovery Effectiveness from the Standpoint of Therapeutic Mechanisms and Indications.” *Nature Reviews Drug Discovery*, October. <https://doi.org/10.1038/nrd.2017.194>.
12. Nguyen, Phuong A., David A. Born, Aimee M. Deaton, Paul Nioi, and Lucas D. Ward. 2019. “Phenotypes Associated with Genes Encoding Drug Targets Are Predictive of Clinical Trial Side Effects.” *Nature Communications* 10 (1): 1–11. <https://doi.org/10.1038/s41467-019-09407-3>.
13. Ding, Li, Matthew H. Bailey, Eduard Porta-Pardo, Vesteynn Thorsson, Antonio Colaprico, Denis Bertrand, David L. Gibbs, et al. 2018. “Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics.” *Cell* 173 (2): 305–320. e10. <https://doi.org/10.1016/j.cell.2018.03.033>.
14. Hon, Chung-Chau, and Piero Carninci. 2020. “Expanded ENCODE Delivers Invaluable Genomic Encyclopedia.” *Nature* 583 (7818): 685–86. <https://doi.org/10.1038/d41586-020-02139-1>.
15. Moore, Jill E., Michael J. Purcaro, Henry E. Pratt, Charles B. Epstein, Noam Shoshitaishvili, Jessika Adrian, Trupti Kawli, et al. 2020. “Expanded Encyclopaedias of DNA Elements in the Human and Mouse Genomes.” *Nature* 583 (7818): 699–710. <https://doi.org/10.1038/s41586-020-2493-4>.
16. Kim-Hellmuth, Sarah, François Aguet, Meritxell Oliva, Manuel Muñoz-Aguirre, Silva Kasela, Valentin Wucher, Stephane E. Castel, et al. 2020. “Cell Type–Specific Genetic Regulation of Gene Expression across Human Tissues.” *Science* 369 (6509). <https://doi.org/10.1126/science.aaz8528>.
17. Uhlén, Mathias, Linn Fagerberg, Björn M. Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, et al. 2015. “Tissue-Based Map of the Human Proteome.” *Science* 347 (6220): 1260419. <https://doi.org/10.1126/science.1260419>.
18. Wang, Dongxue, Basak Eraslan, Thomas Wieland, Björn Hallström, Thomas Hopf, Daniel Paul Zolg, Jana Zecha, et al. 2019. “A Deep Proteome and Transcriptome Abundance Atlas of 29 Healthy Human Tissues.” *Molecular Systems Biology* 15 (2): e8503. <https://doi.org/10.15252/msb.20188503>.
19. Wang, Mingcong, Christina J. Herrmann, Milan Simonovic, Damian Szklarczyk, and Christian von Mering. 2015. “Version 4.0 of PaxDb: Protein Abundance Data, Integrated across Model Organisms, Tissues, and Cell-Lines.” *PROTEOMICS* 15 (18): 3163–68. <https://doi.org/10.1002/pmic.201400441>.
20. Emmerich, Christoph H., Lorena Martinez Gamboa, Martine C. J. Hofmann, Marc Bonin-Andresen, Olga Arbach, Pascal Schendel, Björn Gerlach, et al. 2021. “Improving Target Assessment in Biomedical Research: The GOT-IT Recommendations.” *Nature Reviews Drug Discovery* 20 (1): 64–81. <https://doi.org/10.1038/s41573-020-0087-3>.

References (continued)

21. Cao, Chen, and John Moulton. 2014. “GWAS and Drug Targets.” *BMC Genomics* 15 (4): S5. <https://doi.org/10.1186/1471-2164-15-S4-S5>.
22. Patron, Jonas, Arnau Serra-Cayuela, Beomsoo Han, Carin Li, and David Scott Wishart. 2019. “Assessing the Performance of Genome-Wide Association Studies for Predicting Disease Risk.” *PLOS ONE* 14 (12): e0220215. <https://doi.org/10.1371/journal.pone.0220215>.
23. <https://www.genome.gov/About-Genomics/Introduction-to-Genomics>
24. Jansen, An, and Kevin J. Verstrepen. 2011. “Nucleosome Positioning in *Saccharomyces Cerevisiae*.” *Microbiology and Molecular Biology Reviews* 75 (2): 301–20. <https://doi.org/10.1128/MMBR.00046-10>.
25. https://gnomad.broadinstitute.org/gene/ENSG00000141510?dataset=gnomad_r2_1
26. Collins, Ryan L., Harrison Brand, Konrad J. Karczewski, Xuefang Zhao, Jessica Alfoldi, Laurent C. Francioli, Amit V. Khera, et al. 2020. “A Structural Variation Reference for Medical and Population Genetics.” *Nature* 581 (7809): 444–51. <https://doi.org/10.1038/s41586-020-2287-8>.
27. NCBI Gene of CFTR: <https://www.ncbi.nlm.nih.gov/gene/1080>
28. Hughes, Maria F., Yvonne M. Lenighan, Catherine Godson, and Helen M. Roche. 2018. “Exploring Coronary Artery Disease GWAS Targets With Functional Links to Immunometabolism.” *Frontiers in Cardiovascular Medicine* 5. <https://doi.org/10.3389/fcvm.2018.00148>.
29. Consortium, The GTEx. 2020. “The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues.” *Science* 369 (6509): 1318–30. <https://doi.org/10.1126/science.aaz1776>.
30. Baralle, Marco, and Francisco Ernesto Baralle. 2018. “The Splicing Code.” *Biosystems, Code Biology*, 164 (February): 39–48. <https://doi.org/10.1016/j.biosystems.2017.11.002>.

References (continued)

32. https://en.wikipedia.org/wiki/Haemophilia_in_European_royalty
33. Smemo, Scott, Juan J. Tena, Kyoung-Han Kim, Eric R. Gamazon, Noboru J. Sakabe, Carlos Gómez-Marín, Ivy Aneas, et al. 2014. “Obesity-Associated Variants within FTO Form Long-Range Functional Connections with IIRX3.” *Nature* 507 (7492): 371–75. <https://doi.org/10.1038/nature13138>.
34. Nelson, Matthew R., Hannah Tipney, Jeffery L. Painter, Judong Shen, Paola Nicoletti, Yufeng Shen, Aris Floratos, et al. 2015. “The Support of Human Genetic Evidence for Approved Drug Indications.” *Nature Genetics* 47 (8): 856–60. <https://doi.org/10.1038/ng.3314>.
35. Ardlie, Kristin G., Leonid Kruglyak, and Mark Seielstad. 2002. “Patterns of Linkage Disequilibrium in the Human Genome.” *Nature Reviews Genetics* 3 (4): 299–309. <https://doi.org/10.1038/nrg777>.
36. Brennan, Richard J. 2017. “Target Safety Assessment: Strategies and Resources.” In *Drug Safety Evaluation: Methods and Protocols*, edited by Jean-Charles Gautier, 213–28. *Methods in Molecular Biology*. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-7172-5_12.
37. Lambert, Samuel A., Arttu Jolma, Laura F. Campitelli, Pratyush K. Das, Yimeng Yin, Mihai Albu, Xiaoting Chen, Jussi Taipale, Timothy R. Hughes, and Matthew T. Weirauch. “The Human Transcription Factors.” *Cell* 172, no. 4 (February 8, 2018): 650–65. <https://doi.org/10.1016/j.cell.2018.01.029>.
38. Dann, Emma, Erin Teeple, Rasa Elmentaite, Kerstin B. Meyer, Giorgio Gaglia, Frank Nestle, Virginia Savova, Emanuele de Rinaldis, and Sarah A. Teichmann. 2024. “Estimating the Impact of Single-Cell RNA Sequencing of Human Tissues on Drug Target Validation.” medRxiv. <https://doi.org/10.1101/2024.04.04.24305313>.
39. Minikel, Eric Vallabh, Jeffery L. Painter, Coco Chengliang Dong, and Matthew R. Nelson. 2024. “Refining the Impact of Genetic Evidence on Clinical Success.” *Nature* 629 (8012): 624–29. <https://doi.org/10.1038/s41586-024-07316-0>.

Compare drug-induced teratogenicity and liver injury (DILI)

[The DILIRank dataset](#) comprises 1336 FDA-approved drugs. Each drug is assigned to one of four categories based on its potential to cause drug-induced liver injury (DILI):

- Most DILI concern (217)
- Less DILI concern (351)
- No DILI concern (414)
- Undetermined causality/Ambiguous DILI concern (354)

The classification is derived from a comprehensive review of hepatotoxicity descriptions in FDA-approved labeling and causality evidence from the literature.

Assume that we have an assay with 90% sensitivity and 85% specificity. How useful is it for drug discovery? Comparing this with the example of teratogenicity, what's your conclusion?

An example of complementary views

We want to work on hepatocarcinoma (liver cancer) and have the following information about a potential target X:

- X is a receptor expressing on the surface of most cell types;
- Upon binding ligands, X activates innate immune response;
- Gene sequence of X is conserved in primates but *not* in rodents;
- Protein X interacts with protein Y, which is essential, namely Y knockout causes lethal embryos;
- Asian population has a unique genetic variant in the non-coding region of X;

Discussion: what are the consequences of having these information?

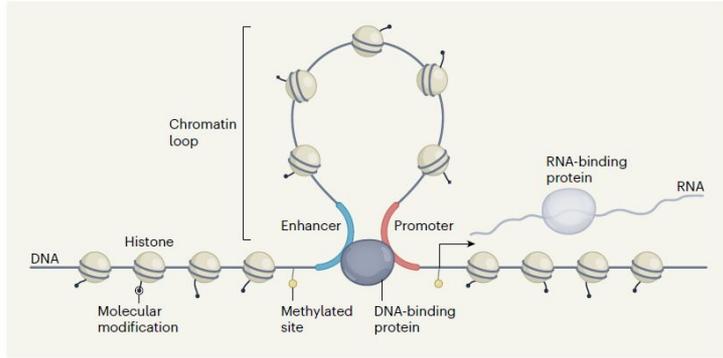
Questions from courses

- Why I recommended the GOT-IT paper? How can academia and industry work together towards good targets?
- Did sequencing cost always follow the Moore's law?
- What happens if there are ATGs (AUGs in RNA) in 5'-untranslated region?
 - In some cases, there are alternative start codons;
 - However, in most cases, the ATGs in 5'-untranslated region seem to be always ignored by the translational machinery. [A study](#) (Rogozin, Bioinformatics, 2001) suggested that those AUGs may ensure low basal expression and generate regulatory elements.
- Which target level (gene or protein) is more useful for the target identification?
In the lecture, gene-level approach seems not so promising.
- Is blue eyeness a marker of Neanderthalian origin? *The story of OAC2*

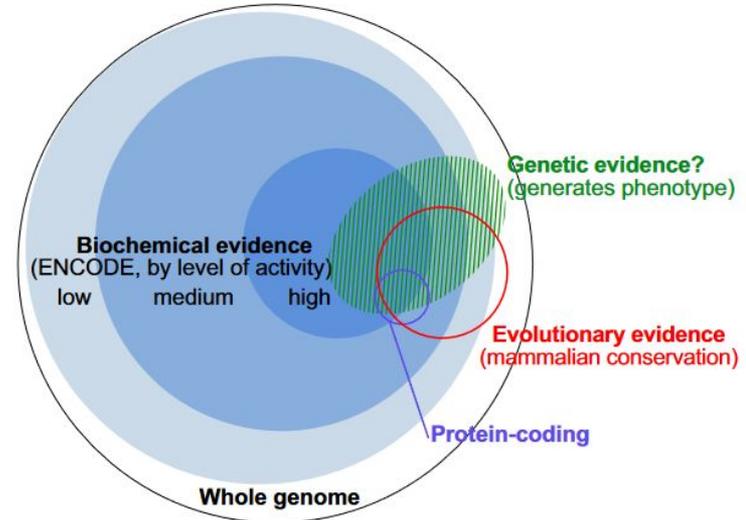
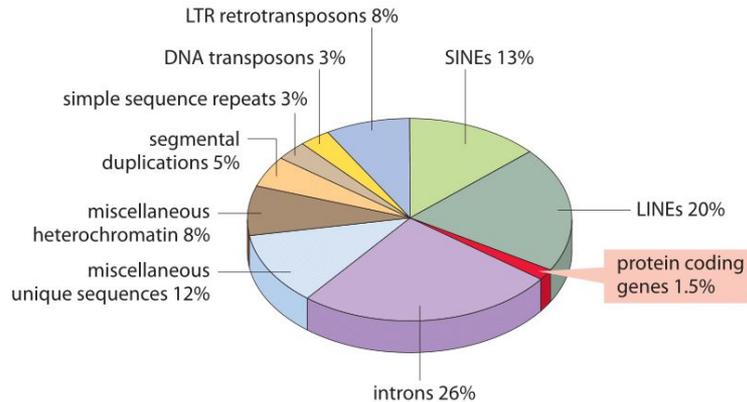
Why autoimmune diseases are more prevalent in females, though one X chromosome is randomly inactivated?

- Sex hormone signaling plays an important role in immune functions, especially estrogens. The hormone signalling apparently explains a lot, but not all, sex differences in autoimmune diseases.
- Mutations of genes on the X-chromosome, as expected, cause many primary immunodeficiencies only in males, because they have only one copy of the X chromosome.
- One of the two X chromosomes in females indeed get inactivated during the embryo stage. However, about 15-20% genes regularly escape the inactivation, among others important genes involved in innate and adaptive immune response, including TLR7 and CD40L.
- There are a few other hypotheses besides X-inactivation escaping, including loss of mosaicism, reactivation, and haploinsufficiency.

Much of the genome is junk, some is regulatory



main components of the human genome



1. Gregory, T. R. Synergy between sequence and size in Large-scale genomics. *Nat Rev Genet* 6, 699–708 (2005).
2. Kellis, M. et al. Defining functional DNA elements in the human genome. *Proceedings of the National Academy of Sciences* 111, 6131–6138 (2014).

Genome biology in one screenshot

The Human Genome and Variations

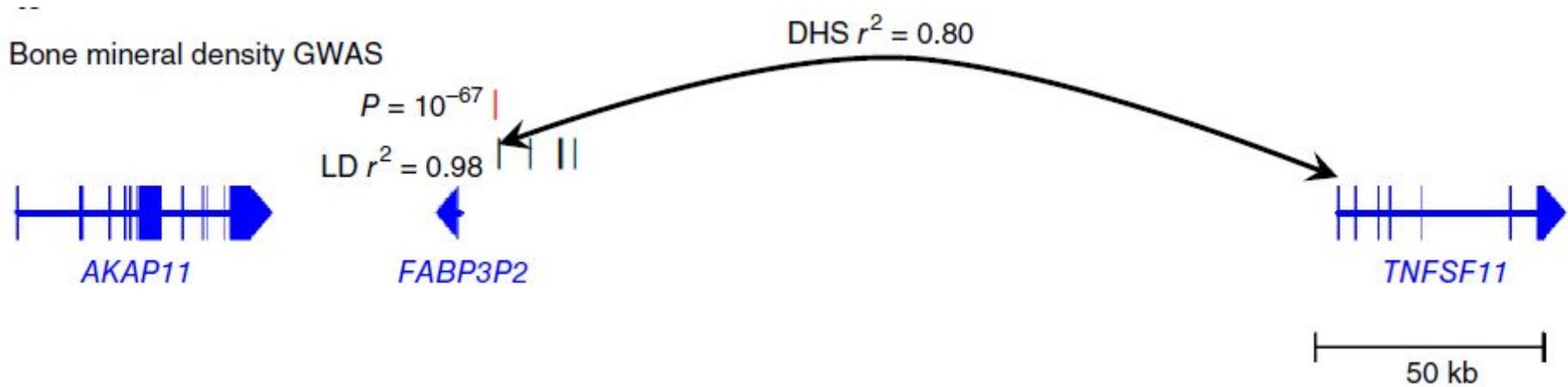
Gene Structure and gene expression

DNA and RNA sequencing

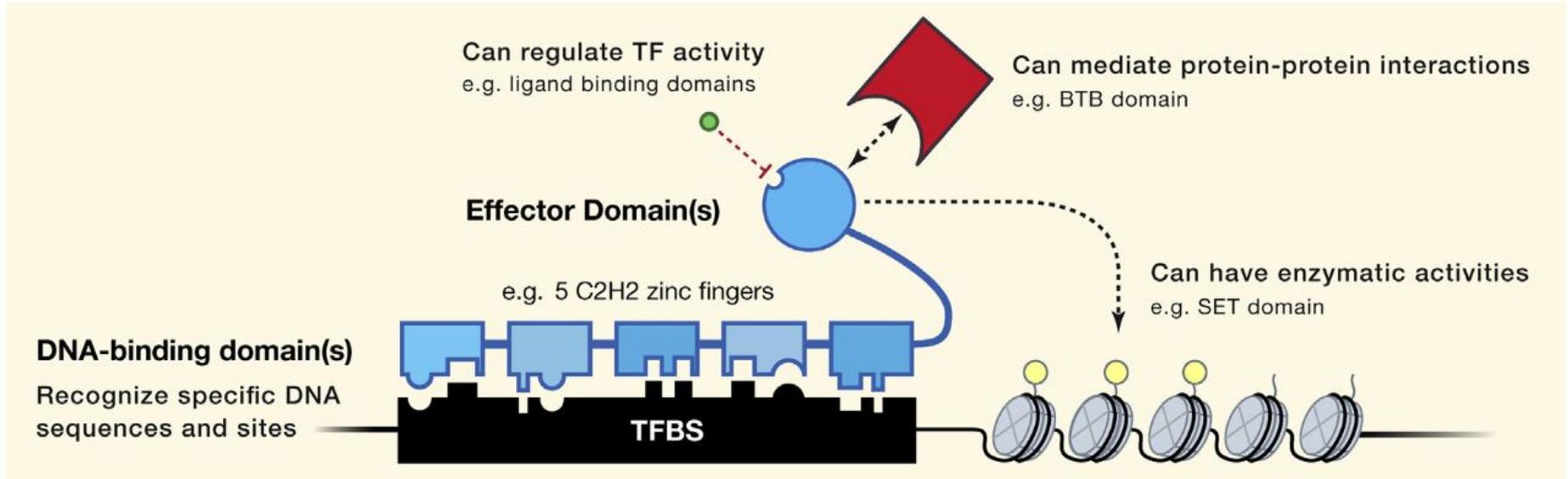


ACE2 viewed in [NCBI Genome Browser](https://www.ncbi.nlm.nih.gov/genome/browser)

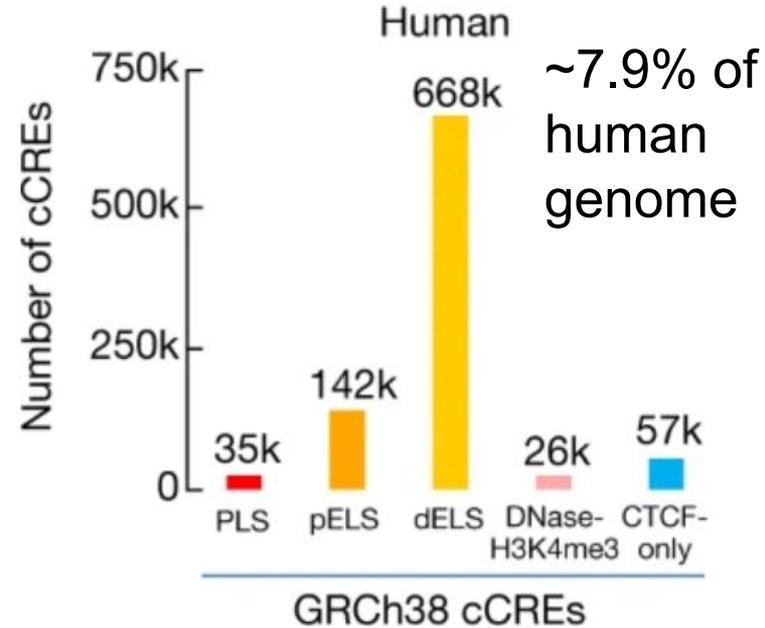
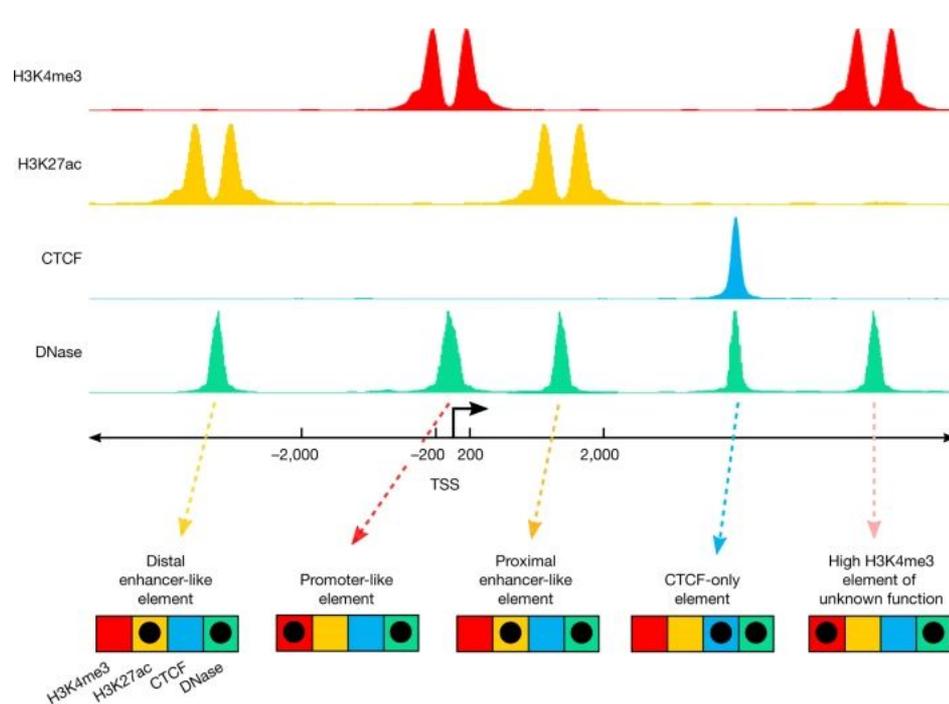
Correlation between DNase I hypersensitive (DHS) sites helps linking genetic variants with genes



Transcription factors induce gene expression



TFs bind to candidate cis-regulatory elements (cCRE) to regulate gene expression



<https://screen.encodeproject.org/>

GOT-IT recommendations for target-disease linkage

Assessment blocks



AB1: target-disease linkage (human targets)

1. Is the target perturbation a cause or consequence of the human disease process?
2. Is the therapeutic relevance (such as human connection) of models used sufficiently high for decision-making?
3. Is the target expression pattern known (that is, within the anticipated patient population)?
4. Is the target manipulation process clinically relevant?
5. Is the read-out used to detect target-dependent processes disease-relevant?
6. Is the stimulus used to activate or influence target-dependent processes disease-relevant?
7. Are the biological consequences of an observed effect size known?

Public resources for target assessment

AB1: target–disease linkage (human targets)

1. Is the target perturbation a cause or consequence of the human disease process?
2. Is the therapeutic relevance (such as human connection) of models used sufficiently high for decision-making?
3. Is the target expression pattern known (that is, within the anticipated patient population)?
4. Is the target manipulation process clinically relevant?
5. Is the read-out used to detect target-dependent processes disease-relevant?
6. Is the stimulus used to activate or influence target-dependent processes disease-relevant?
7. Are the biological consequences of an observed effect size known?

- [OpenTargets](#)
- [Online Mendelian Inheritance in Man](#) (OMIM)

- Scattered in diverse information sources such as [Wikipedia](#) and literature

- Health: [GTEx](#), [The Human Protein Atlas](#)
- Disease: [Gene Expression Atlas](#), scattered

Public resources for target safety assessment

AB2: target-related safety (human targets)

8. Is the target selective and not genetically linked to other diseases (or phenotypes or organ systems)?
9. Is there prior knowledge on safety of the target or reported evidence for the role of the target in a known pathway and/or physiological process that may be harmful if disrupted?
10. Are in vitro or pharmacologically relevant animal models available for safety testing?
11. Do models used for safety testing translate well to humans?
12. Are safety biomarkers available and can adverse effects be monitored and/or predicted by safety biomarkers?
13. Is there sufficient confidence that a necessary safety window has been or can be established?
14. Is the disease life-threatening (at what stage of the disease is the target of relevance)?
15. Is the tissue distribution of the target known (in humans or in animals)?

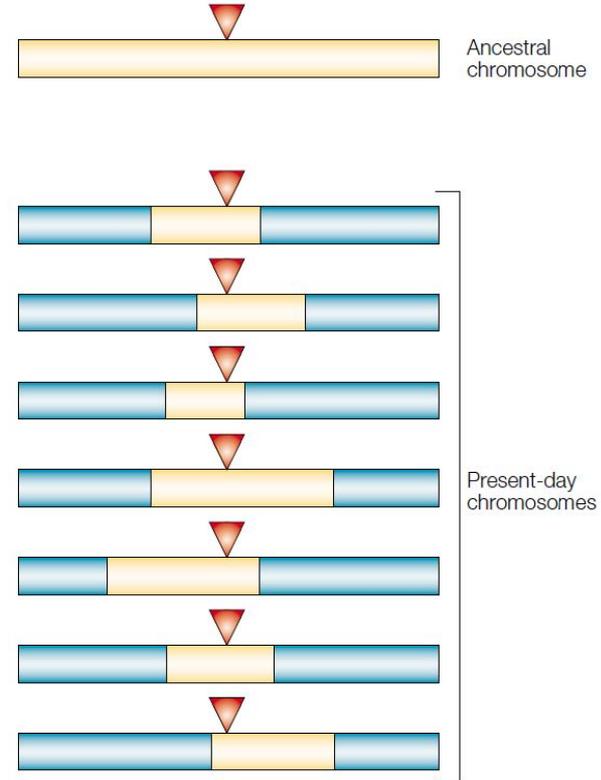
- [Comparative Toxicogenomics Database \(CTD\)](#)

- [DrugBank, DrugCentral](#)
- [FDA Adverse Event Reporting System \(FAERS\)](#)

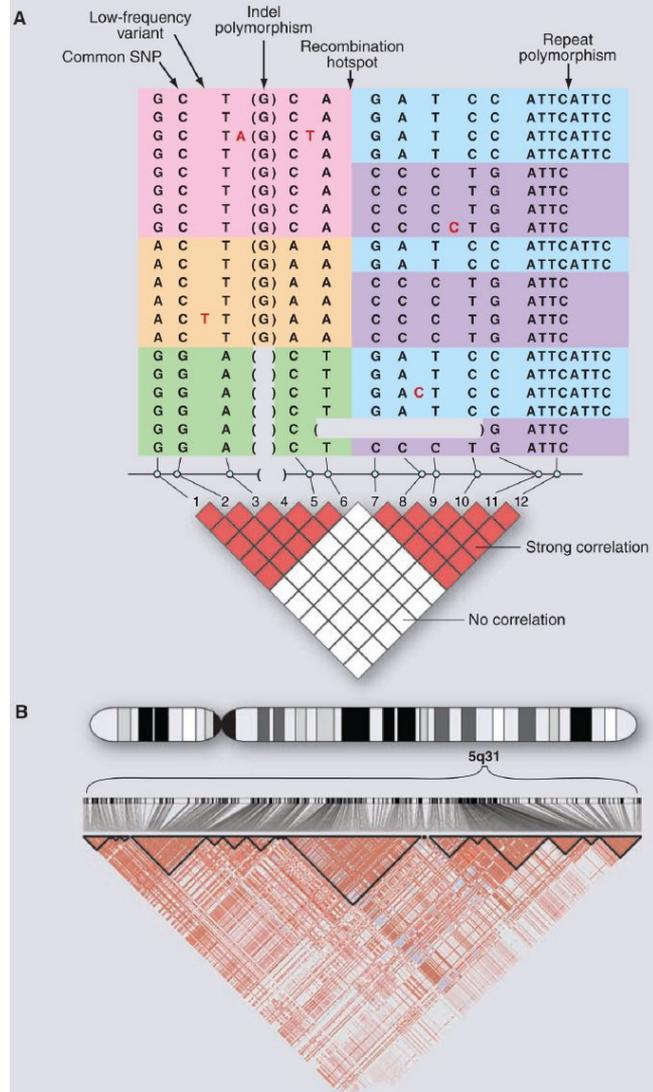
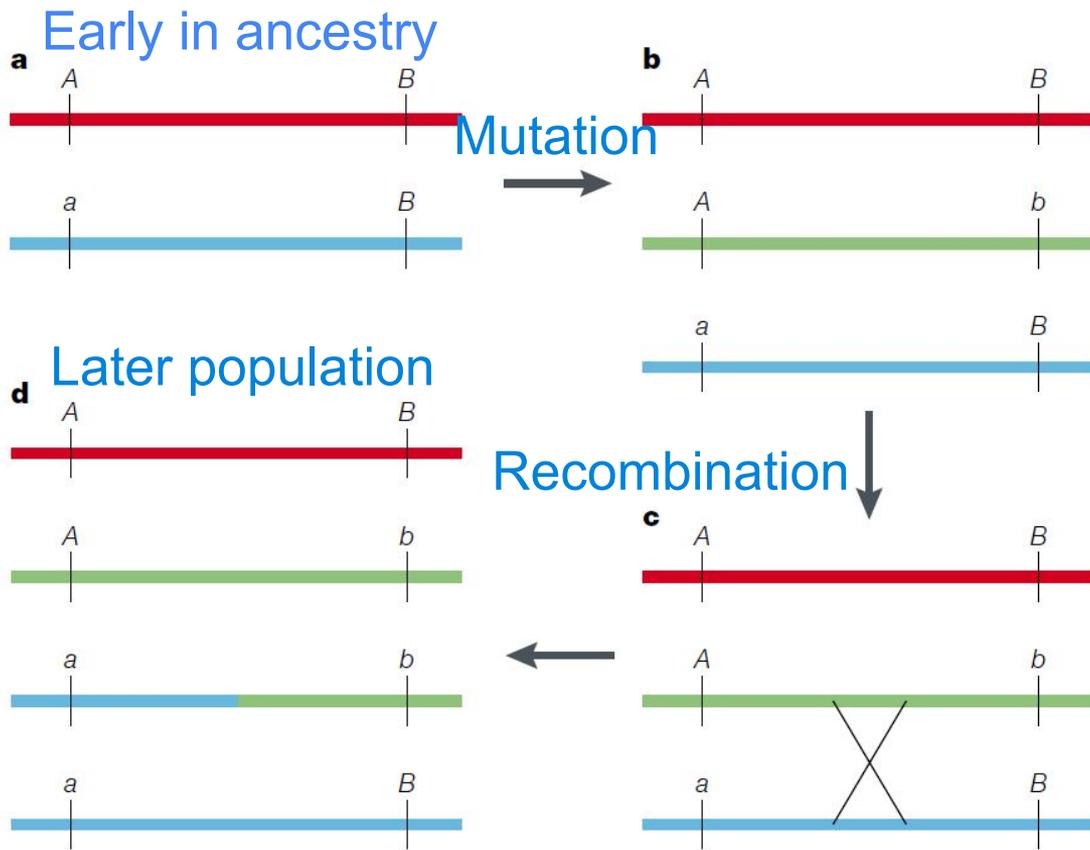
- [NCBI HomoloGene](#)
- [ENSEMBL ComparaGenom](#)
- [Mouse Genome Informatics \(MGI\)](#)

Linkage Disequilibrium in human genome

Particular alleles (single gene copies) at neighbouring loci tend to be co-inherited. For tightly linked loci, this might lead to associations between alleles in the population. This property is known as linkage disequilibrium (LD).



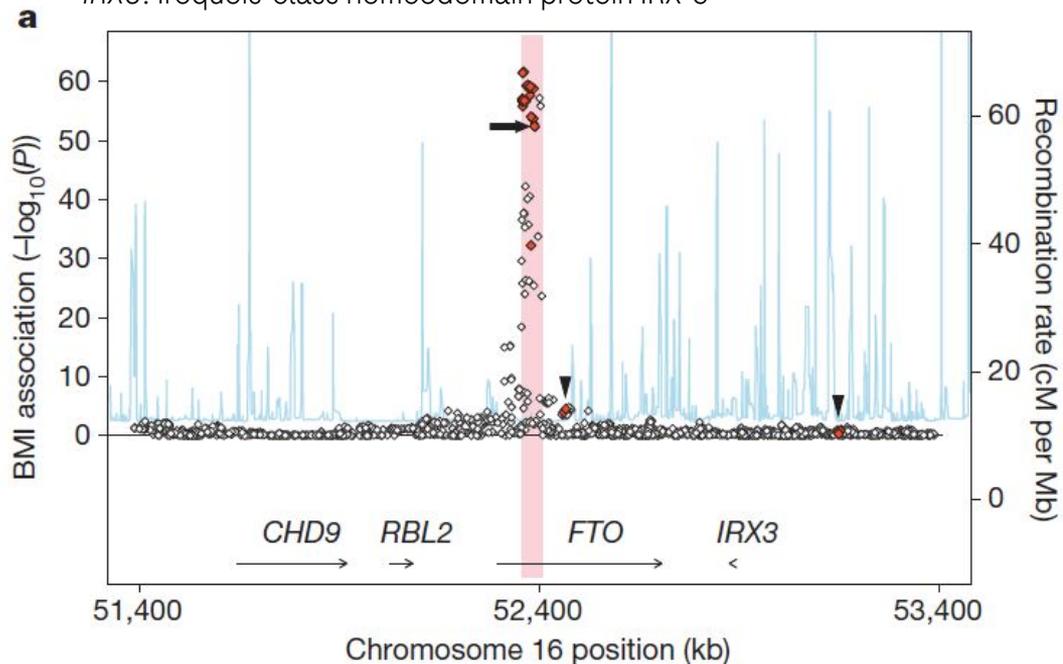
Population genetics helps with disease mapping



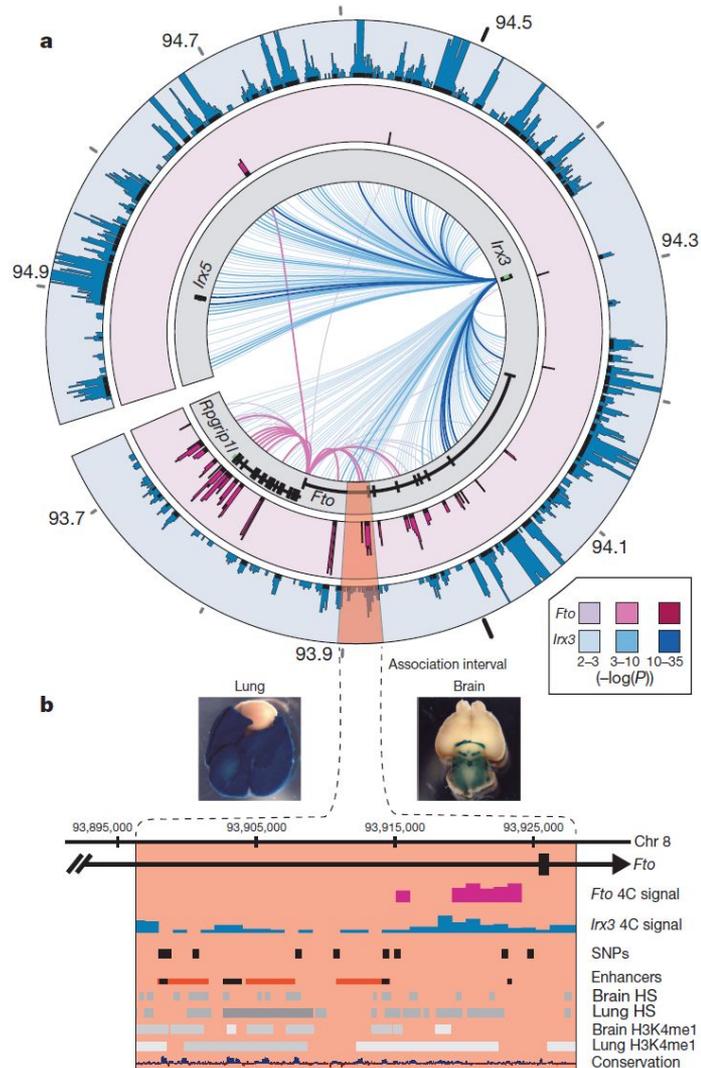
Is FTO a good target for obesity?

FTO: fat mass and obesity-associated gene, which hosts [rs9930506](#)

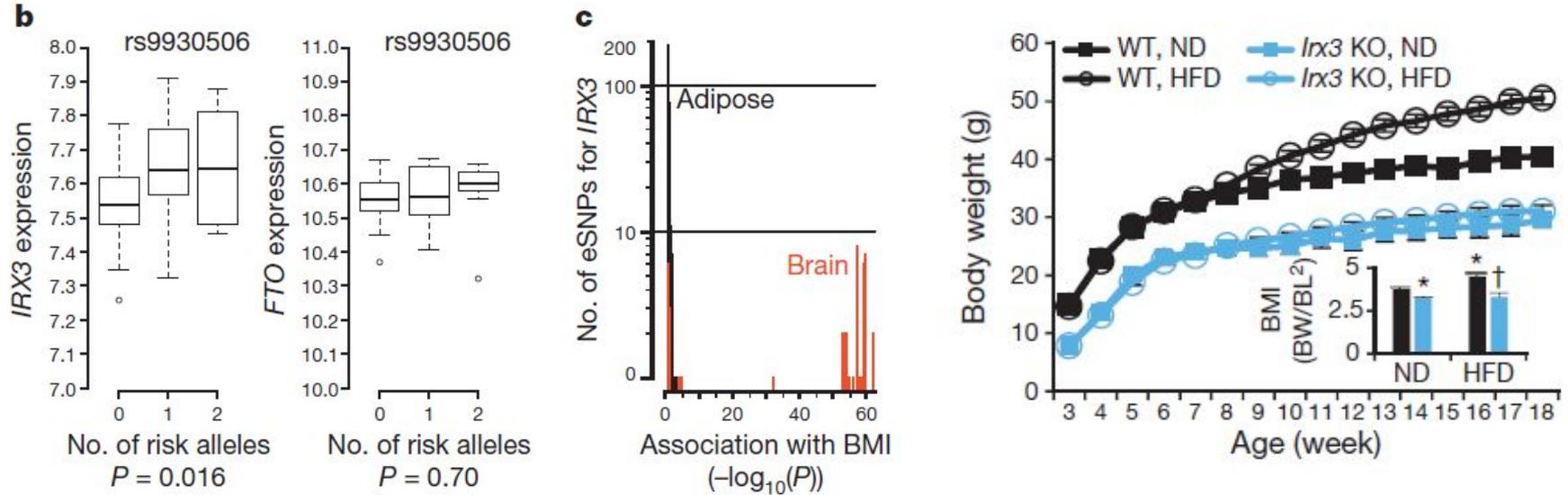
IRX3: Iroquois-class homeodomain protein IRX-3



Smemo, S. et al. Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*. *Nature* 507, 371–375 (2014).



SNPs in *FTO* gene interacts with promoter of *IRX3*, the knockout of which reduces body weight



Other important information resources

- **Genomic variations:** [gnomAD](#), [dbSNP](#), and [TCGA](#) for oncology;
- **Protein domain and static structure:** [InterPro](#), [Pfam](#), and [PDB](#);
- **Interaction network and pathway:** [BioGRID](#), [IntAct](#), [Reactome](#), and [KEGG](#);
- **Gene expression profiles associated with the target:** [NCBI GEO](#) (Gene Expression Omnibus), [ARCHS4](#)