Emerging technologies in genomics

Mark Corbett, Ph.D

mark.corbett at adelaide.edu.au

THE UNIVERSITY of ADELAIDE

Nature *vs.* Nurture
(Genes *vs.* Environment)

Jerome

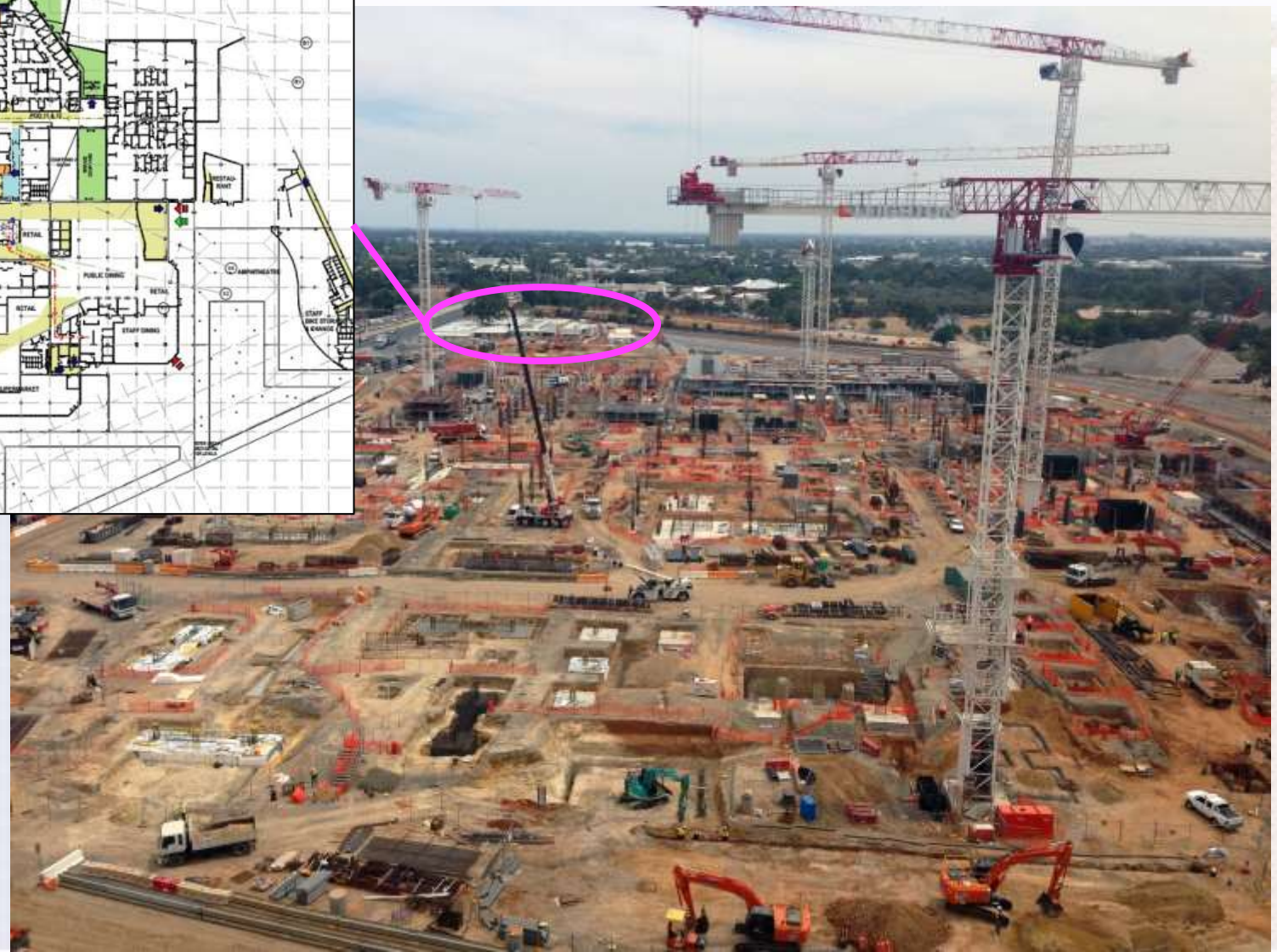
- Selective breeding
- “Ideal” genome
- Poor environment
- Tendency towards self harm

Vincent

- Naturally born
- High risk for heart disease
- Strong willed

Hypothesis to test?

What is a genome?



Instructions for life

Can have notes or modifications, eg. Change the size of the surgery (epigenome)

What's in a human genome?

- 3,609,003,417 base pairs (*i.e.*. x2)
  - Genes (1.5%)
  - Regulatory sequences (6-8%)
  - Sequences with no, or as yet unknown function (All the other bits)
  - 329,816,417 known possible small variations
  - 6,013,009 known possible large variations

Feature	Primary	Alternate
Coding genes	20,376	2,954
Small non-coding genes	5,363	293
Long non-coding genes	14,720	964
"Misc." non-coding genes	2,222	177
Pseudogenes	14,692	1,740
Gene transcripts	203,903	N/A

Note: These numbers change regularly!
http://asia.ensembl.org/Homo_sapiens/Info/Annotation

Health = Balance between the environment and what the genome can or can't protect us from

Gene to gene interaction

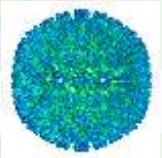
Protection

Genes
(Variants)

Susceptibility

Trauma

Pathogens



Viral

Environment

Chemical

Diet

Genome variation and common disease

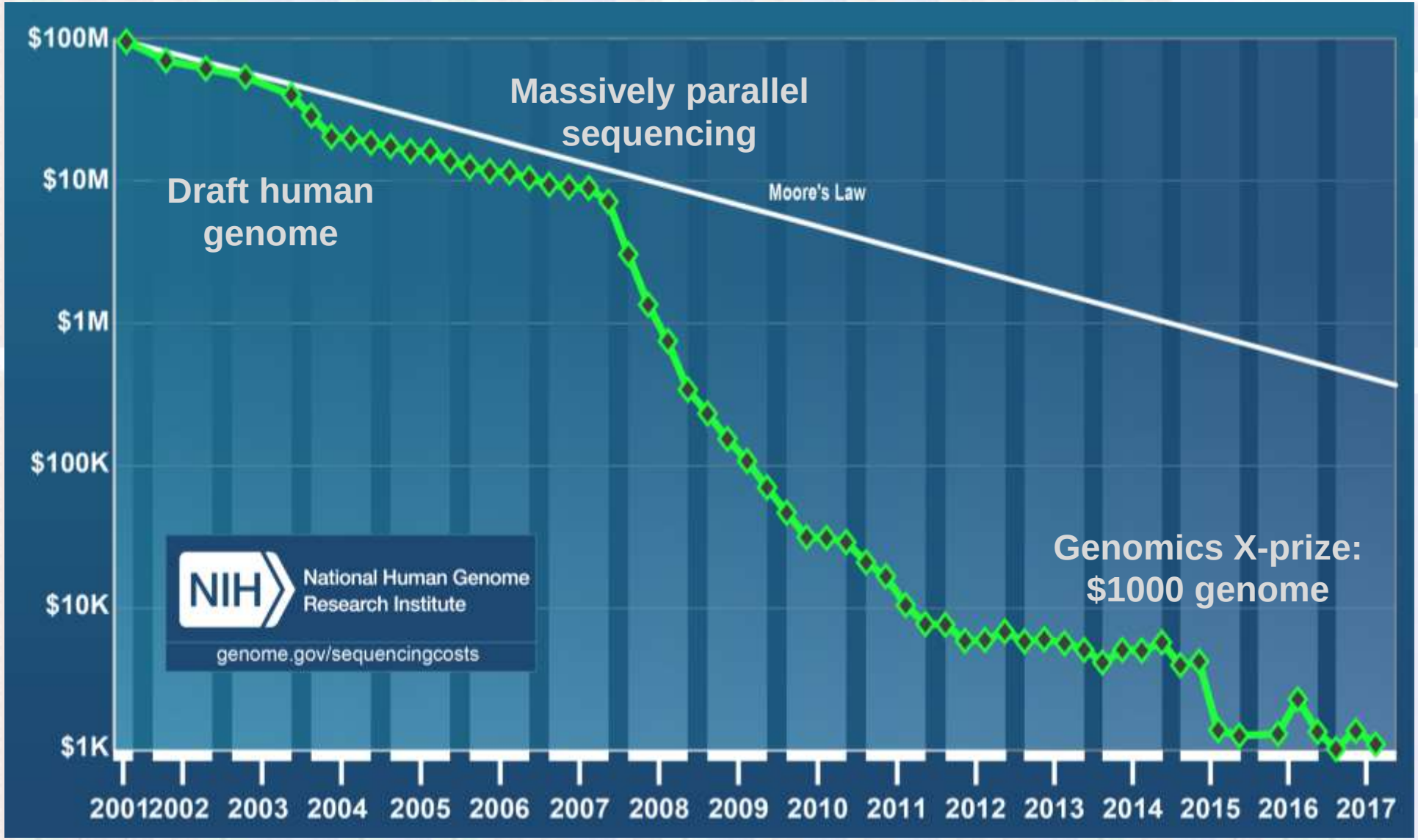
Genome wide association (GWAS) catalogue

- Each dot represents a variation in the genome significantly associated with a trait
- **Risk factors only**

- 828 loci implicated in the nervous system
  - **47** Bipolar disorder
  - **137** Schizophrenia
  - **86** Autism spectrum disorder
  - **126** Level of educational attainment

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

Costs per genome

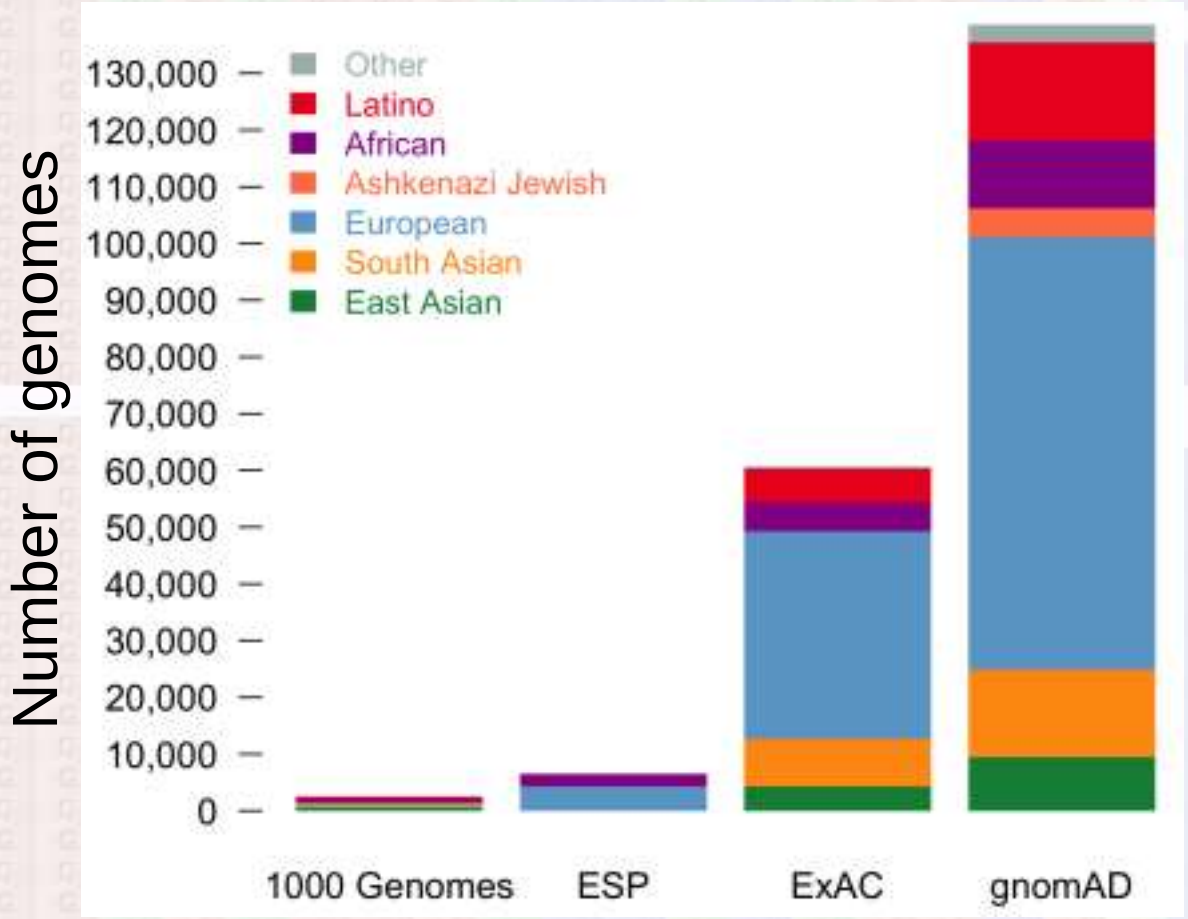


www.genome.gov/sequencingcostsdata. Accessed: 30/04/2018

Born with your genome and epigenome

- **Parental** height, weight, age, socio-economic status, risk behaviour (eg. Smoking, alcohol)
- **Baby** length, weight, conditions at birth (breathing, response, mobility), metabolic profile in blood
- Whole genome sequencing
  - **Anonymised**, federated, aggregate data
  - Genetic risk factors
  - Early detection and intervention
  - Lifetime health plan
- Epigenome
  - Environmental initial conditions

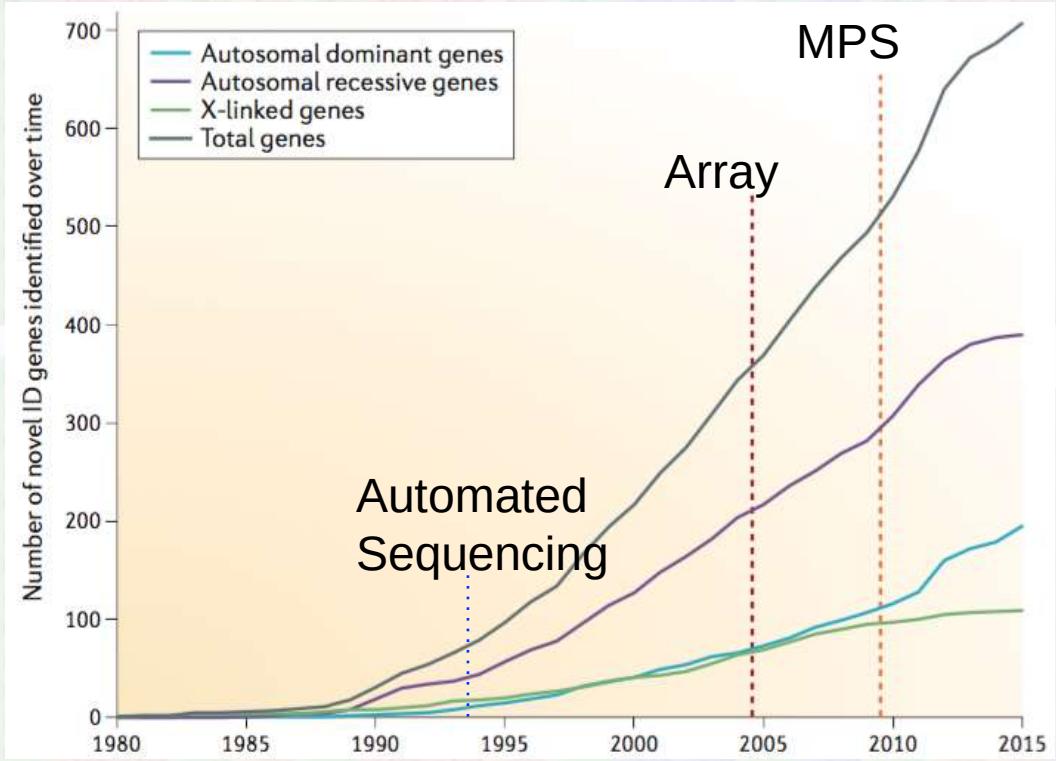
Genome variation data aggregation

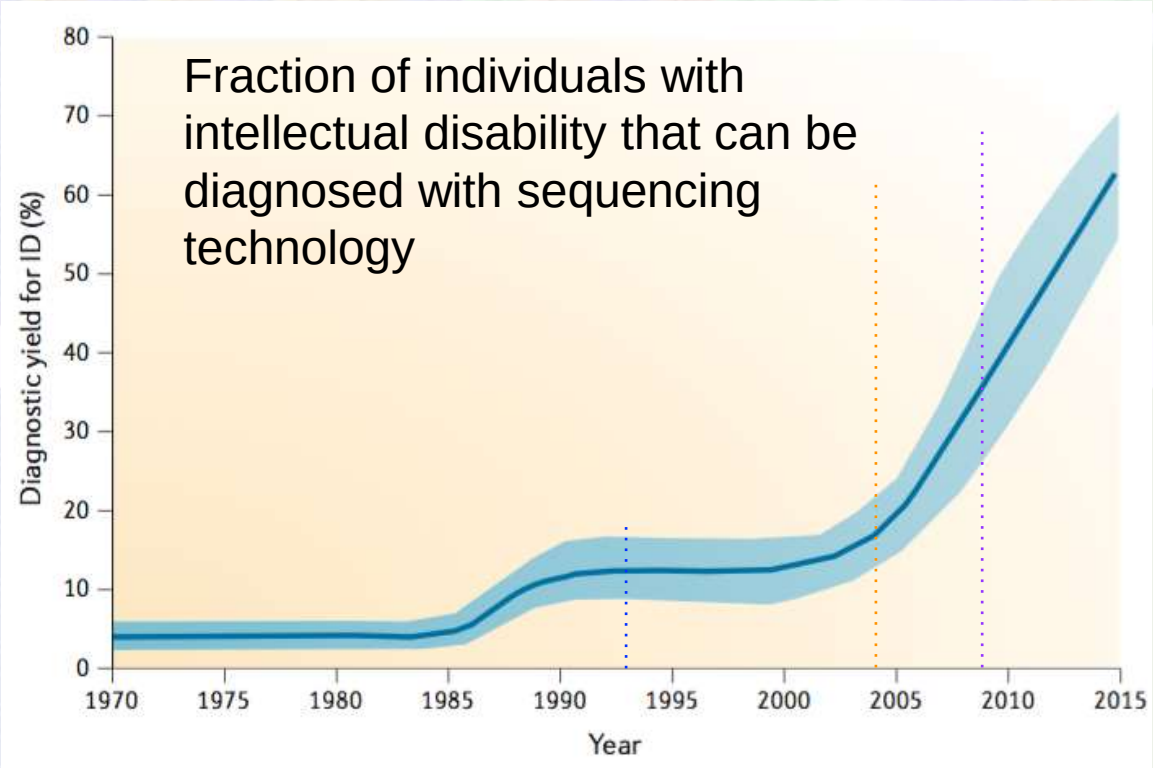


- Worldwide, country or population specific
  - Migration / changing ethnicity
- Which parts of the human genome can vary and which are constrained?
  - Different to evolutionary (across species) constraint
  - More genomes = better estimates
  - Where to look for severe disease causing variants

https://macarthurlab.org/2017/02/27/the-genome-aggregation-database-gnomad/
http://gnomad.broadinstitute.org/

Identification of causes of severe inherited neurodevelopmental diseases



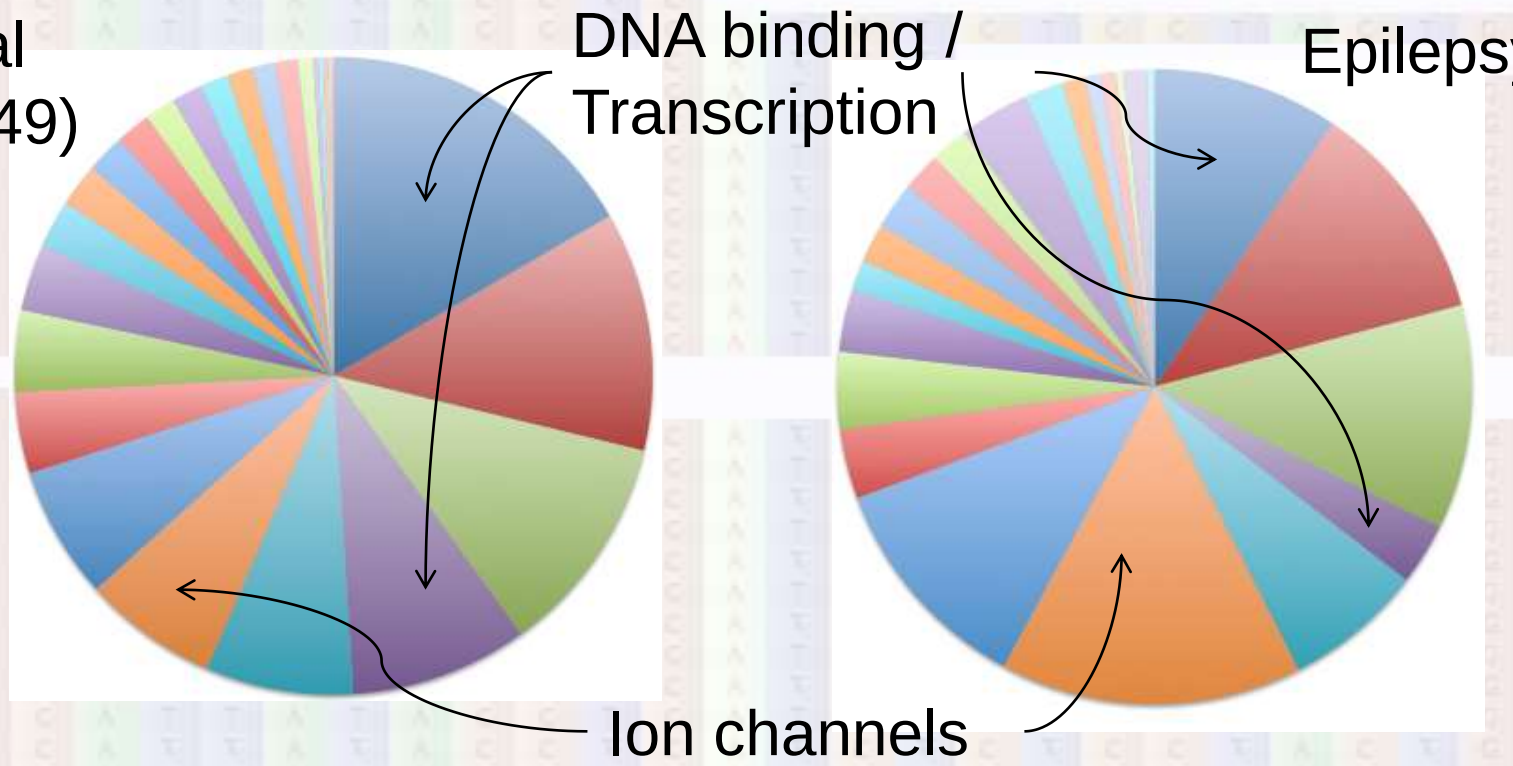


- These variants are disease drivers not risk factors but carriers don't necessarily have the same severity of symptoms

MPS: Massively parallel sequencing

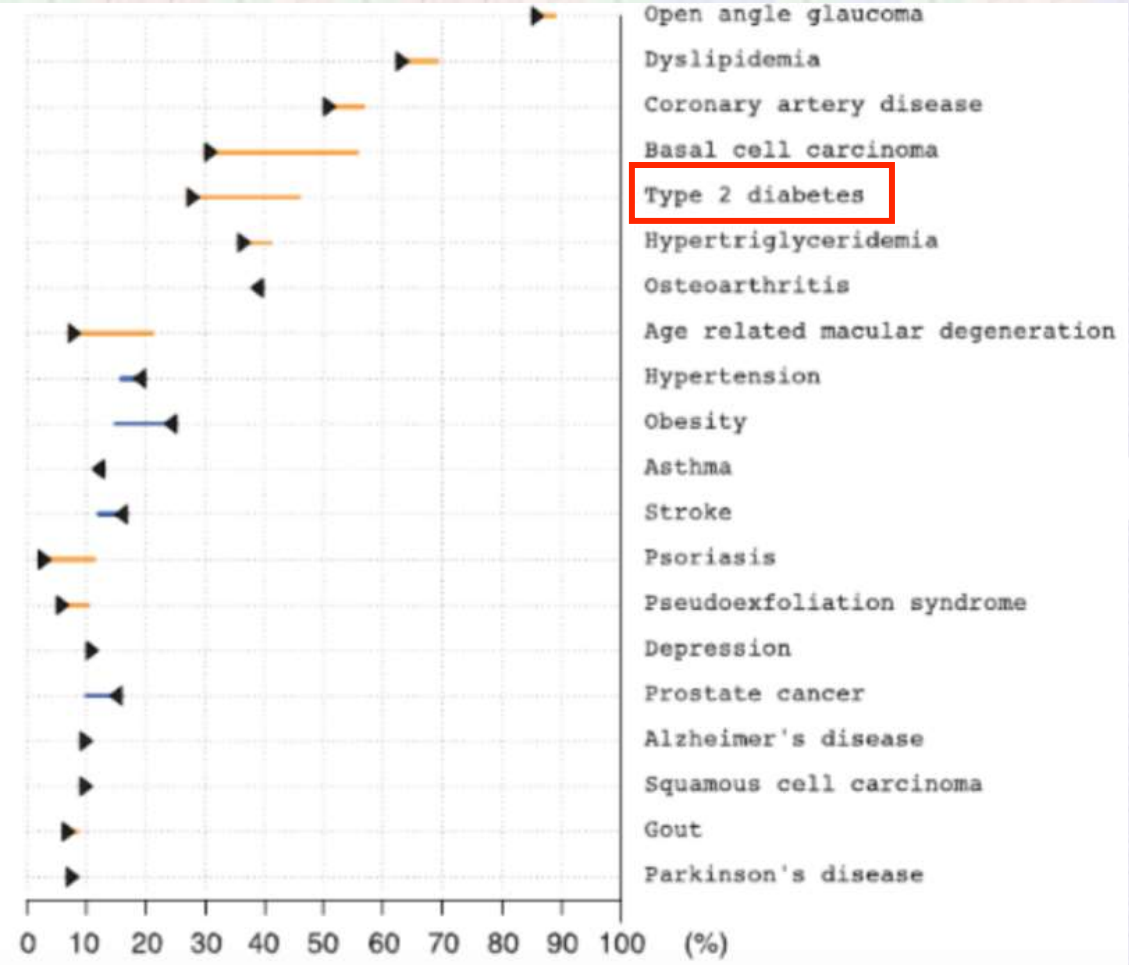
Vissers, L.E.L.M., et al., (2016). Nat. Rev. Genet., 17(1), pp

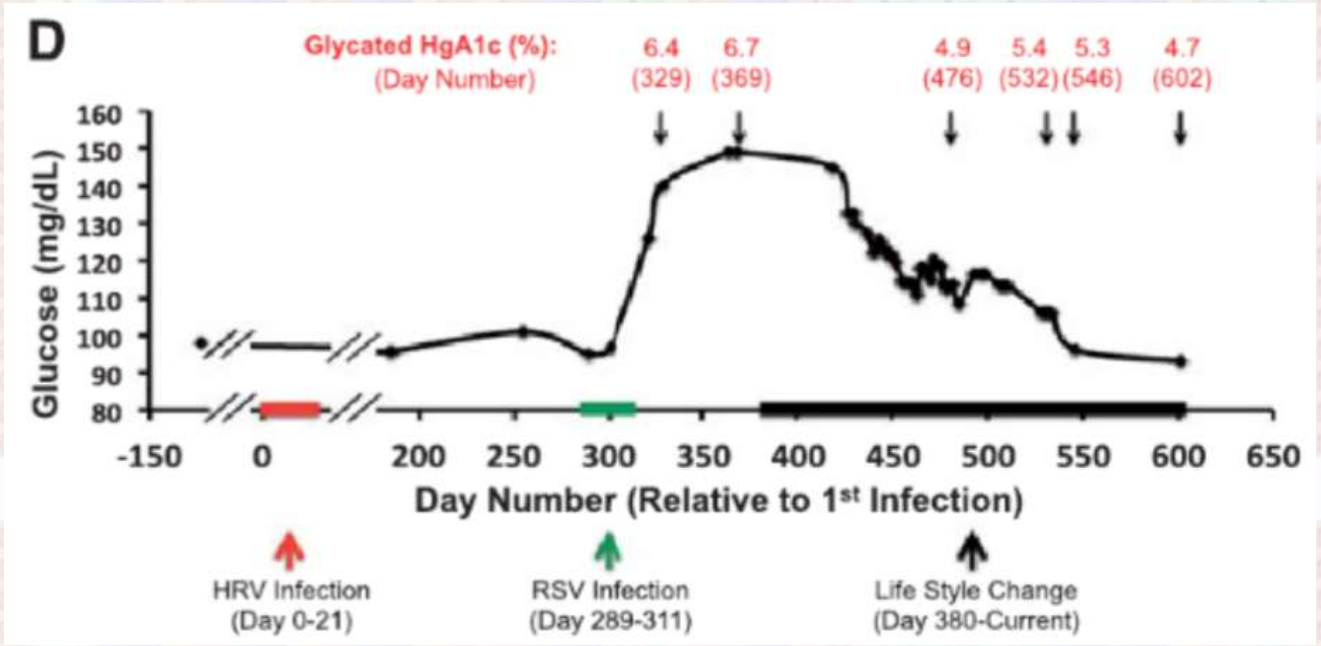
Shared and unique genetic aetiology of neurodevelopmental disorders



- **153** genes shared between both disorders
- Clinical and genetic heterogeneity
  - There is more than one molecular pathway involved in epilepsy or ID
  - No simple or unifying drug that will cure all

Real-time personal “omics” monitoring





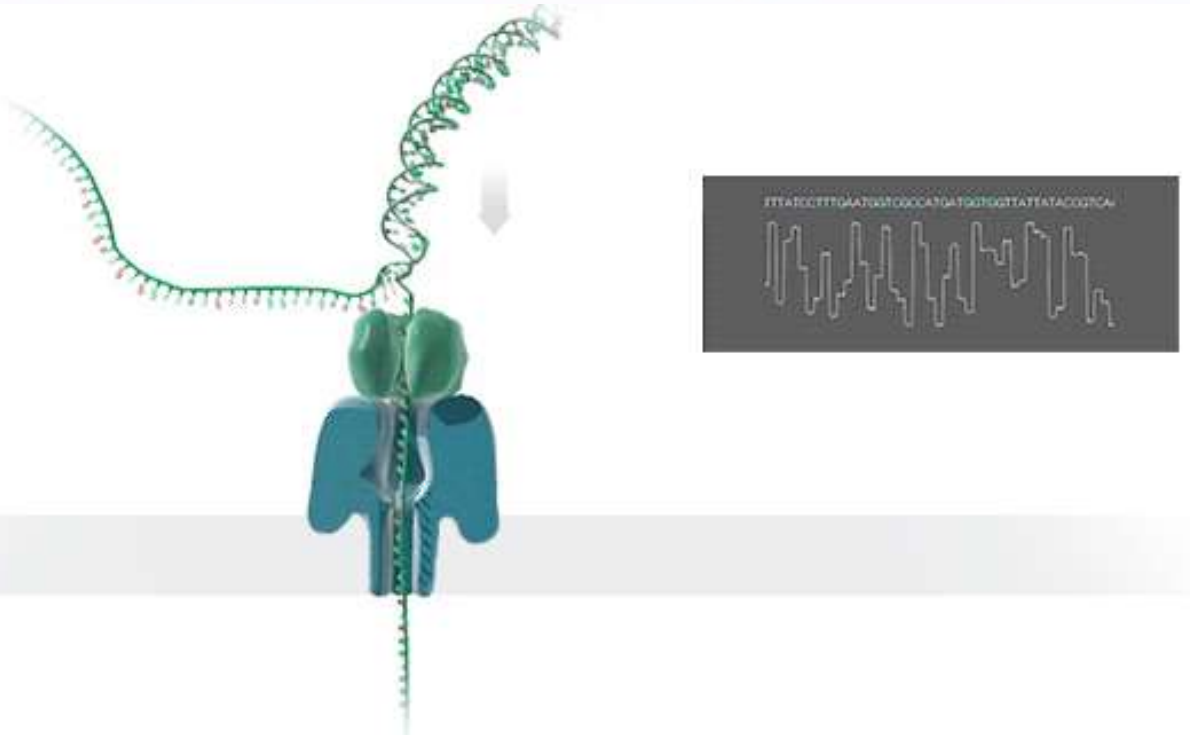
- Genome sequencing predicts elevated risk for type II (adult onset) diabetes
- Subsequent blood glucose monitoring indicates T2D onset triggered by an environmental event (infection)
- Diet change restores blood glucose level

Chen, R et al. 2012, Cell 148; 1293-1307

Real-time genome monitoring

Portable hardware

- Currently requires laboratory equipment for sample preparation.
- Provides DNA and RNA sequences and some epigenetic signatures.
- Detect pathogens.



Future of health care

Pinnacle of healthcare today

Pinnacle of healthcare in the future

For discussion

Drivers

- Health budget
- Personalised medicine
- Empowerment

Societal Trends

- Social media / sharing
- Socioeconomic gap
- Behavioural risk factors

Barriers

- Training for bioinformatics
- Storage costs
- Data security / misuse
- Genetic discrimination

Knowledge Gaps

- Response of genome variation in all conditions
- Catalogue of gene functions
- Data compression
- Embedded hardware