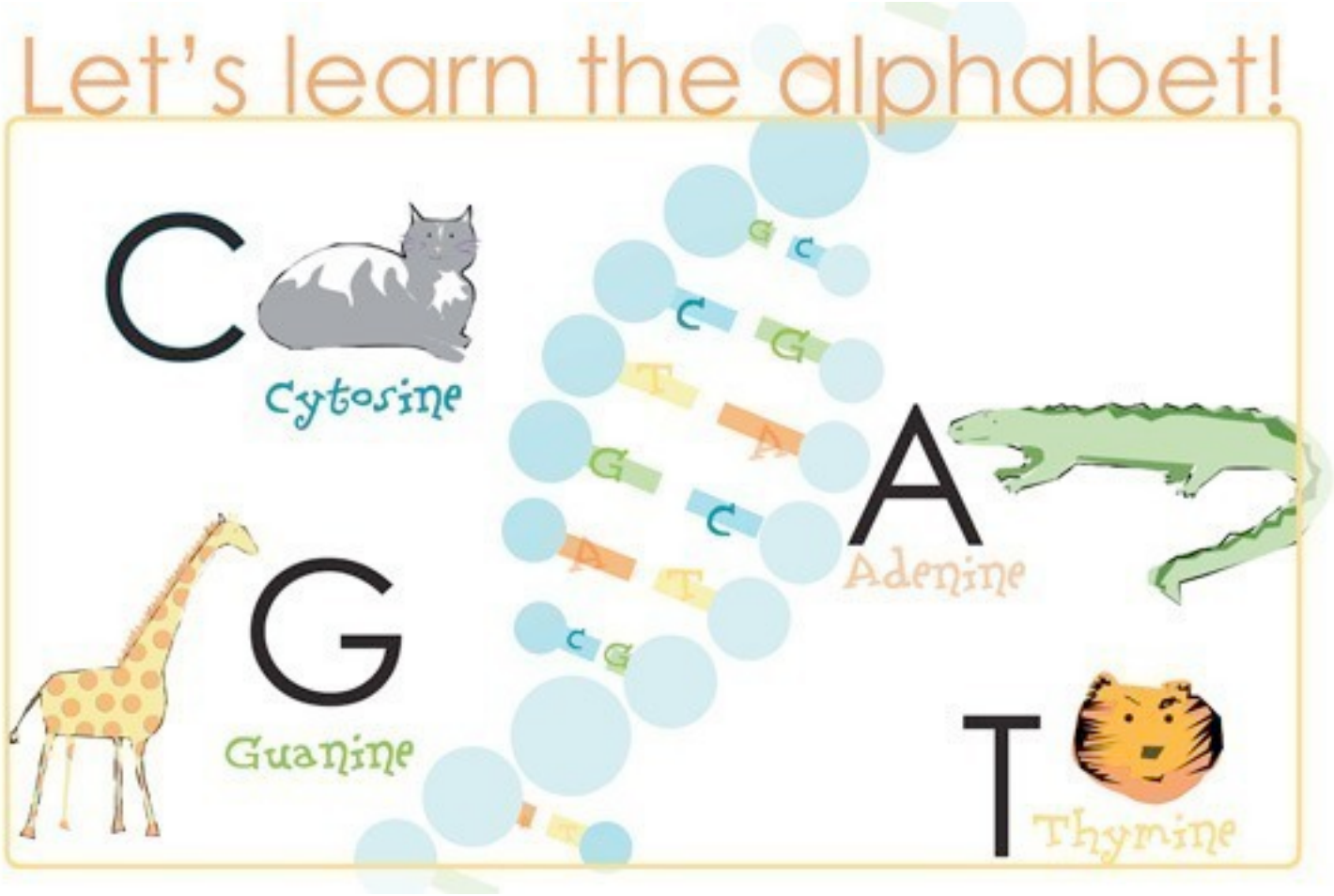
Applied Pediatric Genome Analysis

05 March 2018

Wyeth W. Wasserman
University of British Columbia
BC Children's Hospital

THE BASICS



http://www.nerdybaby.net/dna-for-babies/

BC Children's Hospital
An agency of the Provincial Health Services Authority

OVERVIEW

- Technology shifting in clinical genome analysis
- Introduction via case studies
- Almost no statistics

BC Children's Hospital
An agency of the Provincial Health Services Authority

OLD TECHNOLOGY FOR GENOME ANALYSIS: KARYOTYPING

HOW DOES YOUR GENOME DIFFER FROM THE REFERENCE GENOME

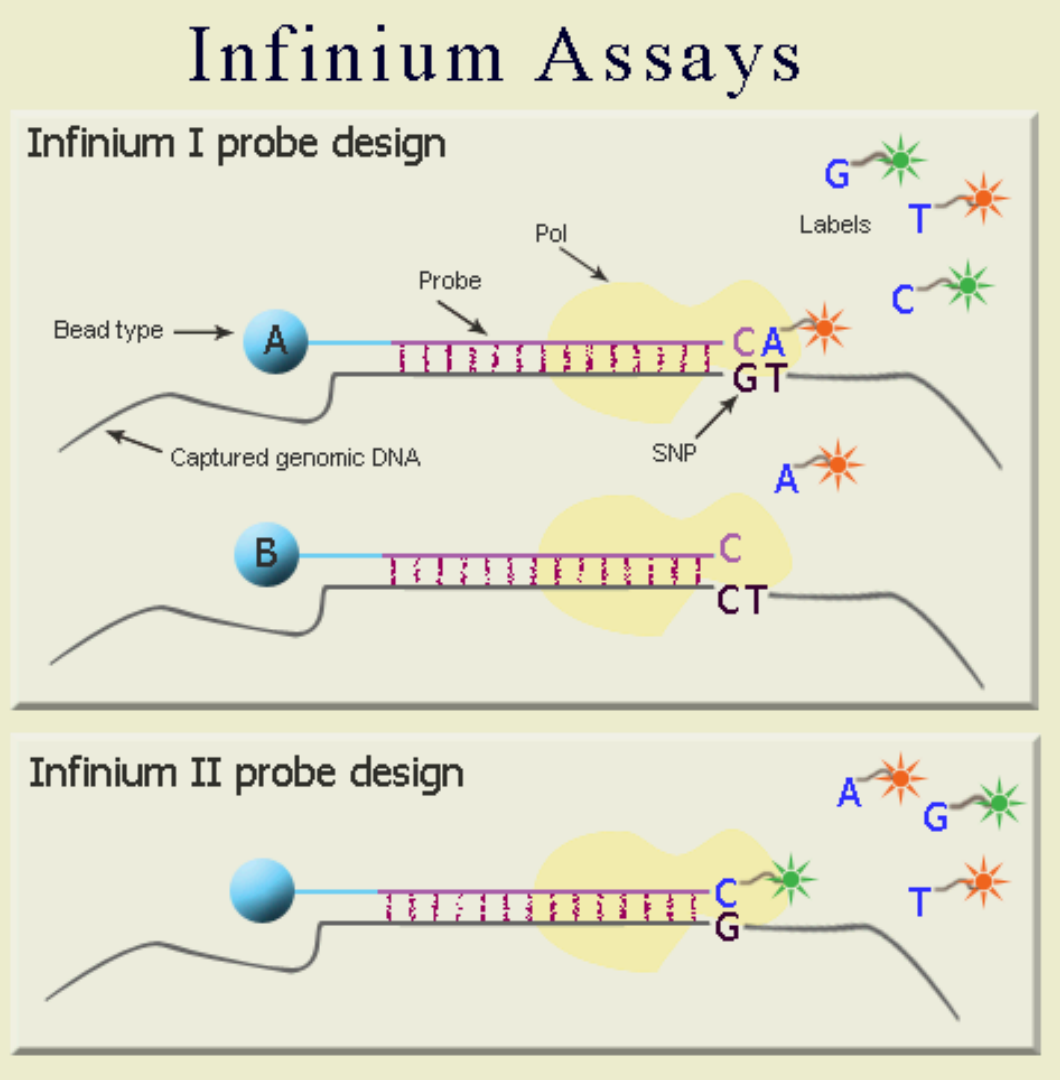
- structural variants (impacting ~20Mbp)
  - ~1,000 large deletions
  - ~160 copy-number variants
  - ~1100 “repeat” insertions (Alu, L1, etc)
  - ~4 NUMTs (mitochondria genome)
  - ~10 inversions
- Base-level
  - ~3.5M (~4.5M for AFR) single nucleotide variants
    - Only ~60,000 (~150k) extremely rare (<0.5%)
  - ~0.5M small indels

https://www.nature.com/articles/nature15393

SNPs vs SNVs

- Single Nucleotide Polymorphism vs Single Nucleotide Variant
- All SNPs are SNVs, but not all SNVs are SNPs
- SNP refers to a polymorphic variant – a variant that occurs over a specified percentage in a specified population (usually 1% at present)
- Do not say "SNP" unless you **mean** SNP

SNP Testing



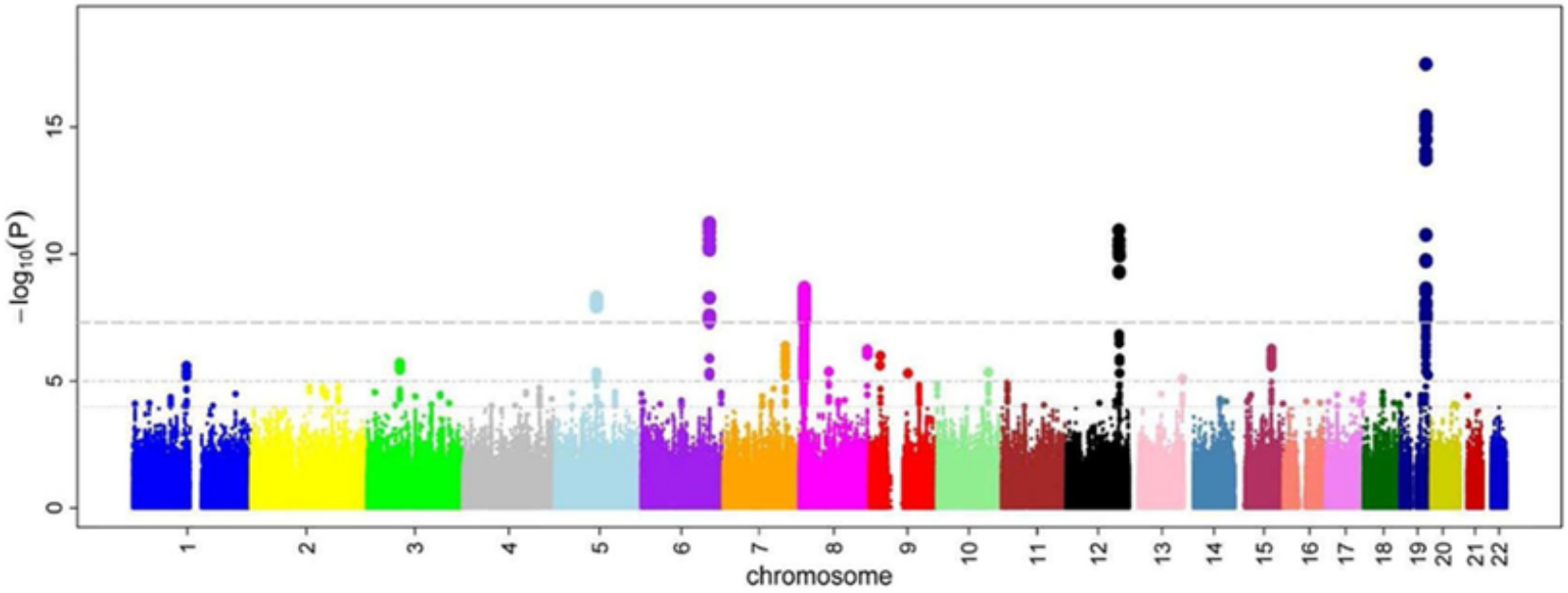
https://www.ncbi.nlm.nih.gov/probe/docs/distrillumina/

BC Children's Hospital
An agency of the Provincial Health Services Authority

Uses of SNP Genotyping

- Population Determination
- GWAS
- Linkage Analysis
- Clinical MicroArray for Copy Number changes

GWAS

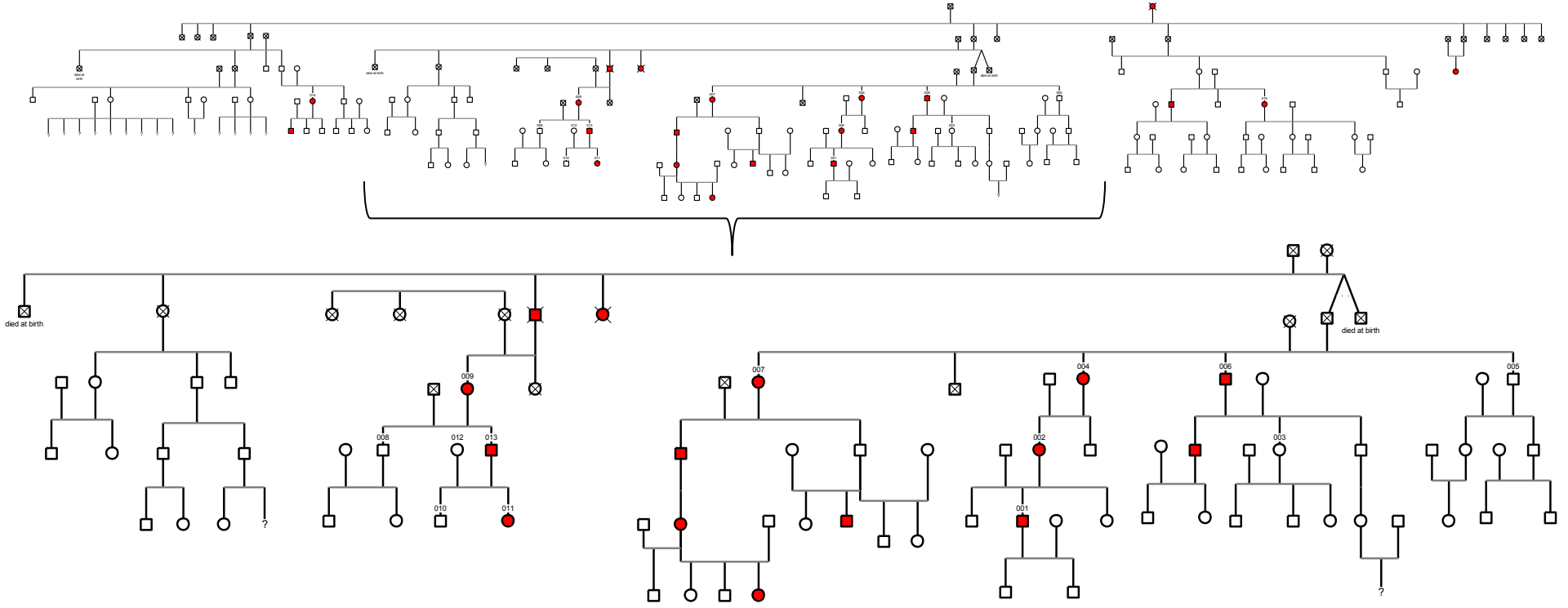


https://upload.wikimedia.org/wikipedia/commons/thumb/1/12/Manhattan_Plot.png/819px-Manhattan_Plot.png

BC Children's Hospital
An agency of the Provincial Health Services Authority

Pedigree-based Linkage Analysis

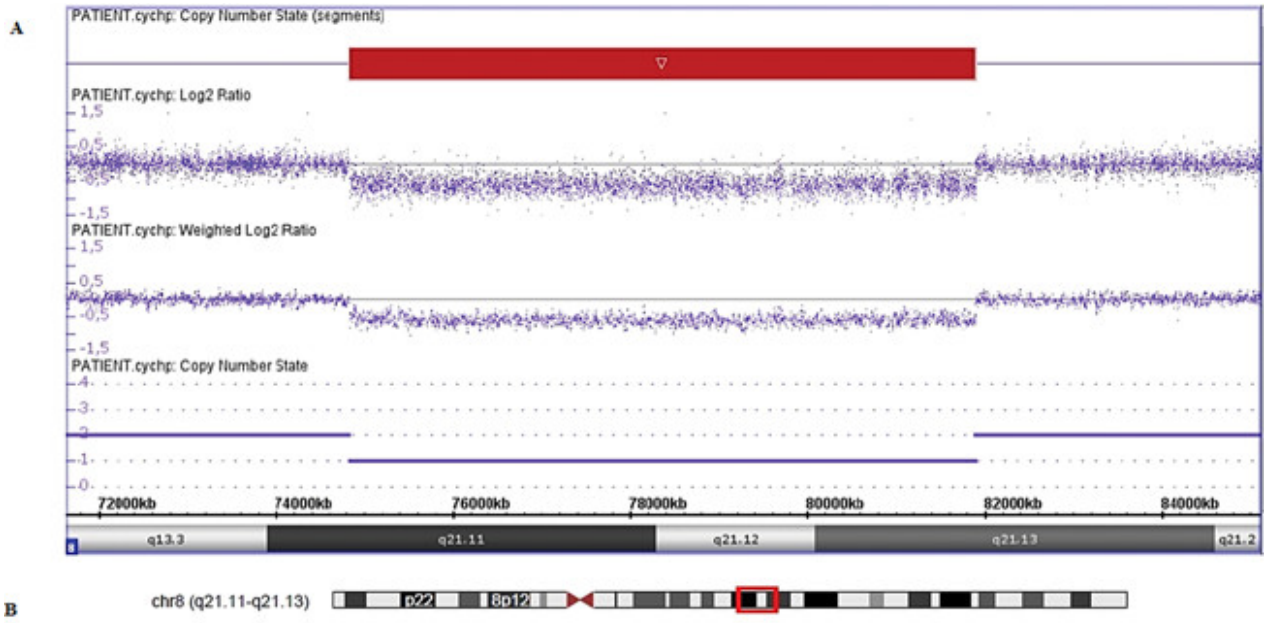
From Pedigree to Critical Region



EXAMPLE OF LOD SCORES

https://www.researchgate.net/figure/Genome-wide-linkage-analysis-of-the-pedigree-using-affected-members-only-with-10K-A_fig4_38055346

Chromosomal MicroArrays



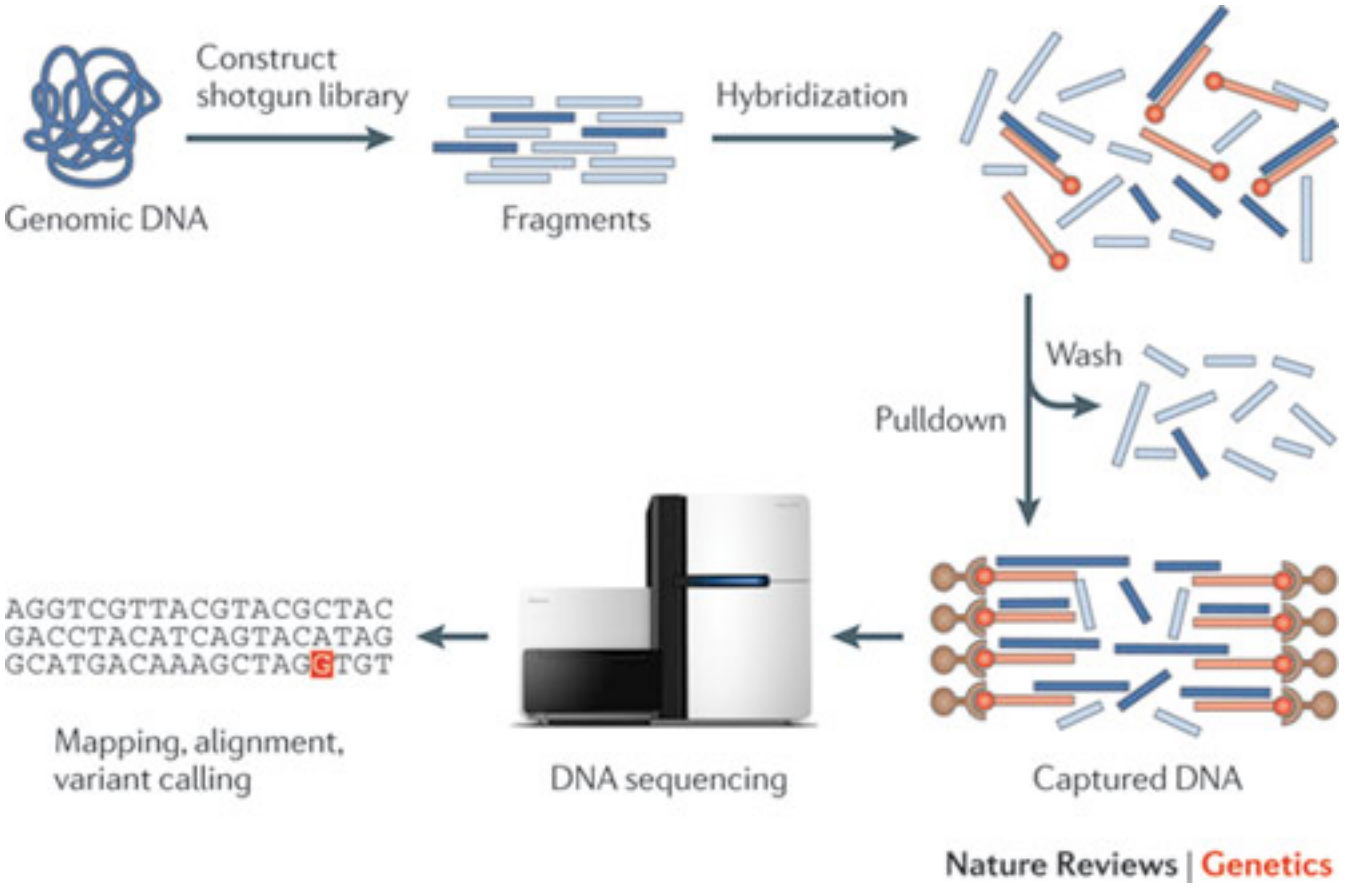
Chromosomal Microarrays are the clinical standard for detecting copy number variants. They have reliable sensitivity down to ~100,000bp (and pretty good down to ~20,000).

https://www.researchgate.net/figure/Microarray-based-copy-number-analysis-performed-with-the-Affymetrix-Cytogenetics_fig2_275363431

EXOMES

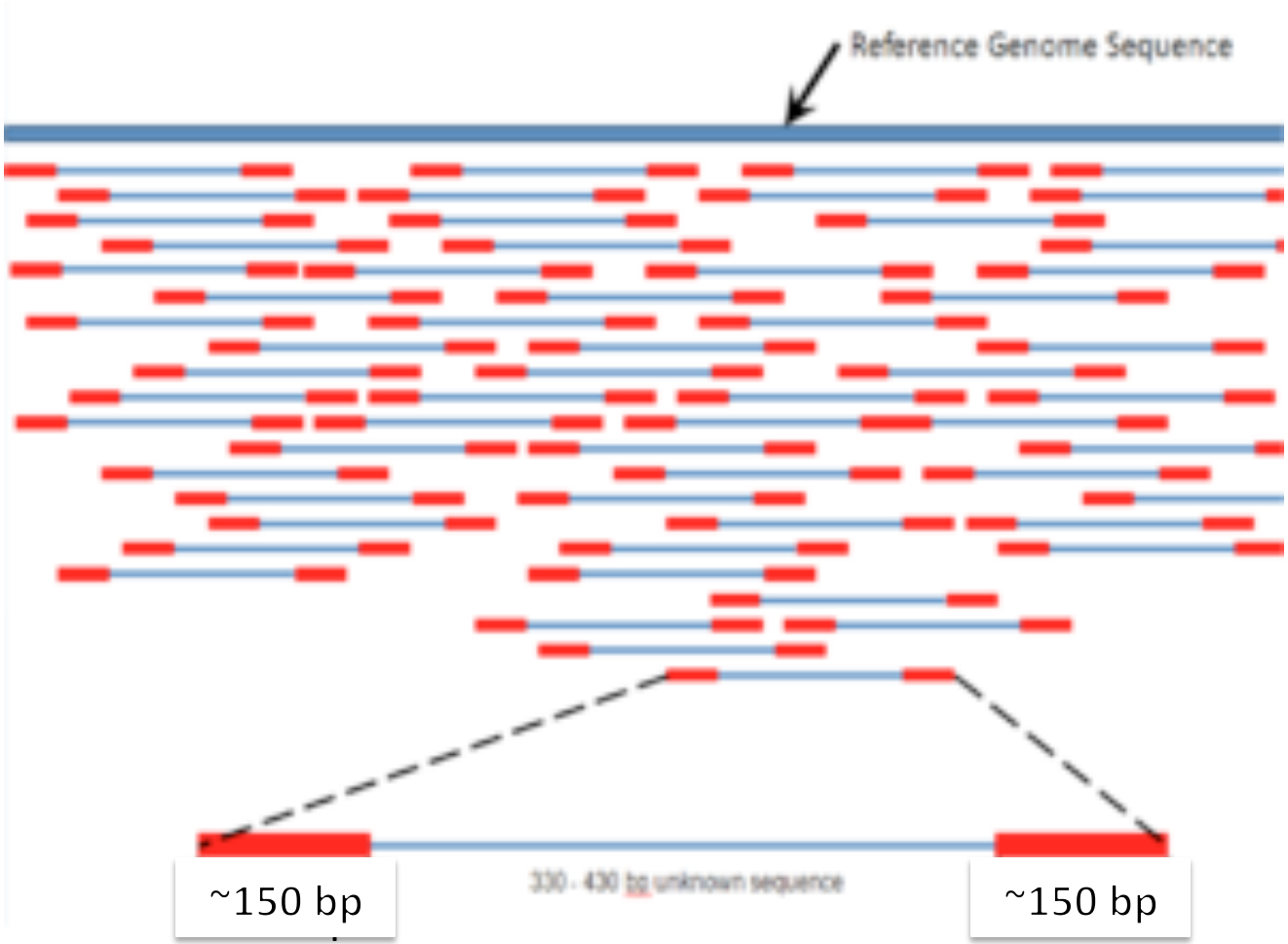
- Until recently the cost of WGS was prohibitive and most individual patients were analyzed for a restricted portion of the genome – the exons.
- "Exomes" recover ~5%
- Different companies offer distinct "kits" which contain oligonucleotides designed to hybridize to the regions of the genome specified by the kit designers

EXOME TECHNOLOGY



Michael J. Bamshad et al (2011) Exome sequencing as a tool for Mendelian disease gene discovery. Nature Reviews Genetics 12, 745-755 doi:10.1038/nrg3031

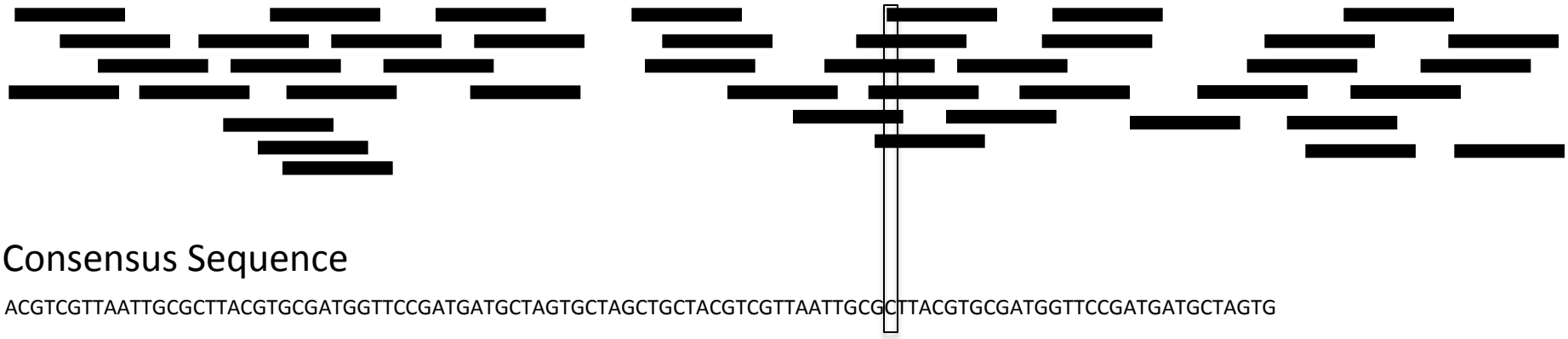
Next Generation Sequencing and Read Alignment



http://en.wikipedia.org/wiki/Genomics

Genotype Calling

Short reads aligned



Consensus Sequence

ACGTCGTTAATTGCGCTTACGTGCGATGGTTCCGATGATGCTAGTGCTAGCTGCTACGTCGTTAATTGCGCTTACGTGCGATGGTTCCGATGATGCTAGTG

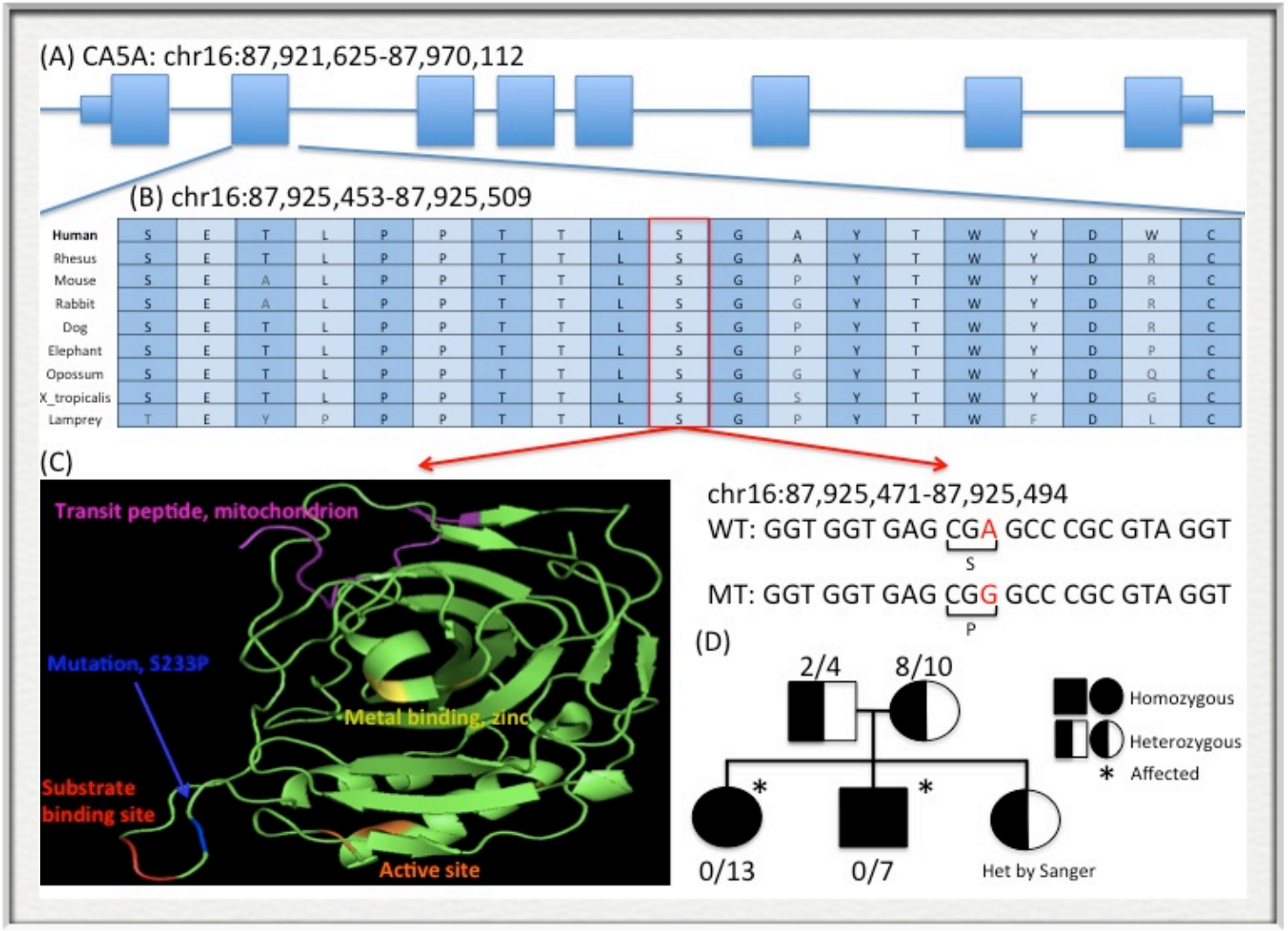
2 A / 4 G = A/G genotype

Most research grade sequencing studies produce an average of 30-100 reads per position with enormous variance across positions

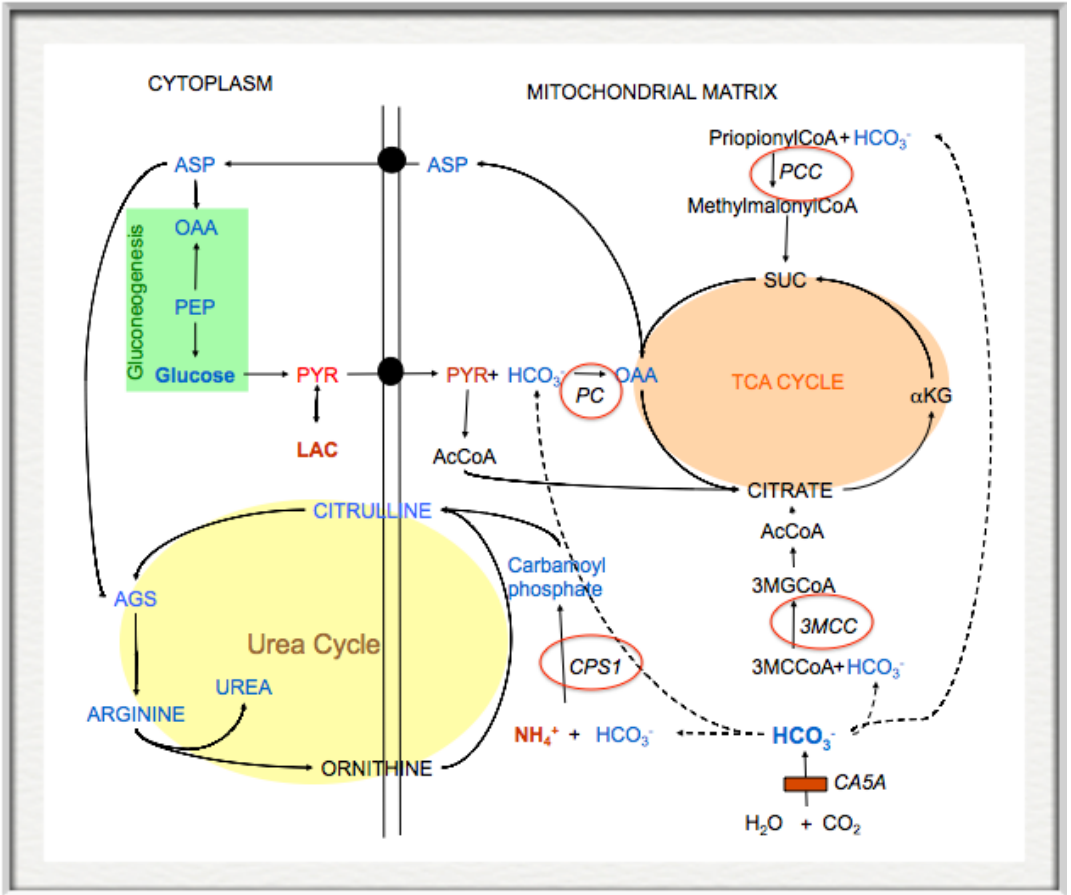
Whole Exome Sequencing EXAMPLE CASE

- 2 affected siblings & unaffected parents
- Agilent SureSelect kit & Illumina HiSeq 2000 (Perkin-Elmer, USA)
- 5.2 billion 100bp pair-end reads
- coverage per base 32X
- Bowtie, BWA and GSNAP: to map reads to hg19 reference genome
- 99% classified as common variants
- 7 candidate genes
  (4 homozygous & 3 compound heterozygous mutations)
- only 1 variant predicted disruptive to protein function
  (Sift and PolyPhen2)

CA5A gene: homozygous S233P mutation



CA-VA at the hub of life: producer & donor HCO_3^- to 4 enzymes



in vitro mutant showed:

decreased enzymatic activity

increased aggregation & thermal lability

WORKFLOW

Patient Selection

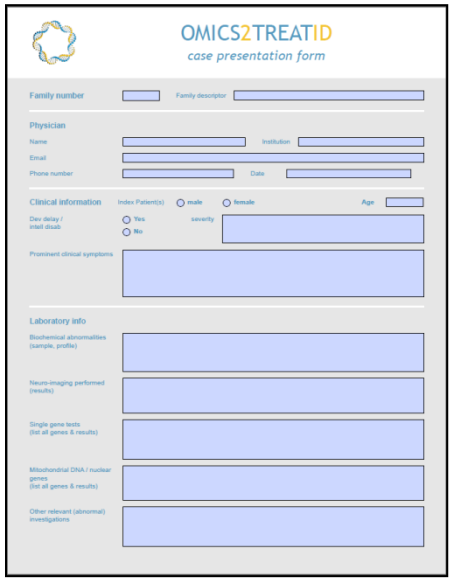
- ID or at risk for ID
- Biochemical phenotype
- oaCGH negative
- Known genes not mutated

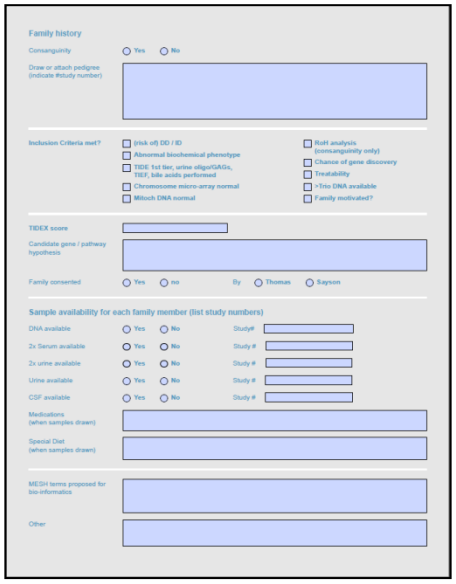
Patient Consent

- TIDEX de-identification study numbers

Sample collection

- Blood/saliva/urine





Sequencing

- DNA prep; WES + WGS

Bioinformatics analysis

- Alignment
- Variant calls
- Variant filtering
- **Variant-phenotype link interpretation**

Candidate Lists to clinicians

Sanger

- Sanger re-sequencing

Experimental

- Experimental, fibroblasts, model organisms, etc.

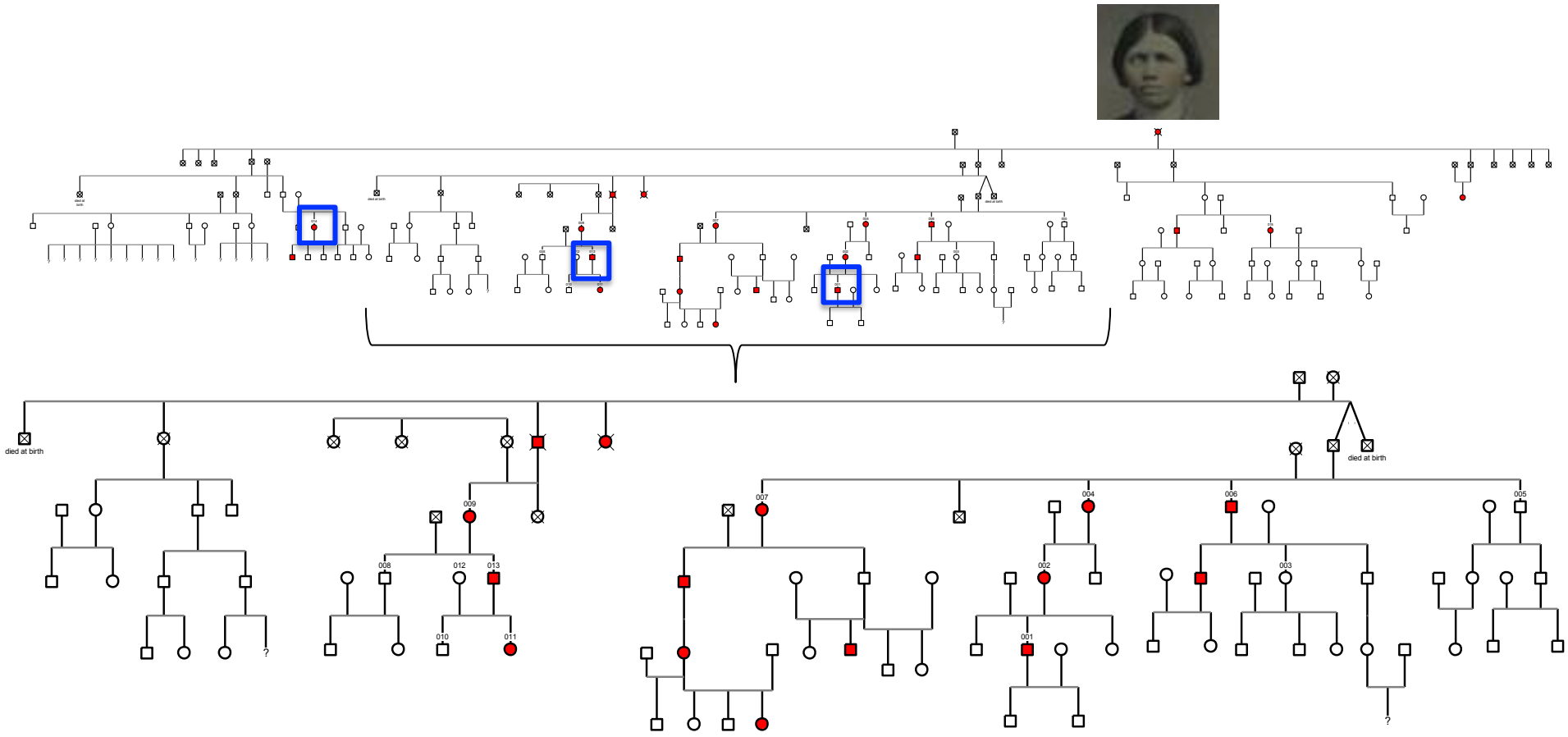
BC Children's Hospital
An agency of the Provincial Health Services Authority

OVERVIEW

WES vs WGS

- Cost no longer prohibitive
- The FEDEX moment is approaching
- WGS has many advantages
  - Copy number calling
  - Structural changes
  - No capture bias
  - Splice altering events in introns
  - Distal regulatory regions

From Critical Region to Mutations



WGS: Reveals shared ~8MB region with no protein coding alterations

BC Children's Hospital
An agency of the Provincial Health Services Authority

Strabismus Outcome

- 5 Mbp critical region
- 30 rare variants
- No protein altering variants
- Two genes nearby with potential causal roles
- Ongoing

BC Children's Hospital
An agency of the Provincial Health Services Authority

OPEN PROBLEMS

- Reliable CNV equivalent to CMA
- Detection of balanced translocations
- Phenotype-to-genotype relationship detection
- Accounting for patient population in the analysis
- Phasing

CLOSING

- WGS has arrived
- Cost is decreasing
- Patients increasingly likely to be diagnosed for simple genetic disorders

Thank You