MEDI6215

UNIVERSITY OF Southampton

Copy Number Variation

Bioinformatics, Interpretation, and Data Quality Assurance in Genome Analysis

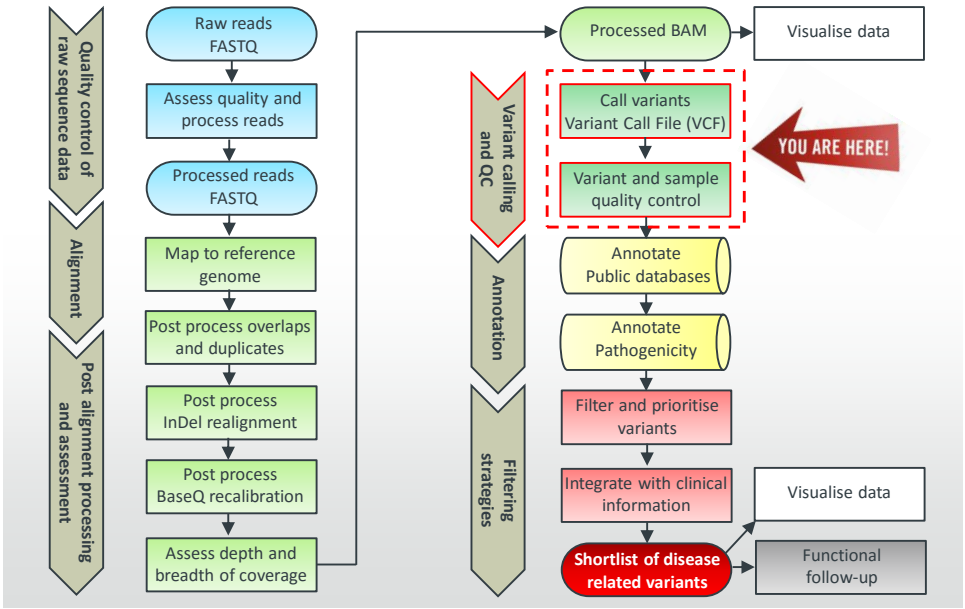
Faisal I. Rezwan

27th February 2017

Lecture outline

- What is Copy number variation (CNV) and structural variation (SV)
- Effect of CNV/SV in diseases and phenotypes
- Detection of CNV/SV
- Strategies for detecting CNV/SV using NGS data
- Tools for detecting CNV/SV

Analysis workflow



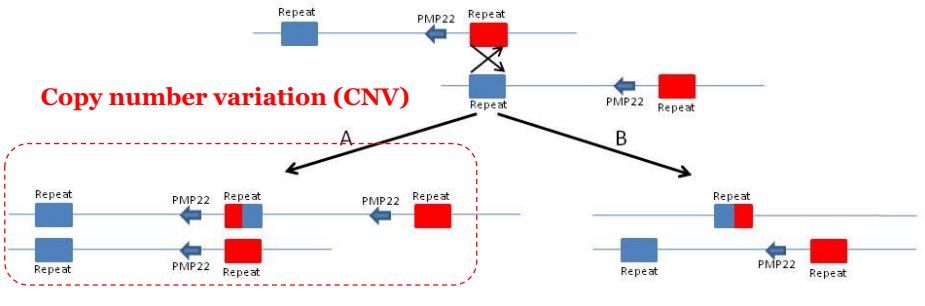
Charcot-Marie-Tooth disease type 1A (CMT1A)

A type of inherited neurological disorder that affects the peripheral nerves.

Cell, Vol. 66, 219–232, July 26, 1991, Copyright © 1991 by Cell Press

DNA Duplication Associated with Charcot–Marie–Tooth Disease Type 1A

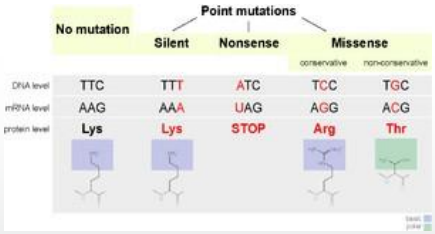
James R. Lupski,*†‖ Roberto Montes de Oca-Luna,* Susan Slaugenhaupt,‡ Liu Pentao,* Vito Guzzetta,* Barbara J. Trask,§ Odila Saucedo-Cardenas,* David F. Barker,¡ James M. Killian,# Carlos A. García,** Aravinda Chakravarti,‡ and Pragna I. Patel**



CNV vs. SNV

University of Southampton

Single-nucleotide variants



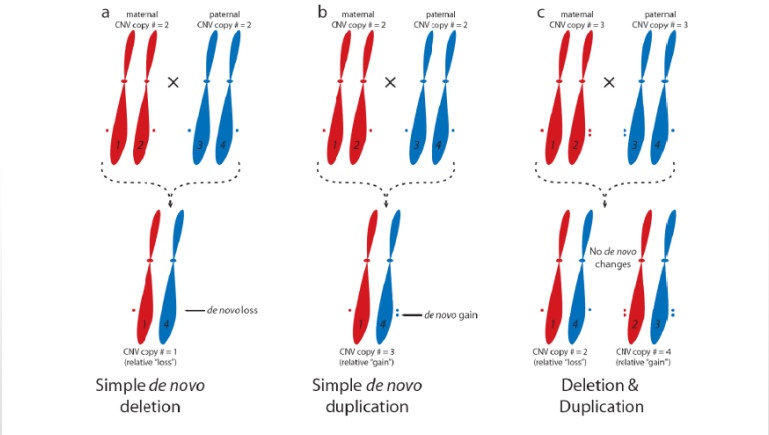
Copy number variants

Copy number loss and gain

University of Southampton

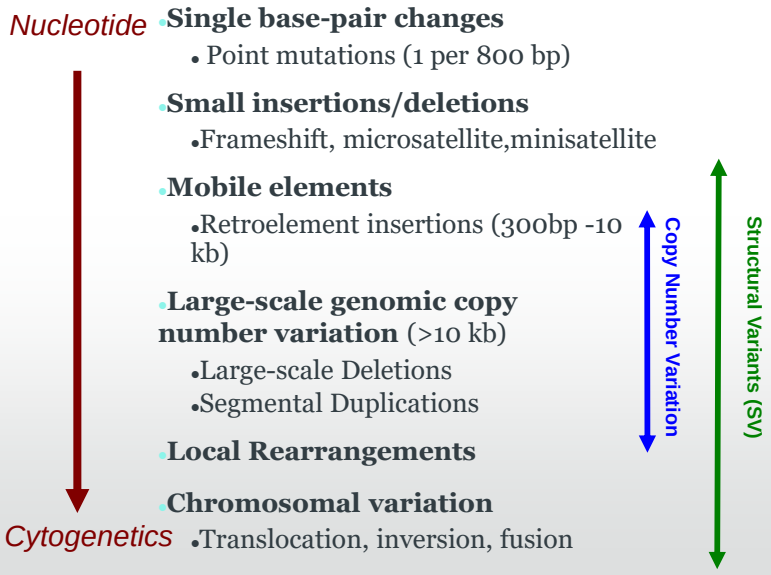
CNVs are losses or gains of genetic material

Redon et al (2006) Supplemental



Redon et al 2006. Global variation in copy number in the human genome. *Nature* 444, 444-454 (23 November 2006)

Continuum of genomic variation

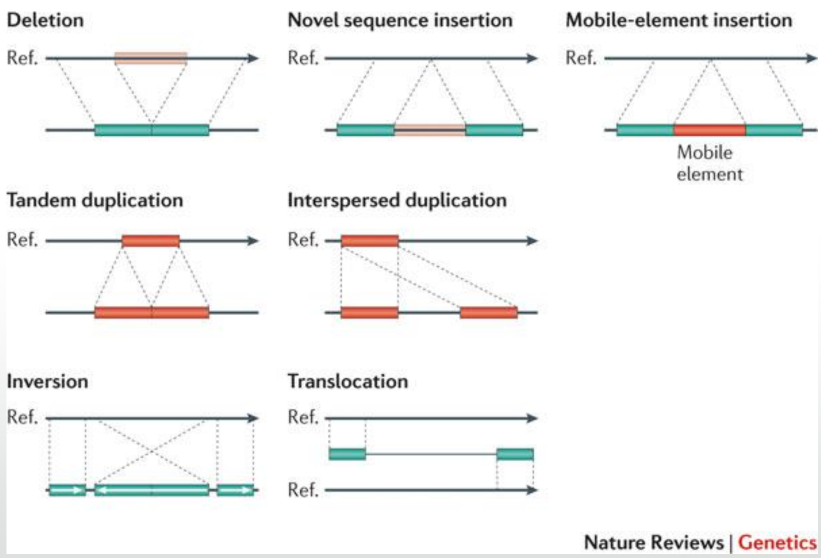


SV Definition

Structural variations (SV) are Genomic rearrangements that effect more than 1 Kb

- Indel:
  - Short insertion or deletion events < 50 bp
- Structural variations:
  - Large insertion
  - TE insertion
  - Inversion
- Copy number variations
  - Large deletion
  - Interspersed duplication
  - Tandem duplication

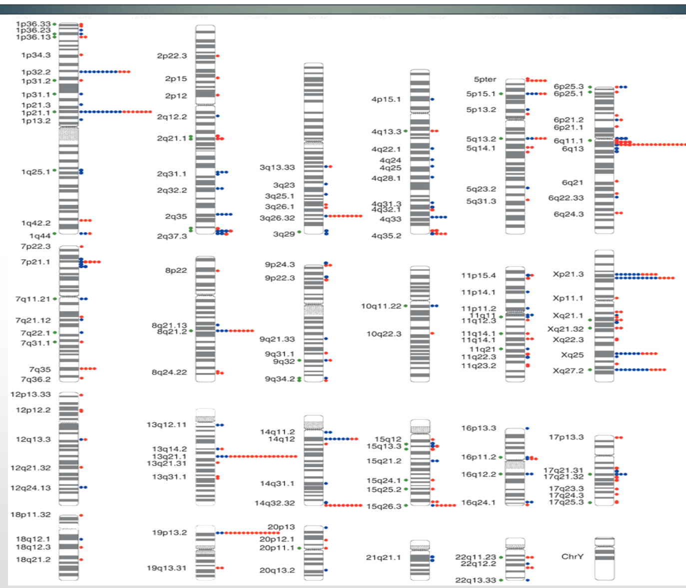
SV Description



Alkan et al. 2011. Genome structural variation discovery and genotyping. *Nature Reviews Genetics* 12, 363-376

Distribution in genome

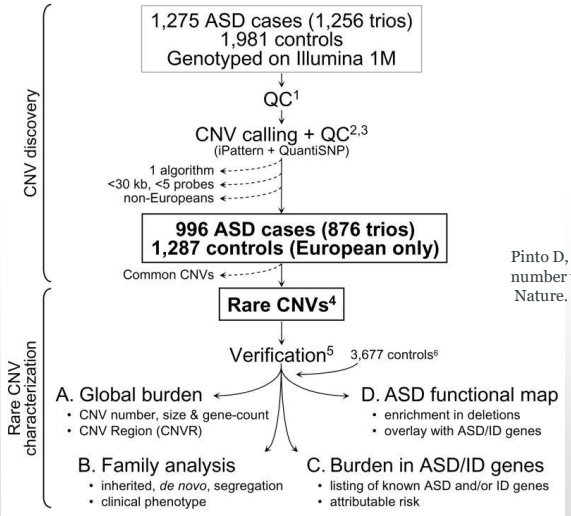
UNIVERSITY OF Southampton



Iafrate et al (2004) Detection of large-scale variation in the human genome. *Nature Genetics* 36, 949-951

Korbel JO et al (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science 318:420–6.

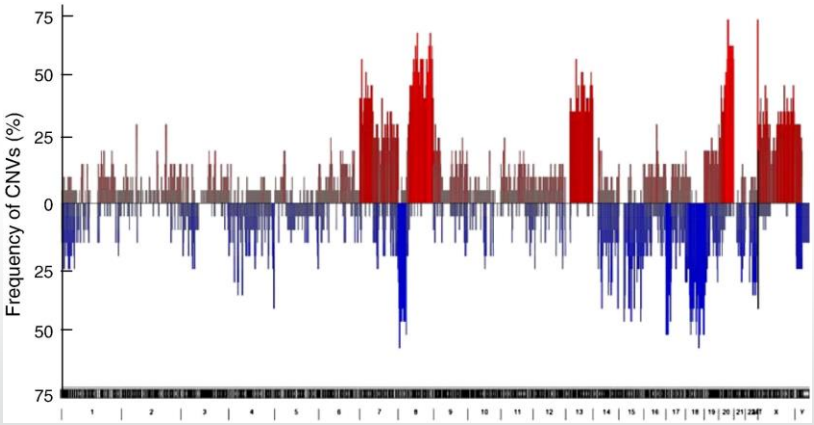
CNV and autism



Pinto D, et al (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010 Jul 15;466(7304):368-72.

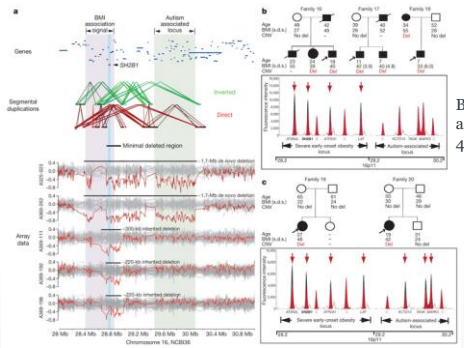
CNV and cancer

Frequency of tumour specific CNVs can explain HNPCC pathogenesis

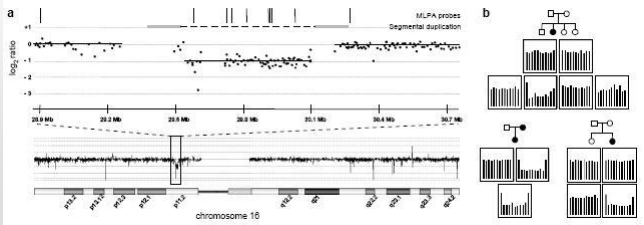


Chen et al. (2013). Identification of chromosomal copy number variations and novel candidate loci in hereditary nonpolyposis colorectal cancer with mismatch repair proficiency. Genomics 102 (2013) 27–34

CNV and obesity

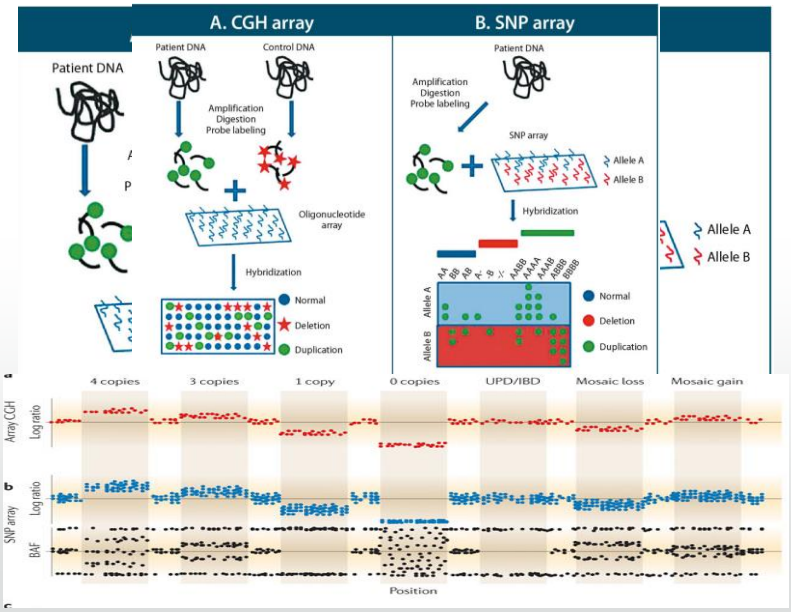


Bochukova EG, et al (2010). Large, rare chromosomal deletions associated with severe early-onset obesity. Nature. 4;463(7281):666-70

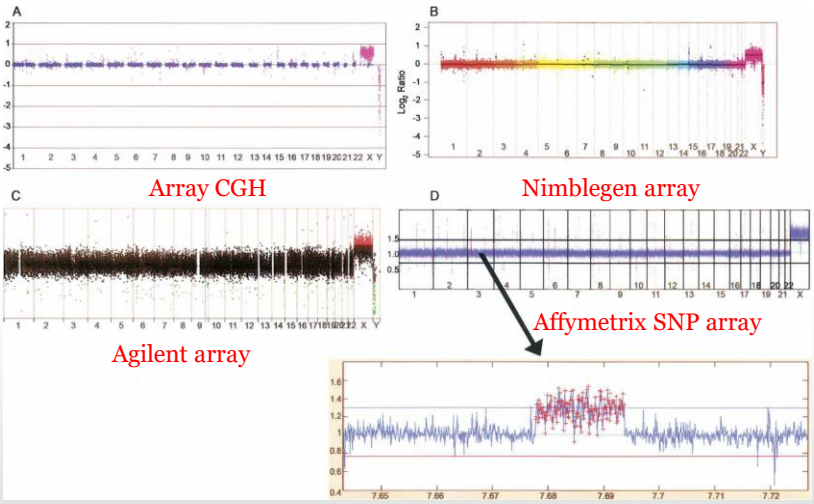


Walters RG et al (2010). A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature.4;463(7281):671-5.

Detection of CNVs

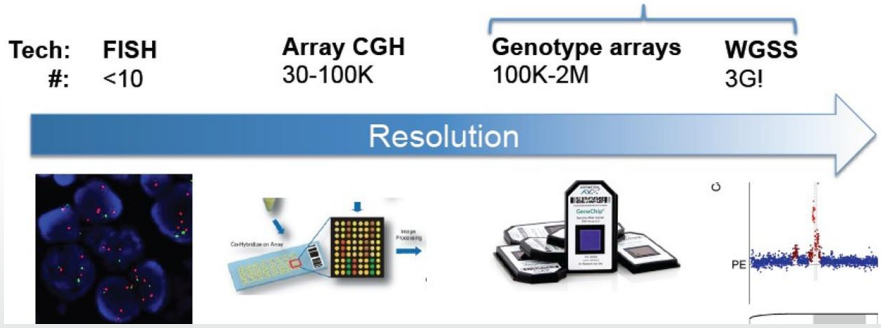


Detection of CNVs

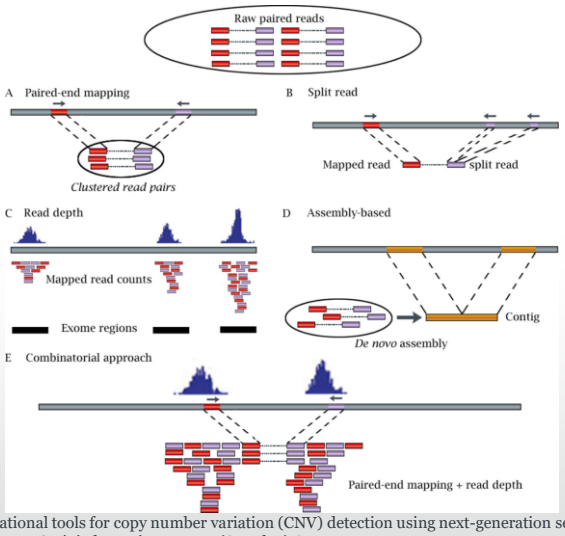


Freeman et al 2006. Copy number variation: New insights in genome diversity. Genome Res. 16(8):949-61.

Evolution of CNV detection technologies

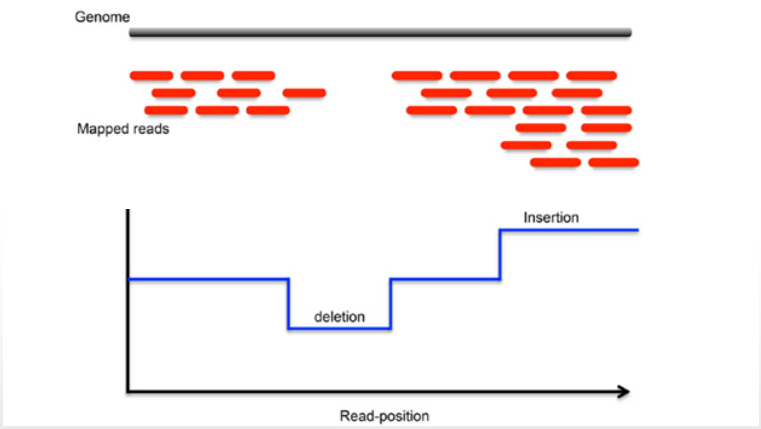


SV/CNV detection using NGS data



Zhao et al 2013. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 2013, 14(Suppl 11):S1

Read depth

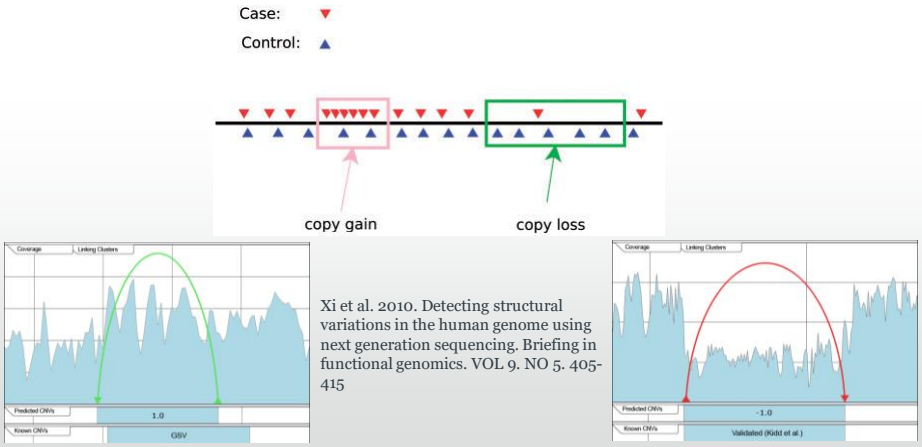


- Mapping Issue
- Poor "sequencability"
- Deletion/duplication

Valsesia et al. 2013. The growing importance of CNVs: new insights for detection and clinical interpretation. Front Genet. 2013; 4: 92.

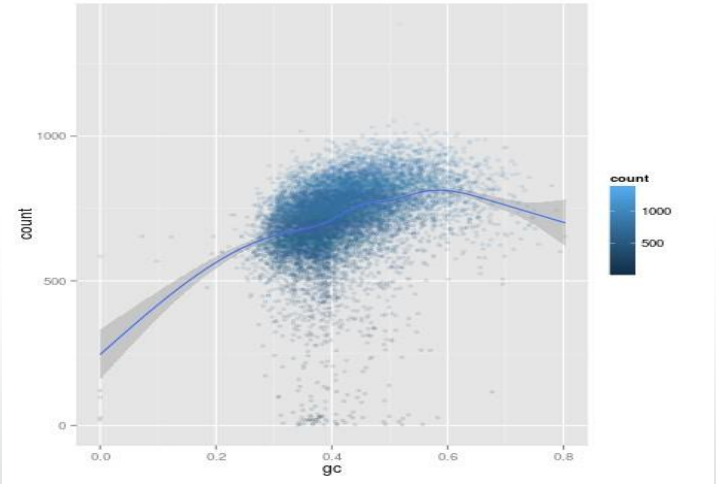
Read depth

An increase/decrease in depth of coverage between two known loci in the reference genome may give an indication of CNV.



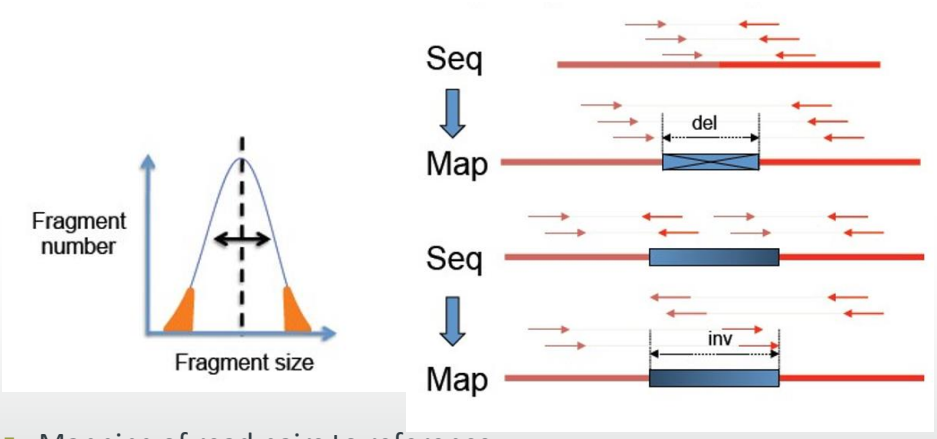
Xi et al. 2010. Detecting structural variations in the human genome using next generation sequencing. Briefing in functional genomics. VOL 9. NO 5. 405-415

Read depth biased by GC content



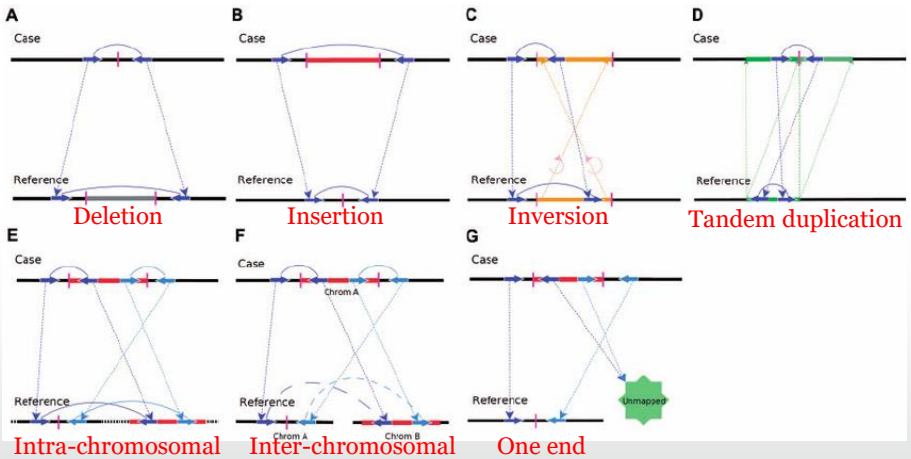
HG00155, 1000 genomes
Peter N. Robinson, Institut fu¨r Medizinische Genetik und Humangenetik Charit´e Universit¨atsmedizin Berlin

Paired end mapping



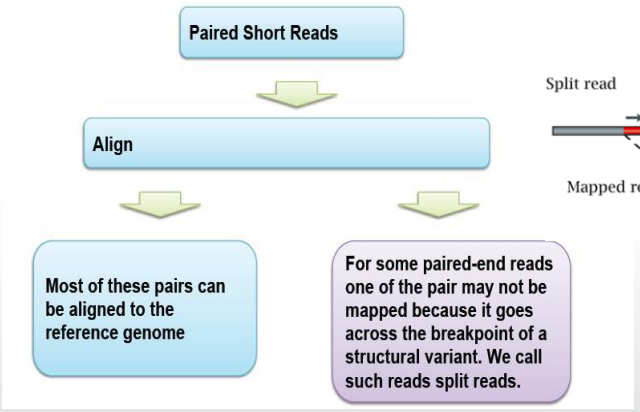
- Mapping of read pairs to reference genome
  - Spanning unexpected distance
  - Unexpected orientation

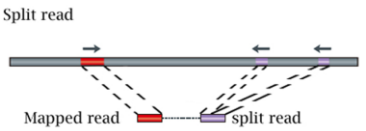
Paired end mapping



Xi et al. 2010. Detecting structural variations in the human genome using next generation sequencing. Briefing in functional genomics. VOL 9. NO 5. 405- 415

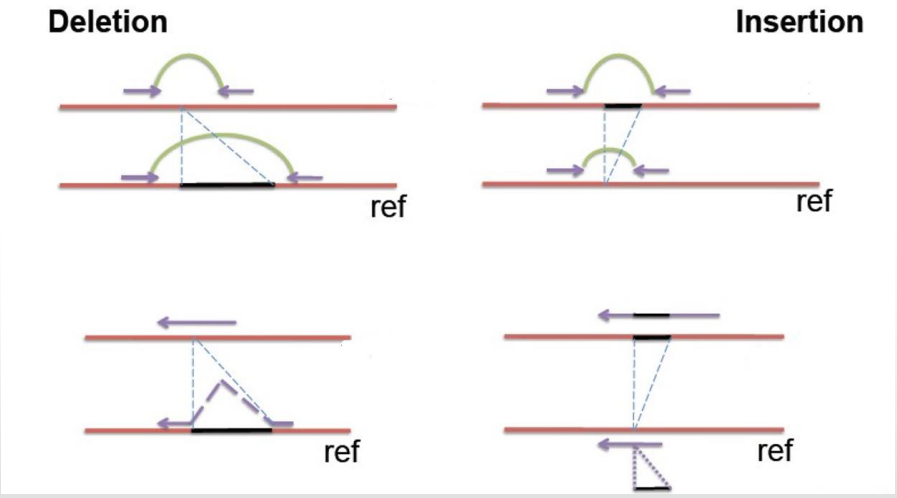
Split read mapping





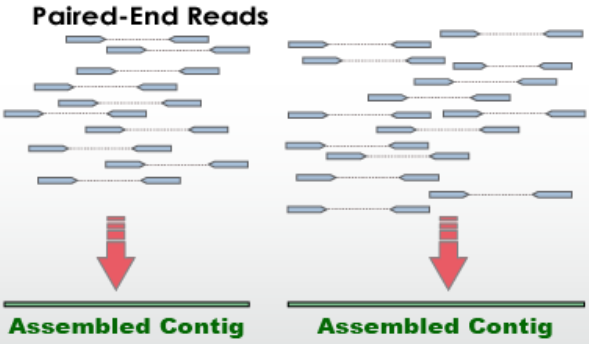
Source: Mathieu Bourgey, *Ph.D* & Louis Letourneau, *M.eng*. **CAGE-KID Cancer Genomics Workshop**

Split read mapping



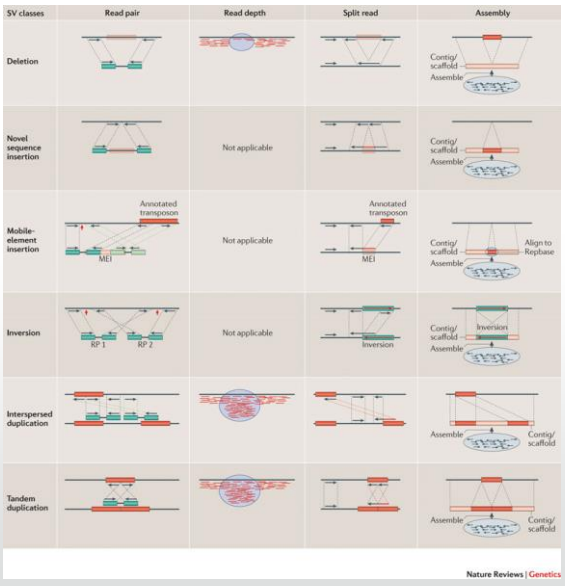
De novo assembly

- Local (de novo) assembly and then align assembled sequences to reference genomes



Issues: Repeat regions and low mappablity

SV detection in a nutshell



Alkan et al. 2011. Genome structural variation discovery and genotyping. *Nature Reviews Genetics* 12, 363-376

SV Detection tools

Table 1 Summary of paired-end mapping (PEM), split read (SR), and *de novo* assembly (AS)-based tools for CNV detection using NGS data

Method	URL	Language	Input	Comments
PEM-based				
BreakDancer				
PEMer				
VariationHunter				
commonLAW				
GASV				
Spanner				
SR-based				
AGE				
Pindel				
SLOPE				
SRiC				
AS-based				
Magnolya				
Cortex assembler				
TIGRA-SV				

*The specific inp
[b]File format fron
[c]The file includir

Table 2 Read depth (RD)-based tools for CNV detection using whole genome sequencing data

Tool	URL	Language	Input	Comments
SegSeq[a]				
CNV-seq[a]				
RDXplorer[b]				
BIC-seq[a]				
CNAseq[a]				
cnMOPS[b]				
JointSLM[b]				
ReadDepth				
rSW-seq[a]				
CNVnator				
CNVnorm[a]				
CMDS[b]				
mrCaNaVar				
CNVeM				
cnvHMM				
ExomeDep				

[a]Tools requi
[b]Tools use n

Table 3 Summary of bioinformatics tools for CNV detection using exome sequencing data

Tool	URL	Language	Input	Comments
Control-FREEC[a]	http://bioinfo-out.curie.fr/projects/	C++	SAM/BAM/pileup/Eland, BED, SOAP, arachne,	Correcting copy number using matched case-control samples or GC
CoNIFER[b]				
XHMM[b]				
ExomeCNV				
CONTRA[c]				
CONDEX				
SeqGene				
PropSeq[c]				
VarScan2[c]				
SVDetect				
inGAP-sv				
Svseq				
Nord et al.				

[a]Control-FRE
[b]Tools use
[c]Tools requi

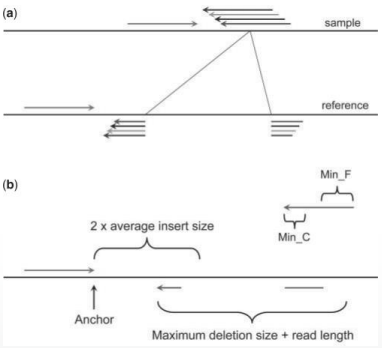
Table 4 Combinatorial bioinformatics tools for CNV detection using NGS data

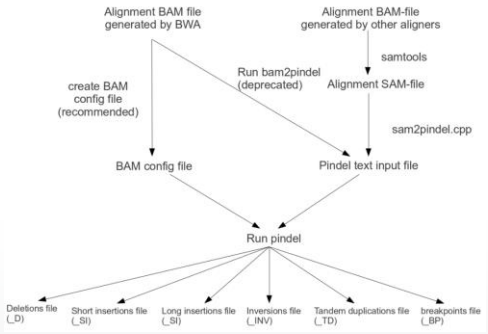
Method	URL	Language	Input	Combination[a]
NovelSeq	http://compbio.cs.sfu.ca/strvar.htm	C	FASTA/SAM	PEM+AS
HYDRA	http://code.google.com/p/hydra-sv/	Python	Discordant paired-end mappings	PEM+AS
CNVer	http://compbio.cs.toronto.edu/CNVer/	Perl, C++	BAM/aligned positions	PEM+RD
GASVPro	http://code.google.com/p/gasv/	C++	BAM	PEM+RD
Genome STRIP	http://www.broadinstitute.org/software/genomestrip/genome-strip	Java, R	BAM	PEM+RD
SVDetect	http://svdetect.sourceforge.net/	Perl	SAM/BAM/ELAND	PEM+RD
inGAP-sv	http://ingap.sourceforge.net/	Java	SAM	PEM+RD
Svseq	http://www.engr.uconn.edu/~jiz08001/svseq.html	C	FASTQ/BAM	PEM+SR
Nord et al.	N/A	N/A	N/A	RD+SR

[a]RD: read depth-based approach; PEM: paired-end mapping approach; SR: split read approach; AS: *de novo* assembly approach.

Zhao et al 2013. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 2013, 14(Suppl 11):S1

Pindel

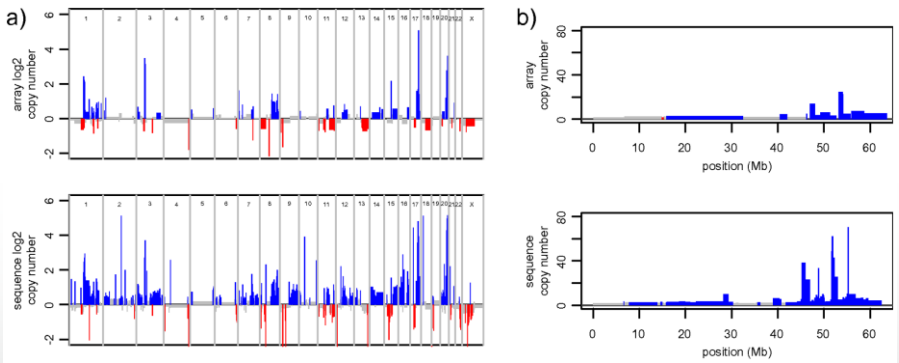





```
D 321    chr1 56173880          56174202              Supports: 15
AAGAGTTTGGTGAGTTTATAGAAATATAGGggccg<311>ataggACAAGGTACAAGGAATGGCTGAAGGAGAGAGGTTG
          GAGTTTATAGAAATATAGG                ACAAGGTACAAGGAATG       +     56173670
        GTGAGTTTATAGAAATATAGG                ACAAGGTACAAGGAA         +     56173677
          GAGTTTATAGAAATATAGG                ACAAGGTACAAGGAATG       +     56173681
       TGGTGAGTTTATAGAAATATAGG               ACAAGGTACAAGG           +     56173687
          GAGTTTATAGAAATATAGG                ACAAGGTACAAGGAATG       +     56173690
        GTGAGTTTATAGAAATATAGG                ACAAGGTACAAGGAA         +     56173695
    AGTTTGGTGAGTTTATAGAAATATAGG              ACAAGGTACAAGGA          +     56173697
        GTGAGTTTATAGAAATATAGG                ACAAGGTACAAGGAA         +     56173700
          AGTTTATAGAAATATAGG                 ACAAGGTACAAGGAATGG      +     56173710
     TTTGGTGAGTTTATAGAAATATAGG               ACAAGGTACAA             -     56174339
          TGAGTTTATAGAAATATAGG               ACAAGGTACAAGGAATG       -     56174356
```

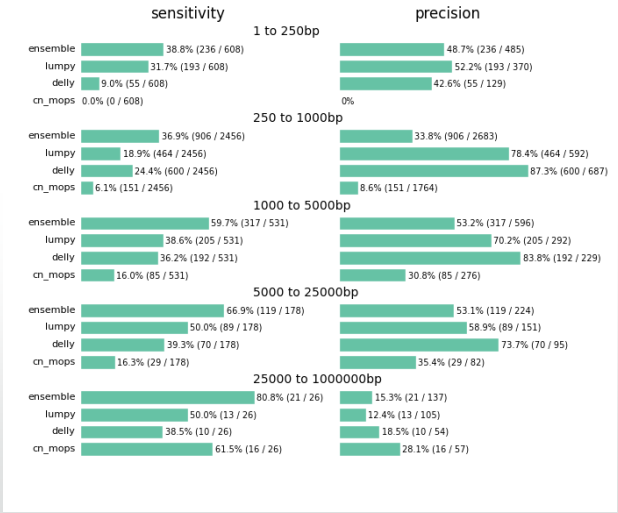
Ye et al. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009 Nov 1;25(21):2865-71

ReadDepth



Miller et al. (2011) ReadDepth: A Parallel R Package for Detecting Copy Number Alterations from Short Sequencing Reads. PLoS ONE 6(1): e16327.

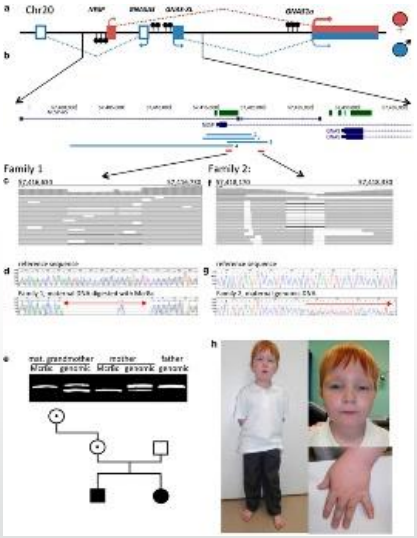
Ensemble of SV tools



Source: https://bcbio.wordpress.com/tag/cnv/

Combining multiple callers helps to detect additional variants

Ensemble of SV tools



Combination of read depth and split read method, Pindel helped to identify two deletion in PHP1B patients

Rezwan et al 2015. Very small deletions within the *NESP55* gene in pseudohypoparathyroidism type 1b. E*uropean Journal of Human Genetics* (2015) **23,** 494–499

Summary

- SVs/CNVs are as important genetic variations as SNVs
- SVs/CNVs are related to several diseases and phenotypes
- Several comparative genomic hybridisation and SNP array methods available but NGS offers more.
- NGS methods have statistical challenges
- Single NGS method/algorithm may not be suitable for identifying SVs/CNVs

UNIVERSITY OF Southampton

?