Structural Variation

Tools like GATK, samtools can't currently detect larger structural changes easily, but…

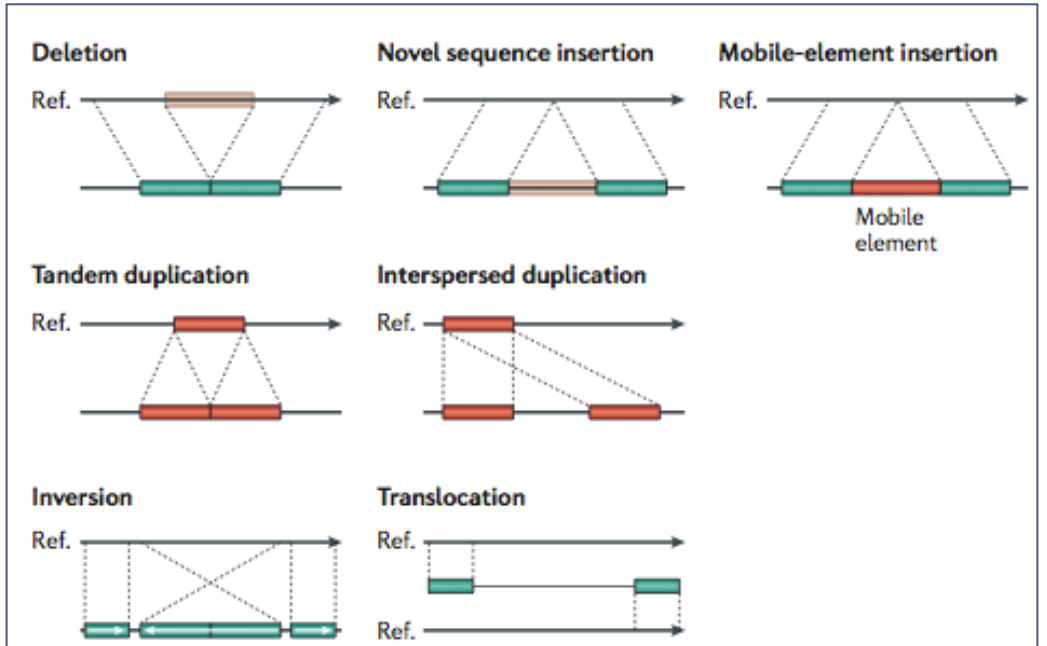
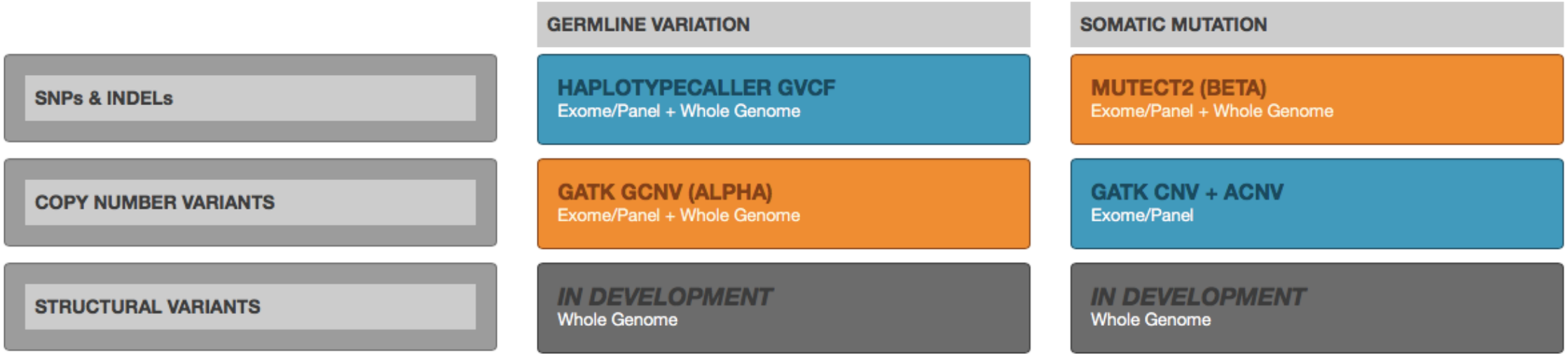


Figure 1 | **Classes of structural variation.** Traditionally, structural variation refers to genomic alterations that are larger than 1 kb in length, but advances in discovery techniques have led to the detection of smaller events. Currently, >50 bp is used as an operational demarcation between indels and copy number variants (CNVs). The schematic depicts deletions, novel sequence insertions, mobile-element insertions, tandem and interspersed segmental duplications, inversions and translocations in a test genome (lower line) when compared with the reference genome.

Alkan *et al*, Nature Genetics **12**:363, 2011

Structural Variation

GATK (v3.7 or higher), this is changing:



https://software.broadinstitute.org/gatk/best-practices/

Structural Variation

Detection using NGS data generally requires multi-layer analyses, may focus on specific SV types

Common tools:

- CNVnator – gross detection of CNVs
- BreakDancer, Cortex-Var – breakpoint detection
- Pindel – large deletions
- Manta

Recent tools (lumpy-sv) integrate approaches

Structural Variation Strategies

Read depth

- Region deviates from expected read depth

Read pair discordance

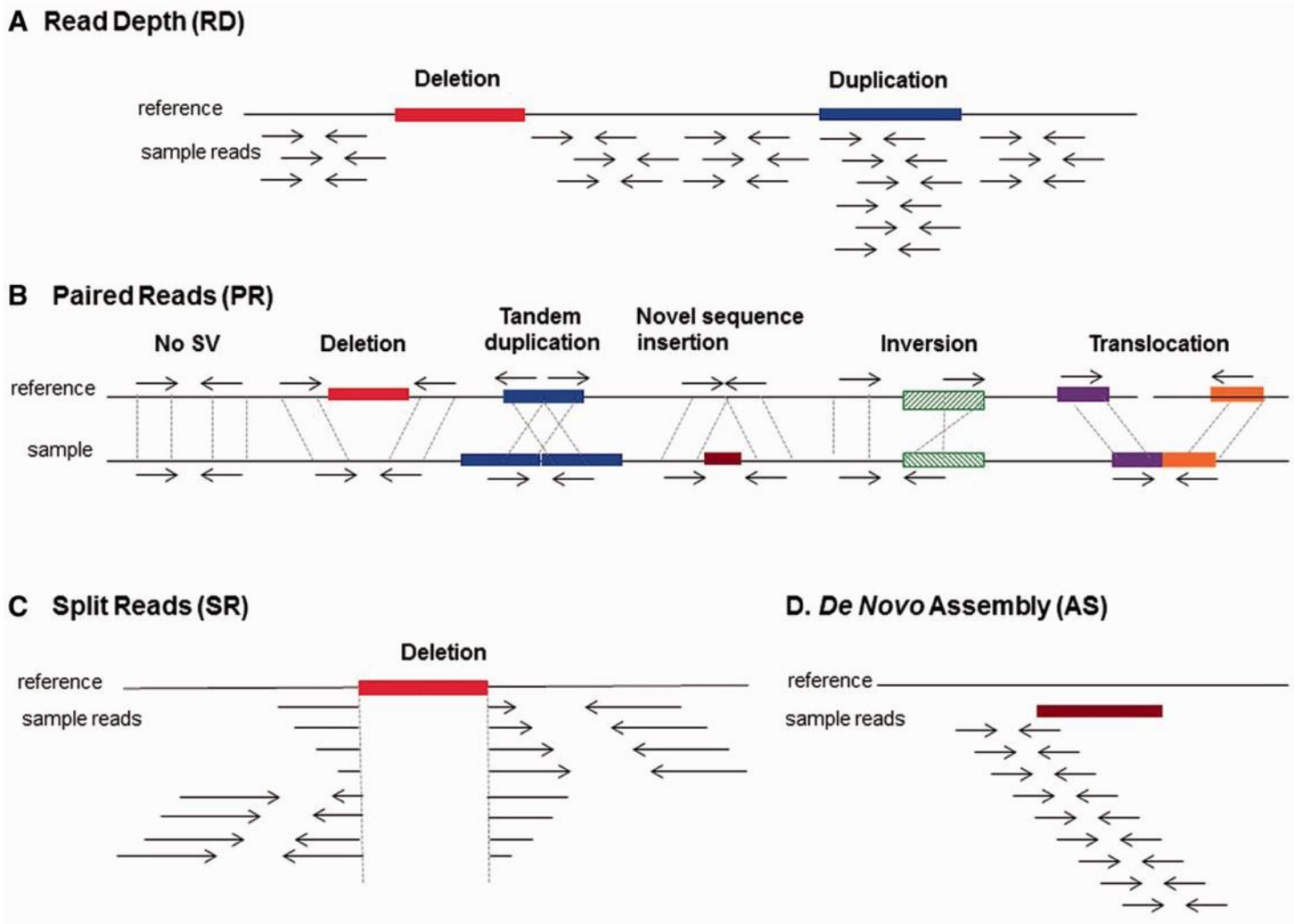
- Insert size is off, orientation of reads is wrong

Split reads

- Single read is split, parts align in two distinct unique locations

Assembly

- Reference-based local assemblies indicate inconsistencies



G. Escaramís, E. Docampo, R. Rabionet, Briefing Func. Genomics 14(5) September 2015

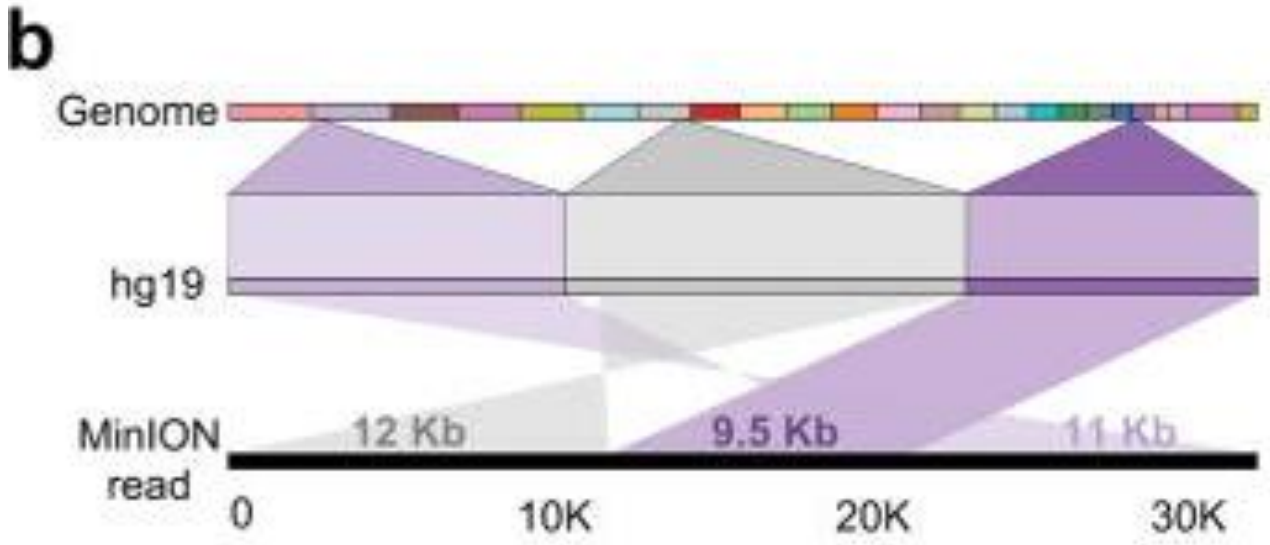
Structural Variation

Still an active area of research

Problems:

- Lots of false positives
- Hard to compare methodologies

More recent publications use long-read technology for structural variant detection ($$$)



(**b**) A complex SV of two translocations detected by *Picky* from a **32.5 Kb nanopore read.** The alignments were visualized by Ribbon (https://github.com/MariaNattestad/ribbon)

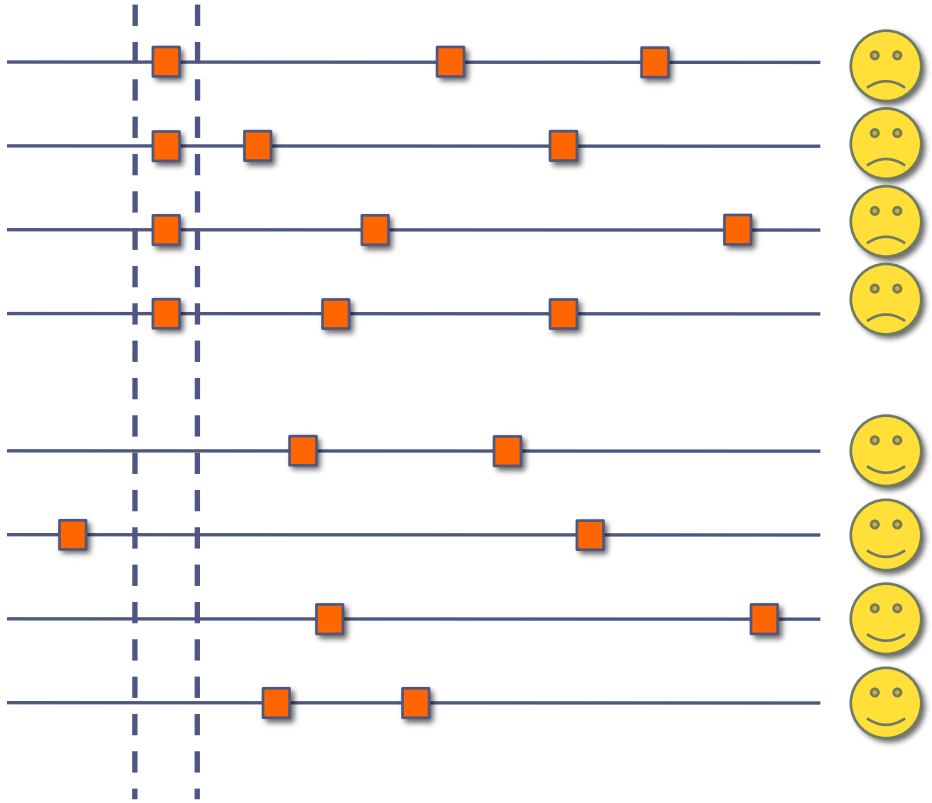
Association Studies

Genome-wide association studies (GWAS)

Trying to determine whether specific variant(s) in many individuals can be associated with a trait

Ex: comparison of groups of people with a disease (cases) and without (controls)

Finding the causal variant in ideal situations*



Spot the variant that is common amongst all affected but absent in all unaffected

This variant is in a gene with known function and causes the protein to be disrupted

* e.g. some rare autosomal disease