

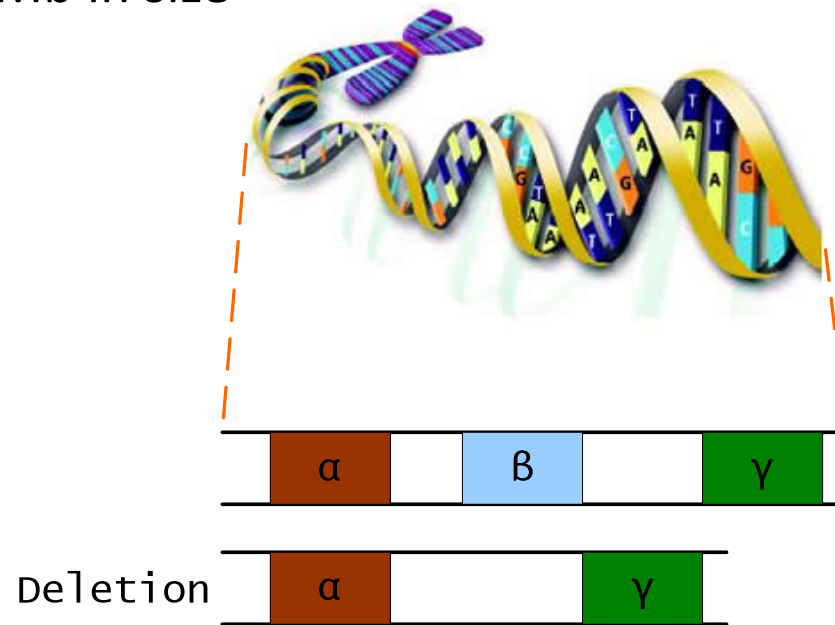
Structural Variations

02-710 Computational Genomics

Seyoung Kim

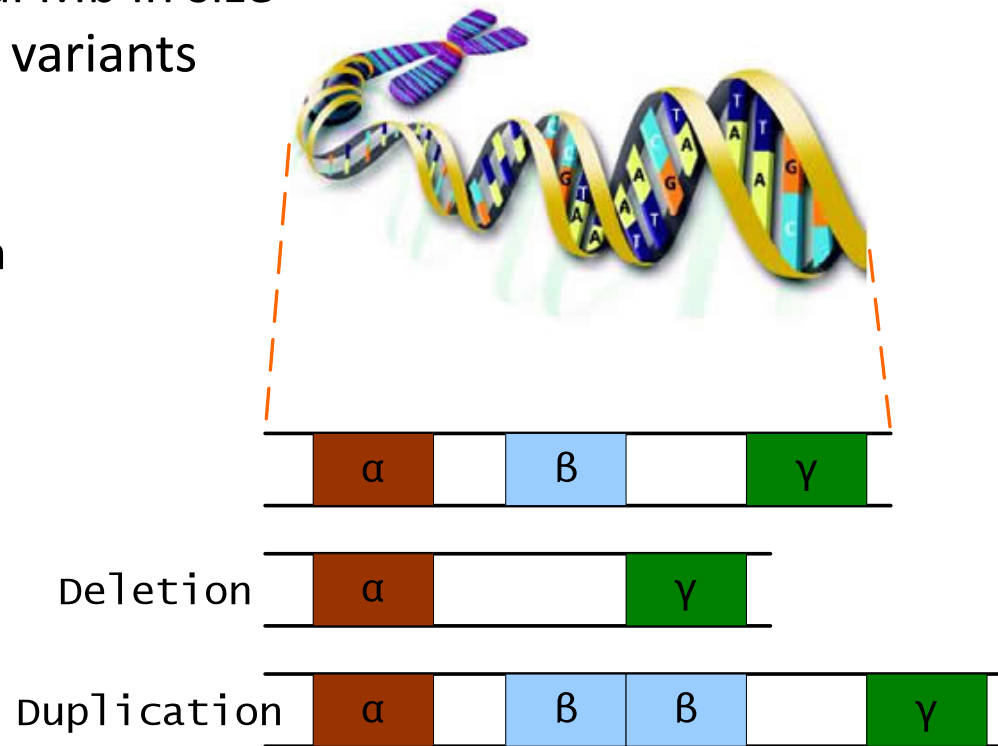
Genomic Rearrangements/ Structural Variations (SVs)

- 1 Kb to several Mb in size



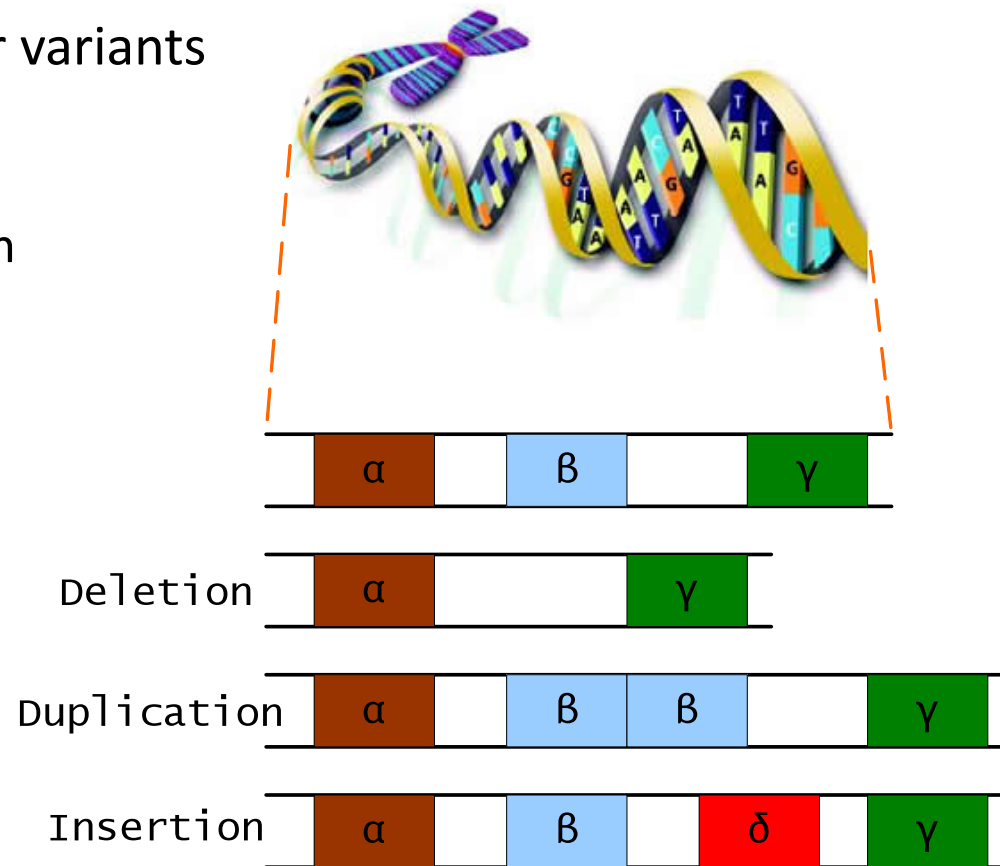
Genomic Rearrangements/ Structural Variations (SVs)

- 1 Kb to several Mb in size
- Copy number variants (CNVs)
 - Deletion
 - Duplication



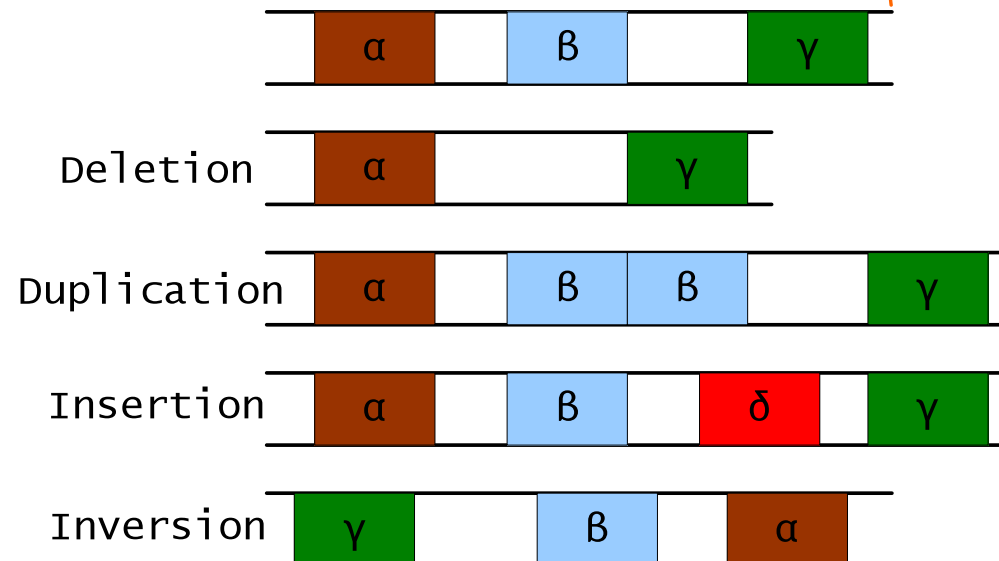
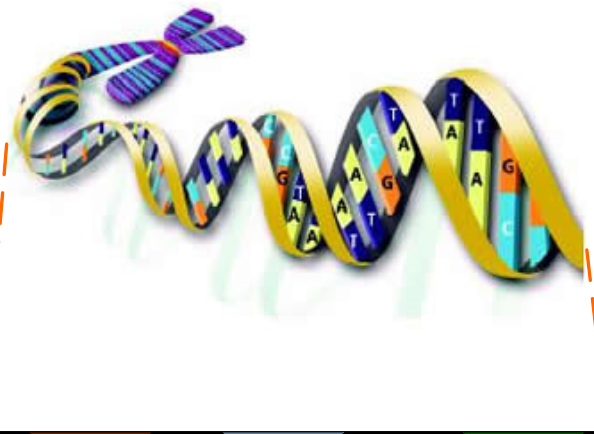
Genomic Rearrangements/ Structural Variations (SVs)

- 1 Kb to several Mb in size
- Copy number variants (CNVs)
 - Deletion
 - Duplication
- Insertion



Genomic Rearrangements/ Structural Variations (SVs)

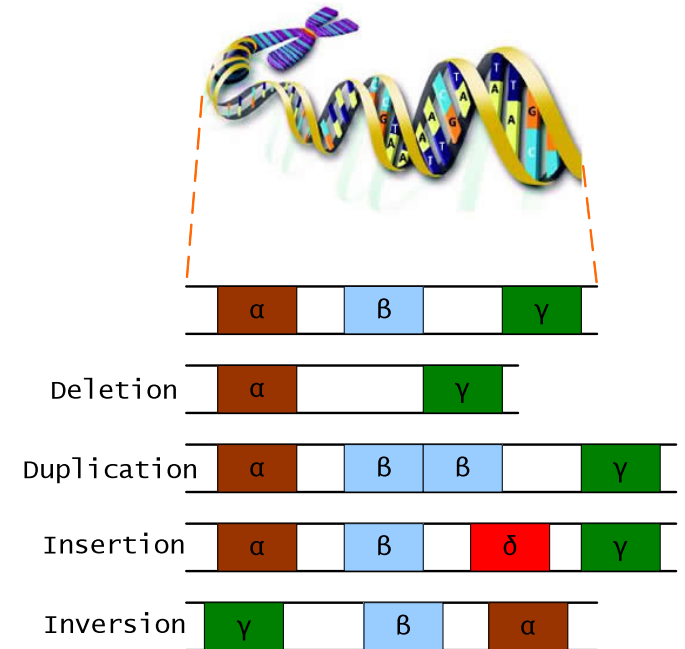
- 1 Kb to several Mb in size
- Copy number variants (CNVs)
 - Deletion
 - Duplication
- Insertion, Inversion



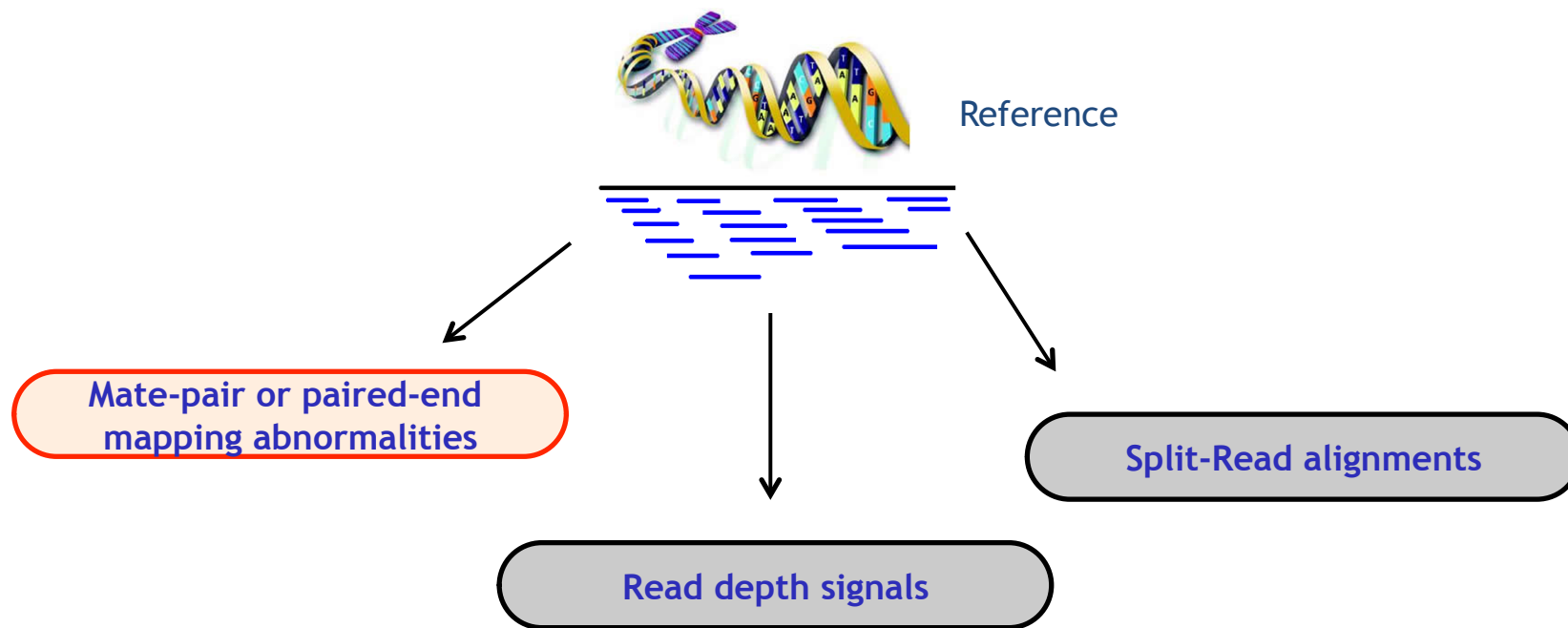
courtesy of Tobias Rausch (EMBL)

Genomic Rearrangements/ Structural Variations

- 1 Kb to several Mb in size
- Copy number variants
 - Deletion
 - Duplication
- Insertion, Inversion, Translocation
- Either neutral or non-neutral in function
- Non-neutral mechanisms
 - Disrupting genes
 - Creating fusion genes
 - Copy number changes of dosage-sensitive genes

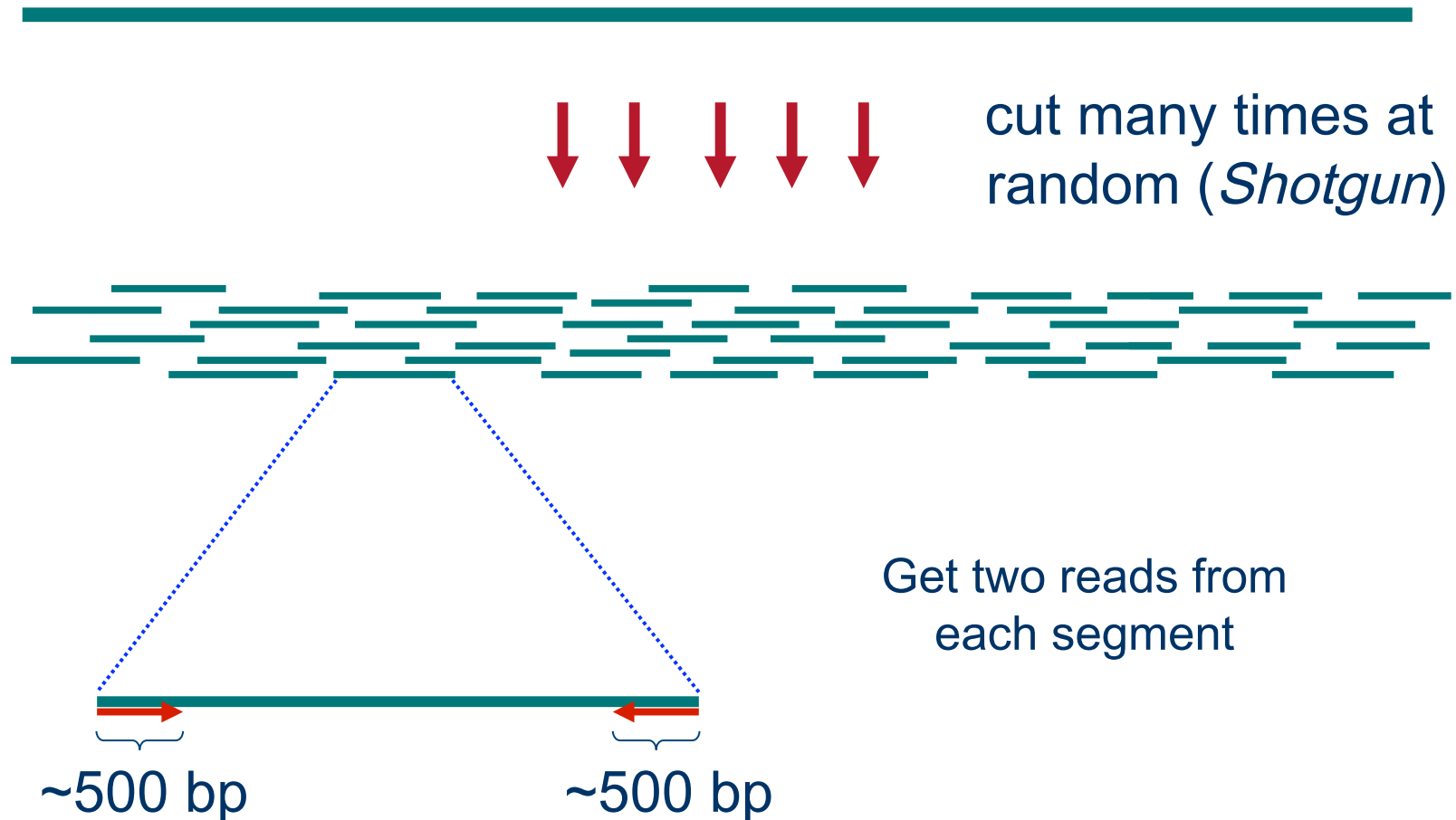


Computational Methods for Detecting Genomic Rearrangements



Paired-end Sequencing

genomic segment



Definition of Coverage



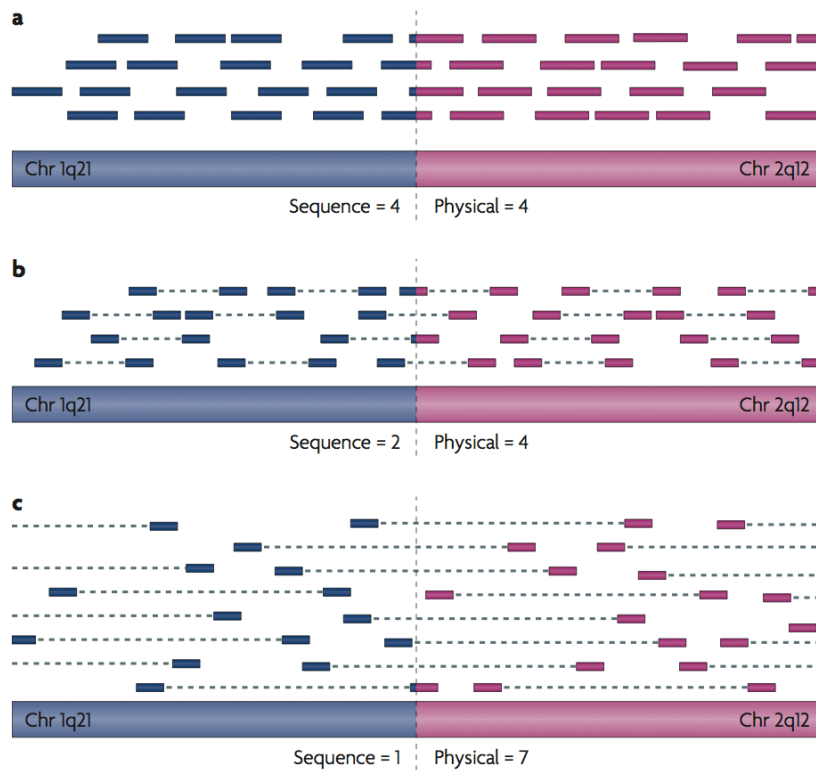
Length of genomic segment: L

Number of reads: n

Length of each read: l

Definition: Coverage $C = n l / L$

Depth of Coverage and Physical Coverage



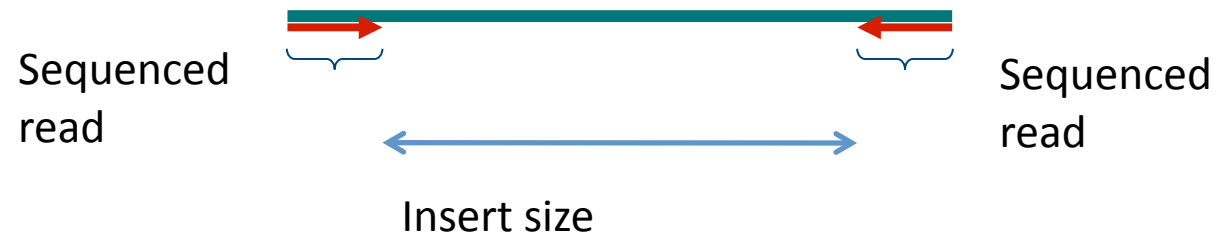
- Single-end sequencing

- Paired-end sequencing

- Paired-end sequencing

Insert Size for a Mate Pair

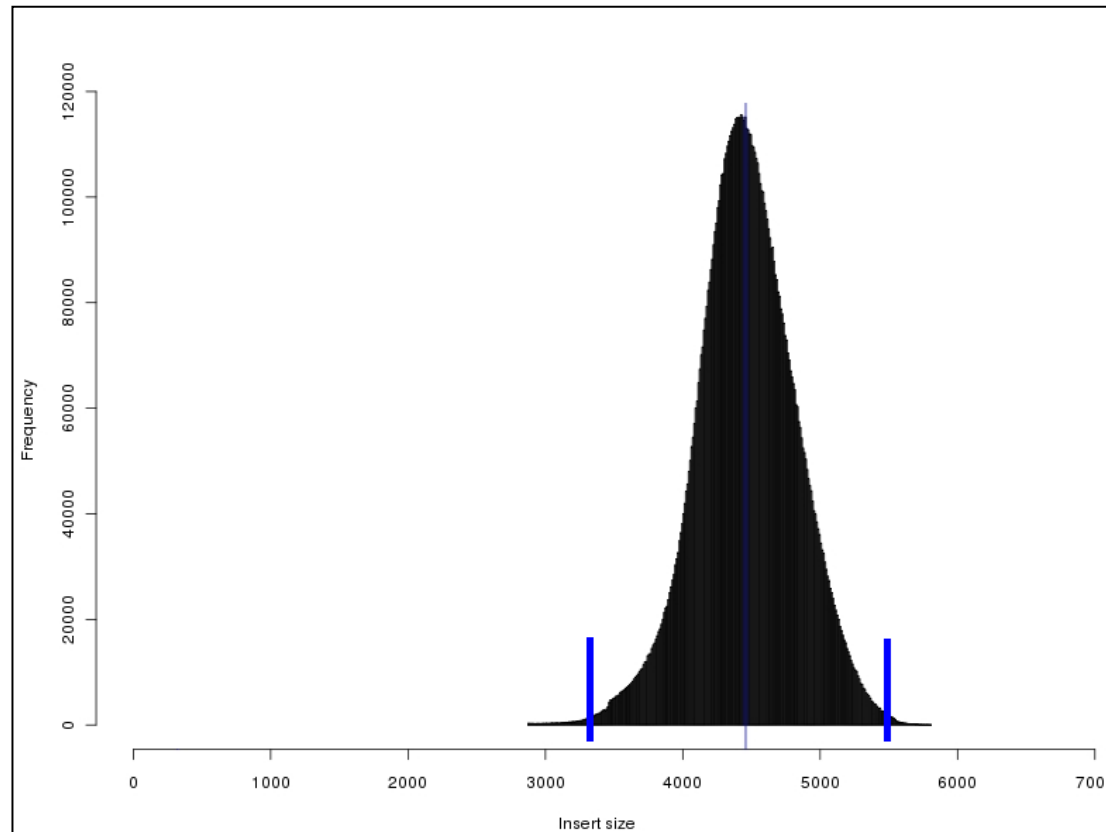
For each fragment



Let's assume we know the insert size distribution.

Paired-end data

- Paired-end NGS (insert size distribution known due to fragment size selection)

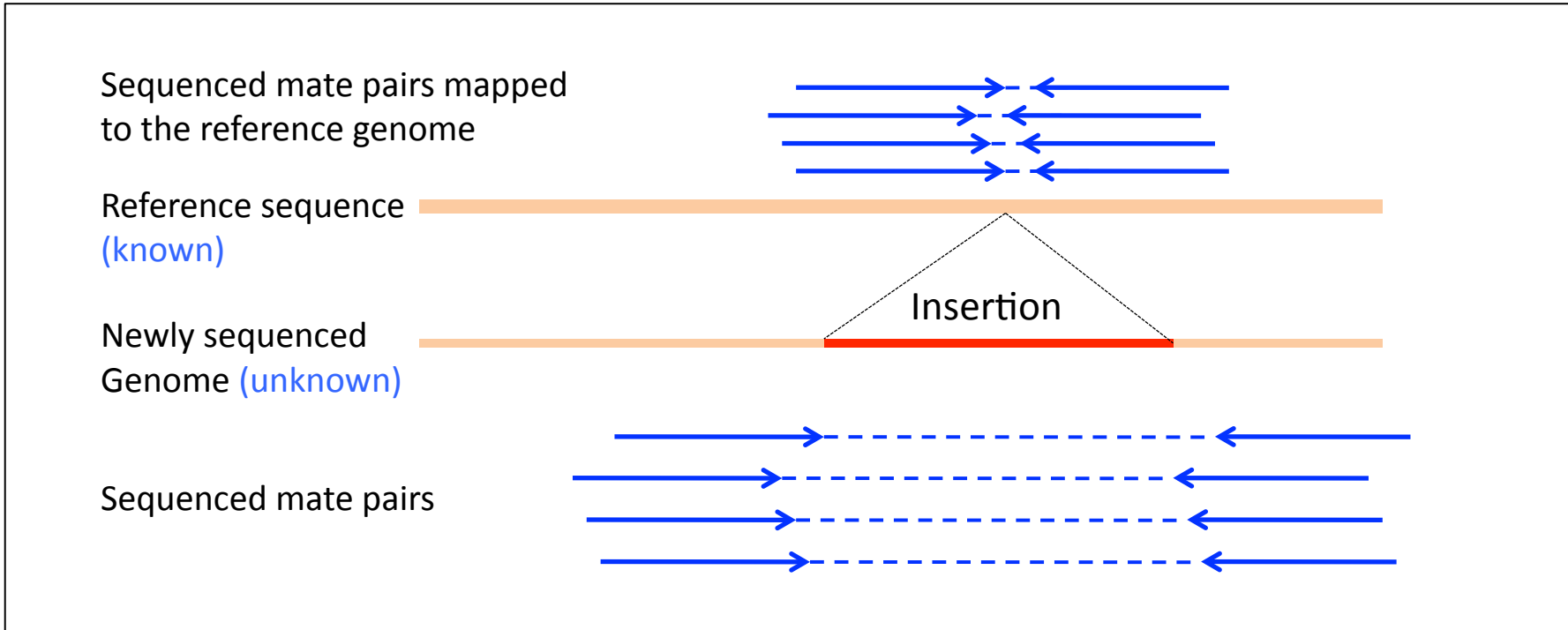
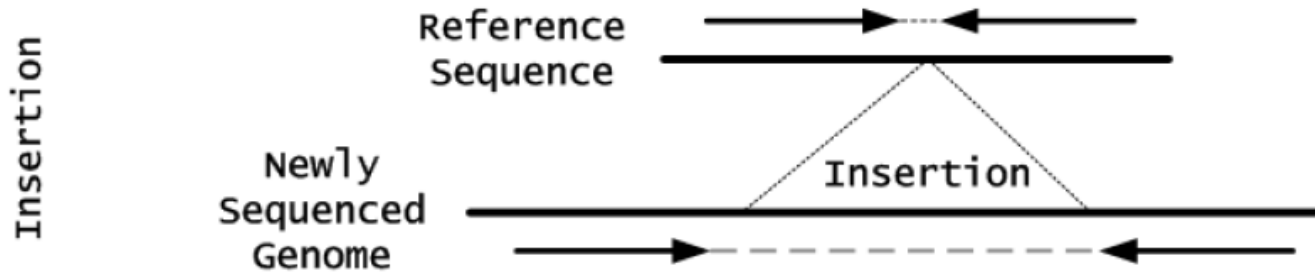


Mate-pair or paired-end mapping abnormalities

Read-depth signals

Split-read alignments

Local assembly



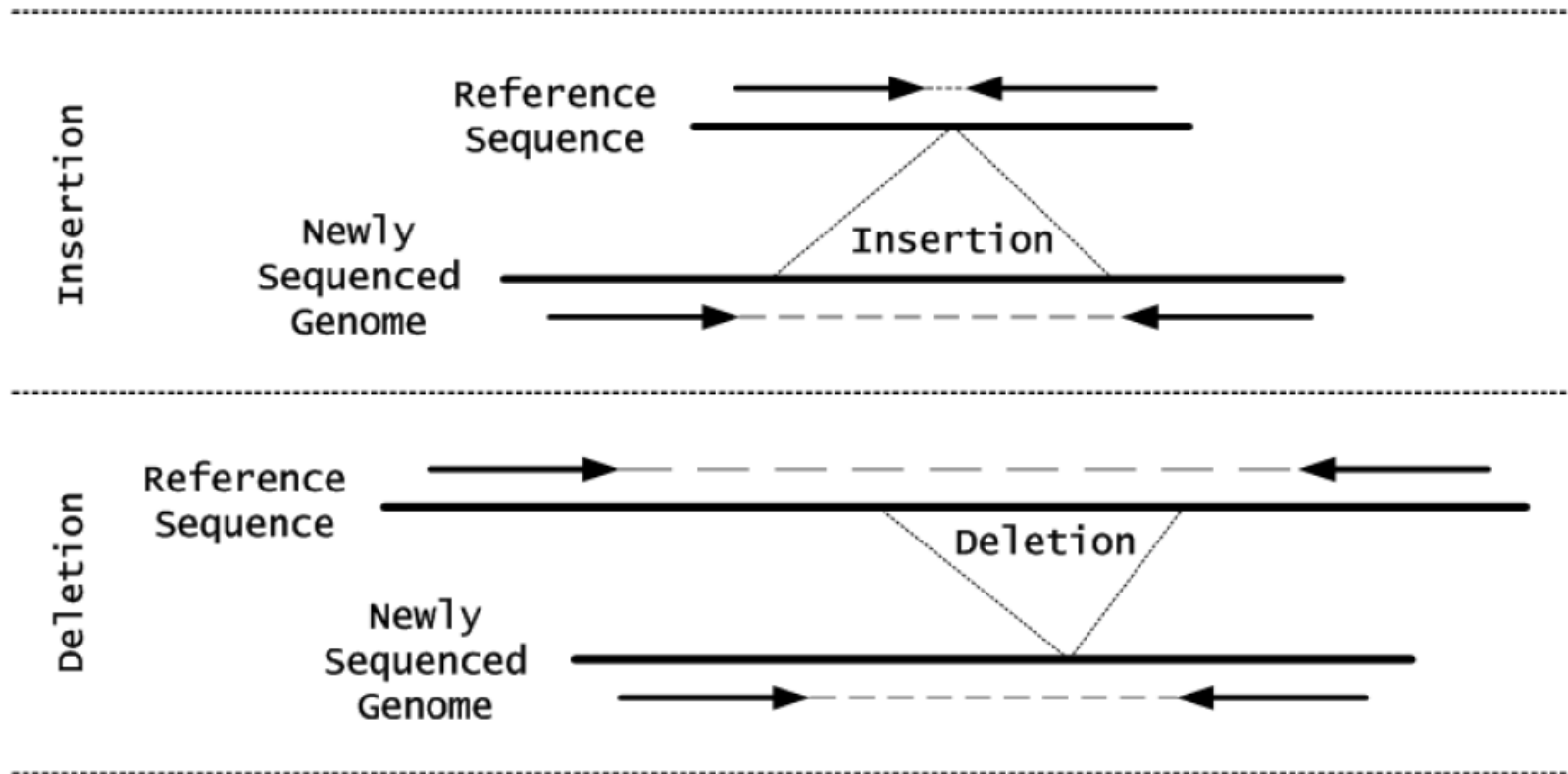
courtesy of Tobias Rausch (EMBL)

Mate-pair or paired-end mapping abnormalities

Read-depth signals

Split-read alignments

Local assembly



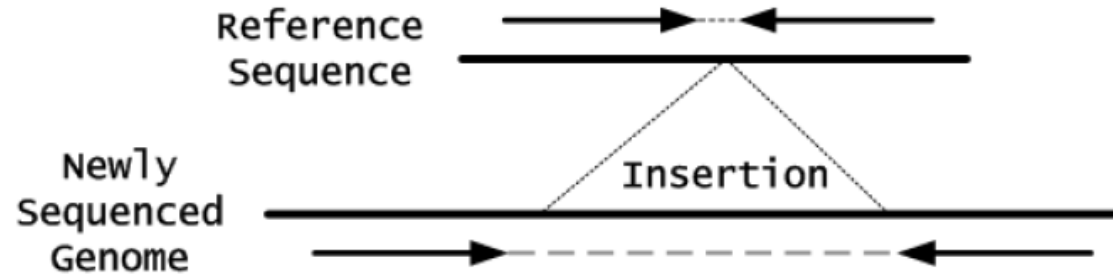
Mate-pair or paired-end mapping abnormalities

Read-depth signals

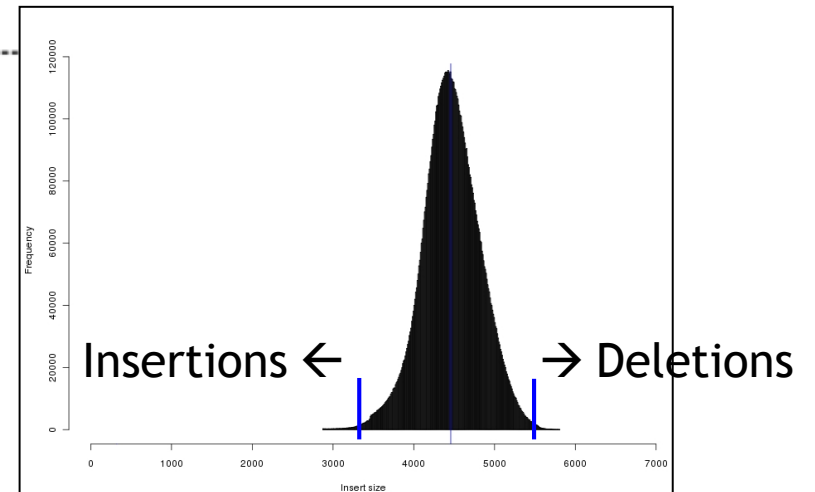
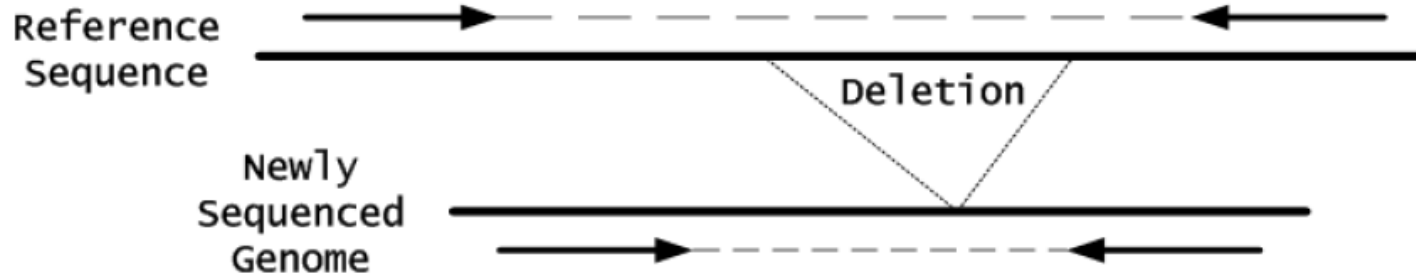
Split-read alignments

Local assembly

Insertion



Deletion



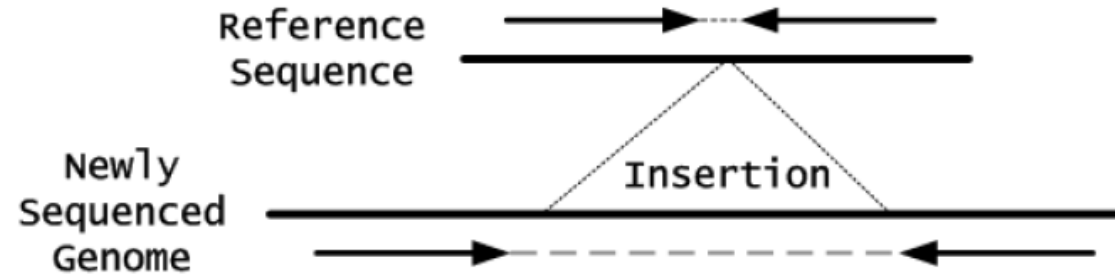
Mate-pair or paired-end mapping abnormalities

Read-depth signals

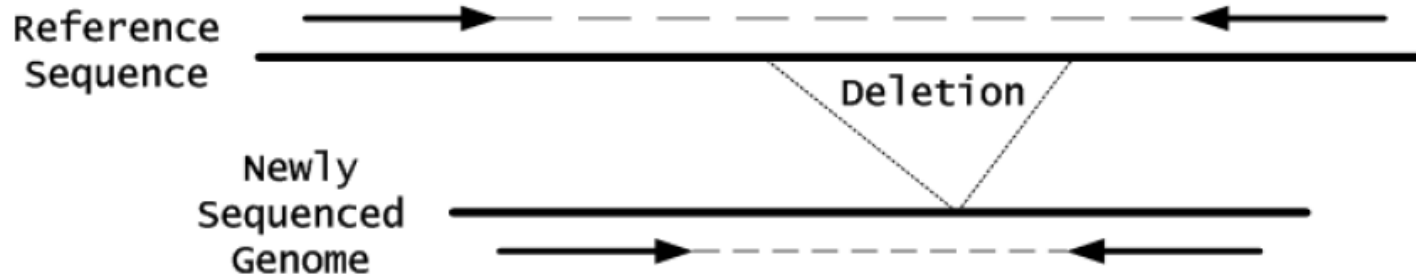
Split-read alignments

Local assembly

Insertion



Deletion



Inversion



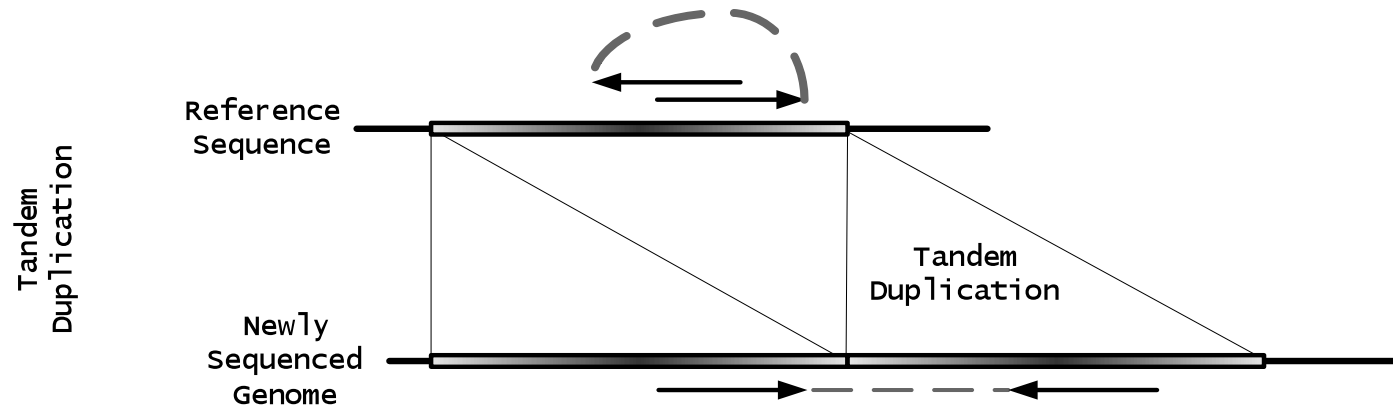
courtesy of Tobias Rausch (EMBL)

Mate-pair or paired-end mapping abnormalities

Read-depth signals

Split-read alignments

Local assembly

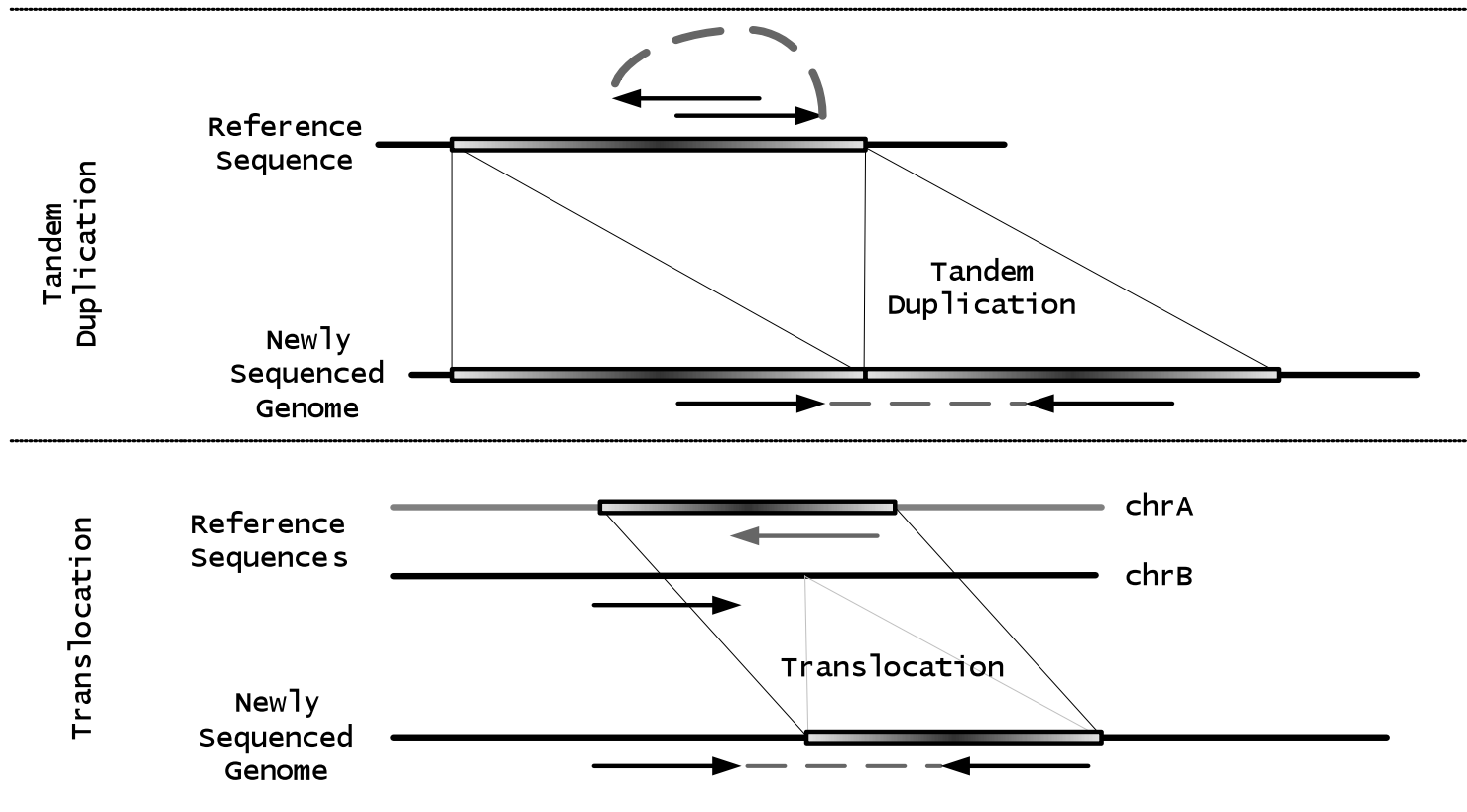


Mate-pair or paired-end mapping abnormalities

Read-depth signals

Split-read alignments

Local assembly

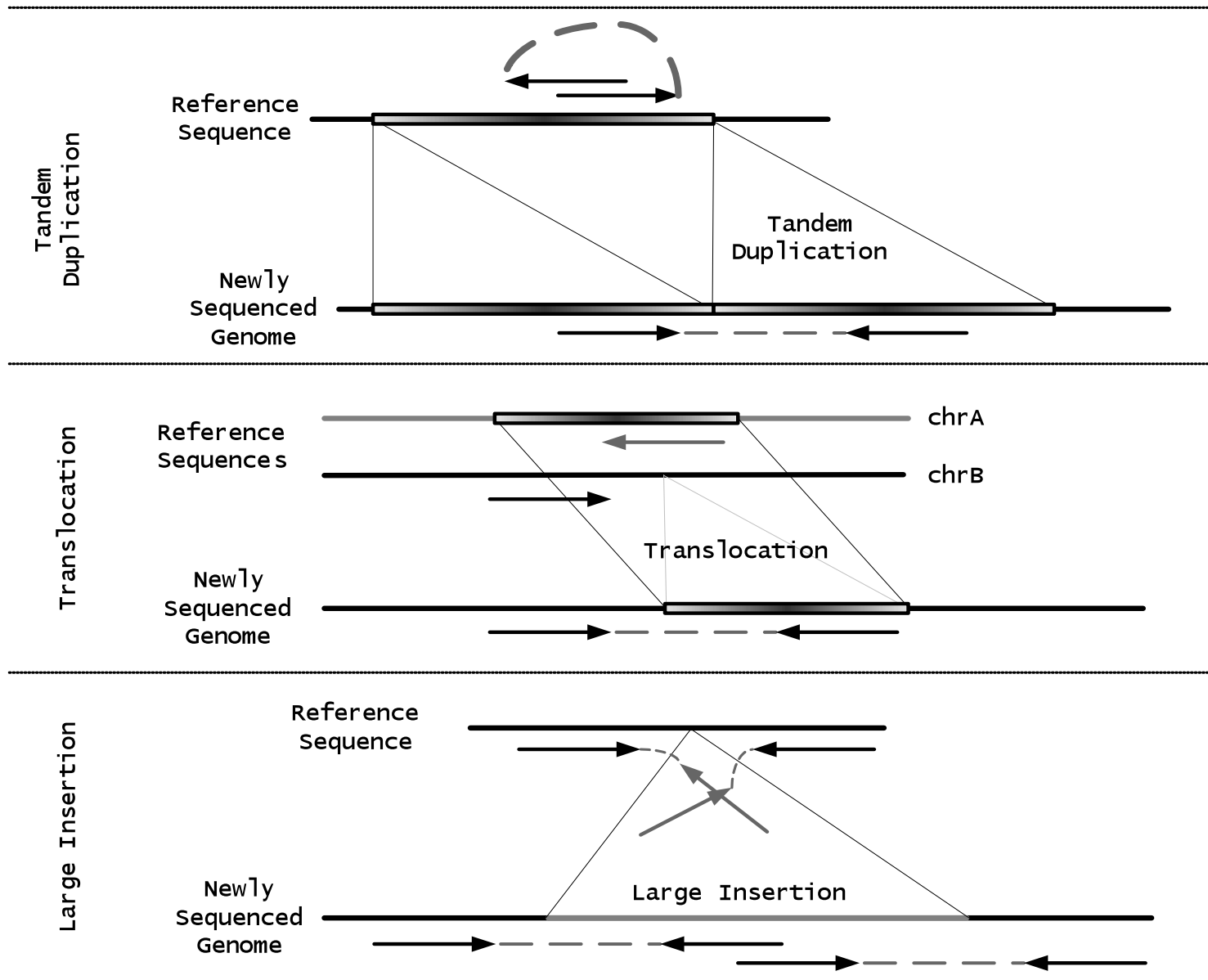


Mate-pair or paired-end mapping abnormalities

Read-depth signals

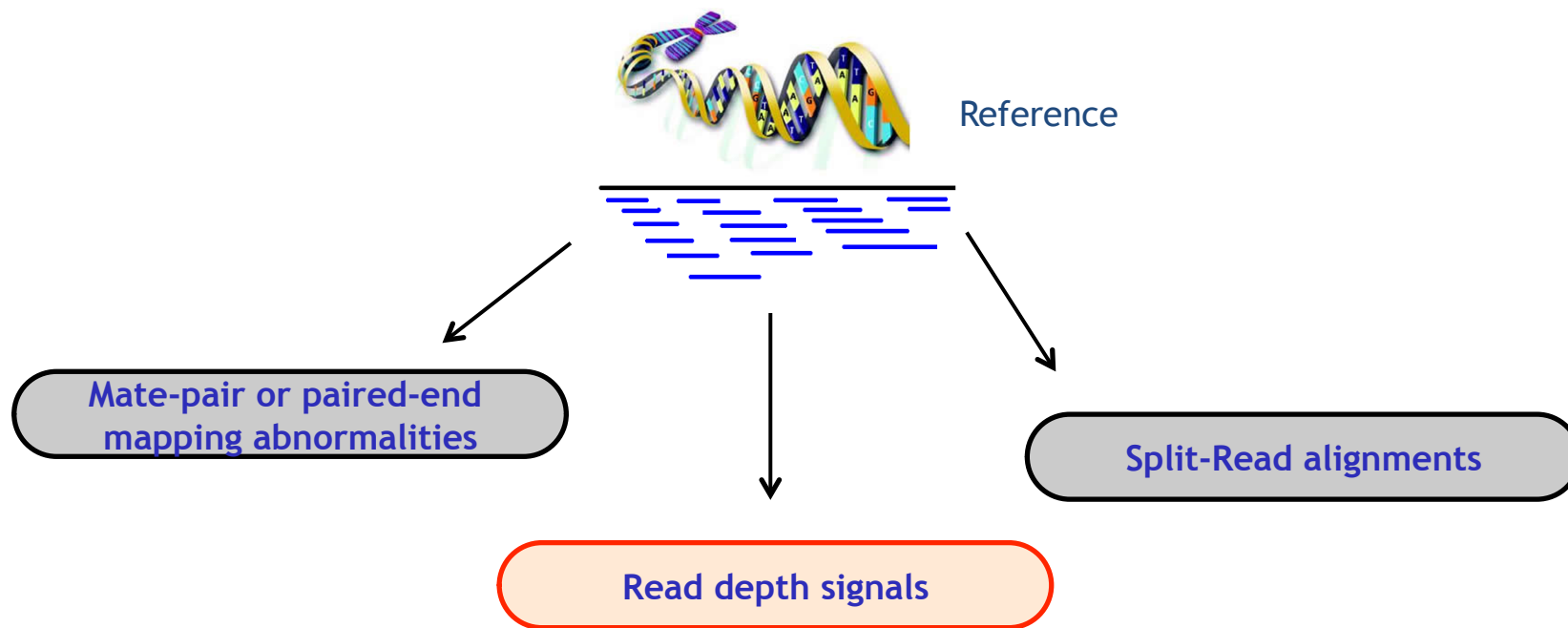
Split-read alignments

Local assembly

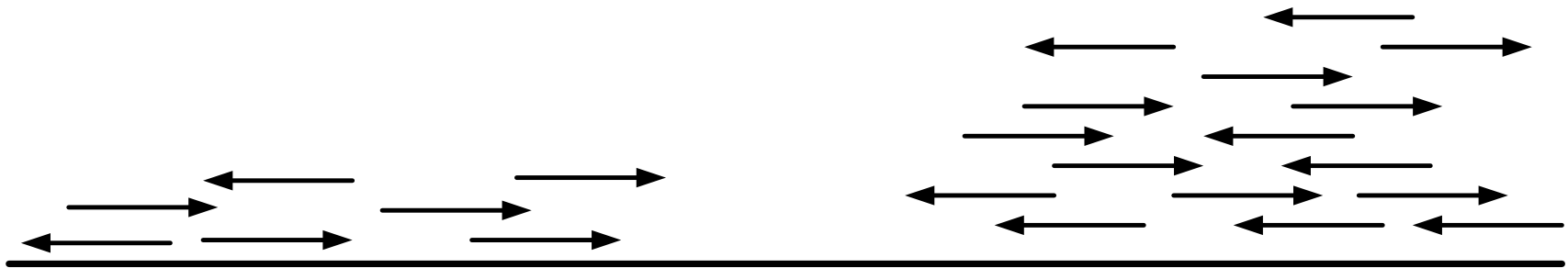


courtesy of Tobias Rausch (EMBL)

Computational Methods for Detecting Genomic Rearrangements

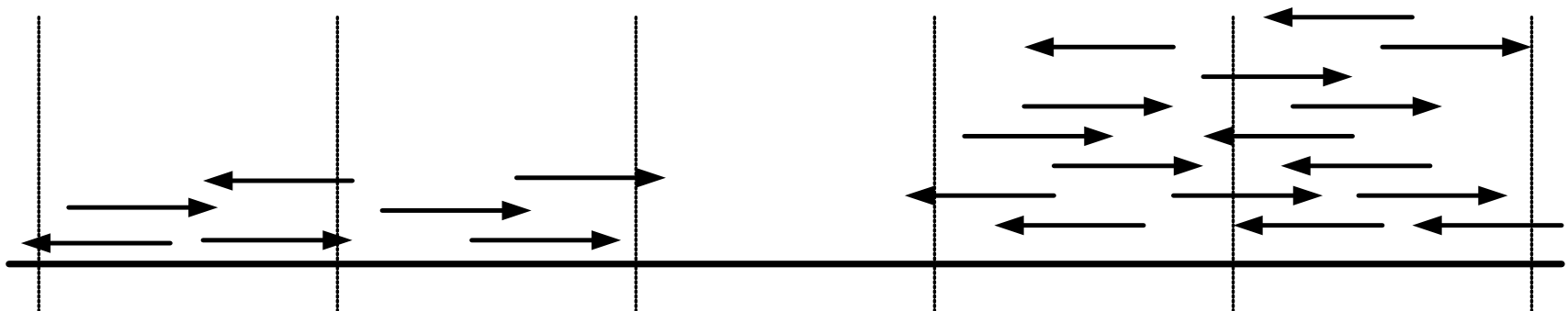


| | | | |
|---|---------------------------|-----------------------|----------------|
| Mate-pair or paired-end mapping abnormalities | Read-depth signals | Split-read alignments | Local assembly |
|---|---------------------------|-----------------------|----------------|



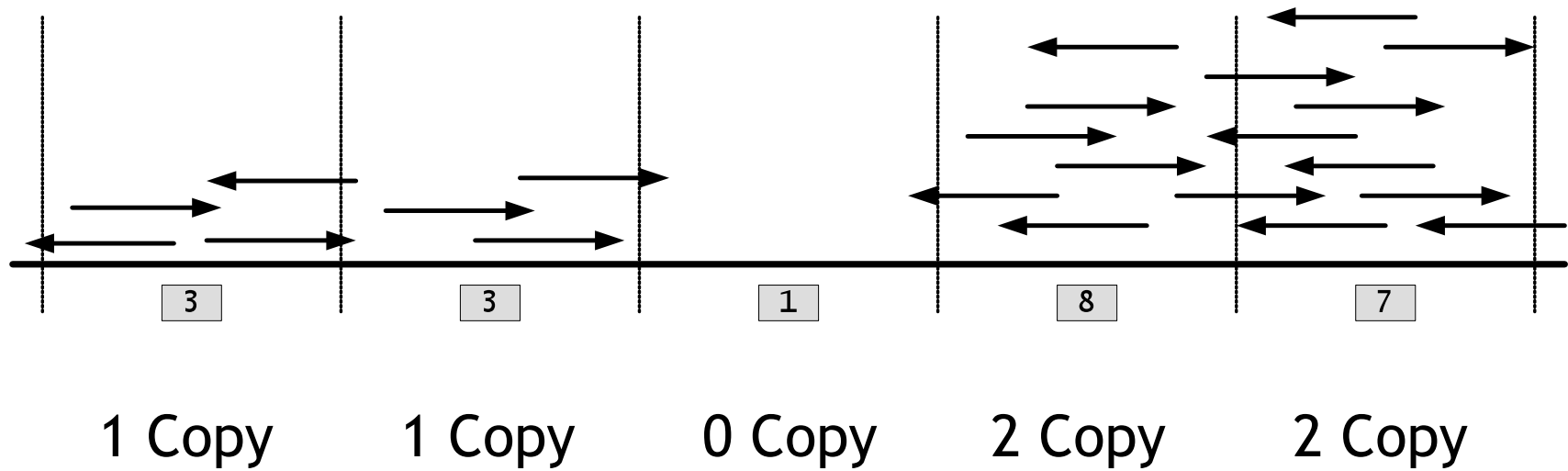
courtesy of Tobias Rausch (EMBL)

| | | | |
|---|---------------------------|-----------------------|----------------|
| Mate-pair or paired-end mapping abnormalities | Read-depth signals | Split-read alignments | Local assembly |
|---|---------------------------|-----------------------|----------------|



courtesy of Tobias Rausch (EMBL)

| | | | |
|---|---------------------------|-----------------------|----------------|
| Mate-pair or paired-end mapping abnormalities | Read-depth signals | Split-read alignments | Local assembly |
|---|---------------------------|-----------------------|----------------|



courtesy of Tobias Rausch (EMBL)

□ Chiang et al. (2009)

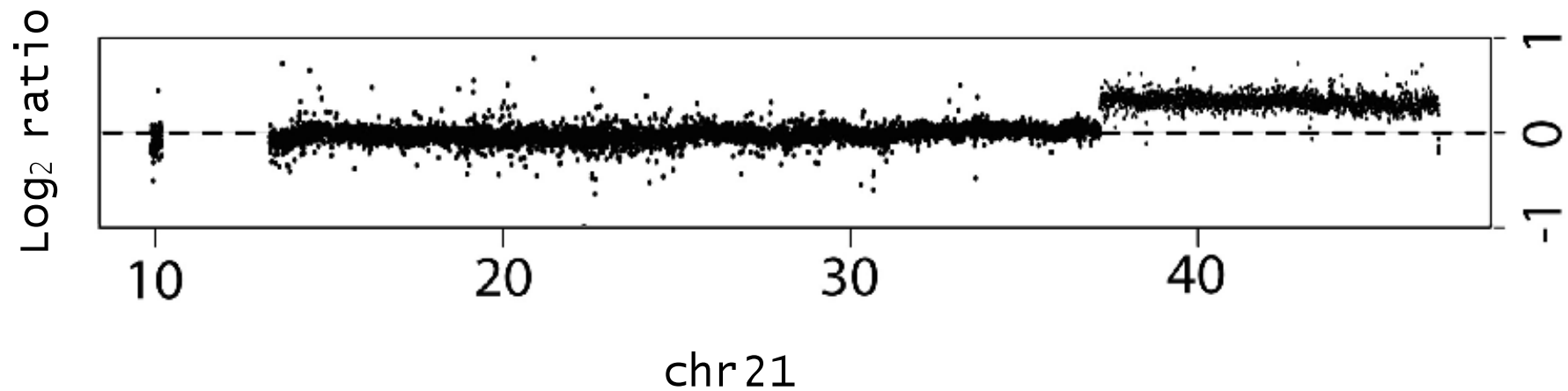
Mate-pair or paired-end mapping abnormalities

Read-depth signals

Split-read alignments

Local assembly

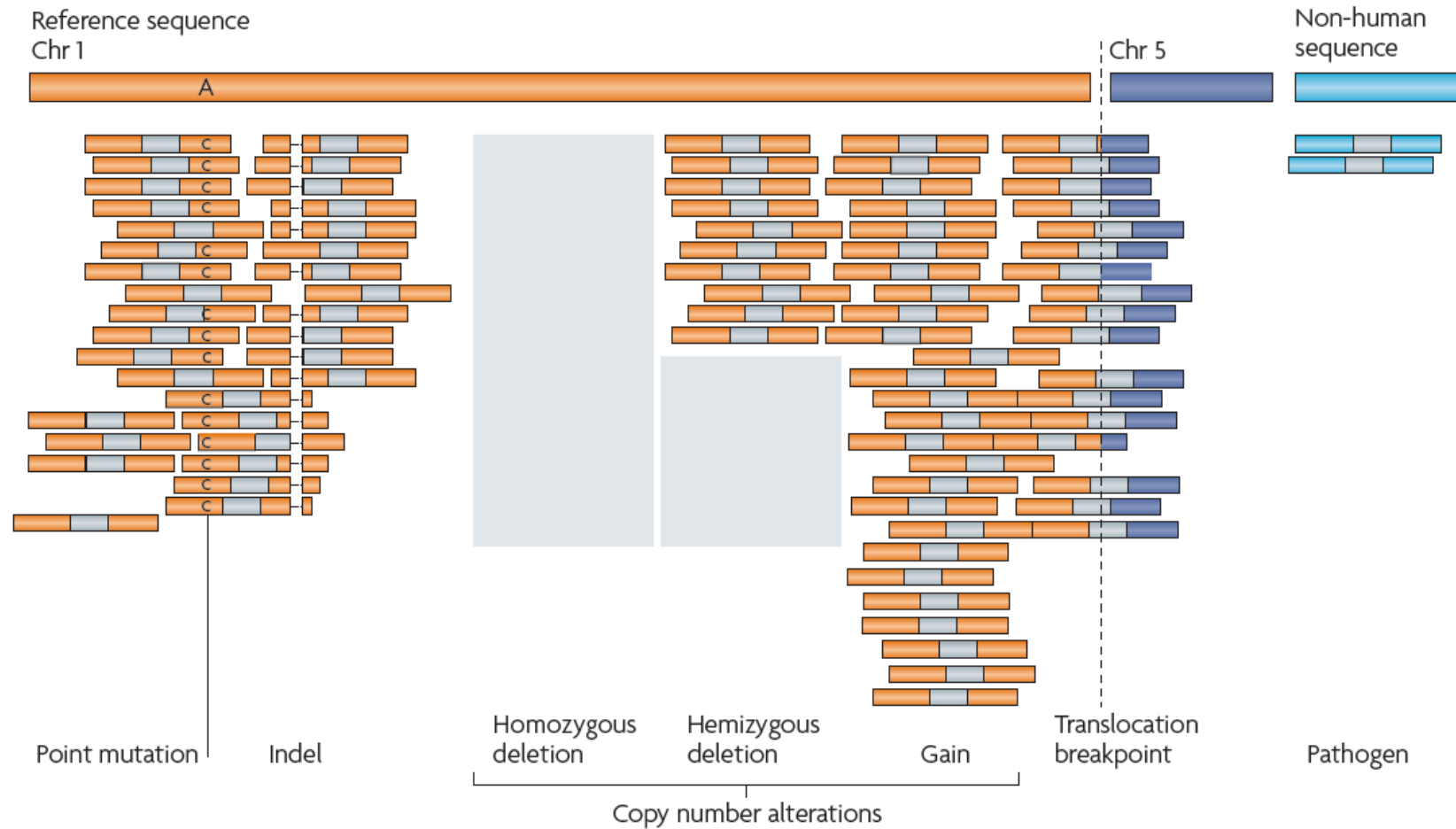
- Down-Syndrom
 - Partial Trisomie 21



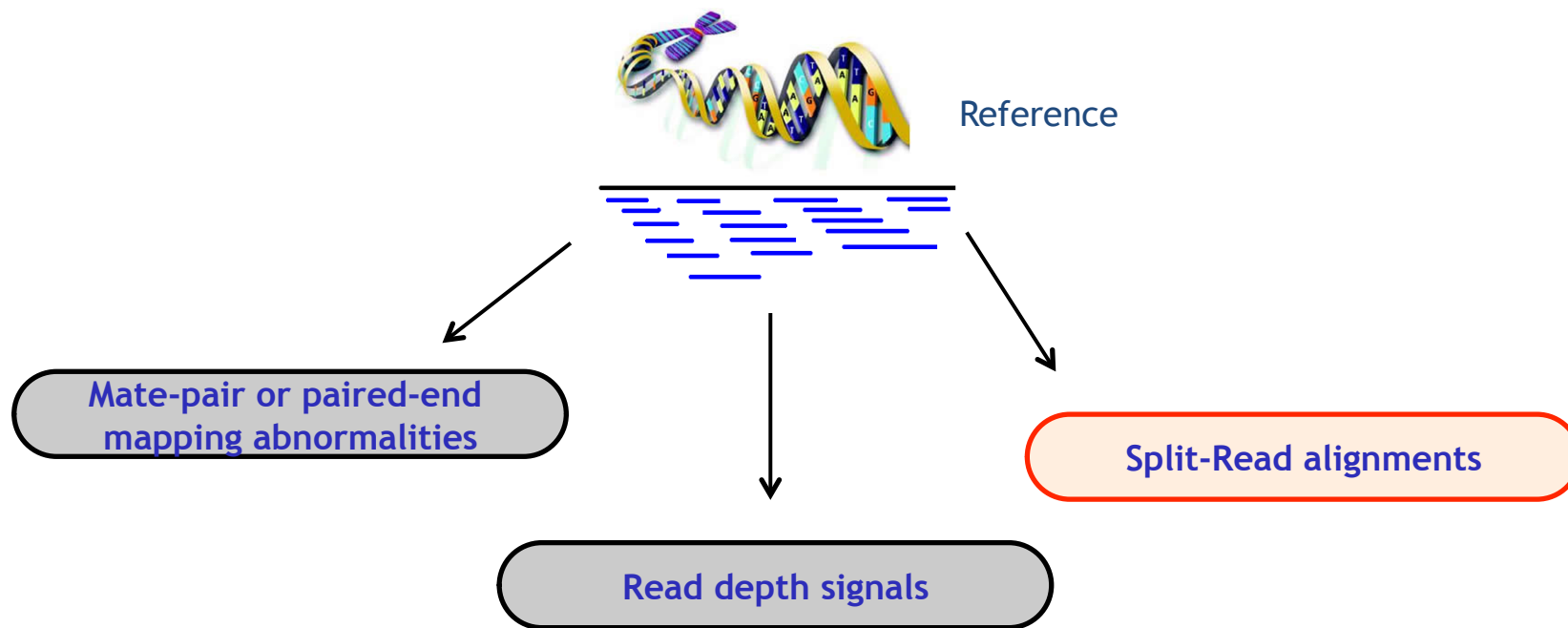
$$\log_2 \frac{\# \text{ Reads}_{Disease}}{\# \text{ Reads}_{Normal}}$$

□ Xie et al. (2009)

Detecting Genetic Polymorphisms from Shotgun Sequencing

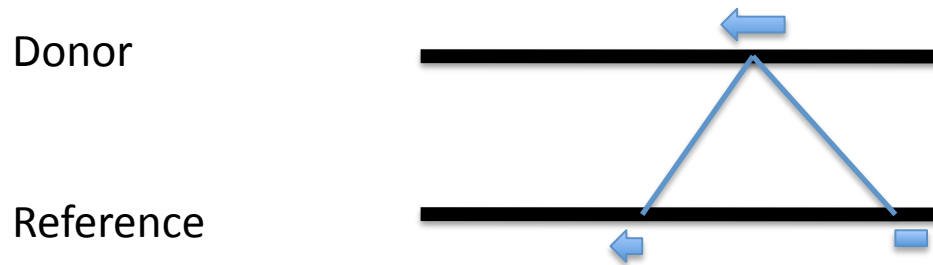


Computational Methods for Detecting Genomic Rearrangements



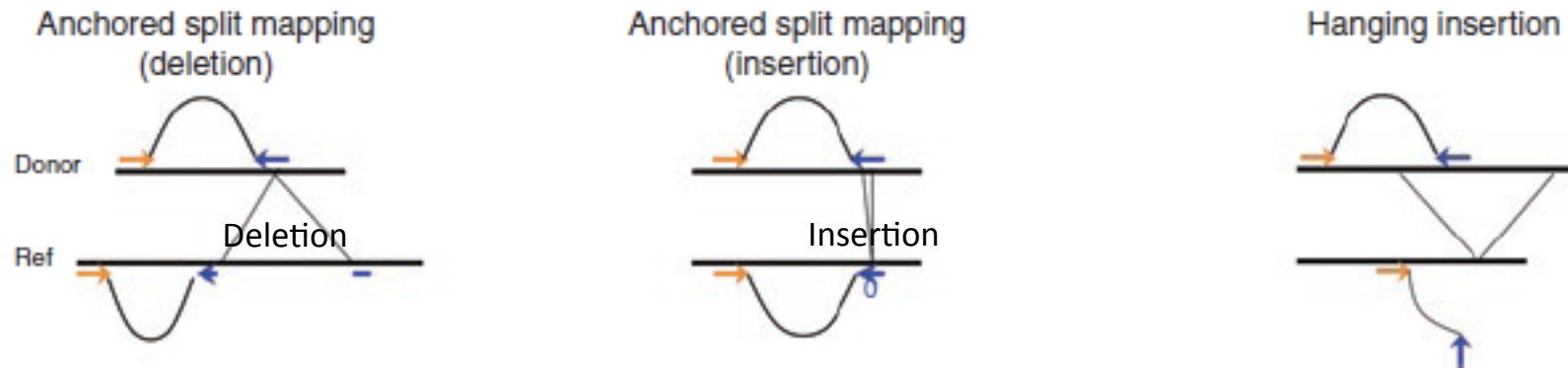
With reads of length 40-100 bps are we able to find the exact breakpoint of a structural variation?

Yes – using split-read mapping



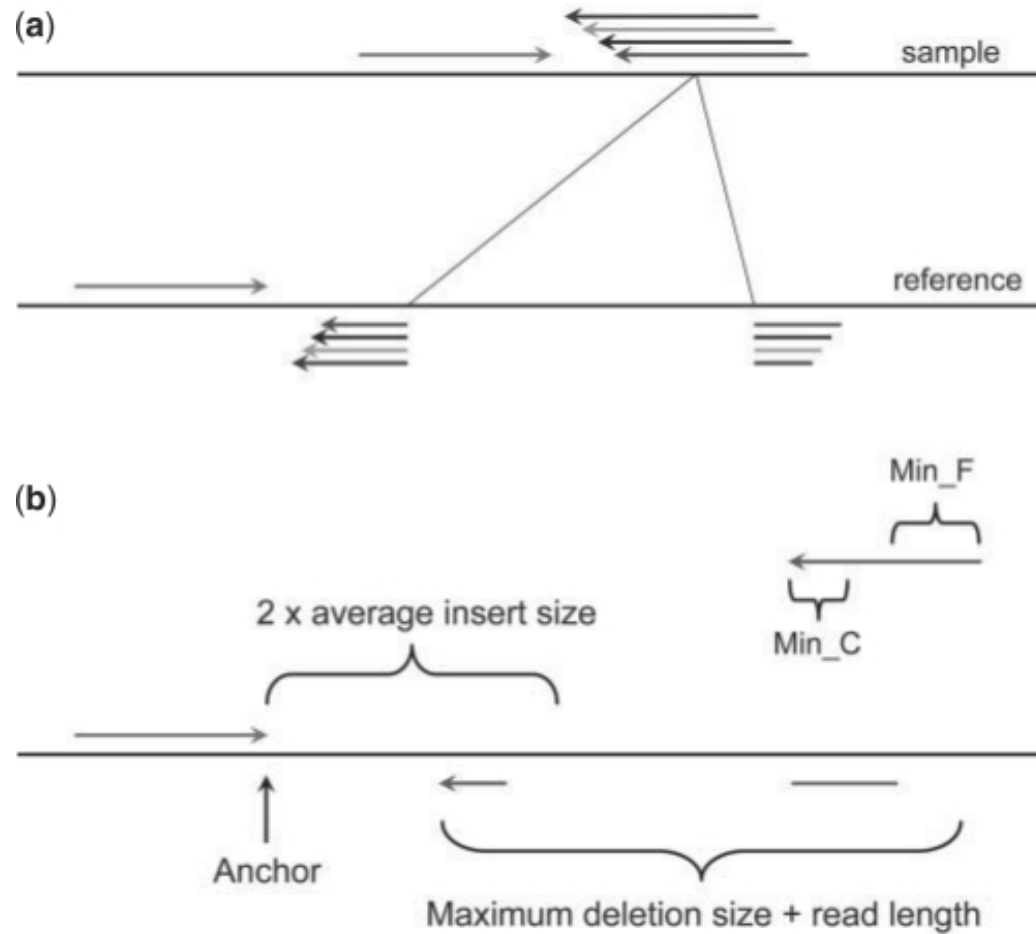
With reads of length 40-100 bps are we able to find the exact breakpoint of a structural variation?

Yes – using anchored split-read mapping



Mappable read mate provides anchor to narrow down search space

The Pindel algorithm (Deletions)



The Pindel algorithm (Real Data)

```

93      D 2299  NT 0 "" ChrID 20      BP 412719      415019 BP_range 412719 415024 Supports 3      + 3      - 0      S1 4      S2 118.634      SUM_MS 99      NumSupSamples 2 COLO-829 1      COLO-829-BL 2
AGGGGGTTGGCAGCAGCAATCACTGGTGAGCCTGCCAGAGGTCCAAGGGGTGAAGGGTTGGCACTGAAGAAATCcttga<2289>actgaCCTGACATCATATTGGCGTGCTGACGGGCGTACTGCCCCCTGGCATGCTAGATGAACC
      AAGGGGTGAAGGGGTGGCACTGAAGAAATC      CCTGACCTCACATTGGCGTGCTGACGGGTGACTGCCCCCTGGC
      GGGGTGAAGGGGTGGCACTGAAGAAATC      CCTGACATCACATTGGCGTGCTGACGGGTGACTGCCCCCTGGCA
GCCAGAGGTCCAAGGGGTGAAGGGTTGGCACTGAAGAAATC      CCTGACATCACATTGGCGTGCTGACGGGTGTA

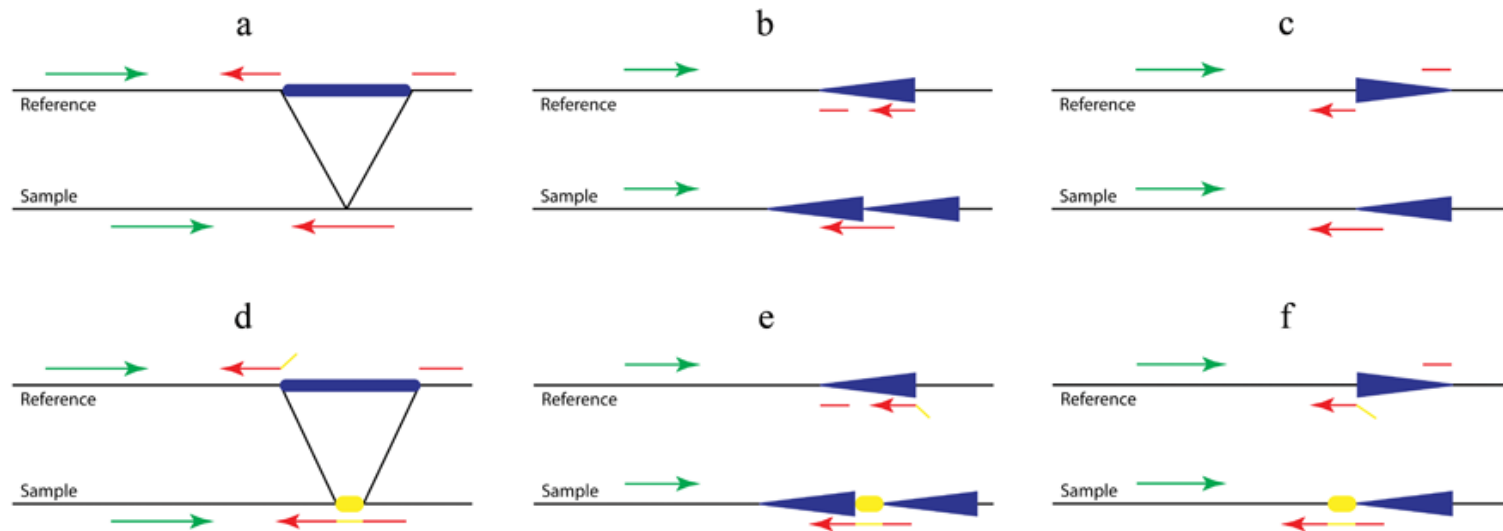
```

```

257     D 1671  NT 0 "" ChrID 20      BP 1337143      1338815 BP_range 1337143      1338818 Supports 11      + 7      - 4      S1 40      S2 434.993      SUM_MS 359      NumSupSamples 2 COLO-829 3
COLO-829-BL 8
GTGGTTAAGCCTTGCCTGTAGCCTAGGTAAGTAAAAATGTGCAAGGTTGTCAGGGCACCAGCATGGAGGAGTTCCCCCTacaad<1661>tcagaACAGTCTTTCCAATACCTATGTGGTCTCTGGAACAGACATTTCCATCCAGTAGCCATTCC
      ACCAGCATGGAGAAGTTCCCCCT      ACAGTCTTTCCAATACCTATGCGGTCTCTGGAACAGACATTTCCATCCAGTACC      -      1339193 25      COLO-829-B
      GTTGTCAAGGCACCAGCATGGAGGAGTTCCCCCT      ACAGTCTTTCCAATAACTATGTGGTCTCTGGAACAGACATT      -      1339198 37      COLO-829-BL
      TAGTCCCCCT      ACAGTCTTTCCAATACCTATTTGGTCTCTGGAACAGACATTTCCATCCAGTAGCCATTCCCTTTCCA      +      1337013 37      COLO-829-BL
      GCATGGAGGAGTTCCCCCT      ACAGTCTTTACAATACCTATGTGGTCTCTGGAACAGACATTTCCATCAAGTAGCCATT      -      1339178 25      COLO-829-BL
      GCATCAGCATGGAGGAGTTCCCCCT      ACAGTCTTTCCAATACCTATGTGGTCTCTGGAACAGACATTTCCATCCAGTA      +      1336784 37      COLO-829
      TCAGGTTACCAGCATGGAGGAGTTCCCCCT      ACAGTCTTTCCAATTCCTATGTGGTCTCTGGAACAGACATTTCCATC      +      1336797 37      COLO-829-BL
      AATGTGCAAGGTTGTCAGGGCACCAGCATGGAGGAGTTCCCCCT      ACAGTCTTTCCAATACCTATGTGGTCTCTGGA      +      1336638 37      COLO-829
      AAAAAATGTGCAAGGTTGTCAGGGCACCAGCATGGAGGAGTTCCCCCT      ACAGTCTTTCCAATACCTATGTGGTCTCT      +      1336657 37      COLO-829-BL
      TGCAAGGTTGTCAGGGCACCAGCATGGAGGAGTTCCCCCT      ACAGTCTTTCCAATACTTATGTGGTCTCTGGAACAG      -      1339117 25      COLO-829-BL
      GTTGTCAAGGCACCATCATGGAGGAGTTCCCCCT      ACAGTCTTTCCAATACCTATGTGGTCTCTGGAACAGACATTT      +      1336754 37      COLO-829
      CTAGGAAAAAATGTGCAAGGTTGTCAGGGCACCAGCATGGAGGAGTTCCCCCT      ACAGTCTTTCCAATACCTATGTG      +      1336766 25      COLO-829-BL

```

The Pindel algorithm for complex variants



- a) large deletion
- b) tandem duplication
- c) inversion
- d-f) same as a-c with non-template sequence (yellow part)

Acknowledgements

- Tobias Rausch (EMBL)
- Kai Ye (Leiden University Medical Center)
- Anne-Katrin Emde (Freie Universität Berlin)