

# Fusion transcript detection in rare genetic disease



**A. Collin Osborne**

*Senior Bioinformatician*

*Mayo Clinic Quantitative Health Sciences*

*Mayo Clinic Center for Individualized Medicine*

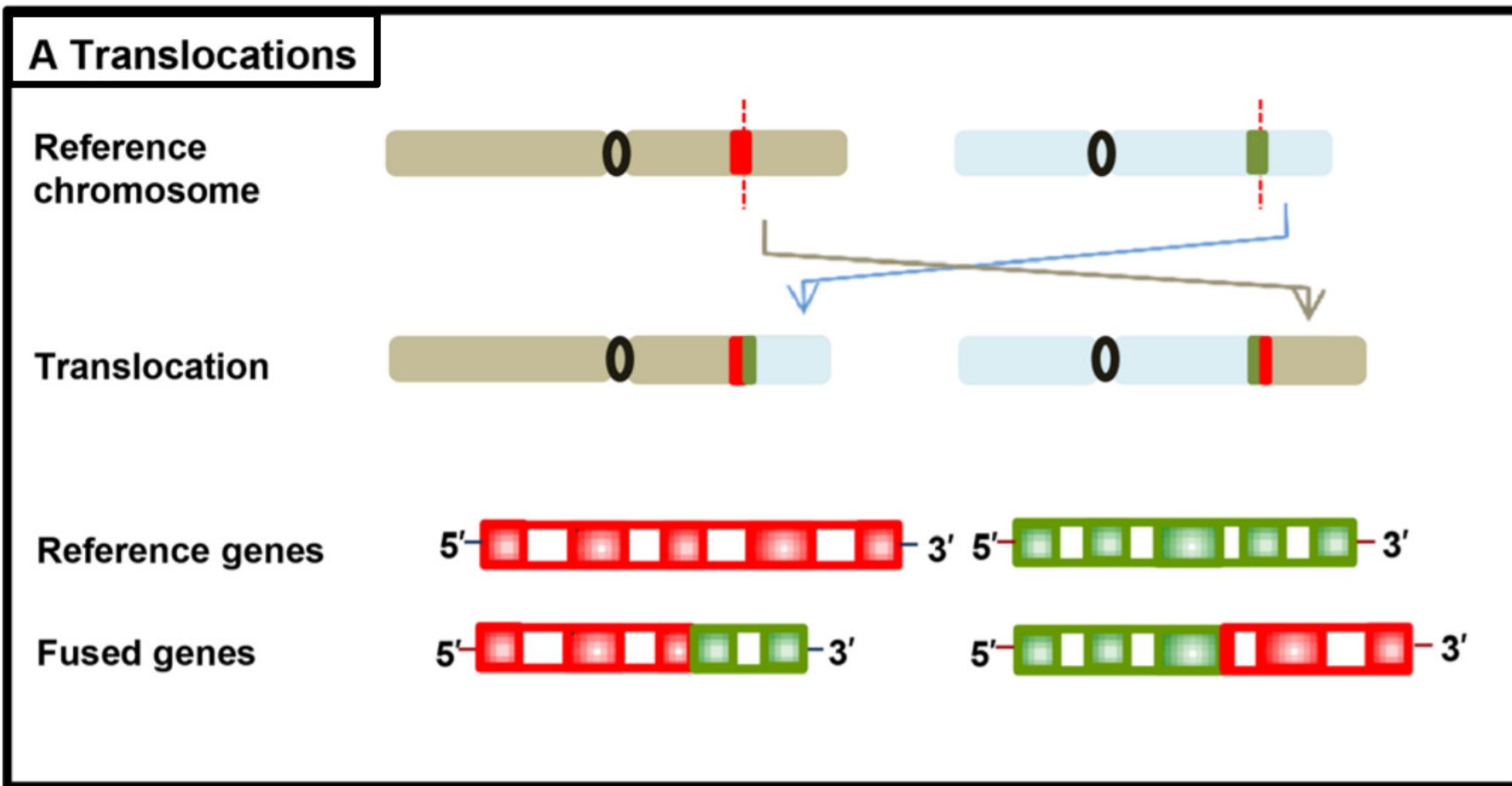
Authored by: **Gavin Oliver**

# A working definition



- Fusion transcription involves the aberrant conjoining and expression of normally discrete genic material
- Therefore a fusion can be considered *“Aberrantly conjoined and expressed genic material that exists separately under normal conditions”*
- More simply: *pieces of multiple genes are expressed as one*
- Caused by a variety of abnormalities at the DNA level as well as (debatably) at the RNA level

# Mechanisms of formation



# Mechanisms of formation



## B Insertions

Reference chromosome



Inserted and deleted chromosome



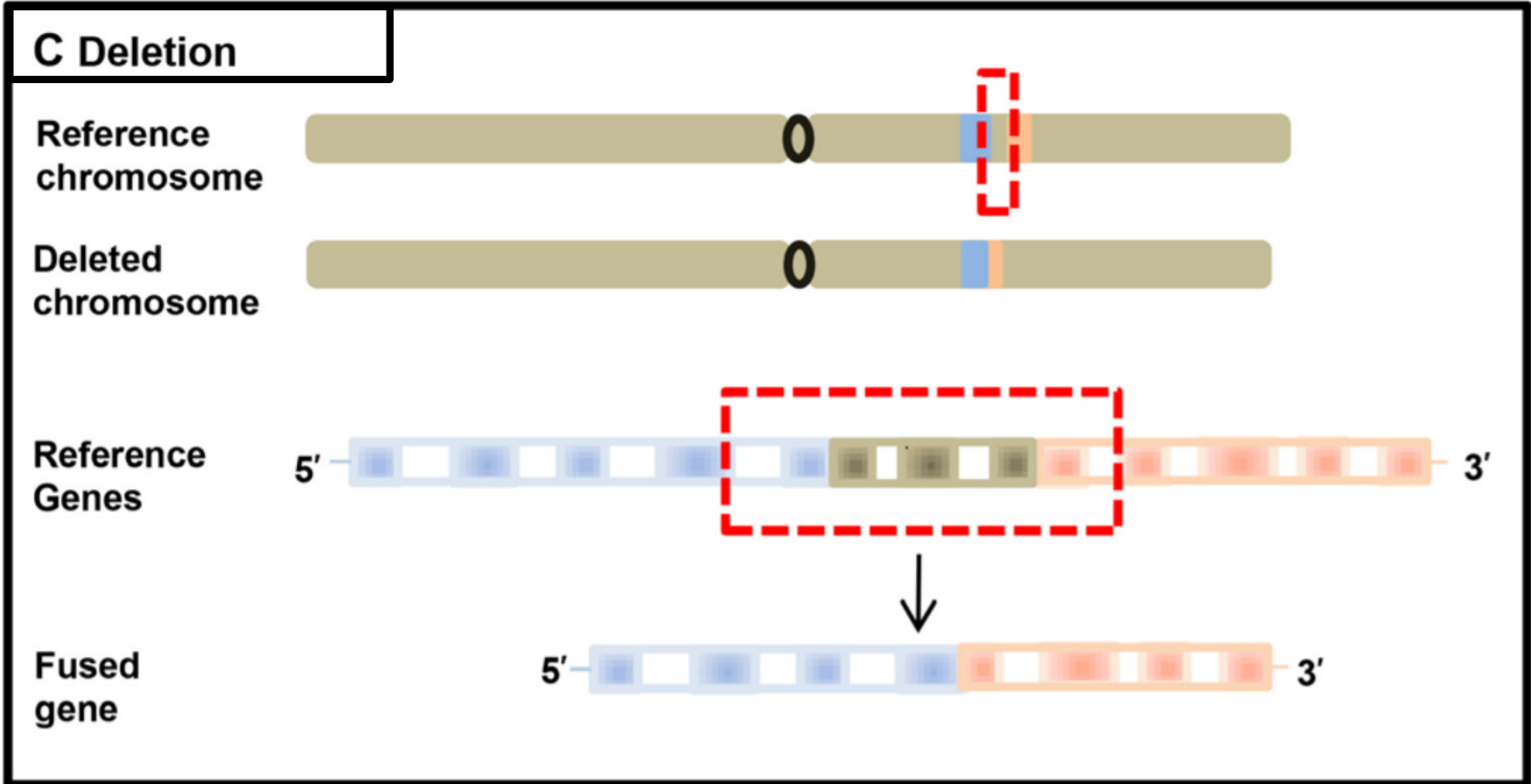
Reference genes



Fusion genes



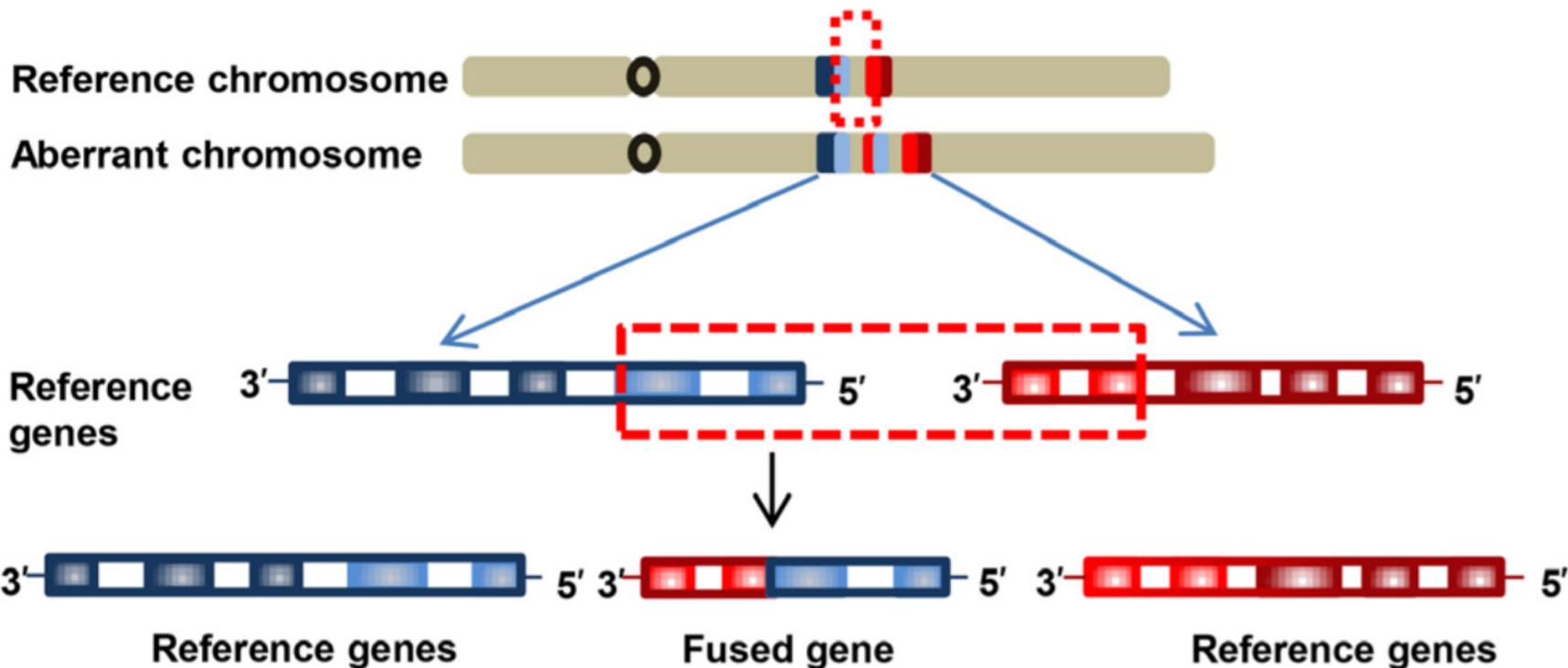
# Mechanisms of formation



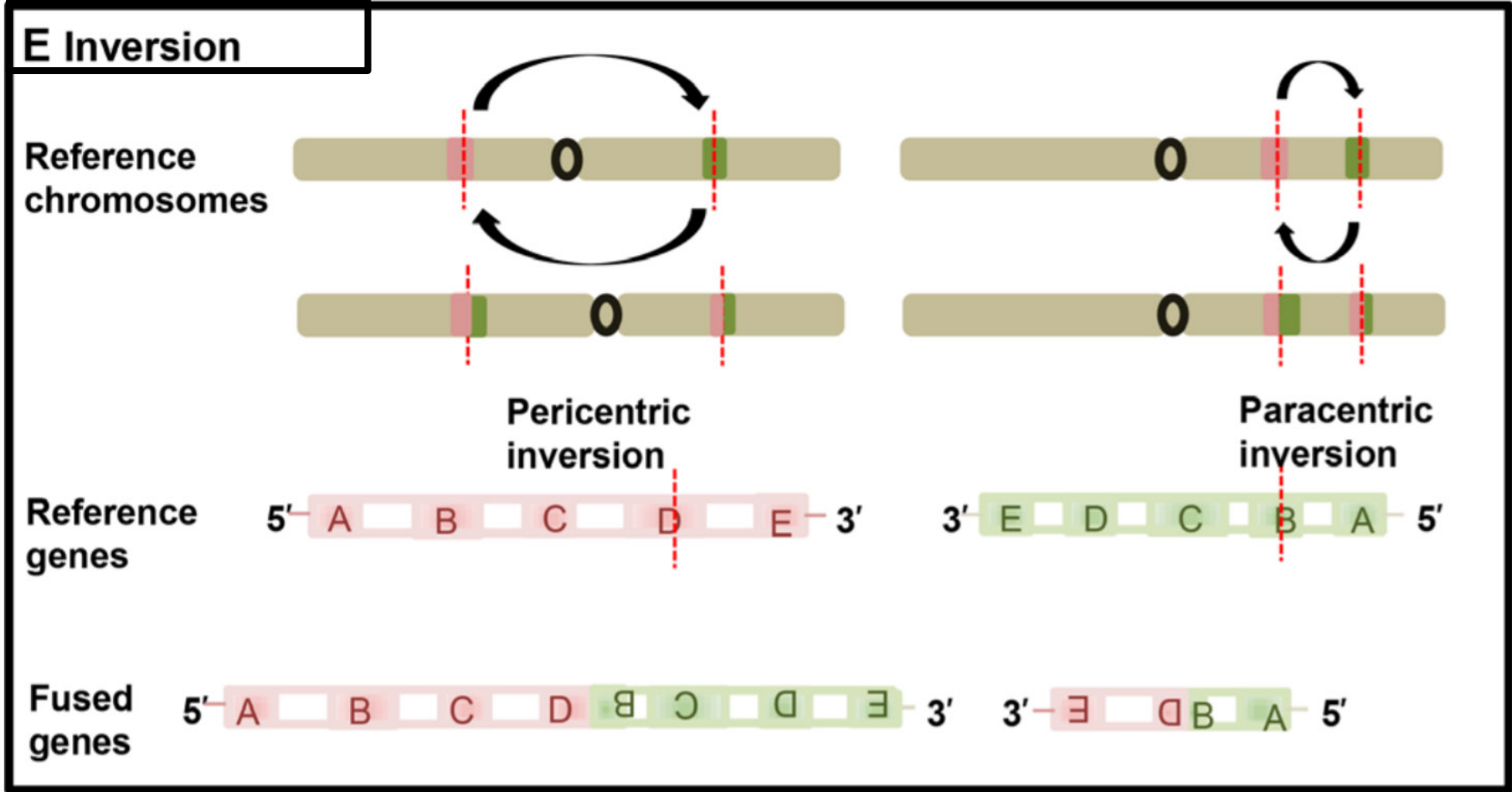
# Mechanisms of formation



## D Tandem duplications



# Mechanisms of formation



# Mechanisms of formation



## F Chromothripsis

Reference chromosome/genes



Catastrophic event  
(Chromosome shattering)



Rearranged chromosome/genes

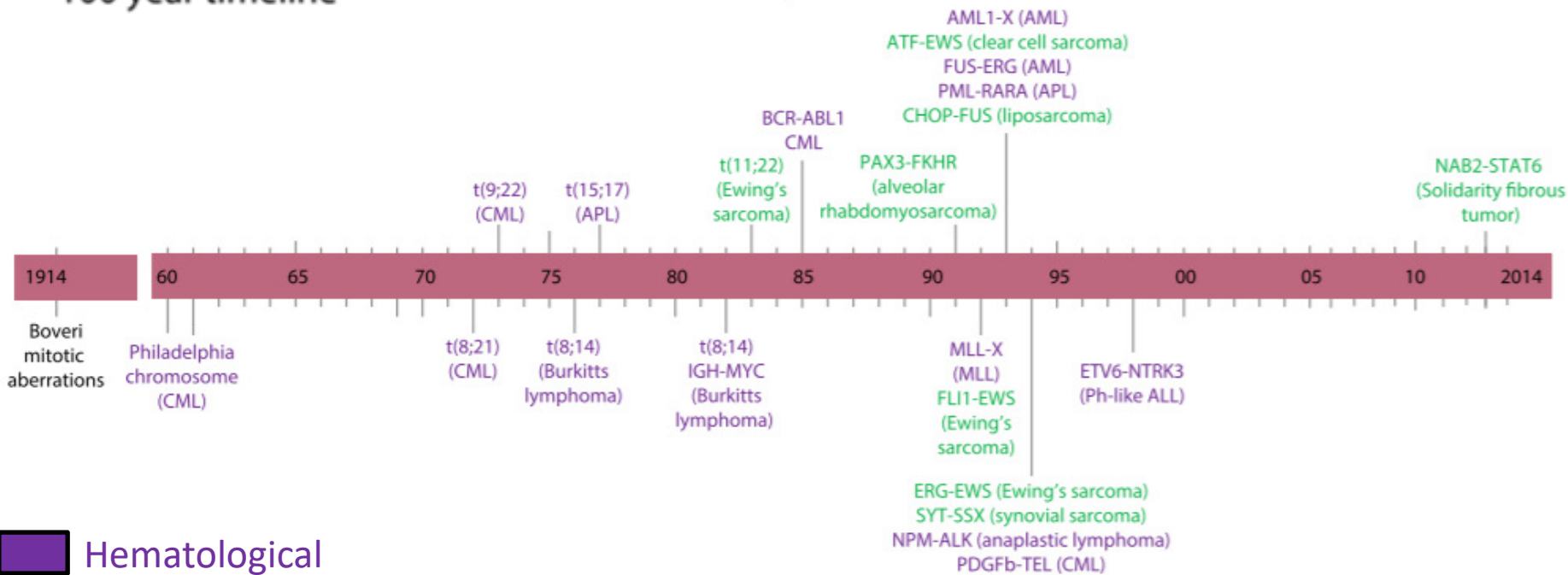


Lost genomic regions



# An oncogenic phenomenon?

## Gene fusions in cancer 100 year timeline



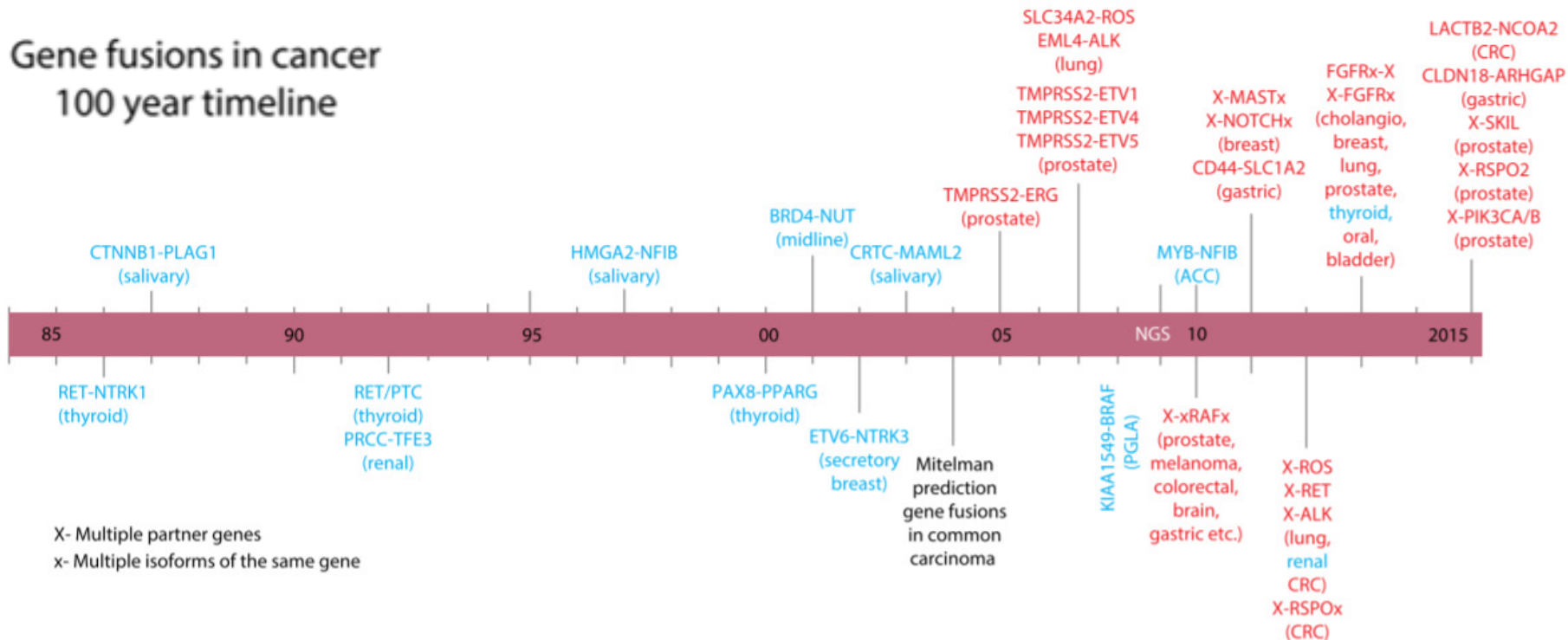
 Hematological

 Soft tissue

# An oncogenic phenomenon?



## Gene fusions in cancer 100 year timeline



Rare epithelial

Common epithelial

- Commonly
  - involve fusion of a downstream kinase
  - or transcription factor
    - with a more highly expressed upstream gene
      - leading to increased expression of the downstream gene or a functional component of it
- Protein formation dependent on in-frame translation

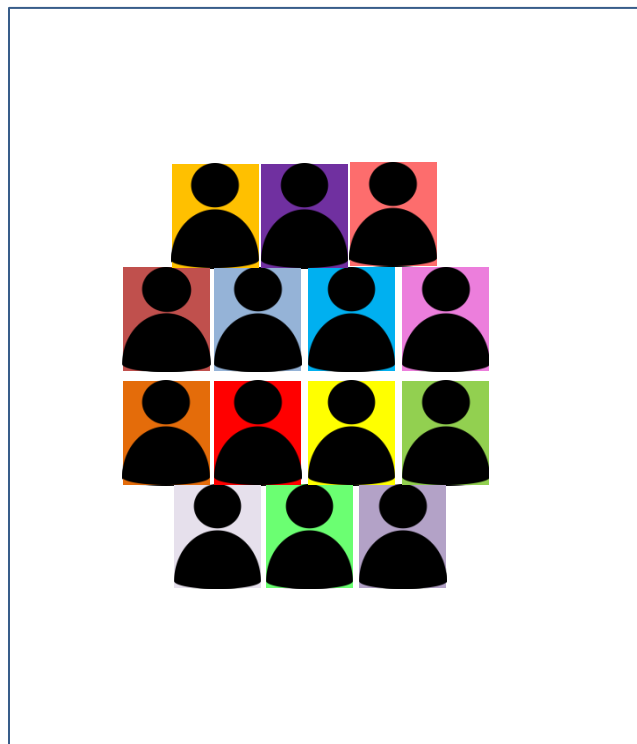
- Most reported gene fusions pertain to gain-of function aberrations imparting neoplastic phenotypes
- Loss of function of tumor suppressors such as TP53 and PTEN have also been identified
- Fusion transcripts are recognized as having diagnostic, prognostic and therapeutic (druggable) relevance in oncology
- Detection of gene fusions is increasingly incorporated into the standard workflow for genomic characterization of tumors in both research and clinical settings

# Fusions in inherited disease



- 18-40% unsolved cases are solved by exome sequencing
- RNA-Seq has recently been proposed as a supplementary diagnostic tool
- Cummings *et al.* achieved a 35% diagnostic increase by profiling aberrant splicing and allele specific expression
- Kremer *et al.* added gene expression quantification to the testing repertoire and demonstrated a 10% increase
- Isolated reports exist in the literature of fusion transcripts being detected in cases of brain malformation, intellectual disability, schizophrenia, ASD and more
- Fusion transcription had not been systematically profiled in inherited disease

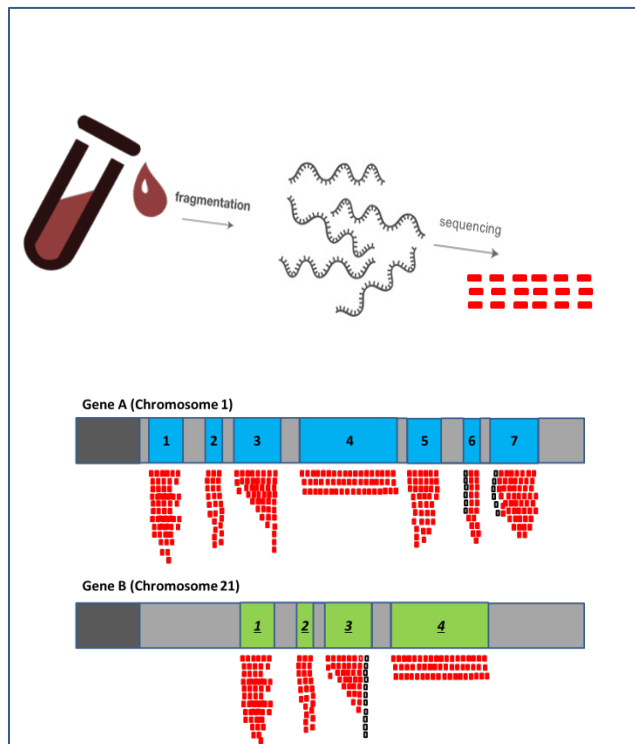
## Patient Cohort



- 47 patients
- Prior exome-sequencing
- 23 M, 24F
- Ages 9 months – 68 years ( median 11)
- Diverse phenotypes
  - Neurological
  - Muscular
  - Gastrointestinal
  - Skeletal
  - Connective tissue disorders



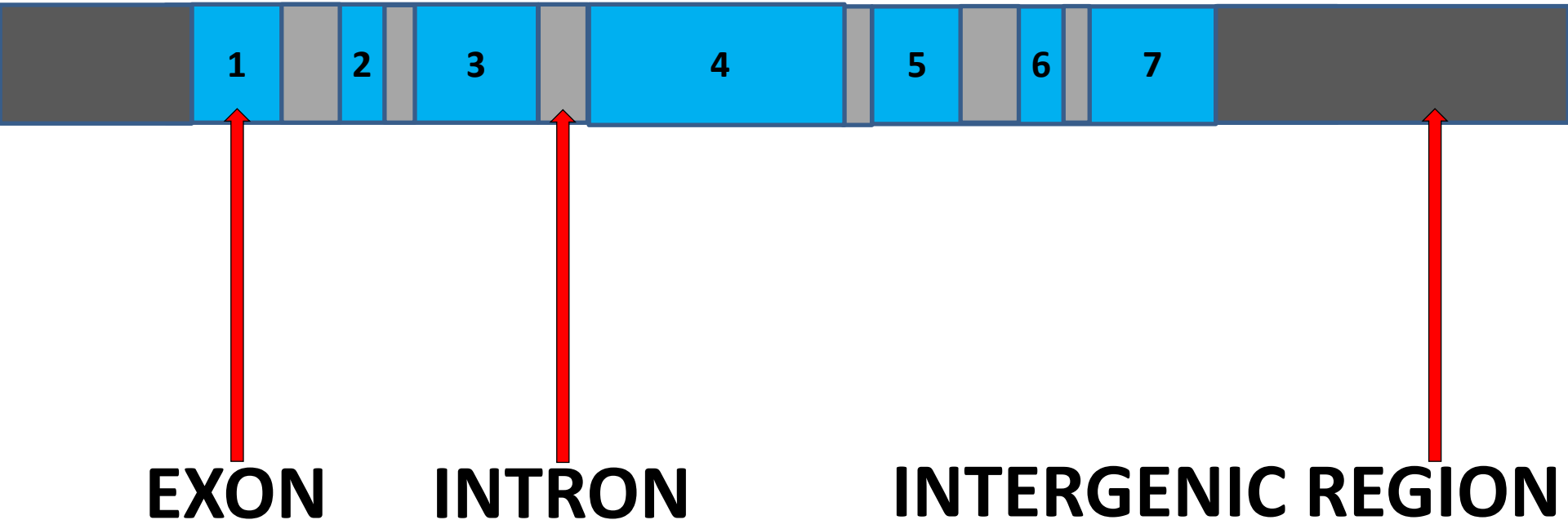
## RNA-Sequencing



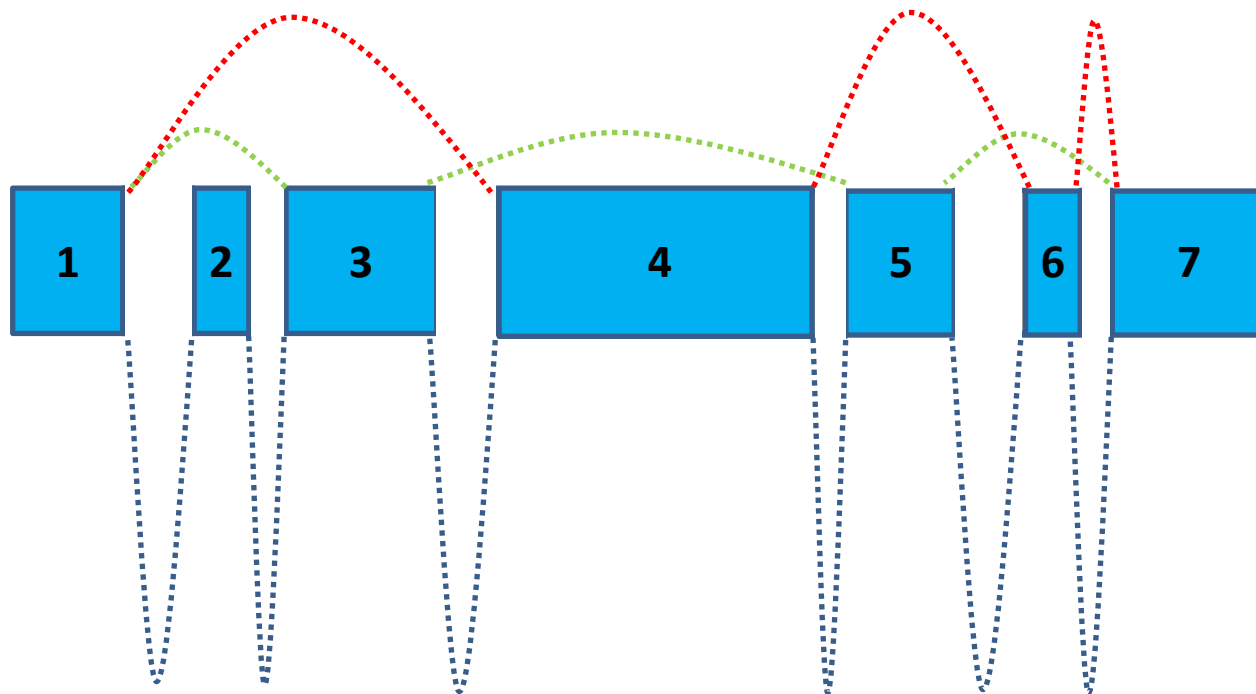
- Patient whole blood
- Illumina HiSeq 2500
- 200 million 100bp PE reads per sample

- Fusion detection increased diagnosis of rare disease
  - Two cases confirmed solved
    - SCID
    - Multiple exostoses
  - 4.3% increase in diagnostic yield
  - Experimentally validated existence of fusion events in disease-relevant genes with potential phenotypic relevance in five additional cases

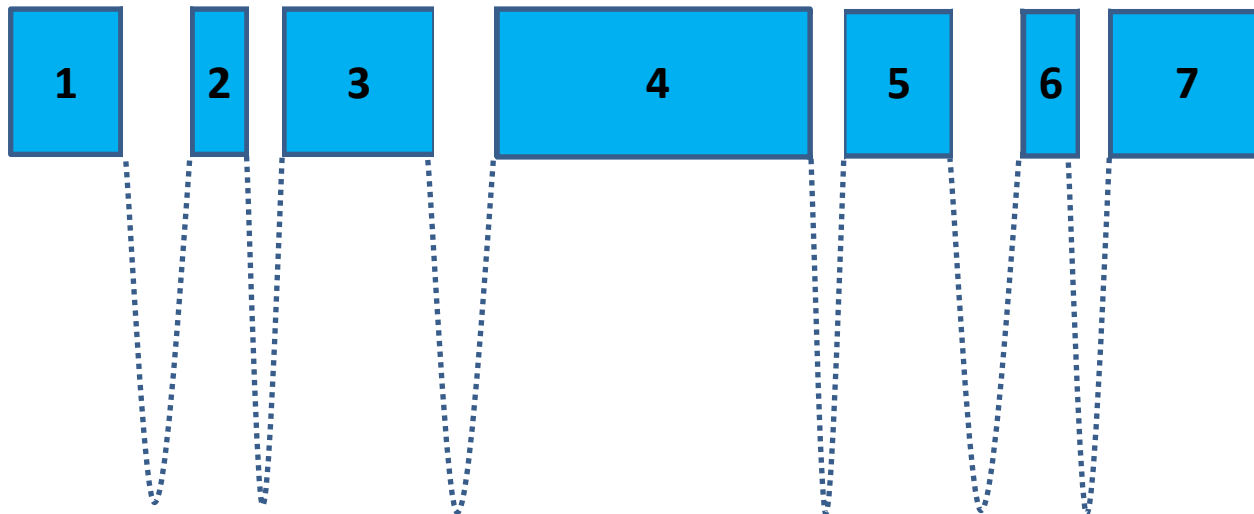
# Software solution overview



# Software solution overview



# Software solution overview



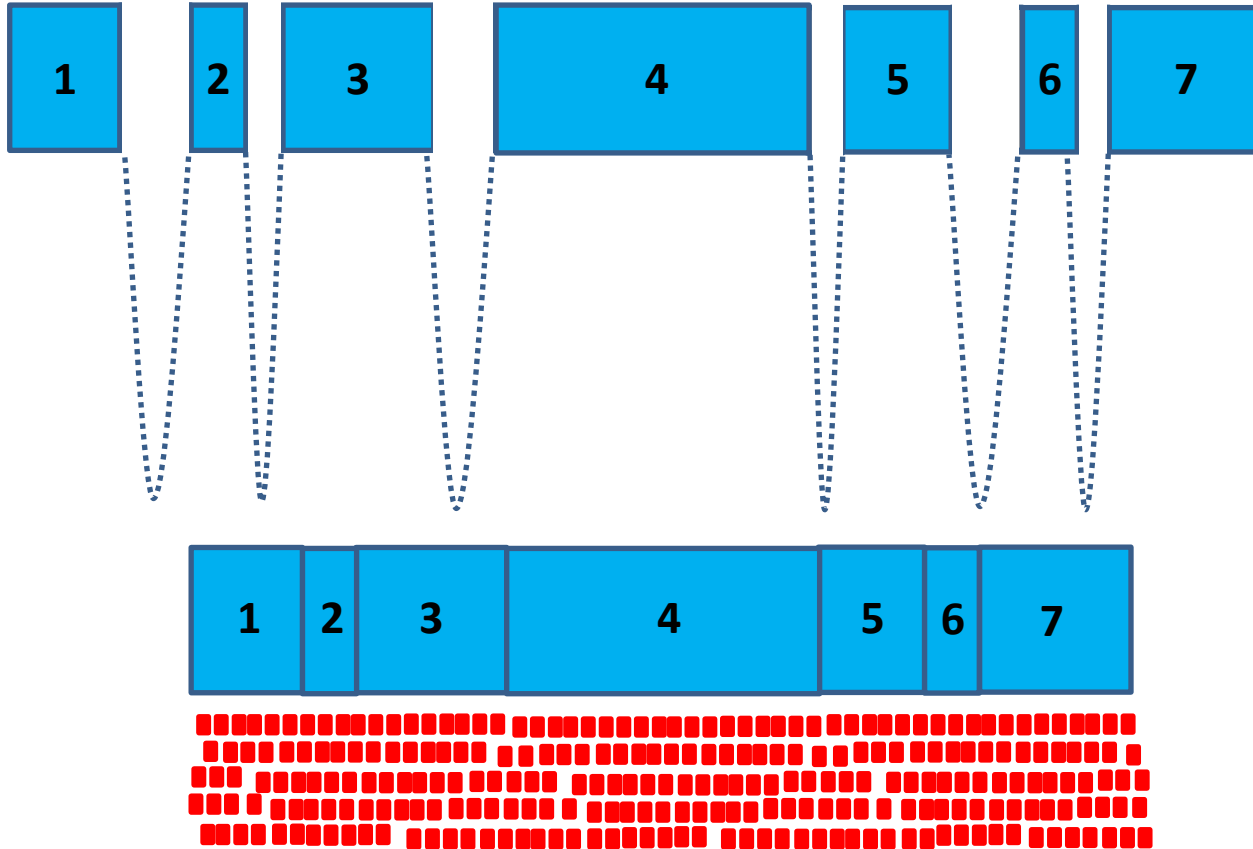
fragmentation



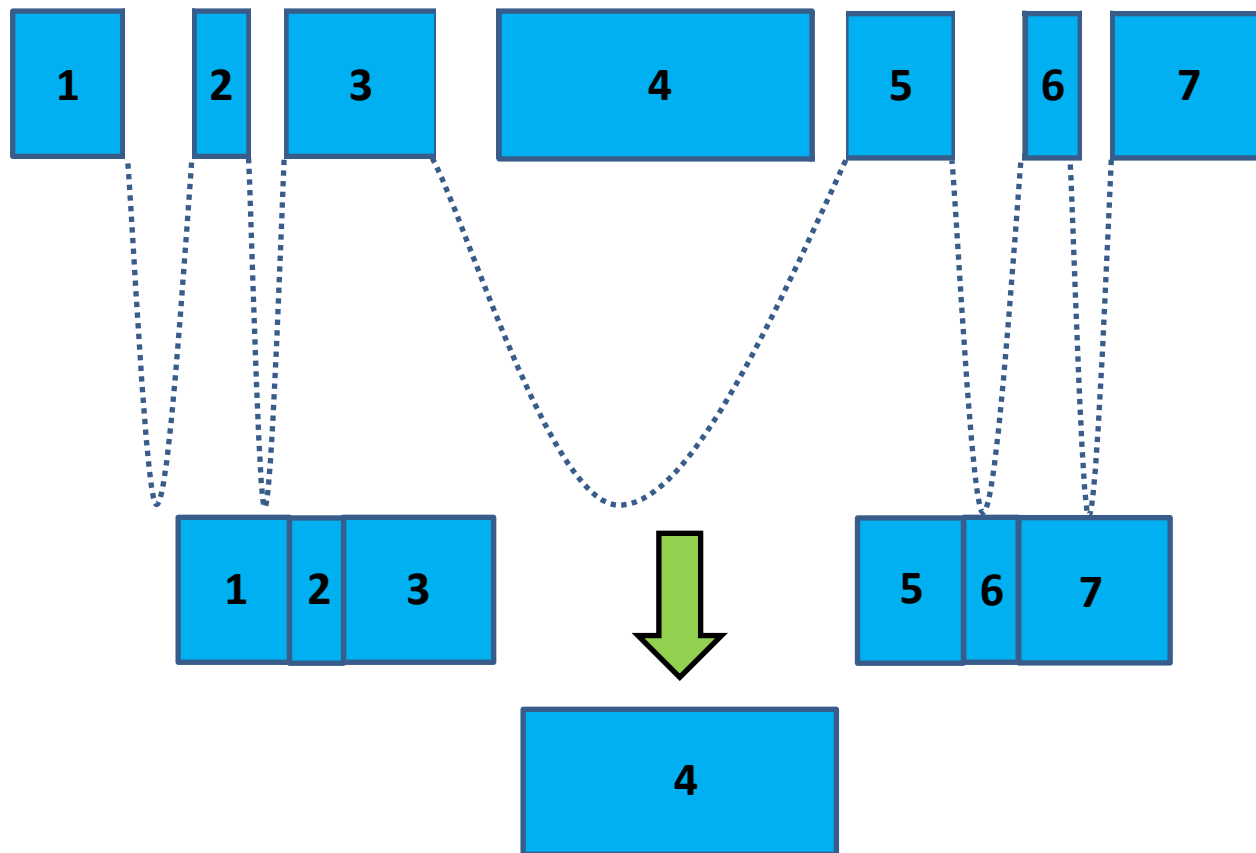
sequencing



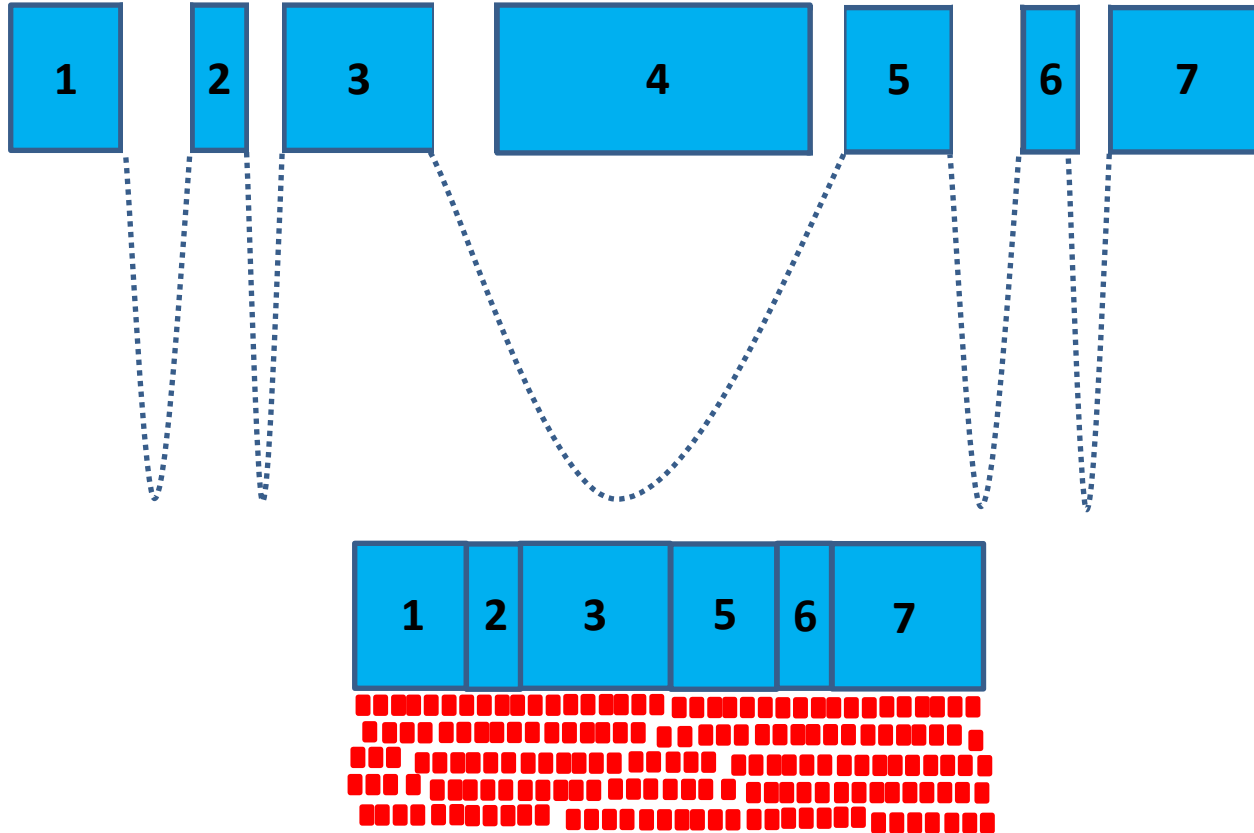
# Software solution overview



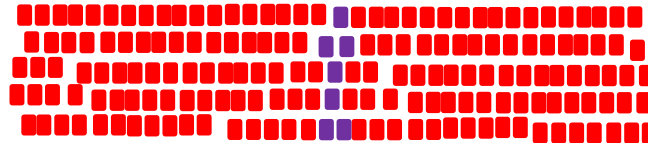
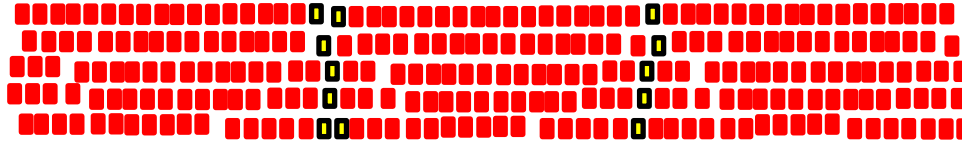
# Software solution overview



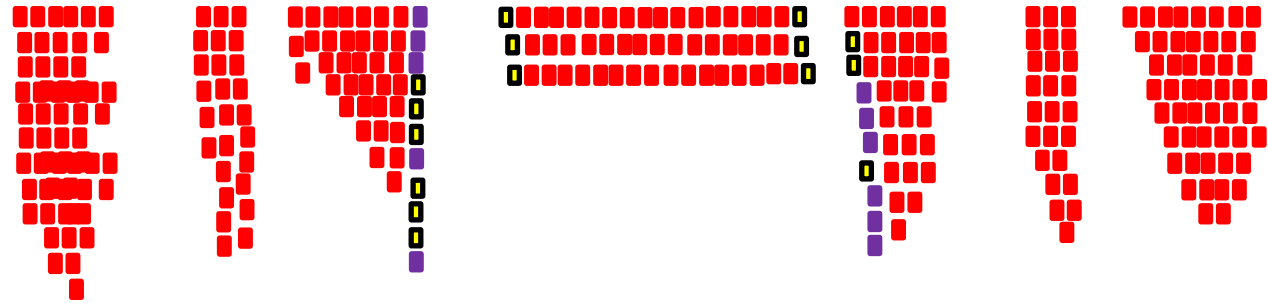
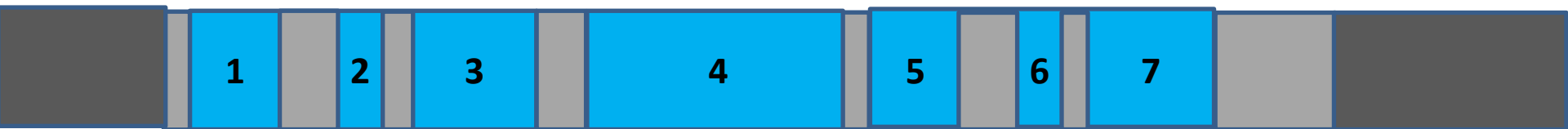
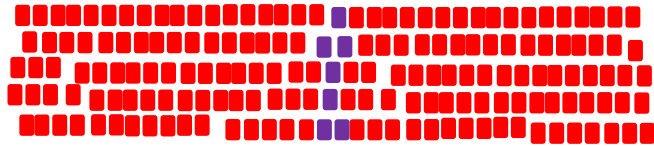
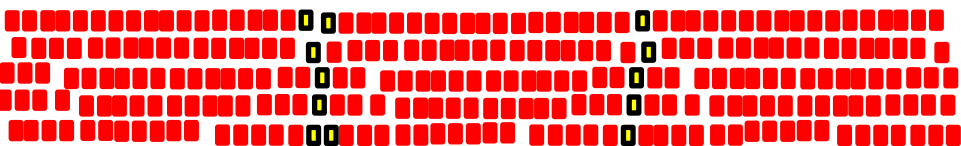
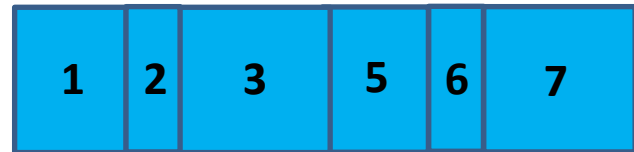
# Software solution overview



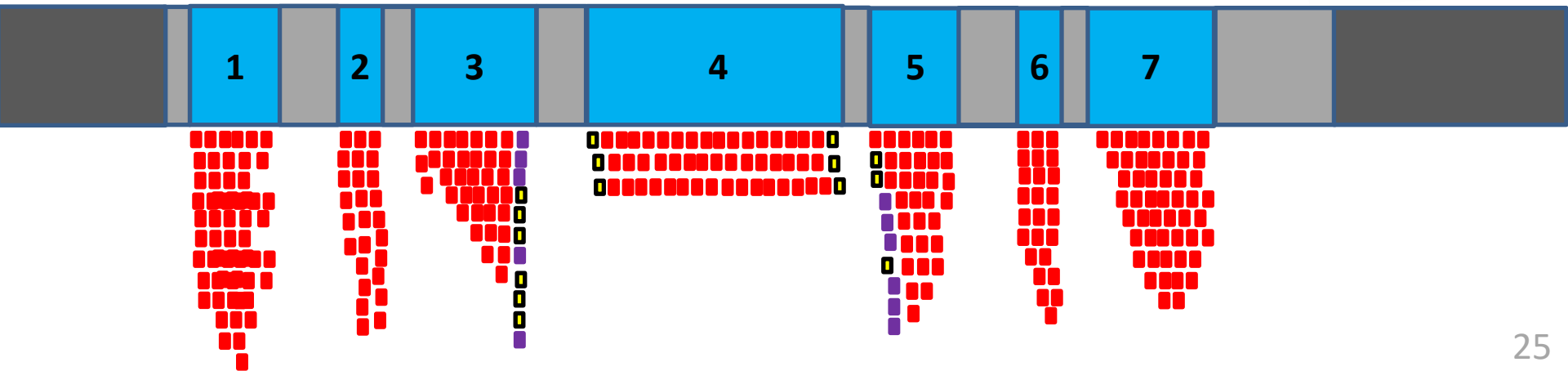
# Software solution overview



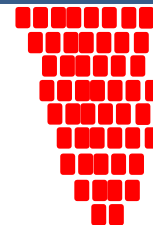
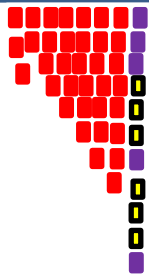
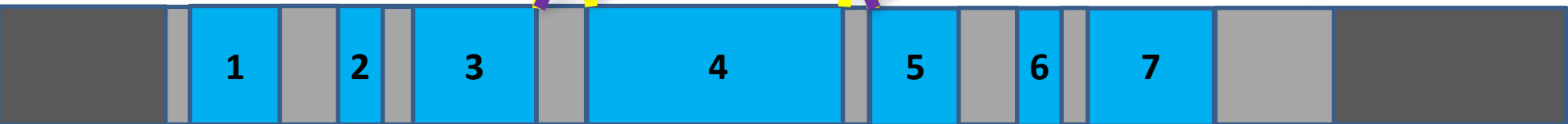
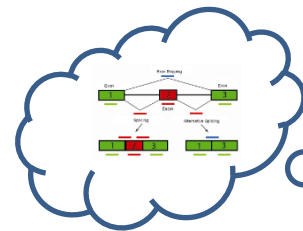
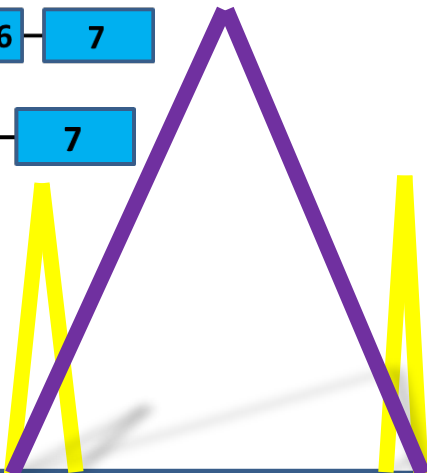
# Software solution overview



# Software solution overview



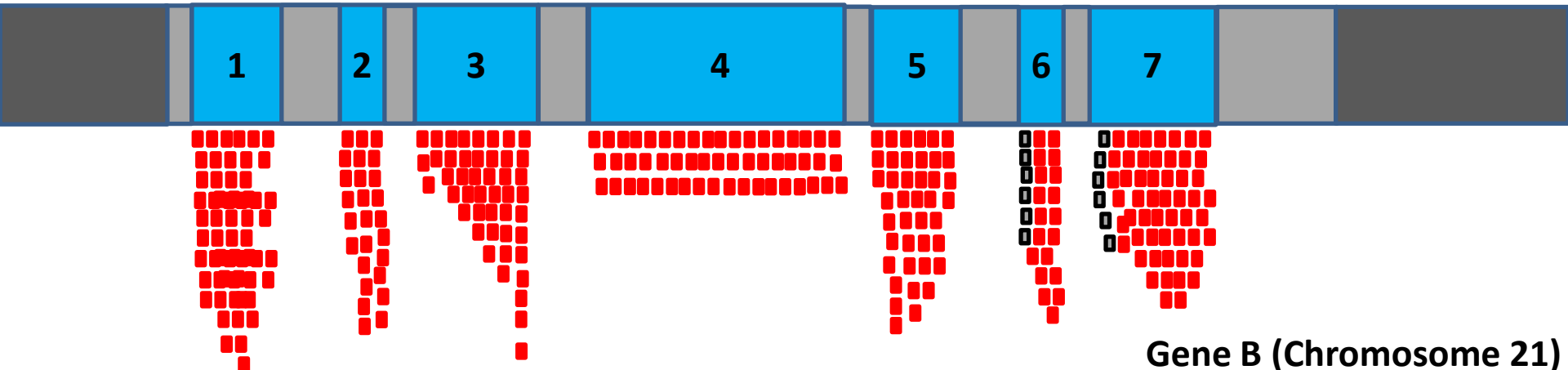
# Software solution overview



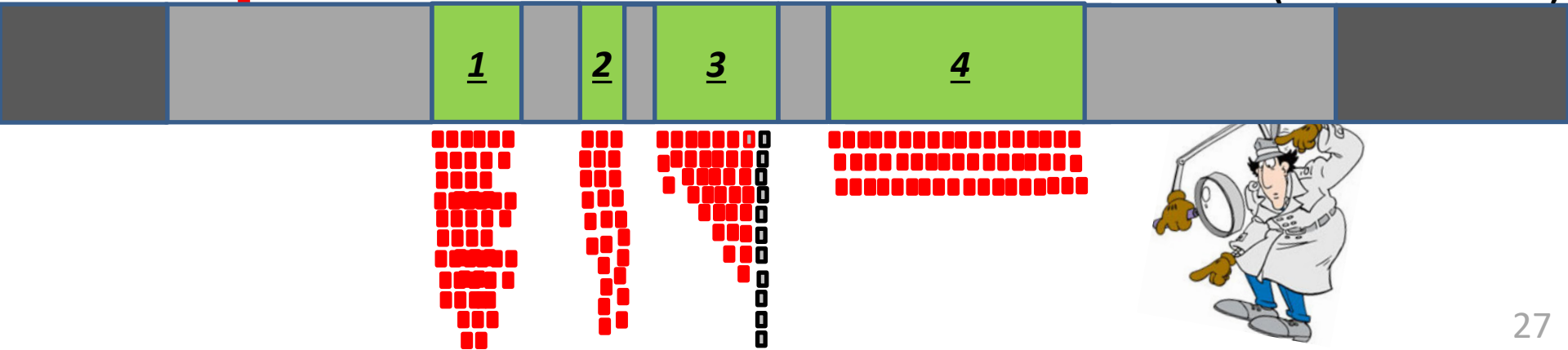
# Fusion transcripts



## Gene A (Chromosome 1)



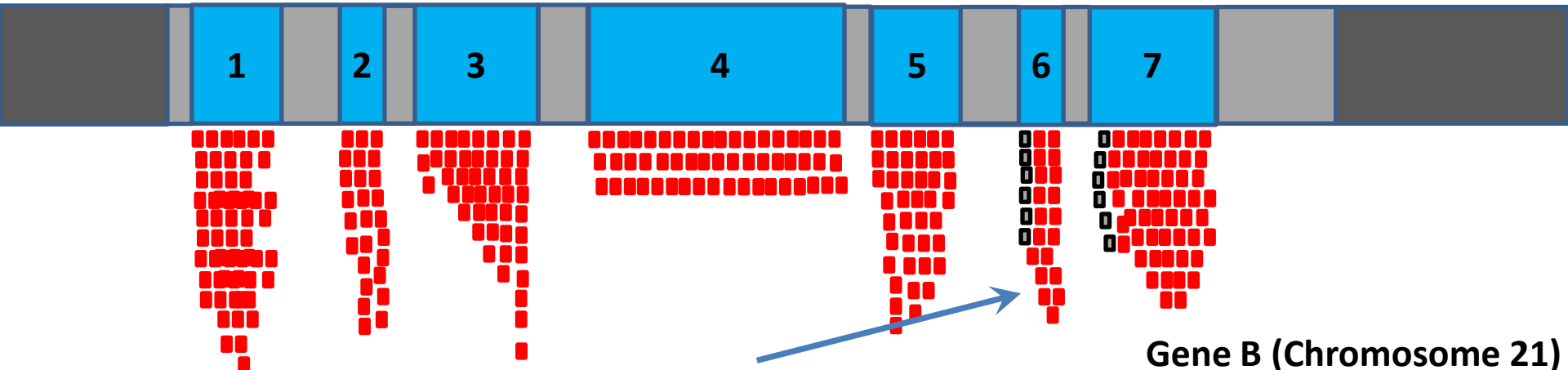
## Gene B (Chromosome 21)



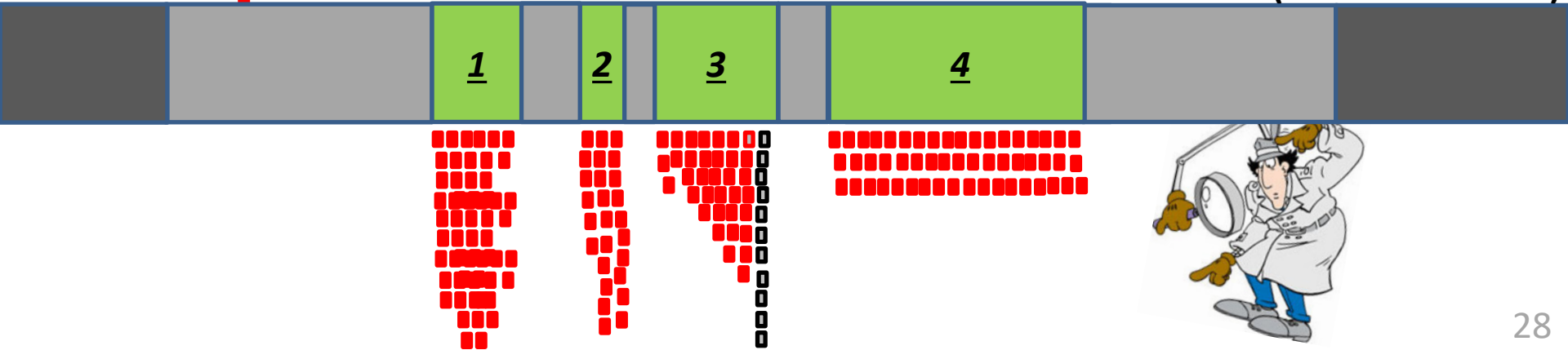
# Fusion transcripts



Gene A (Chromosome 1)



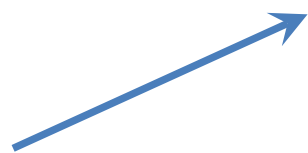
Gene B (Chromosome 21)



# Fusion transcripts



Gene A (Chromosome 1)



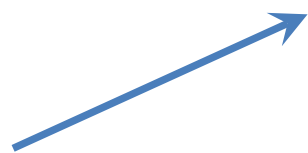
Gene B (Chromosome 21)



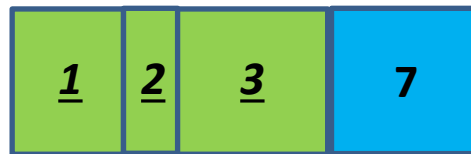
# Fusion transcripts



Gene A (Chromosome 1)



Gene B (Chromosome 21)





# Fusion calling challenges



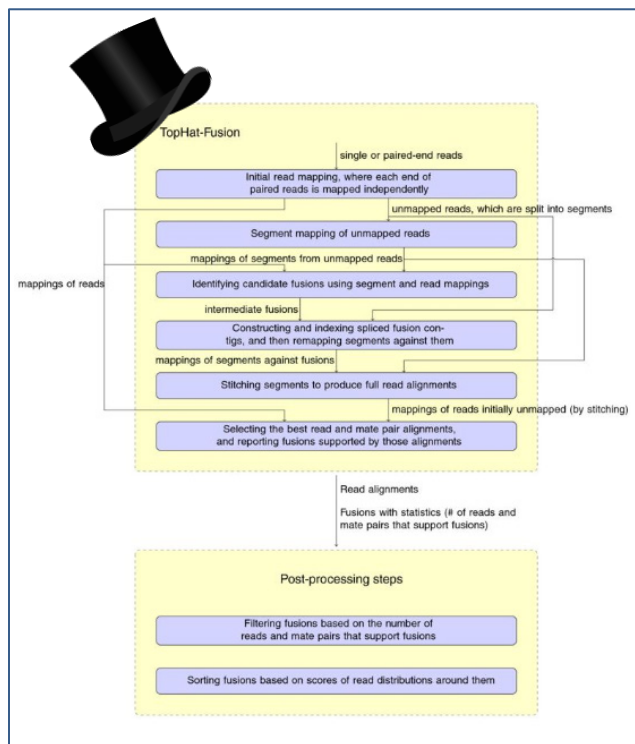
- Numerous software solutions exist for fusion detection
  - e.g. STAR-Fusion, Tophat-Fusion, PRADA, Fusioncatcher
- Technical comparisons demonstrate limited overlap and no caller is fully inclusive
  - Partially because FPs are abundant & outputs require filtering
  - Filters are trained using *in-silico*, tumor or cell-line data & performance falters on alternative data types
- It is recommended to select a caller on the basis of the data being profiled however none are trained on inherited disease

# Fusion calling challenges

- Any attempt to detect fusions in inherited disease thus requires:
  - Inherent sensitivity
  - A means of deprioritizing biologically and phenotypically unimportant fusion candidates



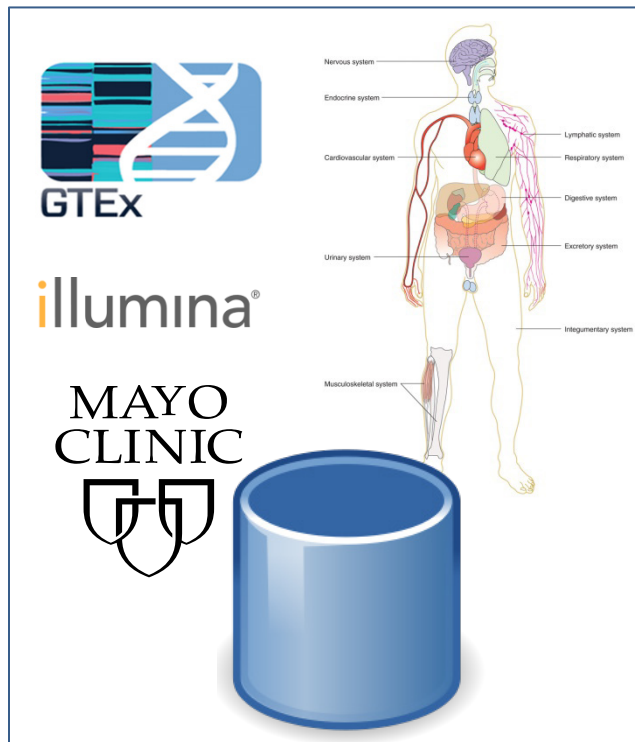
## Read support (basic)



- TopHat Fusion (Kim & Salzberg 2011)
  - Equally applicable to other callers
- Omitted all TopHat filtering steps (cancer cell-line derived)
- Employed a very minimal depth filter (2 reads)



## Normal DB comparison

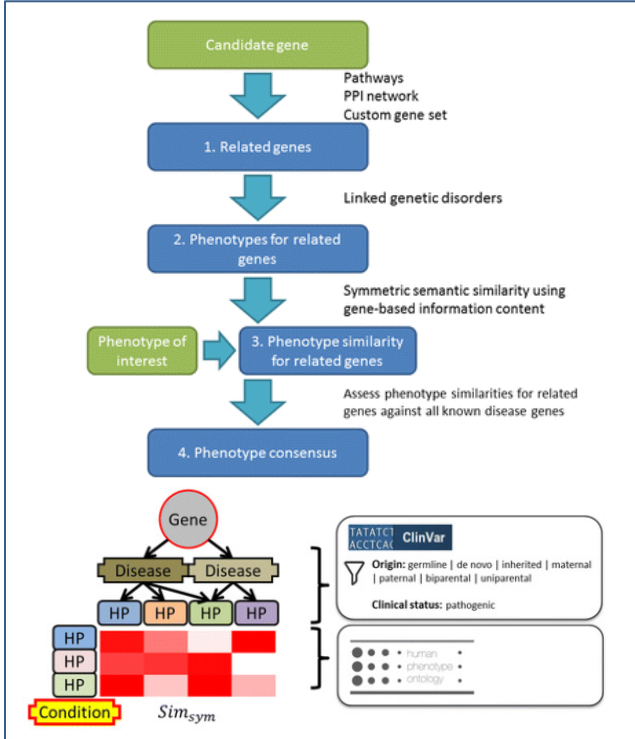


- Compared fusion candidates to a database of candidates from normal tissues
- Fusion calling on samples from GTEx, Illumina Human Bodymap, Mayo Clinic
- Approx. 800 samples, 30 tissues
- Any fusion candidates occurring in DB or more than one cohort sample were categorized as normal/recurrent



# Filtering / Prioritization

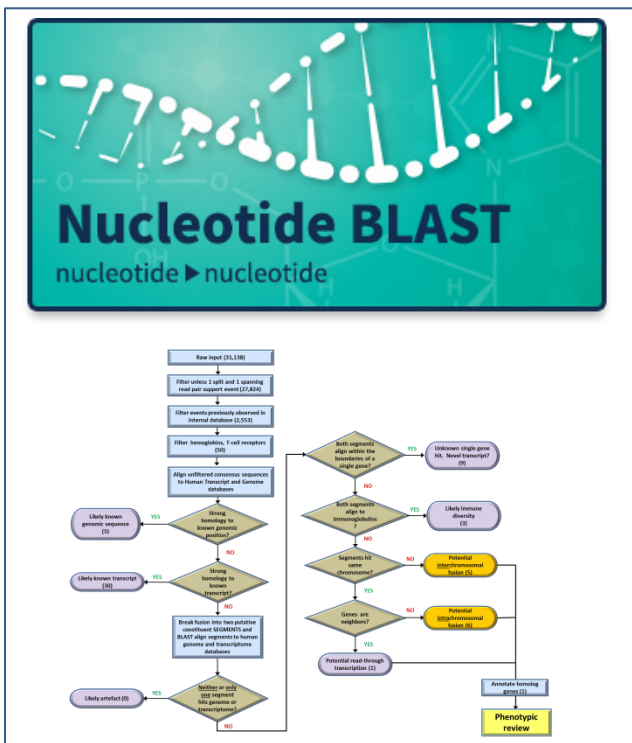
## Phenotypic Prioritization



- Dual approach
  - Manual (Literature, OMIM, Genecards)
  - *In-silico*
    - PCAN: phenotype consensus analysis to support disease-gene association (Godard & Page, 2016)
  
- Generated phenotypically prioritized events for follow-up validation

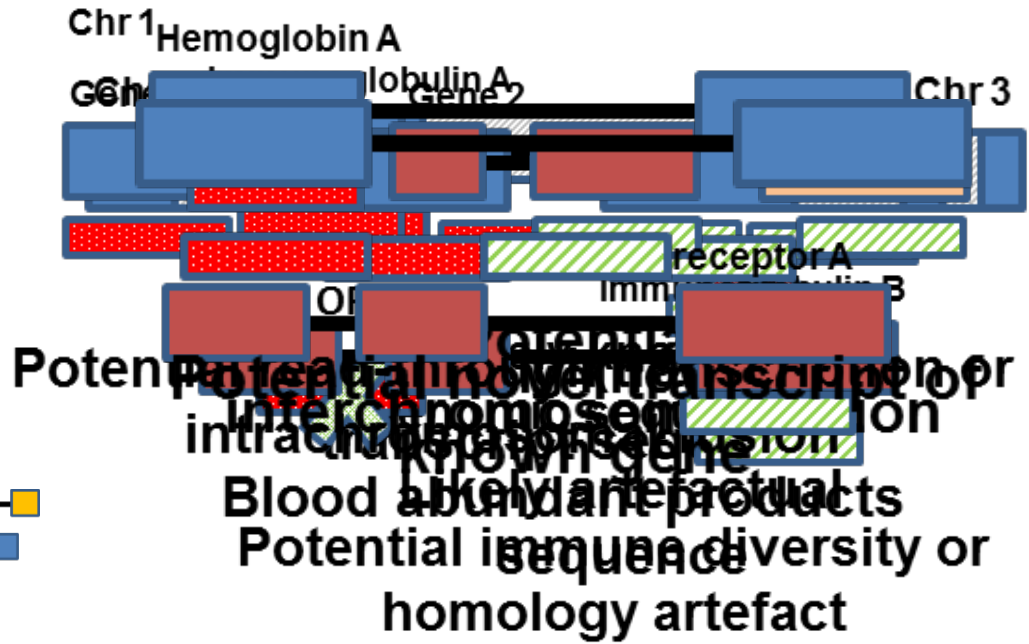
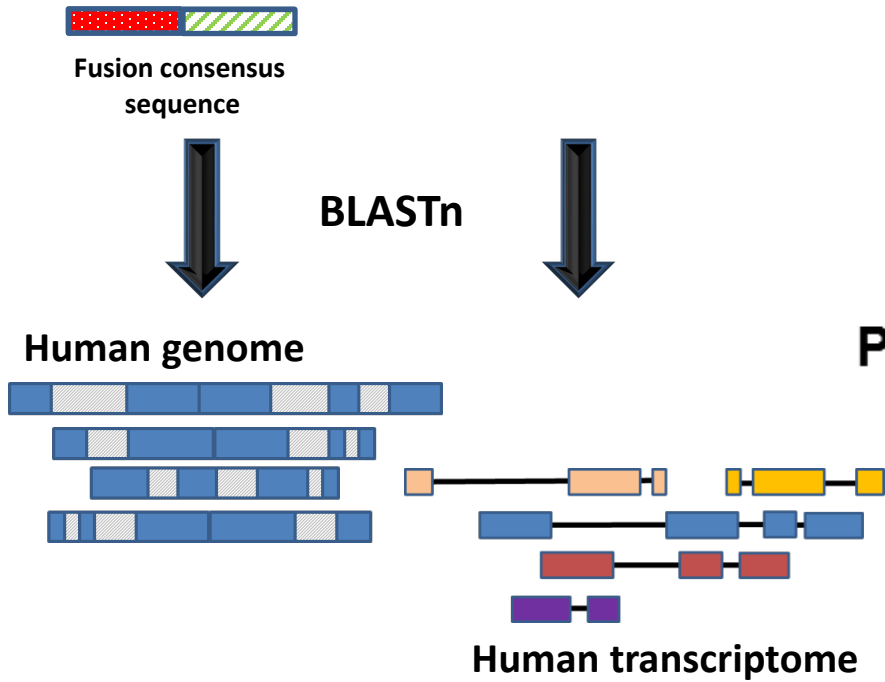


## BLAST categorization



- Fusion consensus sequences generated by TopHat Fusion used as input
  - Algorithm dependent
- Devised custom categorization pipeline based on BLASTn
- Categorization logic based on best alignments

# Candidate Categorization



# Now let's try it...



# Questions