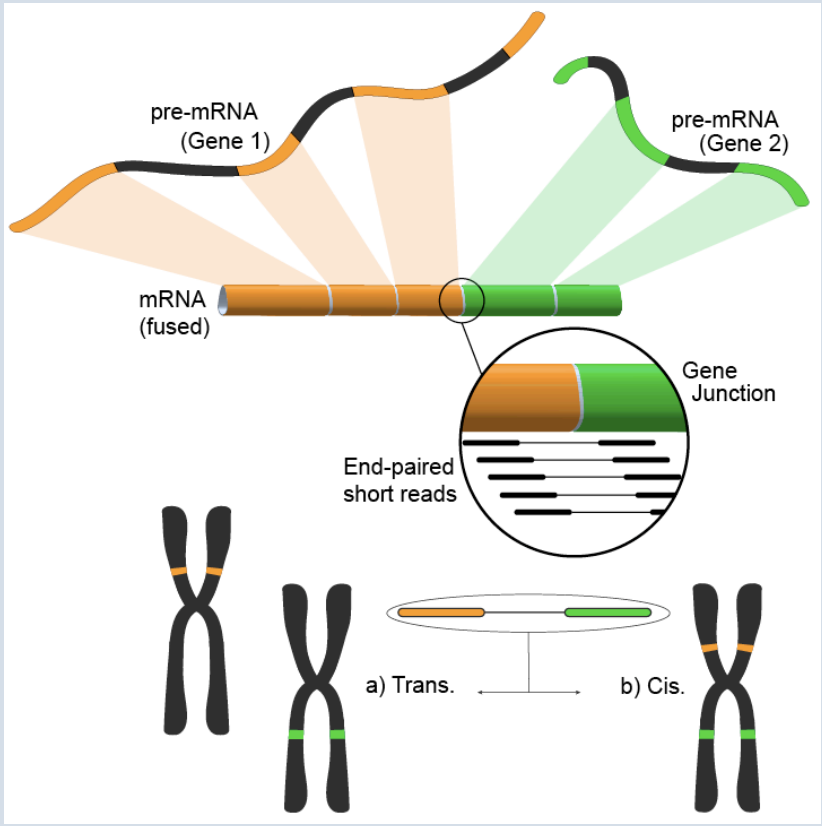
Broad Cancer Program Bootcamp: Fusion Transcript Discovery

Brian Haas
May 9, 2018



Brian:

bhaas@broadinstitute.org

https://github.com/NCIP/Trinity_CTAT/wiki

(Liberal use of material from Andrew McPherson: andrew.mcpherson@gmail.com

http://compbio.bccrc.ca

http://compbio.cs.sfu.ca

Indicated w/ awm)

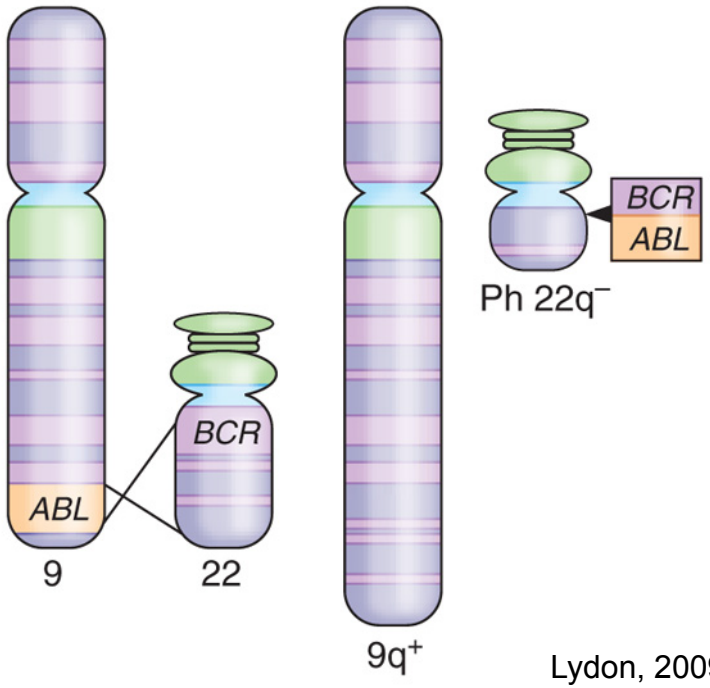
Learning Objectives of Module

- Explore impact of gene fusions in cancer
- Learn about types of evidence for gene fusions
- Understanding of the available detection methods/tools
- Identify common sources of false positives
- Assess a gene fusion's potential function

awm

Definition of a Gene Fusion

- Novel gene formed by fusion of two distinct wild type genes
- In Cancer: produced by somatic genome rearrangements



Lydon, 2009

awm

Diagnostics and Therapeutics Involving Oncogenic Fusion Transcripts in Cancer

- BCR-ABL1 (Philadelphia chromosome)
  - Chronic Mylogenous Leukemia (CML) cases (95% of cases)
  - Treatable with tyrosine kinase inhibitors
- TMPRSS2-ERG
  - prostate cancers (50% of cases)
- EML4-ALK
  - Non small cell lung carcinoma (4% of cases)
  - anaplastic lymphoma kinase (ALK) inhibitors improve patient outcome
- DNAJB1-PRKACA
  - fibrolamellar hepatocellular carcinoma (FL-HCC), 100% of cases, but a rare cancer.
- FGFR3-TACC3
  - found in 8.3% of glioblastoma patients

Evidence Gene Fusions are Initiators of Carcinogenesis

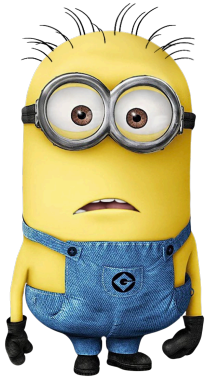
- Correlate with cancer phenotype
- Successful treatment reduces/eradicates fusion products
- Gene fusions produce neoplastic disorders in mouse models
- Silencing fusion transcripts reverses tumorigenesis

awm

How Can Fusions Drive Cancer?

- Activate Tumor Oncogene
- Deactivate Tumor Suppressor

TMTOWTDI



?

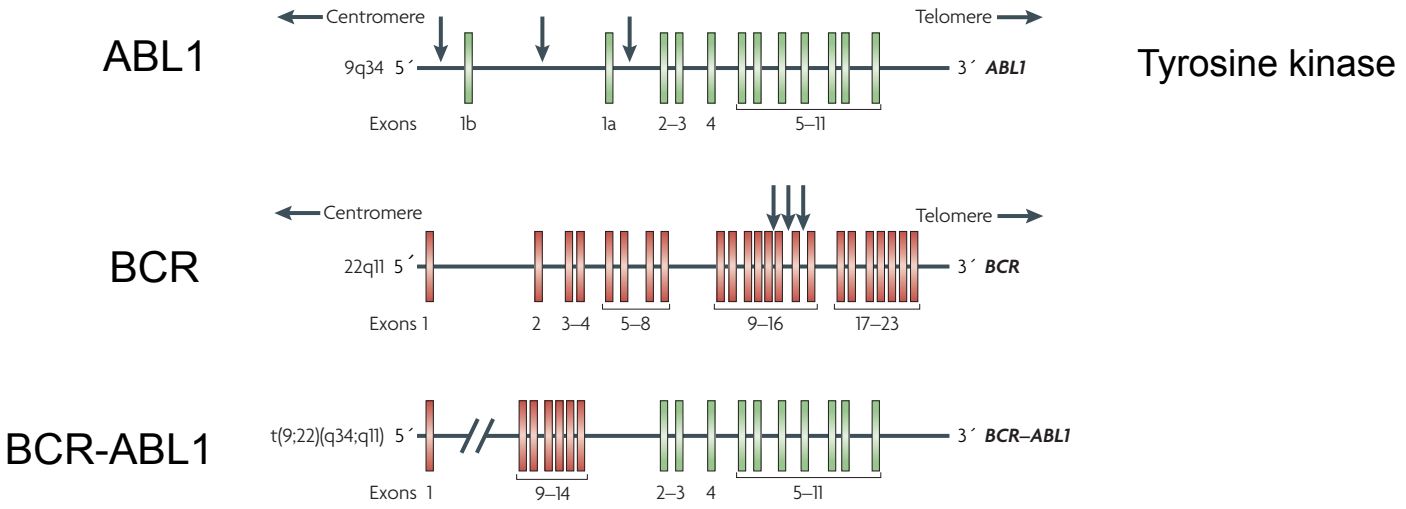
Examples of Fusions With Understood Cancer-driving Mechanisms

1. BCR—ABL1 in chronic myelogenous leukemia
2. EWS—FLI1 in Ewings sarcoma
3. SS18—SSX in synovial sarcoma
4. TMPRSS2-[ERG/ETV1] in prostate cancer
5. IGH-MYC in Burkitt's lymphoma
6. MYB-NFIB in adenoid cystic carcinomas
7. LACTB2-NCOA2 in colorectal cancer
8. MYB-QKI in angiocentric glioma

Classification of Gene Fusion Consequences

Creation of a fusion protein: constitutive kinase

(ex. 1) BCR-ABL1 fusion in Chronic Myelogenous Leukemia



Tyrosine kinase

Mitelman, 2007

Fusion protein is a constitutively active kinase signaling proliferation

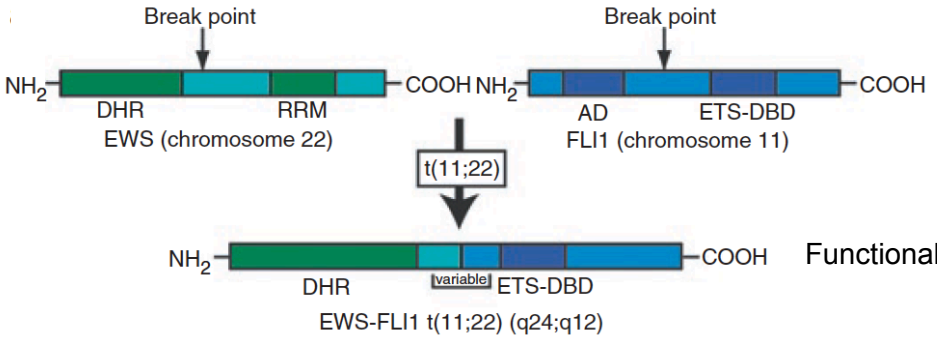
awm

Classification of Gene Fusion Consequences

Creation of a fusion protein: novel Tx factor

(ex. 2) EWS—FLI1 in Ewings Sarcoma

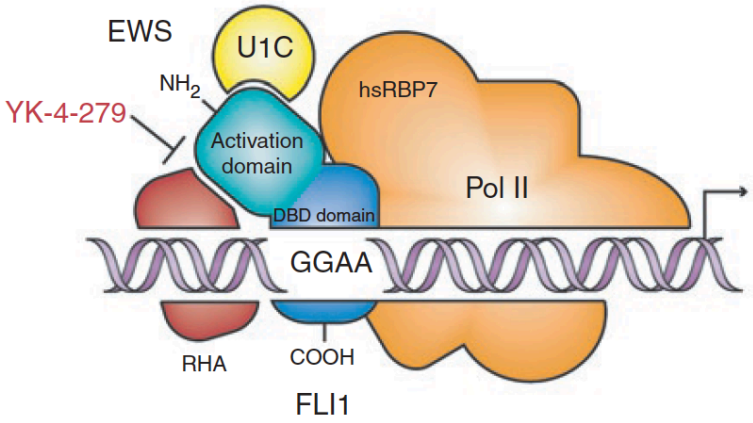
EWS: RNA-binding protein and Tx activator



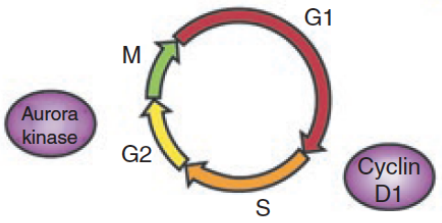
FLI1: ETS-family Tx factor

Functional Fusion Protein

Molecular Mechanism: Aberrant Transcription Factor Inducing Cell Cycle Activation



Transcriptional activation and repression
Alteration of splice site selection
Modulation of RNA half-life



Up-regulation of
Aurora kinase and Cyclin D1

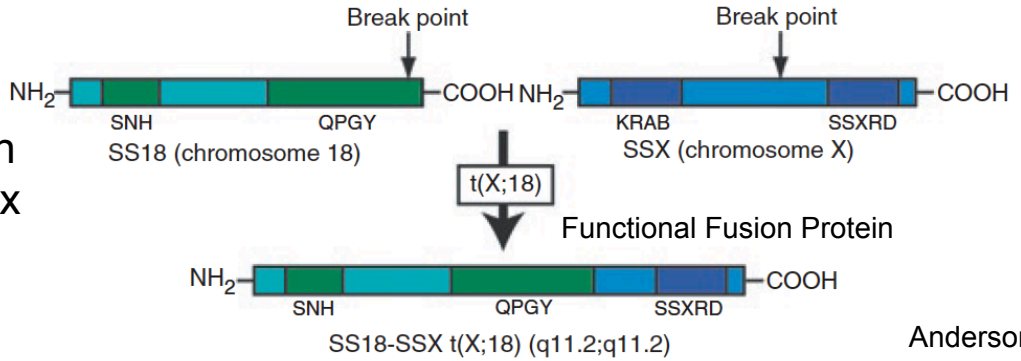
Anderson et al., Pediatric Research, 2012

Classification of Gene Fusion Consequences

Creation of a fusion protein: induces chromatin remodeling

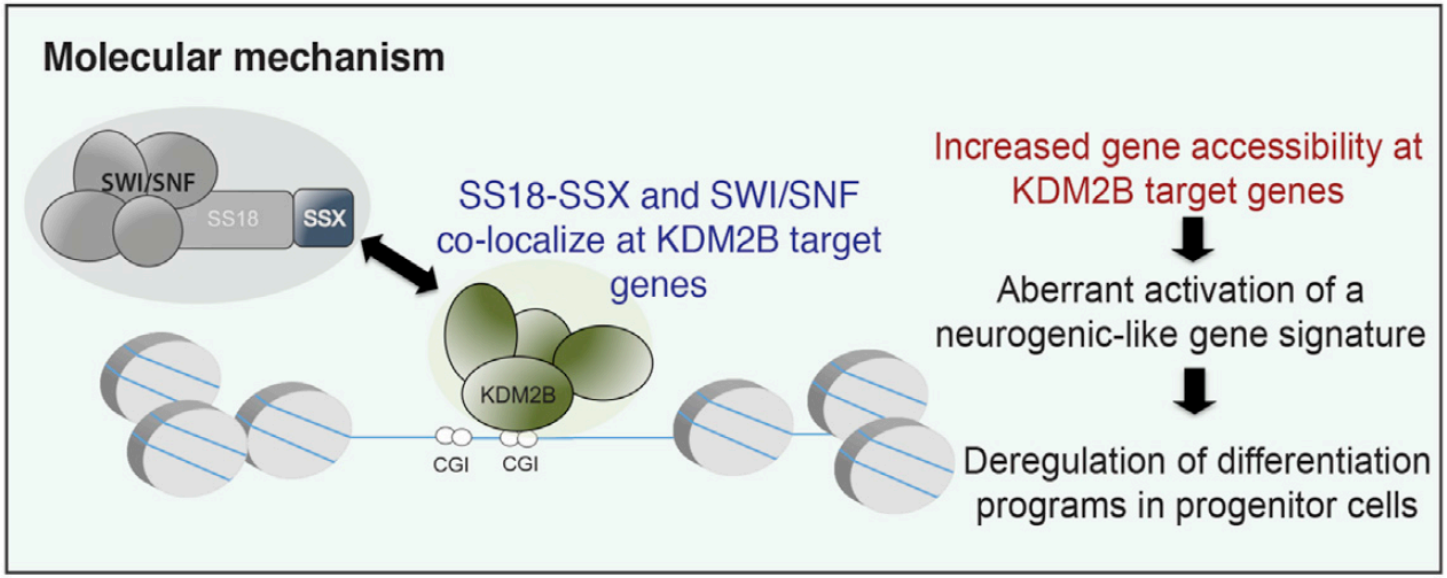
(ex. 3) SS18—SSX in Synovial Sarcoma

SS18: subunit of SWI/SNF chromatin remodeling complex



SSX: Tx factor

Anderson et al., Pediatric Research, 2012



Chromatin remodeling -> Cellular Proliferation

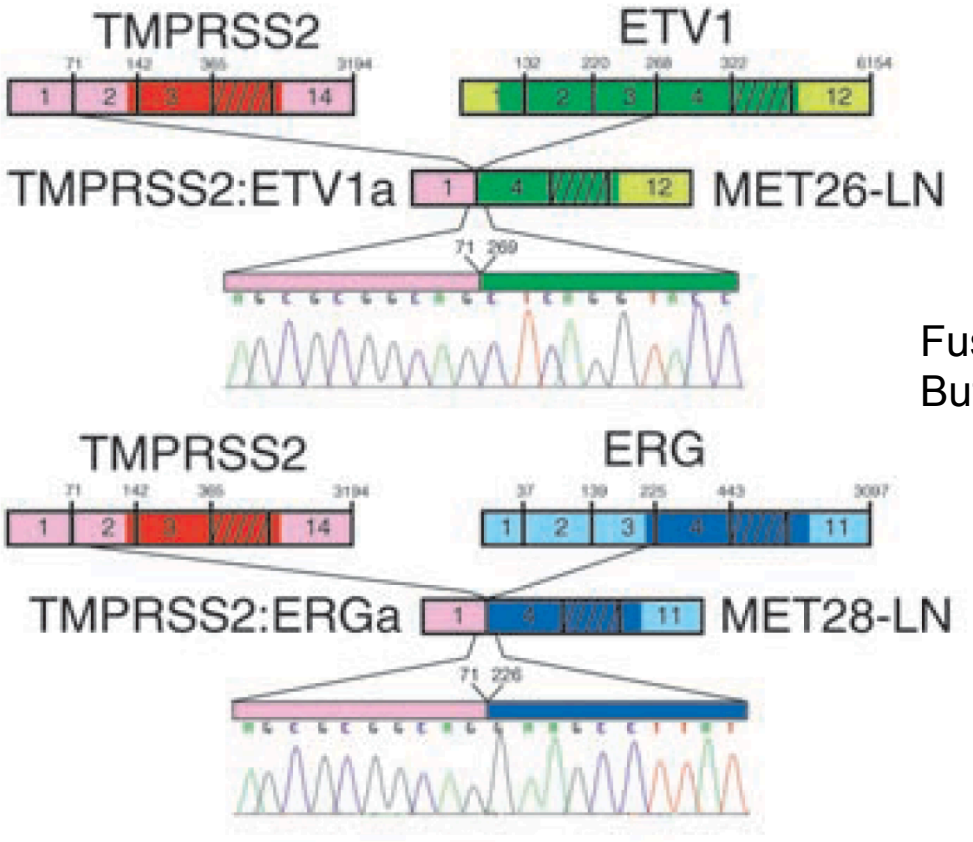
Banito et al., Cancer Cell, 2018

Classification of Gene Fusion Consequences

Deregulation of a proto-oncogene: Tx activation

(ex. 4) TMPRSS2-[ERG/ETV1] in Prostate Cancer

Serine protease
Upregulated in prostate



Tx factor

Fusion Transcript,
But no fusion protein

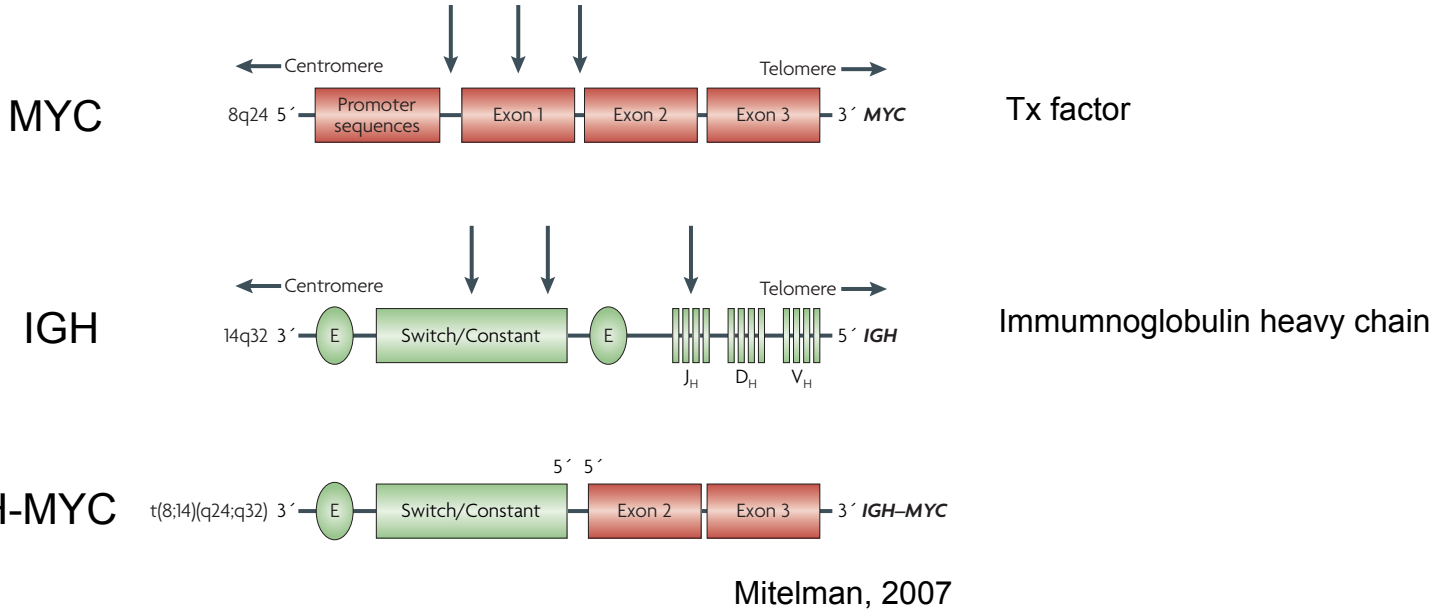
Tomlins, Science 2005

TMPRSS2 promoter drives over-expression of ETS-family Tx Factors

Classification of Gene Fusion Consequences

Deregulation of proto-oncogene: Tx activation

(ex. 5) IGH-MYC fusion in Burkitt's Lymphoma



Mitelman, 2007

MYC now constitutively driven by IGH promoter

awm

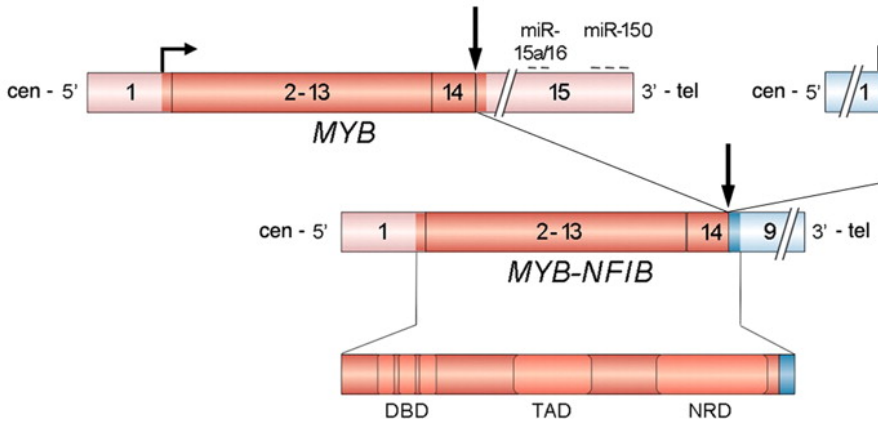
Classification of Gene Fusion Consequences

Deregulation of proto-oncogene: block Tx downregulation

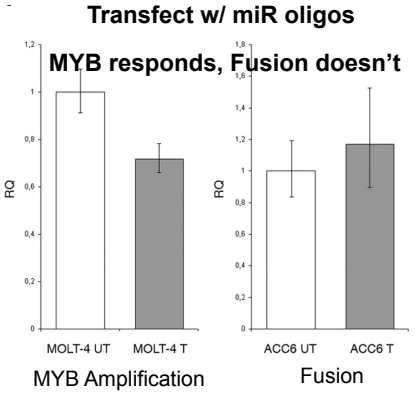
- Alternate forms of deregulation: microRNA binding

(ex. 6) MYB-NFIB in adenoid cystic carcinomas

MYB
Tx factor



NFIB
Tx factor



Fusion removes miRNA-regulation of MYB

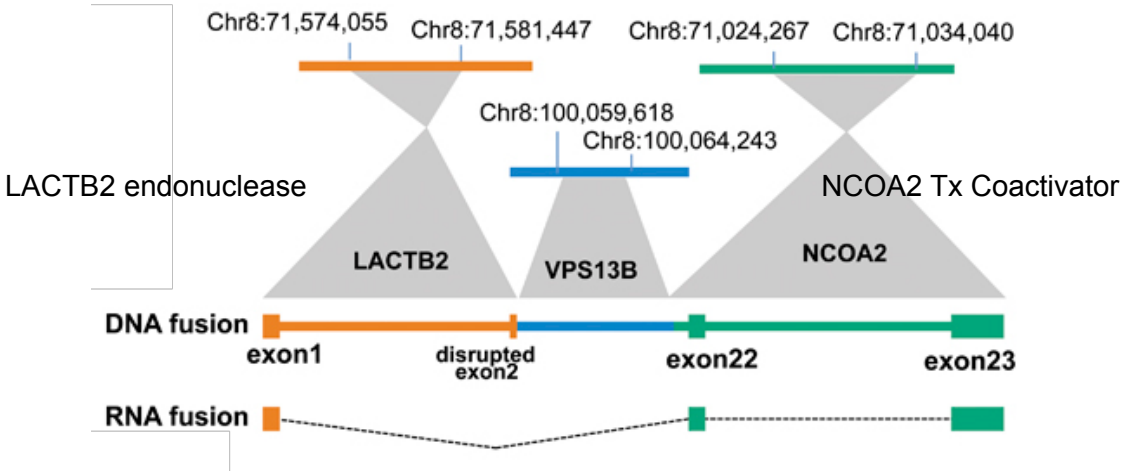
Persson, 2009

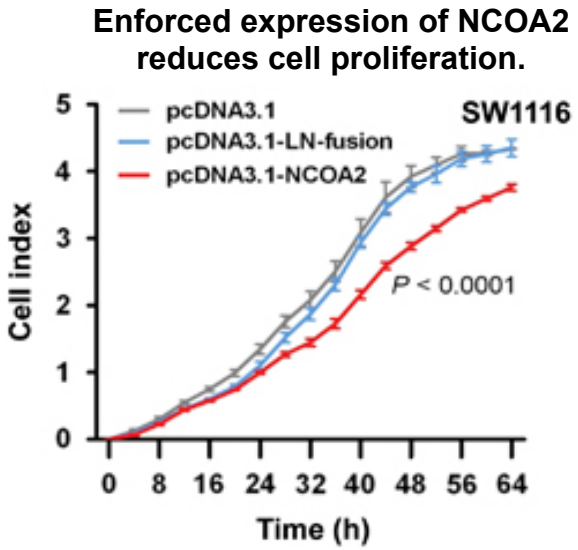
awm

Classification of Gene Fusion Consequences

Disrupt tumor suppressor gene

(ex. 7) LACTB2-NCOA2 Fusion in Colorectal Cancer





Yu, 2015

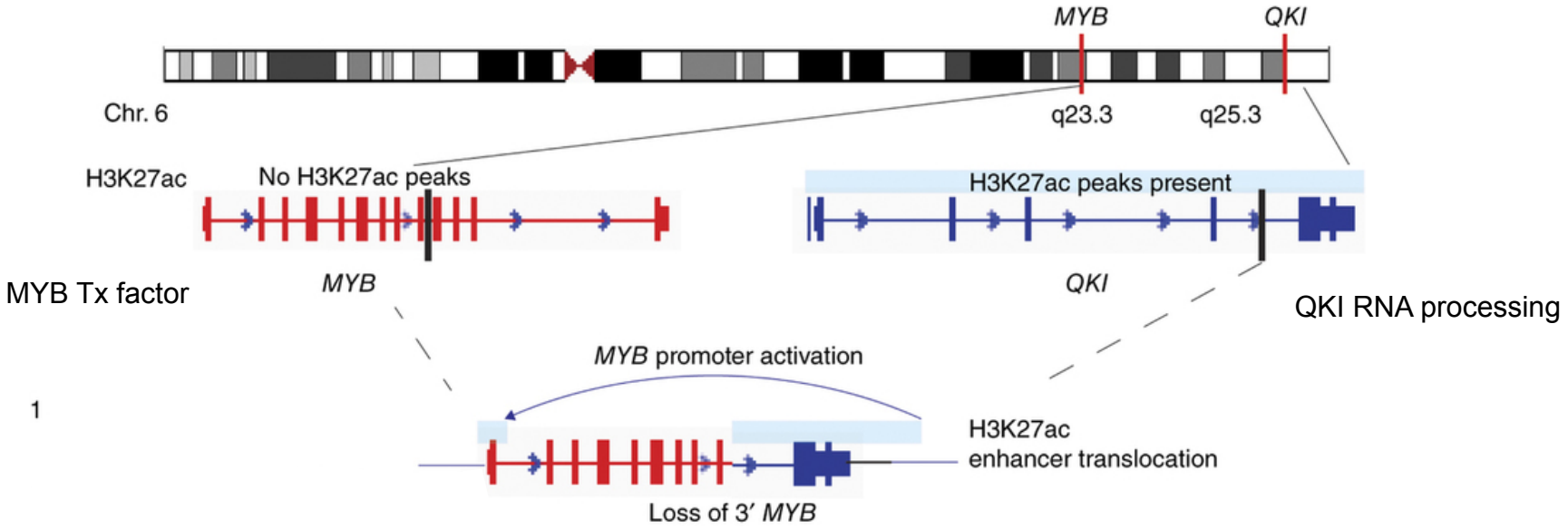
Fusion disrupts NCOA2, reducing its expression and tumor suppression

awm

Classification of Gene Fusion Consequences

Multiple mechanisms are possible

(ex. 8) MYB-QKI rearrangements in angiocentric glioma



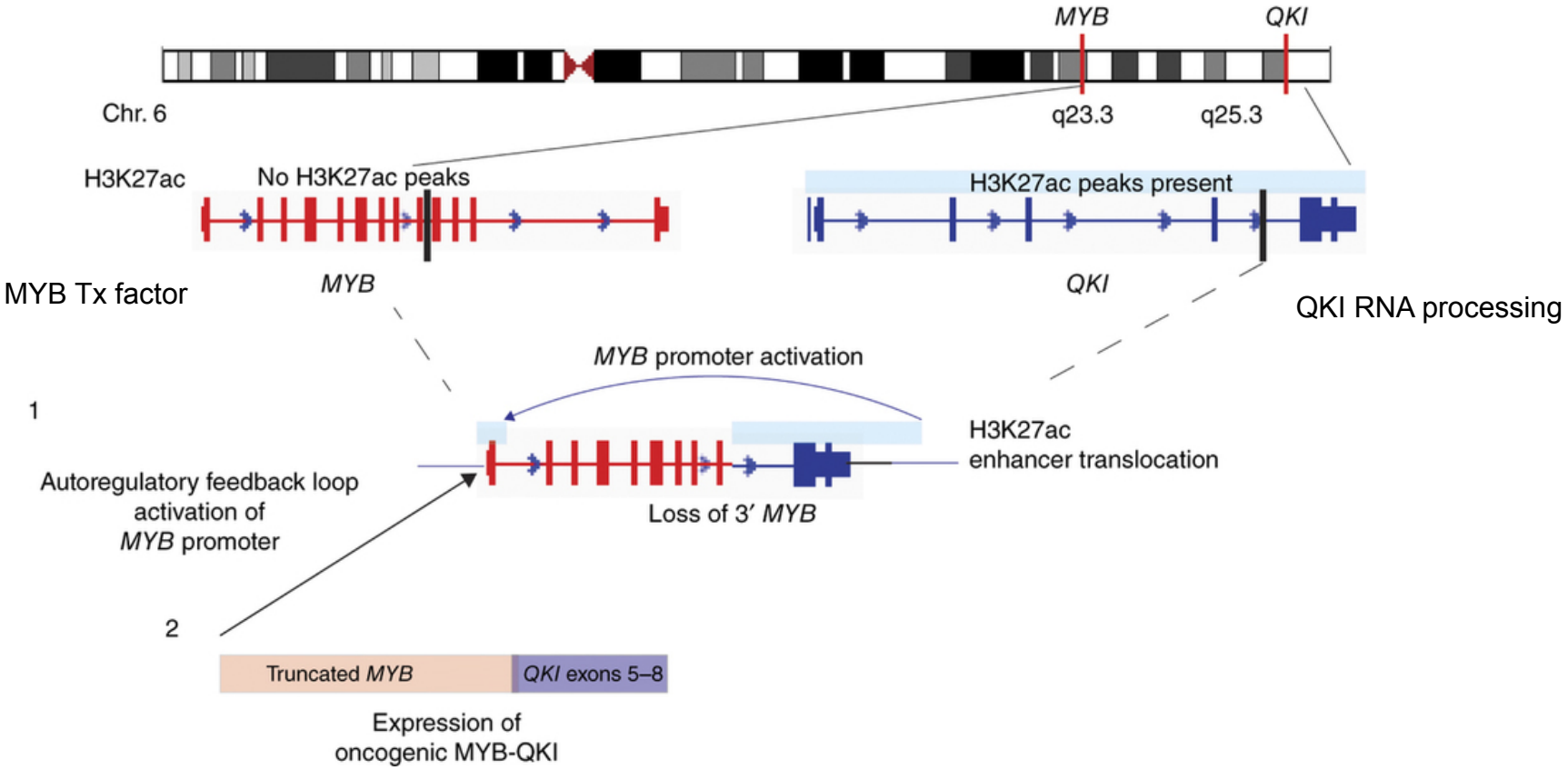
Bandopadhayay, 2016

awm

Classification of Gene Fusion Consequences

Multiple mechanisms are possible

(ex. 8) MYB-QKI rearrangements in angiocentric glioma



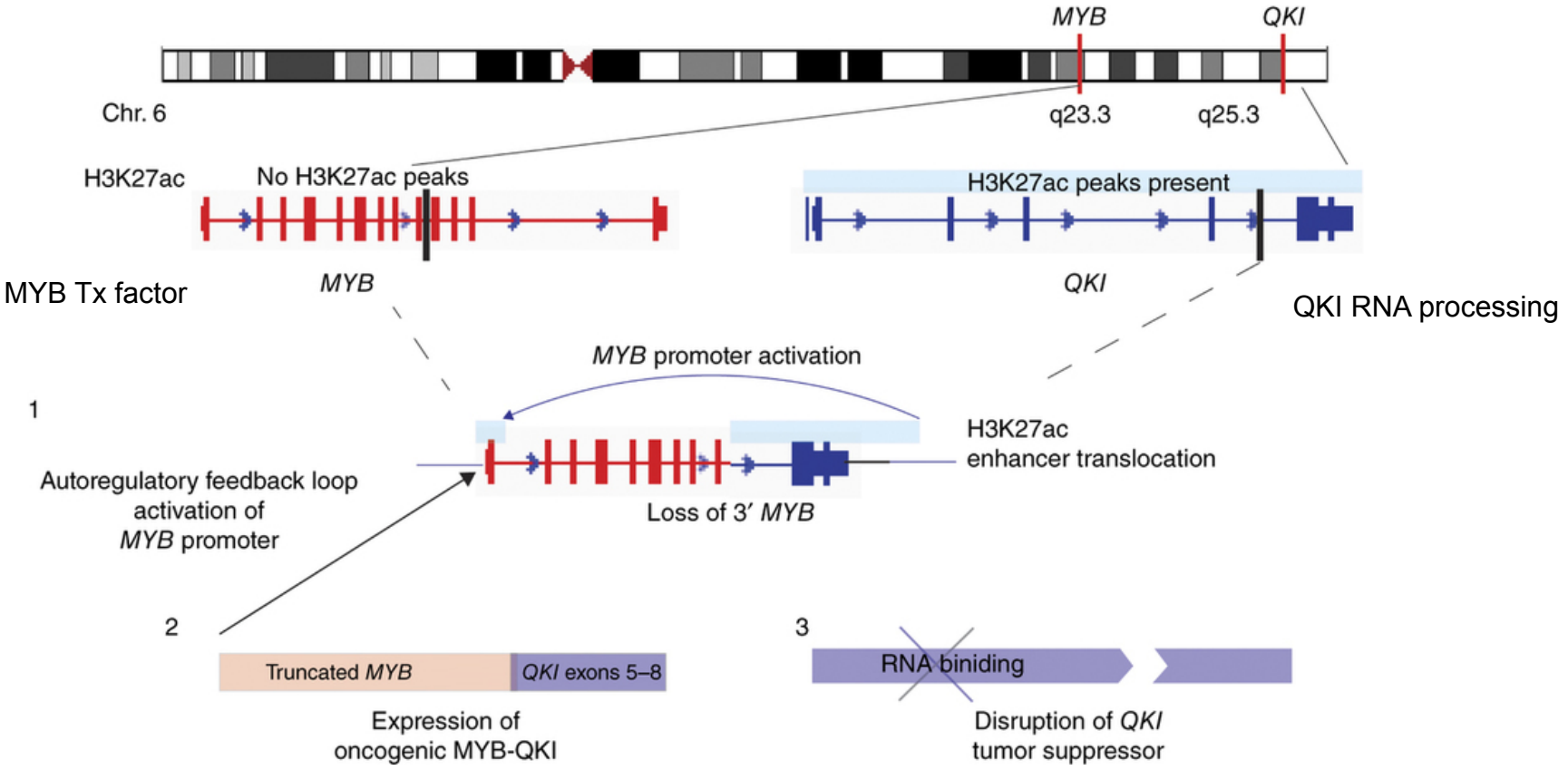
Bandopadhayay, 2016

awm

Classification of Gene Fusion Consequences

Multiple mechanisms are possible

(ex. 8) MYB-QKI rearrangements in angiocentric glioma



Bandopadhayay, 2016

awm

Mechanisms for Fusion to Drive Cancer

Activating cellular proliferation via:

- Protein: cell signaling cascades (eg. kinases)
- RNA:
  - transcriptional activation (eg. Tx factors)
  - Post-transcriptional deregulation
    (eg. removing regulatory motifs)
- DNA: chromatin remodeling
  (eg. Repositioning enhancers,
  altering epigenetic marks)

TMTOWTDI !

Genomic Effects of Gene Fusions

(signatures allowing detection)

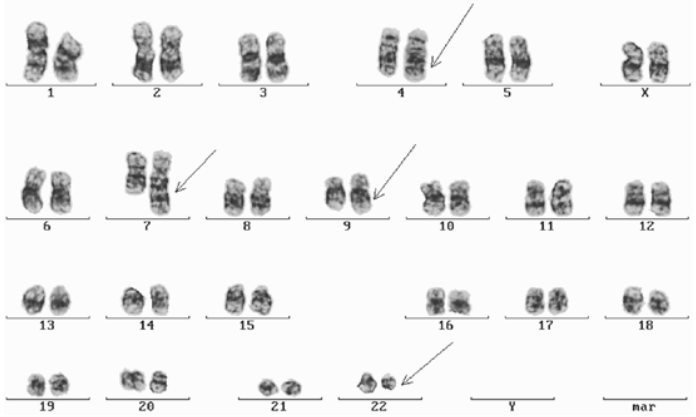
- Chimeric DNA sequence
- Chimeric mRNA sequence
- Expression changes

awm

Discovery Platforms

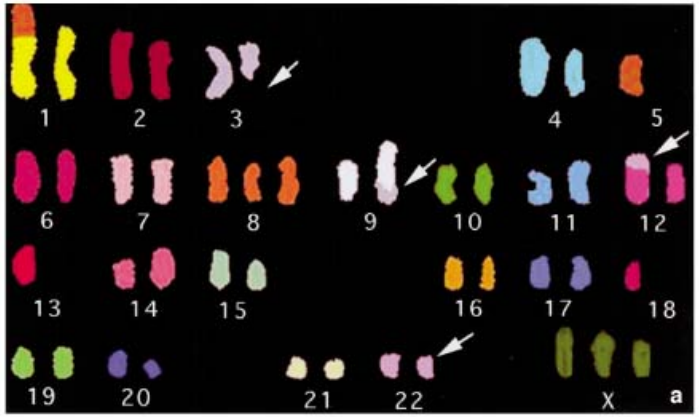
- Cytogenetics - Karyotyping
  - Labour intensive
  - Low throughput

Chromosome Banding Analysis



Jarosová, 2000

Spectral Karyotyping



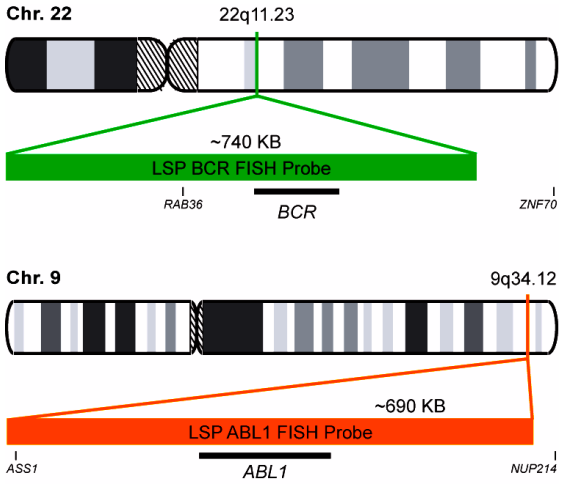
Markovic, 2000

awm

Discovery Platforms

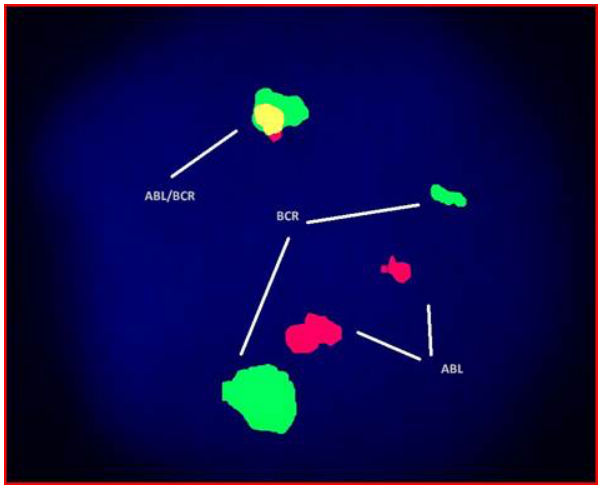
- Cytogenetics – Fluorescence In-Situ Hybridization (FISH)
  - Targeted
  - Low throughput

FISH Probes (Come Together Assay)



courtesy cytotest.com

FISH Image

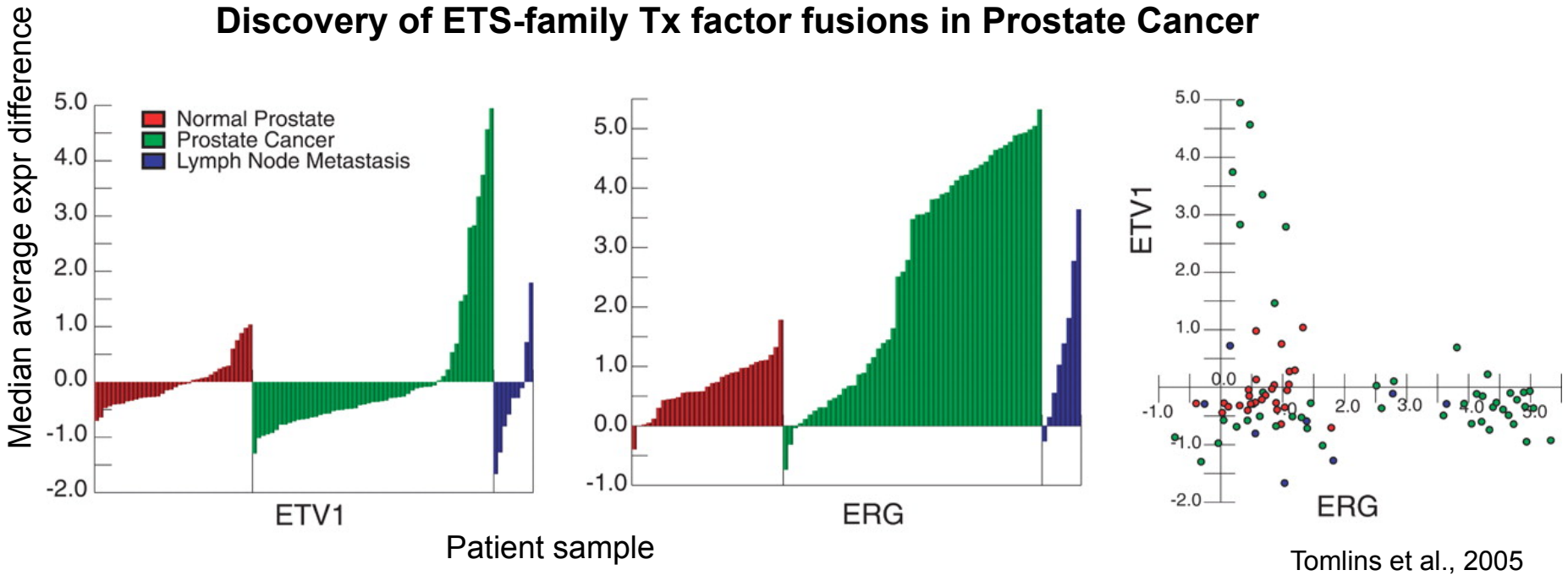


Vaniawala, 2015

awm

Discovery Platforms

- Expression Arrays
  - COPA: Cancer Outlier Profile Analysis
  - Inexpensive
  - No sequence information
  - Use RACE (rapid amplification of cDNA ends) to find potential fusion partner.



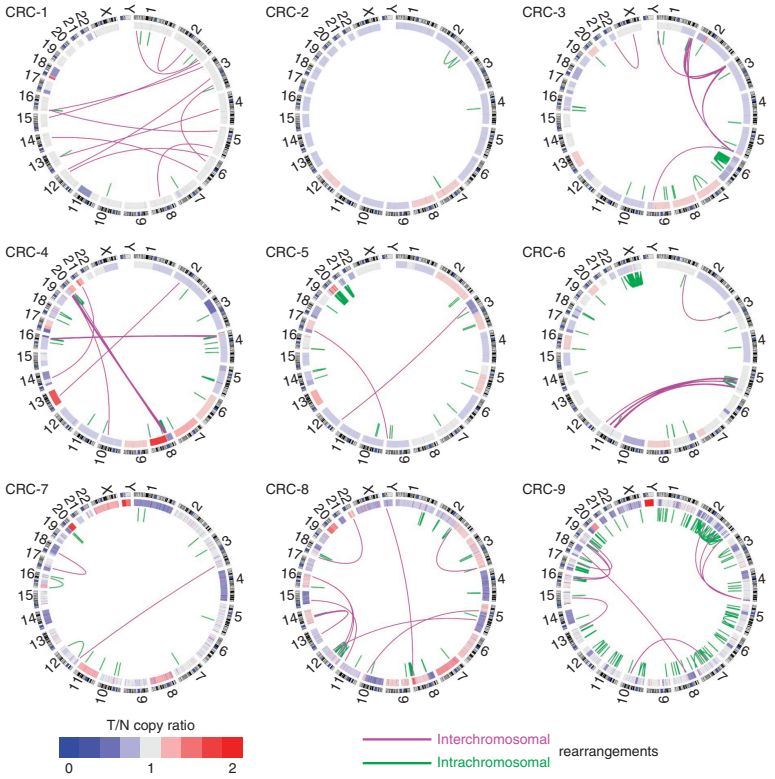
Tomlins et al., 2005

awm

Discovery Platforms

- Genome Sequencing
  - Comprehensive
  - Expensive
  - No expression information

Discovery of *VTI1A-TCF7L2* fusion in colorectal adenocarcinomas



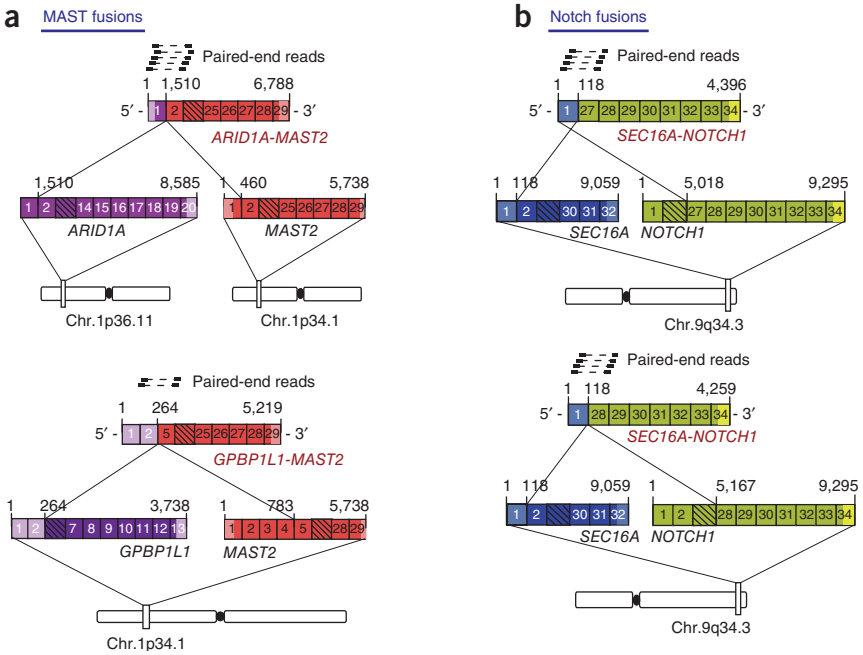
Bass et al., 2011

awm

Discovery Platforms

- mRNA Sequencing (RNA-Seq)
  - Inexpensive
  - Expression information
  - Exact fusion transcript identified
  - Not as comprehensive as genome sequencing

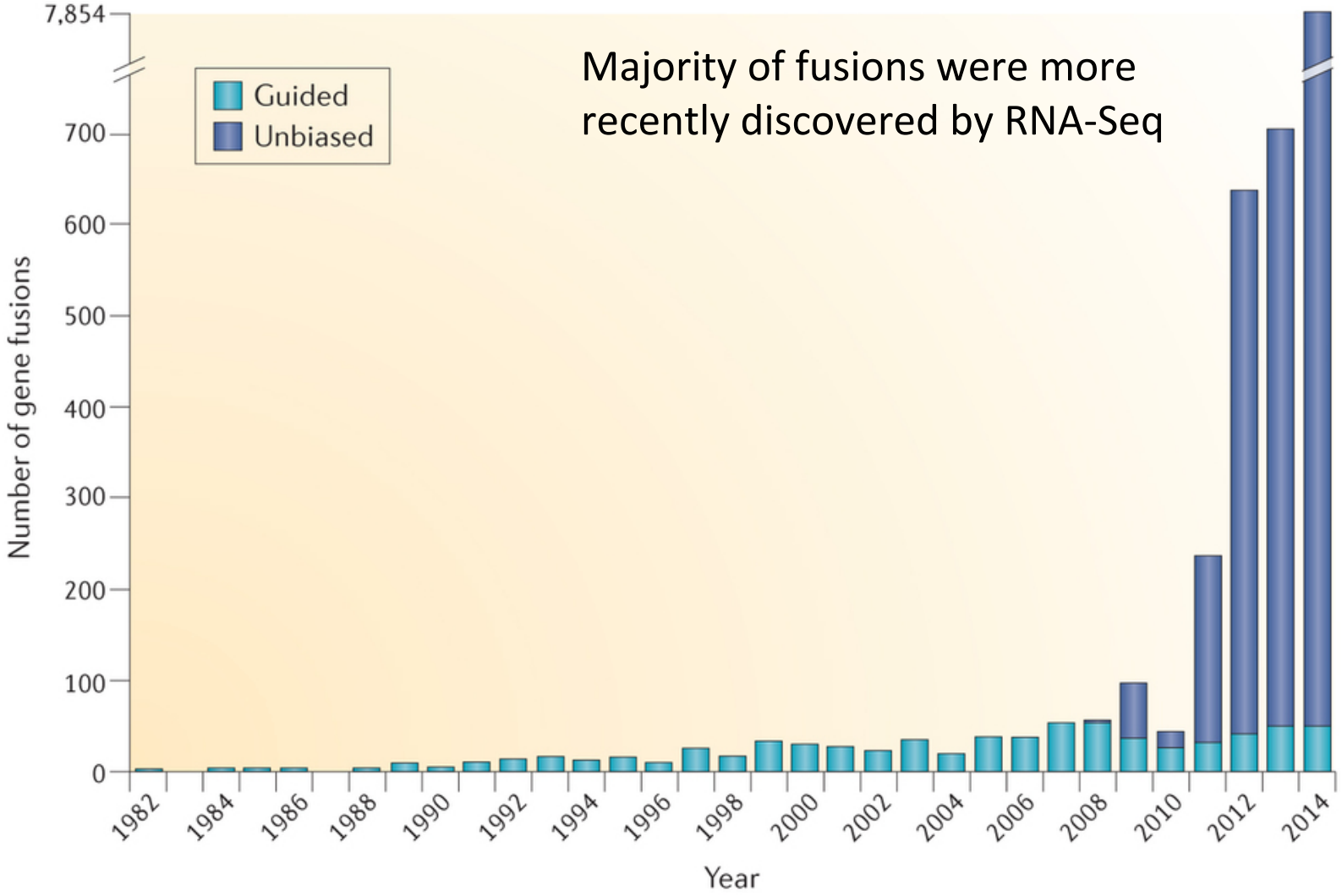
Discovery of *MAST* and *Notch* fusions in breast cancer



Robinson et al., 2011

awm

Massive increase in fusion discovery driven by sequencing



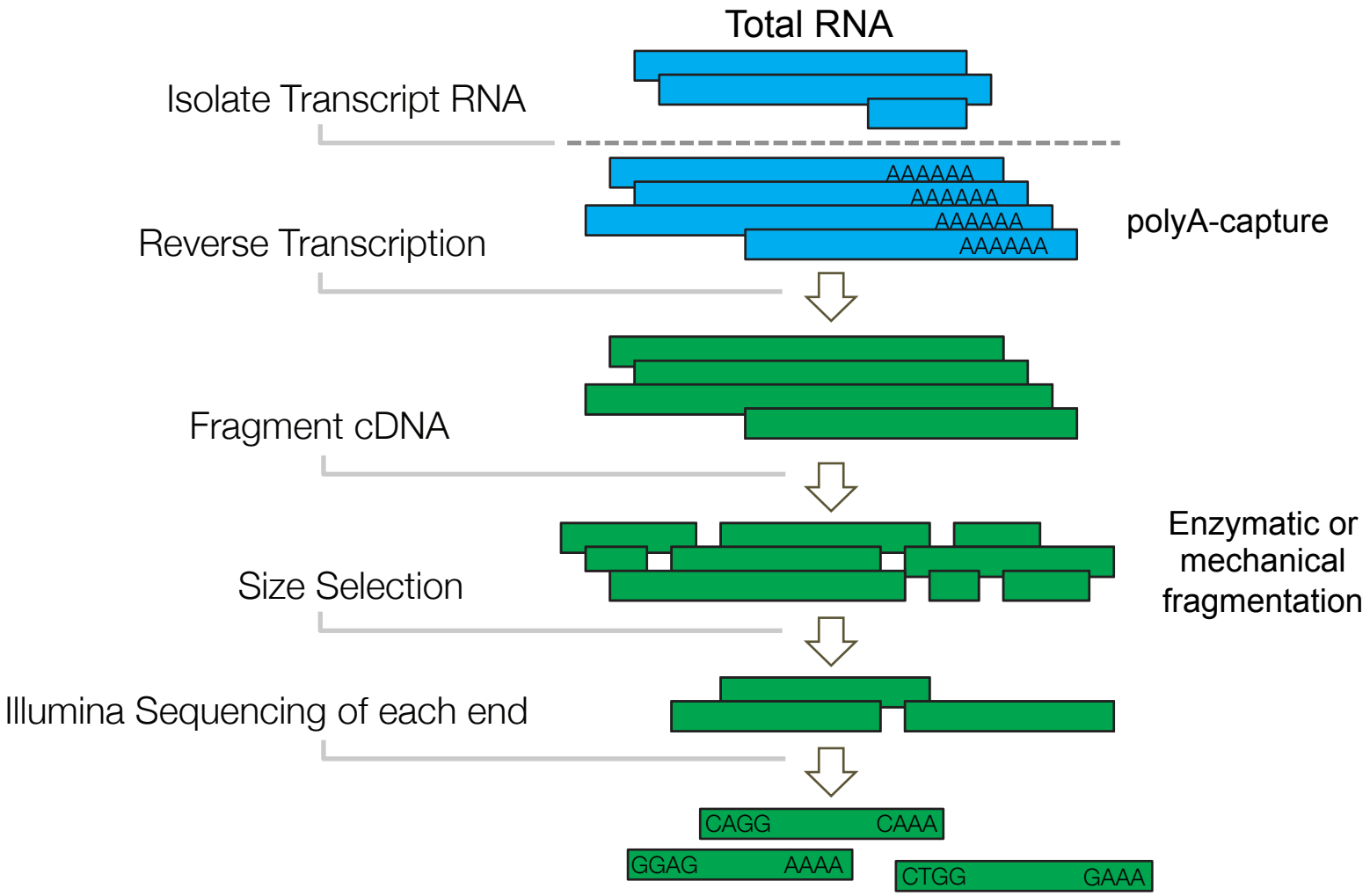
21k

2018

awm

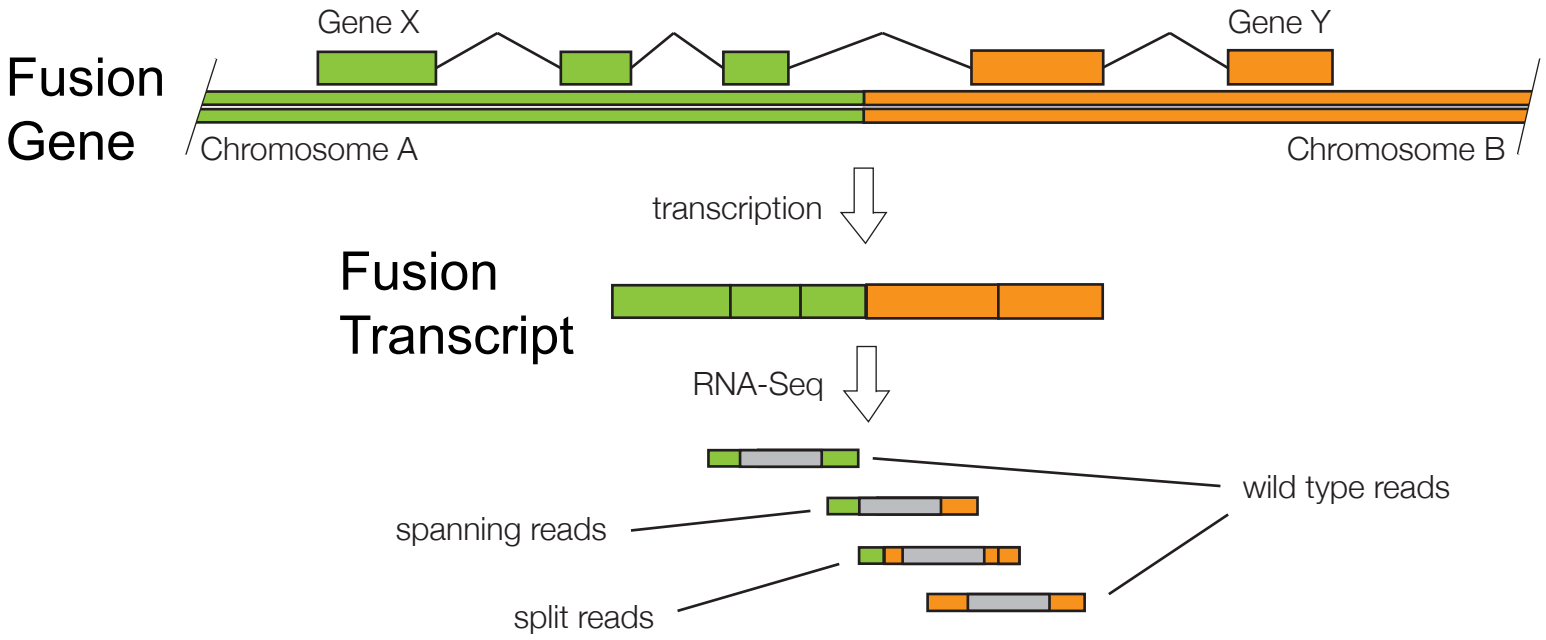
Mertens, 2015

How RNA-seq data is generated



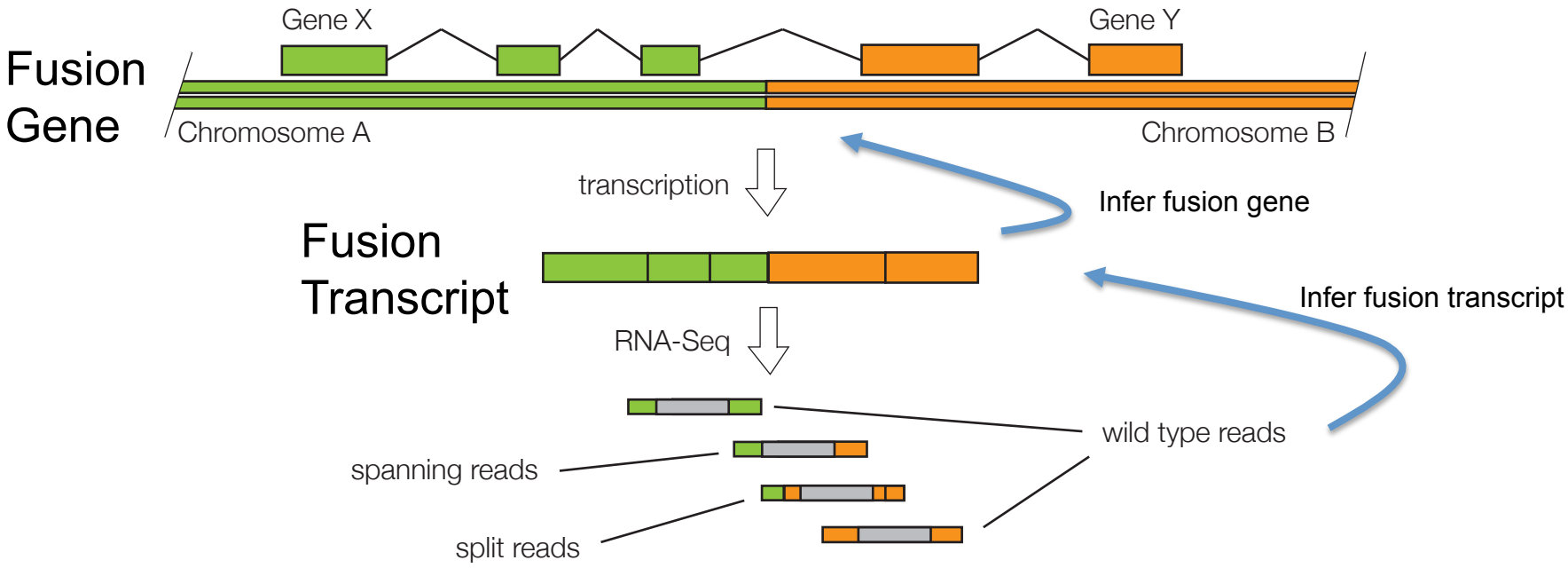
awm

Chimeric Fusion Genes Produce Chimeric RNA-Seq Reads

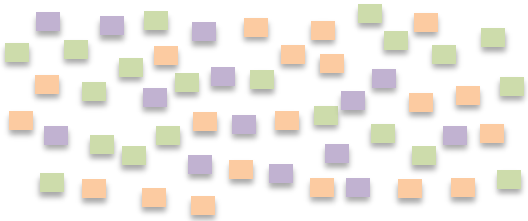


awm

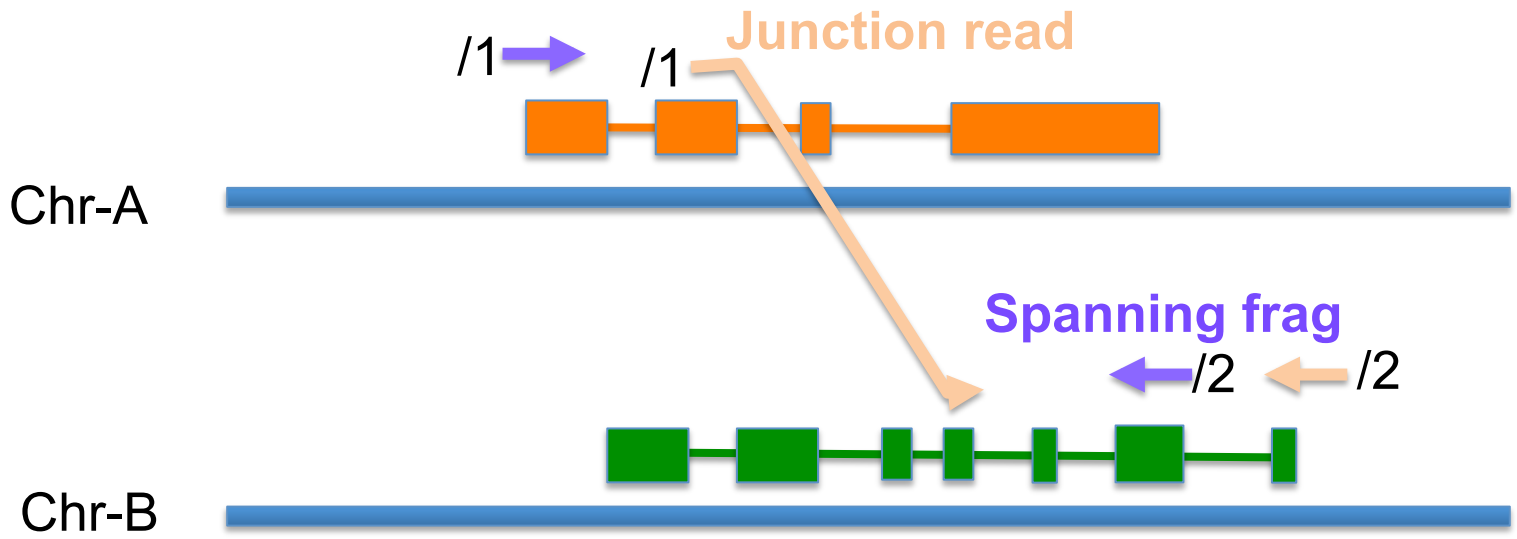
Chimeric Fusion Genes Produce Chimeric RNA-Seq Reads



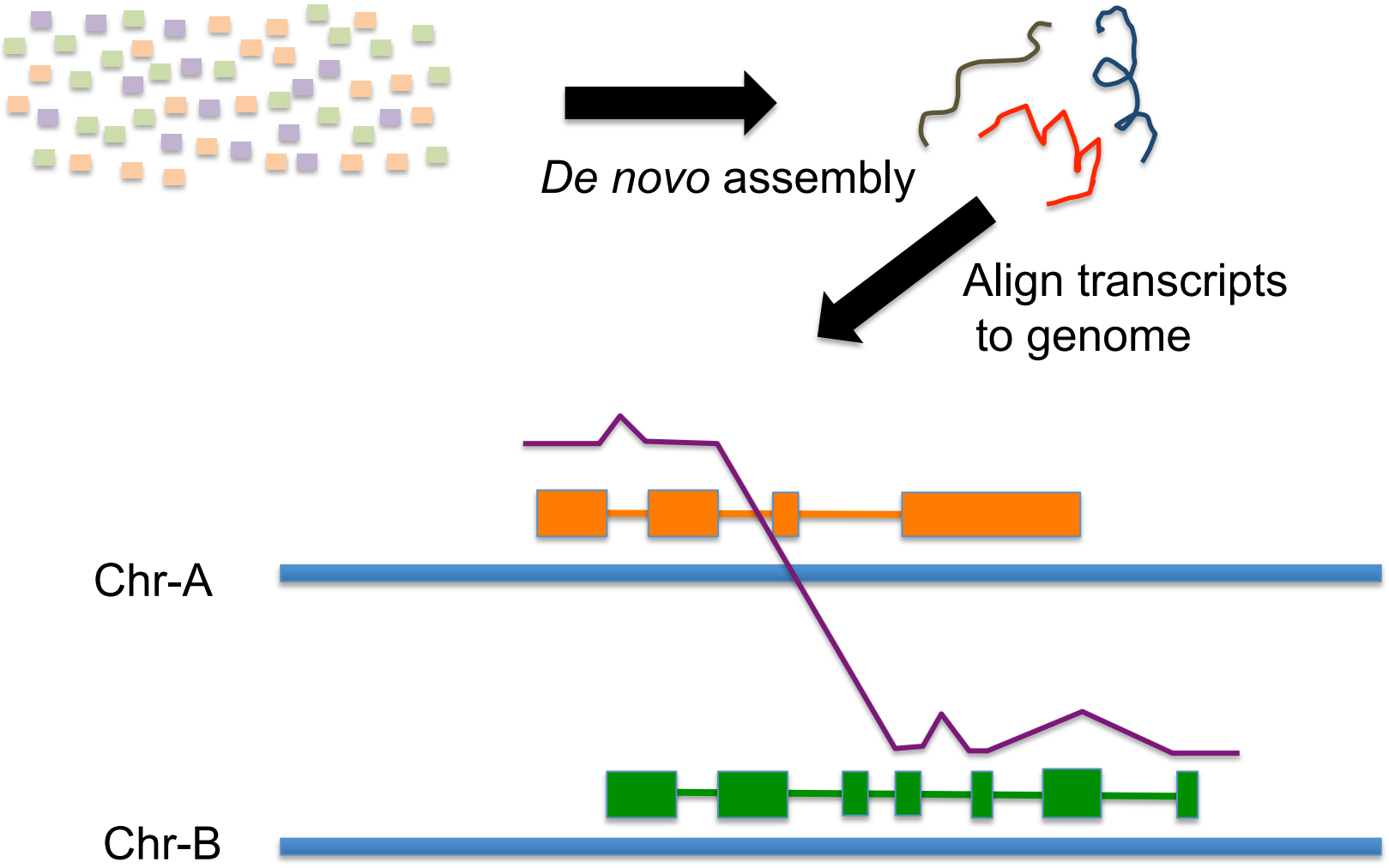
Detecting Fusion Transcripts from Paired-end RNA-Seq Reads (Discordant Spanning reads and Fusion Junction Reads)



Align reads to the genome,
Identify discordant pairs and junction reads.

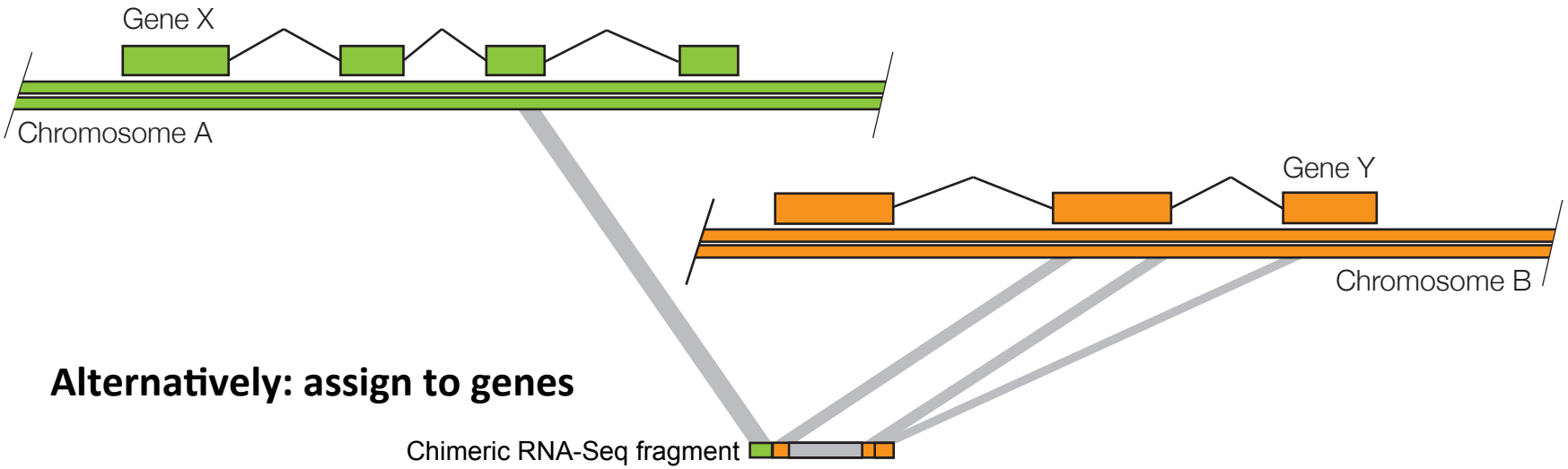


Detecting Fusion Transcripts Using *De novo* Transcript Assemblies

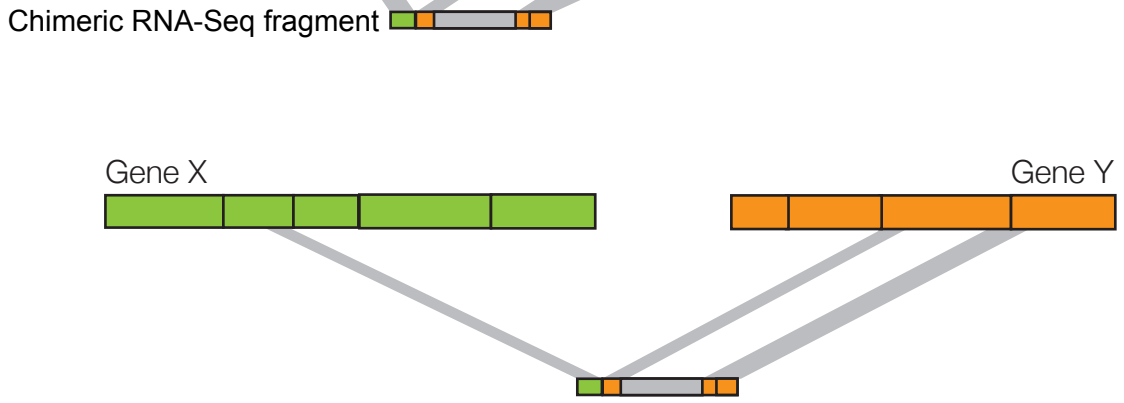


RNA-Seq and the Alignment Problem

- **Problem: assign observed reads to genomic loci**



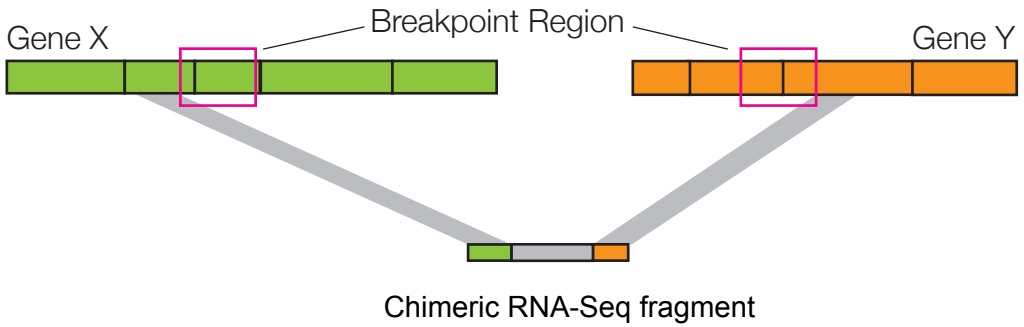
- **Alternatively: assign to genes**



awm

Paired End Read Alignments Inform Fusion Discovery

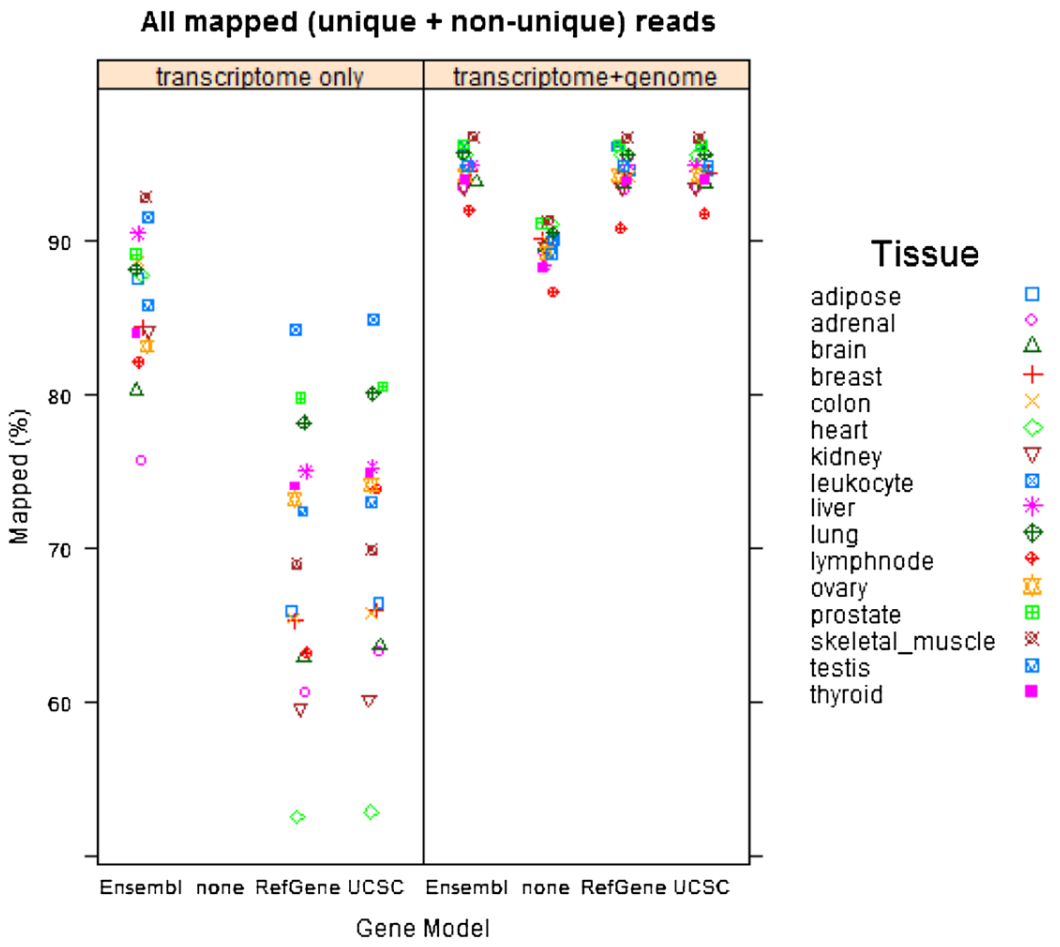
- Paired end read alignments can provide approximate breakpoint information



- Split-reads are required for precise junction identification.

awm

Choice of Reference affects Mapping Rates



Zhao and Zhang, 2016 http://dx.doi.org/10.5772/61197

awm

Choice of Target Influences Read Mappings and Fusion Predictions

Complexity of mapping reads

Genome only

Mis-mapped reads (belong elsewhere in the genome)

Restricted to reference transcripts (no novel exons allowed)

Vs.

Transcriptome only

Vs.

Genome & Transcriptome

(most rigorous)

awm

Computational tools for gene fusion detection using NGS data

Method	URL	Feature
Fusion detection specific		
BreakFusion	http://bioinformatics.mdanderson.org/main/BreakFusion	Identifying gene fusions from paired-end RNA-Seq data
ChimeraScan	http://code.google.com/p/chimerascan/	Detecting fusion transcripts from RNA-Seq data
Comrad	http://code.google.com/p/comrad/	Using both RNA-Seq and WGS data to detect genomic rearrangements and aberrant transcripts
FusionAnalyser	http://www.ilte-cml.org/FusionAnalyser/	Detecting gene fusions from paired-end RNA-Seq data
deFuse	http://sourceforge.net/apps/mediawiki/defuse/	Identifying gene fusions from RNA-Seq data
FusionMap	http://www.omicsoft.com/fusionmap/	Using either WGS or RNA-Seq data to detect fusion genes
FusionHunter	http://bioen-compbio.bioen.illinois.edu/FusionHunter/	Detecting fusion transcripts from RNA-Seq data
FusionSeq	http://archive.gersteinlab.org/proj/rnaseq/fusionseq/	Identifying fusion transcript from RNA-Seq data
ShortFuse	https://bitbucket.org/mckinsel/shortfuse	Identifying fusion transcripts from RNA-Seq data
SnowShoes-FTD	http://mayoresearch.mayo.edu/mayo/research/biostat/stand-alone-packages.cfm	Detecting fusion transcripts from RNA-Seq data
SOAPfusion[a]	http://soap.genomics.org.cn/SOAPfusion.html	Part of the software SOAP, for genome-wide detection of gene fusions from RNA-Seq data
TopHat-Fusion	http://tophat-fusion.sourceforge.net/	An enhanced version of TopHat, for detection of fusion transcripts from RNA-Seq data

* Table from Wang et al. Briefings in Bioinformatics. 2012

And many more since 2012, including:

BARNACLE (2013)
PRADA (2014)
FusionCatcher (2014)
TRUP (2015)
JAFFA (2015)
NCLscan (2015)
IDP-fusion (2015)
INTEGRATE (2016)
InFusion (2016)
ChimeRScope (2017)
ChimPipe (2017)
STAR-Fusion (2017)
Pizzly (2017)

When all else fails…

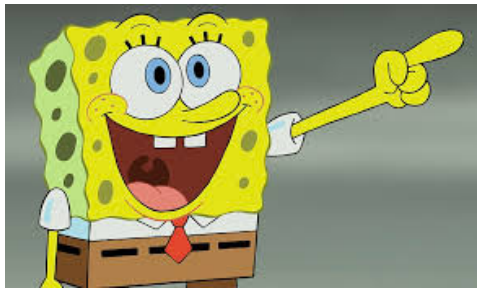
OPEN ACCESS Freely available online

PLOS ONE

The "Grep" Command But Not FusionMap, FusionFinder or ChimeraScan Captures the *CIC-DUX4* Fusion Gene from Whole Transcriptome Sequencing Data on a Small Round Cell Tumor with t(4;19)(q35;q13)

Ioannis Panagopoulos[1,2]*, Ludmila Gorunova[1,2], Bodil Bjerkehagen[3], Sverre Heim[1,2,4]

1 Section for Cancer Cytogenetics, Institute for Cancer Genetics and Informatics, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, **2** Centre for Cancer Biomedicine, Faculty of Medicine, University of Oslo, Oslo, Norway, **3** Department of Pathology, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, **4** Faculty of Medicine, University of Oslo, Oslo, Norway

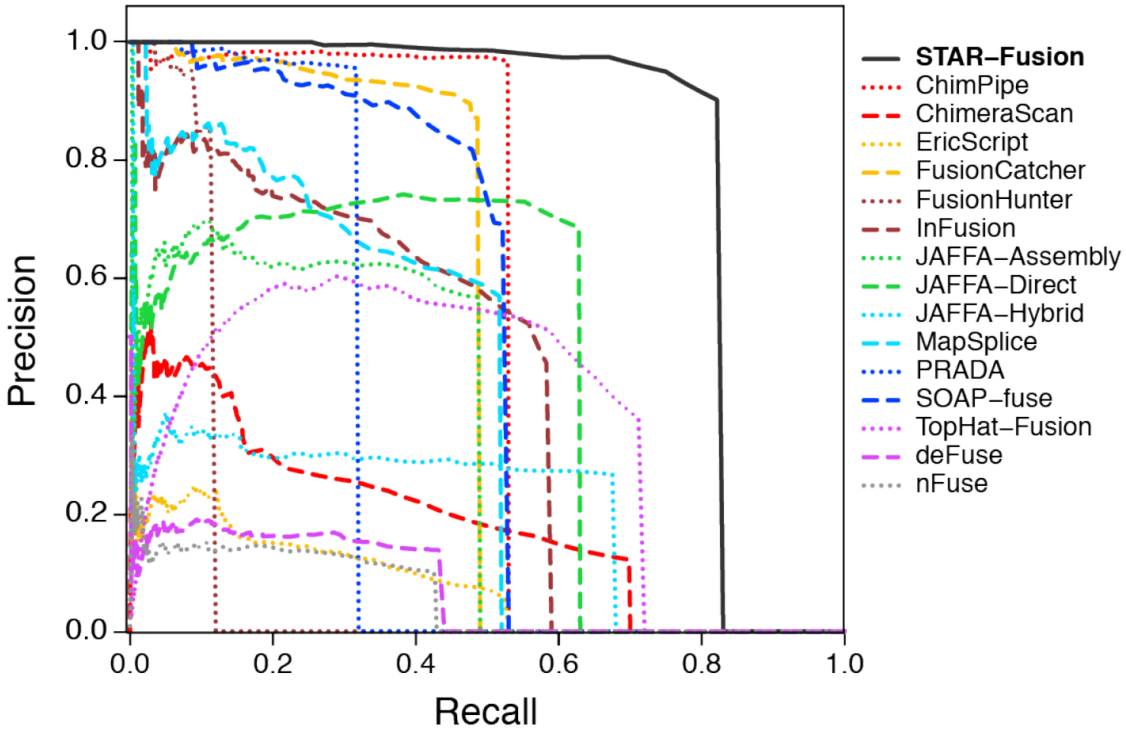


"In conclusion, FusionMap, FusionFinder, and ChimeraScan generated a plethora of fusion transcripts but did not detect the biologically important CIC-DUX4 chimeric transcript. … The "grep" command is an excellent tool to capture chimeric transcripts from RNA sequencing data"

Benchmarking Fusion-finding Tools

- **Simulated data**
  - 5 replicates
  - 2500 Simulated fusions
  - 30M PE sim RNA-Seq data
- **Genuine data**
  - 65 Cancer Cell Lines

Precision – Recall Curves



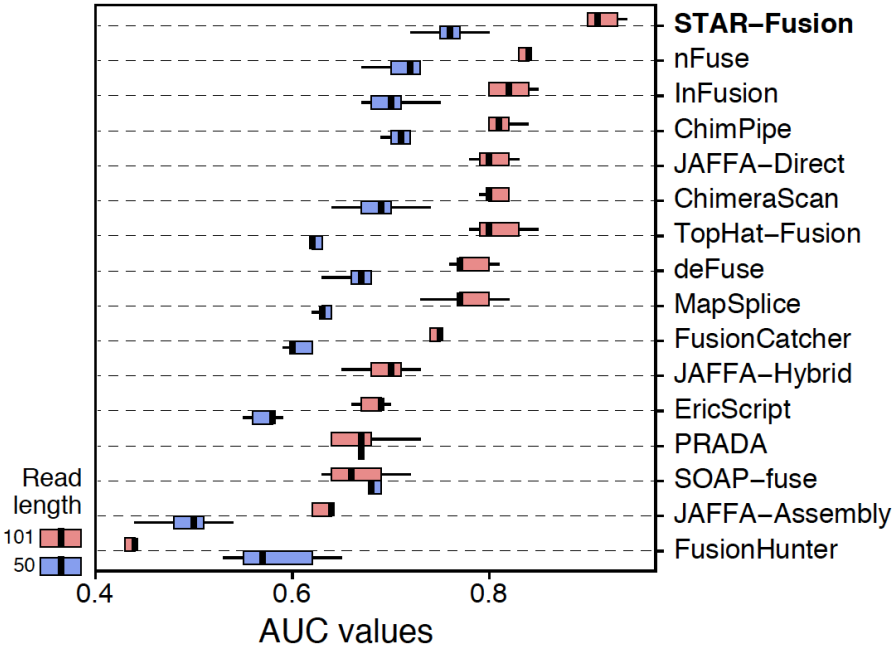
Precision = TP / (TP + FP)
Recall = TP / (TP + FN)

Accuracy = **a**rea **u**nder the **c**urve (AUC)

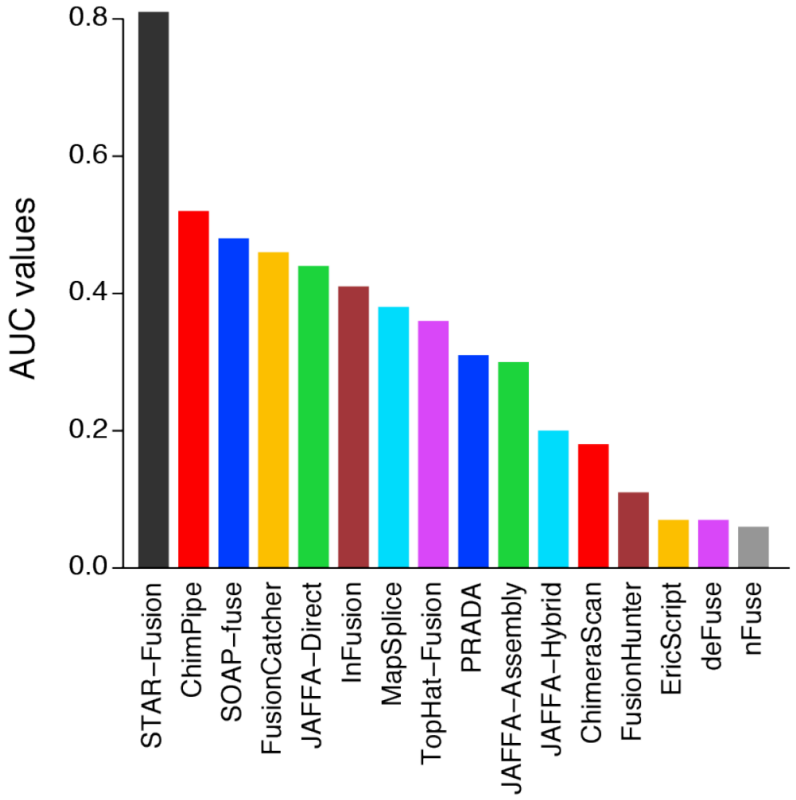
Haas & Dobin et al, bioRxiv, 2017

Benchmarking Fusion-finding Tools

Fusion prediction accuracy on simulated rna-seq data



Fusion prediction accuracy on 65 cancer cell lines



Haas & Dobin et al, bioRxiv, 2017

Sources of False Positives

- Technical artifacts
  - Alignment artifacts
    - Homologous genes
    - High expression (ribosomal RNA) + read errors
  - Chimeric read artifacts
    - Reverse transcriptase template switching
    - Ligation artifacts
- Biological artifacts
  - Natural sources of rearrangement
    - Germline variants, transposons, NUMT
    - Transcription Induced Chimeras (Read-throughs)
    - Trans-splicing

awm

Mitigating Fusion Artifacts

Apply Bioinformatic Filters

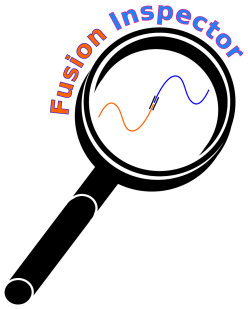
- Screen results against existing databases
  - RepeatMasker, NUMT
  - Gene lists – 'red herrings' (more later…)
  - Remove events involving ribosomal RNA, HLA
- Transcription Induced Chimeras (Read-throughs)
  - Easily identified as fusions between adjacent genes
  - Consider a minimum distance threshold.
- Consider the strength of the evidence
  - Require minimal number of supporting reads / total reads
- Examine the transcript breakpoint
  - Canonical splicing? If not…. greater chance of it being an RT or ligation artifact.

Supervised Fusion Analysis

Given a panel of fusions of interest:

- Capture any evidence for specific fusions
- Characterize the evidence:
  - Scoring the fusion
  - Expression of fusion vs. non-fused alleles
  - Functional impact (frameshift vs. in-frame, etc.)
- Facilitate visualization of the data

Bottom-up Fusion 'In silico Validation' Using FusionInspector

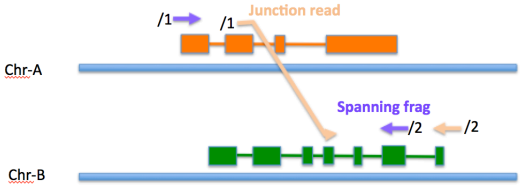


Add to whole genome. Align reads, score and assess.

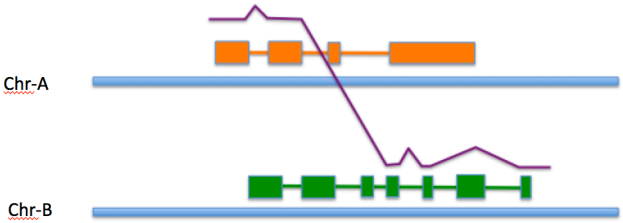
* STAR enhancements to support FusionInspector

Make mini-fusion contigs

STAR-Fusion
PRADA
SOAPfuse
FusionCatcher
…

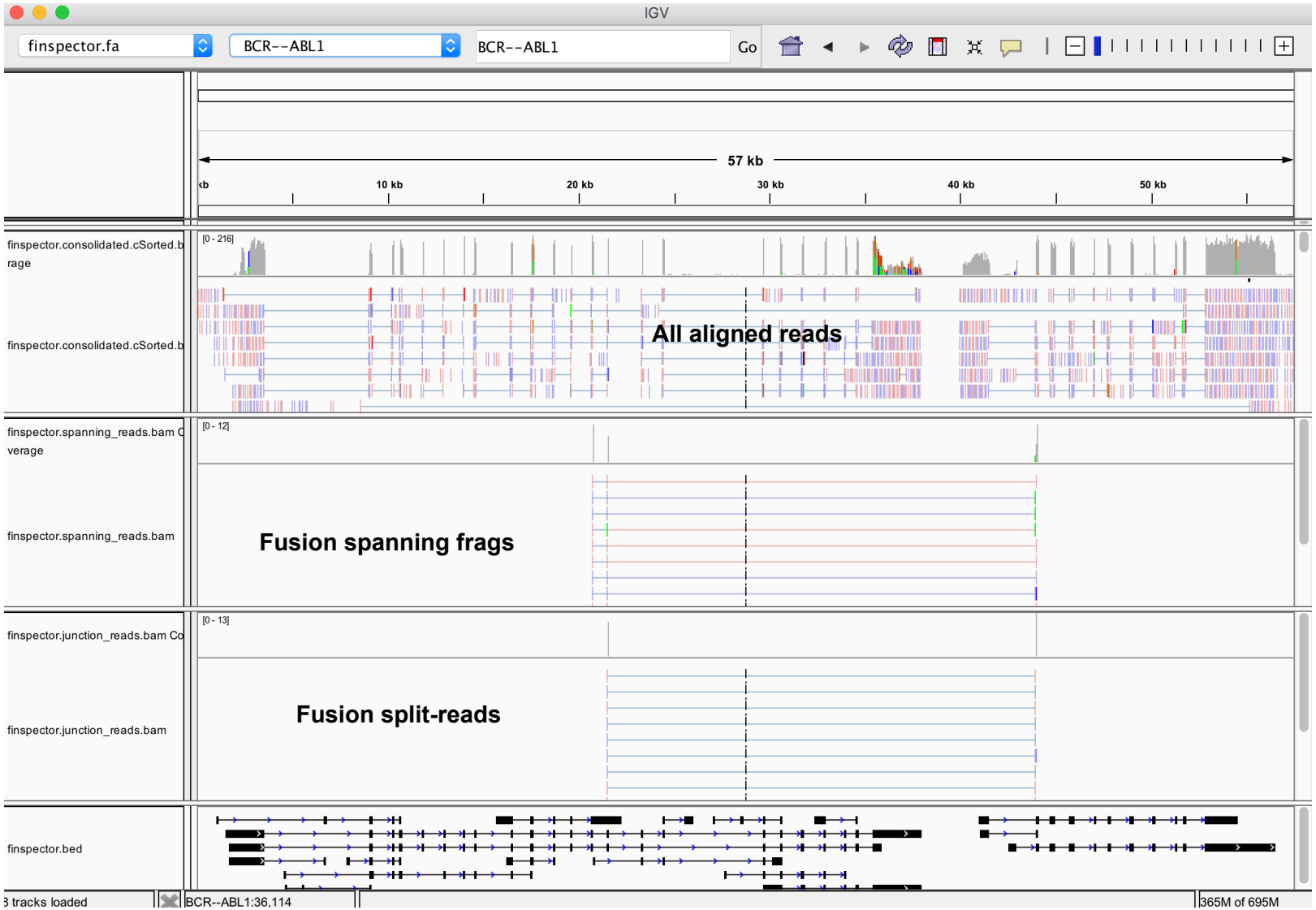


All fusion predictions



TrinityFuse
OasesFuse
JAFFA
DISCASM
…

Using IGV for Visualization with FusionInspector



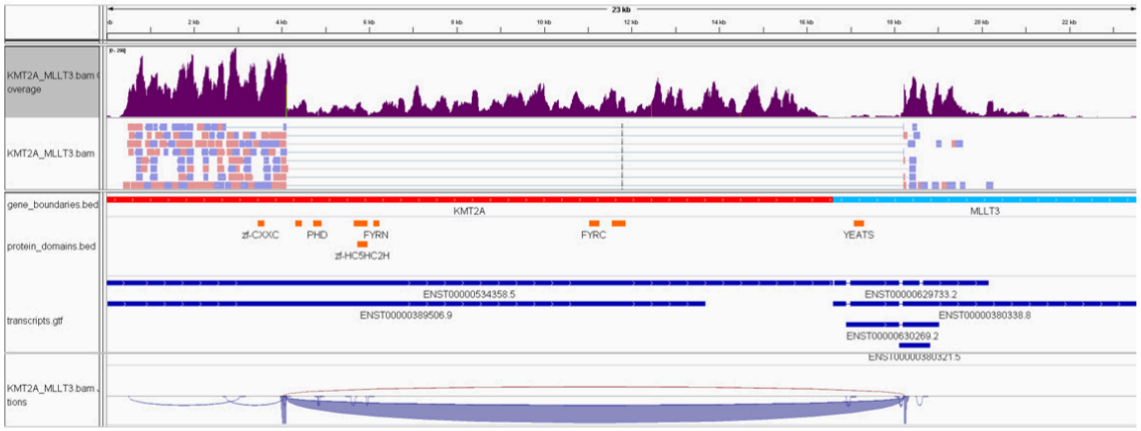
BCR

Condensed gene structures (shrunken introns)

ABL

Supervised Fusion Study via Clinker

B) Clinker Output - IGV



C) Clinker Output - Gviz

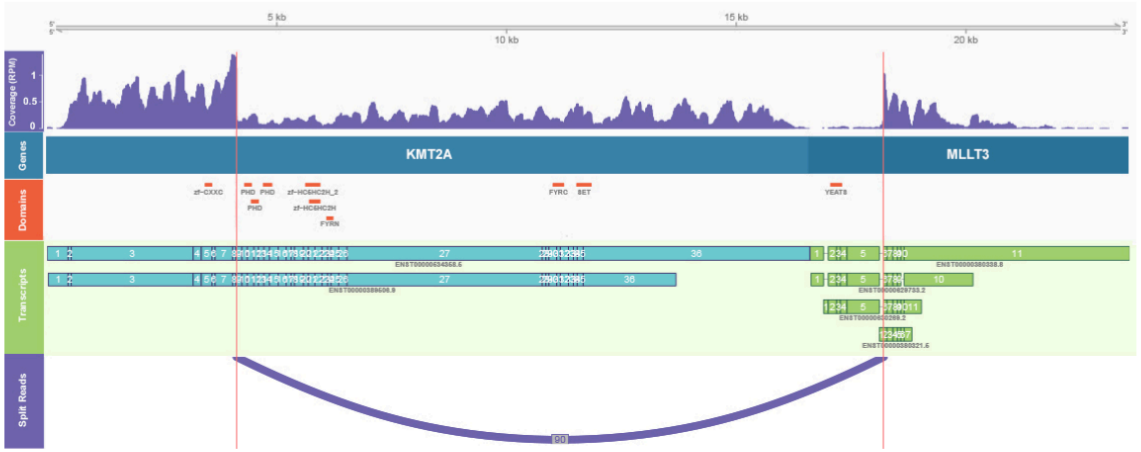
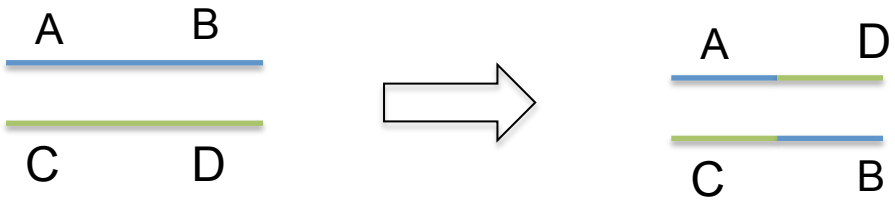


Figure 2. KMT2A-MLLT3 fusion gene visualised in IGV after alignment to the human genome (A). The backgrounds of the IGV tracks are coloured to distinguish between the coverage (purple), aligned reads (white)

Schmidt et al., bioRxiv, 2017

Prioritizing Fusion Candidates (1)

- Expression
  - Expression outlier
  - Wild type expression profile found truncated
- Recurrence
  - Same fusion pair in other samples
  - One of the fused genes is fused in other samples
- Corroborating rearrangement
  - Many balanced rearrangements are known initiators of tumorigenesis



awm

Prioritizing Fusion Candidates (2)

- Gene function
  - Implicated in cancer
  - Involve a kinase in the 3' position
  - Could serve as a drug target

- Function preserved by fusion
  - Fusion preserves reading frame of both genes

- Strength of evidence supporting the fusion call
  - Numbers of RNA-Seq fragments
  - Additional evidence from DNA? (WGS, CNV, etc…)

awm

Gene Fusion Partners

- Tyrosine kinases
  - Transfer a phosphate group from ATP to a protein
  - Transmit signals and regulate complex processes

Genes:
ABL, PDGFRA, PDGFRB, FGFR1, SYK, RET, JAK2, ALK

- Transcription factors
  - Binds to specific DNA sequences
  - Regulates transcription

Genes:
ETV1, ETV4, ETV5, ETV6, ERG, MYC

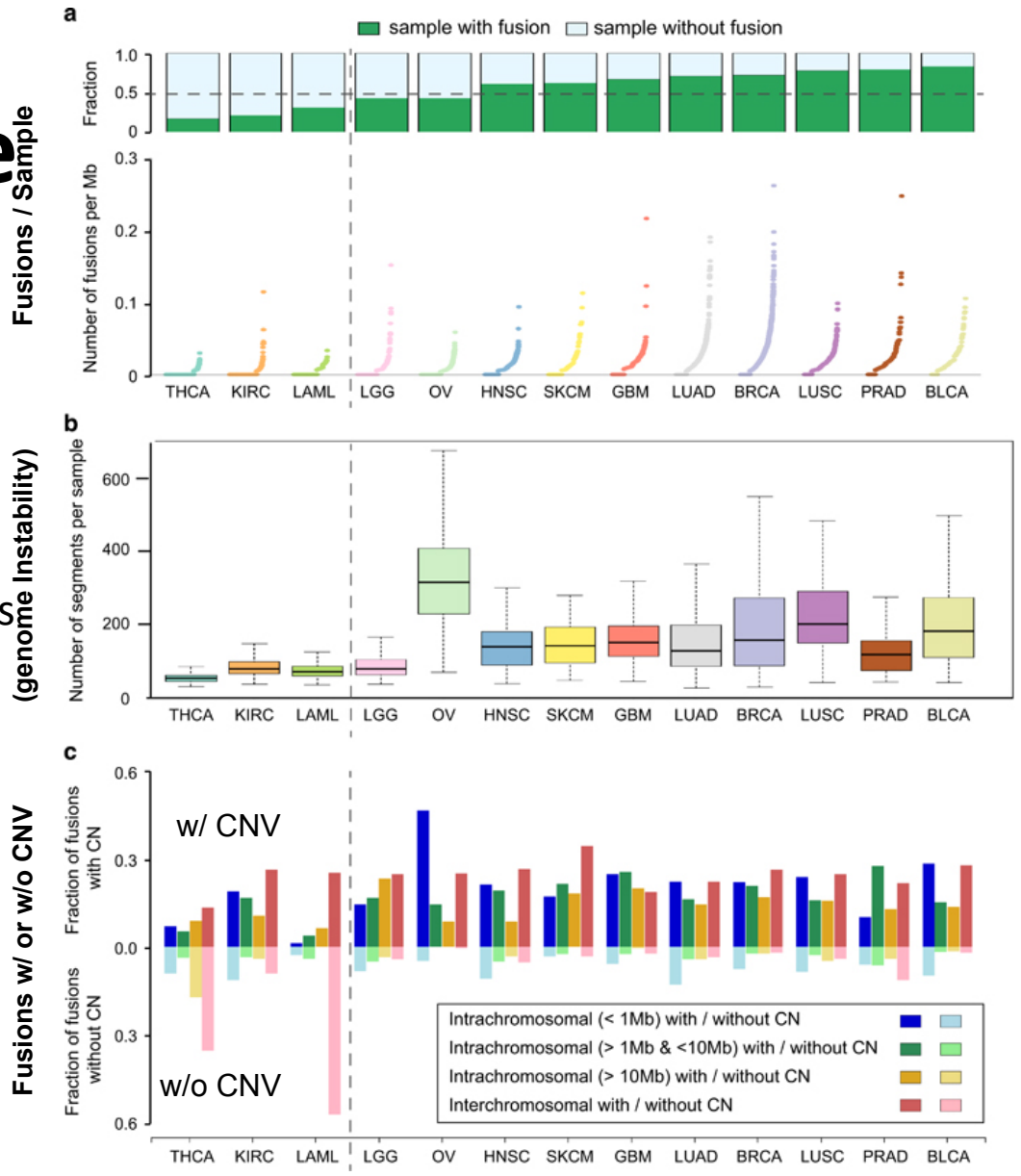
- Oncogenes
  - Regulate apoptosis
  - Promote proliferation

Genes:
BCL2, MYC, BRAF

awm

Fusion Landscape

- Landscape studies of thousands of tumours
  - Yoshihara et al., 2015
  - Stransky et al., 2014 (kinases)
- Genes fused across multiple cancers
- Mechanisms of formation
  - copy number / genome instability
  - balanced?
- Prevalence of fusions for each cancer type



awm

Yoshihara, 2015

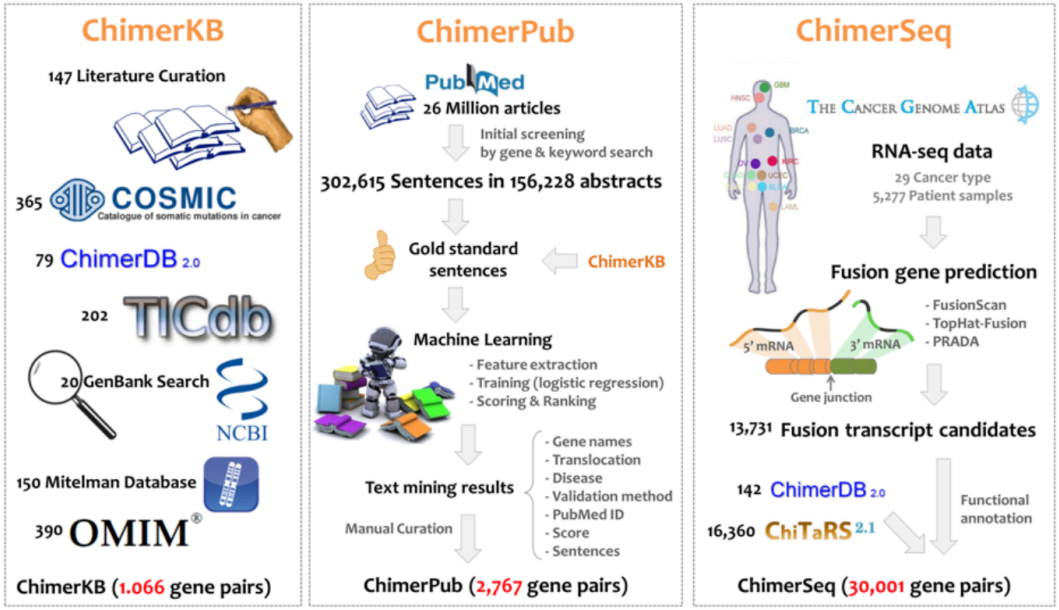
ChimerDB 3: ChimerKB, ChimerPub and ChimerSeq

ChimerDB

Home ChimerKB ChimerPub ChimerSeq Statistic Help Download

About ChimerDB

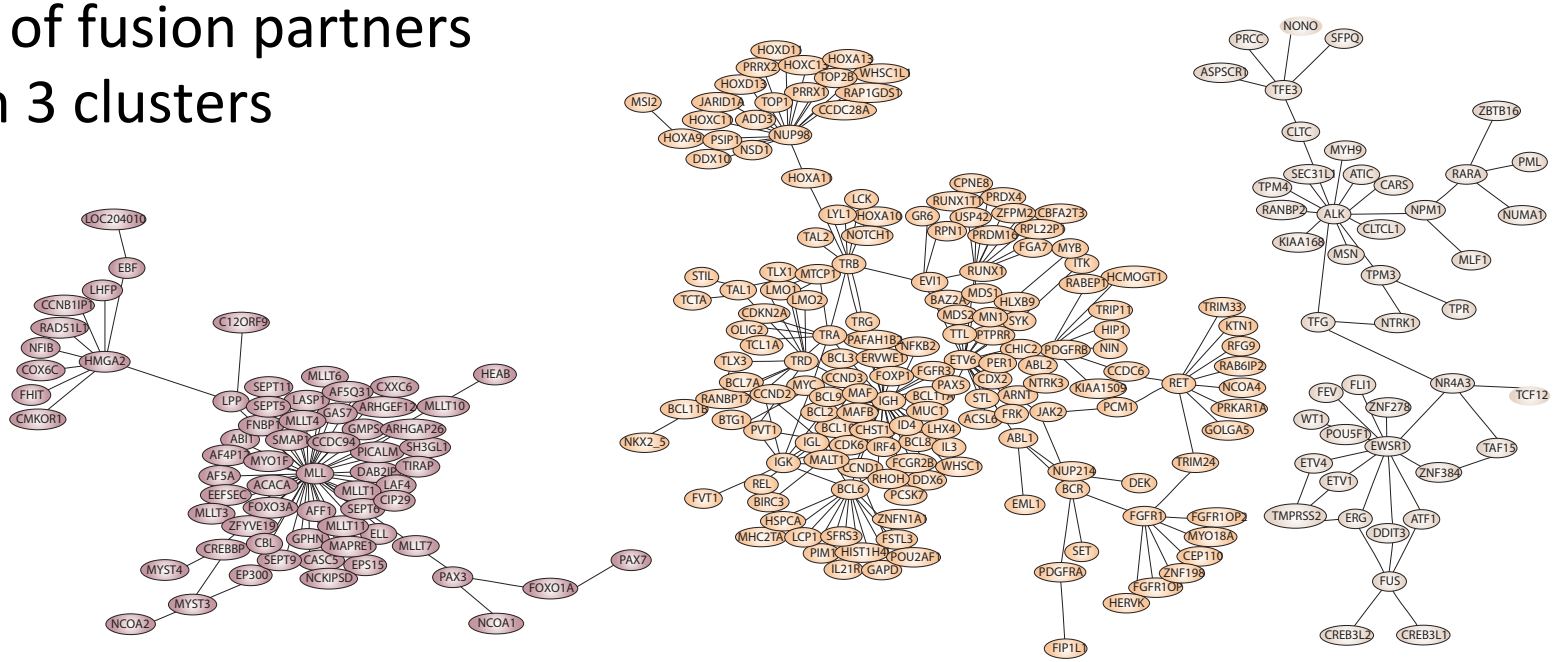
ChimerDB is a comprehensive database of fusion genes encompassing analysis of deep sequencing data and manual curations. In this update, the database coverage was enhanced considerably by adding two new modules of TCGA RNA-Seq analysis and PubMed abstract mining.



http://ercsb.ewha.ac.kr/fusiongene

Some Genes are Promiscuous Fusion Partners

- Gene fusions form a scale-free network
  - Connectivity follows a power-law
- 90% of fusion partners form 3 clusters

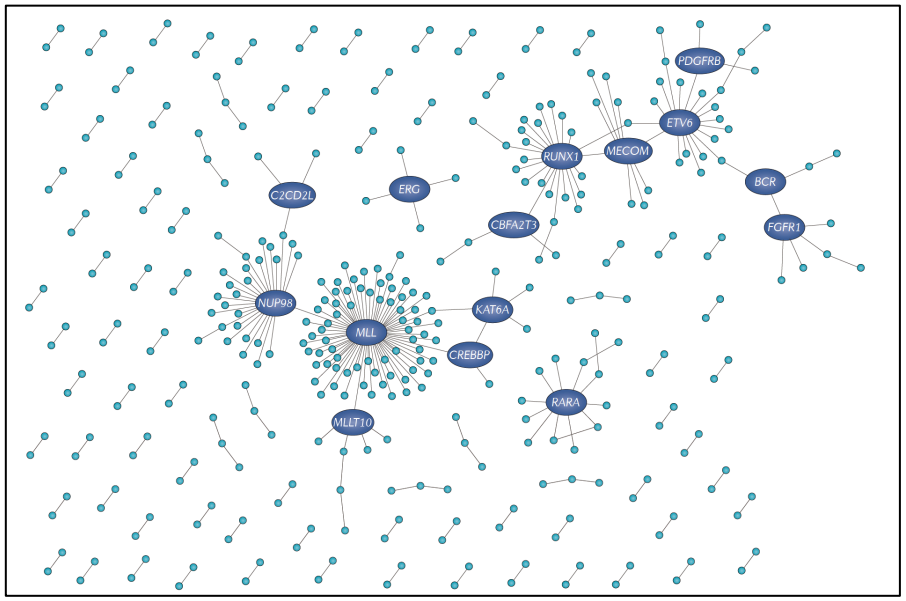


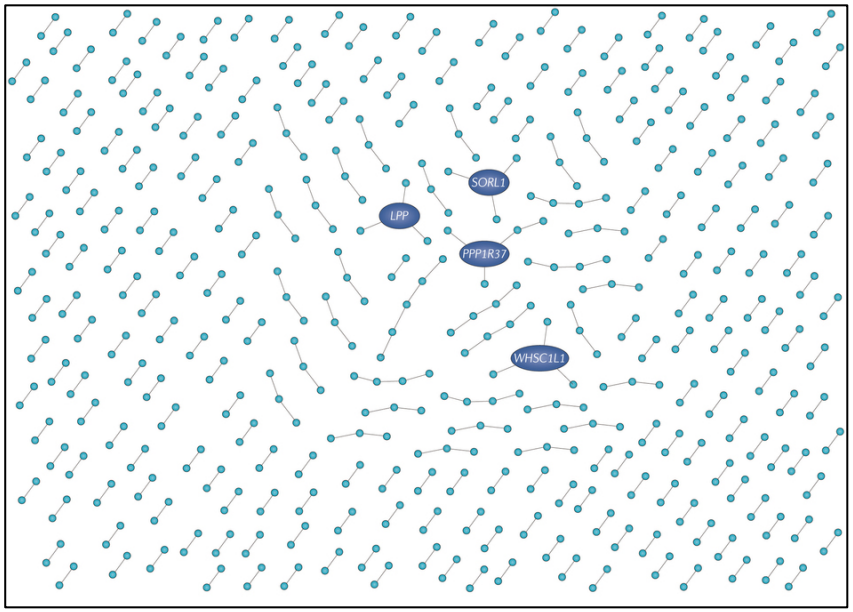
awm

Based on fusions identified pre-NGS era

Mitelman, 2007

Cancer-Specific Fusion Networks





acute myeloid leukaemia (AML)

Ovarian Cancer

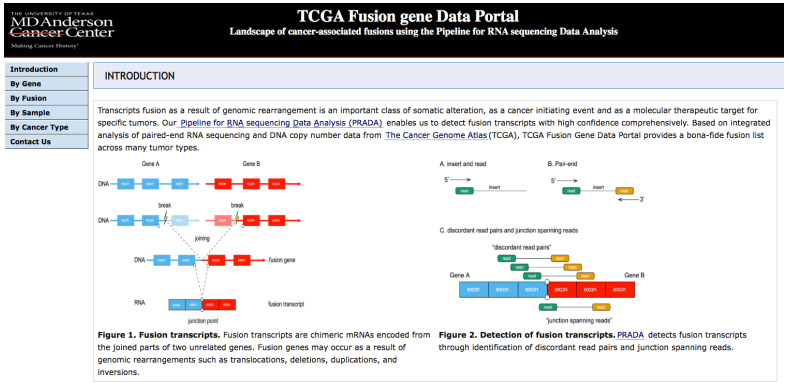
Un-clustered fusions are heavily enriched
for those identified via NGS methods
Passengers vs. Drivers? Artifacts?

Mertens, NRC 2015

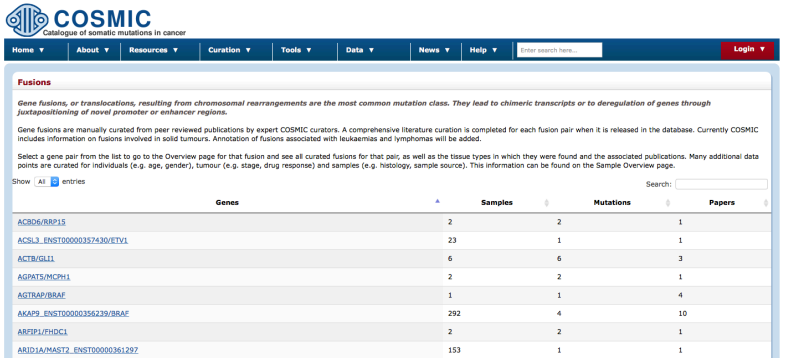
Gene Fusion Databases

- TCGA Gene Fusion portal
  - Predictions from Yoshihara et al., 2015
  - Expression and gene structure
- COSMIC gene fusions
  - Curated from publications
  - Fusion info, tissue, and publications

(Licensing required to access full DB)



http://www.tumorfusions.org



http://cancer.sanger.ac.uk/cosmic/fusion

awm

FusionAnnotator

FusionAnnotator is a simple tool for annotating fusion transcripts, incorporated into the Trinity Cancer Transcriptome Analysis Toolkit utilities, and leveraging databases including CTAT Human Fusion Lib.

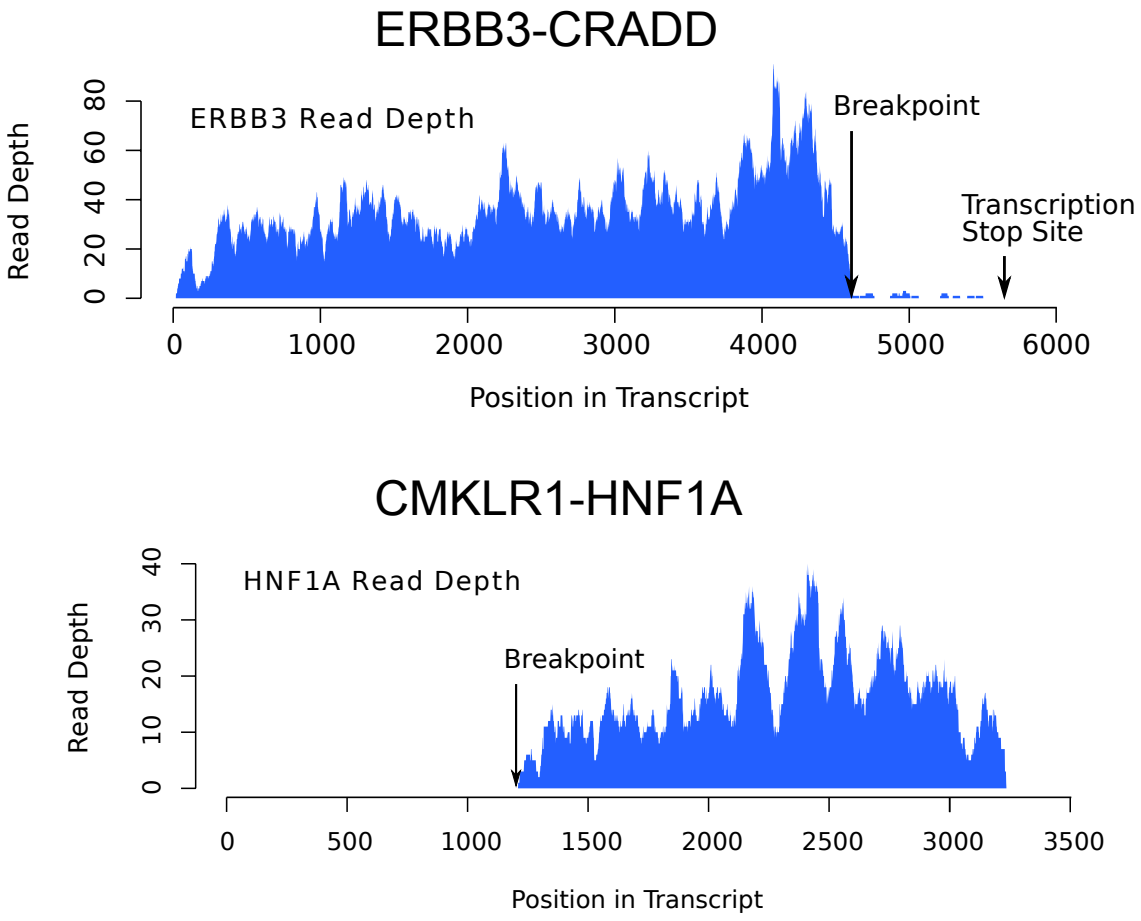
An example of fusion annotations when applying FusionAnnotator using our pre-compiled CTAT_HmanFusionLib:

```
%  FusionAnnotator --genome_lib_dir GRCh37_gencode_v19_CTAT_lib_July192017/ctat_genome_lib_b
                   --annotate fusions.list.txt
```

```
BCR--ABL1       ["chimerdb_pubmed","CCLE","Klijn_CellLines","chimerdb_omim","ChimerSeq","Chi
ACTB--ACTG1     ["ChimerSeq","INTERCHROMOSOMAL[chr7--chr17]"]
ZFP91--RAB6A    ["YOSHIHARA_TCGA","FA_CancerSupp","ChimerSeq","INTRACHROMOSOMAL[chr11:15.00M
ABCB9--ARL6IP4  ["INTRACHROMOSOMAL[chr12:0.00Mb]","NEIGHBORS_OVERLAP:-:+:[1589]"]
KMT2A--EPS15    ["ChimerKB","Cosmic","Mitelman","INTERCHROMOSOMAL[chr11--chr1]"]
PRMT1--AURKB    ["INTERCHROMOSOMAL[chr19--chr17]"]
ZNF580--ZNF510  ["HGNC_GENEFAM","INTERCHROMOSOMAL[chr19--chr9]"]
ETV6--ABL2      ["ChimerKB","ChimerPub","chimerdb_omim","INTERCHROMOSOMAL[chr12--chr1]"]
EML4--ALK       ["ChimerPub","Cosmic","ChimerKB","ChimerSeq","FA_CancerSupp","YOSHIHARA_TCGA
EWSR1--ATF1     ["ChimerSeq","ChimerKB","Cosmic","Mitelman","ChimerPub","INTERCHROMOSOMAL[ch
EWSR1--FLI1     ["CCLE","Klijn_CellLines","ChimerKB","ChimerSeq","FA_CancerSupp","ChimerPub"
TMPRSS2--ETV1   ["ChimerSeq","chimerdb_pubmed","ChimerKB","FA_CancerSupp","Cosmic","ChimerPu
ETV6--NTRK3     ["chimerdb_omim","Larsson_TCGA","YOSHIHARA_TCGA","chimerdb_pubmed","HaasMedC
CD74--ROS1      ["ChimerKB","chimerdb_pubmed","ChimerSeq","FA_CancerSupp","Klijn_CellLines",
HOOK3--RET      ["Cosmic","ChimerKB","ChimerSeq","HaasMedCancer","INTERCHROMOSOMAL[chr8--chr
EML4--ALK       ["ChimerPub","Cosmic","ChimerKB","ChimerSeq","FA_CancerSupp","YOSHIHARA_TCGA
AKAP9--BRAF     ["HaasMedCancer","chimerdb_pubmed","ChimerKB","ChimerSeq","ChimerPub","Cosmi
BRD4--NUTM1     ["Cosmic","ChimerSeq","CCLE","ChimerKB","HaasMedCancer","INTERCHROMOSOMAL[ch
FGFR3--TACC3    ["CCLE","Klijn_CellLines","YOSHIHARA_TCGA","ChimerKB","ChimerSeq","FA_Cancer
```

Localized expression imbalances result from gene fusions

- True gene fusions often associated with truncated expression in fused genes
- Expression analysis of fusion transcripts can identify true positives
- Caveat: balanced rearrangements



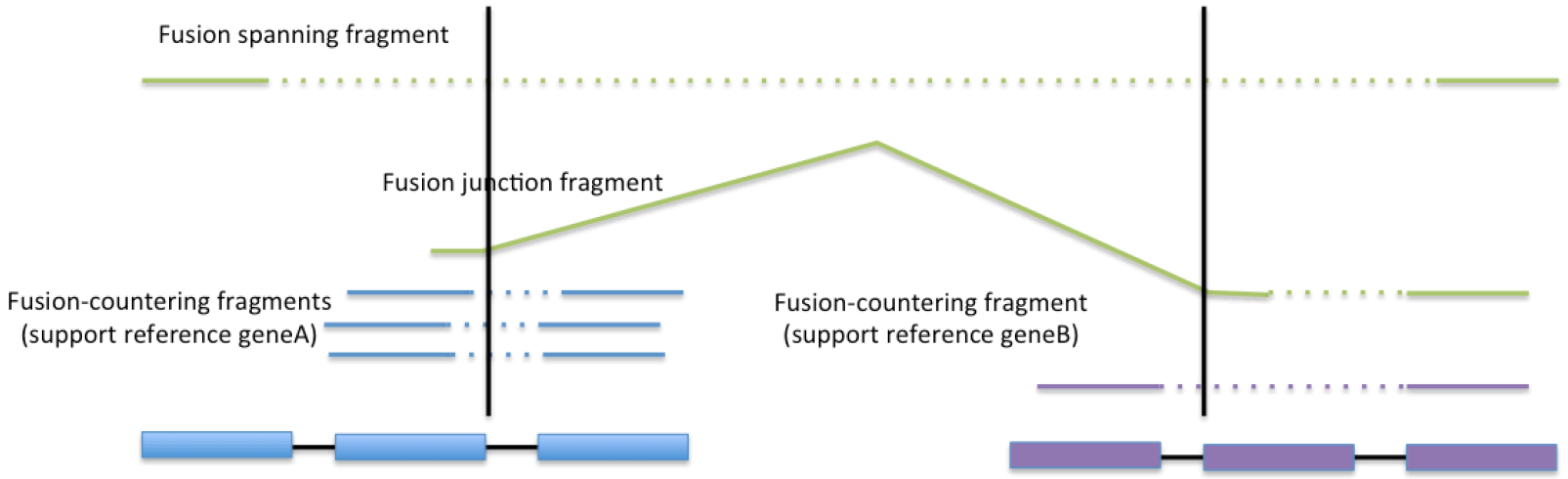
awm

McPherson et al., 2010

Fusion Allelic Ratio from FusionInspector

Insights into Fusion Expression vs. Non-Fused Allele

Computing Fusion Allelic Ratio



$$\text{Fusion Allele Ratio (geneA)} = \frac{2 + \psi}{3 + \psi}$$

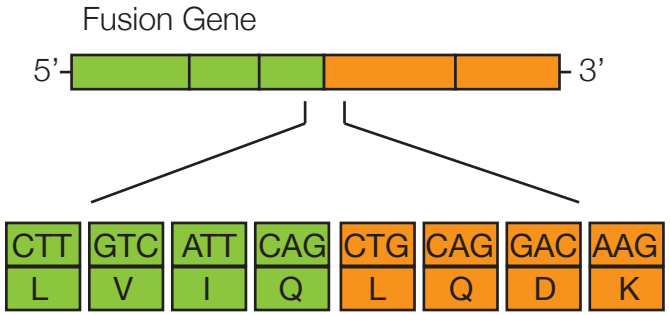
$$\text{Fusion Allele Ratio (geneB)} = \frac{2 + \psi}{1 + \psi}$$

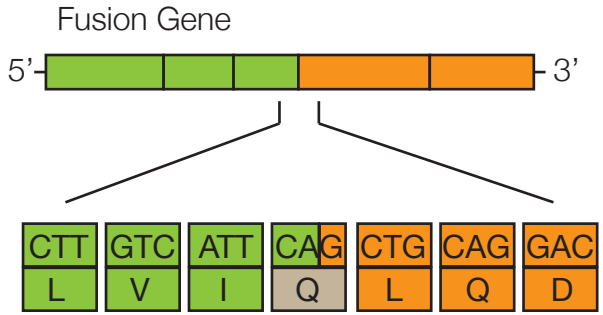
Ψ = pseudocount

(avoid zeros in denominator
and
squeeze FAR towards 1 with few reads)

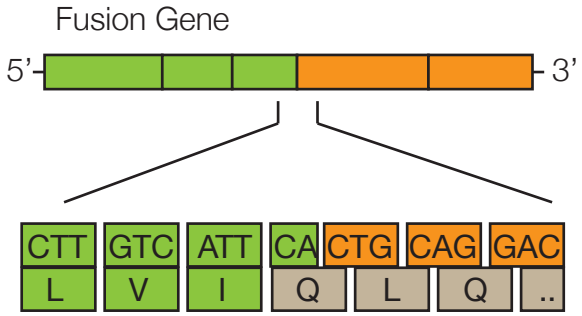
Reading Frame Preservation and 3' Gene Function

- Preserved reading frame results in at most one missense codon





- Frame shift results in nonsense 3' protein sequence



awm

STAR-Fusion –examine_coding_effect

Examine Effect of Fusions on Coding Regions

It is sometimes the case that fusion transcripts generate novel fusion proteins with altered functions. You can further explore the impact of the fusion event on coding regions by invoking the '--examine_coding_effect' parameter.

The coding effect results are appended as additional columns in the STAR-Fusion tab-delimited output file. An example set of columns include:

```
#FusionName         BCR--ABL1
...
CDS_LEFT_ID         ENST00000305877.8
CDS_LEFT_RANGE      1-2782
CDS_RIGHT_ID        ENST00000318560.5
CDS_RIGHT_RANGE     80-3393
PROT_FUSION_TYPE    INFRAME
FUSION_MODEL        chr22|+|[0]23523148-23524426[0]|[1]23595986-23596167[2]|[0]23603137-236
FUSION_CDS          atggtggacccggtgggcttcgcggaggcgtggaaggcgcagttcccggactcagagcccccgcgcatgga
FUSION_TRANSL       MVDPVGFAEAWKAQFPDSEPPRMELRSVGDIEQELERCKASIRRLEQEVNQERFRMIYLQTLLAKEKKSYD
PFAM_LEFT           Bcr-Abl_Oligo|3-75|3.6e-43^RhoGEF|503-689|1.1e-37^IQ_SEC7_PH|724-769|4.
PFAM_RIGHT          SH3_1-PARTIAL|~80-113|1.6e-15^SH3_9-PARTIAL|~80-117|1.3e-11^SH3_2-PARTI
```

Beware of Red Herrings

Red Herrings: Fusion pairs that may not be relevant to cancer, and potential false positives.

Label	Source Info
GTEx	Fusions found recurrently in GTEx as mined using STAR-Fusion and FusionInspector (not published)
BodyMap	Fusions found by STAR-Fusion as applied to the Illumina Human Body Map reference data
DGD_PARALOGS	Duplicated genes as per the Duplicated Genes Database
HGNC_GENEFAM	HGNC gene family membership as per ftp://ftp.ebi.ac.uk/pub/databases/genenames/genefam_list.txt.gz
Greger_Normal	Fusion transcripts (mostly from tandem genes) detected based on analysis of RNA-Seq from 1000 genomes project samples. List derived from Greger et al. PLOS One, 2014
Babiceanu_Normal	Recurrent chimeric fusion RNAs in non-cancer tissues and cells as per Babiceanu et al. NAR, 2016
ConjoinG	Fused transcripts derived from the Conjoined Gene Database

Caveats to Red Herrings

- Why might BCR—ABL show up in GTEx?

  (note, haven't checked the very latest GTEx release yet… but it used to be there!)

- Mimicry or 'Normal' fusions in certain contexts?

A Neoplastic Gene Fusion Mimics Trans-Splicing of RNAs in Normal Human Cells

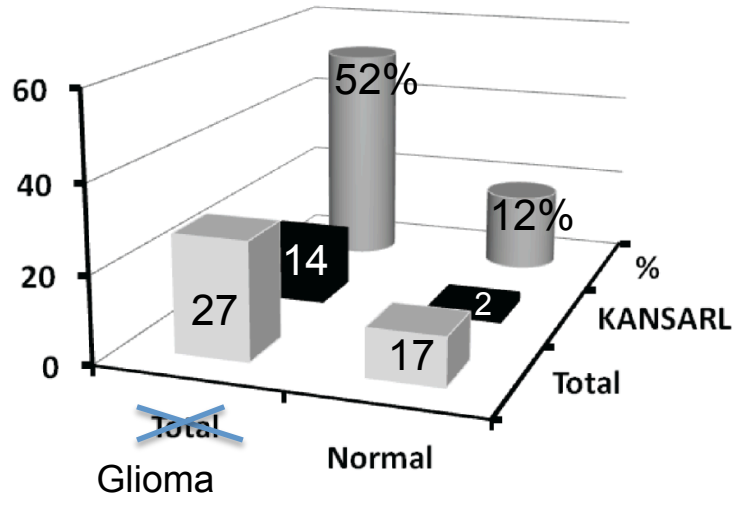
Hui Li,[1] Jinglan Wang,[1] Gil Mor,[2] Jeffrey Sklar[1]*

Chromosomal rearrangements that create gene fusions are common features of human tumors. The prevailing view is that the resultant chimeric transcripts and proteins are abnormal, tumor-specific products that provide tumor cells with a growth and/or survival advantage. We show that normal endometrial stromal cells contain a specific chimeric RNA joining 5′ exons of the *JAZF1* gene on chromosome 7p15 to 3′ exons of the Polycomb group gene *JJAZ1/SUZ12* on chromosome 17q11 and that this RNA is translated into JAZF1-JJAZ1, a protein with anti-apoptotic activity. The *JAZF1-JJAZ1* RNA appears to arise from physiologically regulated trans-splicing between precursor messenger RNAs for *JAZF1* and *JJAZ1*. The chimeric RNA and protein are identical to those produced from a gene fusion found in human endometrial stromal tumors. These observations suggest that certain gene fusions may be pro-neoplastic owing to constitutive expression of chimeric gene products normally generated by trans-splicing of RNAs in developing tissues.

Li et al., Science 2008

The semi-ubiquitous KANSL1-ARL17A fusion

- Found in ~30% of those with European Ancestry
- Statistically enriched in Glioma samples



Zhou et al., Oncotarget. 2017

```
m = matrix(c(14,13,15,2), byrow=T, ncol=2)
fisher.test(m)$p.value
```
P-value = 0.02

Hmm….
What if there were 3 kansarl&normal instead of 2?

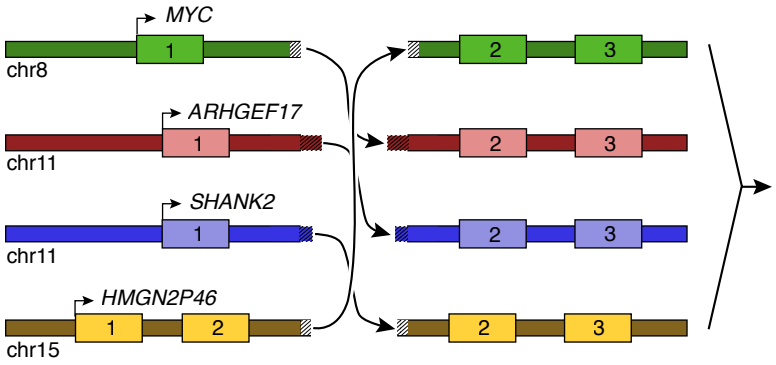
```
m = matrix(c(14,13,14,3), byrow=T, ncol=2)
fisher.test(m)$p.value
```
P-value = 0.06

I'm not convinced!
Let's look at more samples…

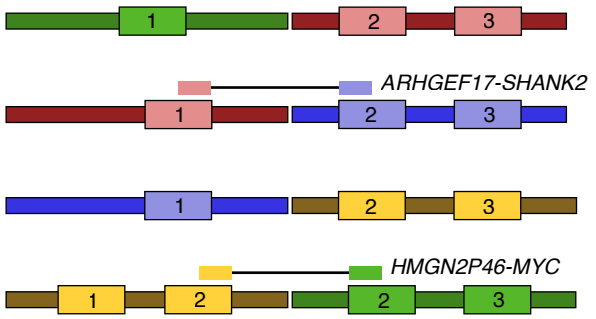
Genomic & Transcriptomic Data Maximize Insights into Cancer Fusions

Complex fusions revealed via breakpoint graphs:

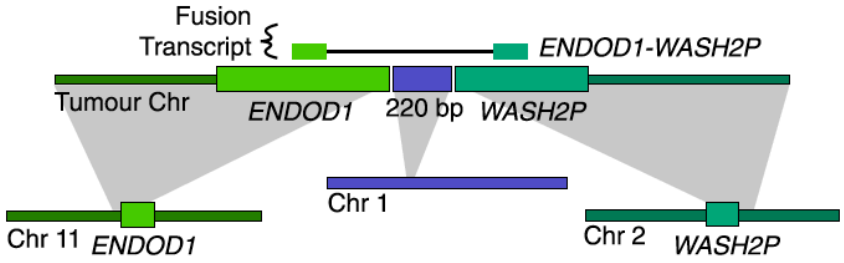
Complex balanced rearrangement in a prostate tumour

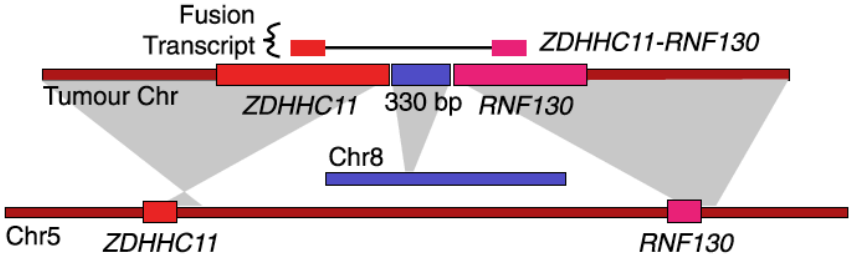


ARHGEF17-SHANK2 and HMGN2P46-MYC fusions



Fusion Transcripts Inform the Biology





McPherson et al., 2012

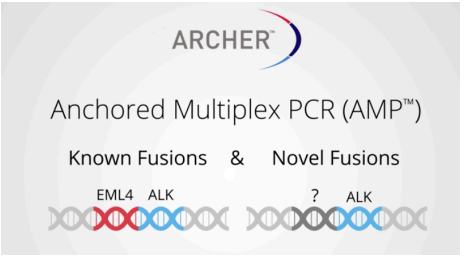
awm

Experimental Design Considerations

- Large cohort size
  - Prohibitive computational cost using some methods
  - Simpler method, filter recurrent artifacts

- Fusion partners known
  - Consider targeted capture

For example:



- How sensitive do you need to be?
  - How many millions of reads?
  - Long reads? Paired ends?

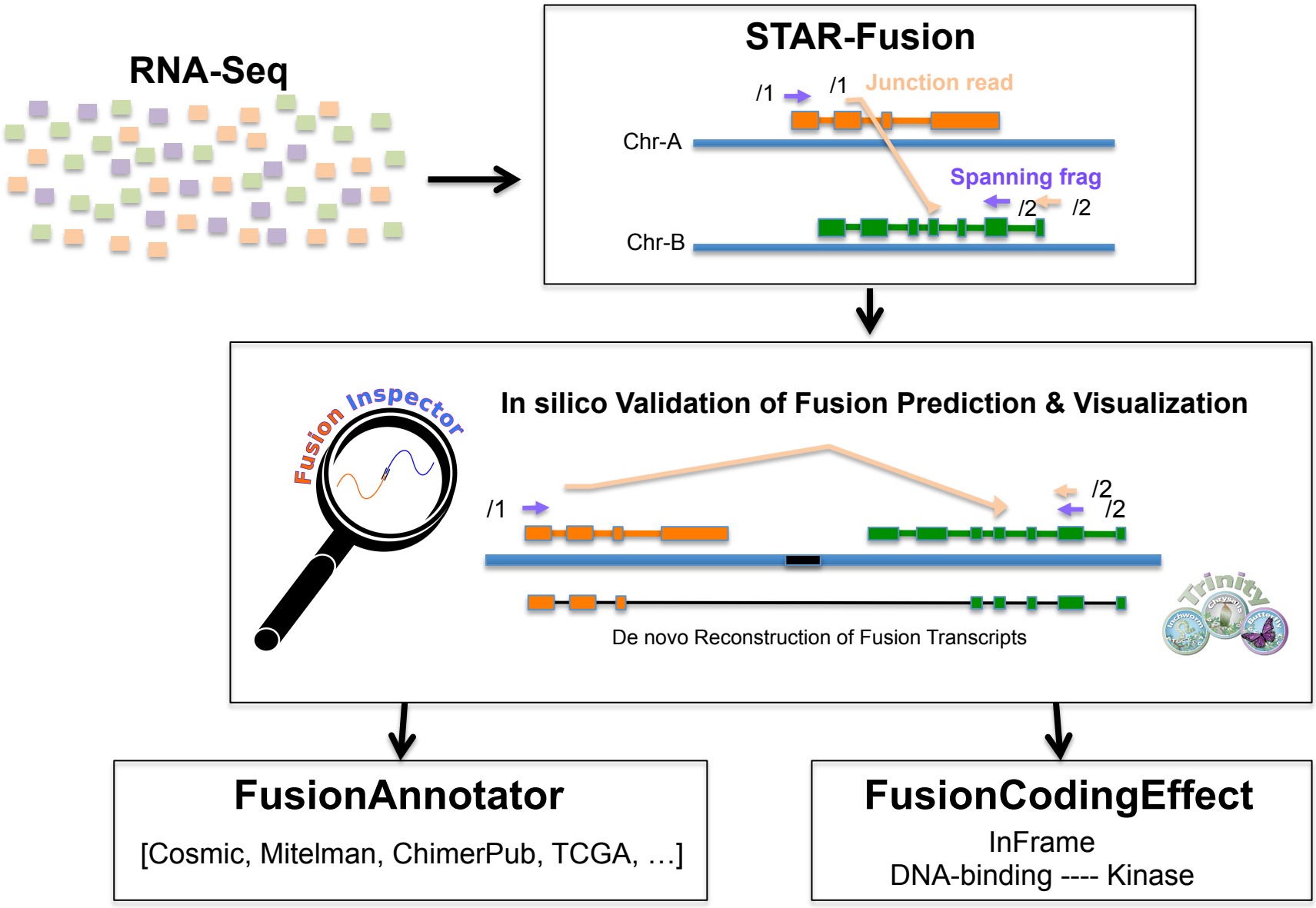
awm

Experimental Design Considerations

- Artifacts are prevalent
  - Leverage multiple computational methods
    - Increased confidence for predictions nominated by multiple methods
    - Use orthogonal techniques (assembly / reference based)
  - Spot-check individual events
    - Manually check supporting reads
  - Always validate predictions
    - Validate a random sampling of each class of event for large sets of predictions

awm

Fusion Discovery and Analysis via Trinity CTAT



Time for Fusion Lab