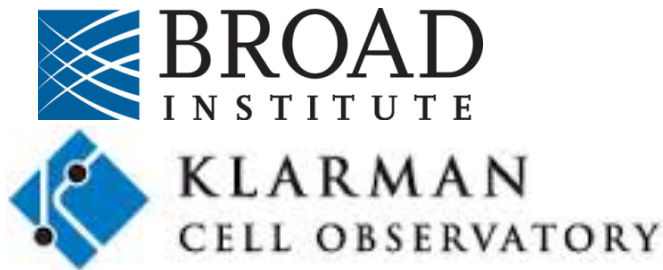




Trinity: Transcriptome Assembly for Genetic and Functional Analysis of Cancer [U24]



PI: Aviv Regev
Co-Investigator: Brian Haas

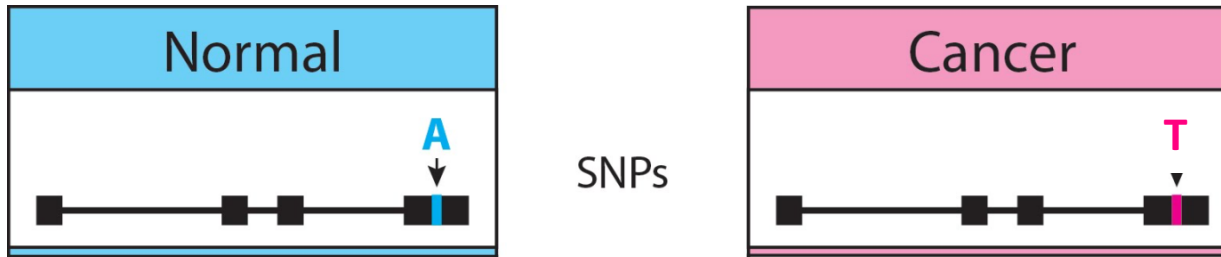
Co-PI: Tom Doak

The Cancer Transcriptome

A window into the (expressed) genetic and epigenetic state of a tumor

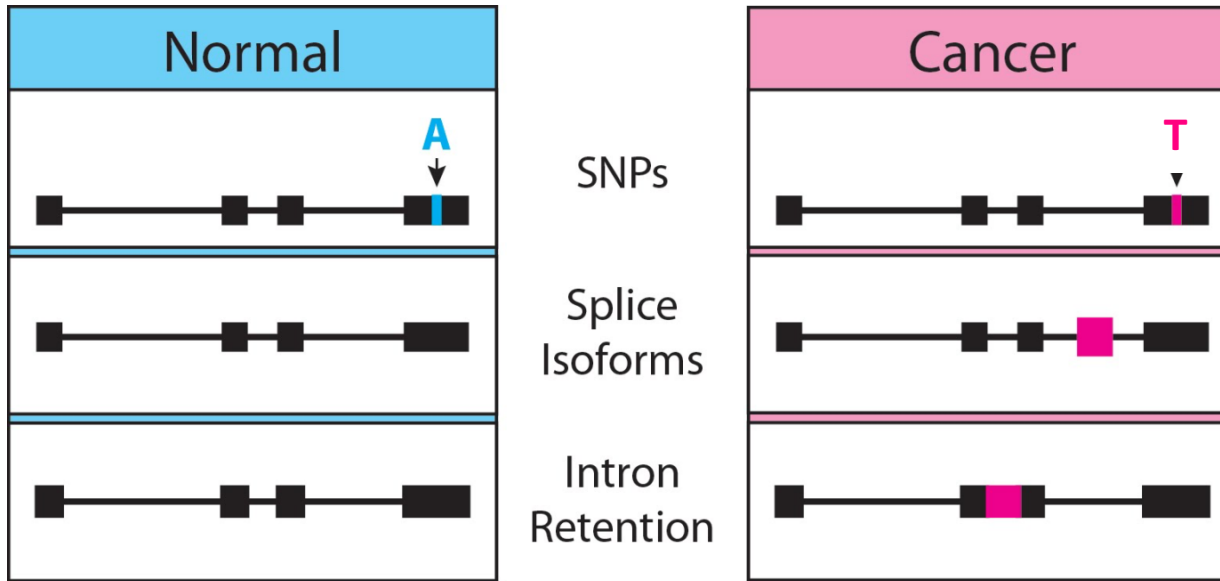
The Cancer Transcriptome

A window into the (expressed) genetic and epigenetic state of a tumor



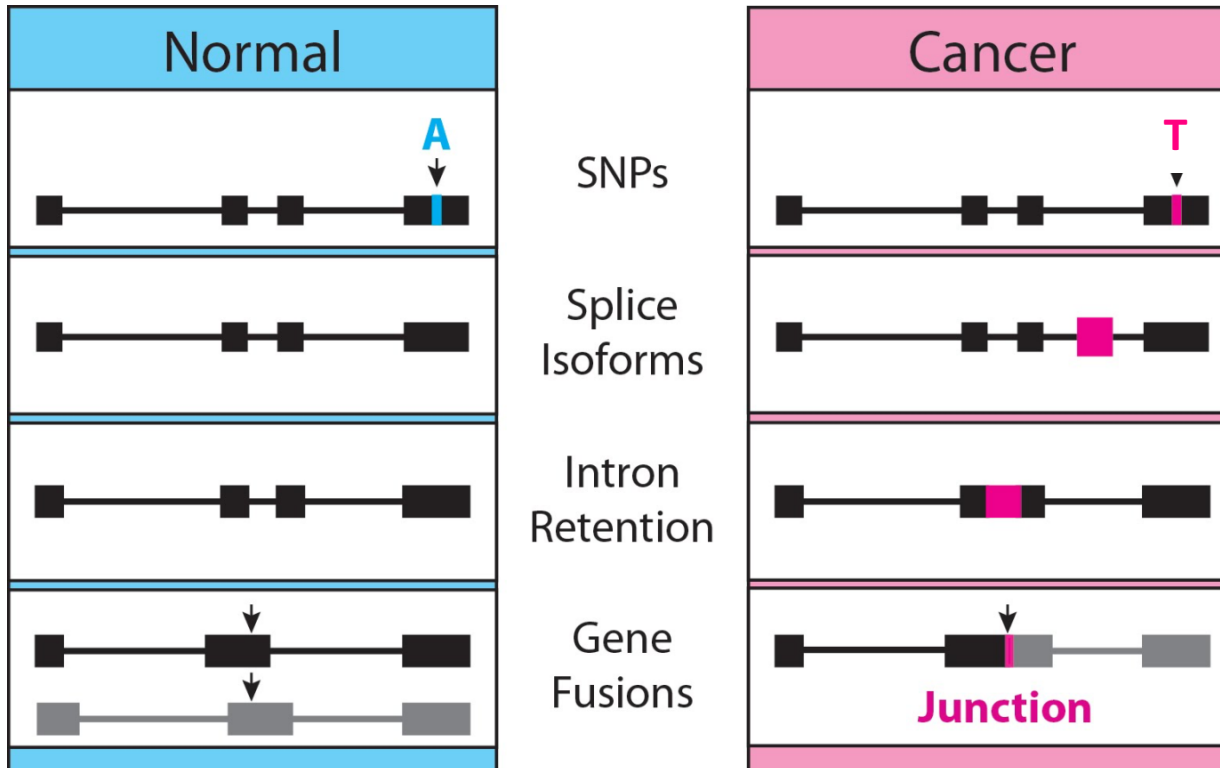
The Cancer Transcriptome

A window into the (expressed) genetic and epigenetic state of a tumor



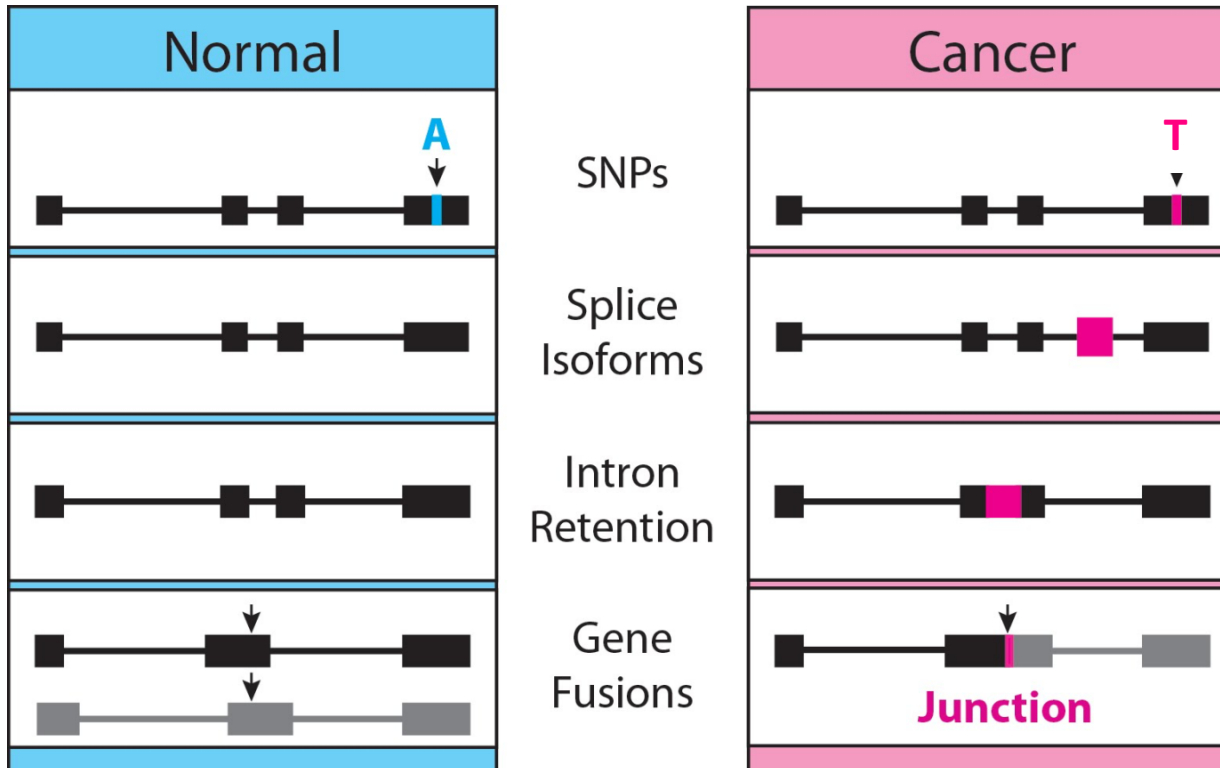
The Cancer Transcriptome

A window into the (expressed) genetic and epigenetic state of a tumor



The Cancer Transcriptome

A window into the (expressed) genetic and epigenetic state of a tumor



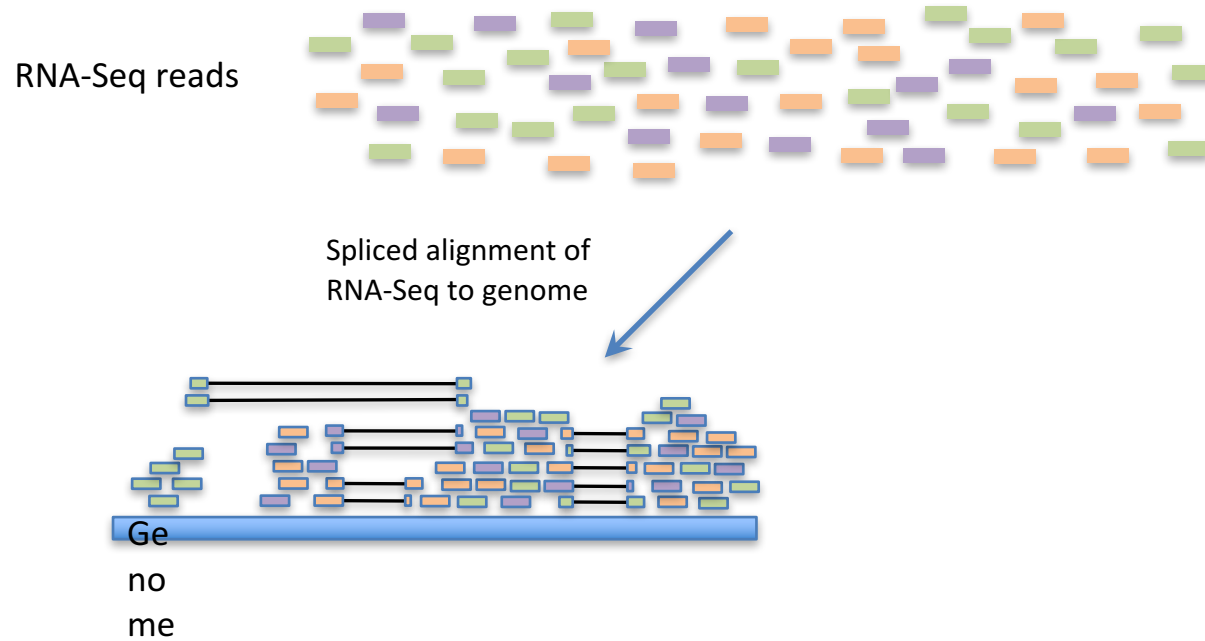
+ the associated microbiome, virome...

Contemporary strategies for transcript analysis from RNA-Seq

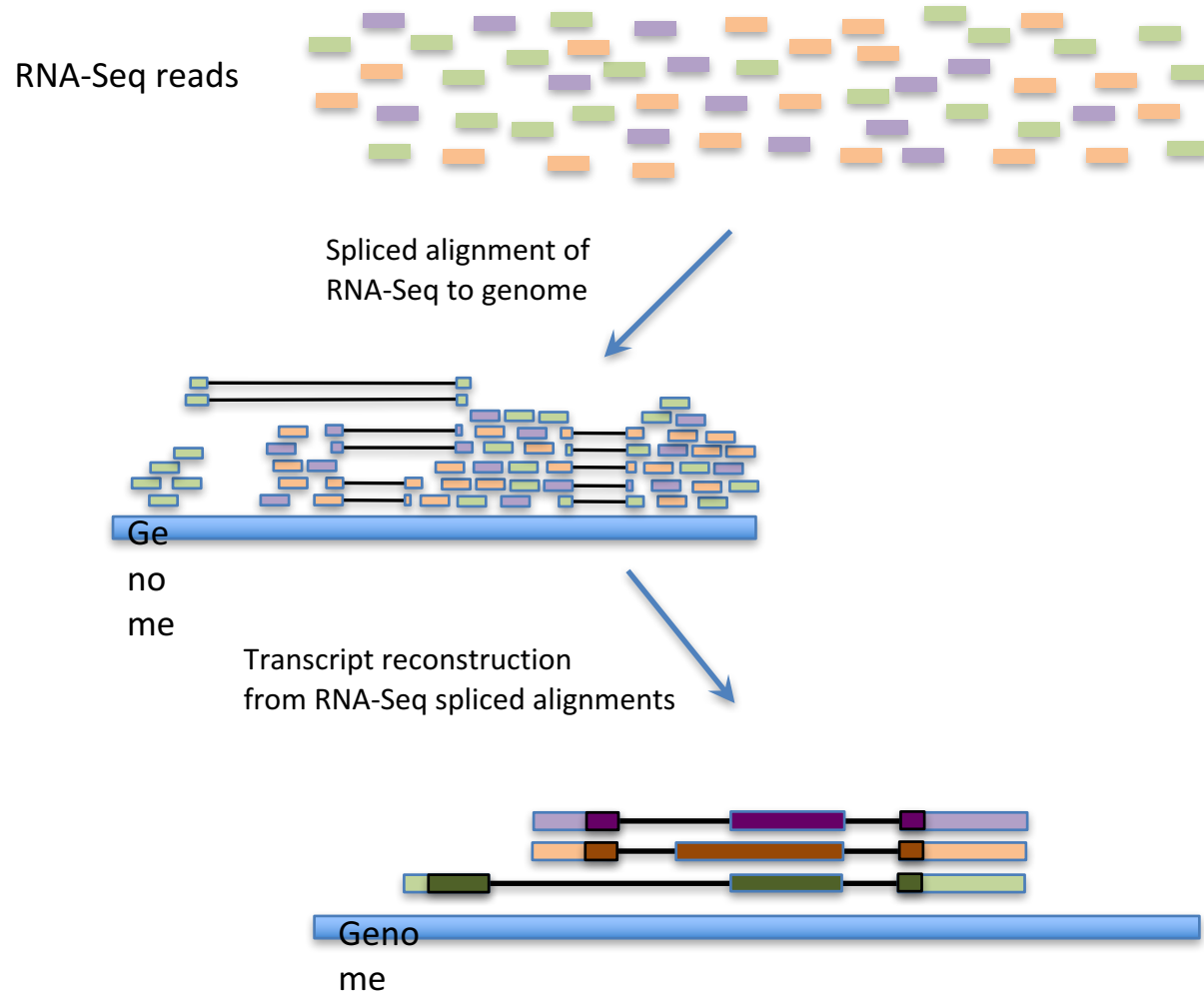


Two paradigms for transcriptome Analysis

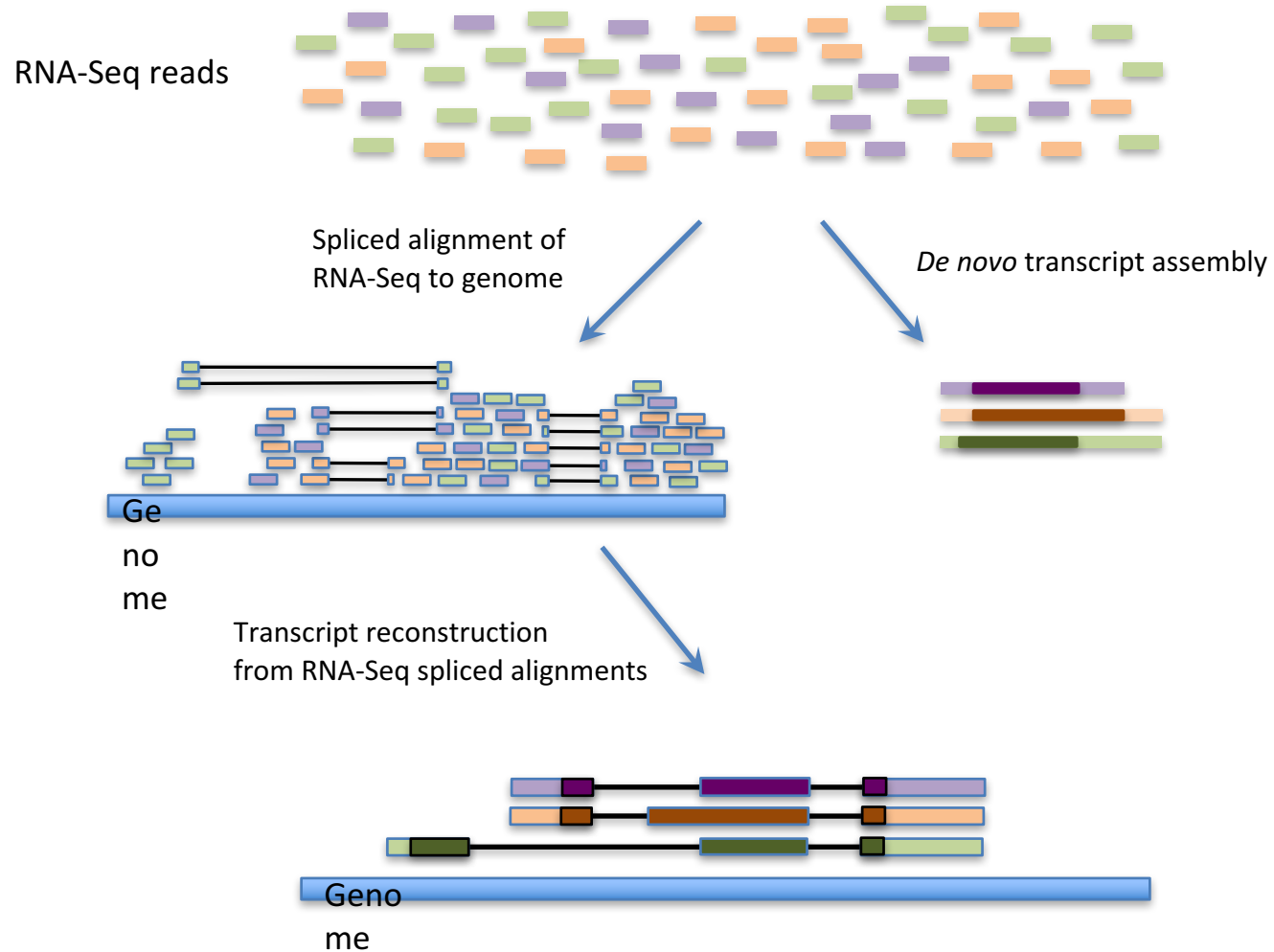
Contemporary strategies for transcript analysis from RNA-Seq



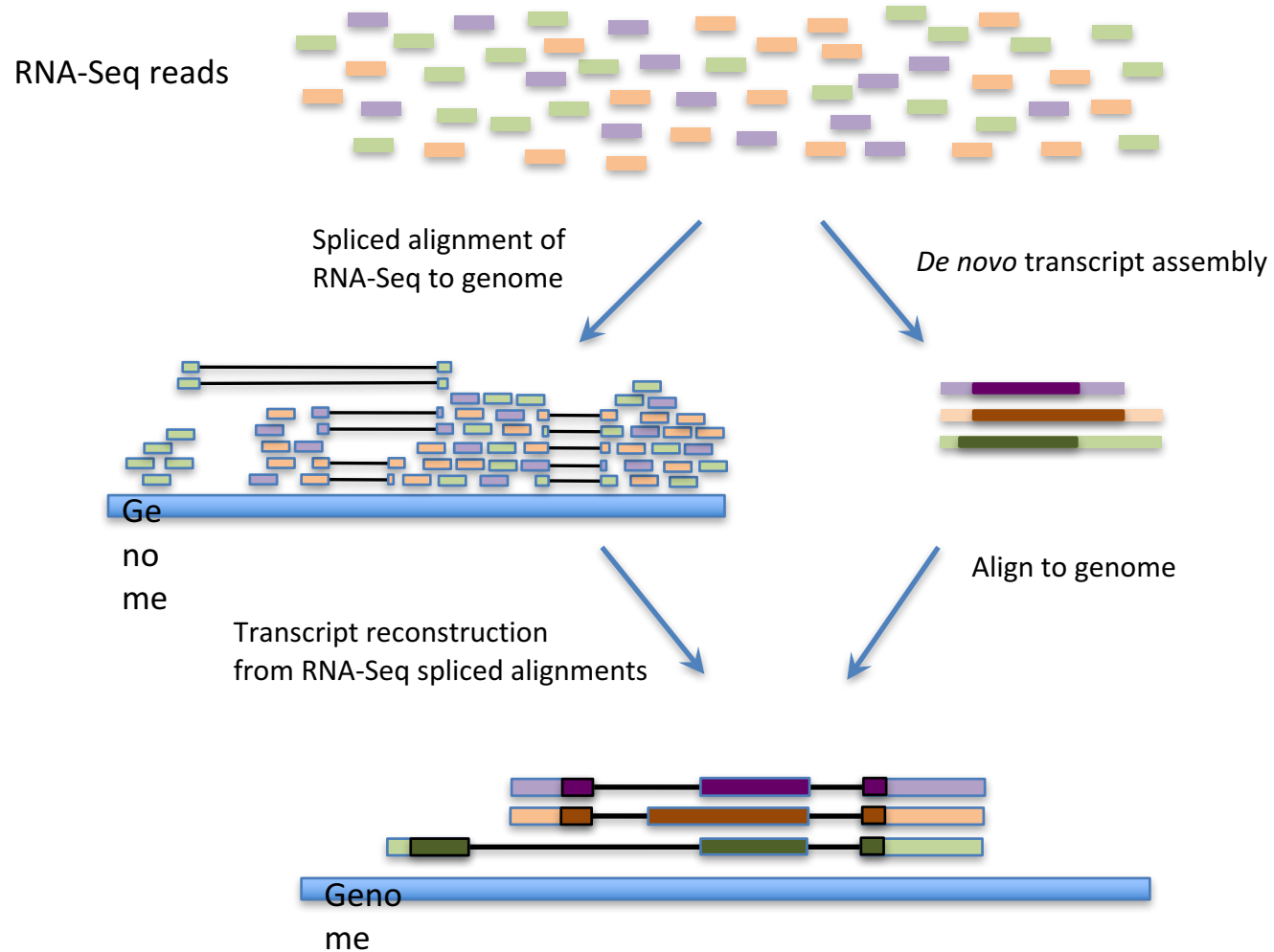
Contemporary strategies for transcript analysis from RNA-Seq



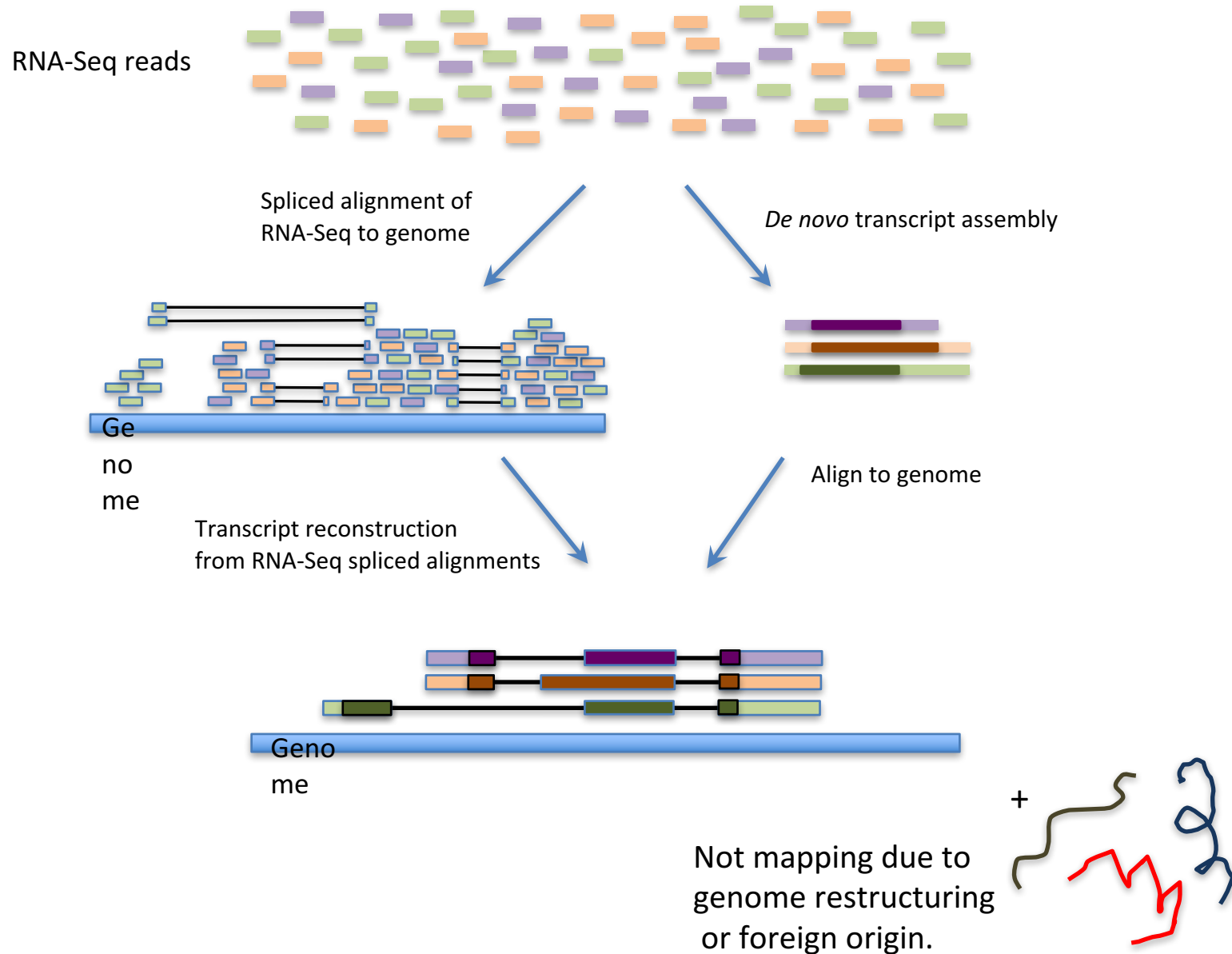
Contemporary strategies for transcript analysis from RNA-Seq



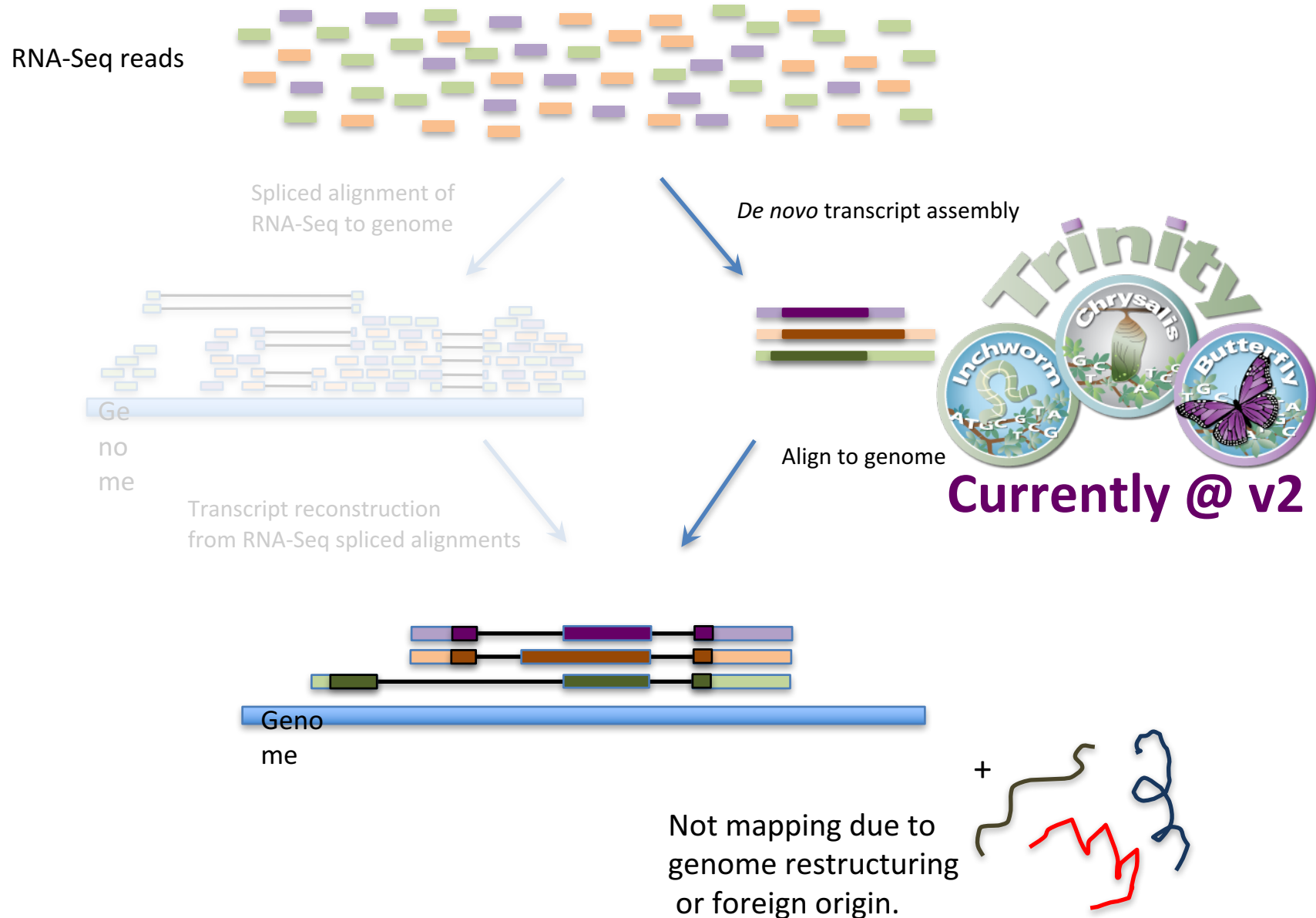
Contemporary strategies for transcript analysis from RNA-Seq



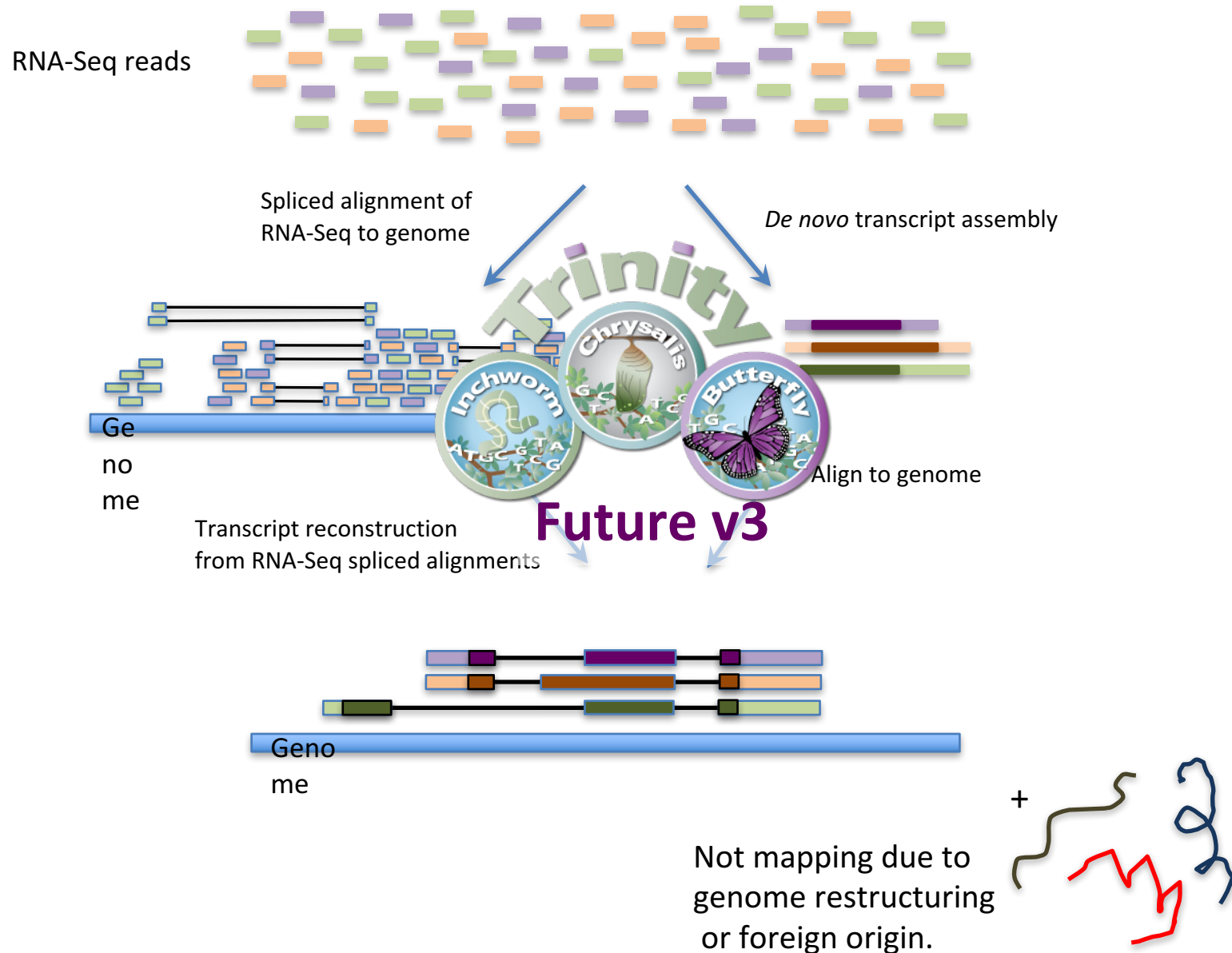
Contemporary strategies for transcript analysis from RNA-Seq



Contemporary strategies for transcript analysis from RNA-Seq



Contemporary strategies for transcript analysis from RNA-Seq



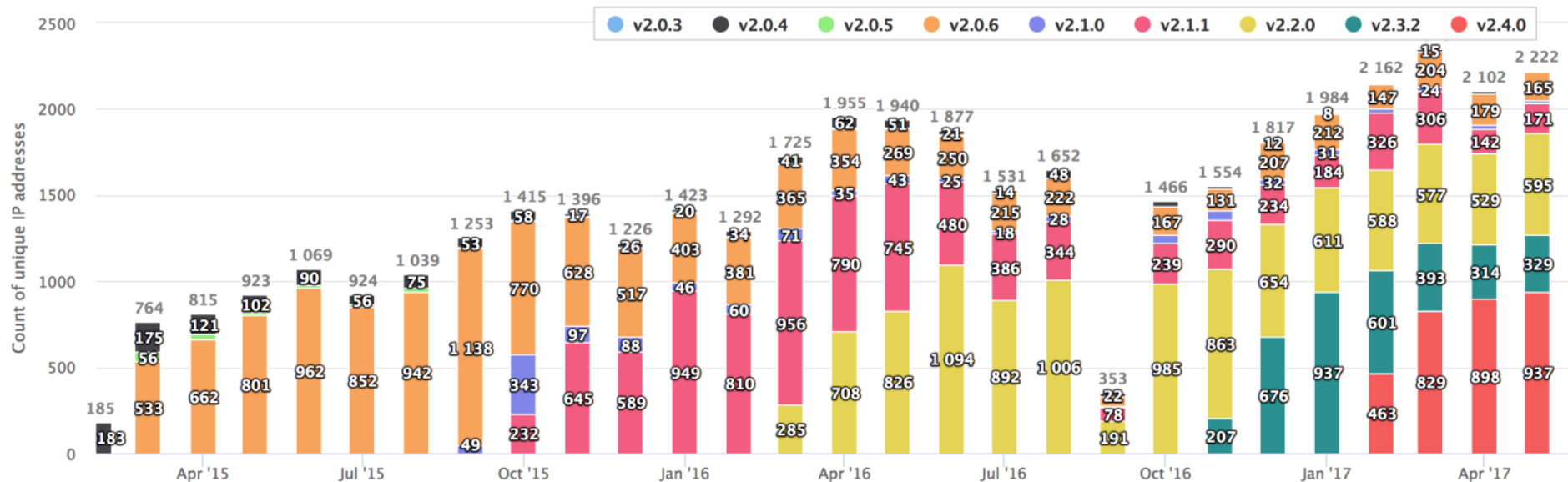
The Ever-Growing Trinity User Community



- ~2k unique users per month
- >4k literature citations (~20% cancer community)
- Open Source software development contributions from the Trinity community.

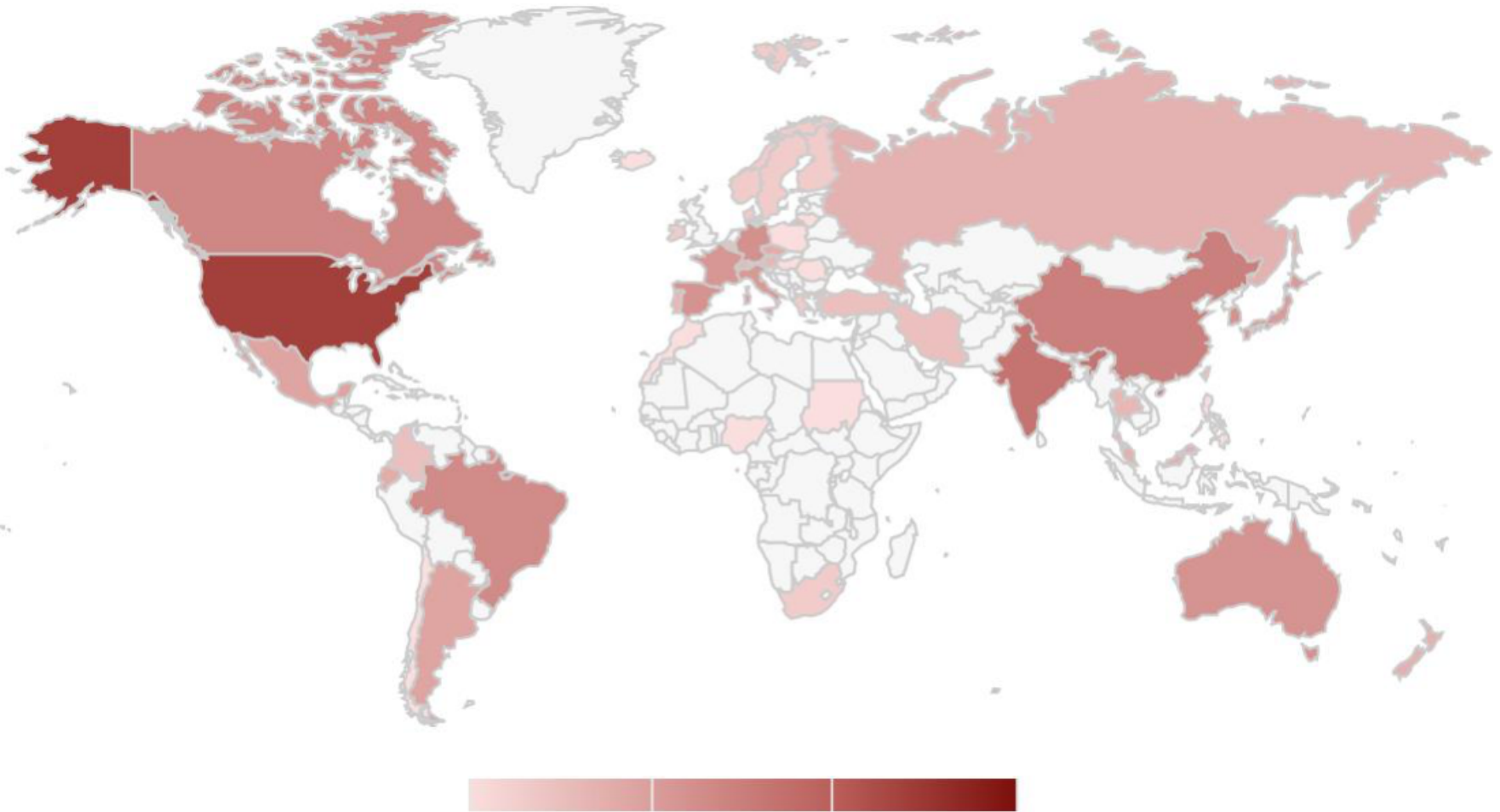


Trinity Usage Tracked by Unique IP Address



Trinity Galaxy Users 2017

Use at 486 institutions in 51 countries



User support and training:

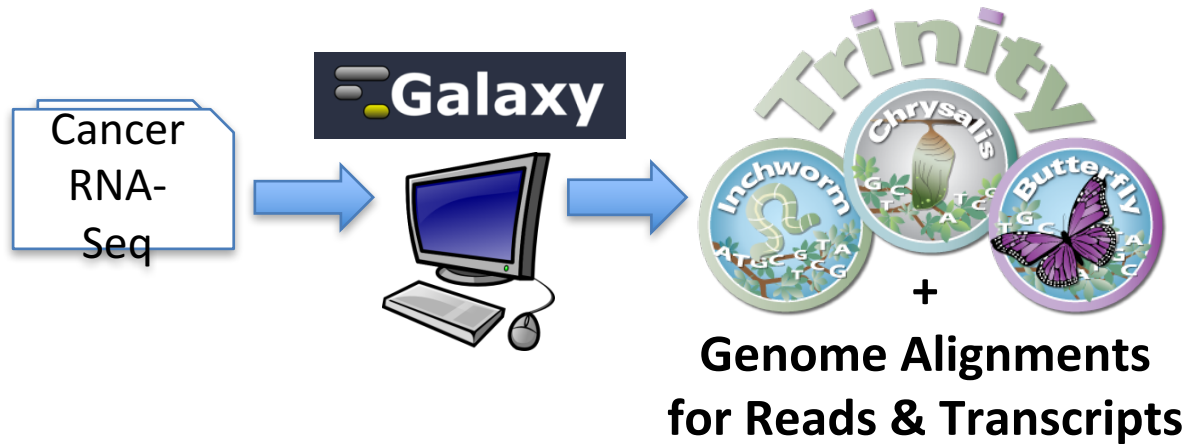
- Google group and Twitter feed for community interaction and support.
- Extensive documentation, user guides, tutorials and protocols
- Demo and training videos
- On-site training workshops



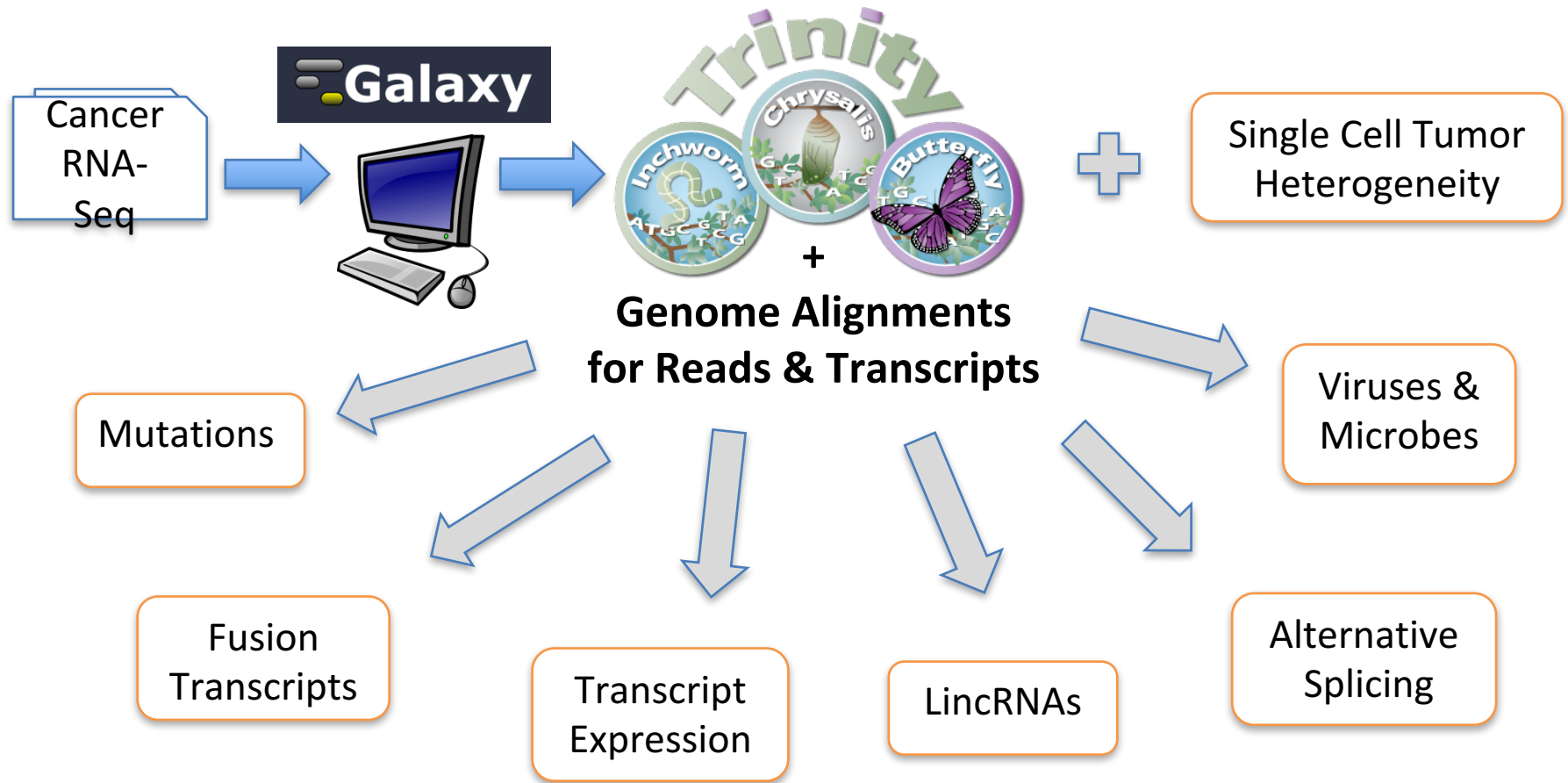
Cancer Transcriptome Analysis Toolkit (CTAT)

Goal: to assist cancer researchers in applying
RNA-Seq to genetic and functional
analyses of cancer

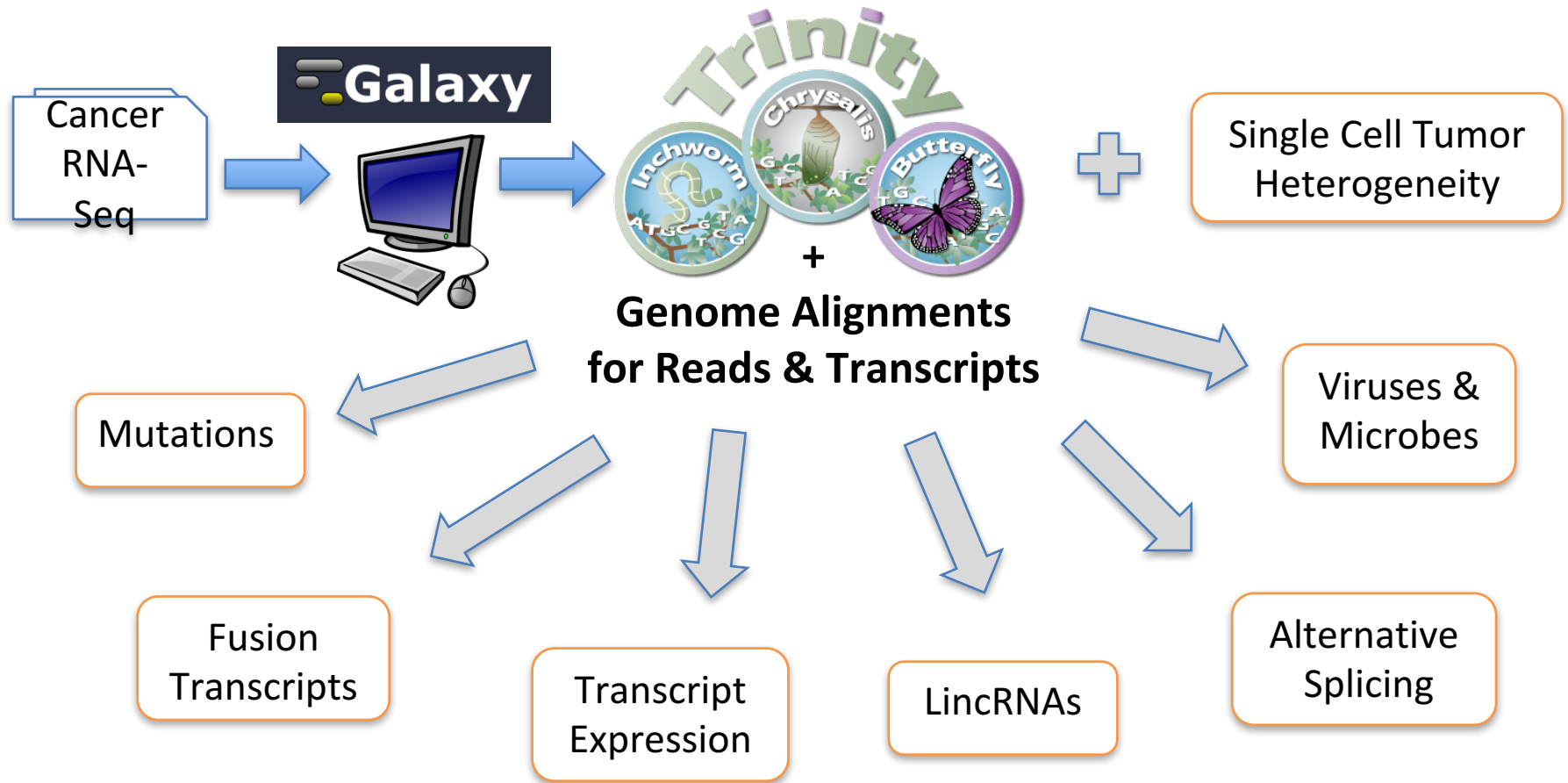
Trinity Cancer Transcriptome Analysis Toolkit



Trinity Cancer Transcriptome Analysis Toolkit



Trinity Cancer Transcriptome Analysis Toolkit



Interactive Visualizations and Summary Reports



Trinity CTAT Available Through Galaxy via NCGAS at Indiana University

The screenshot shows the Trinity Galaxy web interface. On the left, a sidebar lists various tools under categories like 'Import Data', 'Helper Tools', 'Trinity CTAT Assembly Tools', 'Trinity CTAT Mutation Tools', 'Trinity CTAT Fusion Tools', 'Fusion Inspector', 'Star Fusion', 'DISCASM', 'GMAP-fusion', 'Fasta Tools', and 'Workflows'. Three boxes with arrows point to specific tools: 'Transcriptome Assembly' points to 'Trinity CTAT Assembly Tools', 'Mutation Detection' points to 'Trinity CTAT Mutation Tools', and 'Fusion Detection' points to 'Trinity CTAT Fusion Tools'. The main content area features a banner for 'Trinity National Center for Genome Analysis Support', a welcome message, a maintenance notice, and a tweets section. The right sidebar shows a 'History' panel with a list of datasets.

Transcriptome Assembly

Mutation Detection

Fusion Detection

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 11.3 MB

Tools

search tools

Import Data

Helper Tools

Trinity CTAT Assembly Tools

Trinity CTAT Mutation Tools

Trinity CTAT Fusion Tools

Fusion Inspector In silico Validation of Fusion Transcript Predictions

Star Fusion Fusion-finding Pipeline Using the STAR Aligner

DISCASM Discordant and Unmapped Read De novo Transcriptome Assembly

GMAP-fusion Fusion-finding Using De novo RNA-Seq Transcript Assemblies

Fasta Tools

Workflows

All workflows

Trinity National Center for Genome Analysis Support Indiana University Pervasive Technology Institute

Welcome to the Trinity Galaxy Instance

Thank you for choosing Galaxy!

Get started with some help [moving files into Galaxy](#). Feel free to visit our [FAQ page](#) for additional information.

We are committed to helping you succeed with your research. Don't hesitate to contact help@ncgas.org if you need help or if you have questions or suggestions.

This Galaxy instance is running on hardware that is scheduled to be unavailable on the first Tuesday of every month for maintenance. Jobs that are started before this time will resume after maintenance.

Tweets

Follow

Trinity_CTAT @Trinity_CTAT 27 Nov

Trinity Galaxy will be shutting down for updates this weekend. Stay tuned for new content after maintenance this Tuesday, December 1st!

Trinity_CTAT @Trinity_CTAT 13 Nov

Trinity Galaxy will be down over the weekend for updates. We are adding new tools to the Galaxy interface for detecting variants!

History

search datasets

Trinity small test 4 shown 11.31 MB

4: Trinity on data 1 and data 2: Assembled Transcripts

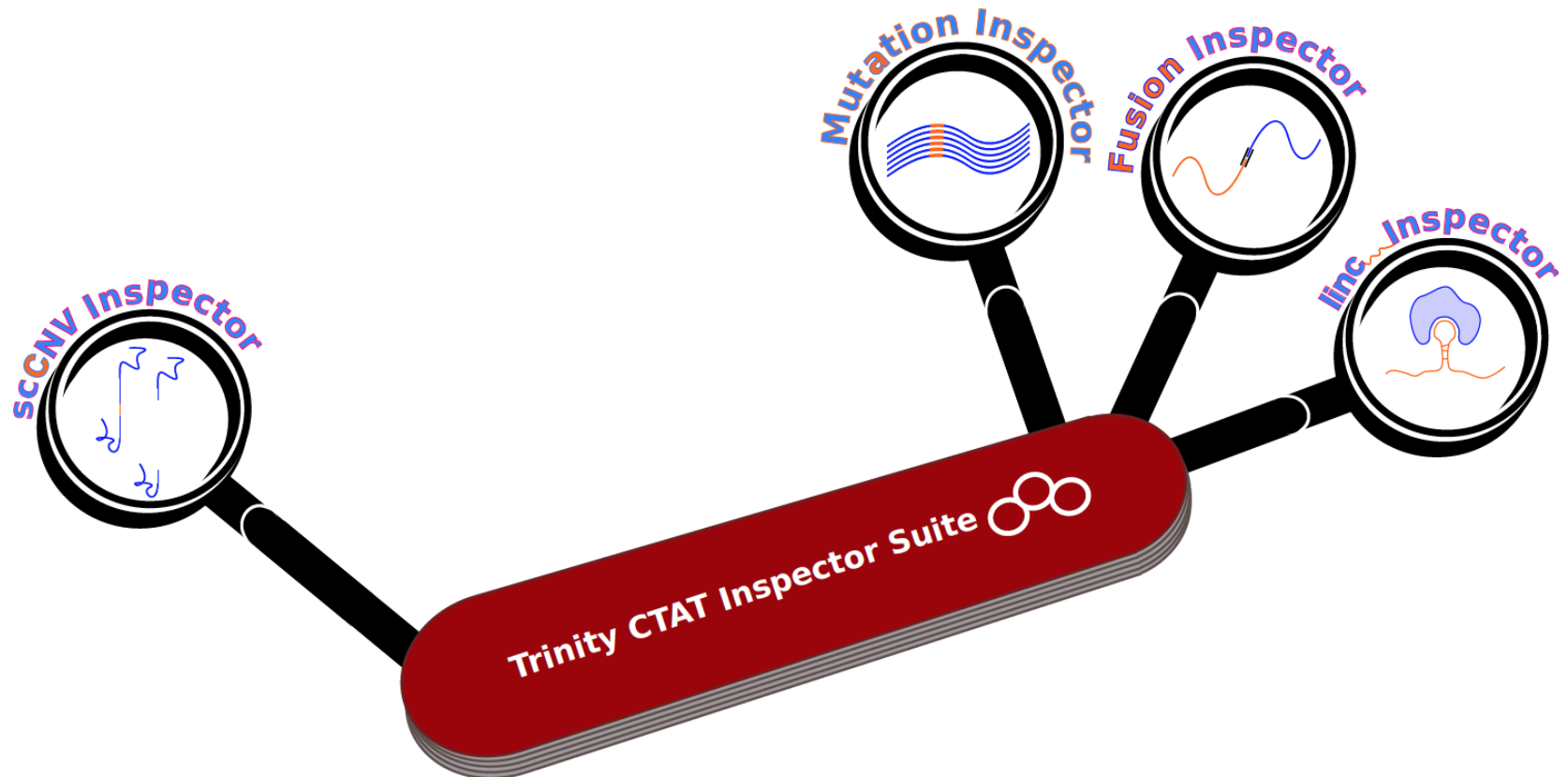
3: Trinity on data 1 and data 2: log

2: reads.right.fq

1: reads.left.fq

Simply Google: Trinity Galaxy or visit: <https://galaxy.ncgas-trinity.indiana.edu/>

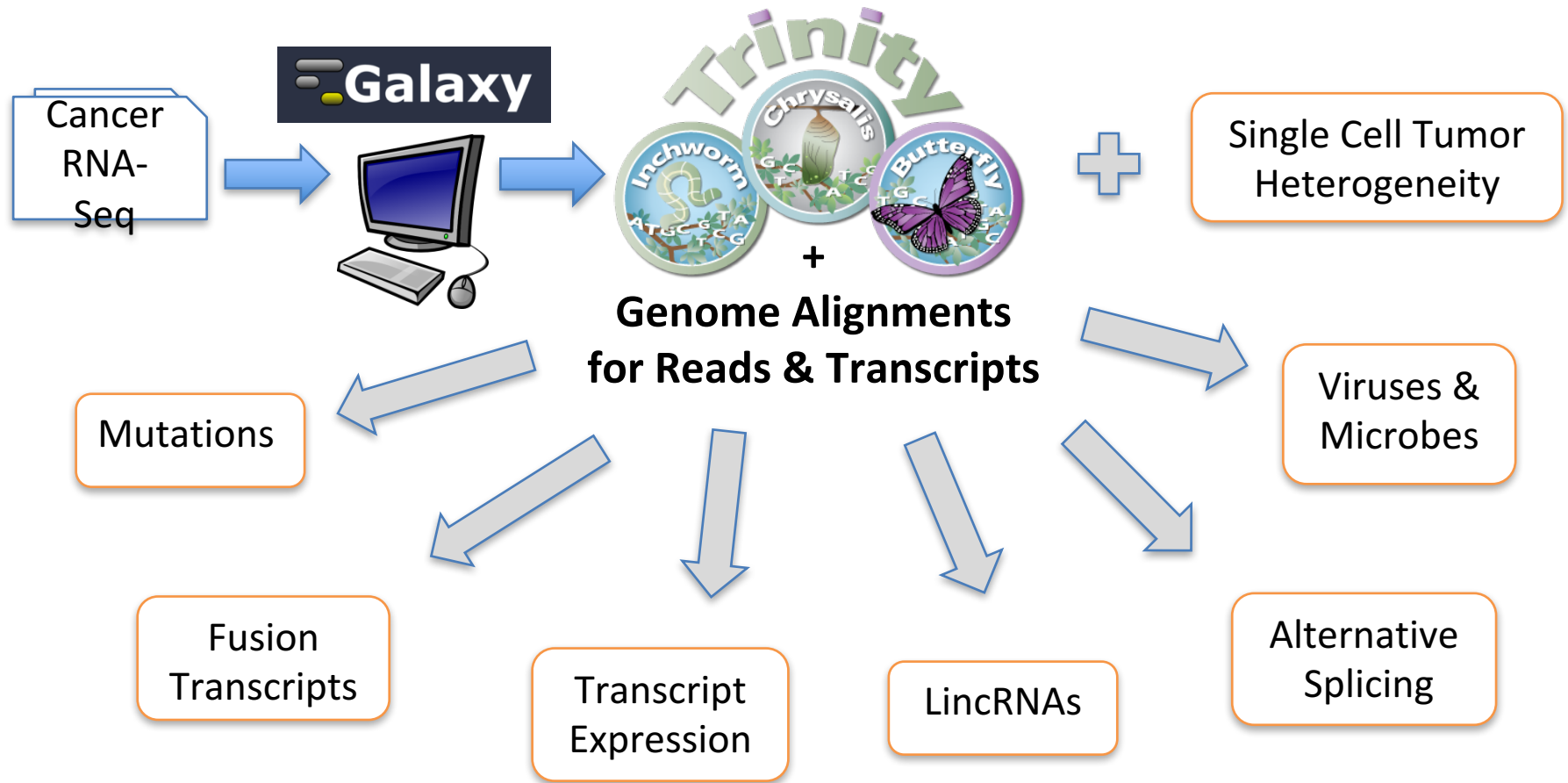
Galaxy Integration of CTAT Inspectors



Interactive Visualizations and Summary Reports



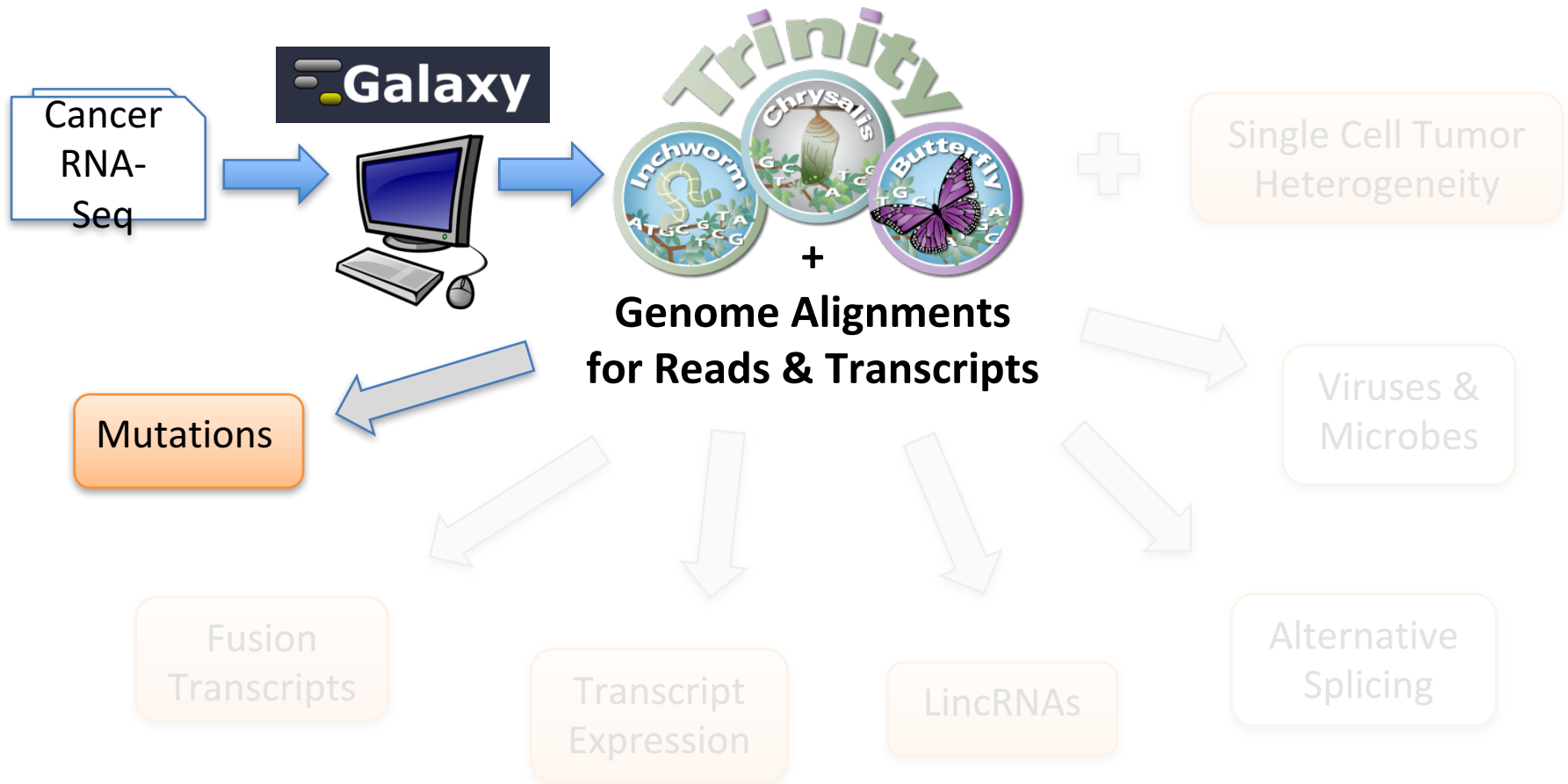
Trinity Cancer Transcriptome Analysis Toolkit



Interactive Visualizations and Summary Reports



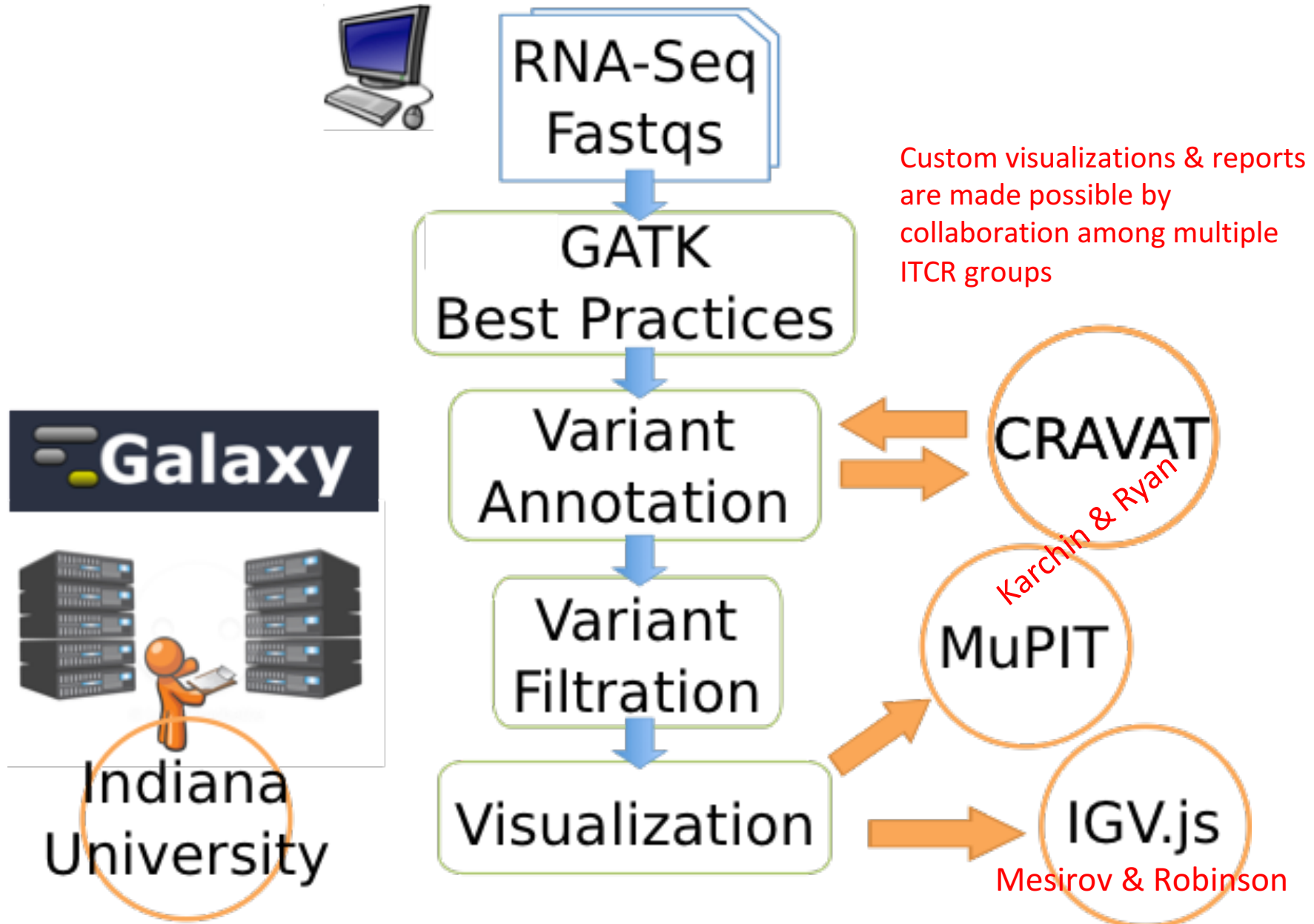
Mutation Detection Using RNA-Seq



Interactive Visualizations and Summary Reports



Trinity CTAT Cancer Mutation Identification Module





Mutation Inspector Report



Exploring a Single Variant



Powered by IGV.js

(ITCR collaborators – Jill Mesirov and James Robinson)

MuPIT Provides Additional 3D Context for Mutation

Visualize variants within protein structure.

← → ↻ mupit.icm.jhu.edu/?gm=chr17:3489155

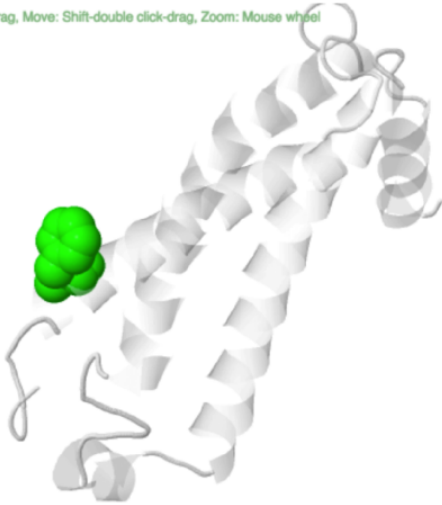
Gene: TRPV1
Structure ID: ENSP00000311692_2
Homology model of TRPV1

Rotate: Left click-drag, Move: Shift-double click-drag, Zoom: Mouse wheel

Help
New Query

Results

TRPV1
ENSP00000311692_2
ENSP00000409627_3
ENSP00000459042_1
NP_061197.4_3
NP_061197.4_4



Style: ☐ Rockets ☒ White ☒ High quality ☐ Show non macromolecules
☒ Cartoons ☐ Custom ☐ Low quality
☐ Trace ☐ Chain ☐ Structure ☐ Temperature

JSmol

Chains

Name	Color	Chain IDs
Modelled_Chain	<input type="checkbox"/>	A

My Mutations

Mutations	Color	Positions in Structure
chr17:3489155	<input checked="" type="checkbox"/>	Modelled Chain Of TRPV1 370

TCGA Mutations

Tissue	TCGA Mutations		TCGA 3D Mutation Hot Regions	
	Color	# Mutated Residues in Structure	Color	Number of TCGA Hot Regions
COAD	<input checked="" type="checkbox"/>	1 +	<input type="checkbox"/>	
ESCA	<input checked="" type="checkbox"/>	1 +	<input type="checkbox"/>	
HNSC	<input checked="" type="checkbox"/>	1 +	<input type="checkbox"/>	
LIHC	<input checked="" type="checkbox"/>	1 +	<input type="checkbox"/>	
PAAD	<input checked="" type="checkbox"/>	1 +	<input type="checkbox"/>	
STAD	<input checked="" type="checkbox"/>	2 +	<input type="checkbox"/>	

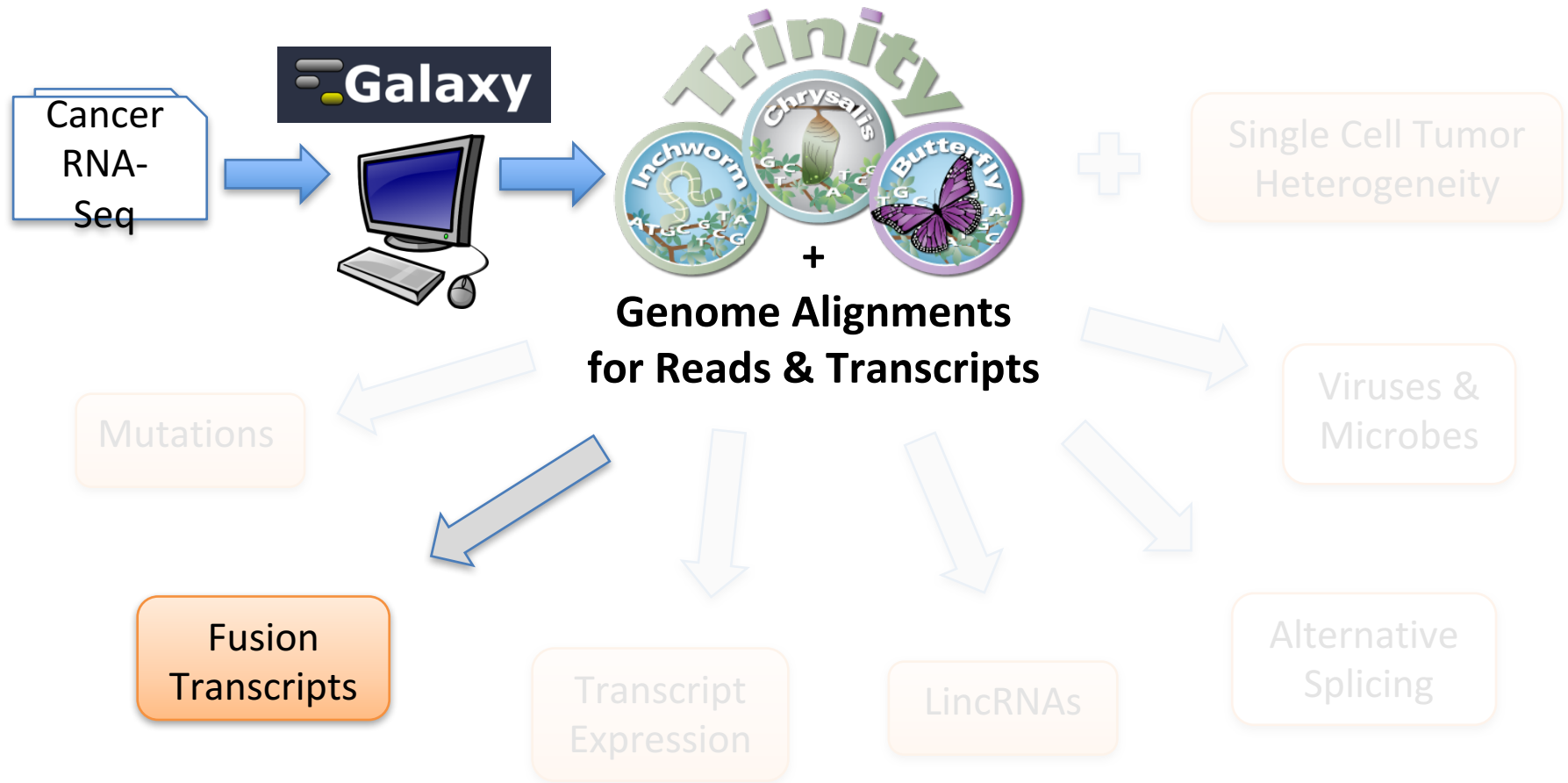
Experimental Mutations of Interest

Description	Color	Number of Feature Ranges
Y_A_Loss_of_s	<input checked="" type="checkbox"/>	1 +

© 2014 Karchin lab | Funding: 5U01CA180956-02, NIH 3U24CA143858-2S1, NIH 5R21CA152432-02 | To cite: PMID: 23793516 | Credits

(ITCR collaborators - Rachel Karchin and Mike Ryan)

Fusion Transcript Detection



Interactive Visualizations and Summary Reports



Top-down Approaches to Fusion Transcript Discovery

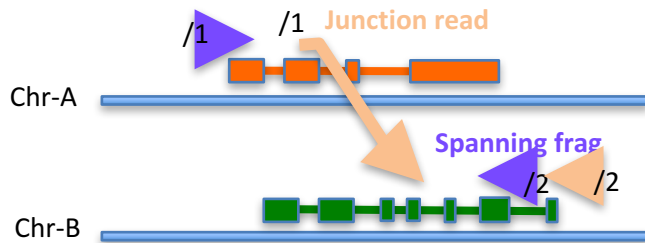
Paired-end
Illumina RNA-Seq



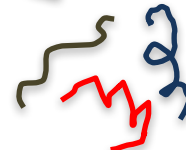
STAR-Fusion

* In collaboration with Alex Dobin,
developer of STAR

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



De novo RNA-Seq assembly

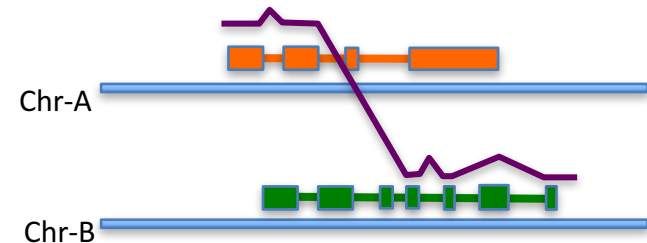


Trinity or Oases (MK)

GMAP-fusion

* In collaboration with Tom Wu,
developer of GMAP

Align **transcripts** to genome,
Identify Fusion Transcripts



Top-down Approaches to Fusion Transcript Discovery

Paired-end
Illumina RNA-Seq

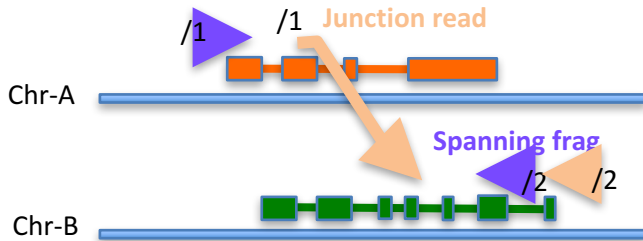


STAR-Fusion

* In collaboration with Alex Dobin,
developer of STAR

STAR-alignments

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



De novo RNA-Seq assembly



Trinity or Oases (MK)

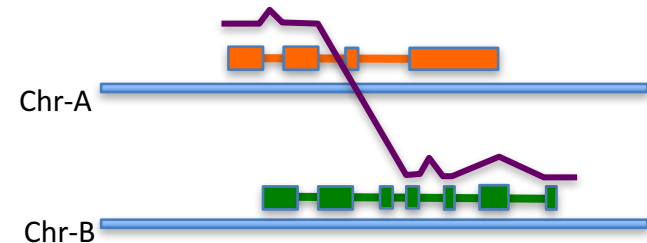
GMAP-fusion

* In collaboration with Tom Wu,
developer of GMAP

DISCASM

Just discordant or
unmapped reads


Align *transcripts* to genome,
Identify Fusion Transcripts





New Results

STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq

 Brian Haas, Alexander Dobin, Nicolas Stransky, Bo Li, Xiao Yang, Timothy Tickle, Asma Bankapur, Carrie Ganote, Thomas Doak, Natalie Pochet, Jing Sun, Catherine Wu, Thomas Gingeras, Aviv Regev

doi: <https://doi.org/10.1101/120295>

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract

Info/History

Metrics

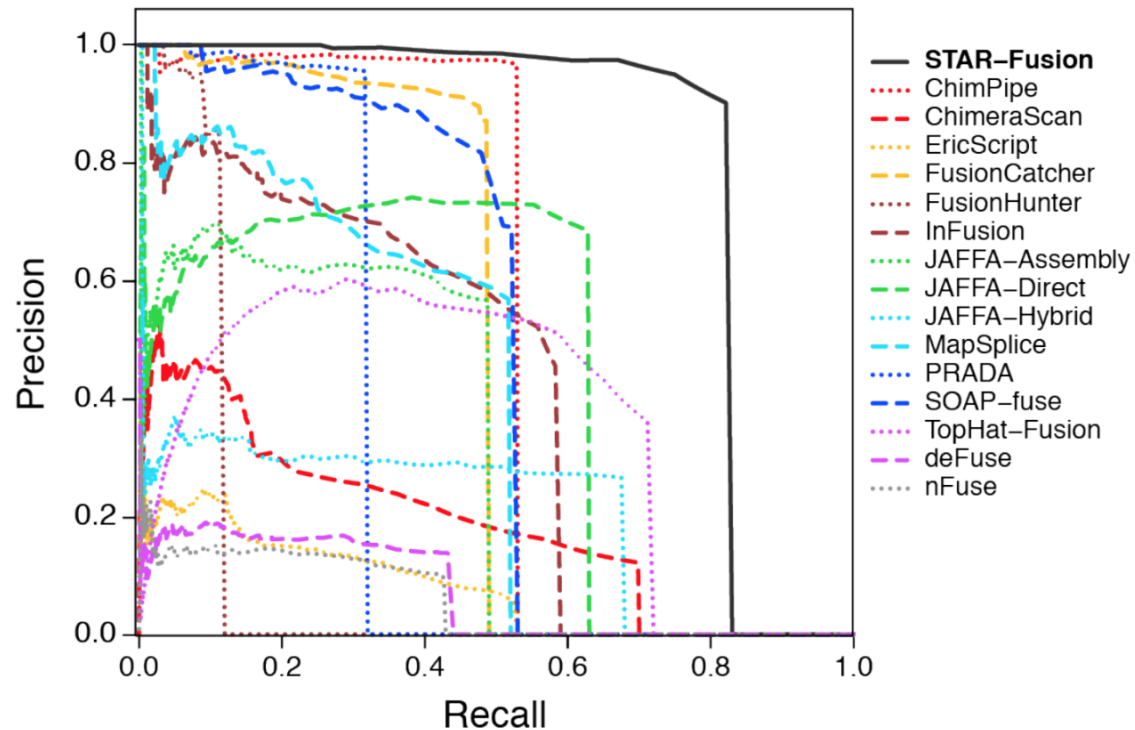
Supplementary material

 Preview PDF

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



$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

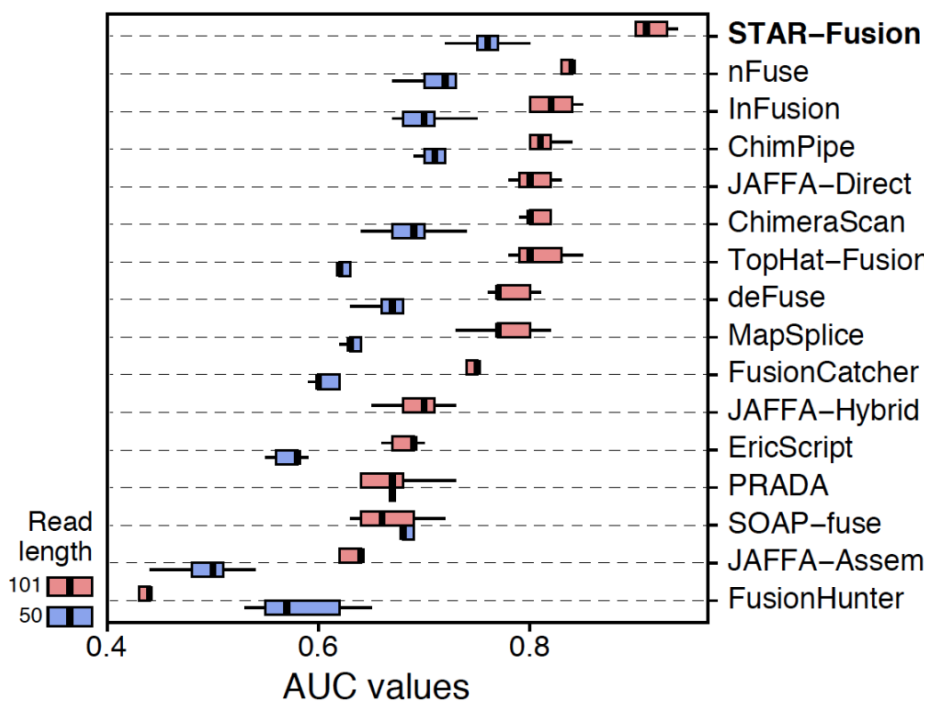
$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Accuracy = area under the curve (AUC)

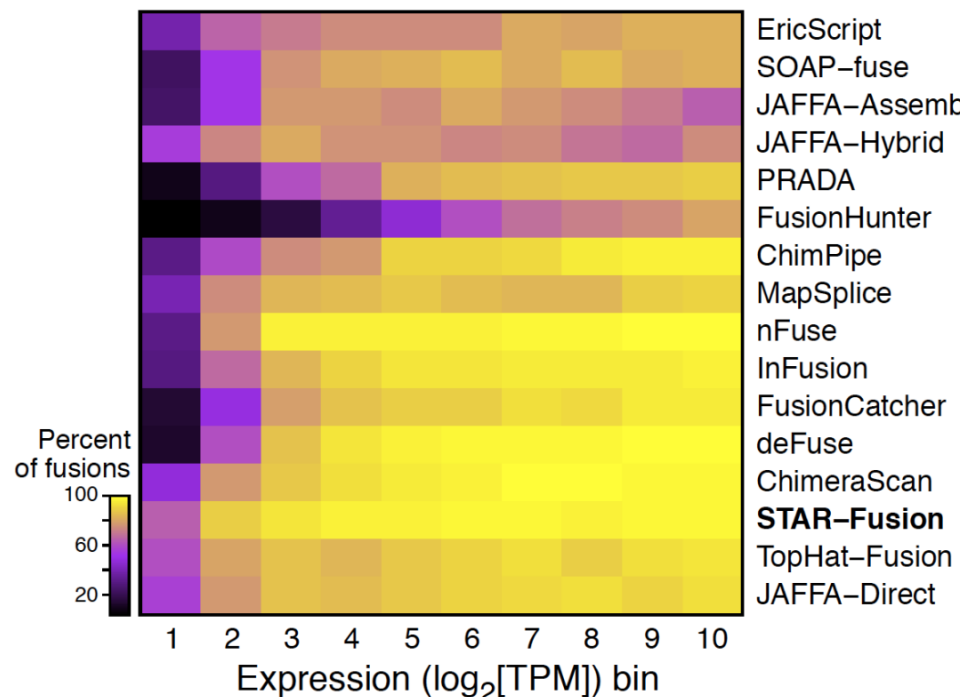
Benchmarking Fusion-finding Tools

(results shown for simulated data)

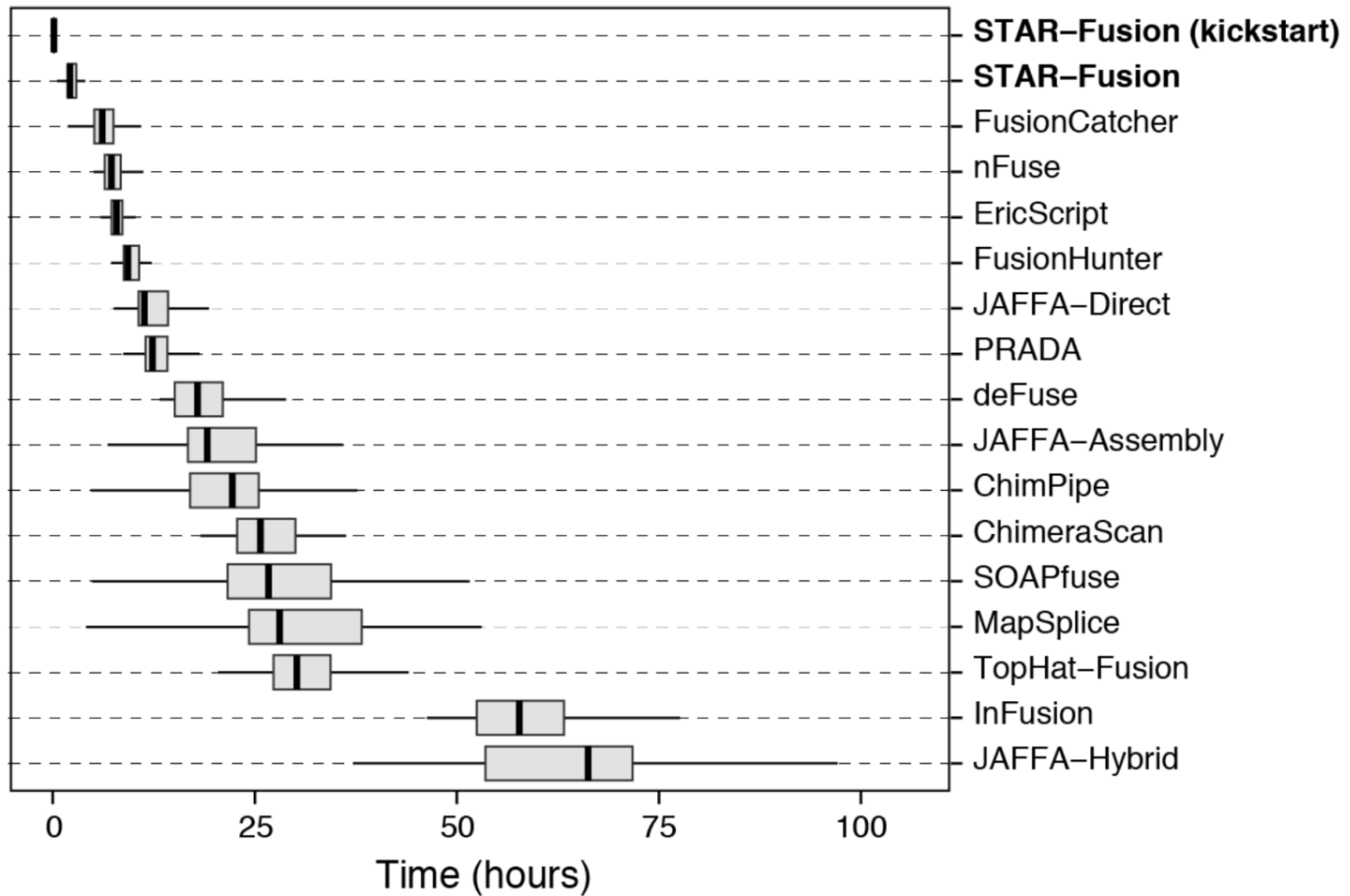
Fusion Prediction Accuracy
(AUC value distribution across 5 replicates)



Fusion Prediction Sensitivity vs. Expression



STAR-Fusion is accurate and ***FAST***




(\$)



New Results

STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq

 Brian Haas, Alexander Dobin, Nicolas Stransky, Bo Li, Xiao Yang, Timothy Tickle, Asma Bankapur, Carrie Ganote, Thomas Doak, Natalie Pochet, Jing Sun, Catherine Wu, Thomas Gingeras, Aviv Regev

doi: <https://doi.org/10.1101/120295>




Reproducible data analysis

- All fusion prediction results from all programs included
- Single command to reanalyze data, generate all figures and tables



New Results

STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq

 Brian Haas, Alexander Dobin, Nicolas Stransky, Bo Li, Xiao Yang, Timothy Tickle, Asma Bankapur, Carrie Ganote, Thomas Doak, Natalie Pochet, Jing Sun, Catherine Wu, Thomas Gingeras, Aviv Regev

doi: <https://doi.org/10.1101/120295>



ICGC-TCGA DREAM Somatic Mutation Calling Challenge – RNA



Bottom-up Fusion 'In silico Validation' Using FusionInspector

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

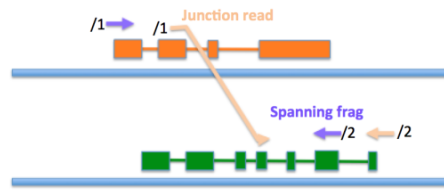


* STAR enhancements to support FusionInspector

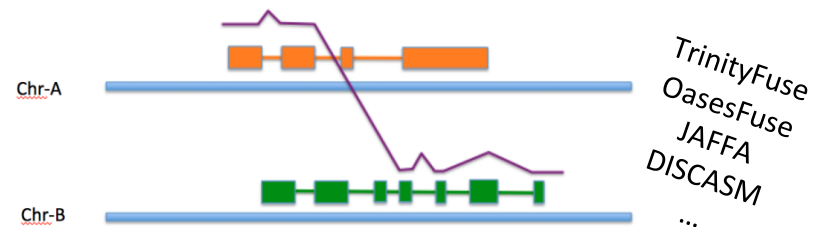
Make mini-fusion contigs



STAR-Fusion
PRADA
SOAPfuse
FusionCatcher
...



All fusion predictions

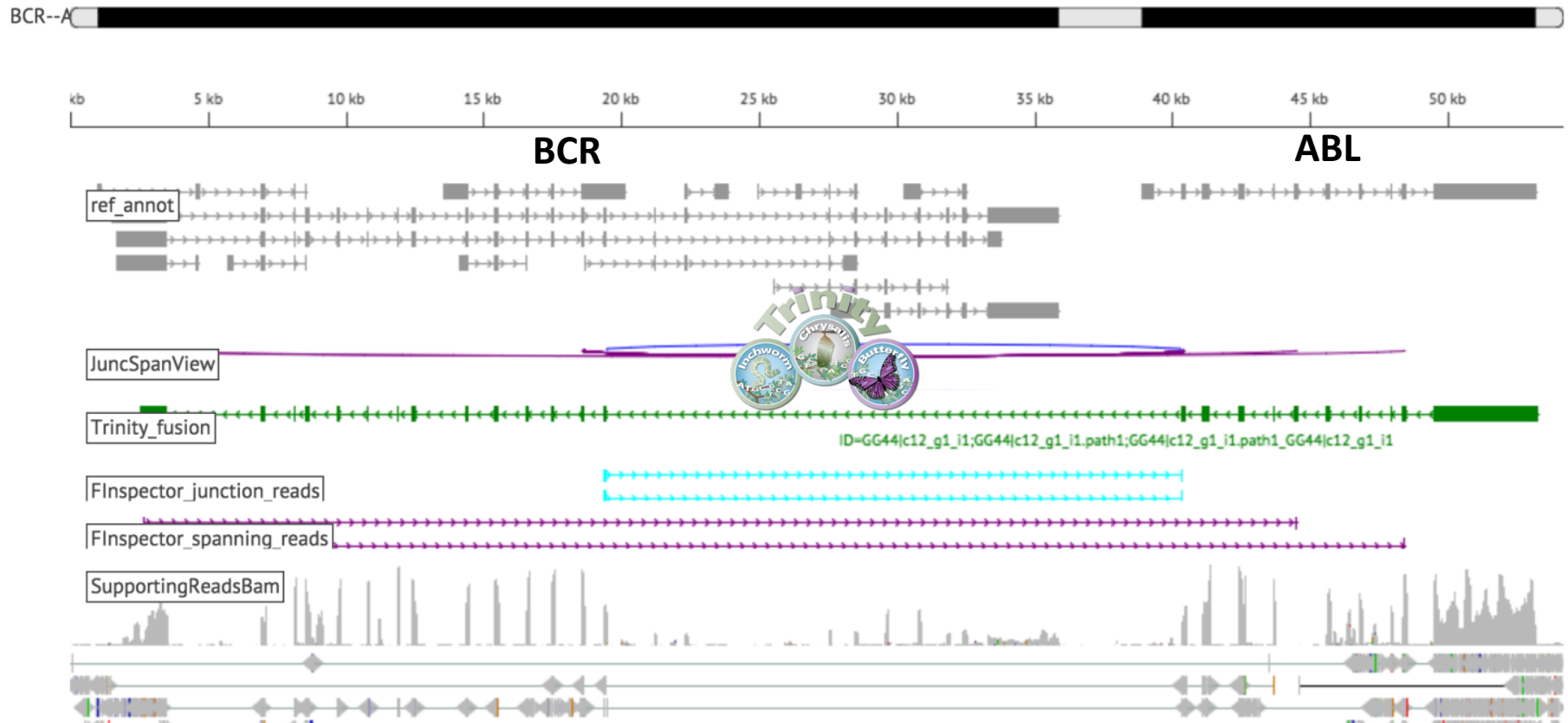


TrinityFuse
OasesFuse
JAFFA
DISASM
...



FusionInspector Fusion View

BCR--ABL1:1-54,203

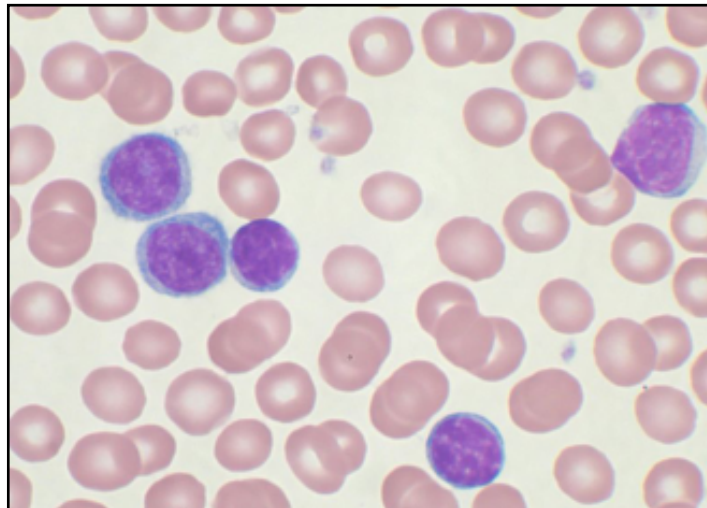


Powered by IGV.js

ITCR collaborators - Jim Robinson and Jill Mesirov

Driving Cancer Project: Search for Gene Fusions in Chronic Lymphocytic Leukemia (CLL)

- A common adult leukemia in Europe and North America
- Tremendous clinical heterogeneity
- Incurable by conventional chemotherapy
- Molecular understanding largely unknown



* Work done in collaboration with Cathy Wu, Dana Farber Cancer Center & BI.

Defining Recurrent Fusion Transcripts in Chronic Lymphocytic Leukemia

270 CLL tumor samples, 18 normal samples

Total fusions

Unique fusions

8322

4898

Filter GTEx

6587 (79%)

4644 (95%)

Remove Normals

5893 (71%)

4558 (94%)

Expression Filter
0.1 FFPM

599 (7%)

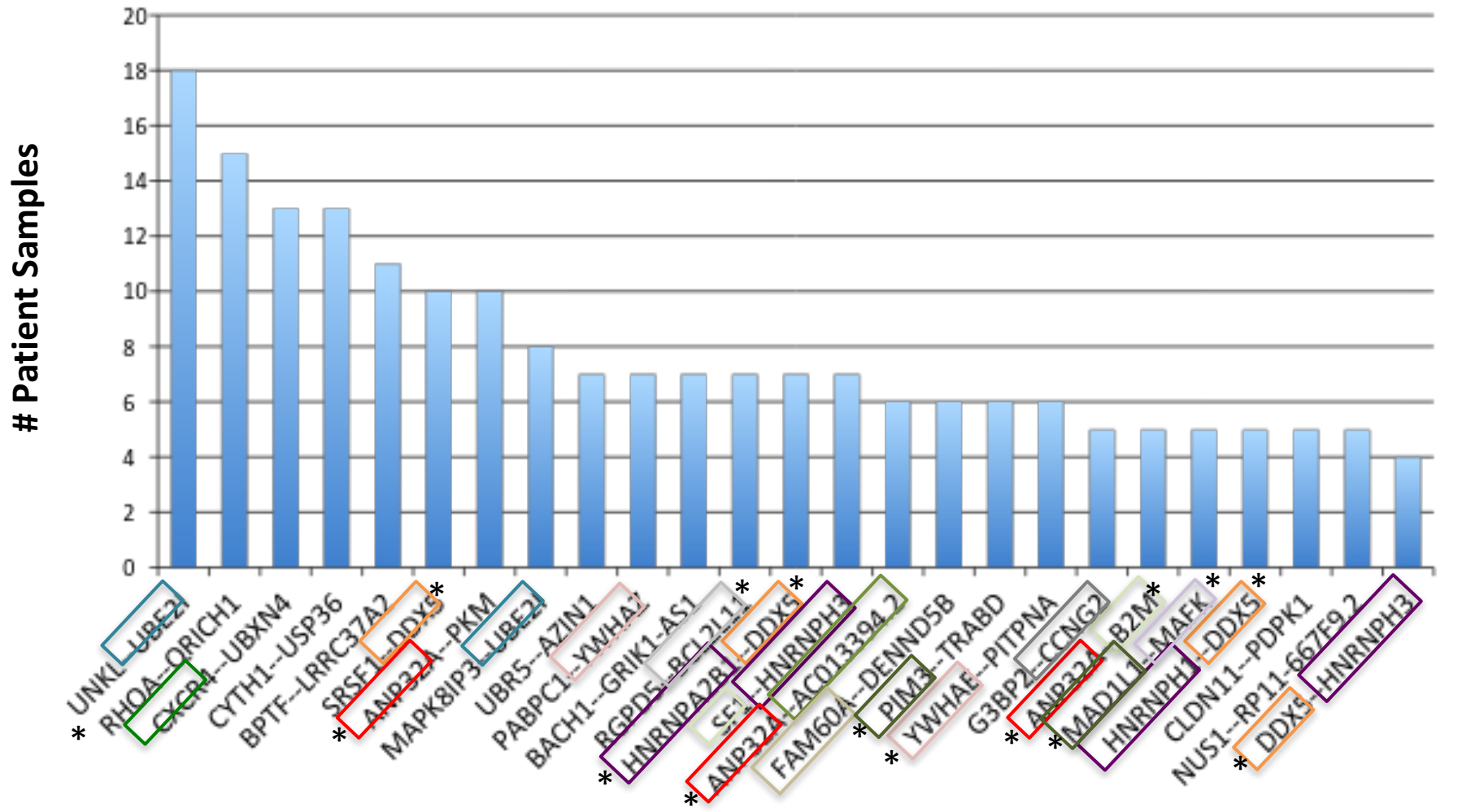
320 (7%)

Define
recurrent

341 (4%)

62 (1%)

Top 25/62 Recurrent Fusions in CLL



(*) Known oncogene

Validating Fusion Predictions via RT-PCR

Pt#, fusion name, Mw

CW105 FNDC3A--ATP7B 196

CW132 KDM5B--AHCYL1 161

CW97 MAML2--PAFAH1B2 159

CW179 PHF11--RNASEH2B 134

CW210 PRKCE--PLEKHH2 123

~~CW105 FAM214A--WDFY2 202~~

CW140 RB1--UNC79 200

Pt#, fusion name, Mw

JN08 TLE4--PVT1 106

CW32 DLEU2--NDFIP2 22

CW107 RNASEH2B--RNASEH2B

CW202 GLCCI1--TRIM73

CW3 INTS6--SMIM14 97

210-L ATXN1--NOTCH4 1

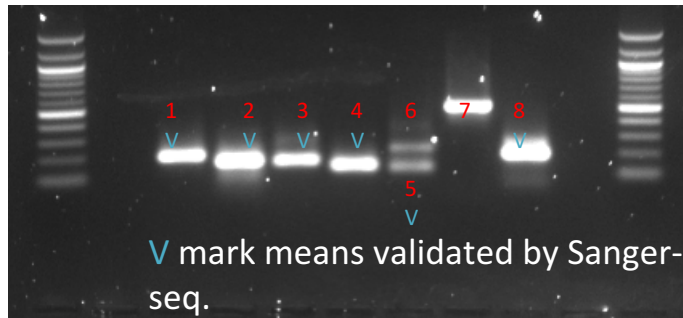
CW105 WDFY2--GLDN 173

CW37 OAZ1--CSNK1G2 92

CW37 USP11--CDK16 277

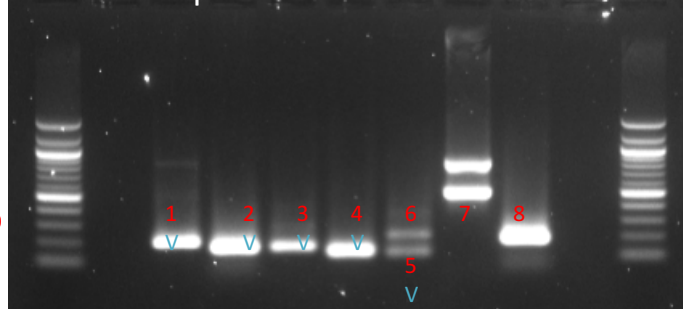
SuperScript 2,
random primer

a



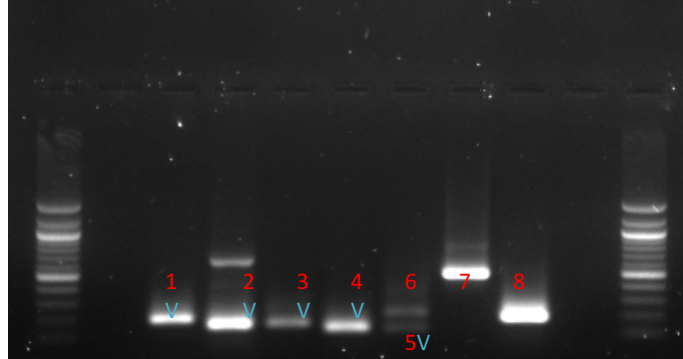
ThermoScript,
random primer

b

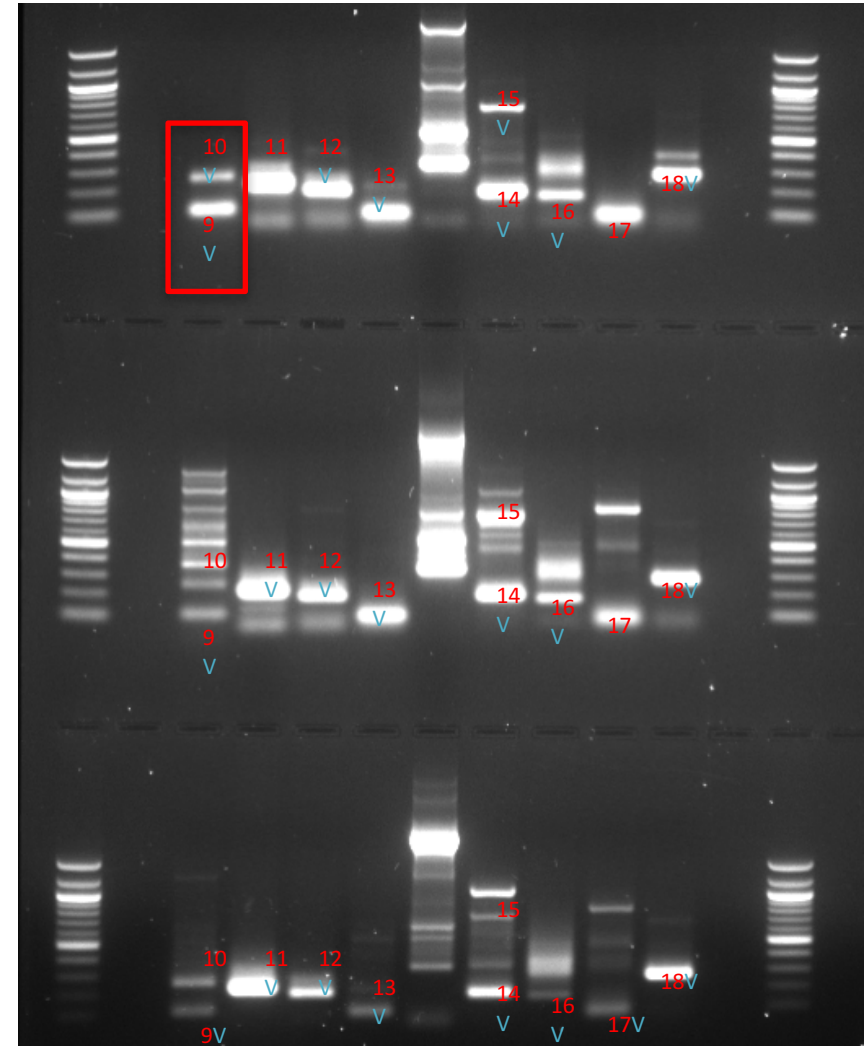


ThermoScript,
origo-dT primer

c



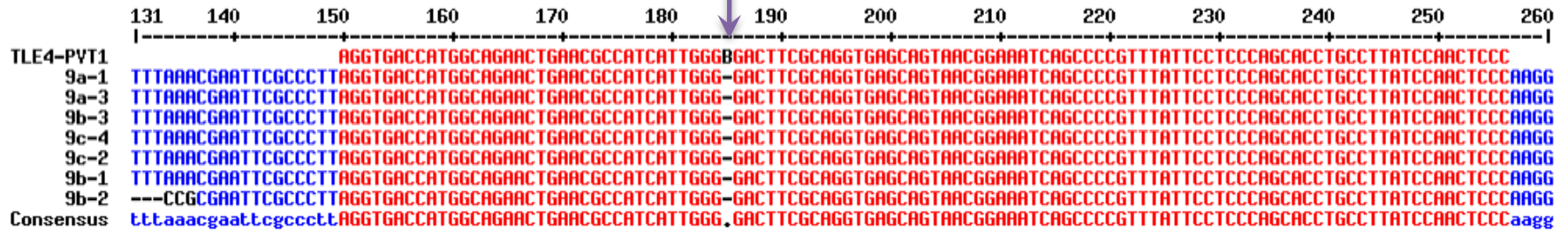
* Work by
Jintaek Kim



Targeted Fusion Transcript Validated by Sanger Sequencing

band 9

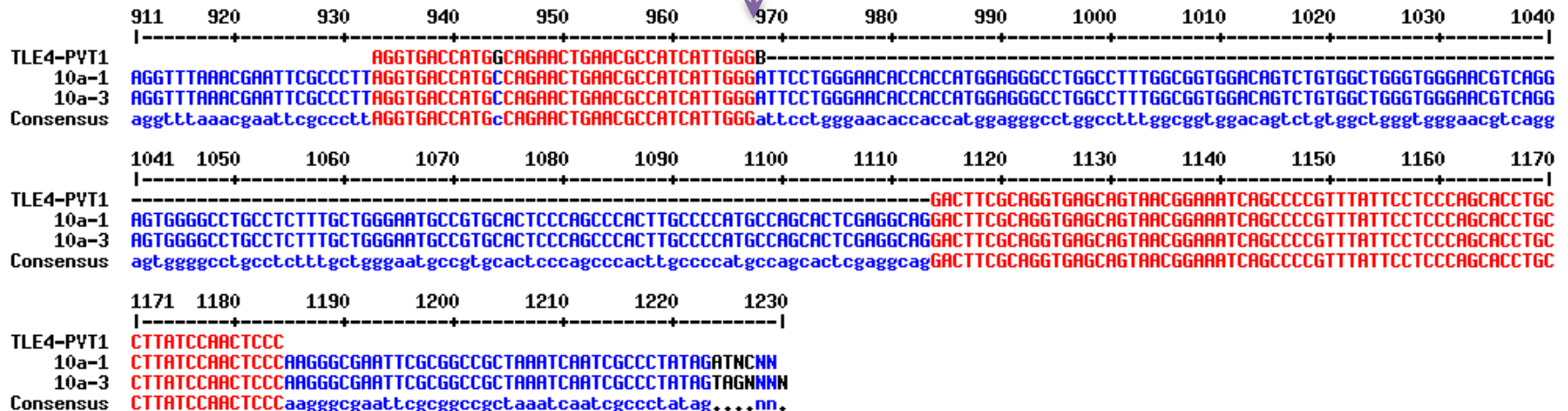
break point



Alternatively Spliced Fusion Transcript Validated

band 10

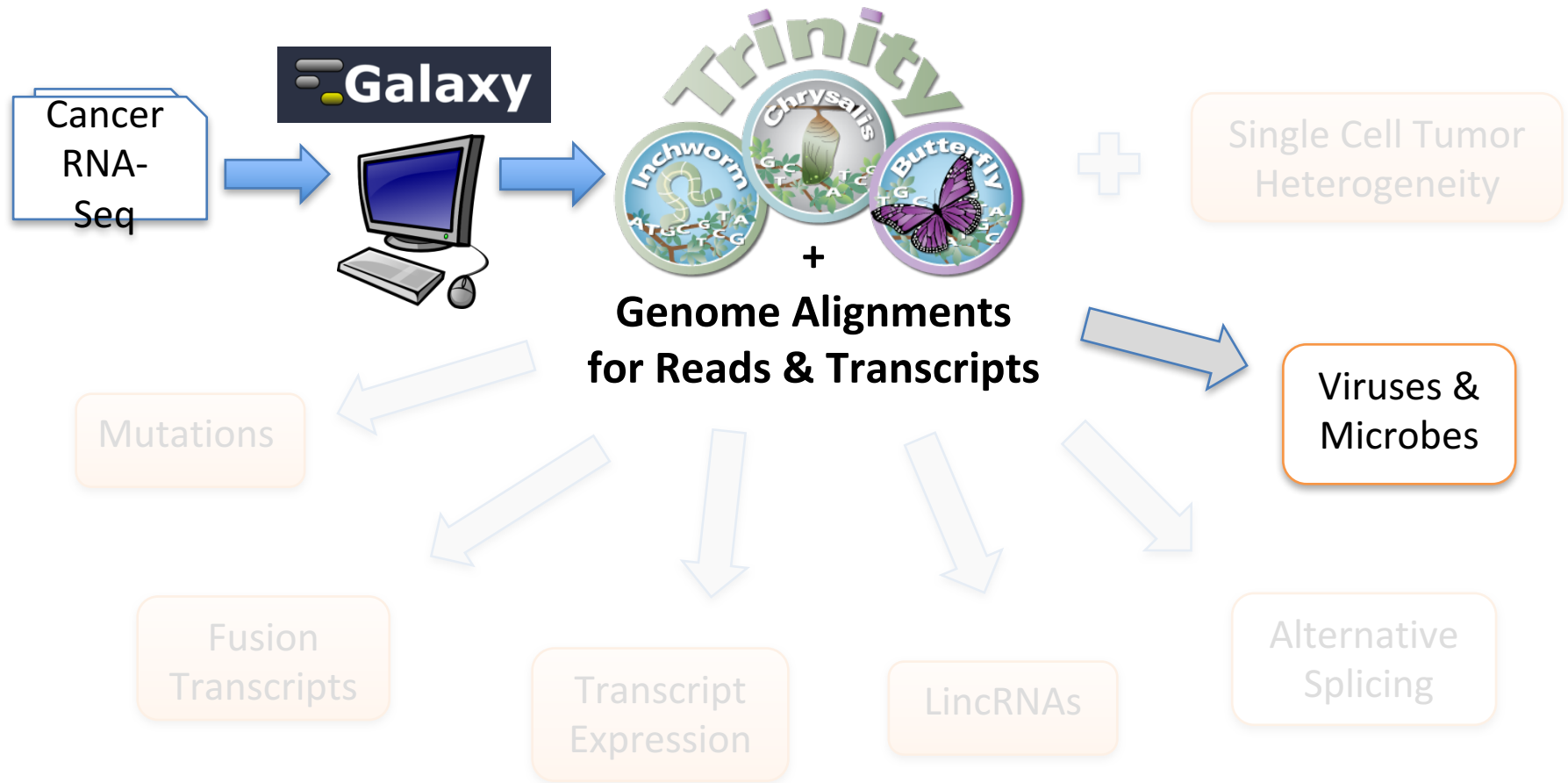
break point



JN08 TLE4--PVT1

CH9--CH8

Single Cell Tumor Heterogeneity



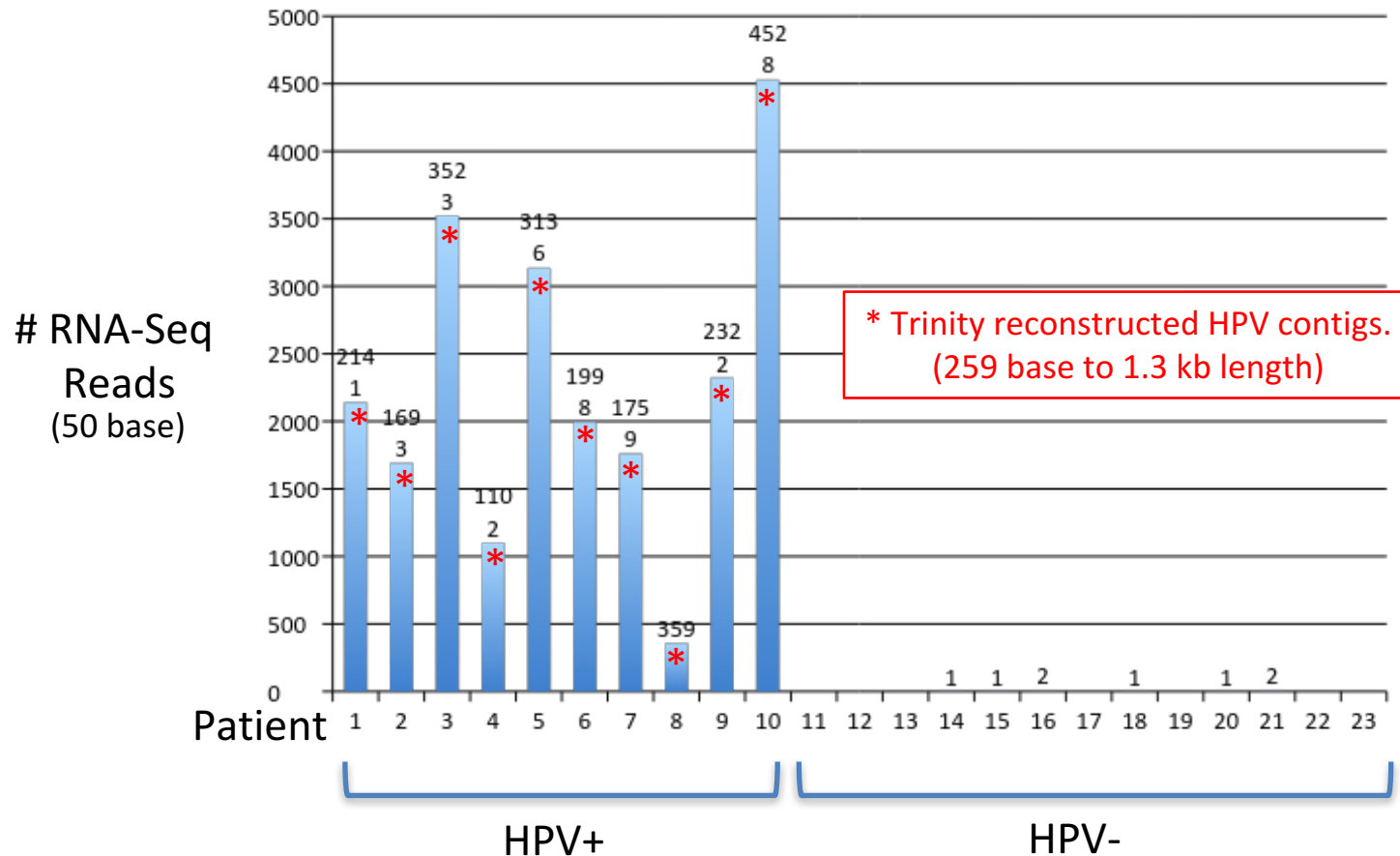
Interactive Visualizations and Summary Reports



Centrifuge +/- Trinity Applied to HPV-Driven HNSCC

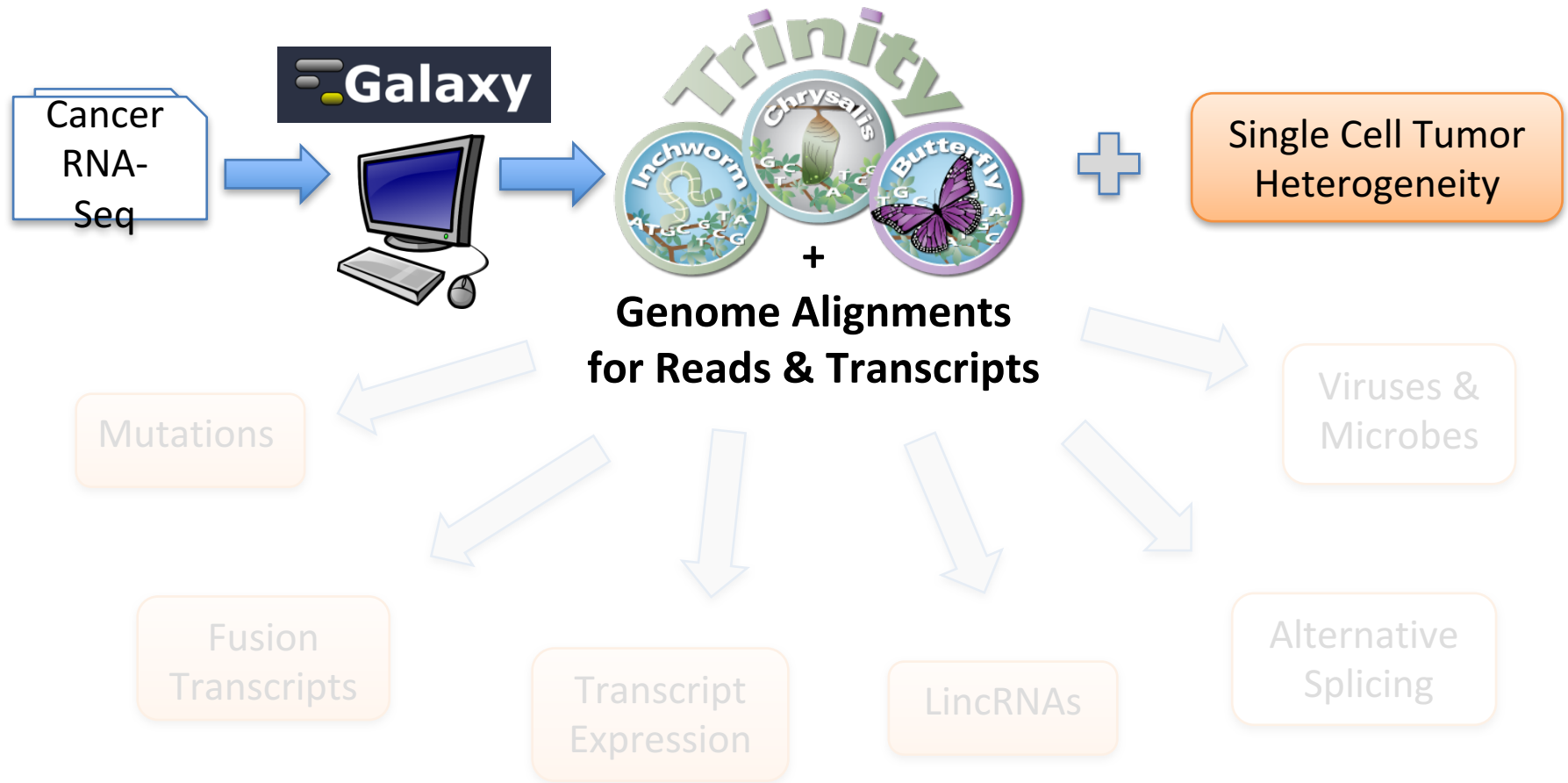
* Collaboration with Steven Salzberg, JHU. Centrifuge: Kim et al. Genome Research, 2016

Longer Sequences and Taxonomic Resolution Using Trinity



* RNA-Seq data from : Gene expression analysis of TIL rich HPV-driven head and neck tumors reveals a distinct B-cell signature when compared to HPV independent tumors. Wood et al. Oncotarget, 2016

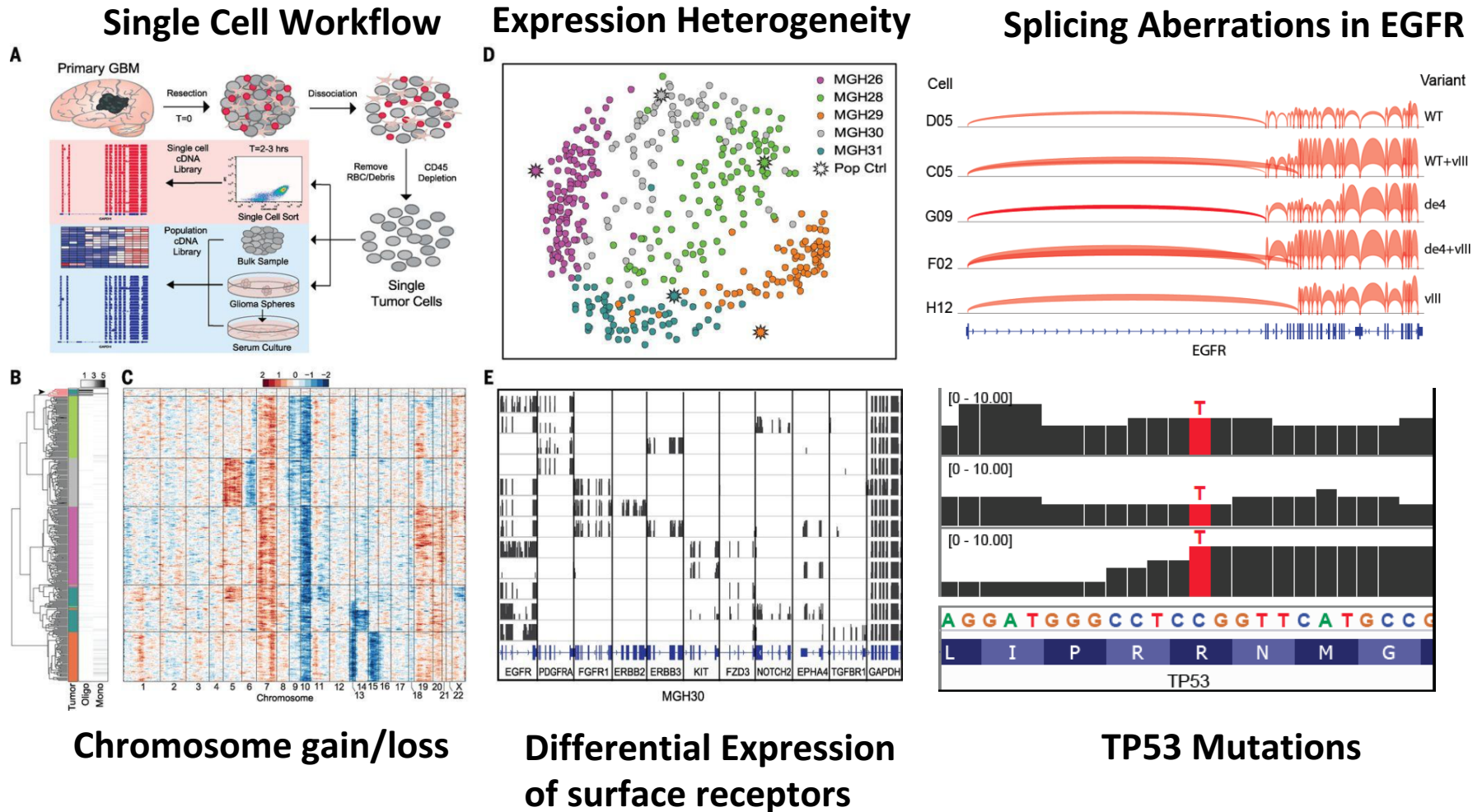
Single Cell Tumor Heterogeneity



Interactive Visualizations and Summary Reports



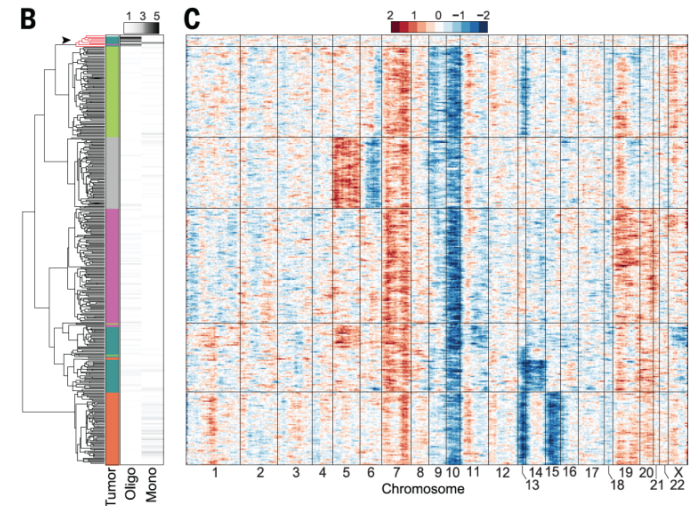
Single Cell Resolution of Tumor Heterogeneity via RNA-Seq



Large-scale Copy Number Variation Inferred from Single Cell RNA-Seq Data

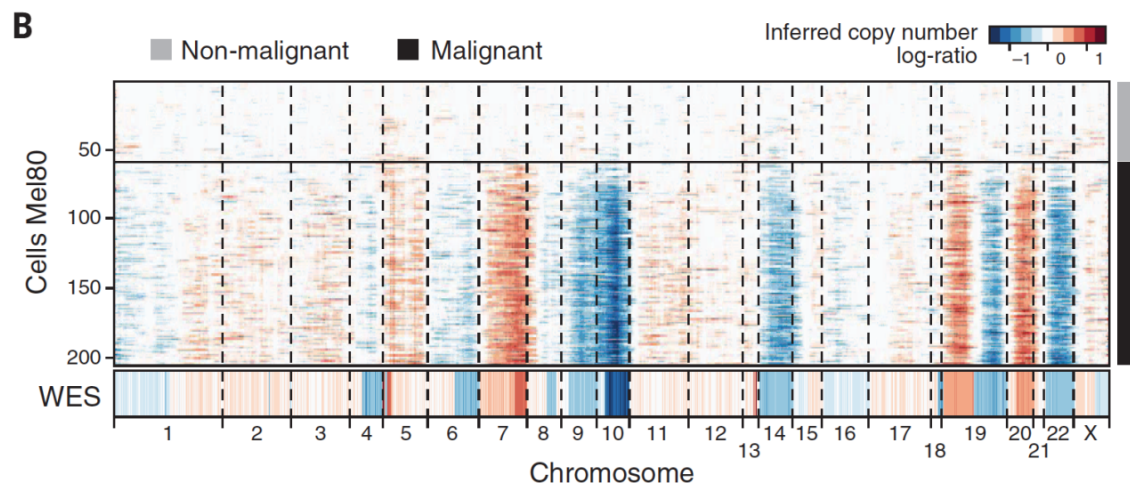
Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma

Patel, Tirosh, ..., Regev, Bernstein; Science 2014



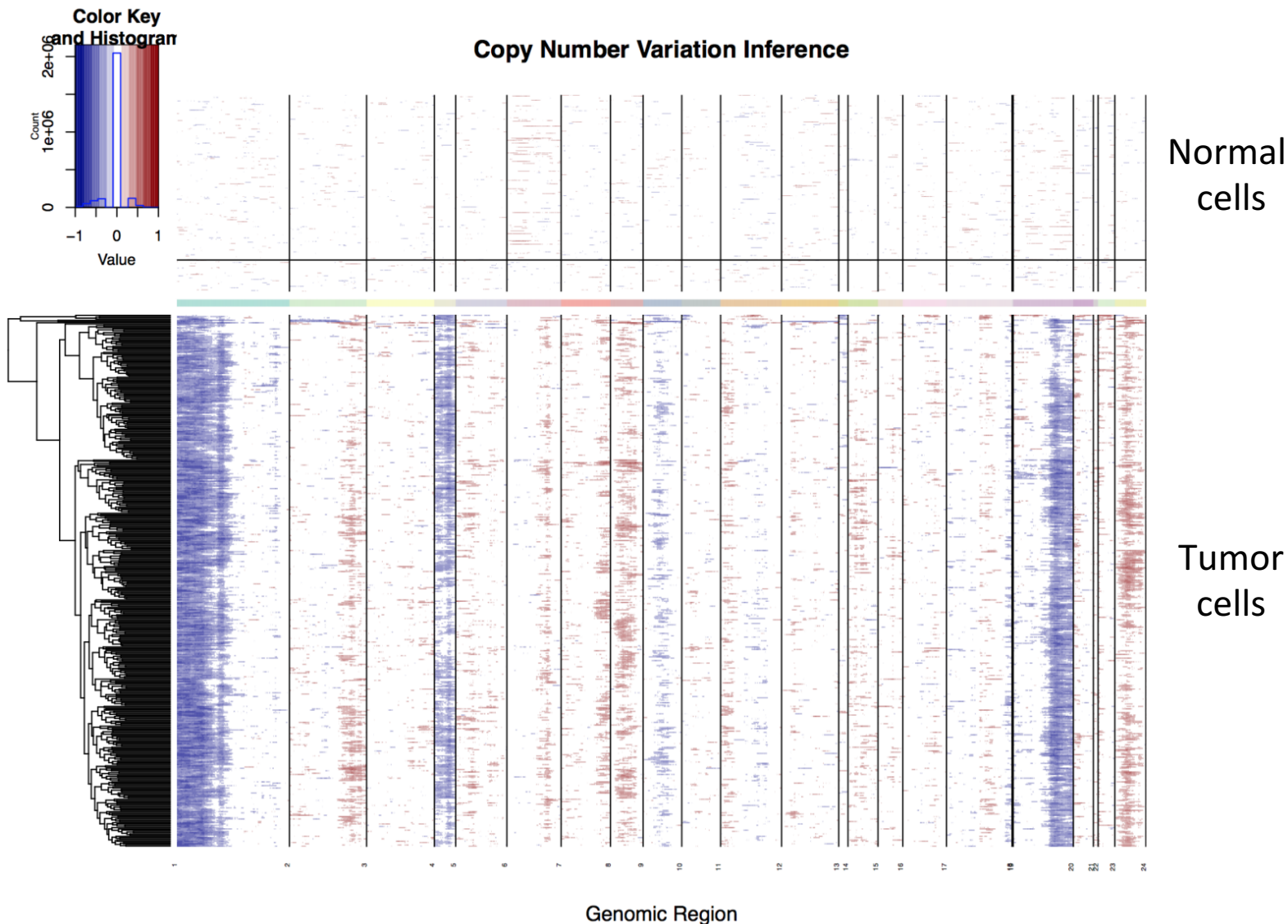
Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq

Tirosh, Izaar,, Regev, Garraway; Science 2016



Trinity CTAT InferCNV: Utility to identify large-scale CNV from single cell RNA-Seq

<https://github.com/broadinstitute/inferCNV>



* Example from oligodendroglioma

First official software release: May 30, 2017

Firecloud

Scalable Cancer Computing Solution
for the
NCI Cloud



- Integration of Trinity CTAT into Docker and WDL workflows
- Process TCGA data
- Shareable workflows and data resources

Integration of Trinity CTAT into:



Clinical Research Sequencing Platform (CRSP)



BROAD
INSTITUTE

GENOMIC
SERVICES

Pilot study in pediatric oncology underway

Got Cancer RNA-Seq? Run Trinity!

Mutation detection

Expression



Fusion transcripts

Splicing

Viruses

Single cell tumor heterogeneity

Lots more to come!!!

<https://galaxy.ncgas-trinity.indiana.edu/>

Acknowledgements



Aviv Regev

Brian Haas
Timothy Tickle
Asma Bankapur
Christophe Georgescu
Ami-levy Moonshine



Jill Mesirov
James Robinson



Steven Salzberg



Alex Dobin



BRIGHAM AND
WOMEN'S HOSPITAL

Nathalie Pochet



Rachel Karchin
Mike Ryan



DANA-FARBER
CANCER INSTITUTE

Cathy Wu
Jing Sun
Peggy Hsu
Sachet Shukla



Thomas Doak
Carrie Ganote
Robert Henschel
Cicada Brokaw



Informatics Technology
for Cancer Research