Informatics Technology for Cancer Research

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

BROAD INSTITUTE

KLARMAN CELL OBSERVATORY

ITCR meeting, June 2016

The Cancer Transcriptome

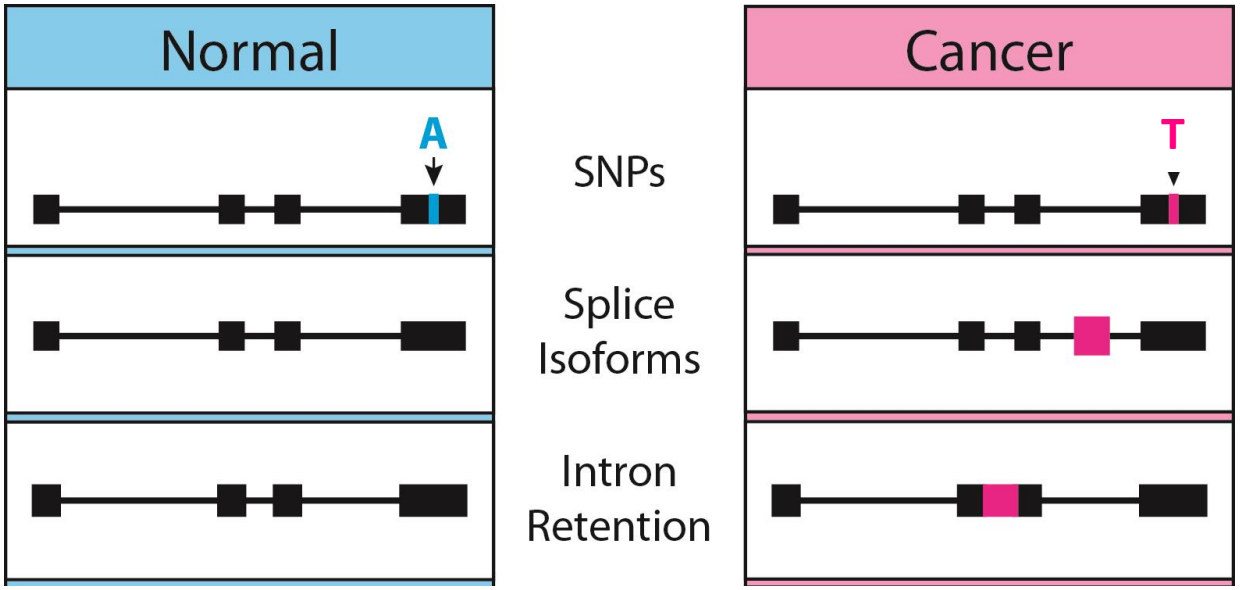
A window into the (expressed) genetic and epigenetic state of a tumor

The Cancer Transcriptome

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

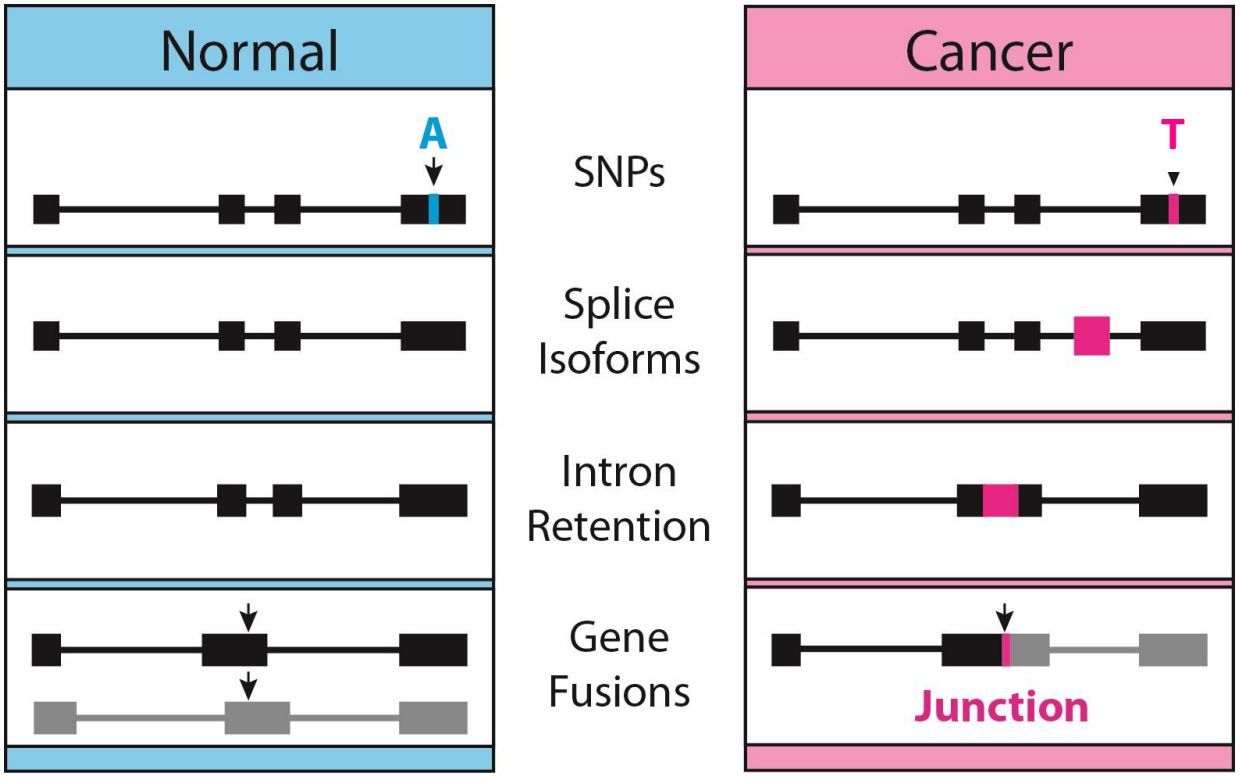
The Cancer Transcriptome

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



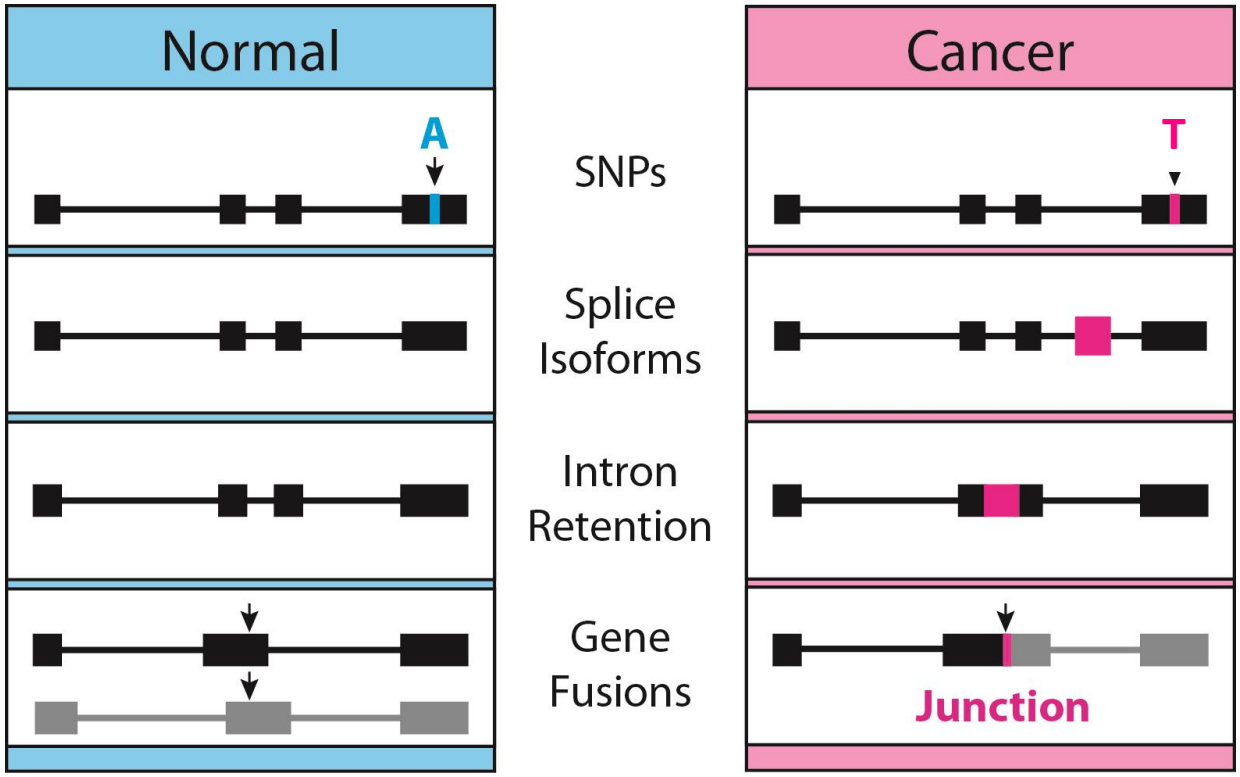
The Cancer Transcriptome

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



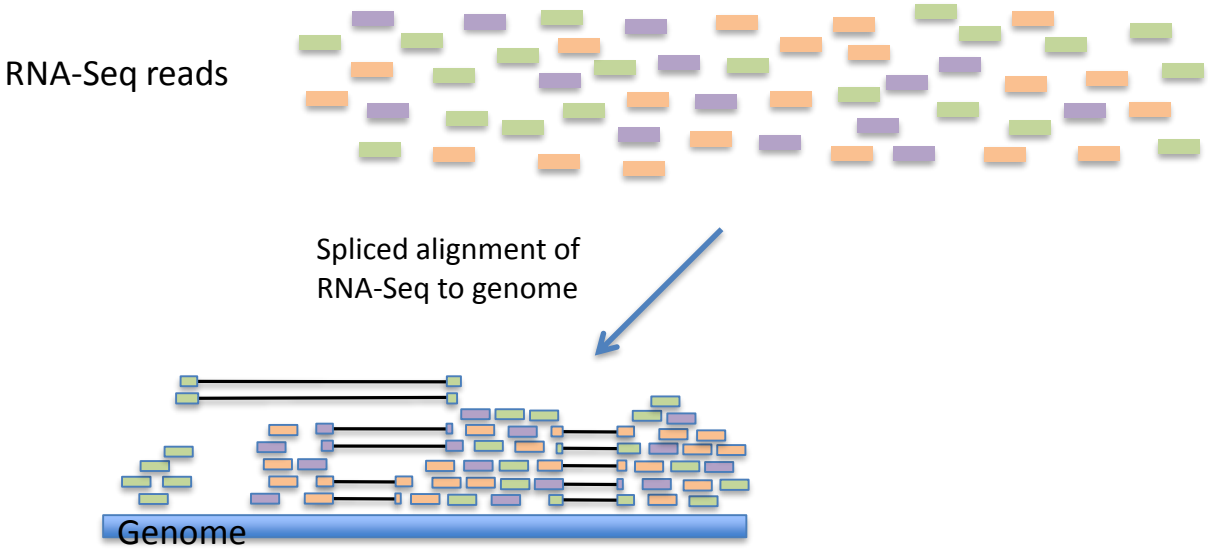
+ the associated microbiome, virome…

Contemporary strategies for transcript analysis from RNA-Seq

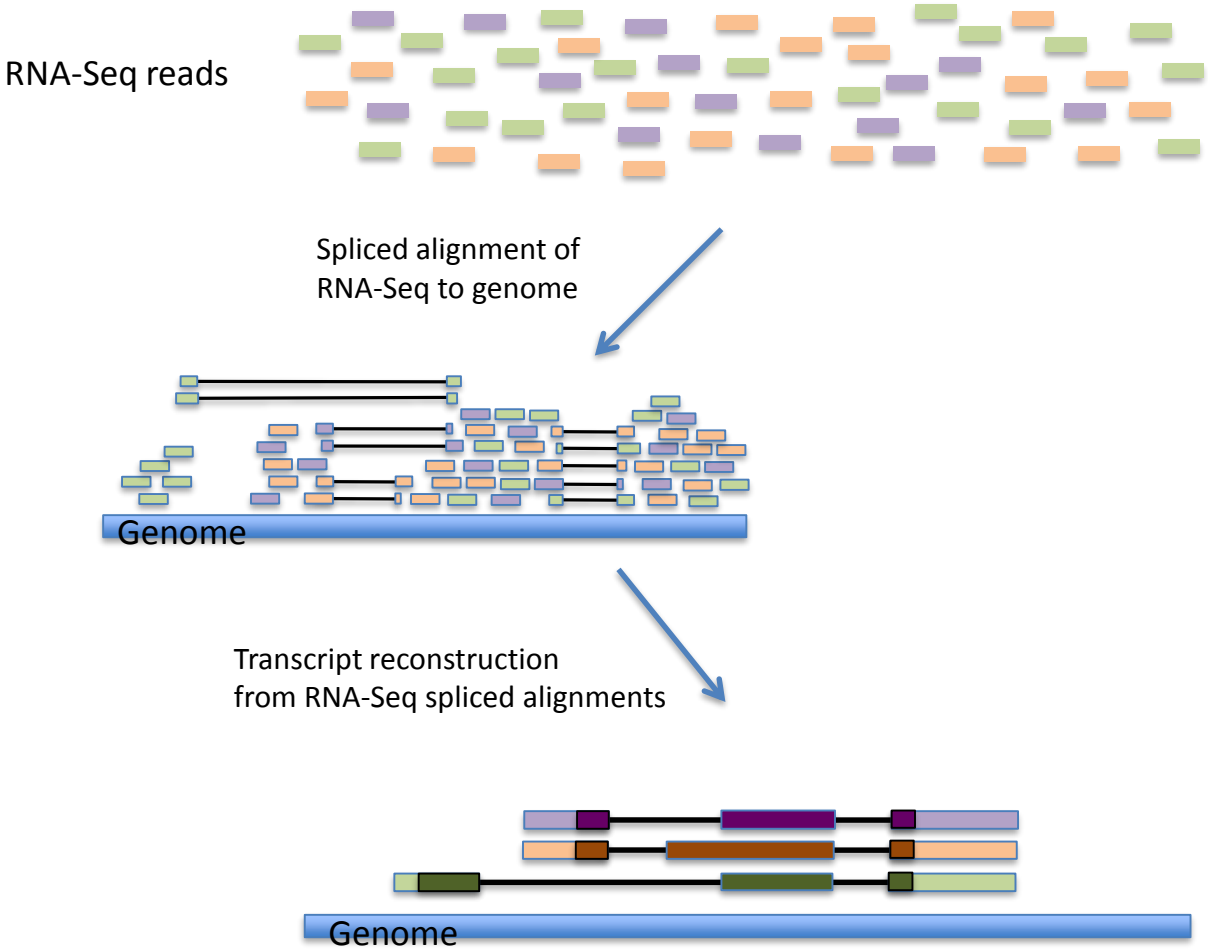
RNA-Seq reads

Two paradigms for transcriptome Analysis

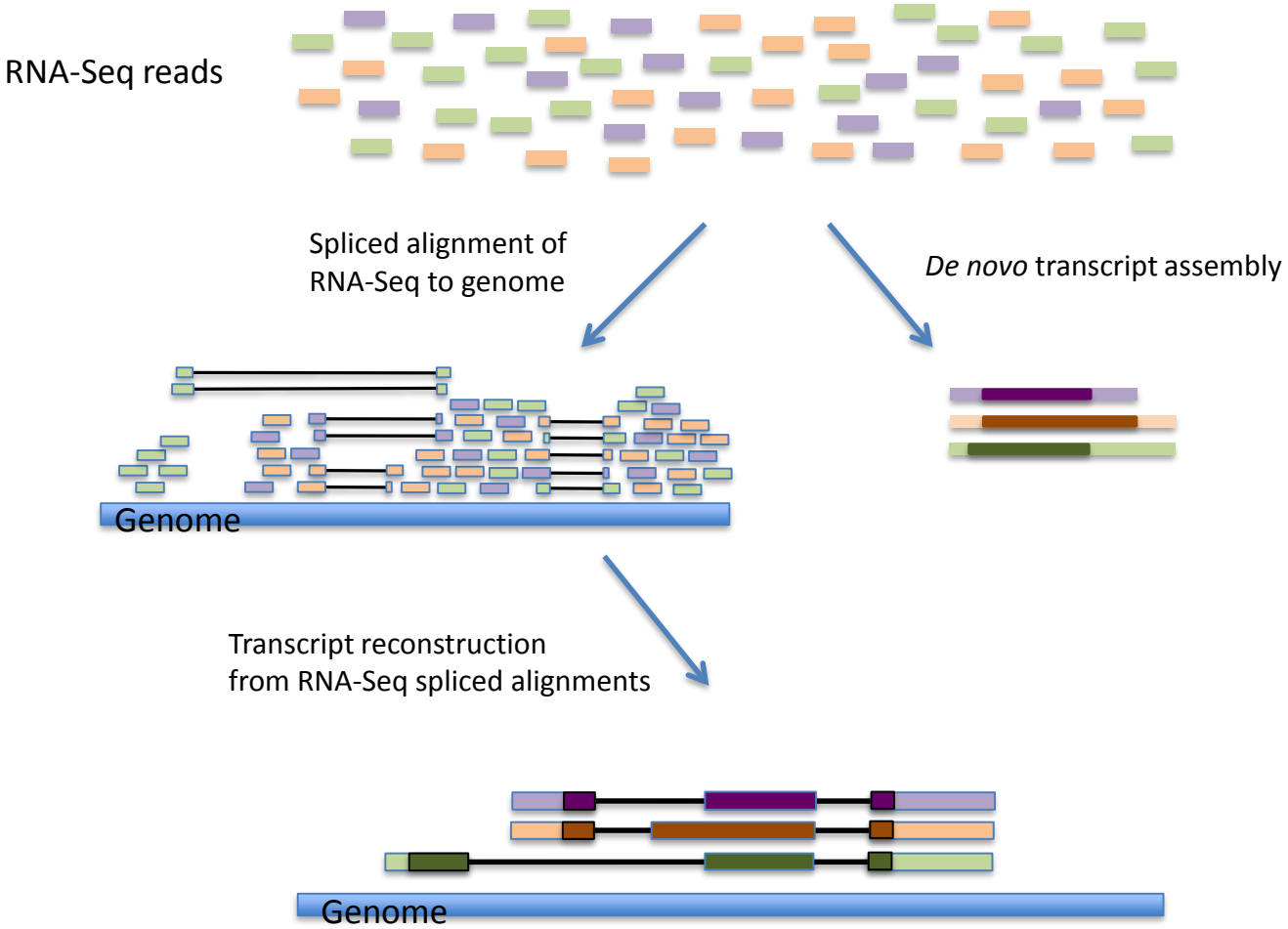
Contemporary strategies for transcript analysis from RNA-Seq



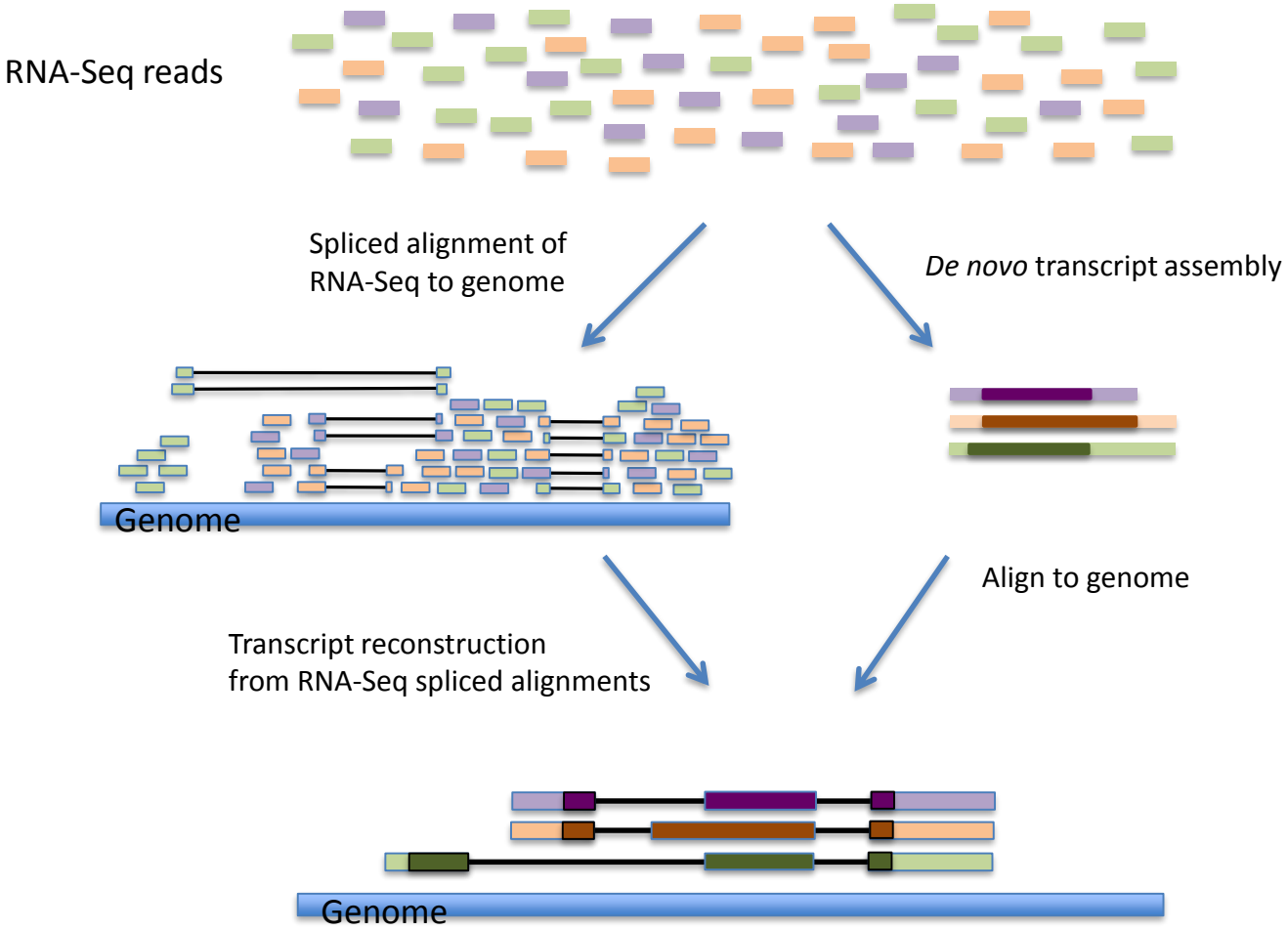
Contemporary strategies for transcript analysis from RNA-Seq



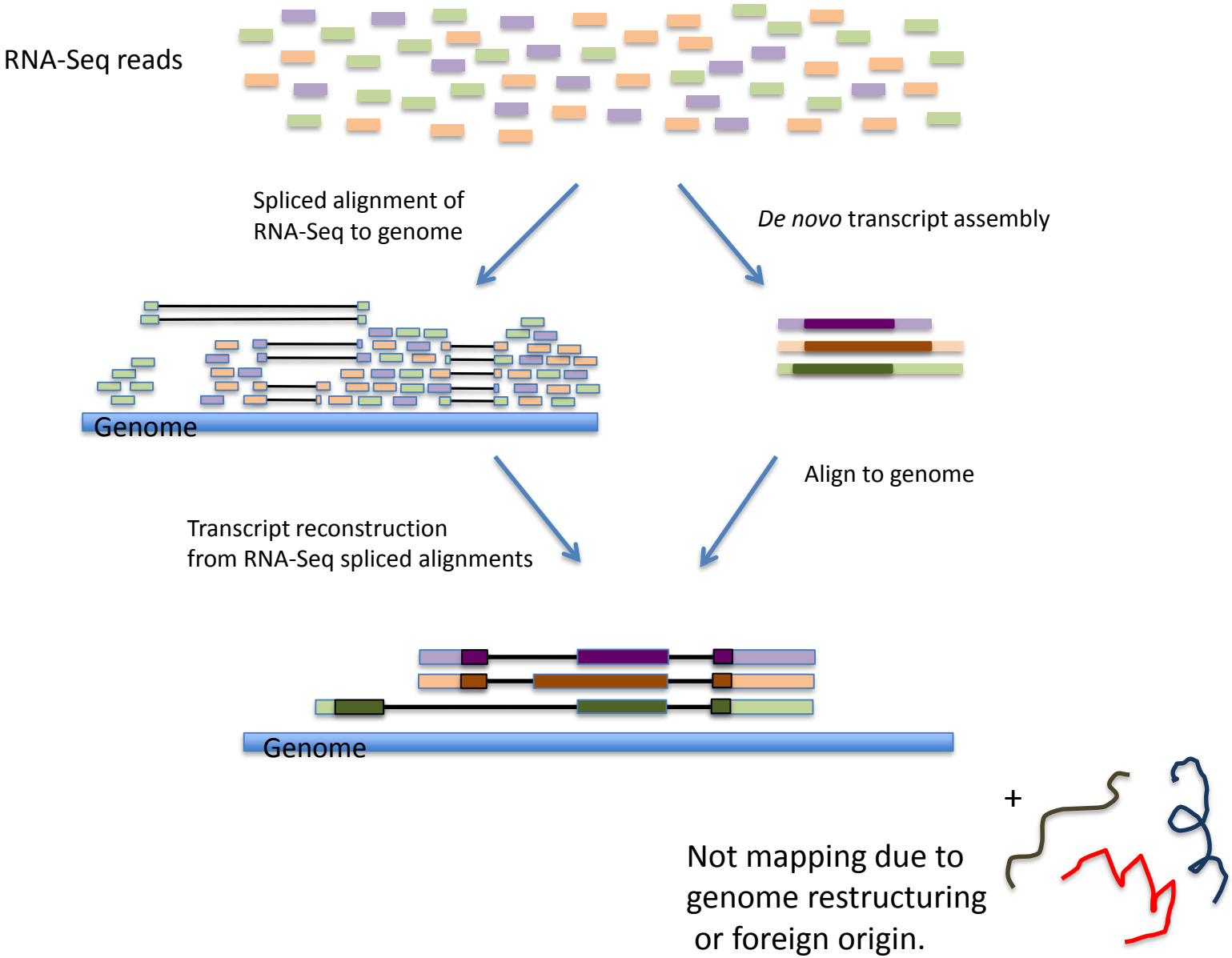
Contemporary strategies for transcript analysis from RNA-Seq



Contemporary strategies for transcript analysis from RNA-Seq

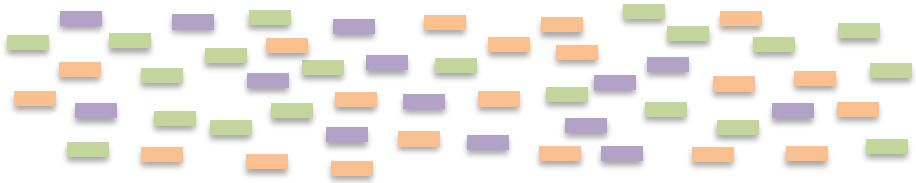


Contemporary strategies for transcript analysis from RNA-Seq



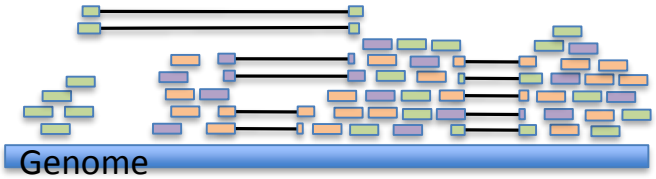
Contemporary strategies for transcript analysis from RNA-Seq

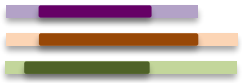
RNA-Seq reads



Spliced alignment of
RNA-Seq to genome

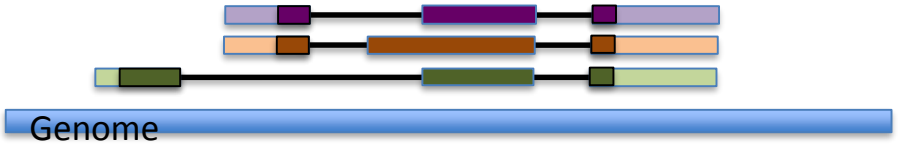
De novo transcript assembly



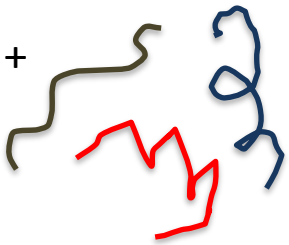


Align to genome

Transcript reconstruction
from RNA-Seq spliced alignments



+



Not mapping due to
genome restructuring
or foreign origin.

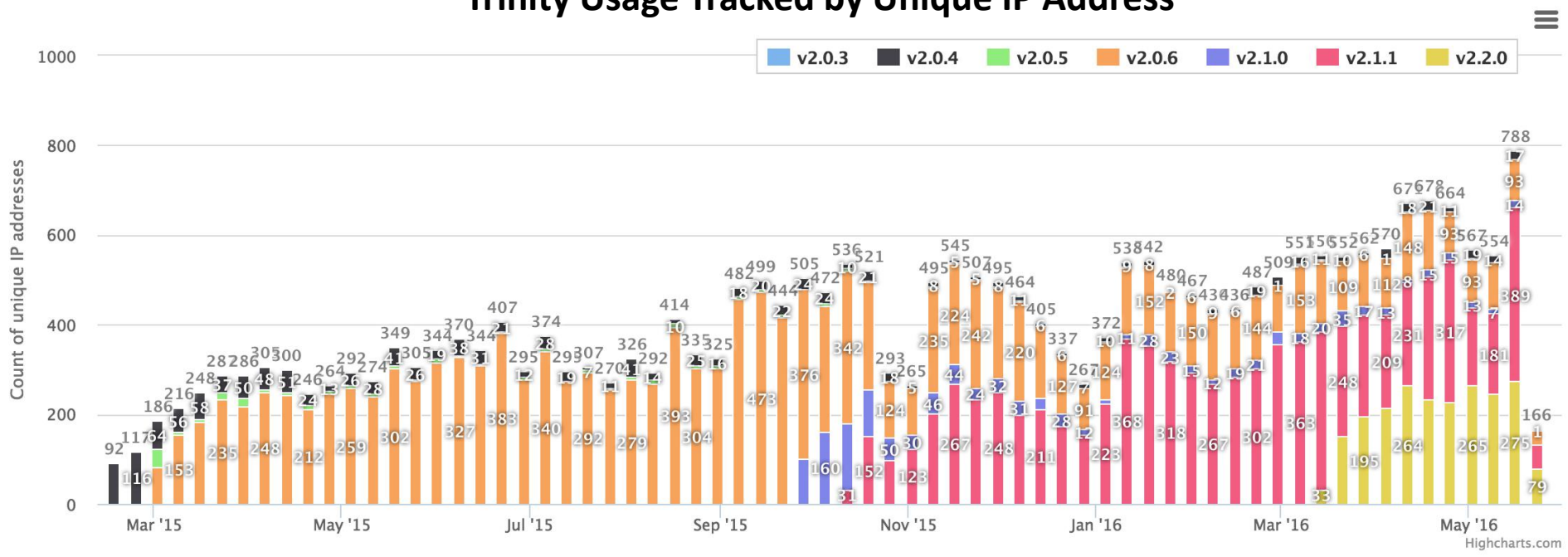
The Ever-Growing Trinity User Community

- ~1.5k unique users per month
- >3k literature citations (~20% cancer community)
- Open Source software development contributions from the Trinity community.

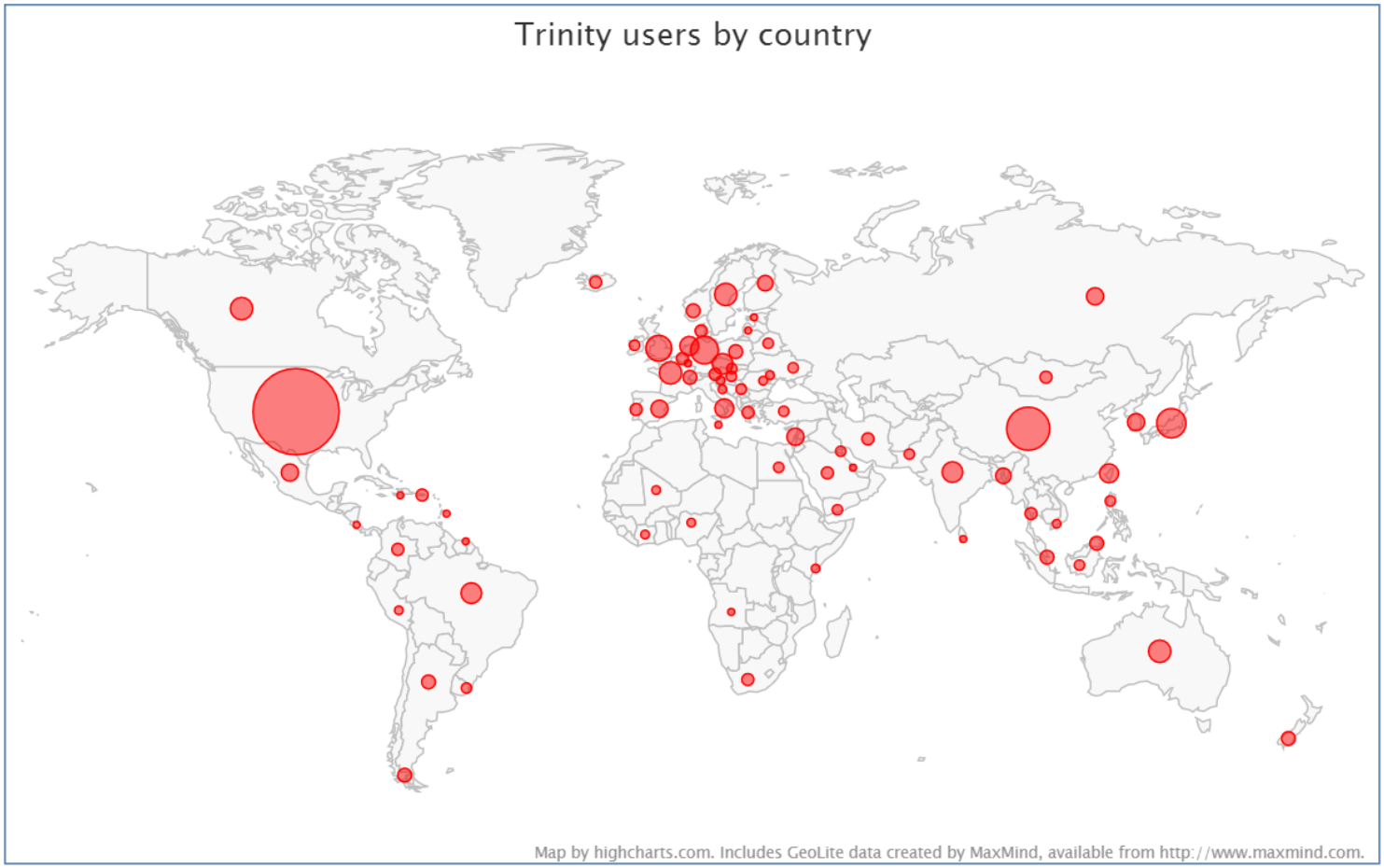
GitHub

http://trinityrnaseq.github.io

Trinity Usage Tracked by Unique IP Address



The Trinity Community is Global



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

Informatics Technology for Cancer Research

Cancer Transcriptome Analysis Toolkit

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

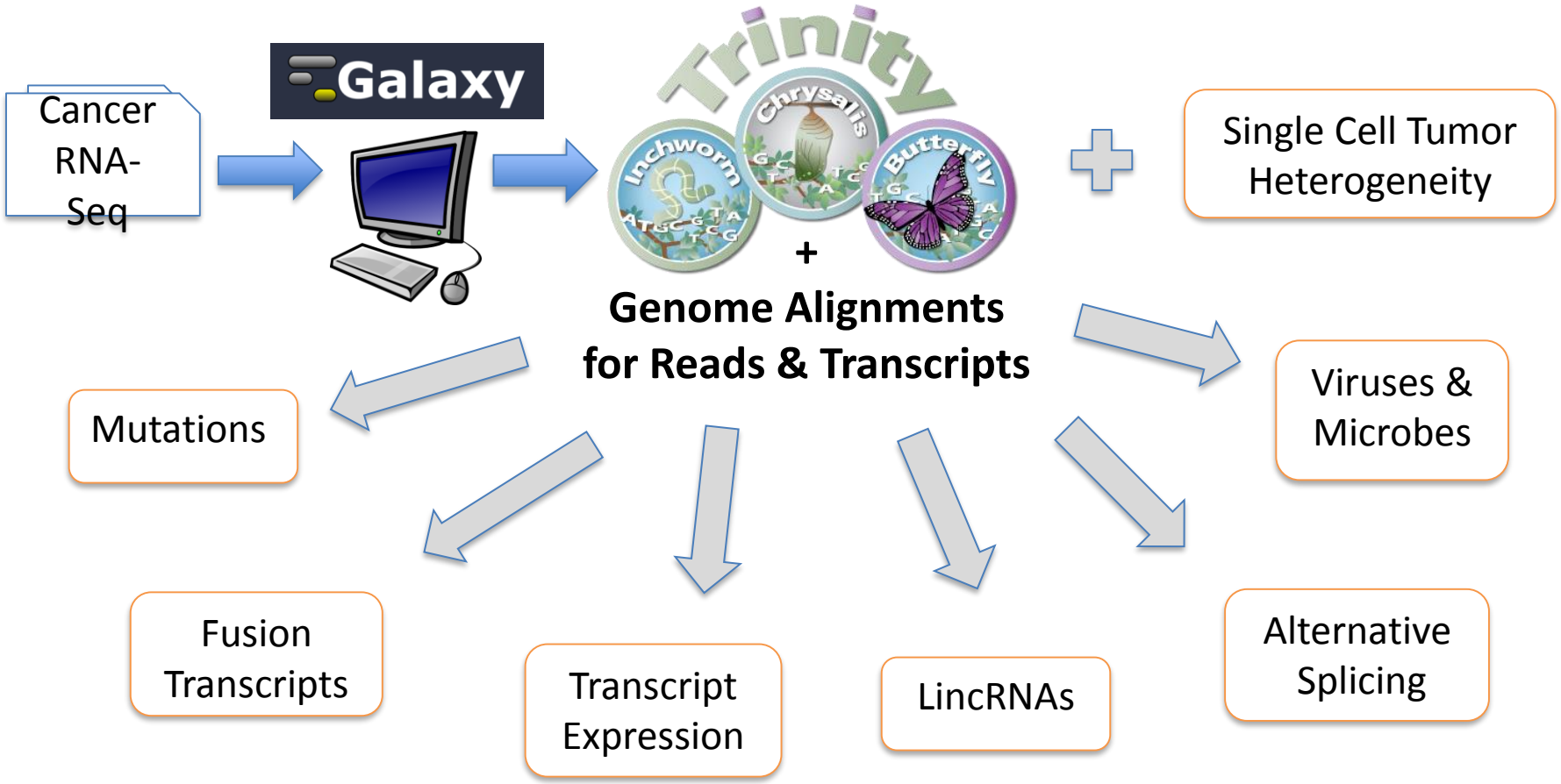
BROAD INSTITUTE

KLARMAN CELL OBSERVATORY

Trinity Cancer Transcriptome Analysis Toolkit

Genome Alignments for Reads & Transcripts

Trinity Cancer Transcriptome Analysis Toolkit



Trinity Cancer Transcriptome Analysis Toolkit

Cancer RNA-Seq

Galaxy

Trinity

Inchworm Chrysalis Butterfly

\+

Genome Alignments for Reads & Transcripts

\+ Single Cell Tumor Heterogeneity

Mutations

Viruses & Microbes

Fusion Transcripts

Transcript Expression

LincRNAs

Alternative Splicing

Interactive Visualizations and Summary Reports

Trinity Cancer Transcriptome Analysis Toolkit

Cancer RNA-Seq

Galaxy

Trinity

Inchworm Chrysalis Butterfly

+

Genome Alignments for Reads & Transcripts

Single Cell Tumor Heterogeneity

Mutations

Viruses & Microbes

Fusion Transcripts

Transcript Expression

LincRNAs

Alternative Splicing

Interactive Visualizations and Summary Reports

Mutation Detection Using RNA-Seq

Cancer RNA-Seq

\+

Genome Alignments for Reads & Transcripts

Mutations

Single Cell Tumor Heterogeneity

Viruses & Microbes

Fusion Transcripts

Transcript Expression

LincRNAs

Alternative Splicing

Interactive Visualizations and Summary Reports

Trinity CTAT Cancer Mutation Identification Module

RNA-Seq Fastqs

GATK Best Practices

Variant Annotation

CRAVAT

Karchin lab

Variant Filtration

MuPIT

Visualization

IGV.js

Mesirov lab

Galaxy

Indiana University

Custom visualization is a product of multiple labs efforts.

Mutation Analysis and Visualization from Within Galaxy

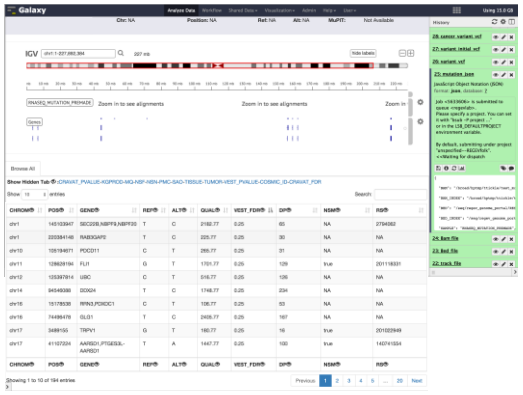
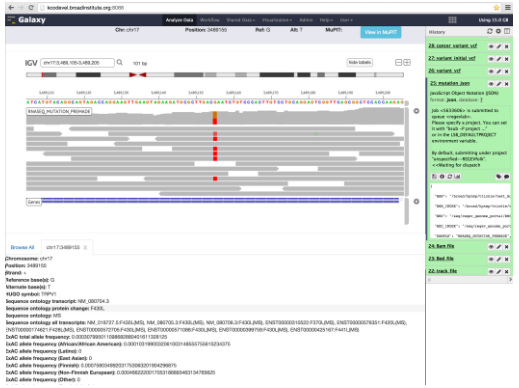
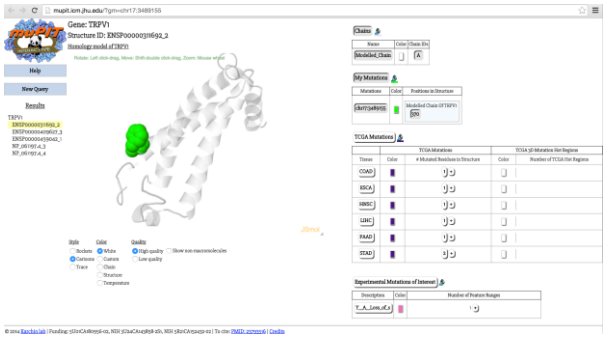


Table of Predicted Variants with scores, attributes and rankings.



Individual mutation report, including genome evidence view and annotations.



Mupit 3D protein structure view

(ITCR - Rachel Karchin and Mike Ryan)

** will demo **

Fusion Transcript Detection

Cancer RNA-Seq

\+

Genome Alignments for Reads & Transcripts

Fusion Transcripts

Transcript Expression

LincRNAs

Alternative Splicing

Mutations

Viruses & Microbes

Single Cell Tumor Heterogeneity

Interactive Visualizations and Summary Reports

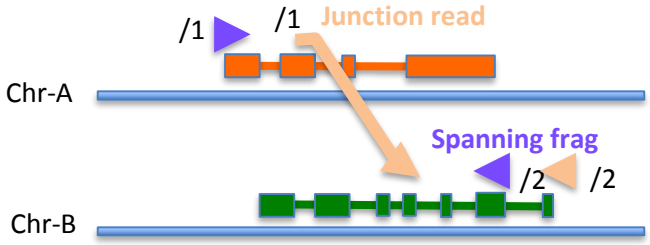
Top-down Approaches to Fusion Transcript Discovery

Paired-end Illumina RNA-Seq

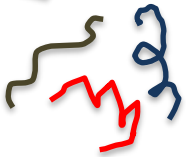
STAR-Fusion

* 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

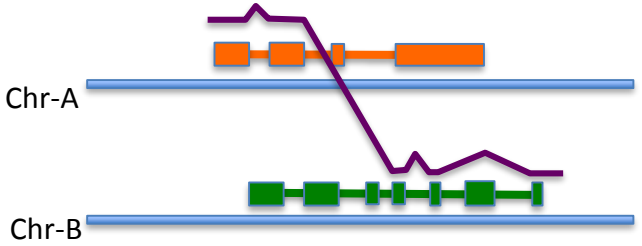


Trinity or Oases (MK)

GMAP-fusion

* In collaboration with Tom Wu, developer of GMAP

Align *transcripts* to genome, Identify Fusion Transcripts



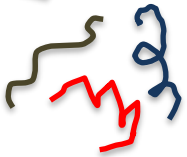
Top-down Approaches to Fusion Transcript Discovery

Paired-end Illumina RNA-Seq

STAR-Fusion

* In collaboration with Alex Dobin, developer of STAR

De novo RNA-Seq assembly



Trinity or Oases (MK)

DISCASM

Just discordant or unmapped reads

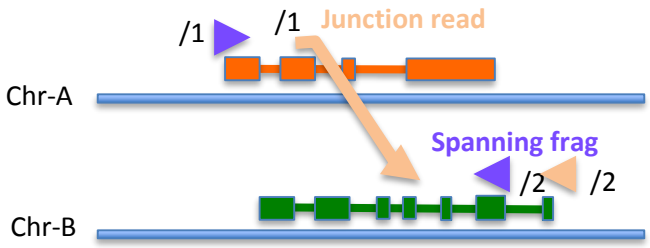
STAR-alignments

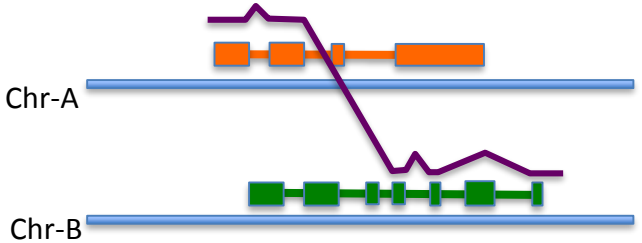
Align *reads* to the genome, Identify discordant pairs and junction/split reads.

GMAP-fusion

* In collaboration with Tom Wu, developer of GMAP

Align *transcripts* to genome, Identify Fusion Transcripts





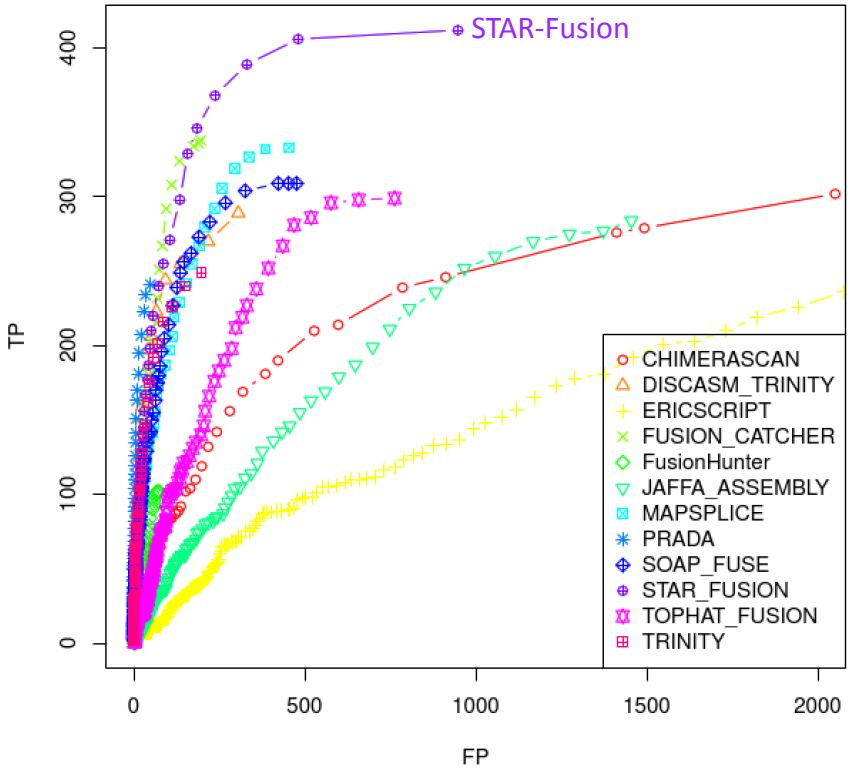
Compare to:

Prada	ChimeraScan	Ericscript	Jaffa
SoapFuse	Defuse	FusionHunter	
TophatFusion	FusionCatcher	Mapsplice	

Evaluation of Fusion-Finding Accuracy

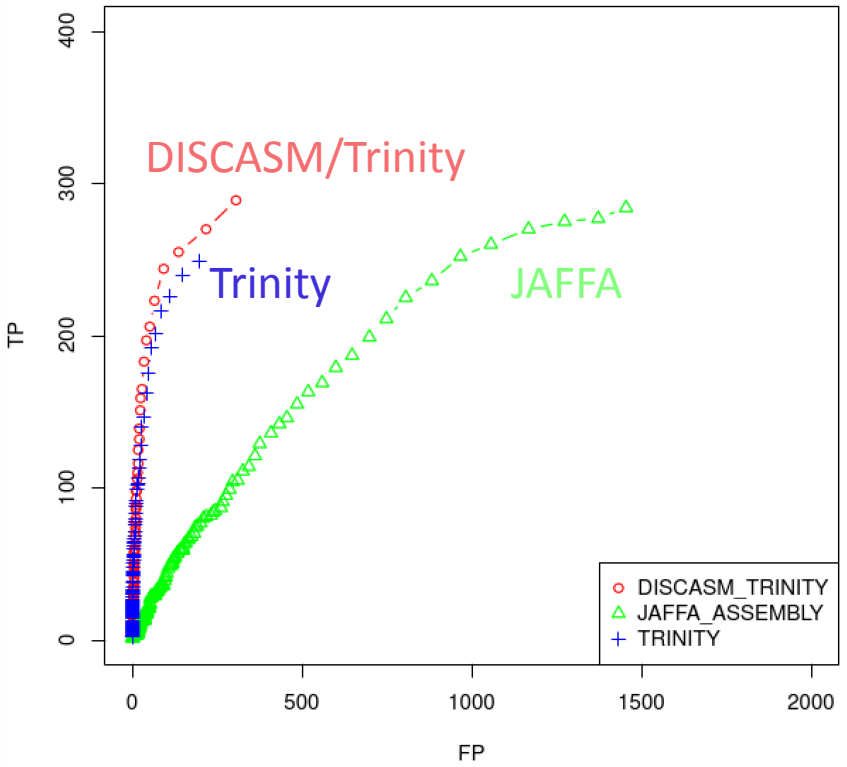
(using 75 Cancer Cell Lines, with TP = min 3 tools agree)

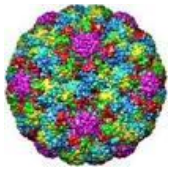
All Fusion Prediction Accuracies



Fusion predictions ranked according to min evidence support.

De novo Assembly-based Fusion Prediction Accuracy

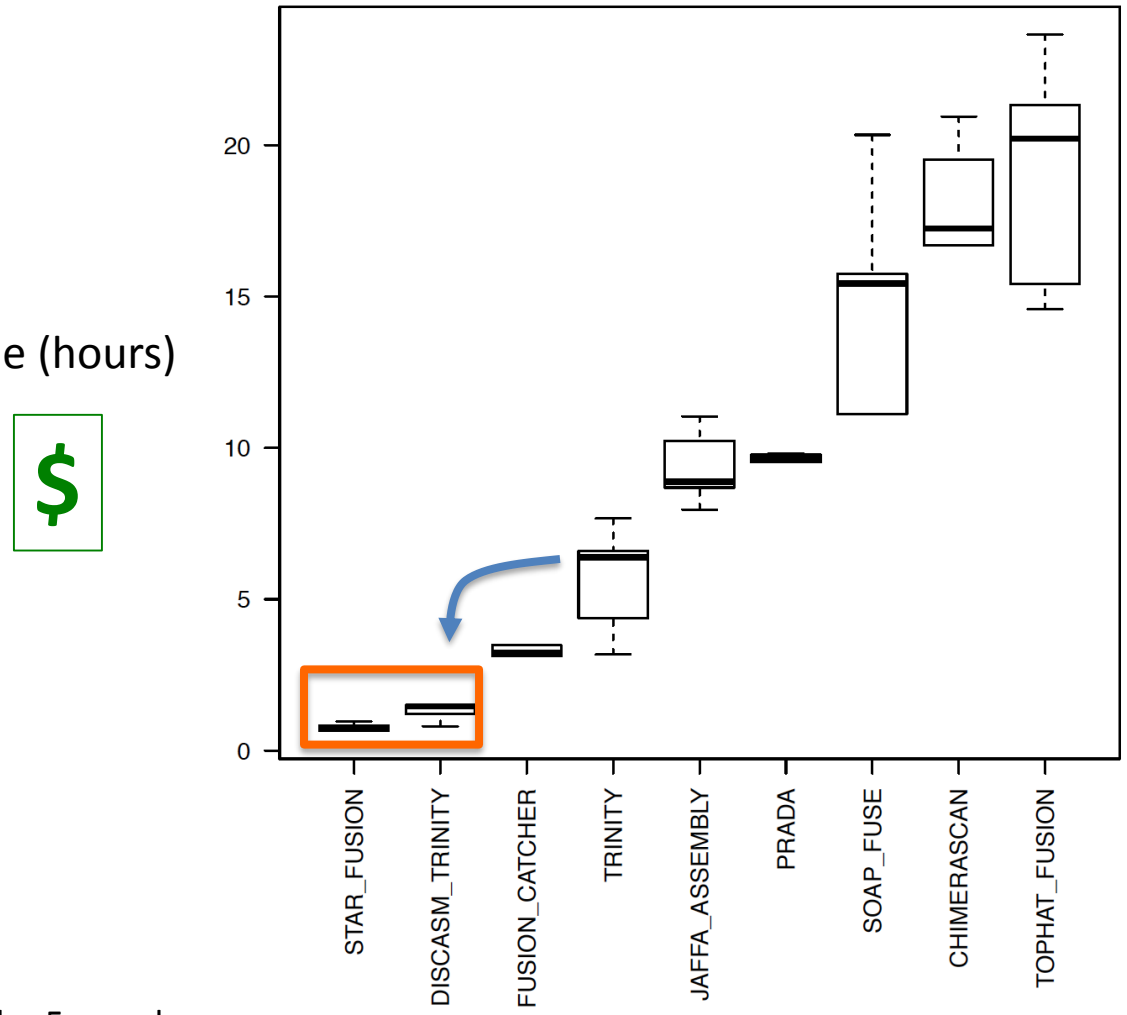




Cancer-associated Viruses and Microbiome

STAR-Fusion and DISCASM/Trinity Improve on both Speed and Accuracy of Fusion Detection

Time (hours)



Using 30M PE reads , 5 samples ea.

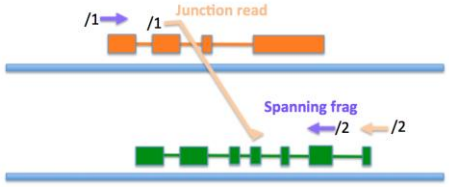
Bottom-up Fusion 'In silico Validation' Using FusionInspector

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

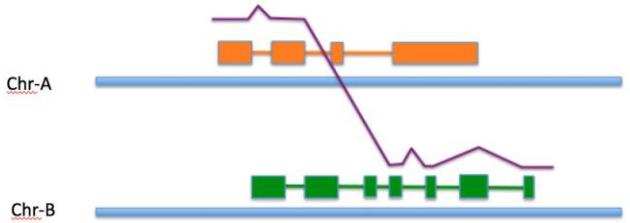
* STAR enhancements to support FusionInspector

Make mini-fusion contigs

STAR-Fusion
PRADA
SOAPfuse
FusionCatcher
...

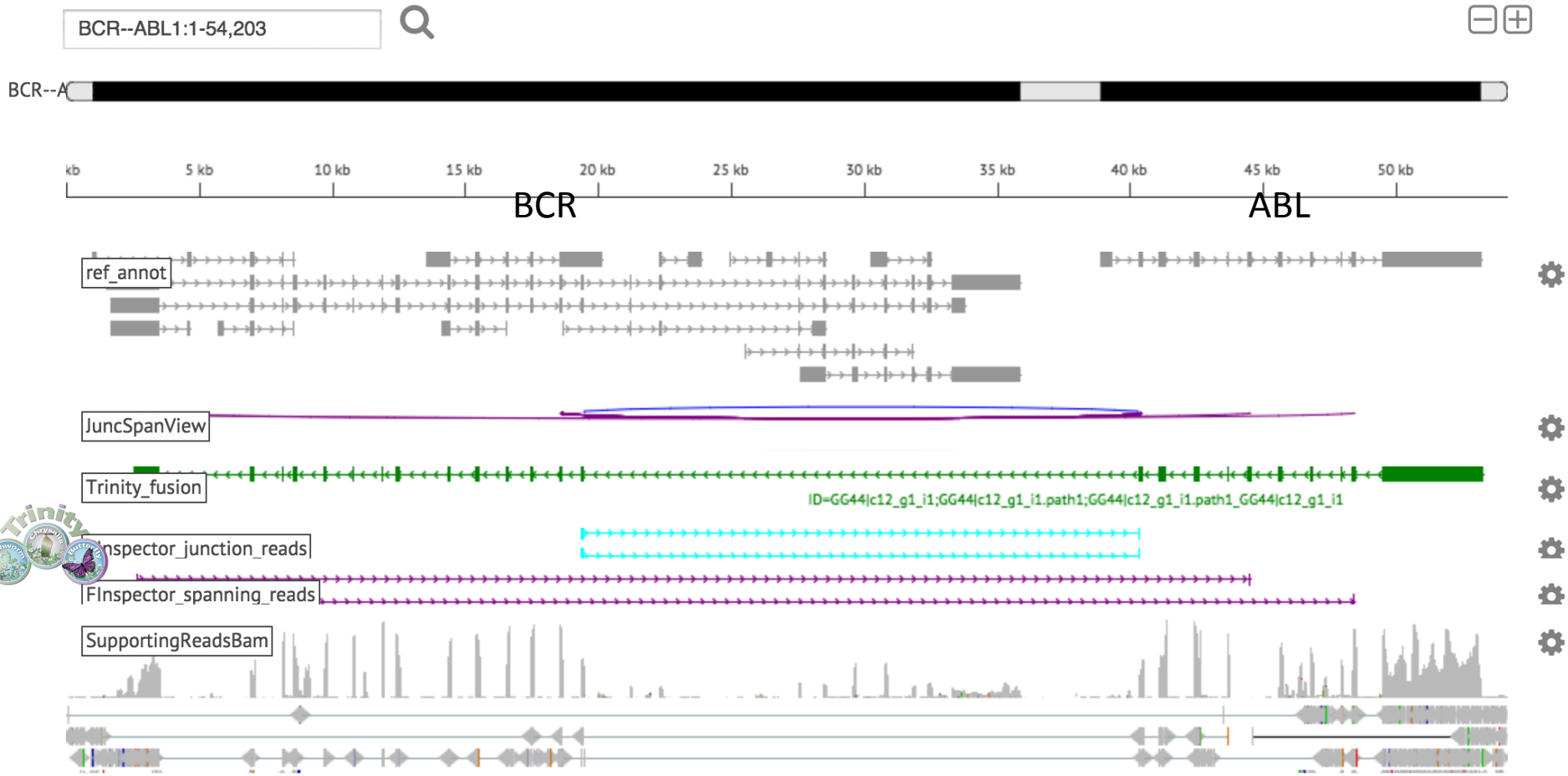


All fusion predictions



TrinityFuse
OasesFuse
JAFFA
DISCASM
...

FusionInspector Fusion View



Powered by IGV.js

ITCR - Jim Robinson and Jill Mesirov

LincRNA Identification

Cancer RNA-Seq

\+

Genome Alignments for Reads & Transcripts

Single Cell Tumor Heterogeneity

Viruses & Microbes

Mutations

Fusion Transcripts

Transcript Expression

LincRNAs

Alternative Splicing

Interactive Visualizations and Summary Reports

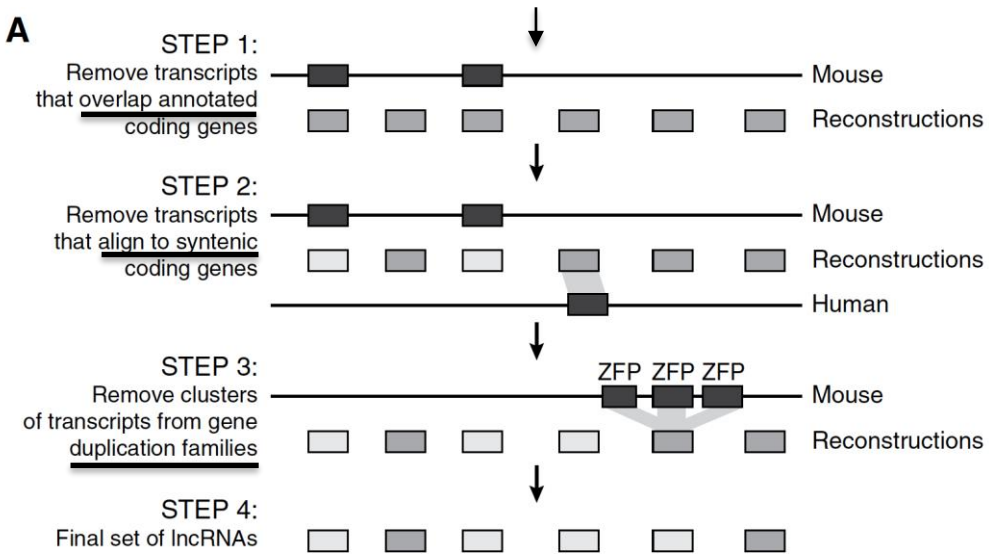
SLNCKY: LincRNA Identification from Reconstructed Transcripts

RESEARCH Open Access

Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs

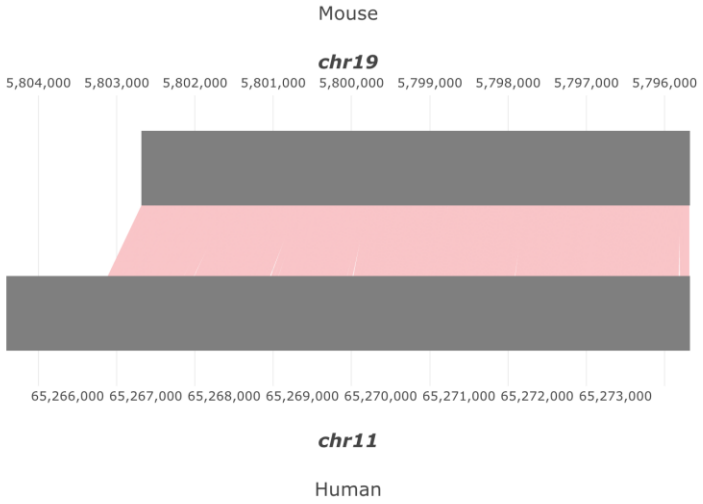
Jenny Chen, …, **Aviv Regev** & Manuel Garber; Genome Biology 2016

Transcripts (reconstructed from RNA-Seq)



Example:

Homo sapiens metastasis associated lung adenocarcinoma transcript 1 (MALAT1), non-coding RNA



Considers conserved ORFs and dN/dS

Freely available, open source: https://slncky.github.io/

SLNCKY - LncRNA Analysis and Visualization from Within Galaxy

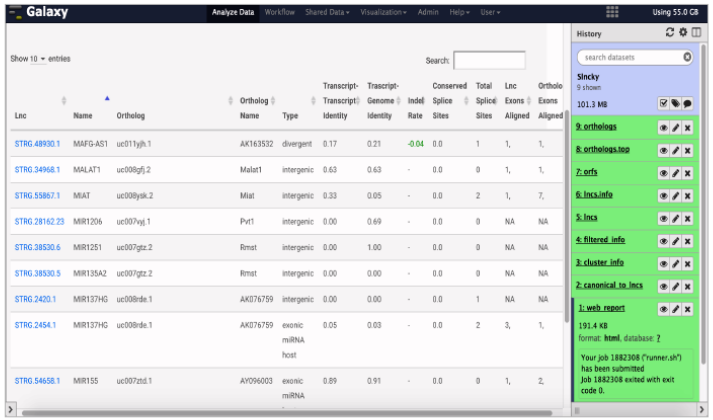
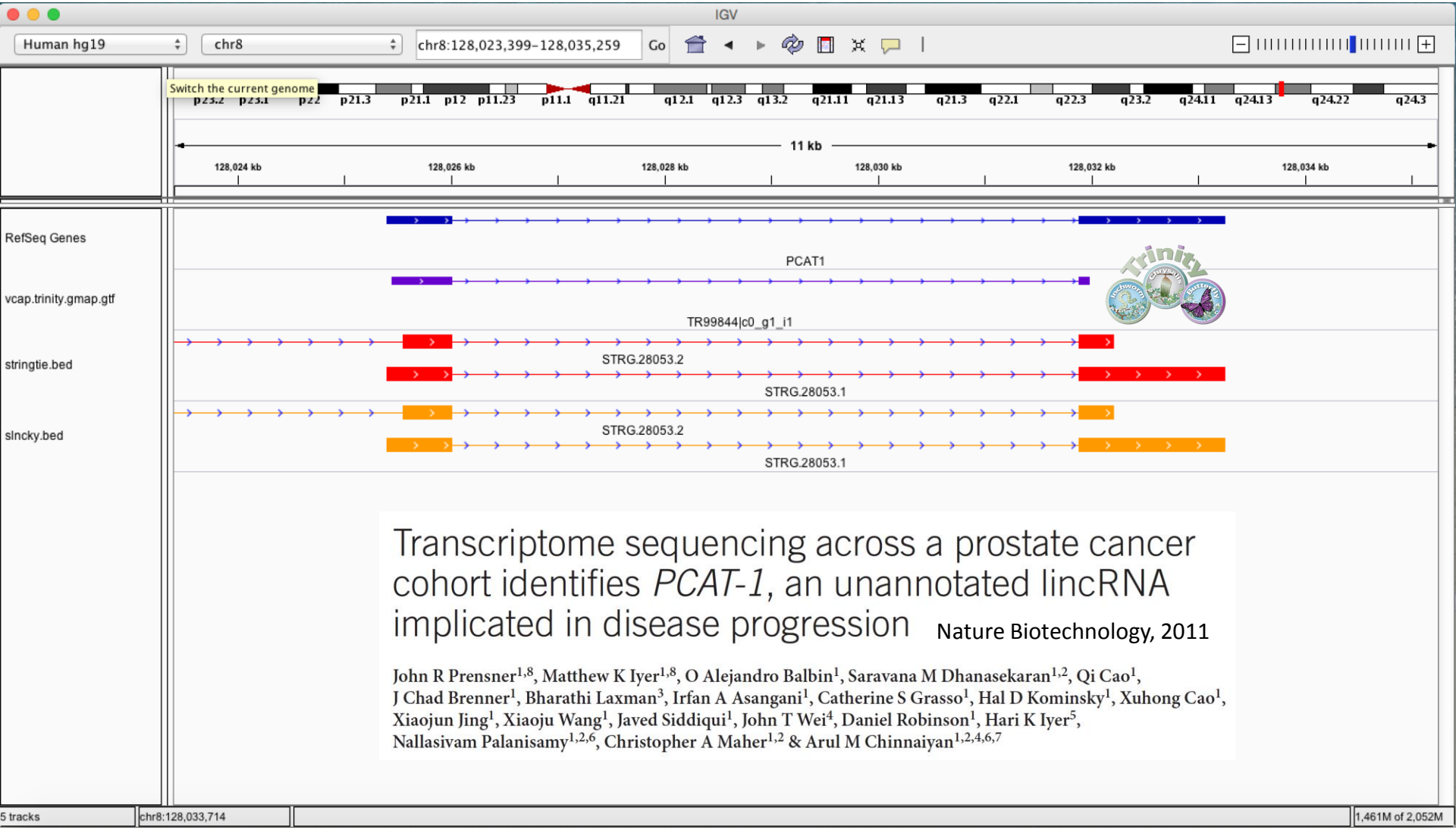


Table of Predicted LncRNAs with their orthologs

Individual LncRNA report, including alignment view and evolutionary metrics.

** will demo **

SLNCKY-based Re-discovery of PCAT1: Prostate Cancer Associated Transcript 1



Transcriptome sequencing across a prostate cancer cohort identifies *PCAT-1*, an unannotated lincRNA implicated in disease progression

Nature Biotechnology, 2011

John R Prensner[1,8], Matthew K Iyer[1,8], O Alejandro Balbin[1], Saravana M Dhanasekaran[1,2], Qi Cao[1], J Chad Brenner[1], Bharathi Laxman[3], Irfan A Asangani[1], Catherine S Grasso[1], Hal D Kominsky[1], Xuhong Cao[1], Xiaojun Jing[1], Xiaoju Wang[1], Javed Siddiqui[1], John T Wei[4], Daniel Robinson[1], Hari K Iyer[5], Nallasivam Palanisamy[1,2,6], Christopher A Maher[1,2] & Arul M Chinnaiyan[1,2,4,6,7]

Single Cell Tumor Heterogeneity

Cancer RNA-Seq

\+

Genome Alignments for Reads & Transcripts

Single Cell Tumor Heterogeneity

Mutations

Viruses & Microbes

Fusion Transcripts

Transcript Expression

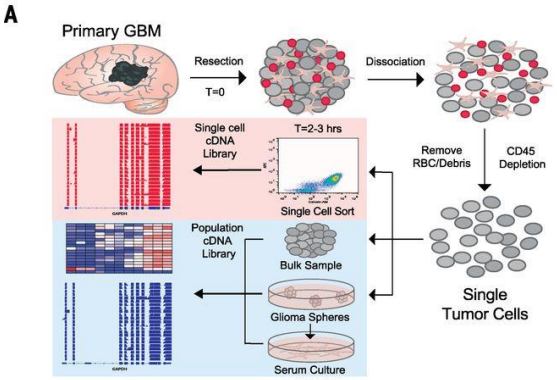
LincRNAs

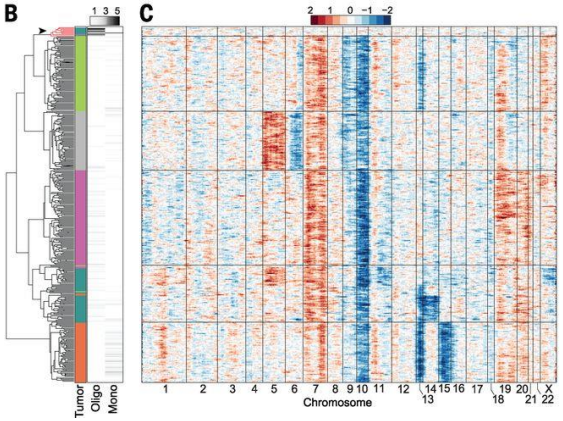
Alternative Splicing

Interactive Visualizations and Summary Reports

Single Cell Resolution of Tumor Heterogeneity via RNA-Seq

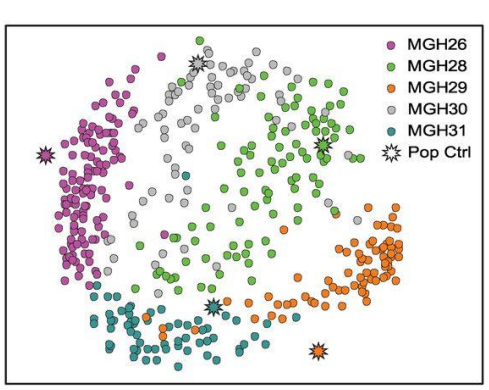
Single Cell Workflow

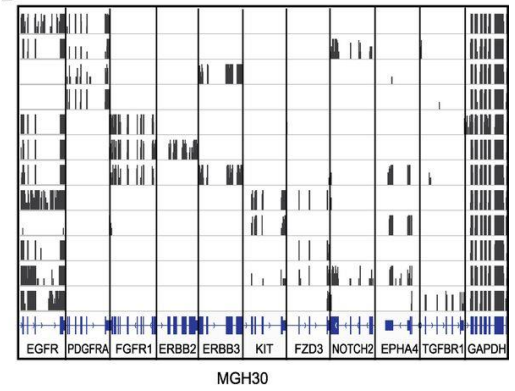




Chromosome gain/loss

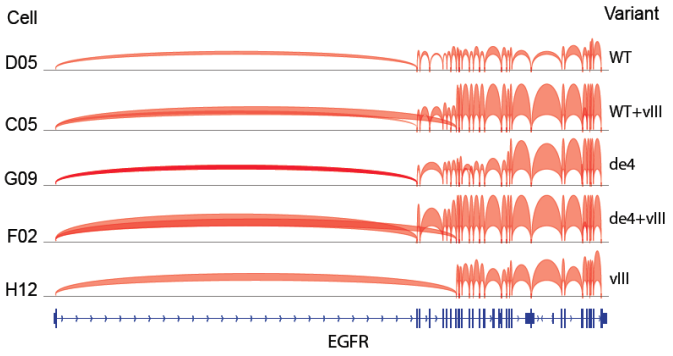
Expression Heterogeneity

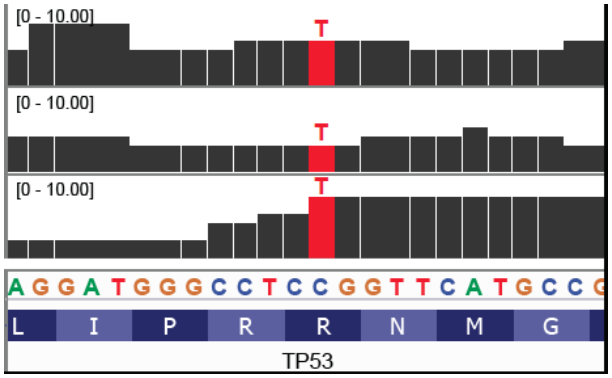




Differential Expression of surface receptors

Splicing Aberrations in EGFR





TP53 Mutations

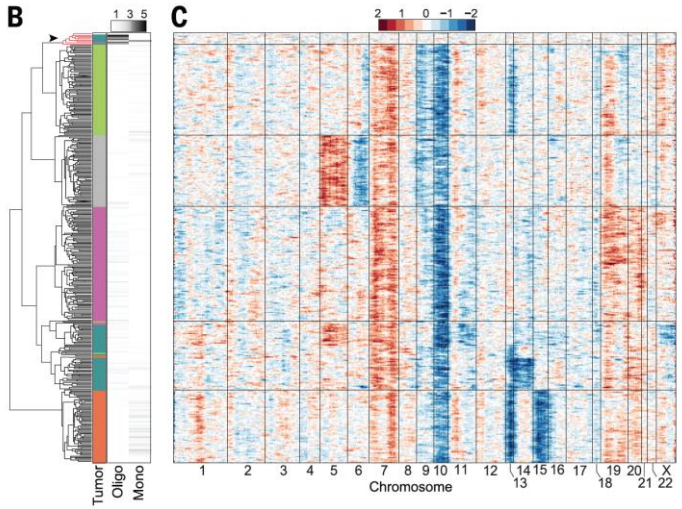
KLARMAN CELL OBSERVATORY

A. Patel, **I. Tirosh**, …, A. Regev, B. Bernstein. Science, 2014

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

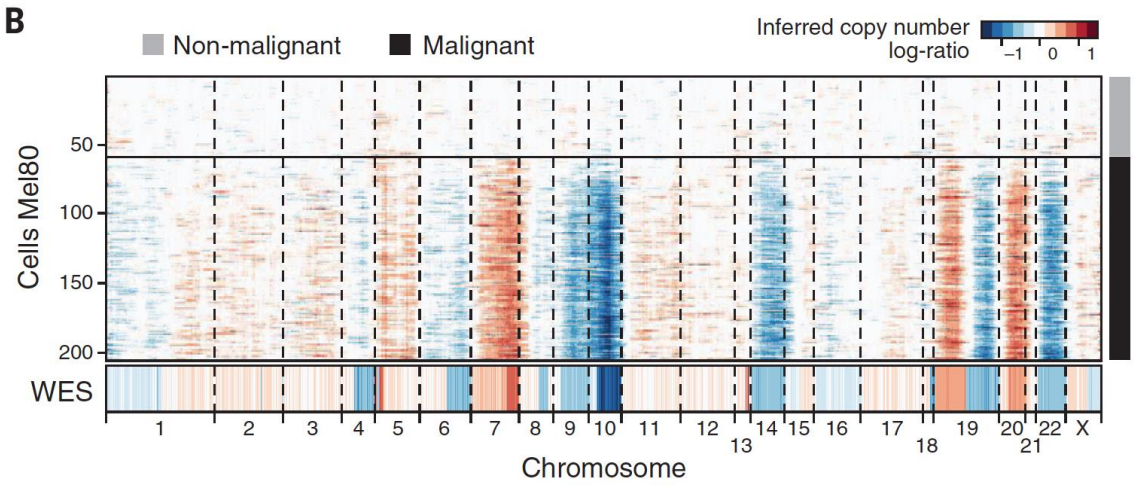
Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma

Patel, Tirosh, …, **Regev**, Bernstein; Science 2014



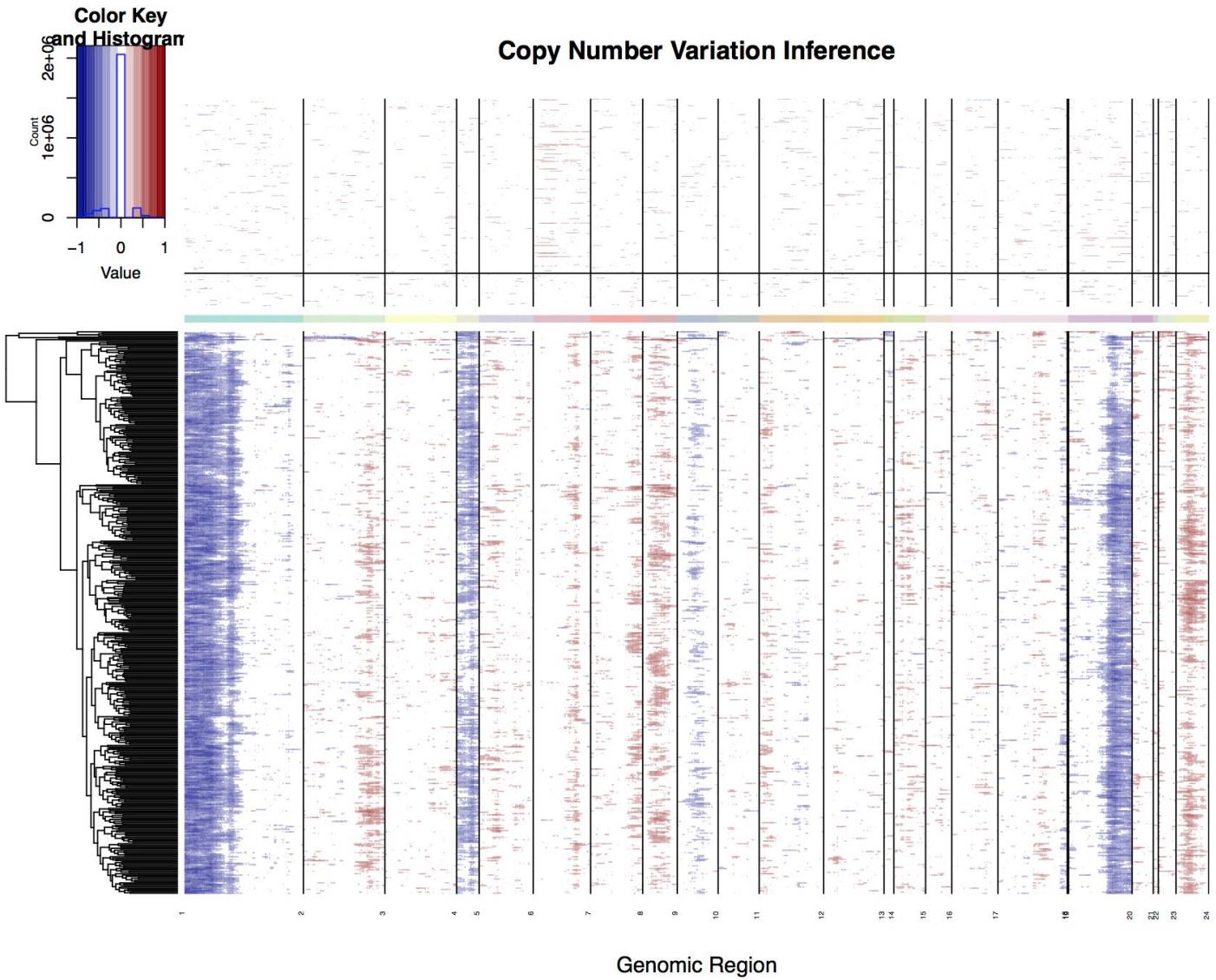
Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq

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

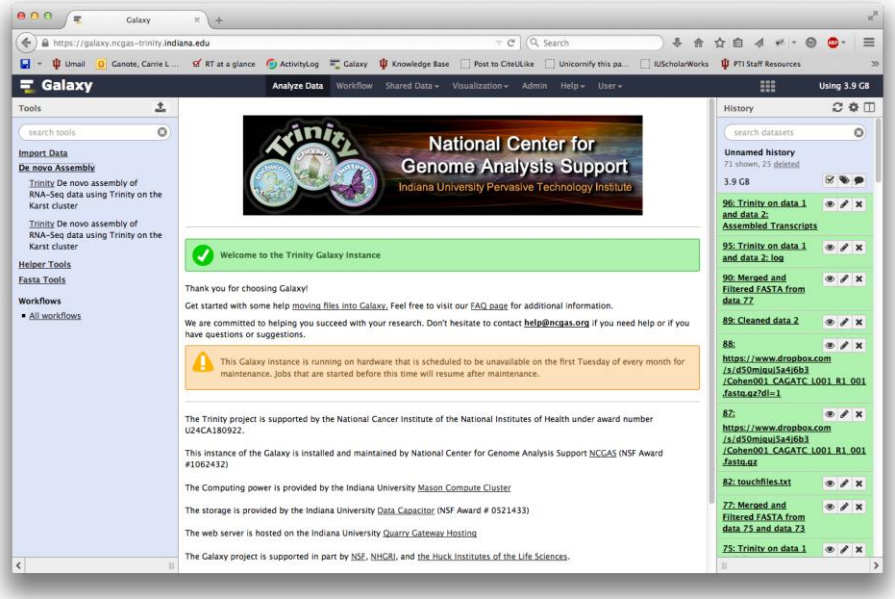


CTAT InferCNV by Tim Tickle and Itay Tirosh

Goal: cancer transcriptome toolkit accessible to ***any*** cancer researcher

Starting point: RNA-Seq data (fastq files)

Access Trinity CTAT via Galaxy

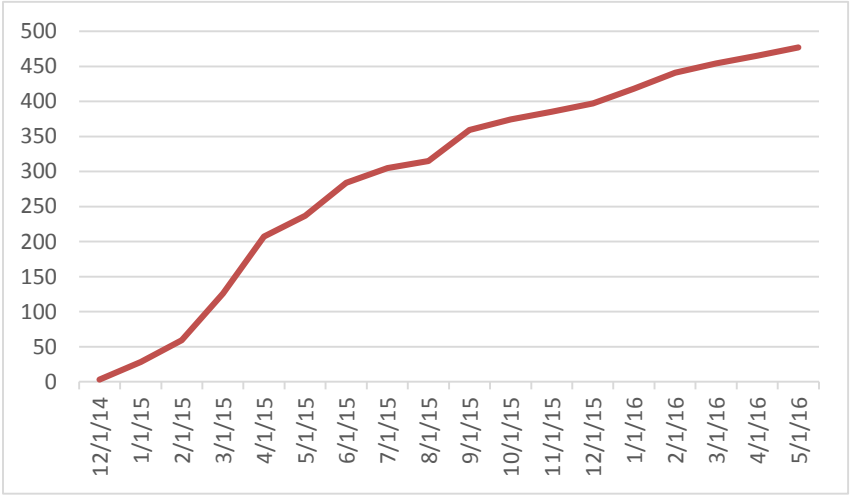


The **National Center for Genome Analysis Support** hosts the public web interface for running Trinity jobs.

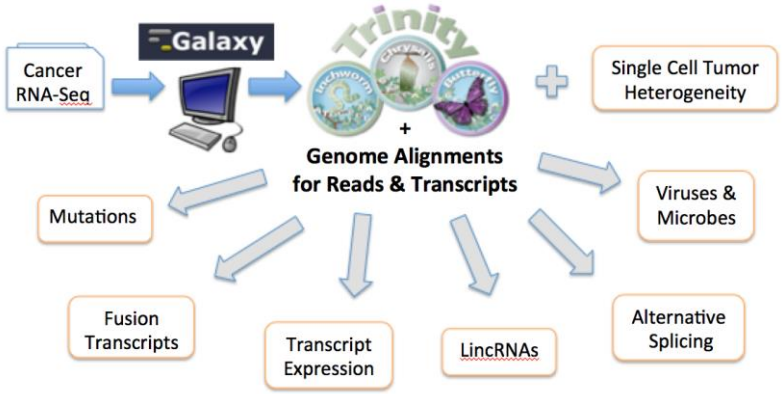
- Backed by three devoted nodes running on the Karst system with 512GB memory each

We're steadily growing since the official launch in January, 2015

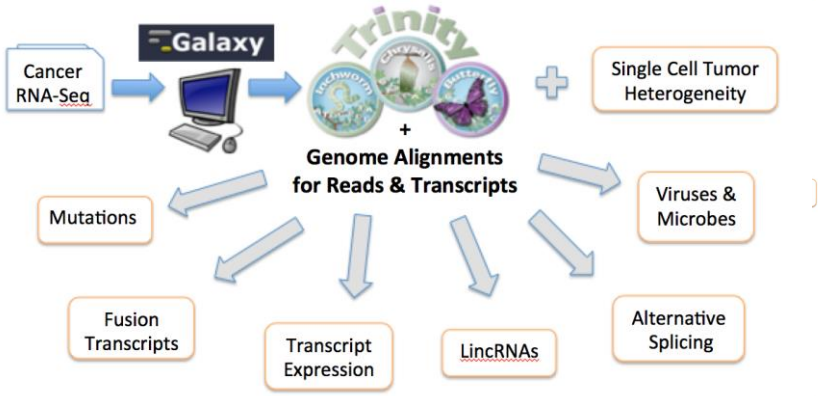
Total Galaxy Users per Month



NATIONAL CENTER FOR GENOME ANALYSIS SUPPORT
INDIANA UNIVERSITY



Scaling from individual samples



Scaling from individual samples **to many samples**

Firecloud

Scalable Cancer Computing Solution

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

Also enables integration of Trinity CTAT into:

Clinical Research Sequencing Platform (CRSP)

BROAD INSTITUTE | GENOMIC SERVICES

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
Ami-levy Moonshine

Jill Mesirov
James Robinson

Alex Dobin

Nathalie Pochet
Nik Obholzer

Tom Wu

Cathy Wu
Jing Sun
Peggy Hsu
Jintaek Kim
Sachet Shukla
Dan Landau

Bill Barnett
Thomas Doak
Carrie Ganote
Robert Henschel
Ben Fulton