Personal Oncology with Galaxy

Jeremy Goecks, H. Jean Khoury, Bassel El Reyes, James Taylor, and Michael R. Rossi
Emory University

An Approach for Personal Oncology

Tumor Analysis Using High-throughput Sequencing

Targeted exome cancer panel of ~600 significant genes
variants
Whole transcriptome / RNA-seq
variant validation, gene expression, fusions

Match Tumor to Similar Cancer Cell Line(s)

Match by variants
Match by expression levels
Match using combination of variants and gene expression levels

Choose Drugs using Known Associations, Public Databases, and High-throughput Screening

Galaxy Visualization & Visual Analysis

Features

› Completely Web-based approach: data fetched as needed from server and rendered in Web browser
› Scales as needed to exceptionally large datasets
› Uses D3 and HTML5 Canvas
› Framework for adding Web-based visualizations to Galaxy
› Can share fully-functional visualizations with the Web
› Can integrate Galaxy tools and visualization for visual analysis

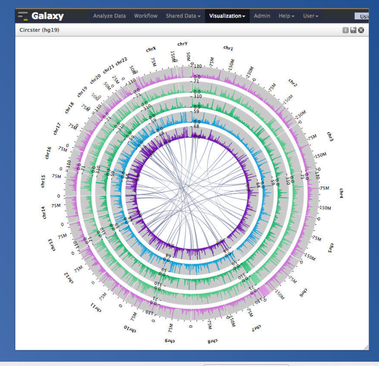

Figure: left: partial parameter sweep for variant calling tool reveals how called variants change in response to parameter changes; lower right: Web-based genome browser showing a variant called from mapped exome reads; top right: interactive Circos plot for tumor gene expression datasets and fusions.

Galaxy Workflows

Features

› Workflows for:
- variant calling in exome or transcriptome sequencing
- fusions identification, and gene expression analysis in transcriptome sequencing
- annotating and reducing variants
- combining called variants with gene expression results

› Workflows are sharable, publishable, downloadable, and completely reproducible

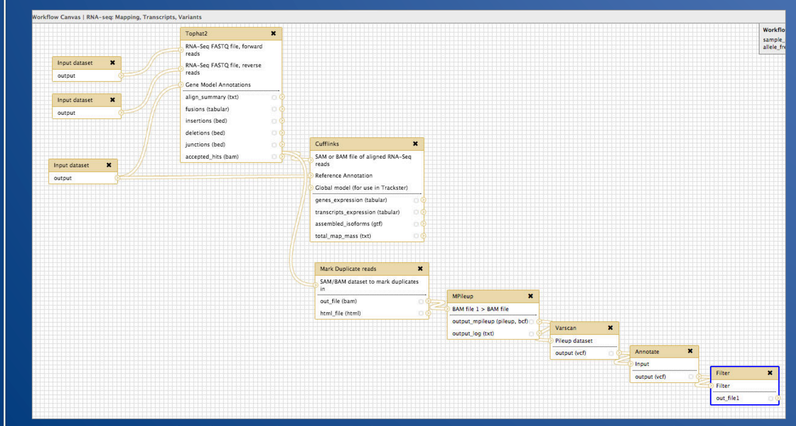

Figure: Galaxy workflow for analyzing tumor RNA-seq data: reads are mapped, gene expression is quantified, and variants are called, annotated, and filtered.

Data and Results

High-throughput Sequencing Data

› 3 common cancer cell lines: MiaPaCa2, PANC1, HPAC: targeted exome + whole transcriptome
› 6 primary pancreatic tumors: whole transcriptome

Results

› Genome-wide expression highly similar across cell lines and patients (Spearman ρ = 0.77), but gene expression similarity amongst cancer panel genes much less (Spearman ρ = 0.51), so likely better to use expression amongst panel genes only to choose similar but distinct cell lines

› Cancer panel gene expression reveals strong correlations
- one patient shows very highly correlation with cell lines in cancer panel (Pearson r = 0.66 vs. HPAC, 0.48 vs. PANC1)
- three patients show extremely high correlation (Pearson r = ~0.9) with each other but no correlation with cell lines

› One very good overall match between a patient and HPAC cell line in cancer panel genes: large number of shared variants (65% shared) and highly correlated gene expression (Pearson r = 0.66)

More Information

The Galaxy Project: http://galaxyproject.org
Public Galaxy Webserver: http://usegalaxy.org
jeremy.goecks@emory.edu

EMORY UNIVERSITY

Galaxy