

# JBrowse

Programming for Biology 2016  
CSHL

Scott Cain

Ontario Institute for Cancer Research

GMOD Coordinator

WormBase Developer

[scott@scottcain.net](mailto:scott@scottcain.net)

# What is GMOD?

- A set of interoperable open-source **software** components for visualizing, annotating, and managing biological data.
- An active **community** of developers and users asking diverse questions, and facing common challenges, with their biological data.

# Who uses GMOD?



# Other GMOD Projects

- GBrowse - “old school” Genome Browser, based on BioPerl
- MAKER - Tool for doing whole genome *de novo* annotation
- Chado - Organism-agnostic database schema for genomics and related data
- Apollo - Web-based feature annotation editor based on JBrowse
- Tripal - Web framework based on Drupal/PHP; provides user interface to data in Chado
- InterMine - Very powerful query engine/web UI for biological data (loaders for GFF, Chado)
- Galaxy - Very powerful workflow editor—lets you create and easily rerun complex workflows.

# What are Genome Browsers good for?

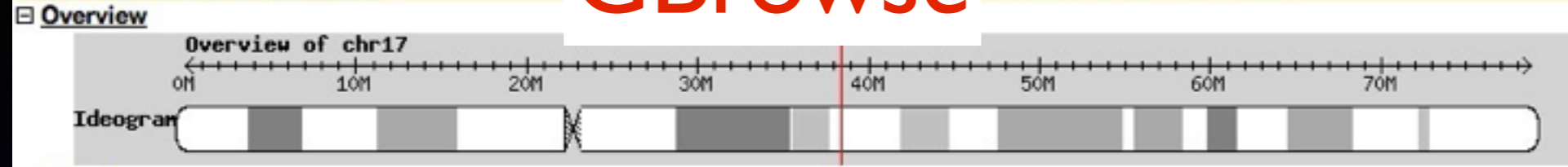
- Visualizing dense data from a multitude of sources (genes from a GFF file, RNASeq data from a BAM file, variation data from a VCF file)
- Homology and gene expression support for gene models

# GBrowse

Landmark or Region: chr17:38350227..384002; Search

Reports & Analysis: Download GFF File... Configure... Go

Data Source: Welcome to eqtl.uchicago.edu! kbp: + >>> Flip



Details

Entrez genes

NM\_025267  
AARS1: alanyl-tRNA synthetase domain containing 1

NM\_173079  
RUNDC1: RUN domain containing 1

Degner, Pai, Pique-Regi et al. (2012):  $-\log_{10}(P)$ , LCLs, 70 Nigerian HAPMAP ids, DNase sensitivity QTLs (dsQTLs) by

Schadt et al. (2007):  $-\log_{10}(P)$ , liver, 427 ids, European descent

Myers et al. (2007):  $-\log_{10}(P)$ , cortex from control brain, 279 ids, European descent

Stranger et al. (2007):  $-\log_{10}(P)$ , LCLs, 210 HAPMAP ids, 4 single populations.

Veyrieras et al. (2008):  $-\log_{10}(P)$ , LCLs, 210 HAPMAP ids, multi-population.

Veyrieras et al. (2008): posterior probability, LCLs, 210 HAPMAP ids, multi-population.

Pickrell et al. (2010):  $-\log_{10}(P)$ , LCLs, 69 Nigerian HAPMAP ids, RNA-Seq for eQTLs.

Innocenti et al. (2011):  $\log_{10}(\text{Bayes Factor})$ , Liver, 266 ids, RNA-Chip for eQTLs.

Pickrell et al. (2010):  $-\log_{10}(P)$ , LCLs, 69 Nigerian HAPMAP ids, RNA-Seq for splicing QTLs.

Montgomery et al. (2010):  $-\log_{10}(P)$ , LCLs, 60 European HAPMAP ids, RNA-Seq for transcript QTLs.

Zeller et al. (2010):  $-\log_{10}(P)$ , Monocytes, 1,490 ids recruited in Germany.

Montgomery et al. (2010):  $-\log_{10}(P)$ , LCLs, 60 European HAPMAP ids, RNA-Seq for exon QTLs.

Dinas et al. (2009):  $-\log_{10}(P)$ , Fibroblasts, 75 Europeans

Dinas et al. (2009):  $-\log_{10}(P)$ , LCLs, 75 Europeans

Dinas et al. (2009):  $-\log_{10}(P)$ , T-cells, 75 Europeans

Track height: Drag/Select:

32.40 Mb .80 Mb

820.11

< SNORA  
< N4BP2L1

P4  
RP11-380B4

32.40 Mb 32.60 Mb 32.80 Mb

processed transcript  
RNA gene

cb 32.38Mb 32.40Mb

IFIT1P1-001 >  
processed pseudogene

BRCA2-002 >  
nonsense mediated decay

BRCA2-004 >  
processed transcript

BRCA2-005 >  
nonsense mediated decay

BRCA2-006 >  
retained intron

AL445212.9 >

AL137247.14 >

< N4BP2L1-001  
protein coding

- Many “specialty” browsers (eg, Biodalliance, Savant)

# Why Install Your Own?

- You have data no one else has
- You want to be able to share it with your group, community, the world (potentially with “less savvy” users)
- You want to have control over how it looks

# Why JBrowse?

- (Fairly) Easy install
- Good user experience (getting close to a browser-desktop hybrid)
- Good community support (mailing lists, tutorials, software updates)



# Installation

- Only requires:
  - Web server (apache, lighttpd, nginx, etc)
  - Perl/make/standard unix-y tools
  - Really easy to get via AWS (Docker too, but we won't be using that)

# JBrowse Attributes

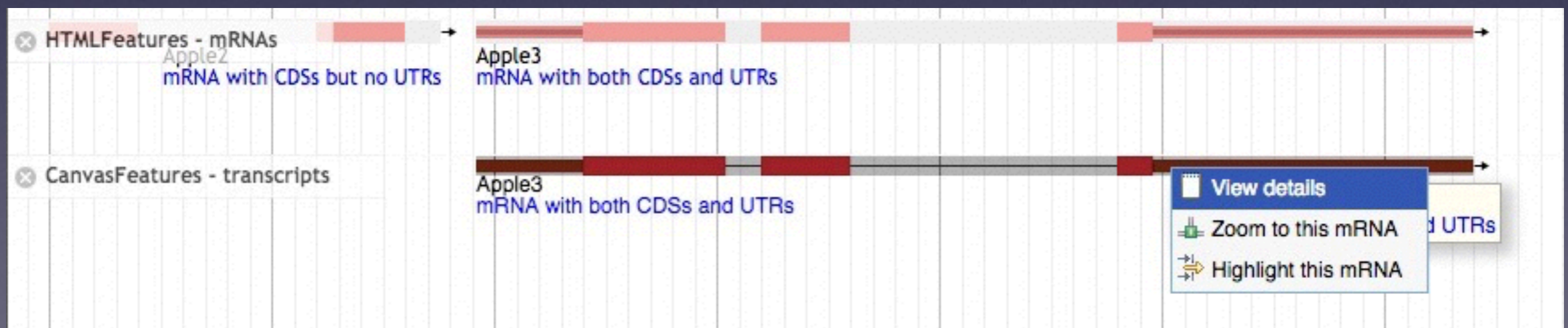
- Do everything possible on the client side, in JavaScript.
- Fast, smooth navigation.
- Supports GFF3, BED, Bio::DB::\*, Chado, WIG, BAM, BigWig, VCF, and UCSC import (intron/exon structure, name lookups, quantitative plots).
- Is stably funded by NHGRI.
- Is open source, of course.
- Did I mention it's fast?

# The JBrowse Project

- free and open source (license: LGPL / Artistic)
- a GMOD project
  - <http://gmod.org>
- developed using git, hosted on GitHub
  - <http://github.com/GMOD/jbrowse>
- Pls most involved: Ian Holmes, Lincoln Stein, Suzi Lewis

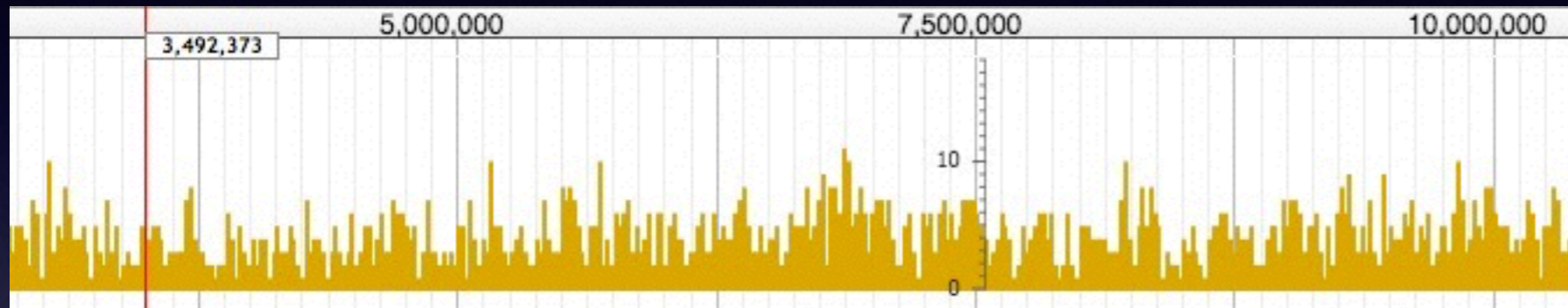
# Feature Tracks

- HTMLFeatures: Rectangles (<div>s) with various fills and heights to represent the feature spans (more compatible)
- CanvasFeatures: Much prettier, more configurable glyphs
- Super-configurable left clicking and right-click

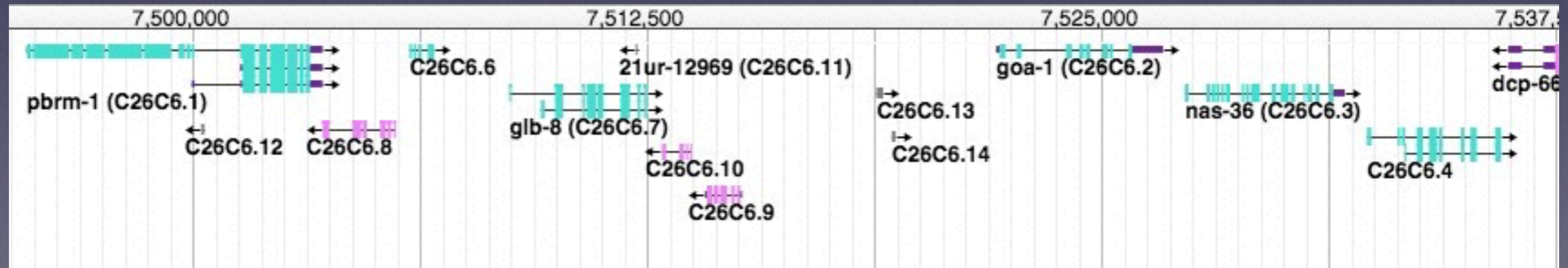


# Feature Density Plots

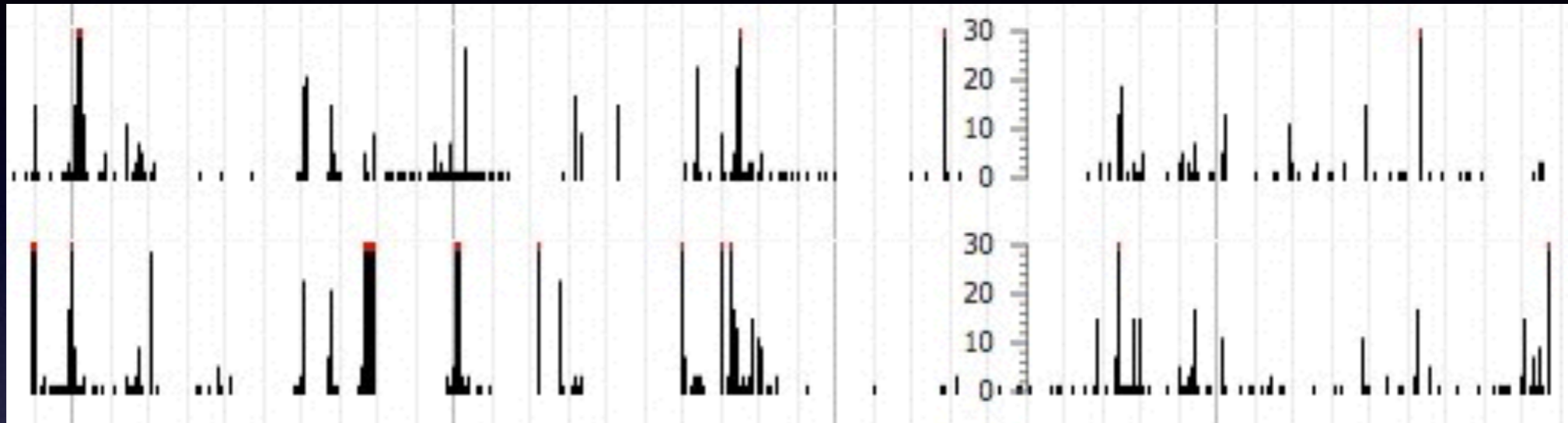
Zoomed out



Zoomed in

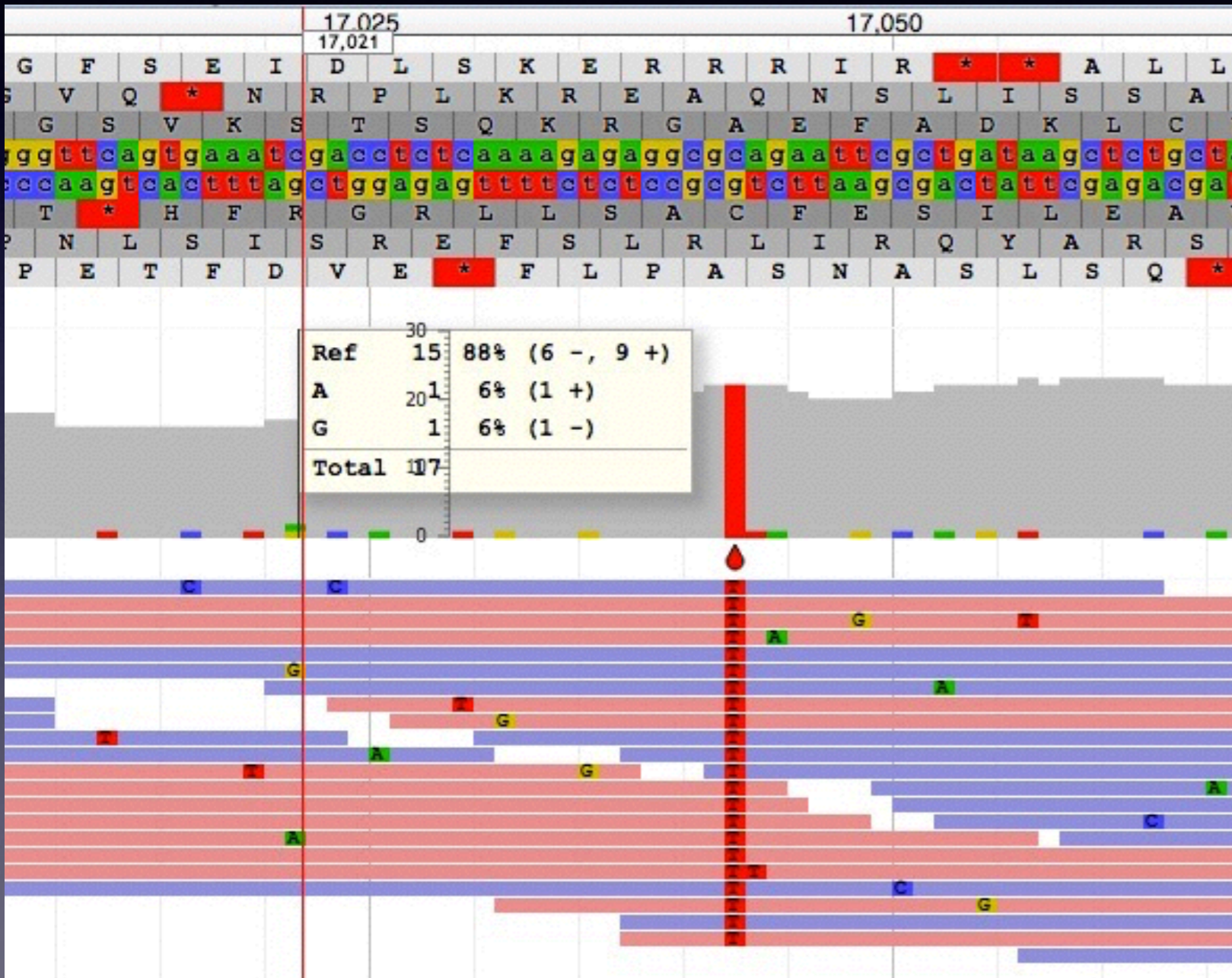


# Wiggle/BigWig Tracks



- Reads small chunks directly from BigWig file.
- Needs only a not-super-old (< 5 yrs) browser, except for Internet Explorer.
- IE requires version 10.

# BAM Alignment Tracks



- Reads small chunks directly from BAM file.
- Coverage and mismatches.

# VCF Tracks

**SNV**

**Primary Data**

Type	SNV
Score	104
Description	SNV A -> T
Position	scf1117875582023:88017..88017
Length	1 bp

**Attributes**

AC1	2
AF1	1
DP	17
DP4	0 0 11 6
FQ	-78
MQ	57
VDB	0.0507
alternative_alleles	T
description	SNV A -> T
reference_allele	A
seq_id	scf1117875582023

**Genotypes (1)**

variant	1	100%
homozygous	1	100%
T variant	1	100%
<b>Total</b>	<b>1</b>	<b>100%</b>

Name	GT	PL	GQ
simulated-sorted.bam	T / T	137 51 0	99

- Reads directly from VCF files compressed and indexed with bgzip and tabix.
- Shows all VCF data: alleles, genotypes, quality, etc.



# Particular Strengths

- Web-based, but fast and smooth easy to set up
- Compressed NGS data: direct-to-browser BAM, BigWig, and VCF
- Optional faceted track selector – efficiently search thousands of tracks
- Open local files directly on client, no data transfer required
- Highly customizable, embeddable, integratable, programmable

# Apollo

<http://gmod.org/wiki/WebApollo>

- Based on JBrowse, using plugin system
- Next generation of the popular Apollo annotation editor
- Online annotation editing and curation!

**Web  pollo**

# Apollo

- Clients receive updates in real time (like Google Docs)
- Saves edits to a central Chado database
- Client side is a JBrowse plugin
- Extensive server-side Java
- Maybe a live demo (later)
- <http://genomearchitect.org/WebApolloDemo/>

# JBrowse Plugins

- Extend JBrowse with your own JavaScript code
- Can do pretty much anything
  - Add your own track visualizations
  - Add your own data backends
  - Add menu items
  - Subscribe to event notifications (pub/sub system)
  - Reach deep into the guts of JBrowse and (carefully!)
    - change anything at all!
- Apollo client is a JBrowse plugin

# Coming in 2.X

- MORE: data types, sorting options, speed
- Graphical configuration
- Multiple views, linked or independent
- Logins, uploading, track sharing
- Circular genomes

# Big Thanks

Ian Holms (UC Berkeley)

Rob Buels

Mitch Skinner

Amelia Ireland

Eric Yao

Lincoln Stein (OICR)

Julien Smith-Roberge

Erik Derohanian

Julie Moon

Natalie Fox

Adam Wright

Suzi Lewis (LBNL)

Gregg Helt

Ed Lee

Justin Reese (UofMo)

Colin Diesh (UofMo)

**NHGRI**

Cold, hard cash

# The End (on to the workshop)

<http://jbrowse.org/>

GMOD: <http://gmod.org/wiki/JBrowse>

Github: <http://github.com/GMOD/jbrowse>