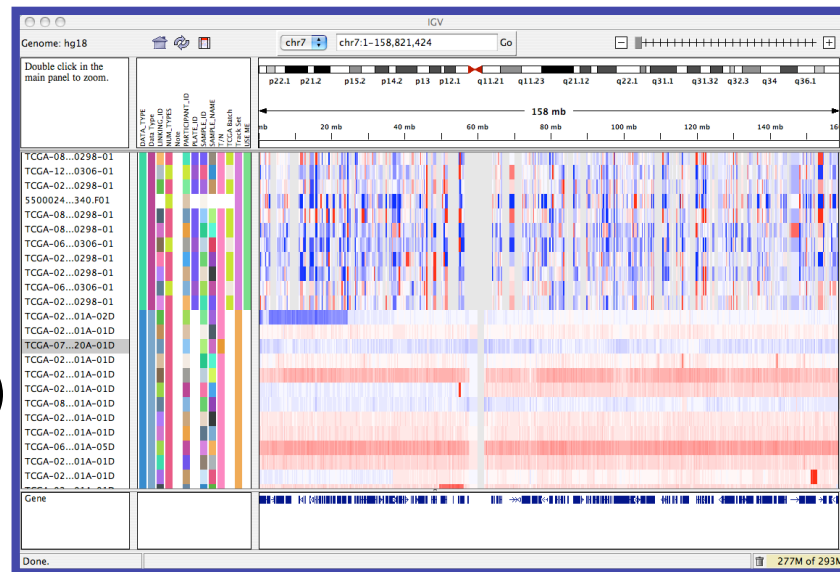


Visualizing Cancer Genome Data

Broad
Integrative
Genomics
Viewer (IGV)



<http://www.broad.mit.edu/igv>

Jill P. Mesirov, Ph.D.

TCGA Steering Committee December 4, 2008



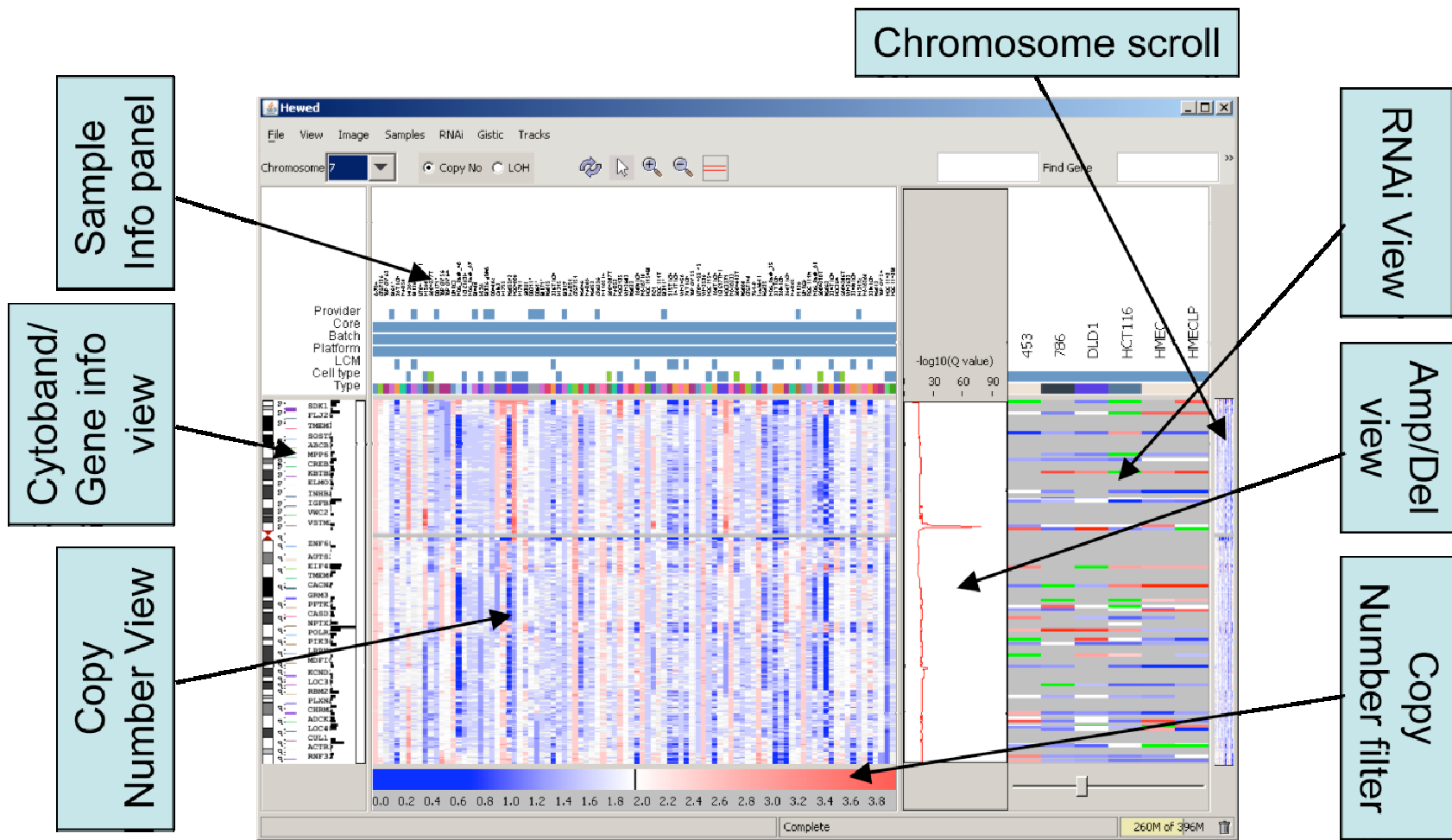
The Eli and Edythe L. Broad Institute

A Collaboration of Massachusetts Institute of Technology, Harvard University and affiliated Hospitals, and Whitehead Institute for Biomedical Research

Requirements for a TCGA Integrative Genomics Viewer

- Simultaneously view and interact with diverse TCGA data types
- Link mutations, amplifications, deletions with
 - Clinical phenotype
 - Transcription profiles, RNAi and epigenetic data
- Scientific goal
 - Provide insights
 - Generate hypotheses

GenePattern's SNPviewer



What the IGV provides

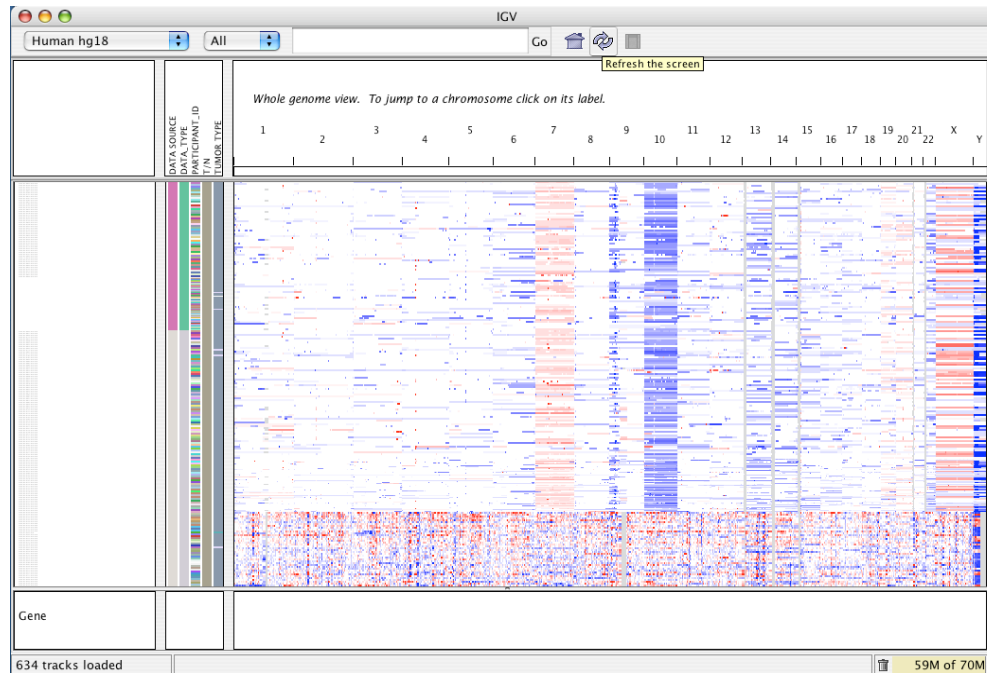
- Integrated views of multiple data types
 - Tied to genome
 - Very large data sets (TCGA)
 - Local or remote
 - Easy integration of new data types, views, layouts
- Interactive views
 - “Real-time”, smooth pan and zoom
 - Whole genome to base pair
 - Visualization options: heatmaps, barcharts, scatter plots, histograms, graphs
 - Sort, filter, group data tracks
 - Different data types - side by side or overlay
- User friendly
 - Intuitive and easy to use
 - Save sessions and share with colleagues

IGV Screen Layout

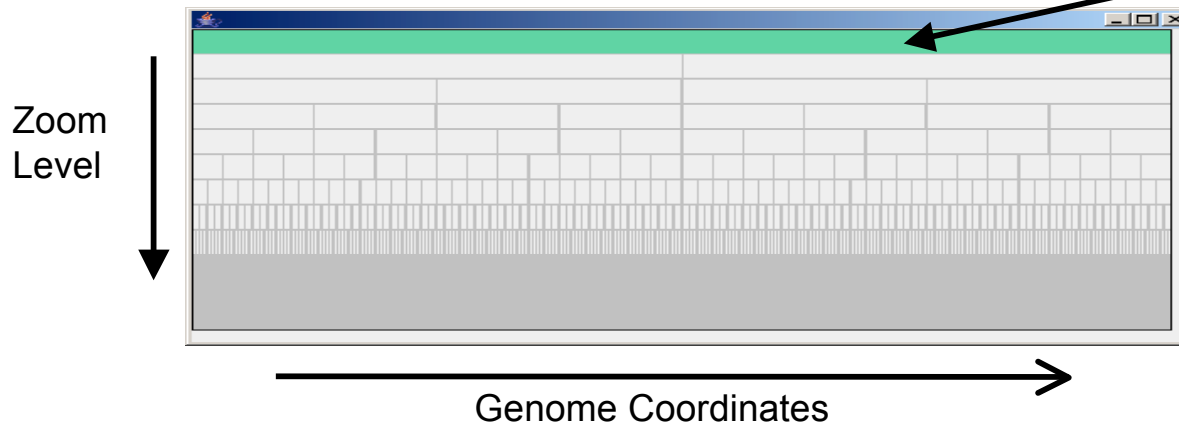
The screenshot displays the IGV interface with the following components labeled:

- Sample Annotation:** Located at the top left, it points to the metadata table containing columns like DATA_TYPE, LINKING_ID, NUM_TYPES, PARTICIPANT_ID, PLATE_ID, SAMPLE_ID, SAMPLE_NAME, T/N, TCGA Batch, Track Set, and USE ME.
- Cytoband:** Located at the top center, it points to the cytoband track showing chromosome bands (p22.1, p21.2, p15.2, p14.2, p13, p12.1, q11.21, q11.23, q21.12, q22.1, q31.1, q31.32, q32.3, q34, q36.1).
- Search Box:** Located at the top center, it points to the search input field containing "chr7:1-158,821,424".
- Zoom Control:** Located at the top right, it points to the zoom slider and navigation buttons.
- Sample Names:** Located on the left side, it points to the list of sample IDs (e.g., TCGA-08...0298-01, TCGA-12...0306-01).
- Genomic Coordinates:** Located on the right side, it points to the genomic coordinate scale at the top of the main panel.
- Data Panel:** Located on the right side, it points to the main heatmap area displaying data points for each sample across the genomic region.
- Annotation Heatmap:** Located on the left side, it points to the vertical heatmap tracks representing various annotations for each sample.
- Genome Features:** Located on the right side, it points to the track at the bottom of the main panel showing gene models and other genomic features.

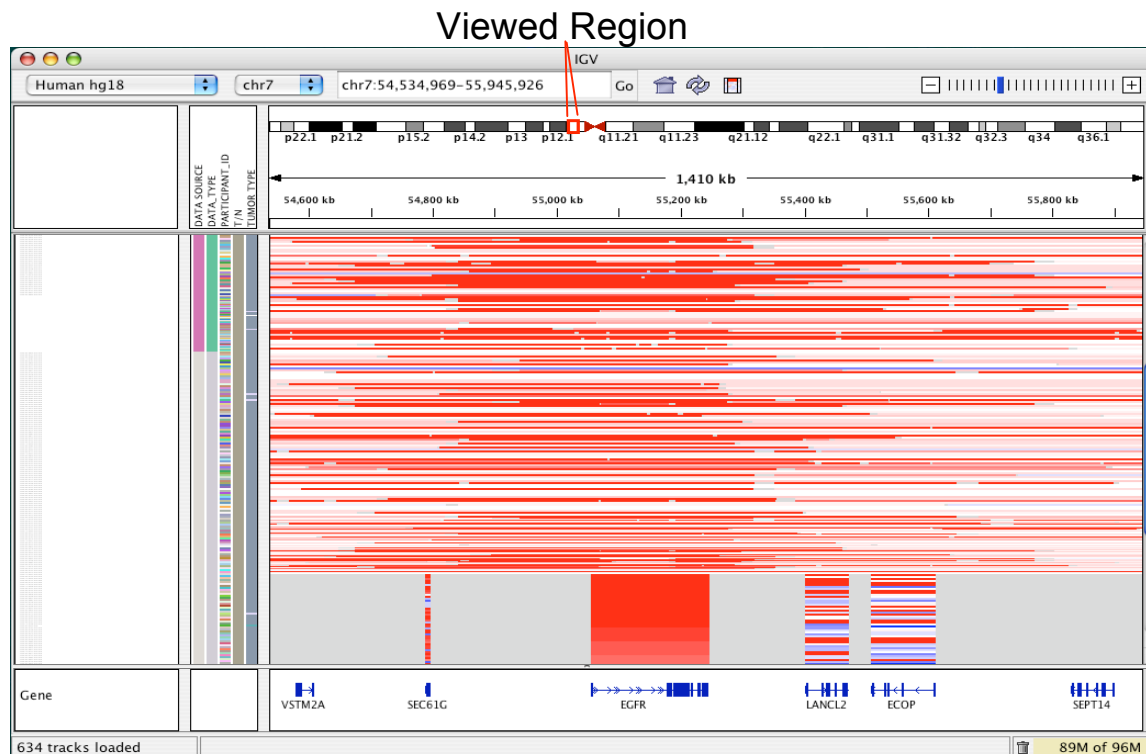
“Real-time” pan and zoom



Whole genome view.
The entire genome is represented by a single low resolution data tile.

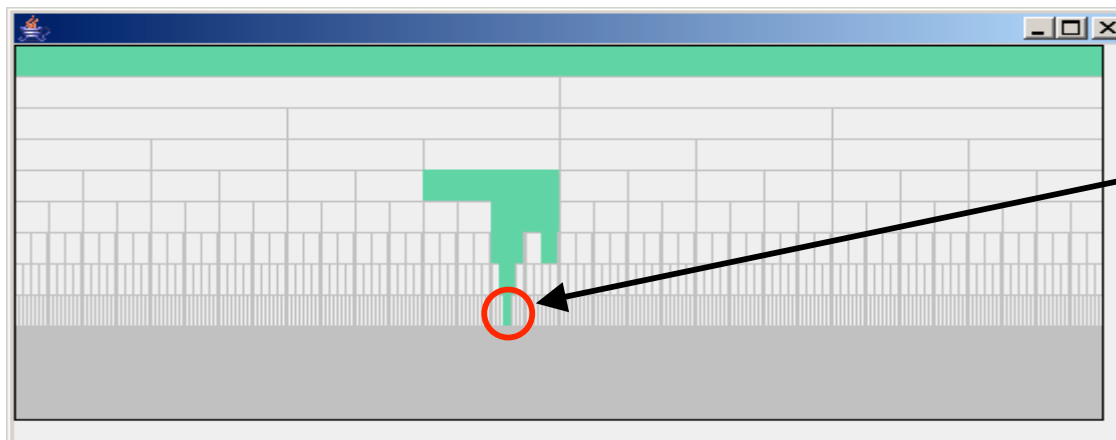


“Real-time” pan and zoom



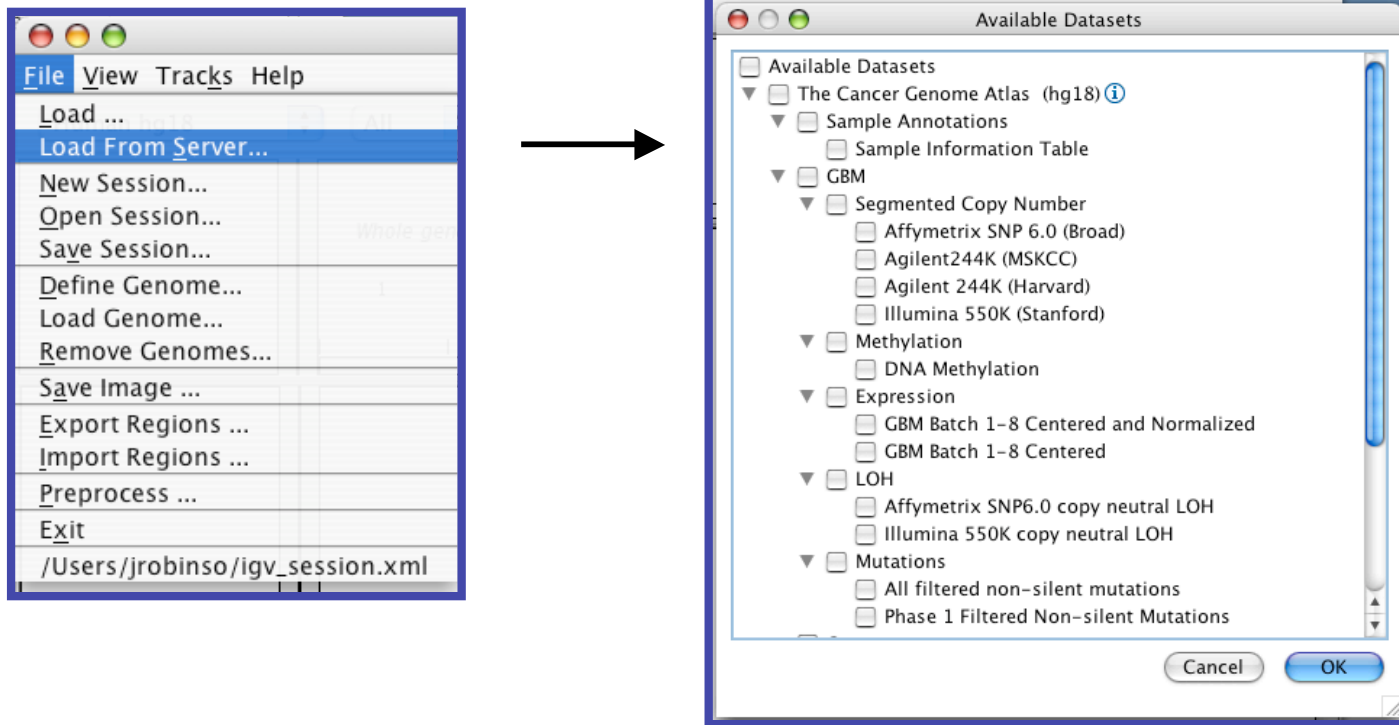
Zoomed-in view.

About 1.5 Mb of
Chr 7 around EGFR



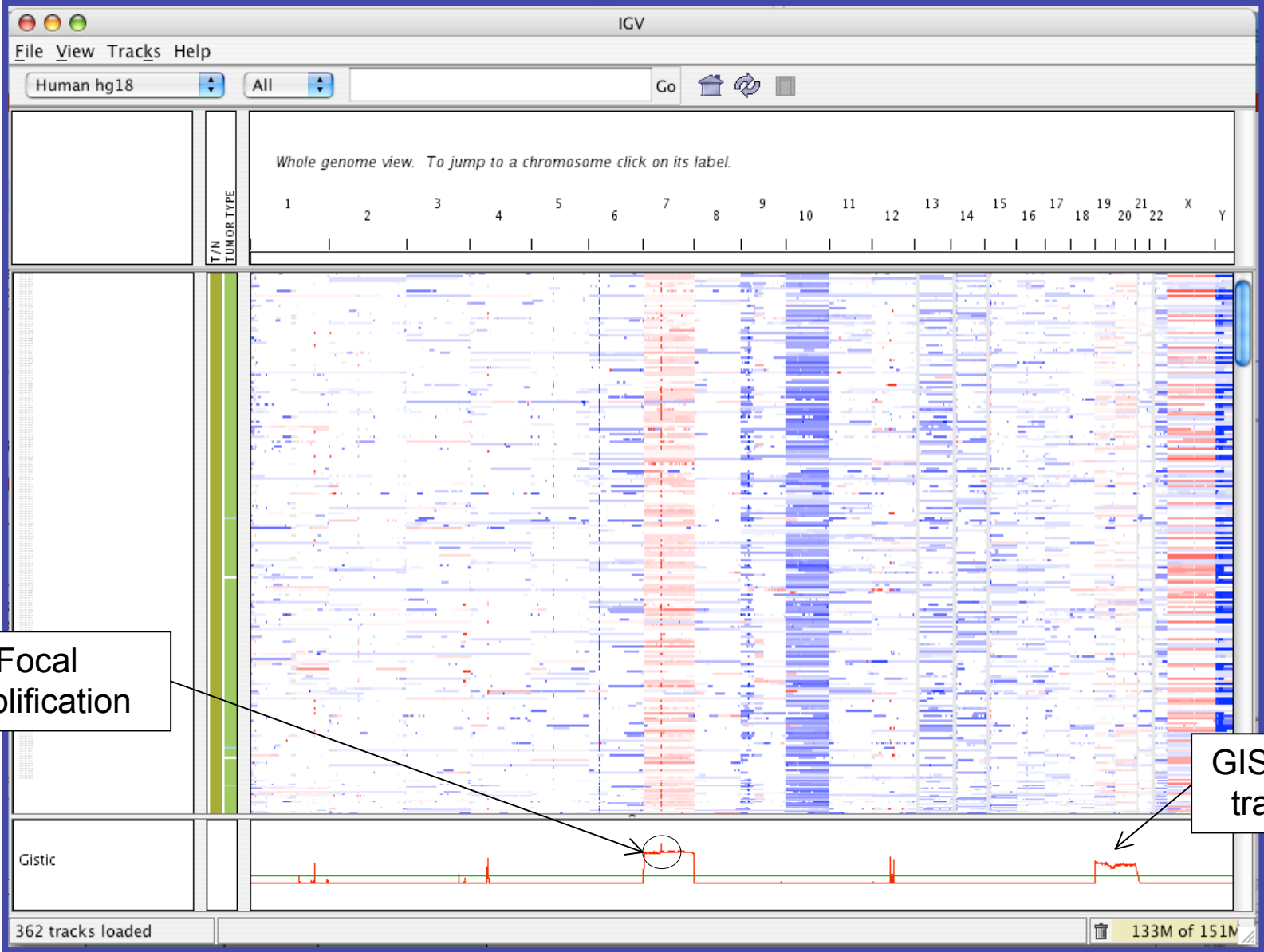
Viewing TCGA data in IGV

1. Start IGV from <http://www.broad.mit.edu/igv>
2. Choose “Load from Server” and select data to view



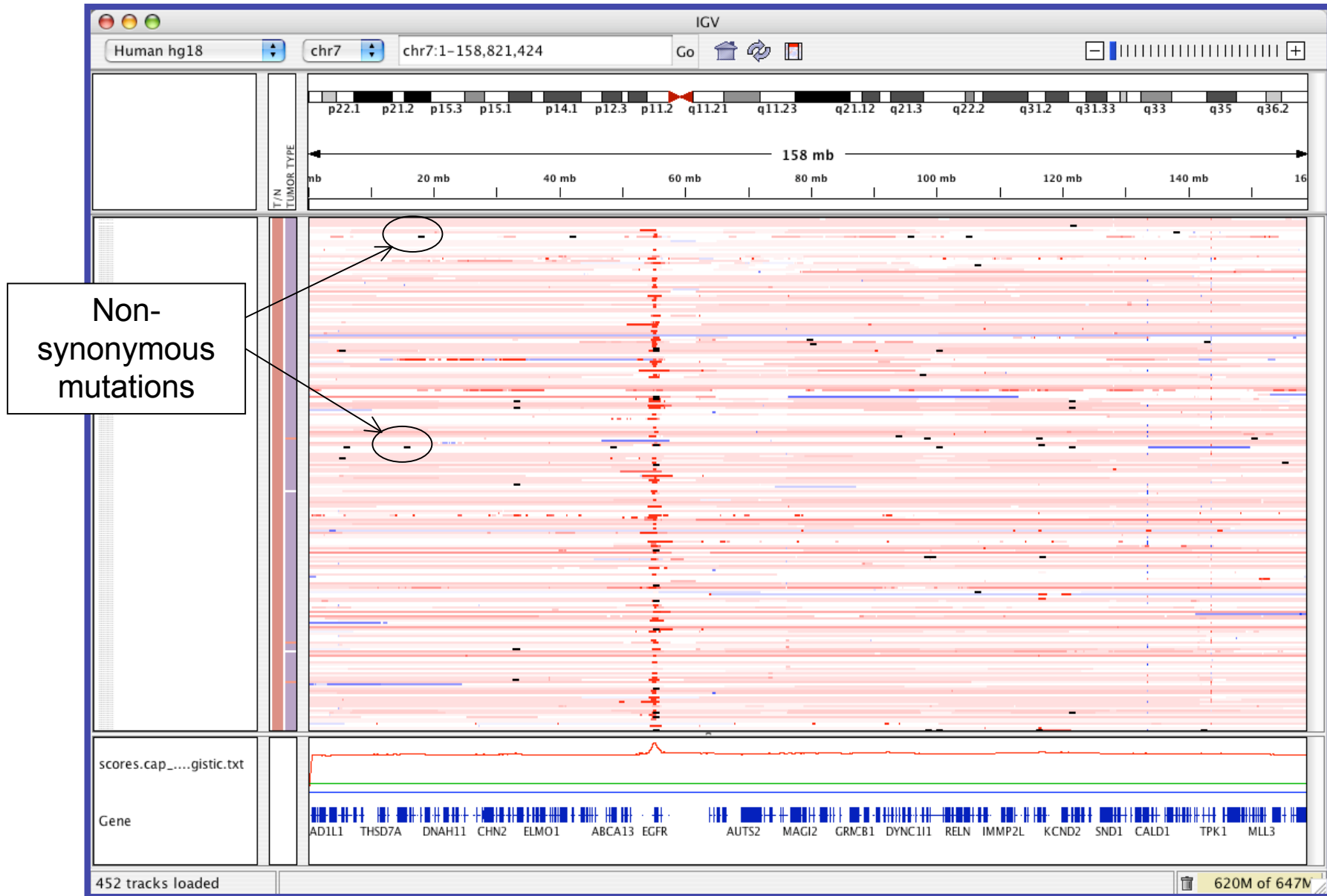
GBM copy number - 100's of TCGA samples

Whole Genome View

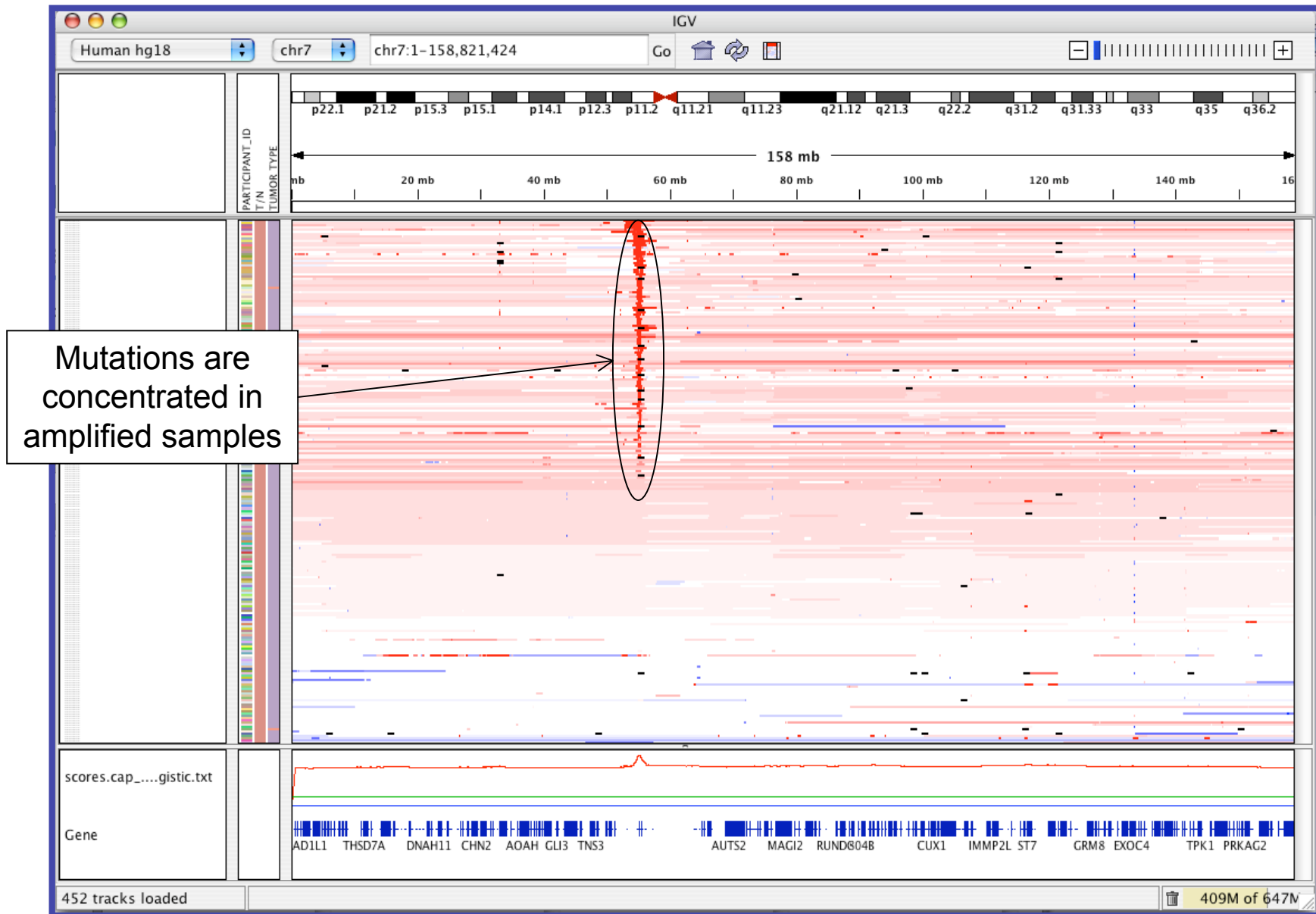


Copy number and mutations

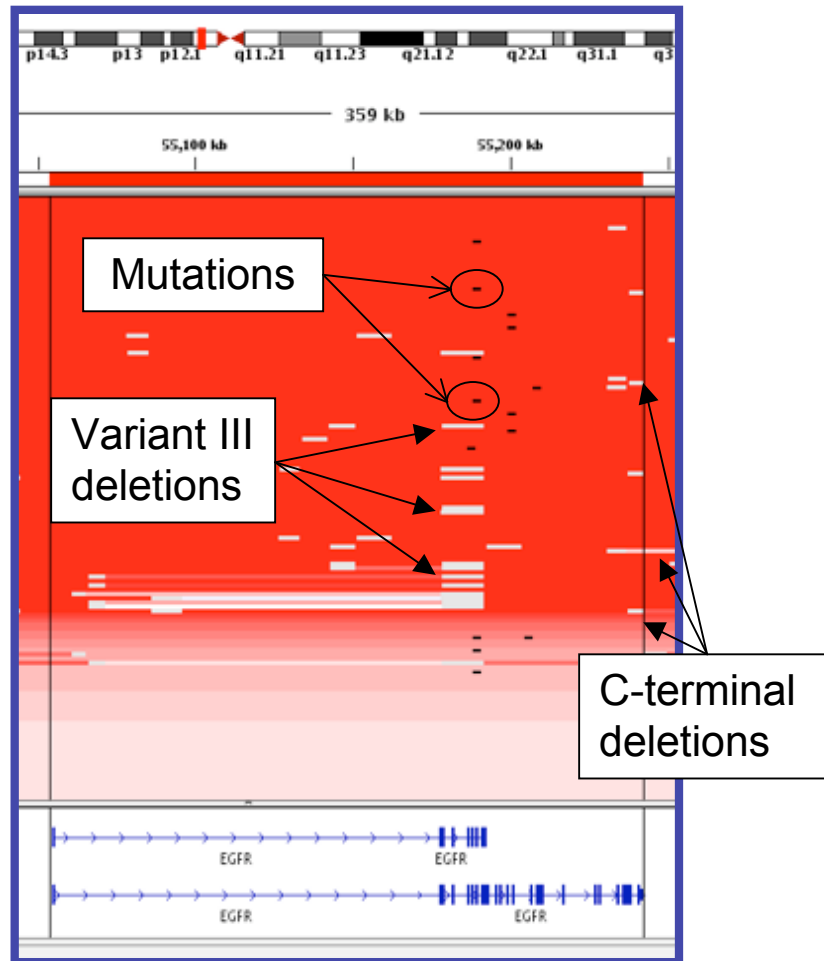
Zoomed to chromosome 7



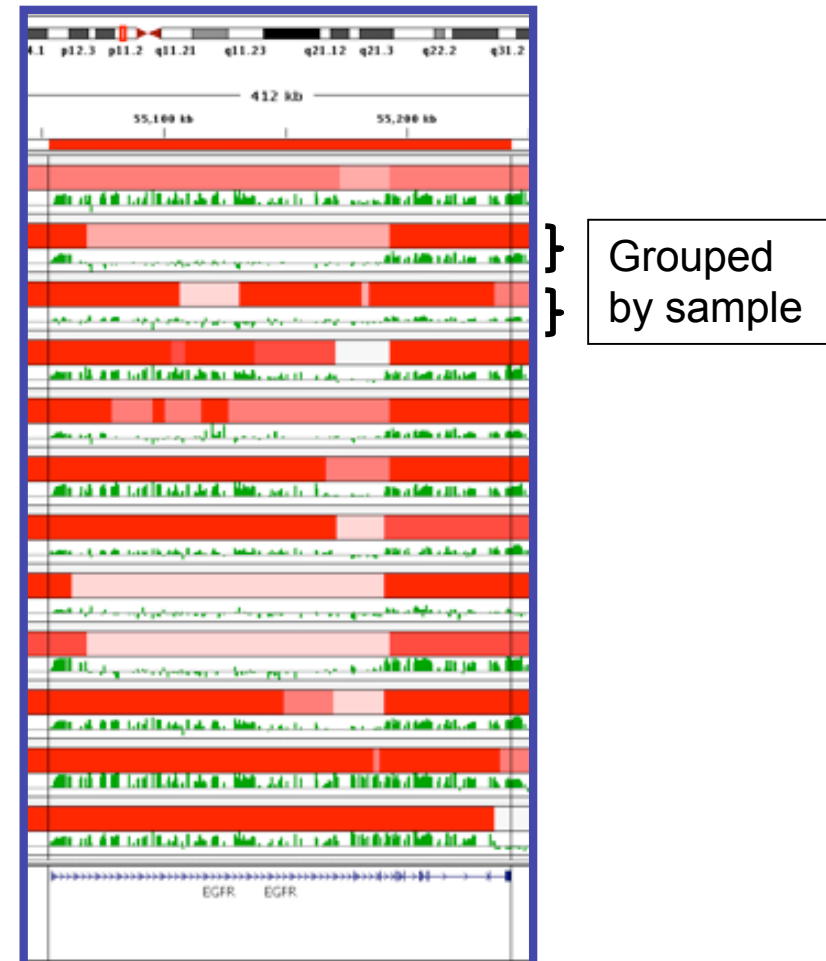
Sorted by amplification in EGFR region



EGFR locus - alternative views

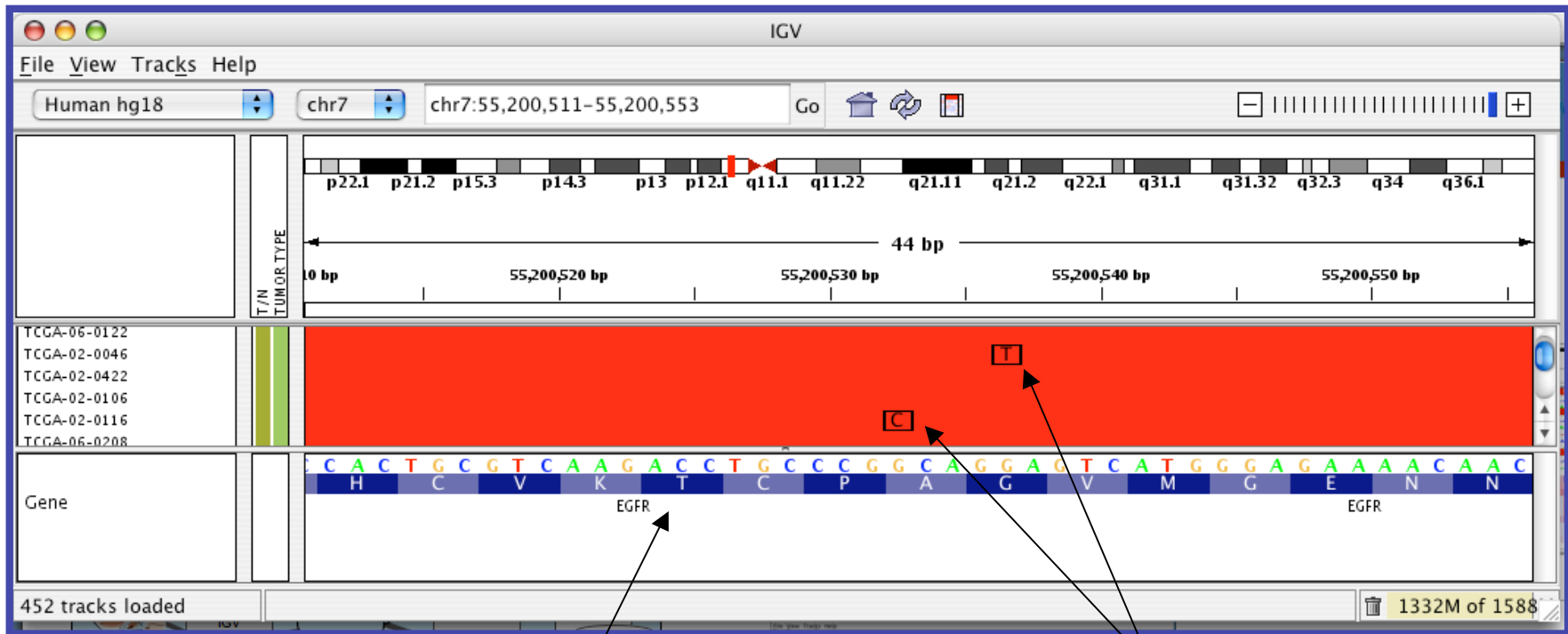


copy number and mutation



copy number and expression

Base-pair level view



Sequence and amino acid tracks

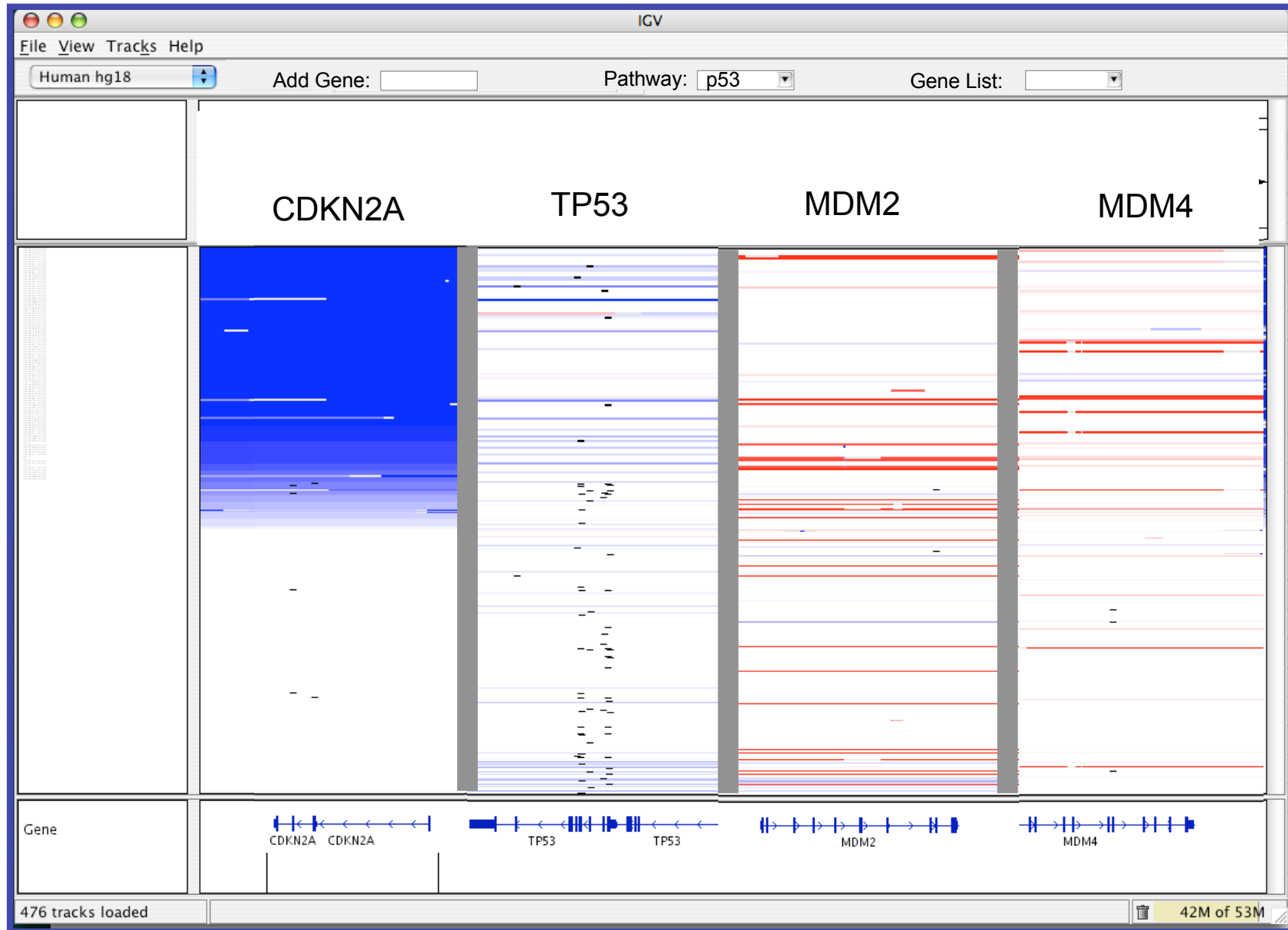
Mutations showing mutated base

Possible IGV enhancements

- Add gene and functionally oriented views
- More sequence support
 - Alignments
 - base counts for allele calling
- Advanced search and filtering
- Plugin APIs - analysis, rendering, etc.
- GenePattern Integration
- Browser-based version

Gene list view mockup

Samples sorted by deletion in CDKN2A



IGV software

<http://www.broad.mit.edu/igv>

- IGV 1.0 - July 2008
 - First public release
 - Standalone application
- IGV 1.1 - September 2008
 - Client – server architecture
 - Genomes and datasets hosted on public server at Broad
 - Direct access to public TCGA datasets preprocessed for IGV
- IGV 1.2 – public beta November 2008
 - Additional genomes hosted & support for user-defined genomes
- User's Guide and Quickstart
- Open source
- 1200+ registered users
- Contact: igv-help@broad.mit.edu

Acknowledgements

<http://www.broad.mit.edu/igv>

Architect and Developer

Jim Robinson

Other contributors - GenePattern team

Cancer and Genome Biology Programs

Jesse Boehm

Elijah Flakes

Gad Getz

Mitchell Guttman

Eric Lander

Matthew Meyerson

Marc-Danie Nazaire

Michael Reich

Helga Thorvaldsdóttir

Barbara Weir

Wendy Winckler



