

Exploring human genetic variation

Aaron Quinlan and Ira Hall

BIMS 6000

04-Sept-2013

Goals

- Explore genetic variation using real DNA sequence data from 3 individuals from the 1000 Genomes project.
- Understand the types and functional impact of different types of genetic variation
- Understand inheritance patterns.
- Get your hands dirty with real data.
- **Remember:** this dataset is always available, so you can repeat this exercise anytime you want.

Studying individuals from CEU population

CEU: Utah residents with Northern and Western European ancestry from the CEPH collection.



FONDATION JEAN DAUSSET

Home ▾ HGDP - CEPH Project ▾ Other Projects ▾ CEPH Biobank

HGDP-CEPH Human Genome Diversity Cell Line Panel

A resource of 1063 cultured lymphoblastoid cell lines (LCLs) from 1050 individuals in 52 world populations and corresponding milligram quantities of DNA is banked at the Foundation Jean Dausset-CEPH in Paris. These LCLs were collected from various laboratories by the Human Genome Diversity Project (HGDP) and CEPH in order to provide unlimited supplies of DNA and RNA for studies of sequence diversity and history of modern human populations. Information for each LCL is limited to sex of the individual and population and geographic origin.

The table provides details of all the LCLs in the resource, uncorrected for duplicates (13 duplicate pairs), 2 genetically atypical samples, **13 pairs of duplicate LCLs and those from two genetically atypical individuals**, and 96 pairs of close relatives (first and/or second degree relative pairs) **LCLs for ("close") relative pairs**. Sixteen LCLs differ in gender indicated on records and that determined by molecular typing. All samples used for this resource were collected with proper informed consent.

The DNAs have been distributed to 107 investigators (as of 1 December 2010) for genotyping and/or resequencing; the results are contributed to a **central database**. To date, the DNAs have been typed genome wide with almost 1 million SNPs, 843 microsatellites, and 51 small indel loci. Some 10,000 CNV calls from two different laboratories are included in the database. Nuclear and mitochondrial DNA regions have been resequenced. High throughput sequencing of entire HGDP-CEPH genomes and array captured targets is underway.

For more information contact **Dr. Howard Cann** , CEPH.

CEU part of 1000 Genomes Project

The screenshot shows a Mac OS X desktop with a window titled "www.1000genomes.org". The main content area displays the "1000 Genomes" logo and a background image of human chromosomes. Below the logo, a navigation bar includes links for Home, About, Data, Analysis, Participants, Contact, Browser, Wiki, and FTP search. A search bar is also present. To the right, there's a sidebar titled "NAVIGATION" with a link to "Frequently Asked Questions". The main content section is titled "LATEST ANNOUNCEMENTS" and features a post from "MONDAY JULY 02, 2012" about "Phase 1 analysis results including chrY and chrMT variant calls". It mentions the availability of analysis results based on the phase1 integrated variant call set, including chrY and chrMT variant calls, functional annotation, and local area ancestry inference for admixed populations. It also links to full details on a webpage, EBI and NCBI data access links, and a README file. Below this, there's a section titled "Recent project announcements" with a post from "THURSDAY SEPTEMBER 06, 2012" about genome accessibility information now available on the 1000 Genomes Browser. It states that two accessibility tracks have been added to the browser. The sidebar also includes links for "All Project Announcements", "Sample and Project Information", "Media Archive", and "Download the 1000 Genomes Pilot Paper".

1000 Genomes
A Deep Catalog of Human Genetic Variation

Home About Data Analysis Participants Contact Browser Wiki FTP search Search

LATEST ANNOUNCEMENTS

MONDAY JULY 02, 2012

Phase 1 analysis results including chrY and chrMT variant calls.

Analysis results based on our phase1 integrated variant call set are now available.

This includes [chrY and chrMT variant calls](#), [functional annotation](#) of our variant calls and local area ancestry inference for our admixed populations.

Full details of the directory contents can be found [on this webpage](#).

Data Access links: [EBI](#) / [NCBI](#)

README: [EBI](#) / [NCBI](#)

Recent project announcements

THURSDAY SEPTEMBER 06, 2012

Genome Accessibility information now available on the 1000 Genomes Browser

Two Accessibility Tracks have now been added to the [1000 Genomes Browser](#)

NAVIGATION

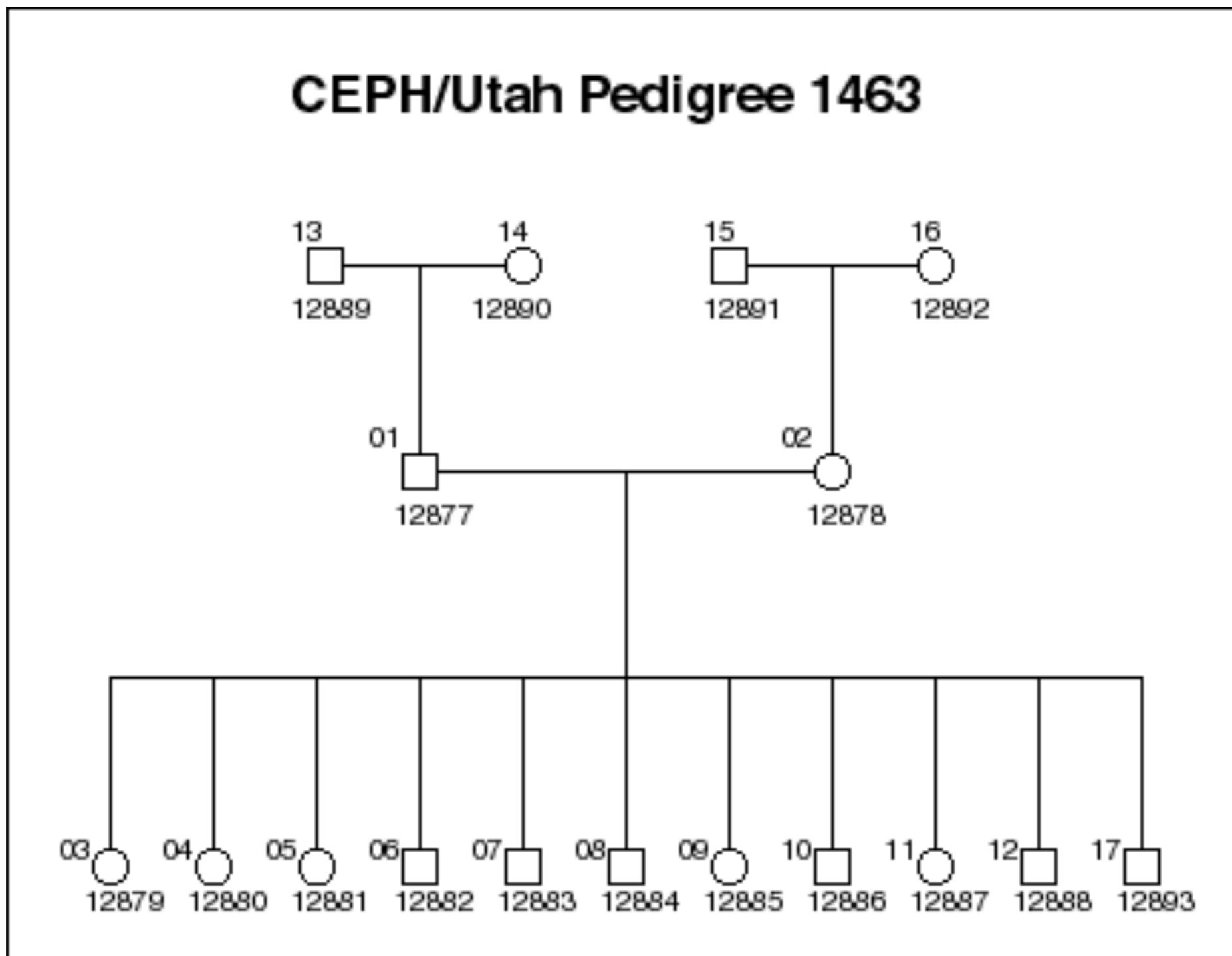
- [Frequently Asked Questions](#)

LINKS

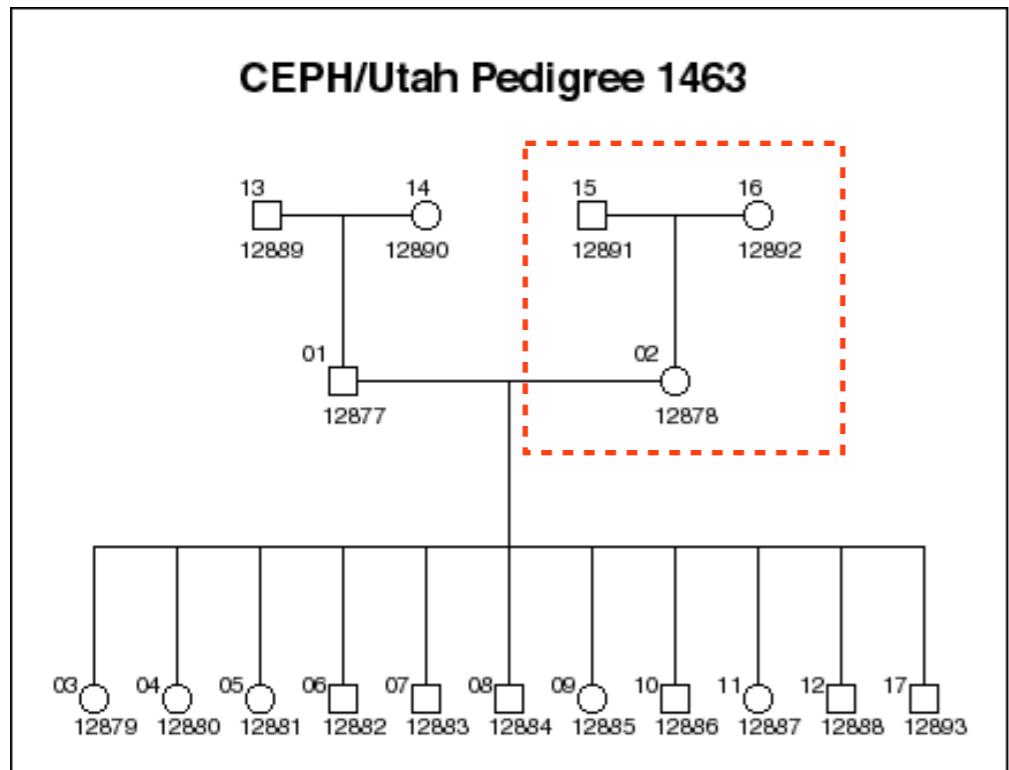
- [All Project Announcements](#)
- [Sample and Project Information](#)
- [Media Archive](#)
- [Download the 1000 Genomes Pilot Paper](#)

This information was built using sequence data from the [phase1 dataset](#).

CEPH Pedigree 1463



CEPH Pedigree 1463



maternal
grandfather

15
12891

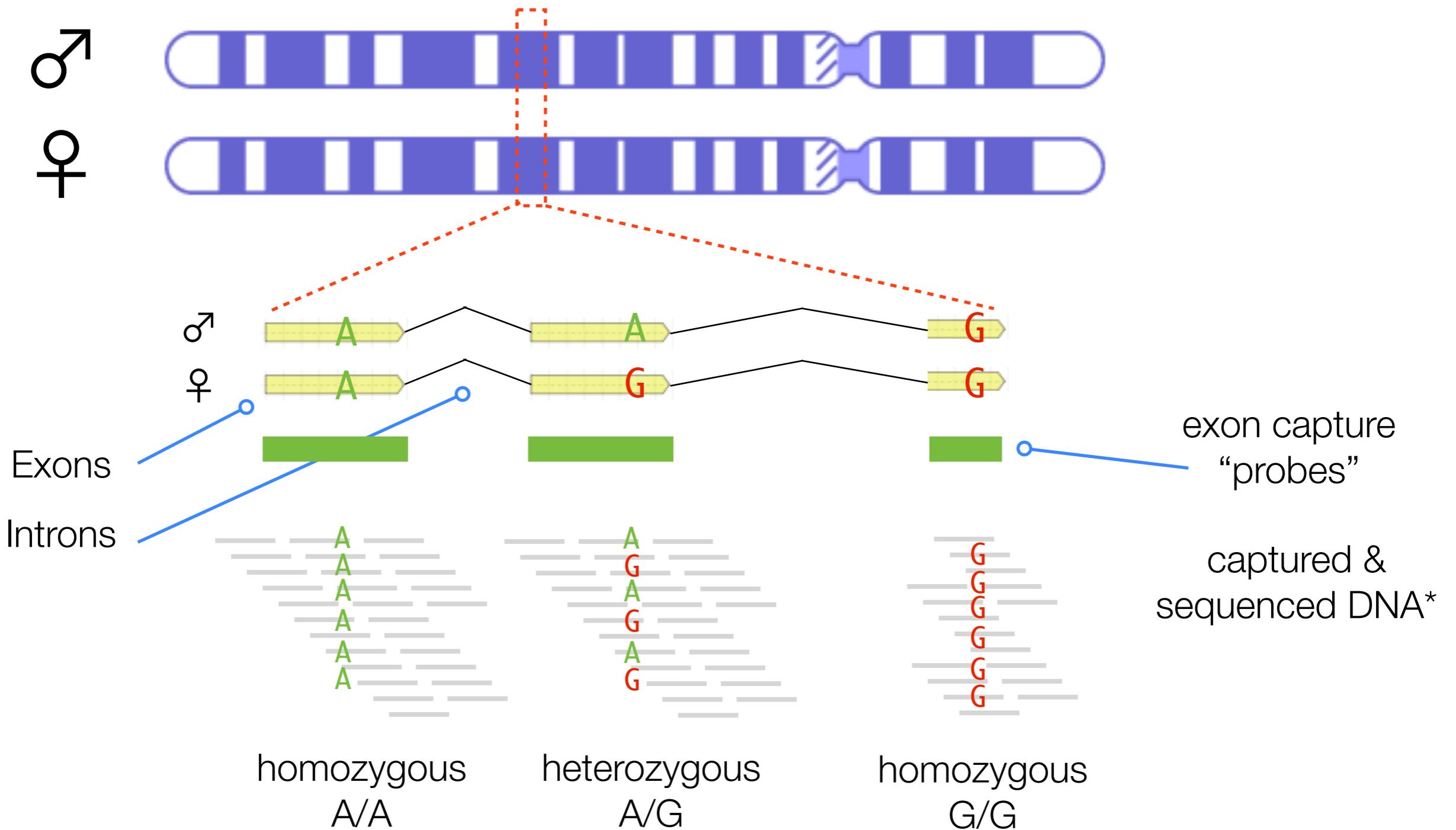
maternal
grandmother

16
12892

02
12878

mother

Exome capture from 1000G



* remember that DNA from each parental chromosome is captured and each sequence represents DNA from one of the two parental chromosomes

Google Maps for the genome



The screenshot shows the IGV software interface. On the left, there is a navigation sidebar with the IGV logo, a search bar, and links to Home, Downloads, Documents, Hosted Genomes, FAQ, IGV User Guide, File Formats, Release Notes, Credits, and Contact. Below this is a "Search website" input field and a "search" button. At the bottom of the sidebar are links to Broad Home and Cancer Program, and the BROAD INSTITUTE logo with the text "© 2012 Broad Institute".

Home

Integrative Genomics Viewer

The main area displays the IGV logo and a screenshot of the software's graphical user interface. The interface features a central panel with a genomic track showing multiple data layers, including tracks for RNA-seq, ChIP-seq, and other genomic features. A legend on the right side of the track panel identifies the different data series.

What's New

NEWS

July 3, 2012. Soybean (*Glycine max*) and Rat (*rn5*) genomes have been updated.

April 20, 2012. IGV 2.1 has been released.
See the [release notes](#) for more details.

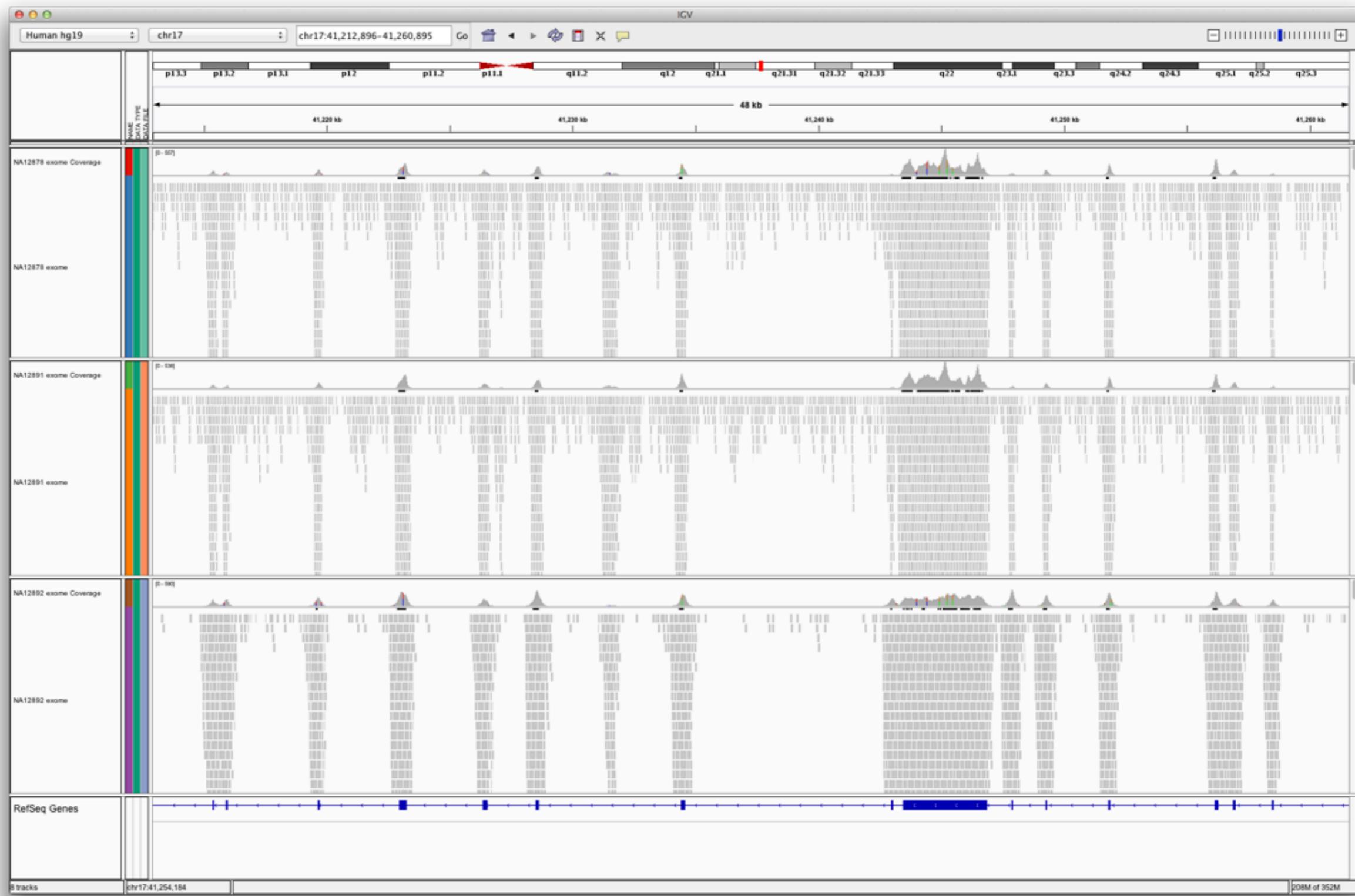
April 19, 2012. See our new [IGV paper](#) in *Briefings in Bioinformatics*.

Citing IGV

To cite your use of IGV in your publication:

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. [Integrative Genomics Viewer](#). *Nature Biotechnology* 29, 24–26 (2011), or
Helga Thorvaldsdottir, James T. Robinson, Jill P. Mesirov.

Google Maps for the genome



g igv browser - Google Search

https://www.google.com/#hl=en&sugexp=les%3B&gs_nf=1&gs_mss=IGV%20brw&cp=8&gs_id=10&xhr=t&q=IGV%20brows...

Calendar Genetics Clan MacMillan - Wiki Share Your Terminal Rare Genomics Institut python - How to mo... Project Dashboard

+You Search Images Maps Play YouTube News Gmail Documents Calendar More

Search

Sign in

IGV browser

igv browser

igv browser download

igv browser broad

broad institute igv browser

Web

[Home | Integrative Genomics Viewer](#)
www.broadinstitute.org/igv/
What's New. July 3, 2012. Soybean (Glycine max) and Rat (rn5) genomes have been updated. April 20, 2012. IGV 2.1 has been released. See the release notes ...

Images

Maps

Videos

News

Shopping

More

[Downloads](#)
Downloads, Documents, Hosted Genomes · FAQ · IGV User ...

[FAQ](#)
Alignments, Q: I loaded a BAM file and don't see anything. What's ...

[Start IGV](#)
Start IGV. To start IGV: Go to the IGV downloads page: http ...

[Viewing Alignments](#)
The preferred file format for viewing alignments in IGV is the ...

[IGV User Guide](#)
IGV User Guide. This guide describes the Integrative ...

[Preferences](#)
To display the Preferences window, click View>Preferences ...

[More results from broadinstitute.org »](#)

Charlottesville, VA

Change location

Show search tools

[Software | Broad Institute of MIT and Harvard](#)
www.broadinstitute.org/scientific-community/software
Use our search function, **browse** the complete software collection or click on one ...
The Argo Genome **Browser** is a production tool for visualizing and manually ...

OK 0:00:11 17.050



Home | Integrative Genomics

www.broadinstitute.org/igv/

Calendar Genetics Clan MacMillan - Wi Share Your Terminal Rare Genomics Insti python - How to mo Project Dashboard

Home

Integrative Genomics Viewer



What's New

NEWS July 3, 2012. Soybean (*Glycine max*) and Rat (rn5) genomes have been updated.

April 20, 2012. IGV 2.1 has been released. See the [release notes](#) for more details.

April 19, 2012. See our new [IGV paper](#) in *Briefings in Bioinformatics*.

Citing IGV

To cite your use of IGV in your publication:

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. [Integrative Genomics Viewer](#). *Nature Biotechnology* 29, 24–26 (2011), or

Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. [Integrative Genomics Viewer \(IGV\): high-performance genomic data visualization and exploration](#)

Log In | Integrative Genomics

www.broadinstitute.org/software/igv/log-in

Calendar Genetics Clan MacMillan - Will Share Your Terminal Rare Genomics Institut python - How to mo Project Dashboard

Home > Log In

Log In

To use IGV, registration is required.
[Click here](#) to register.

If you have already registered for IGV please enter your registration email address below.

email address:



Search website

[Broad Home](#)
[Cancer Program](#)

BROAD INSTITUTE
© 2012 Broad Institute

Downloads | Integrative Genomics Viewer

www.broadinstitute.org/software/igv/download

Calendar Genetics Clan MacMillan - Will Share Your Terminal Rare Genomics Institut python - How to move Project Dashboard

Home > Downloads

Downloads

Integrative Genomics Viewer (Version 2.1)

Notes

Java: IGV 2.1 requires Java 6 or greater.

Chrome: Chrome does not launch "java webstart" files by default. Instead, the launch buttons below will download a "jnlp" file. This should appear in the lower left corner of the browser. Double-click the downloaded file to run.

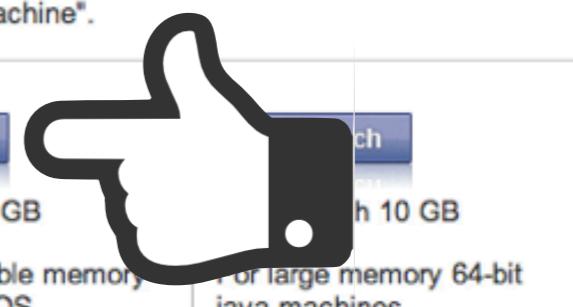
Windows users: To run with more than 1.2 GB you must install 64-bit Java. This is often not installed by default even with the latest Windows 7 machines with many GB of memory. In general trying to launch with more memory than your OS/Java combination supports will result in the obscure error "could not create virtual machine".

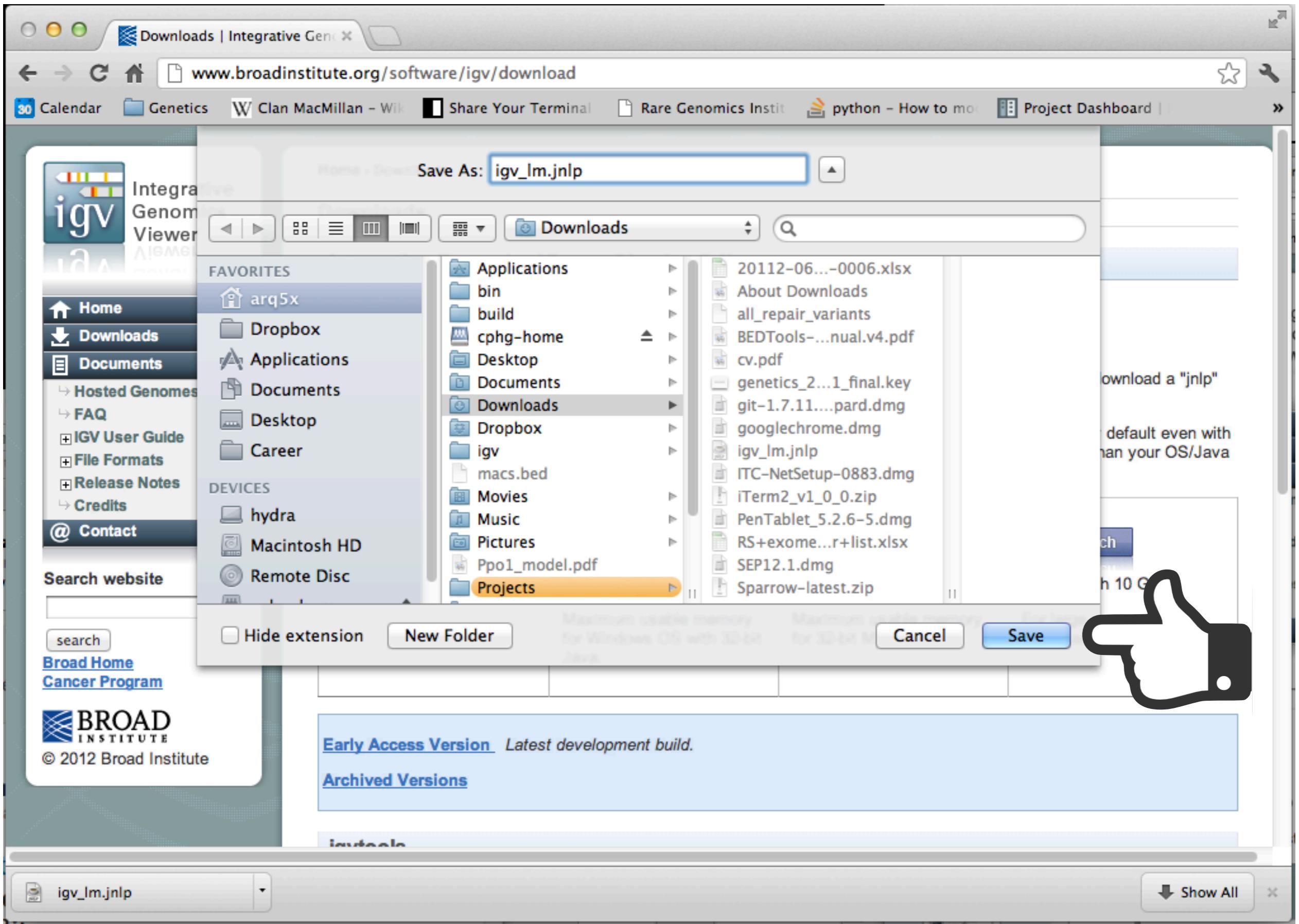
 Launch Launch with 750 MB	 Launch Launch with 1.2 GB Maximum usable memory for Windows OS with 32-bit Java.	 Launch Launch with 2 GB Maximum usable memory for large memory 64-bit java machines.
--	--	--

Early Access Version Latest development build.

Archived Versions

www.broadinstitute.org/igv/projects/current/igv_lm.jnlp





Downloads | Integrative Genomics Viewer

www.broadinstitute.org/software/igv/download

Calendar Genetics Clan MacMillan - Wiki Share Your Terminal Rare Genomics Institut python - How to move Project Dashboard

Home > Downloads

Downloads

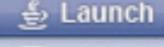
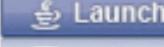
Integrative Genomics Viewer (Version 2.1)

Notes

Java: IGV 2.1 requires Java 6 or greater.

Chrome: Chrome does not launch "java webstart" files by default. Instead, the launch buttons below will download a "jnlp" file. This should appear in the lower left corner of the browser. Double-click the downloaded file to run.

Windows users: To run with more than 1.2 GB you must install 64-bit Java. This is often not installed by default even with the latest Windows 7 machines with many GB of memory. In general trying to launch with more memory than your OS/Java combination supports will result in the obscure error "could not create virtual machine".

 Launch Launch with 750 MB	 Launch Launch with 1.2 GB Maximum usable memory for Windows OS with 32-bit Java.	 Launch Launch with 2 GB Maximum usable memory for 32-bit MacOS.	 Launch Launch with 10 GB For large memory 64-bit java machines.
---	--	---	---

[Early Access Version](#) Latest development build.

[Archived Versions](#)

Downloads

igv_lm.jnlp

Show All



Downloads | Integrative Genomics Viewer

www.broadinstitute.org/software/igv/download

Calendar Genetics Clan MacMillan - Wiki Share Your Terminal Rare Genomics Institut python - How to move Project Dashboard

Home > Downloads

Downloads

Integrative Genomics Viewer (Version 2.1)

Notes

Java: IGV 2.1 requires Java 6 or greater.

Chrome: Chrome does not launch "java webstart" files by default. Instead, the launch buttons below will download a "jnlp" file. This should appear in the

Windows users: To run with the latest Windows 7 machines combination supports will result

The application "IGV 2.1" from "www.broadinstitute.org" is requesting access to your computer.

The digital signature from "Broad Institute" has been verified.

Allow all applications from "www.broadinstitute.org" with this signature

[Launch](#) Launch with 750 MB [Show Details...](#) Deny Allow

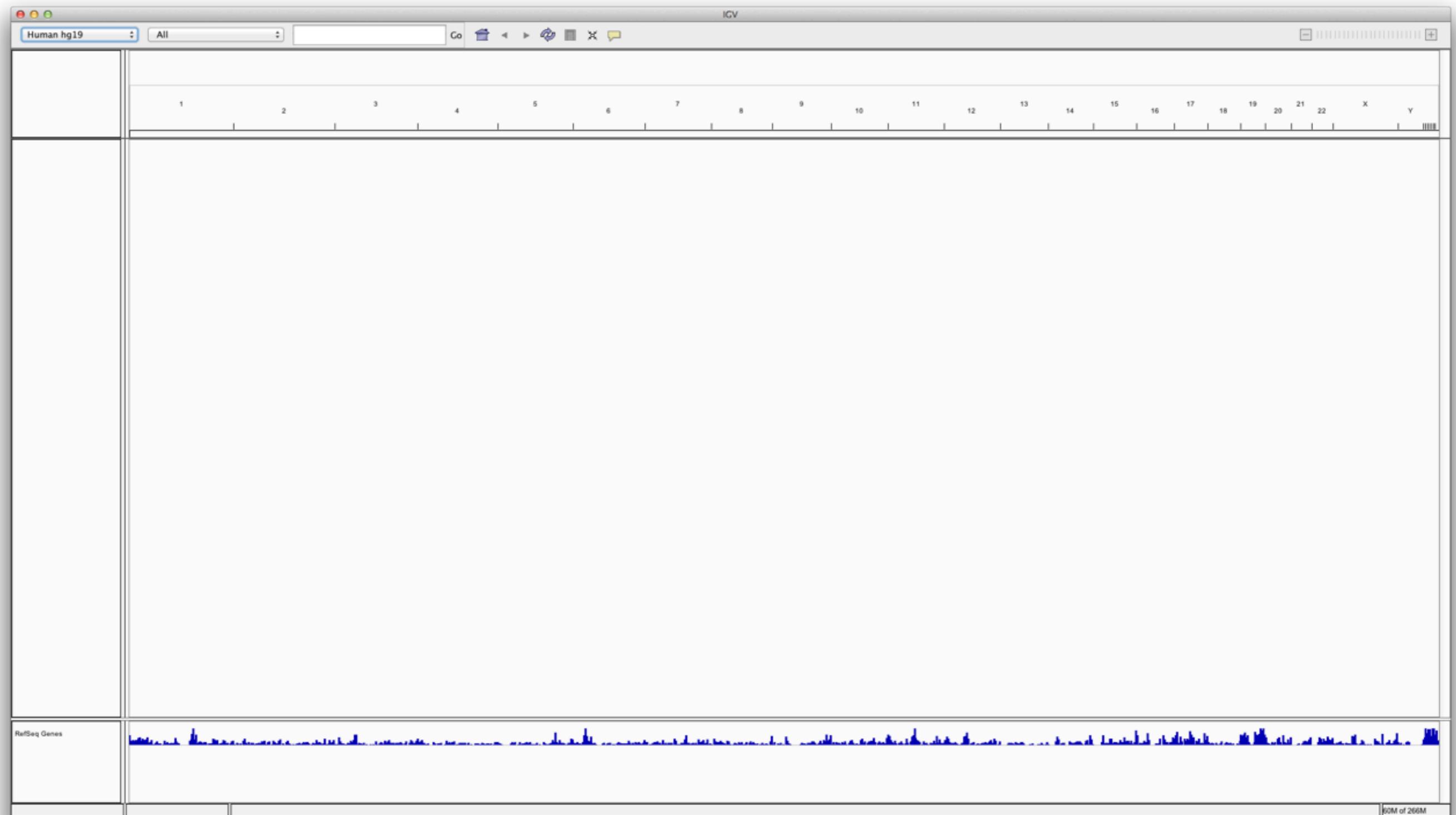
For Windows OS with 32-bit Java. For 64-bit Mac OS X Java machines.

[Early Access Version](#) Latest development build.

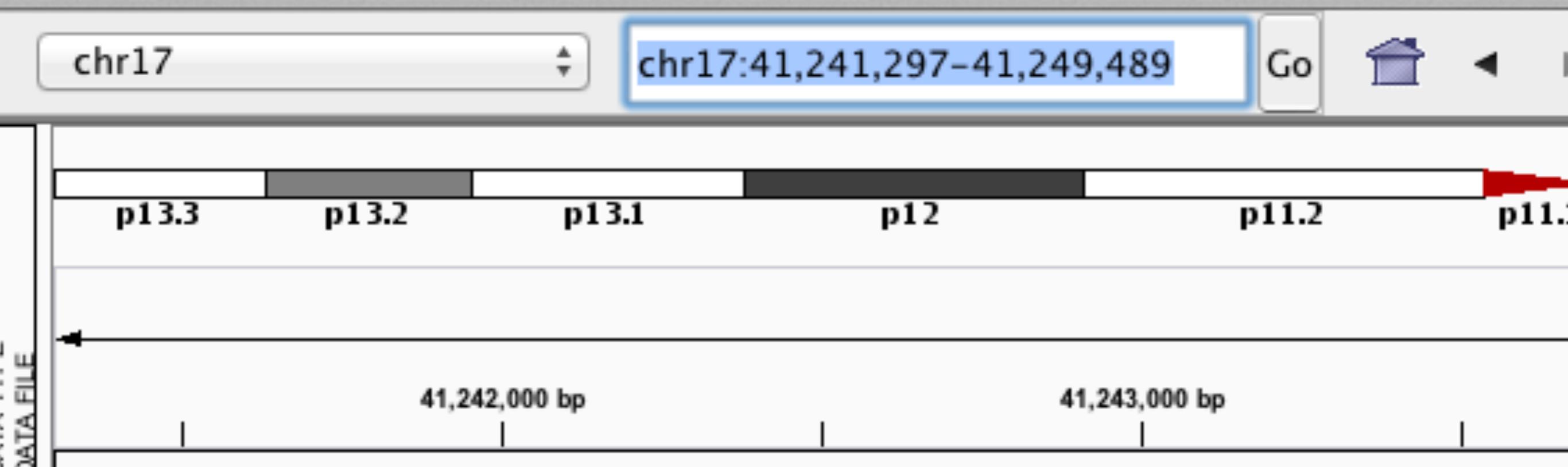
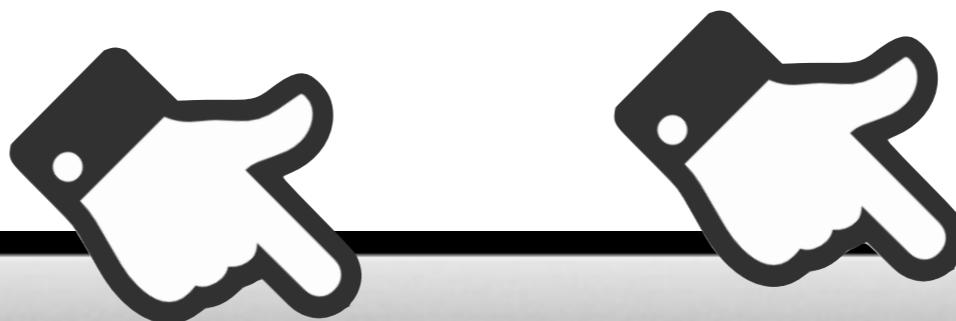
[Archived Versions](#)

igv_lm.jnlp Show All

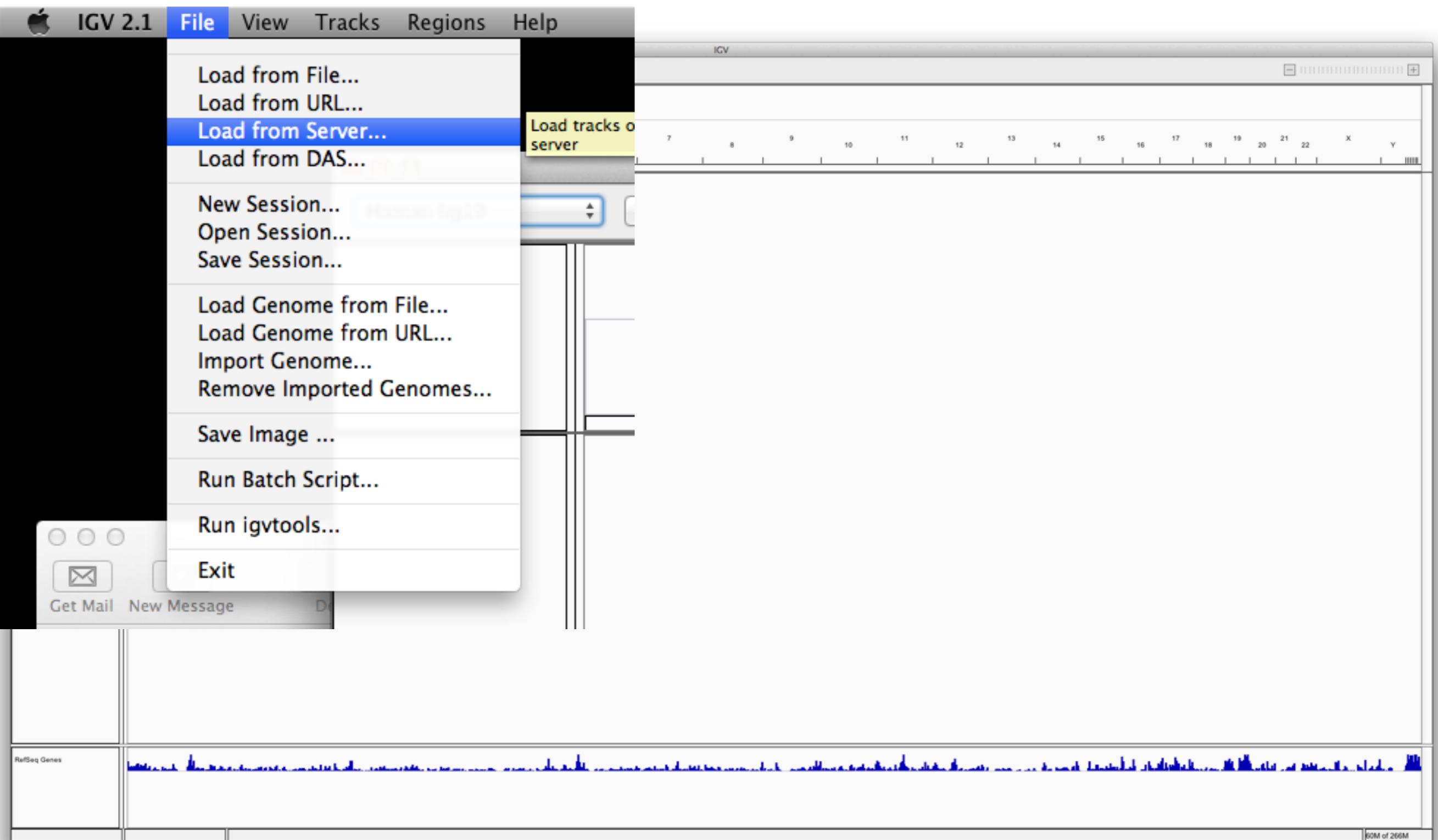
Victory.



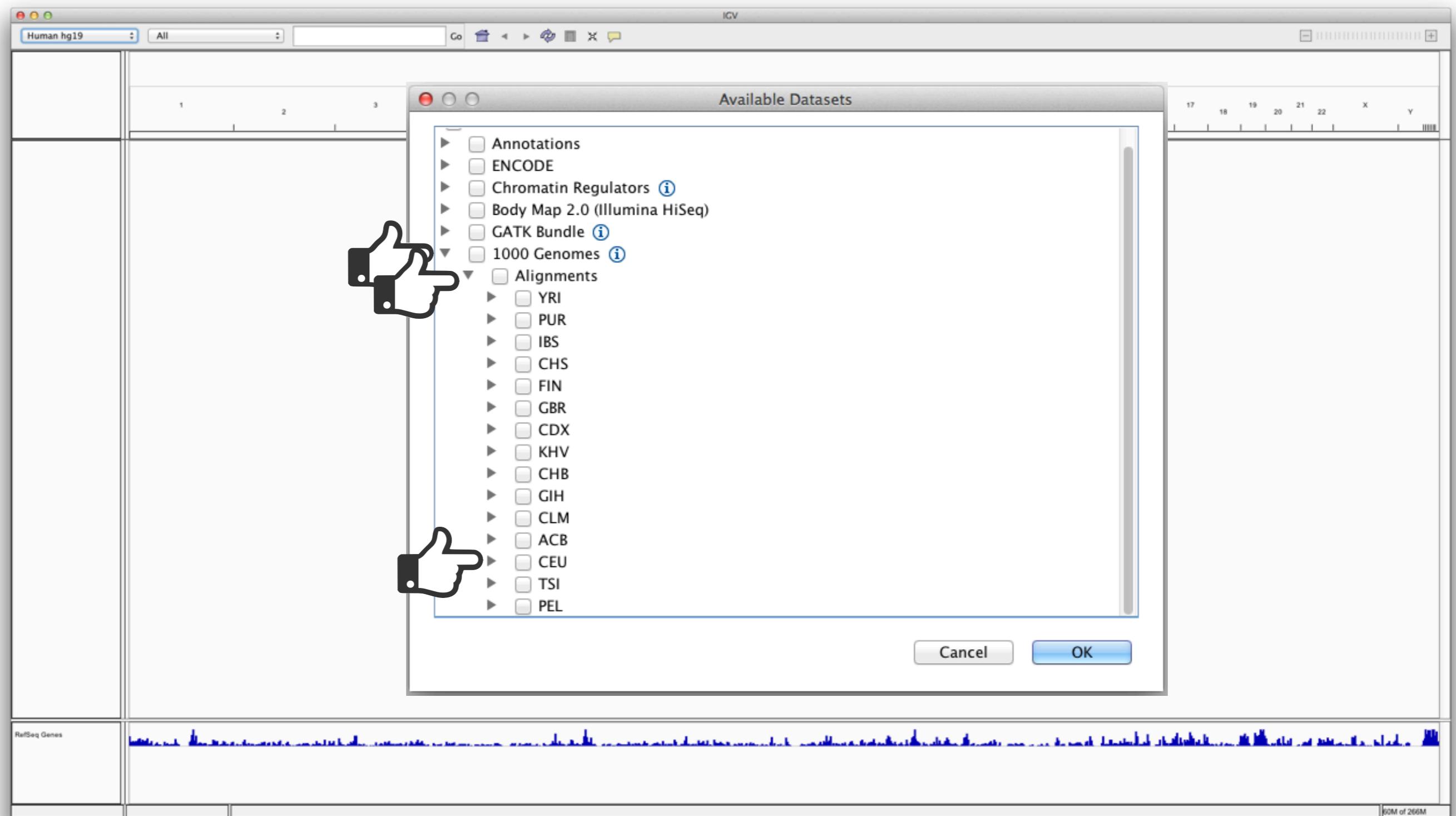
Zoom in.

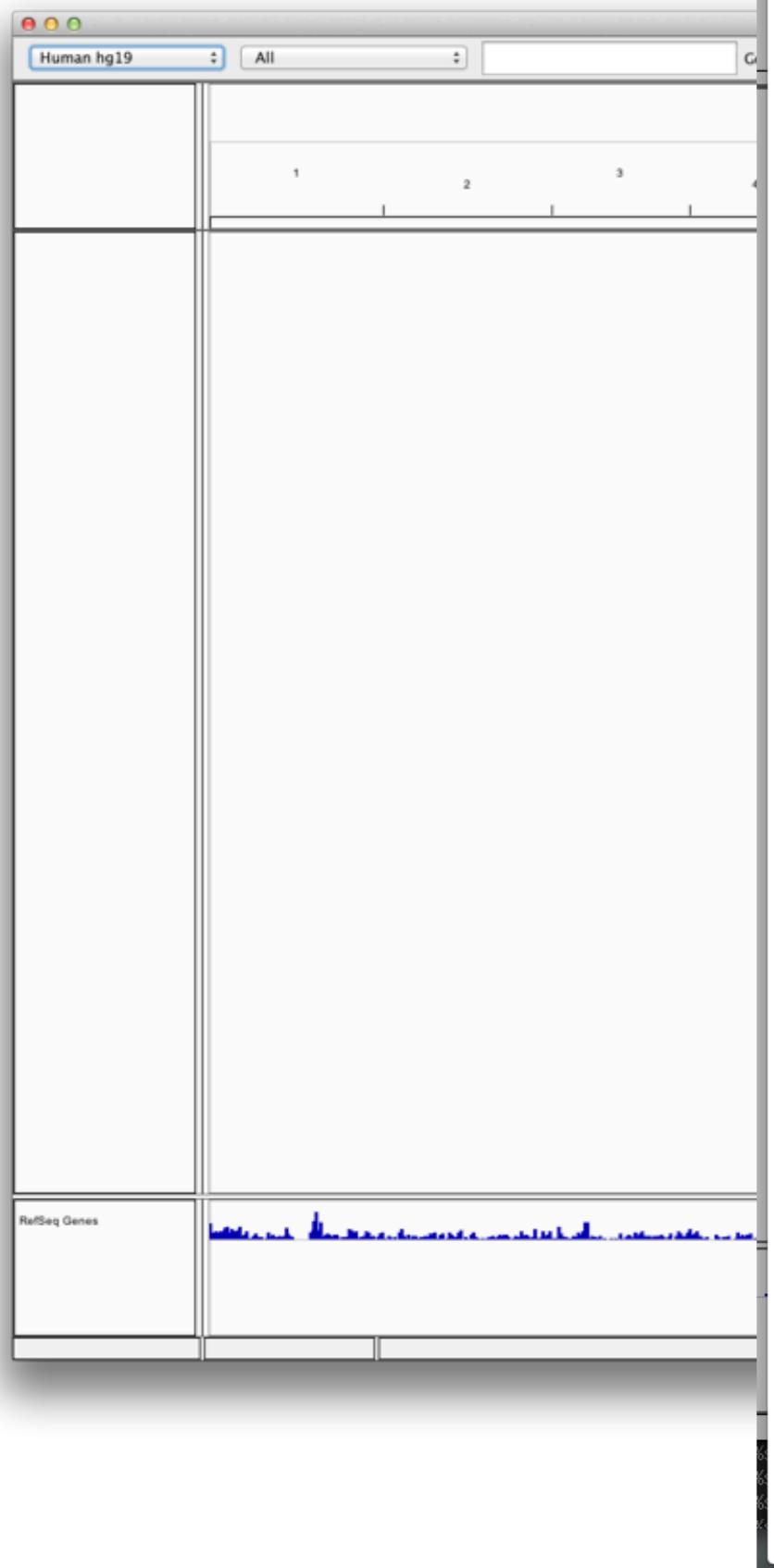


Load our trio.



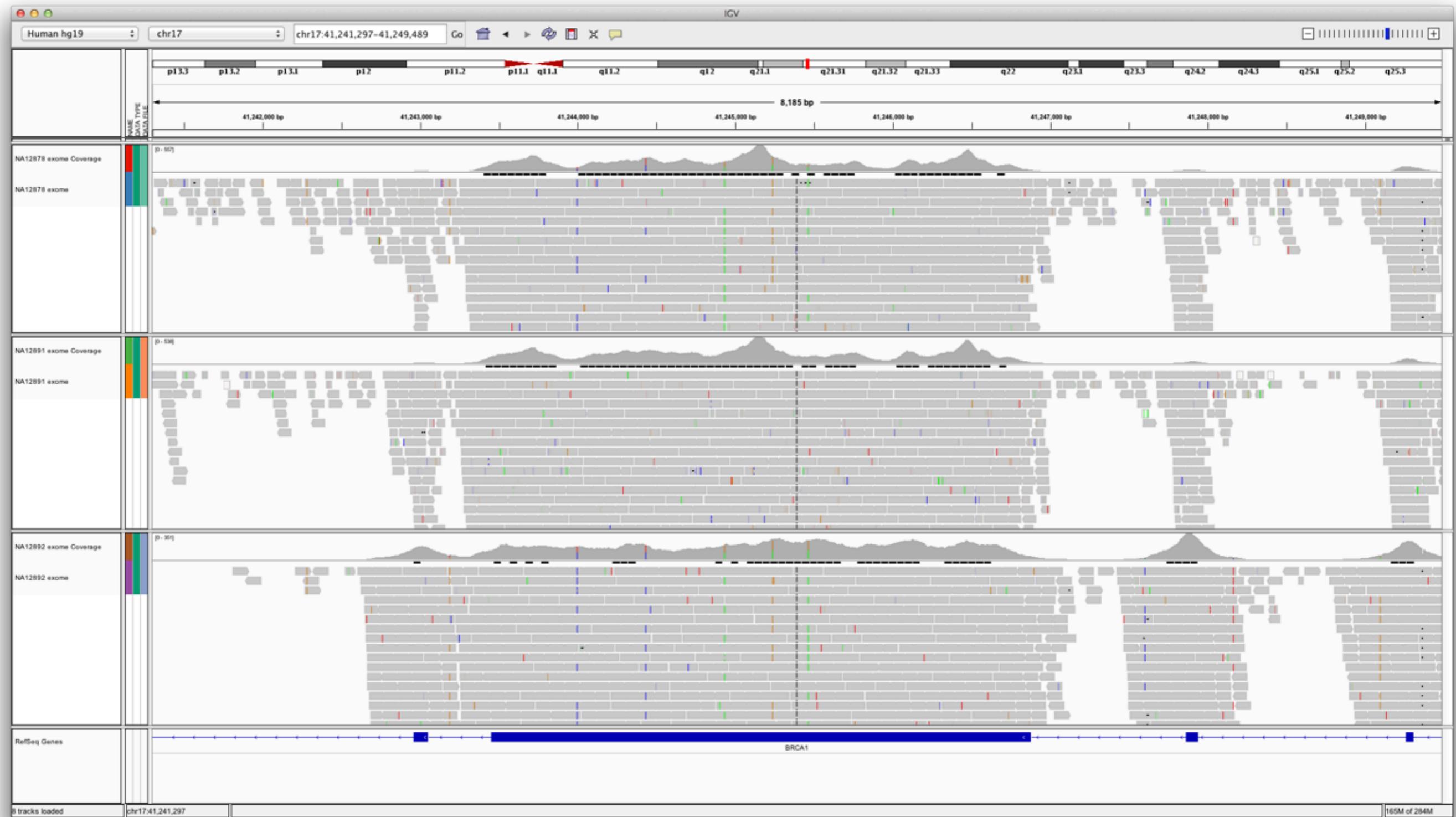
Load our trio.



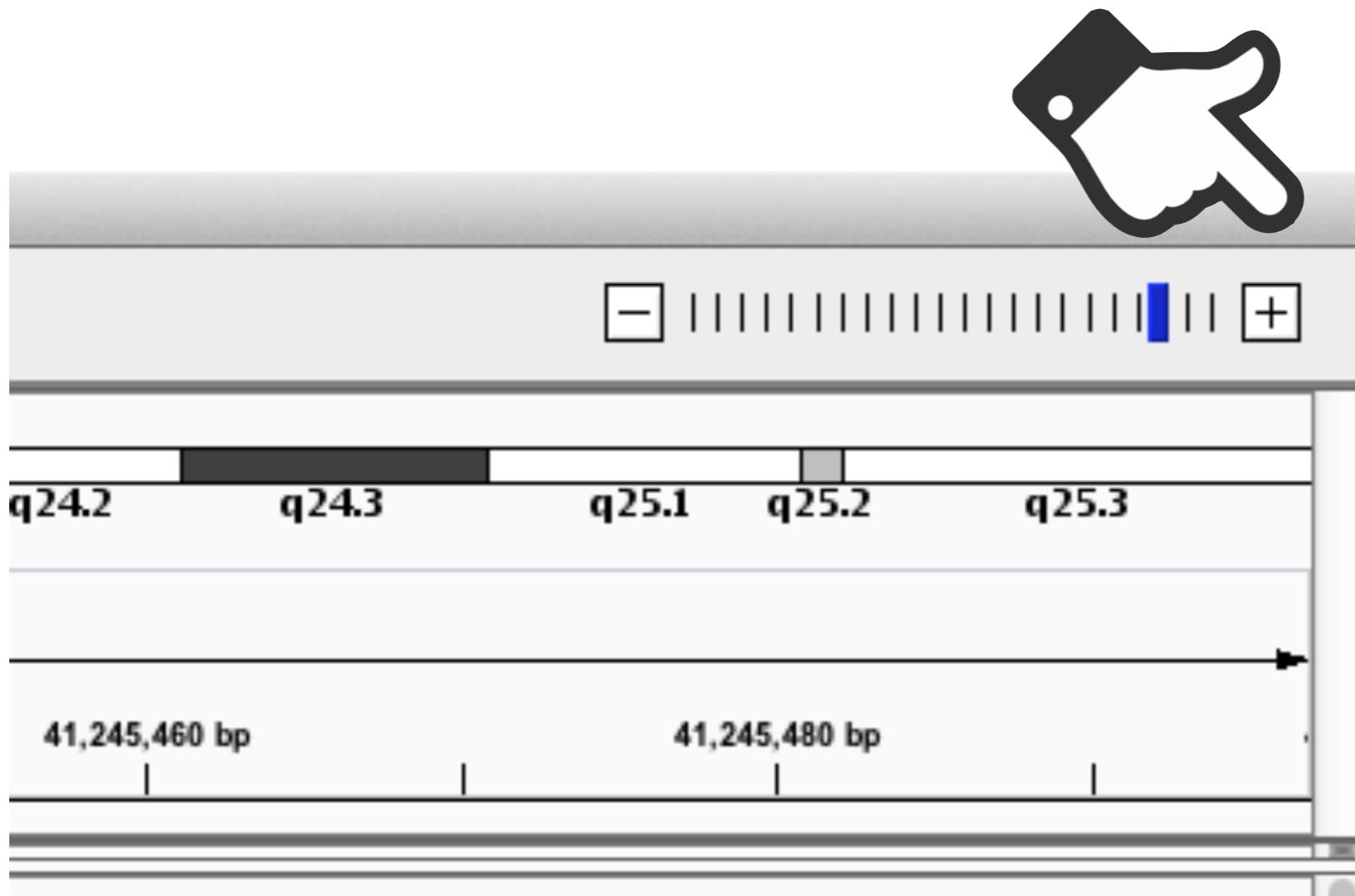


Get a coffee. Internet(s) at work.

Yell if you don't see this.



Zoom in some more.



Click '+' 6
exactly 6 times.

Data. Smells good.

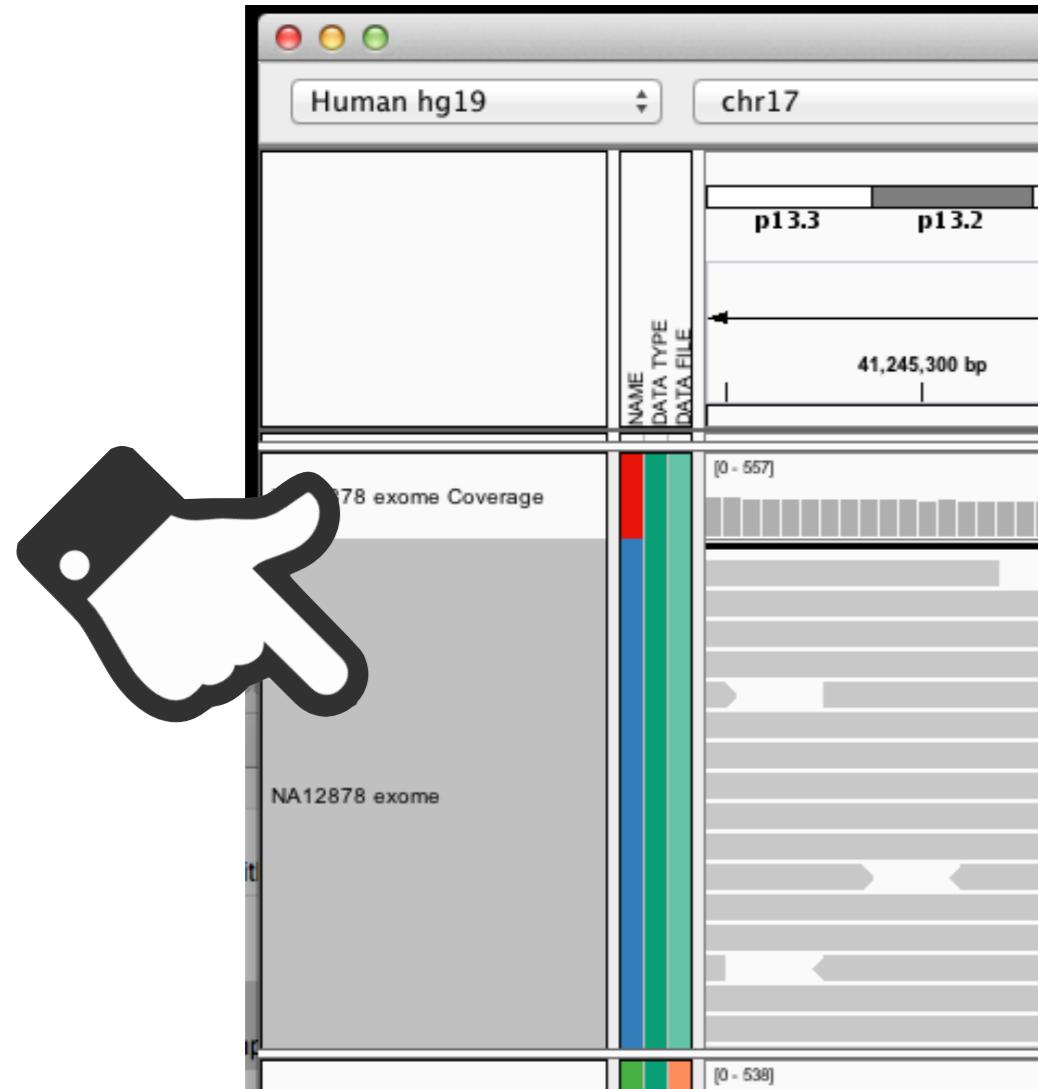


What are we looking at?

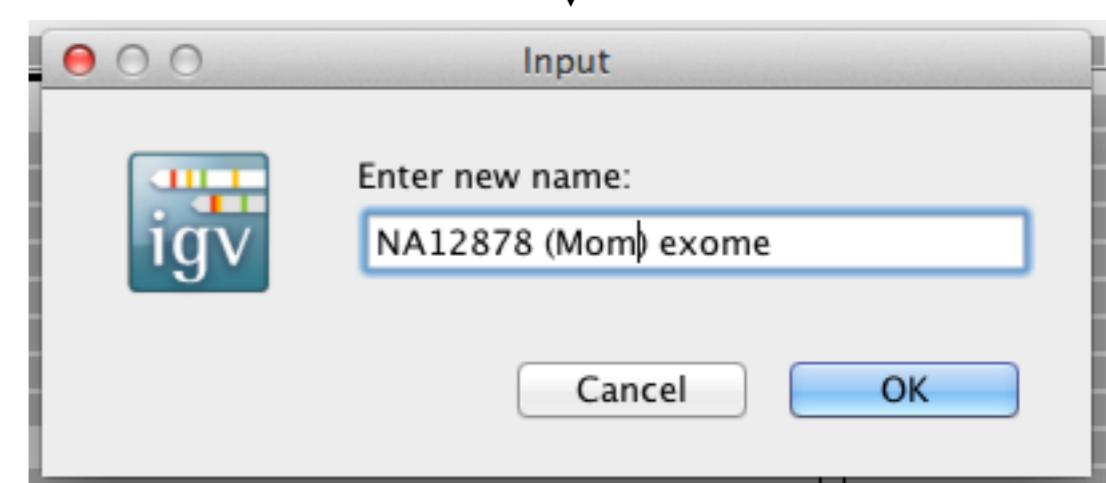
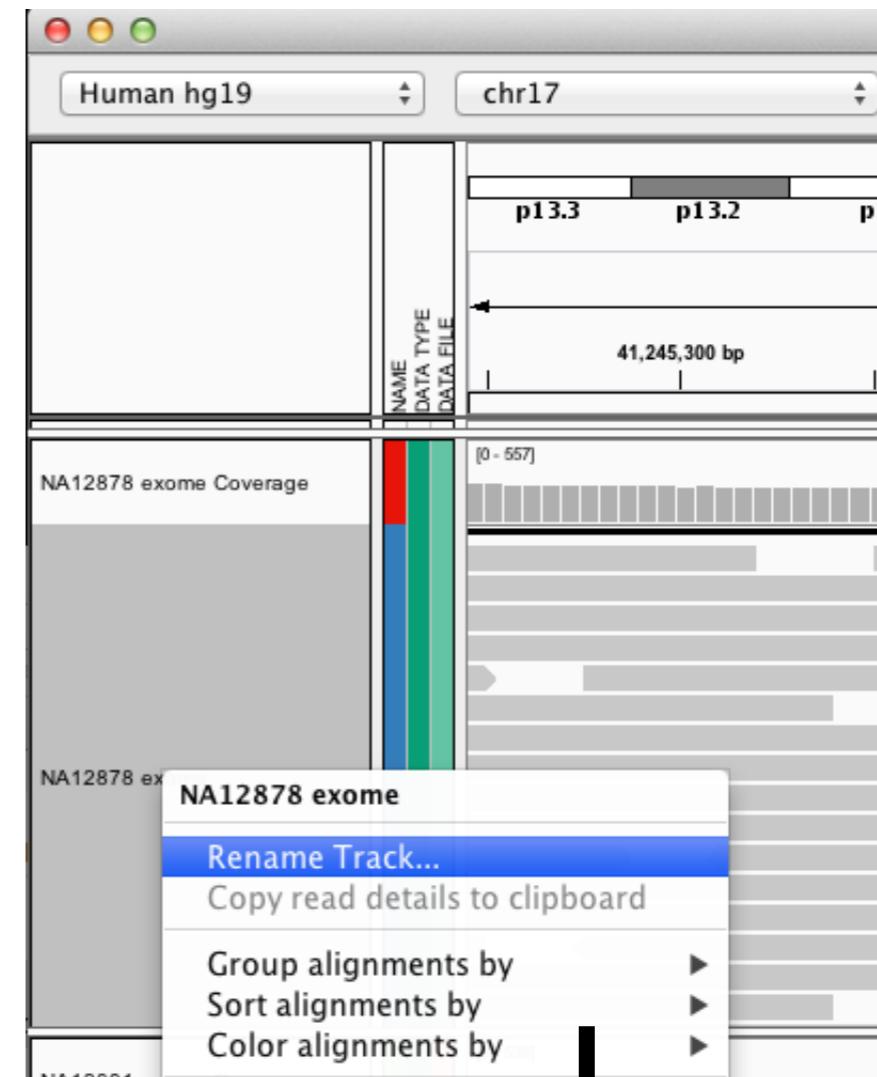
Walk through what the alignments tell you, etc.



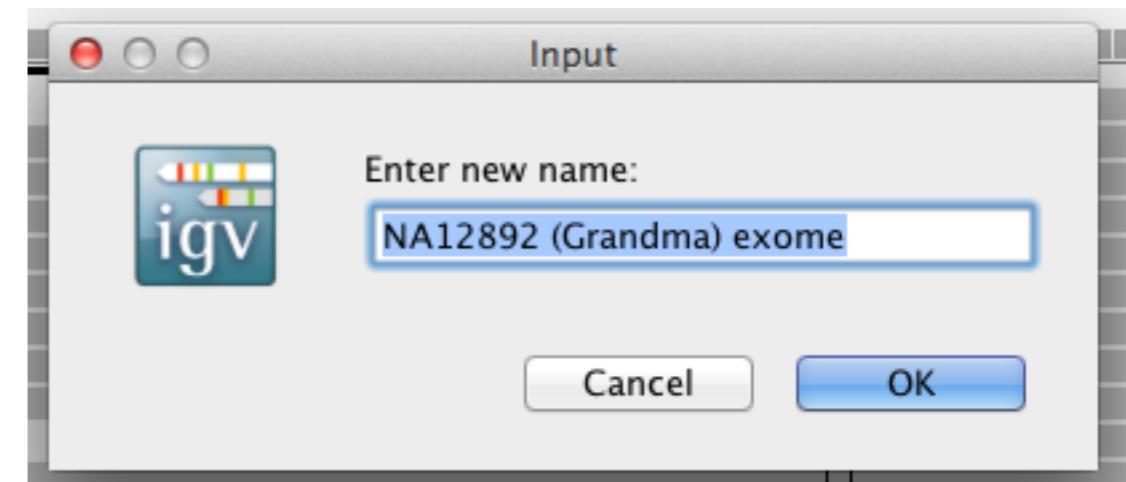
Let's relabel or samples for clarity



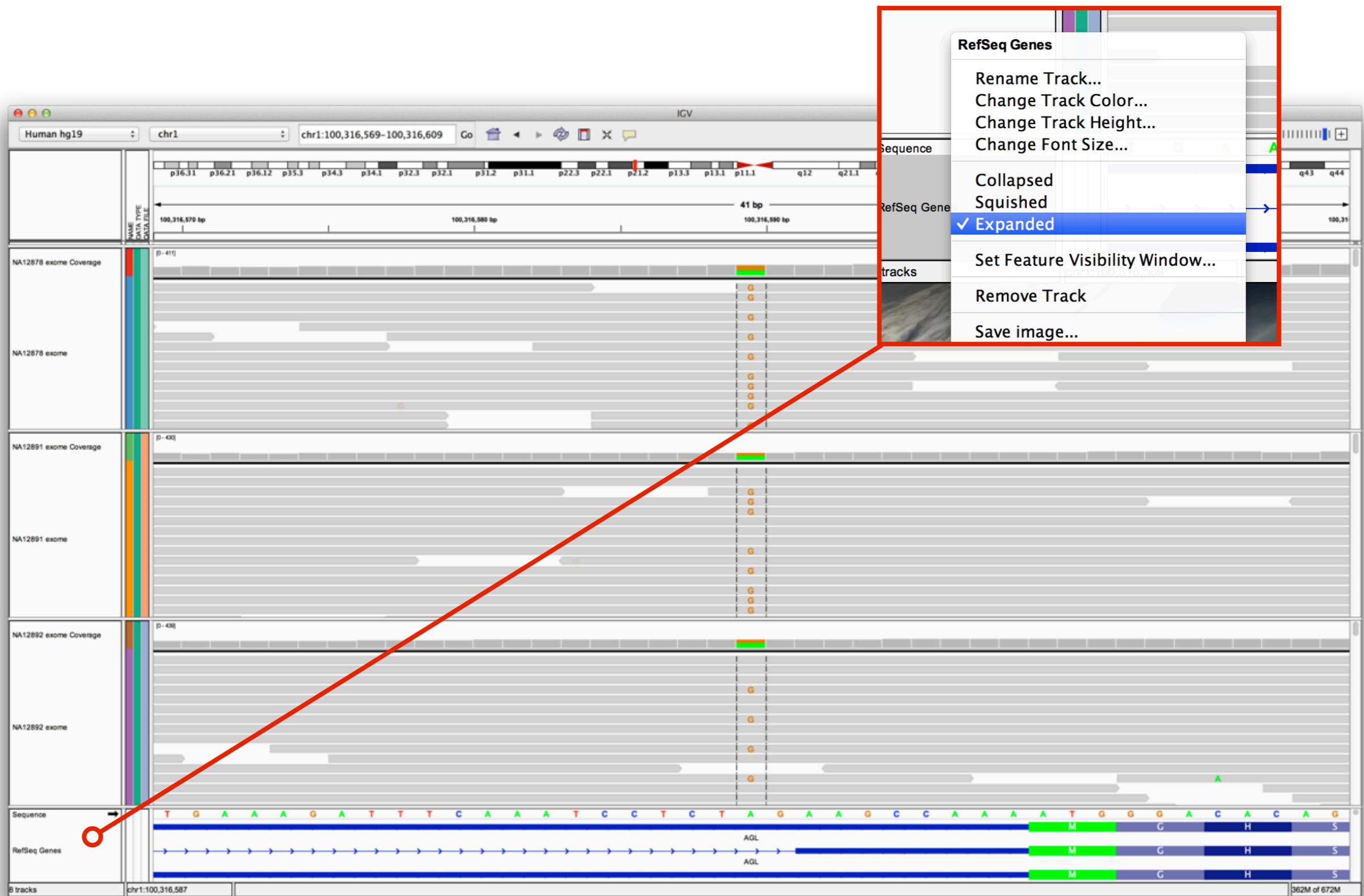
Right click



Repeat for grandma and grandpa



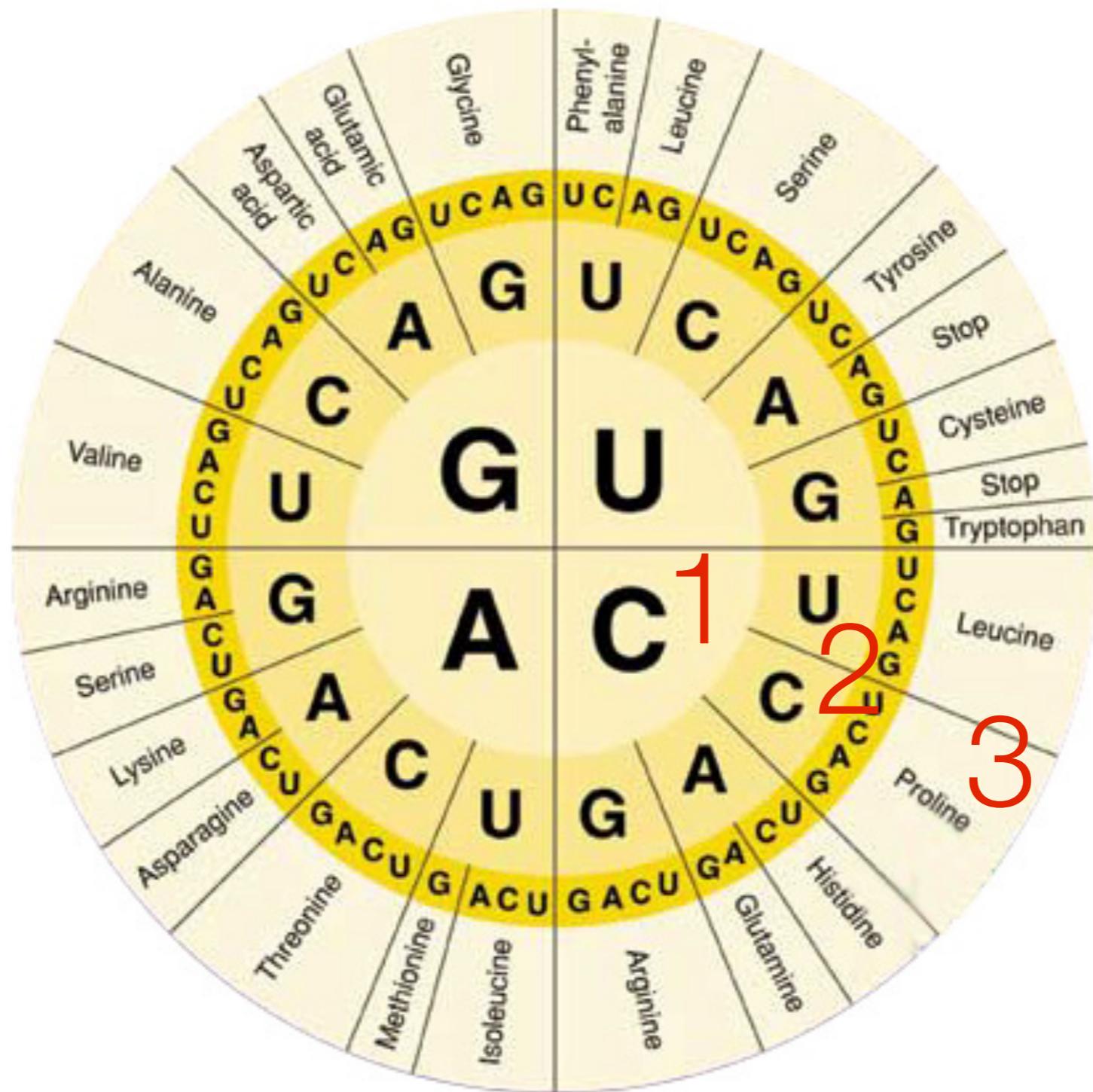
Expand the genes track so that you see all gene transcripts



Question #1

How could we use the alignment data to verify that the samples are the sex we expect them to be?

Genetic code

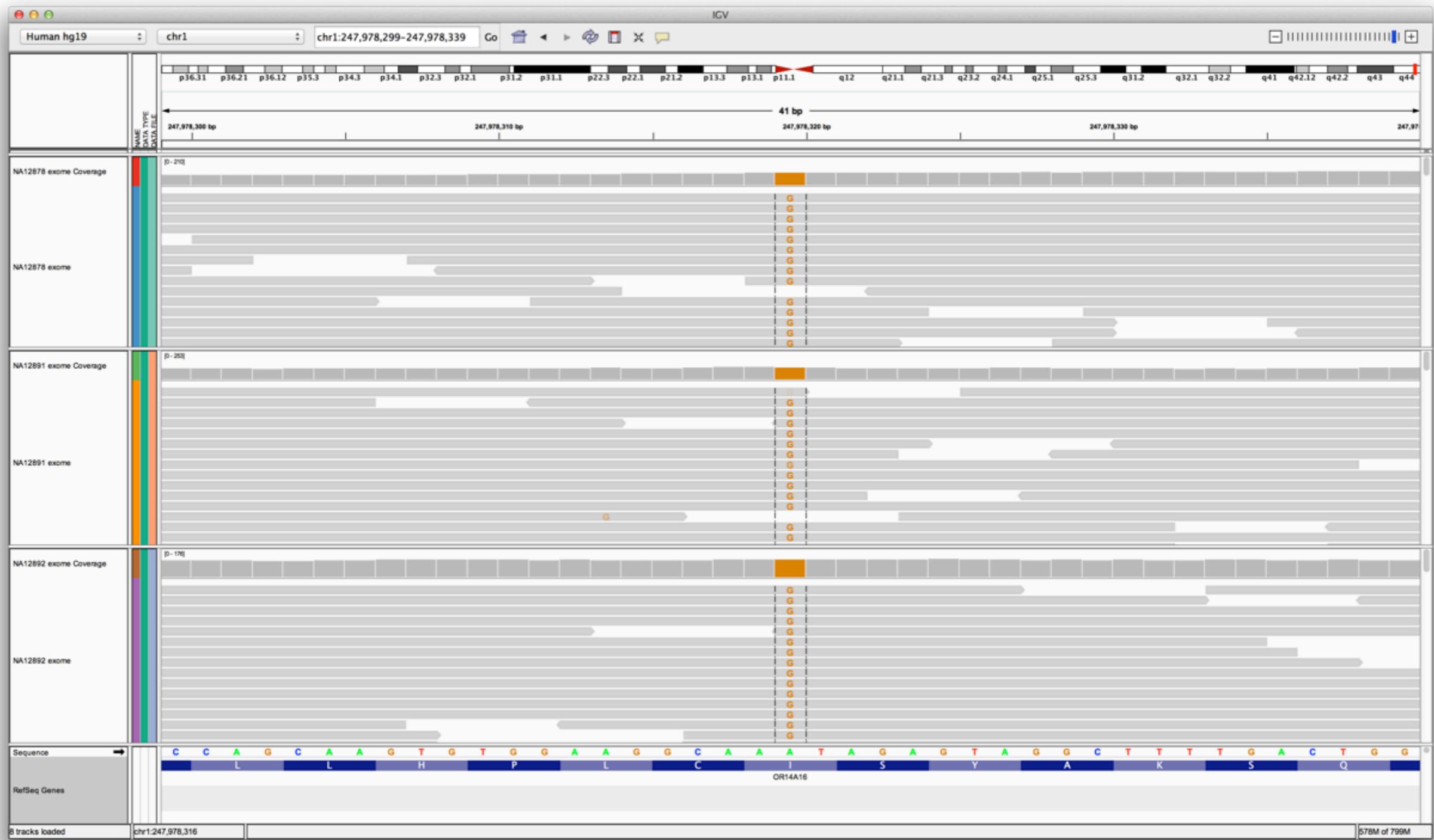


For question #2, inspect each of
the following variants:

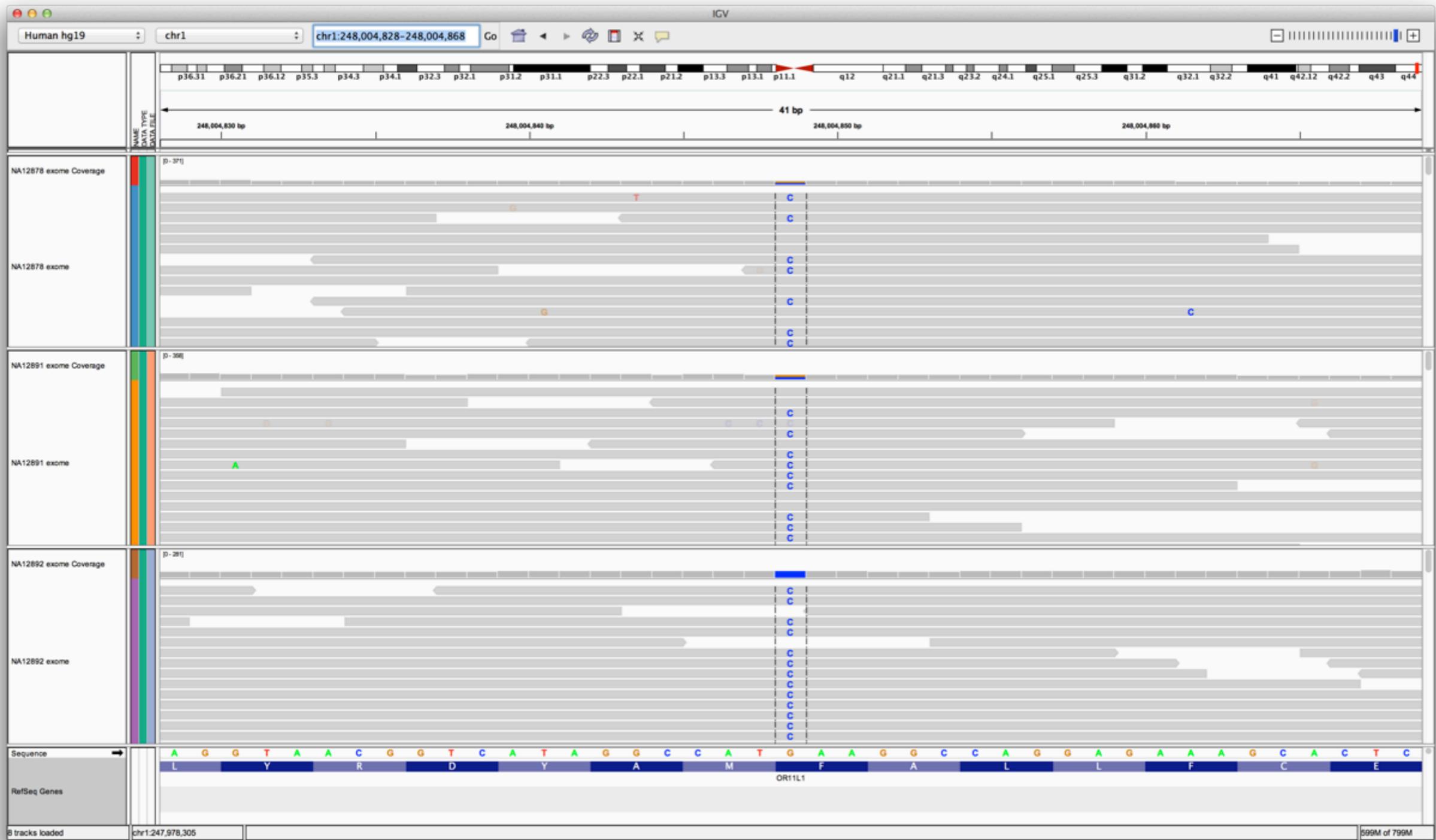
chr1:159,558,223-159,558,293



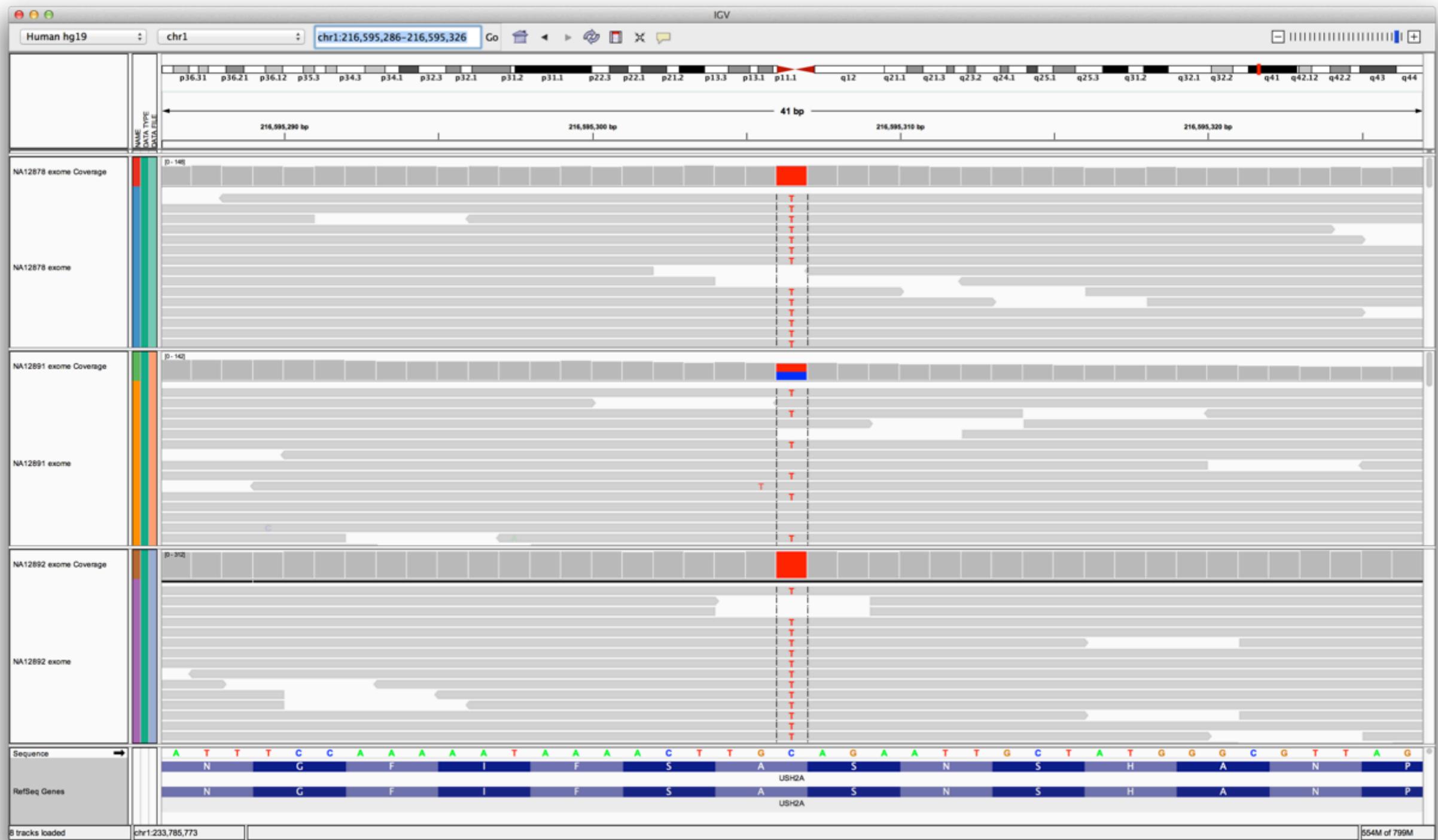
chr1:247978318-247978319



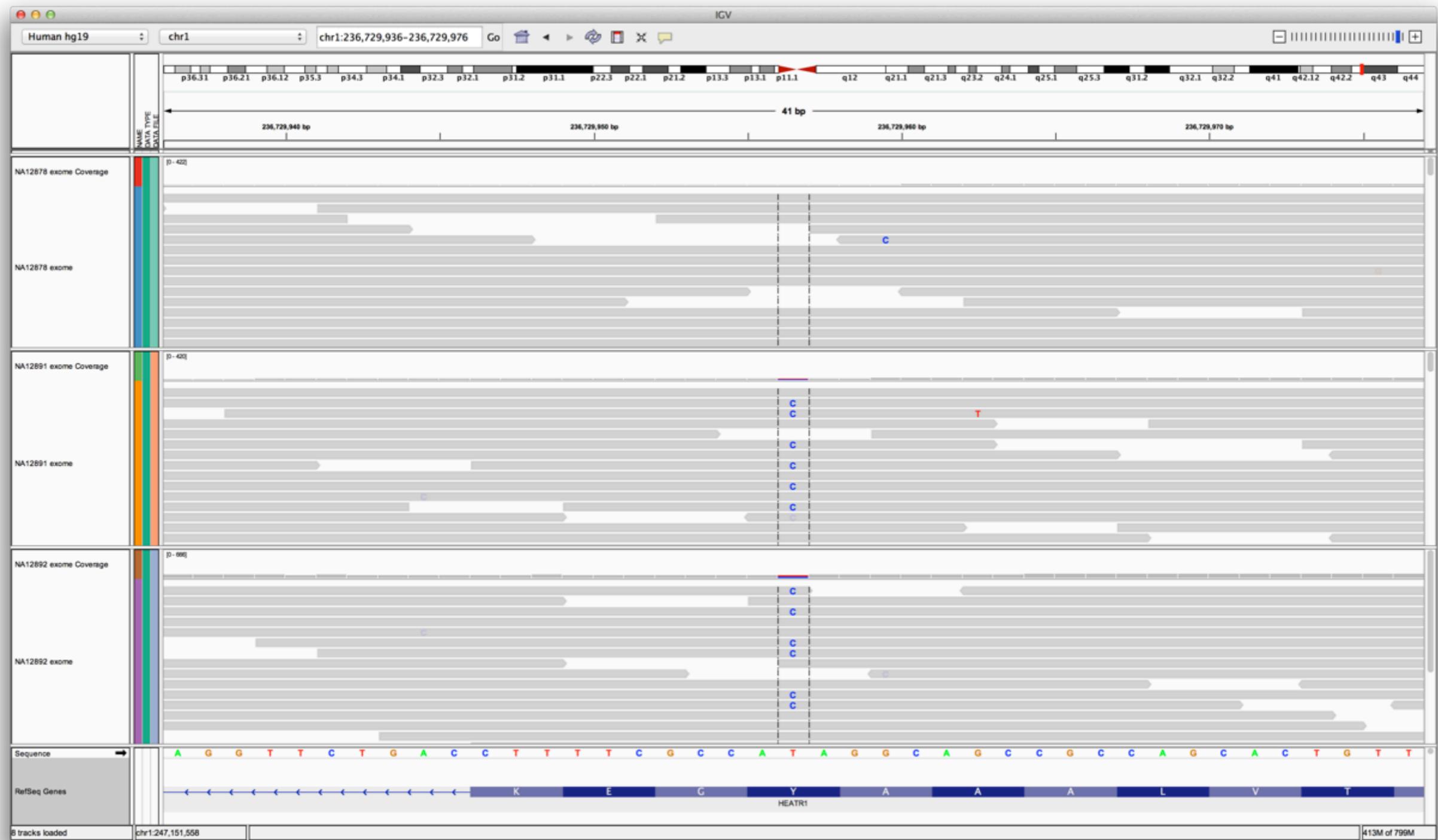
chr1:248004847-248004848



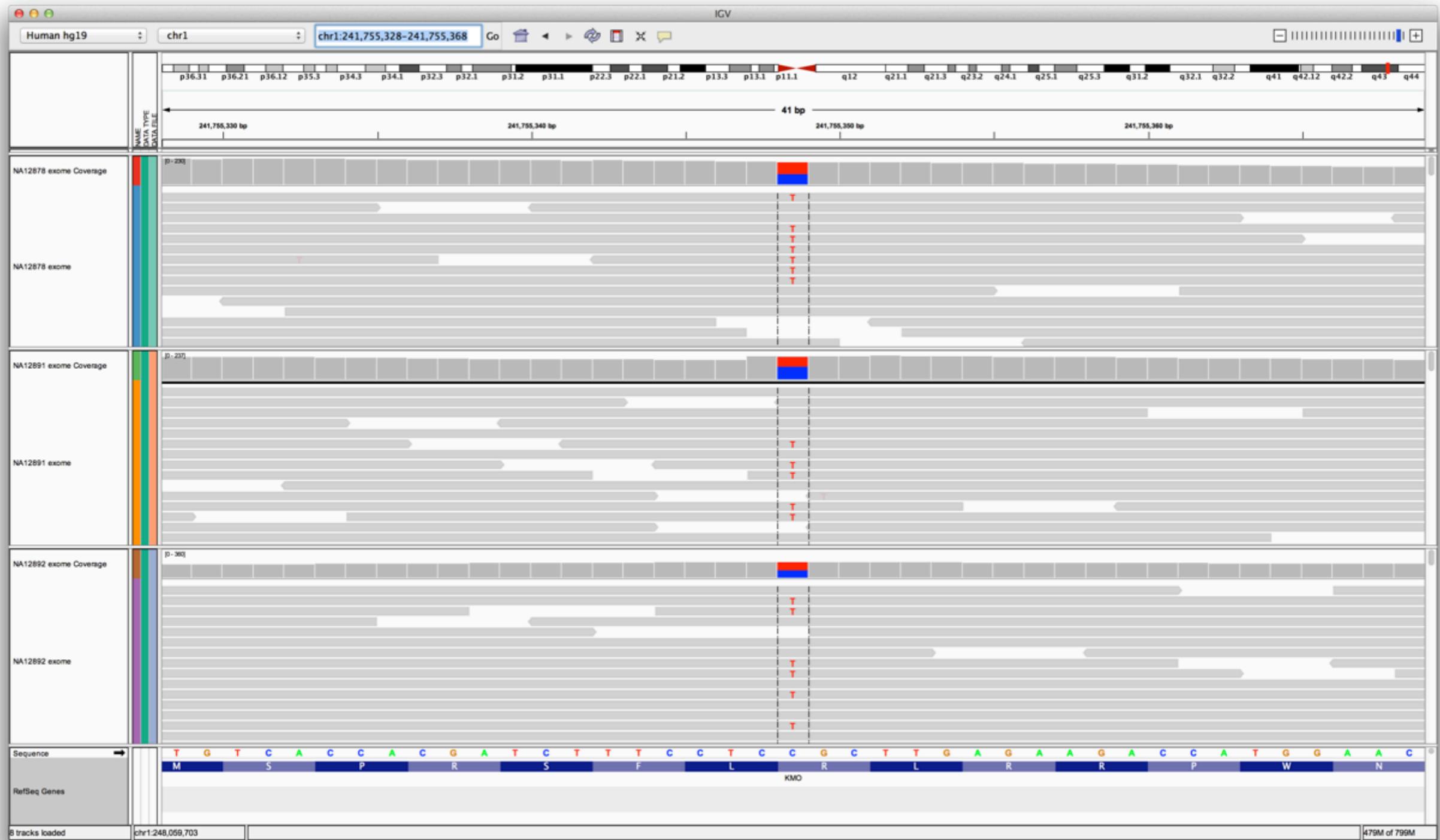
chr1:216595305-216595306



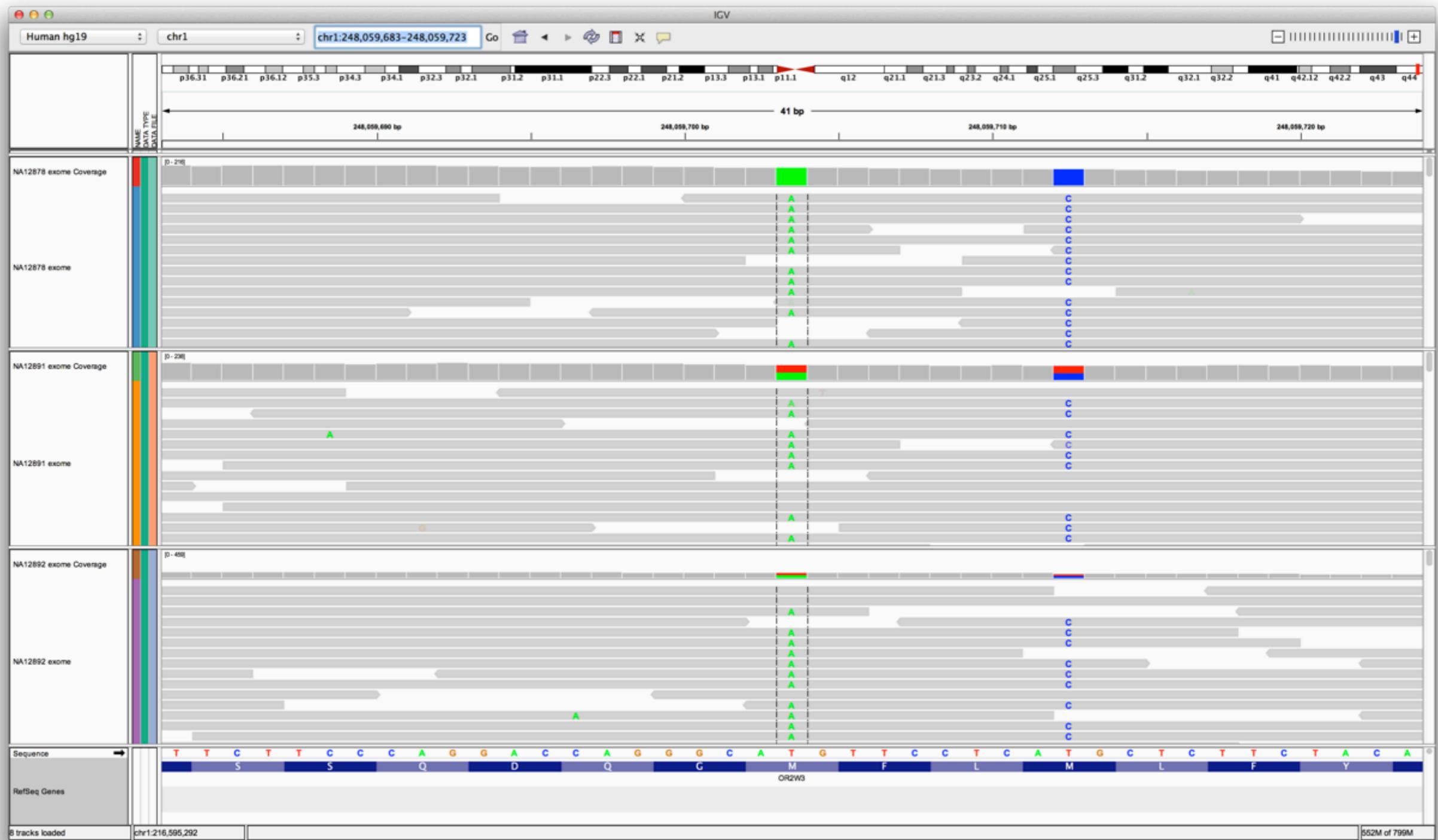
chr1:236729956-236729956



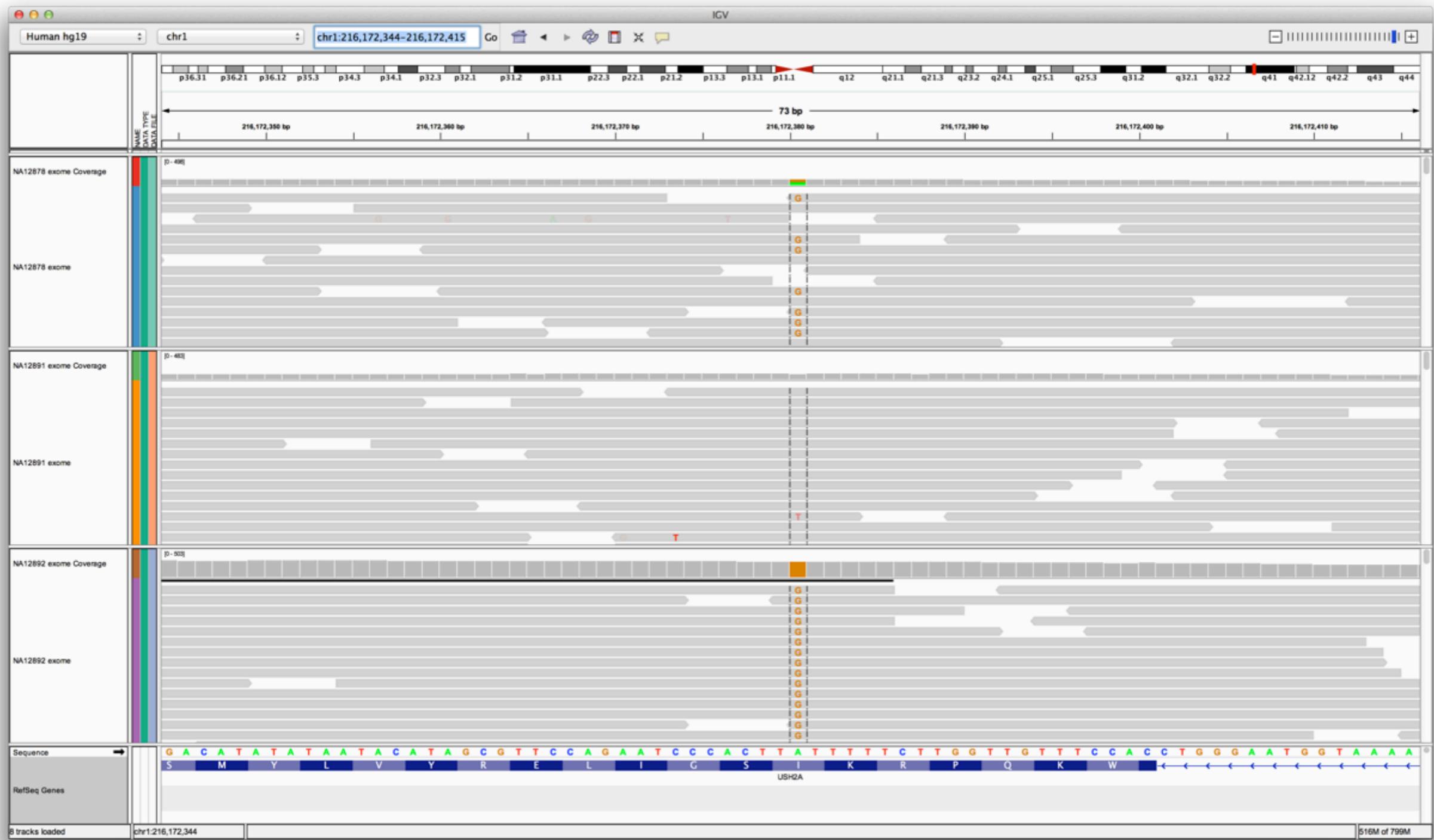
chr1:241755348-241755348



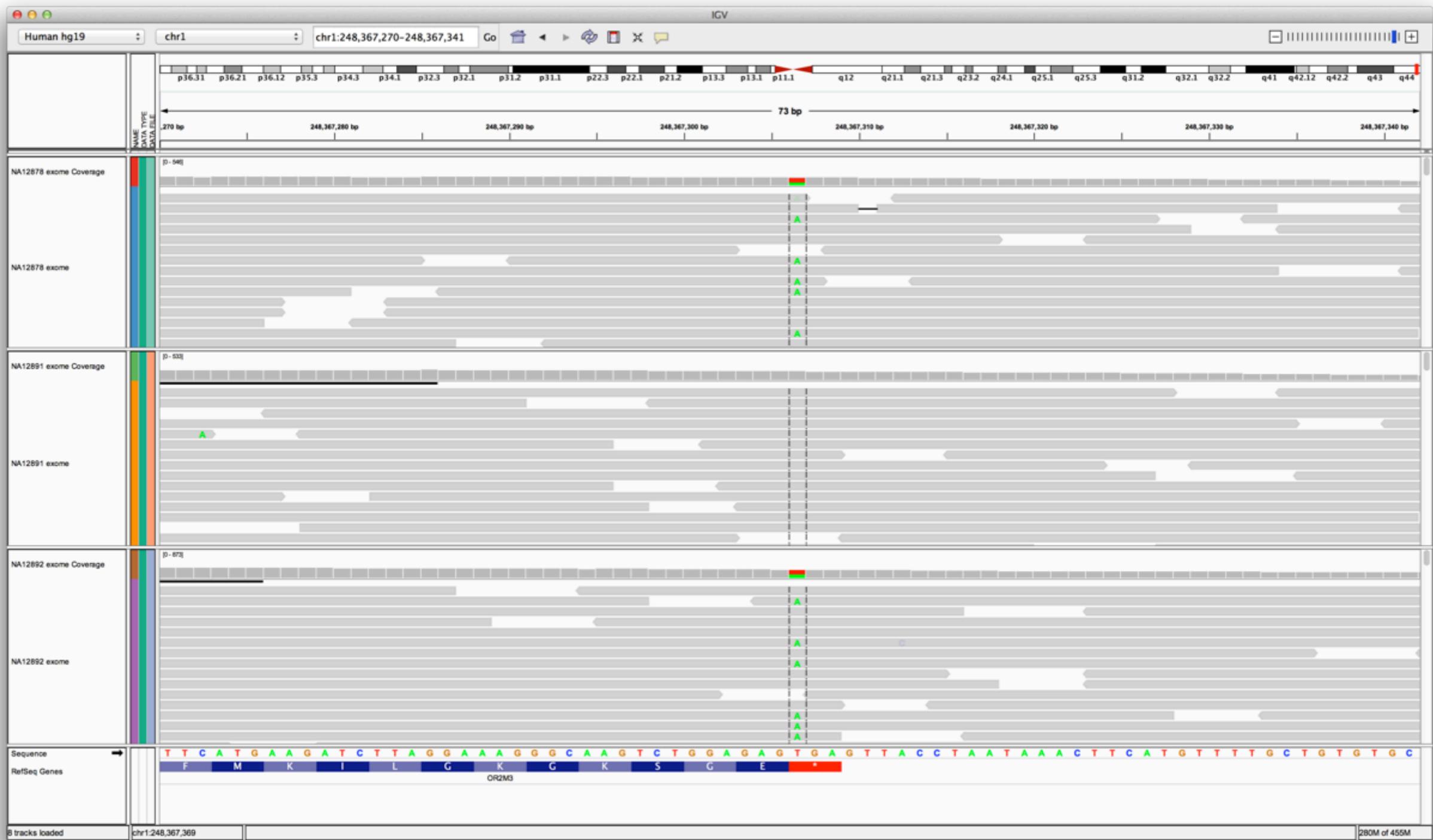
chr1:248059702-248059703



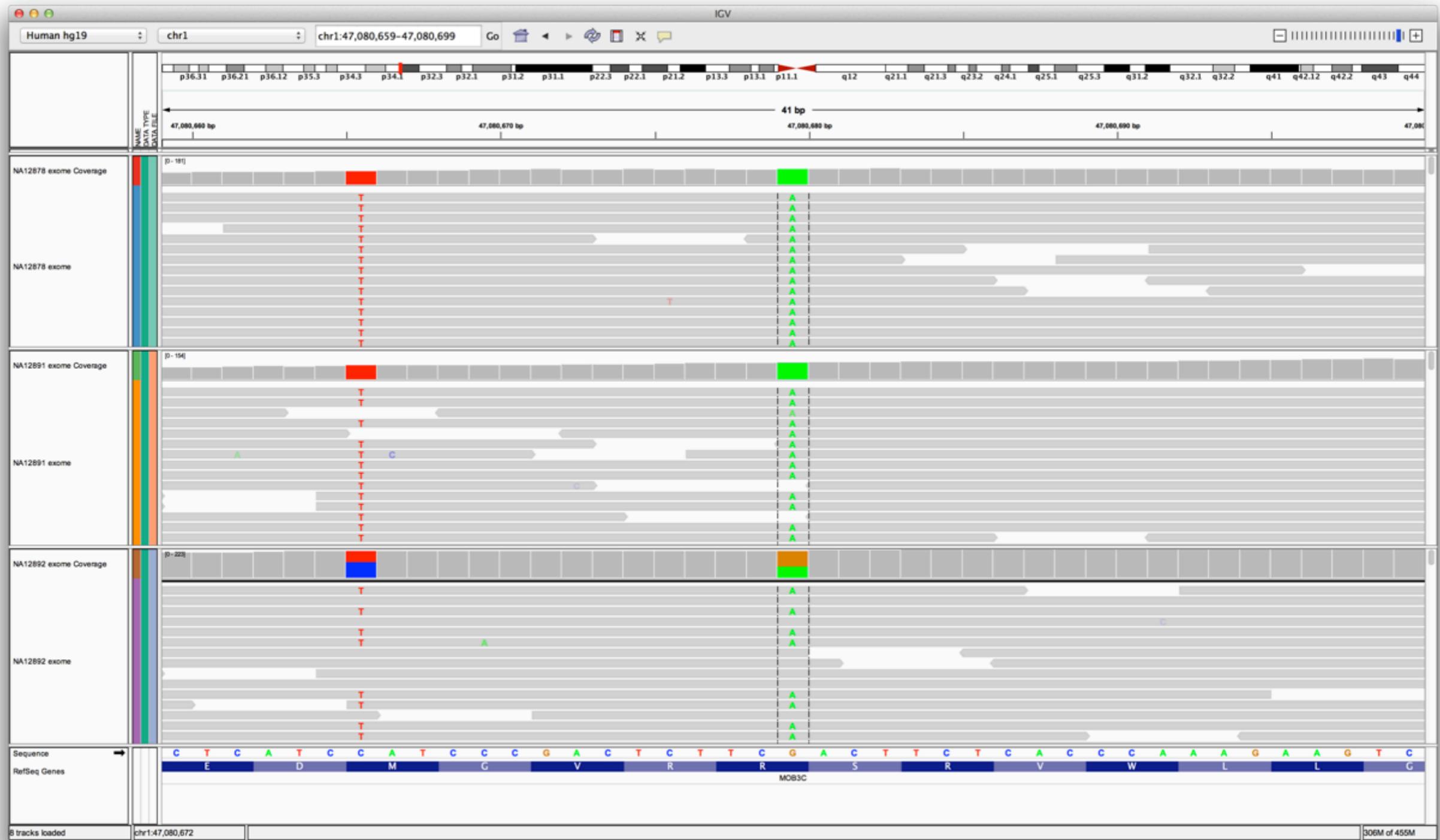
chr1:216172379-216172380



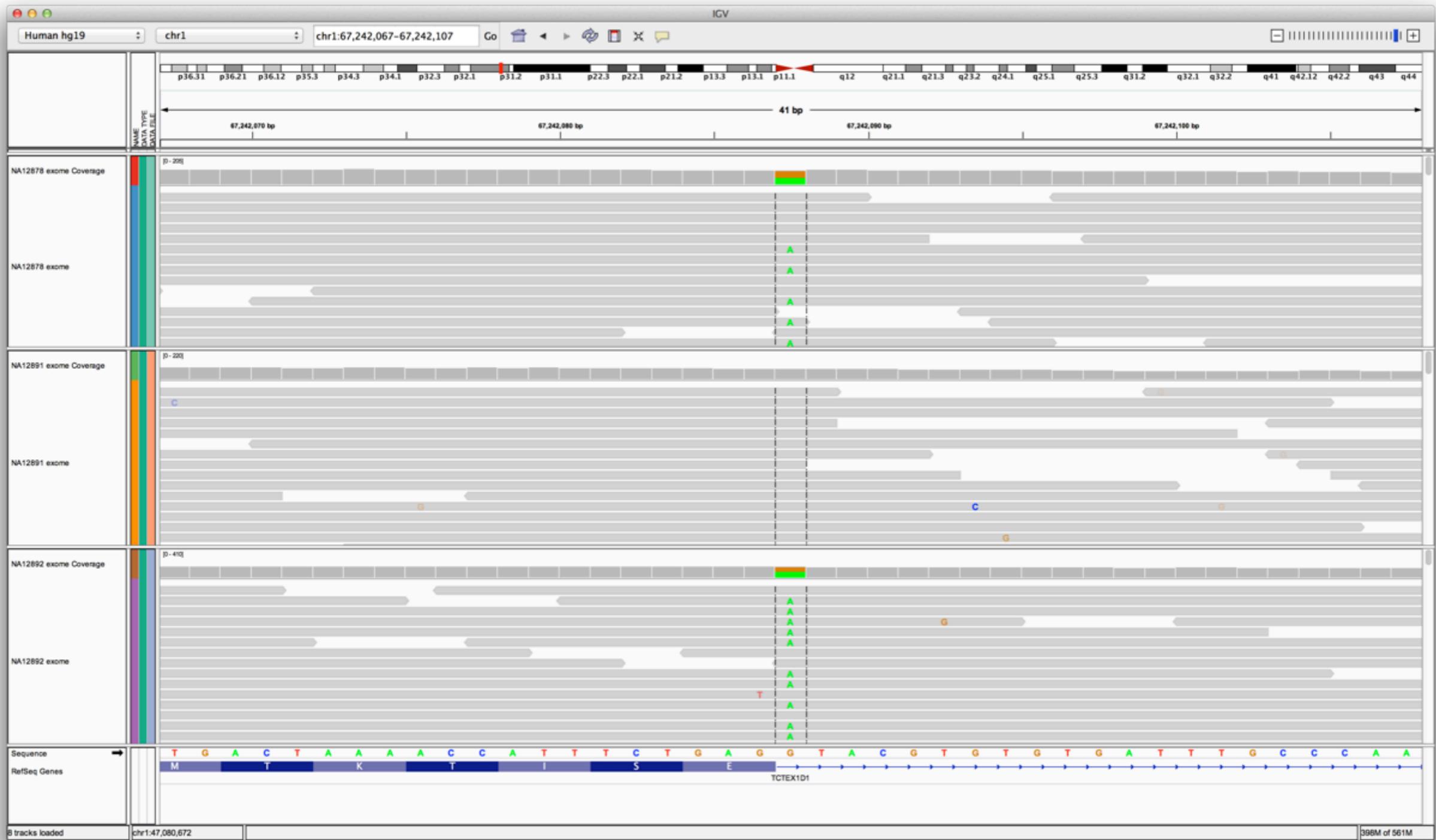
chr1:248367305-248367306



chr1:47080678-47080679



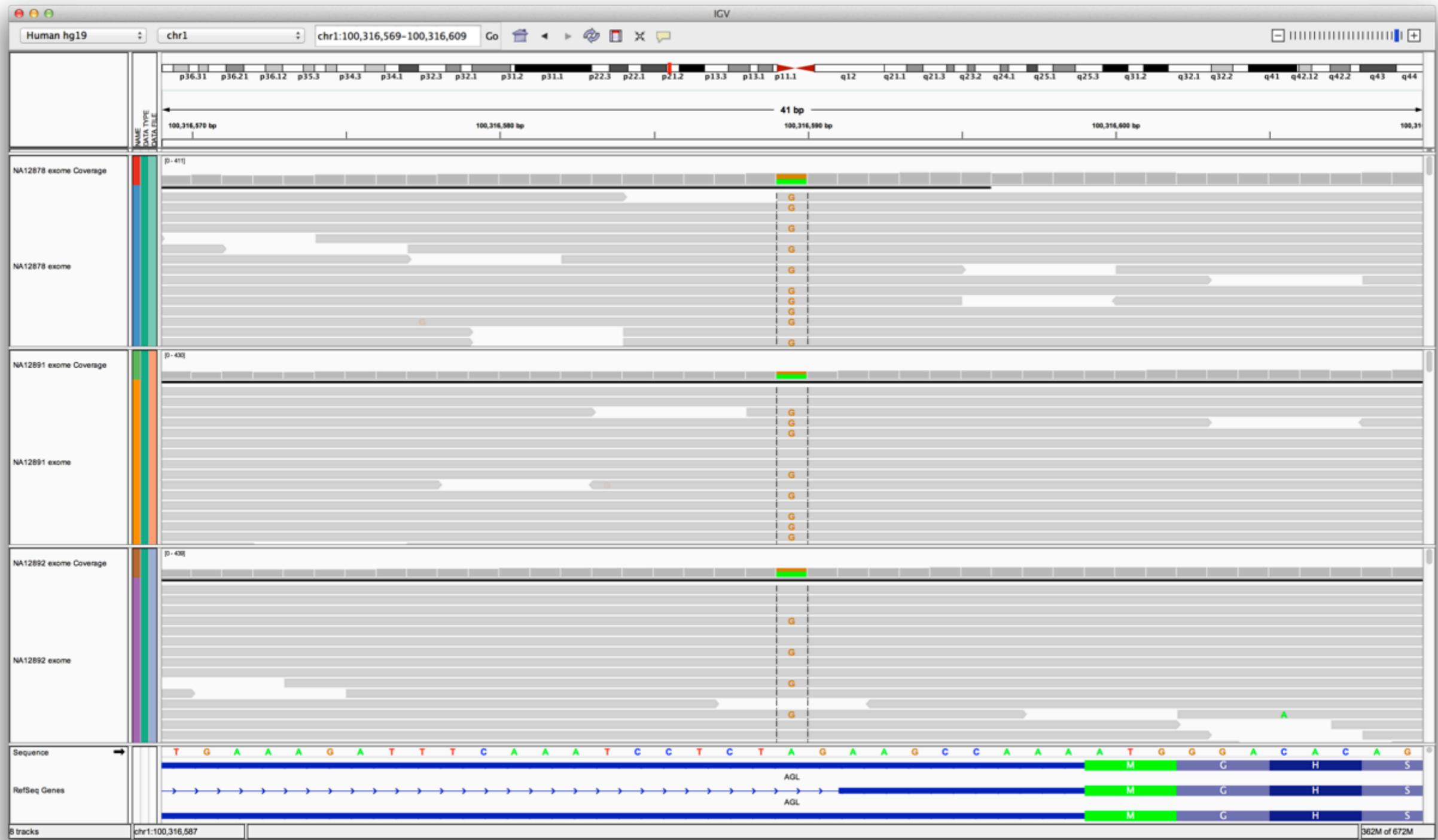
chr1:67242086-67242087



chr1:248,024,002-248,024,144



chr1:100316588-100316589



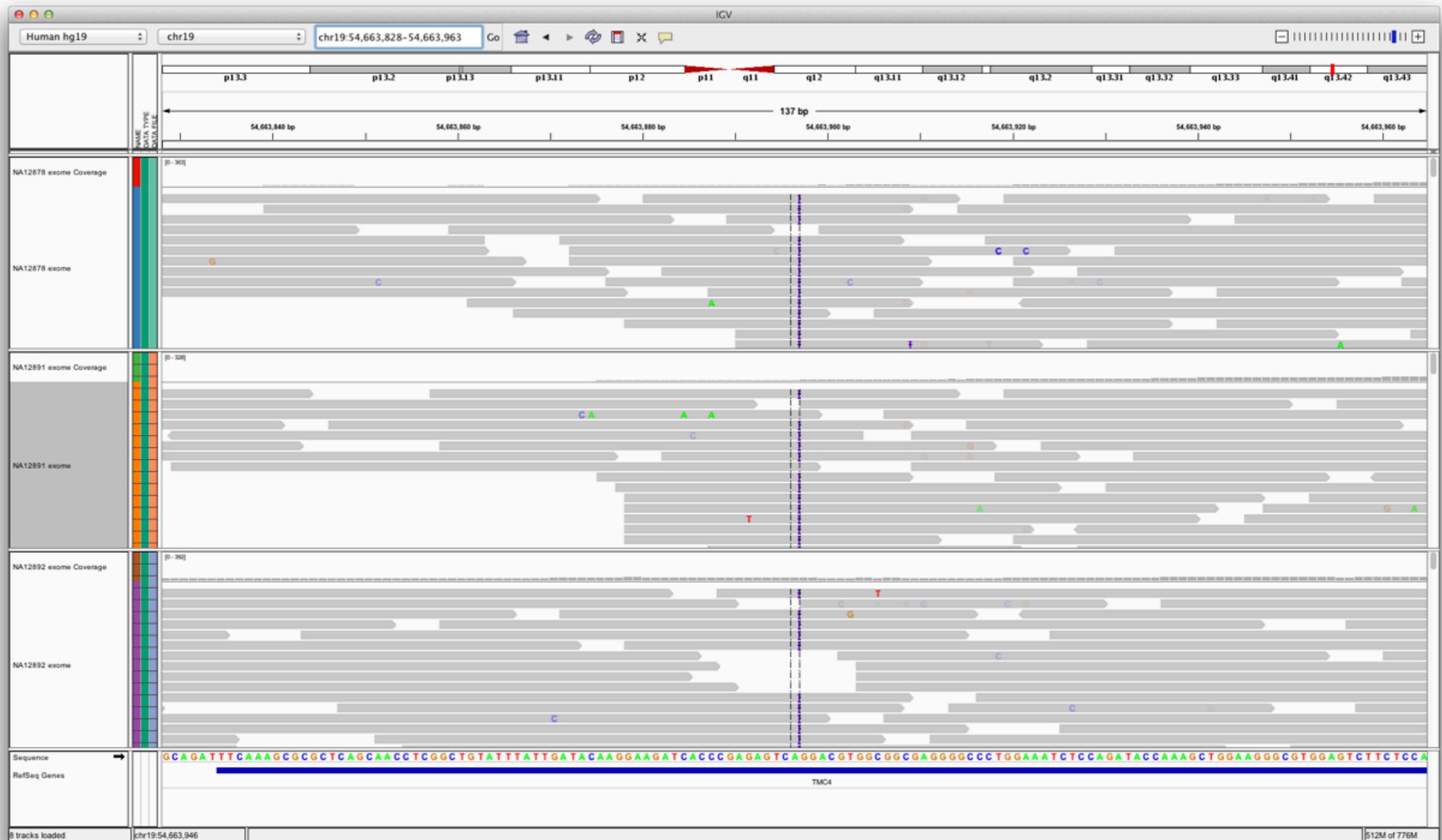
chr14:56763333-56763373



chr16:30,768,388-30,768,596



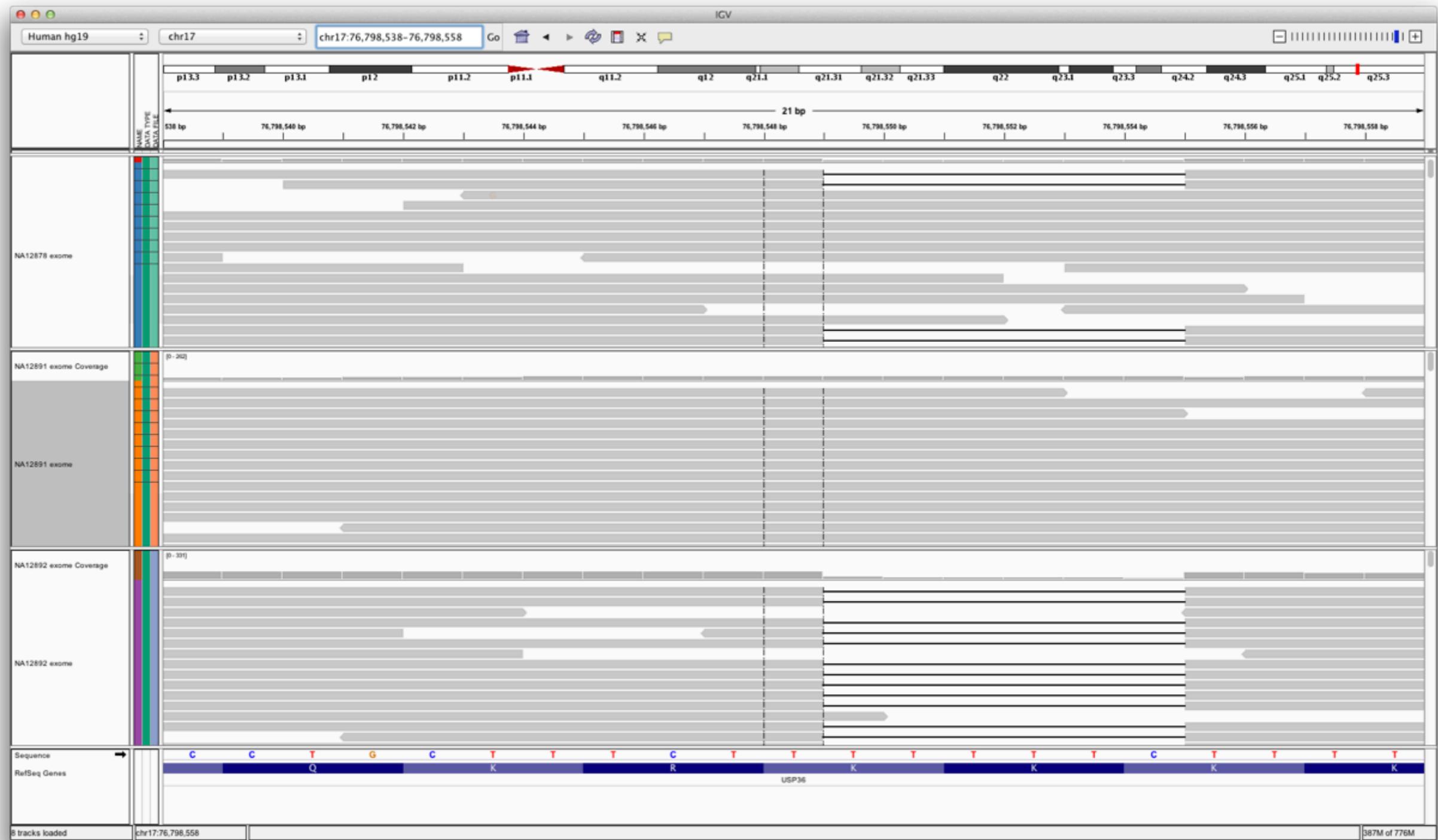
chr19:54,663,828-54,663,963



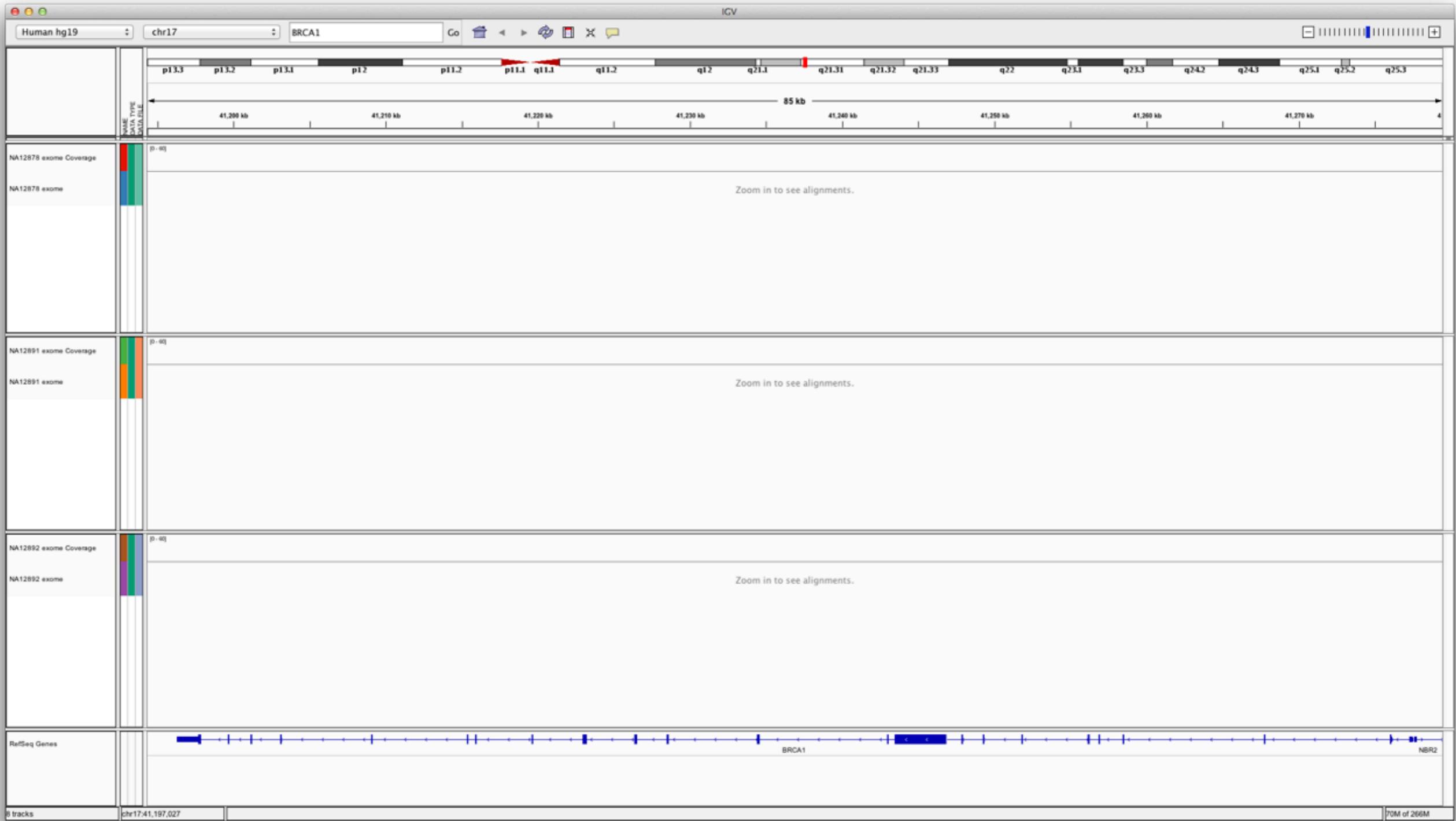
chr19:52,887,077-52,887,212



chr17:76,798,538-76,798,558



Navigate to BRCA1



Where's the data?

