Visualizing proteomics data in genomic context using the UCSC Genome Browser

Kate R. Rosenbloom, Hiram Clawson, Mark Diekhans, Luvina Guruvadoo, Rachel A. Harte[1,2], Donna Karolchik, Brian T. Lee, Brian J. Raney, Ann S. Zweig, David Haussler[1,3], Robert M. Kuhn, W. James Kent

[1] UC Santa Cruz Genomics Institute, and [3]Howard Hughes Medical Institute, University of California Santa Cruz, SC, CA, USA
[2] Invitae, San Francisco, CA, USA

Zhi Sun, Eric W. Deutsch — Institute for Systems Biology, Seattle, WA, USA
Christopher R. Kinsinger, Henry Rodriguez — National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
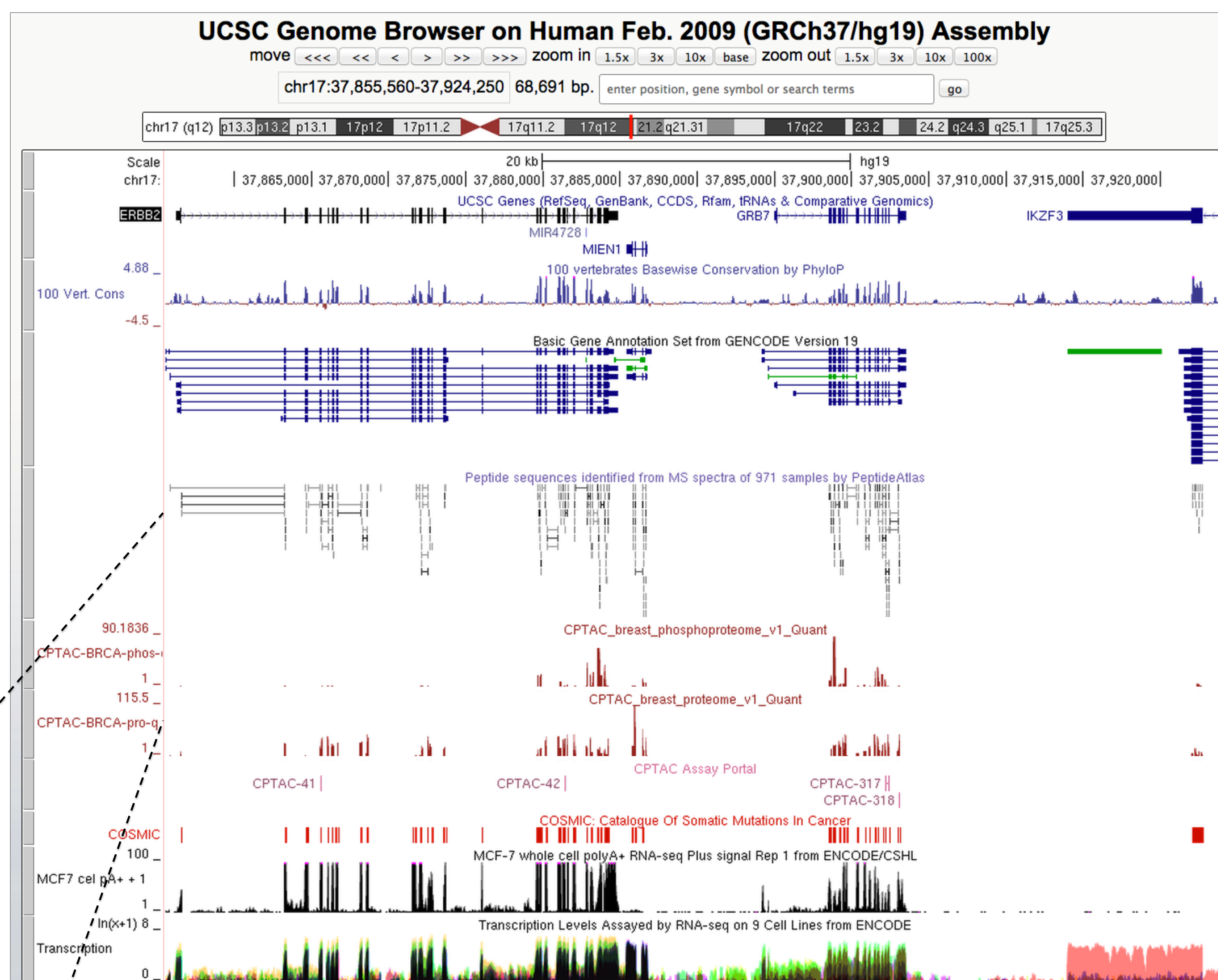David Fenyo — Center for Health Informatics and Bioinformatics, NYU School of Medicine, New York, NY, USA
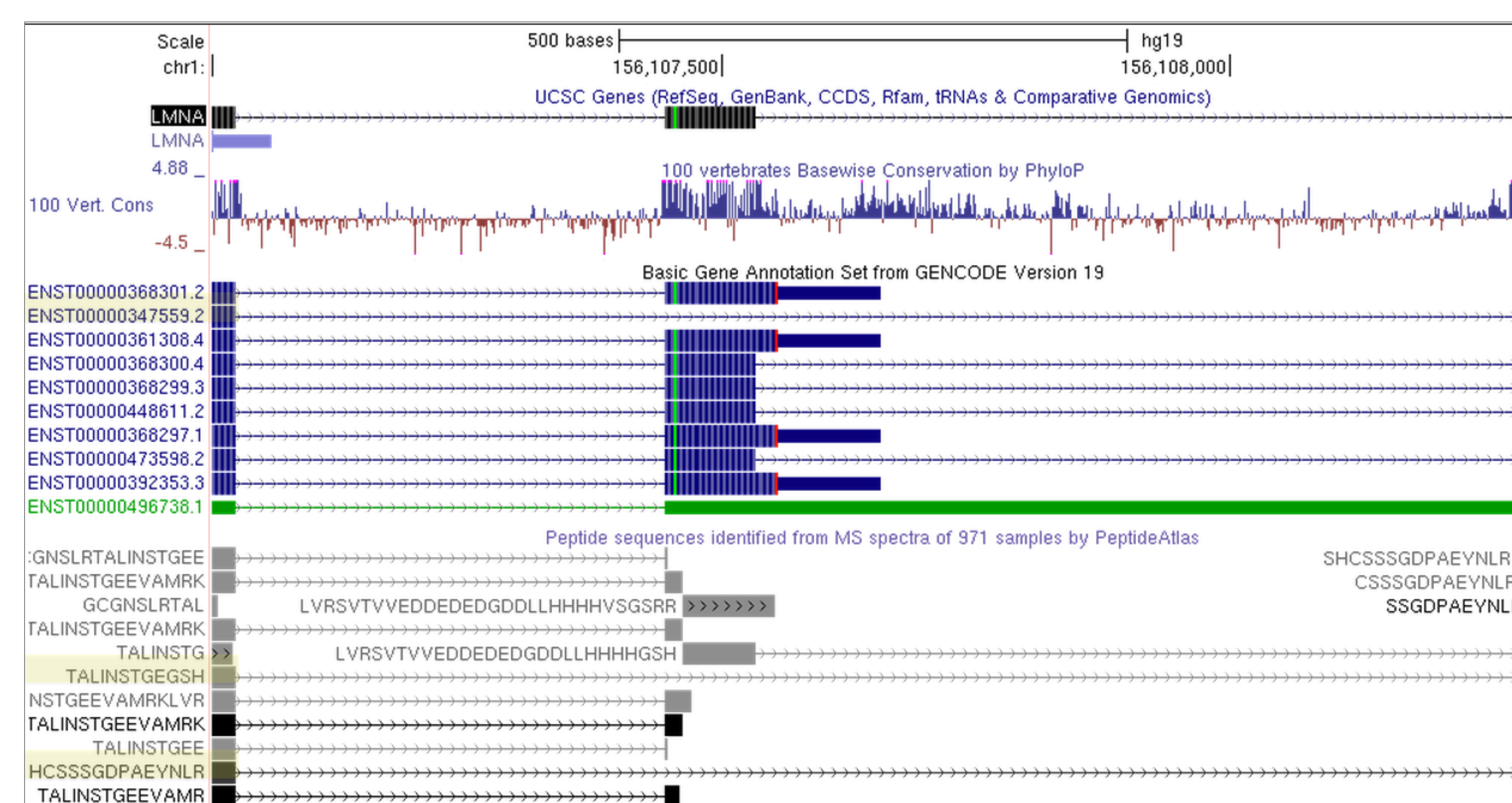
genome.ucsc.edu

Overview

As the comprehensiveness and quality of mass spectrometry peptide identifications improve, there is increasing interest from researchers outside of the proteomics community in accessing, visualizing, and analyzing proteomic datasets in genomic context with gene annotations, epigenetic profiles, genetic variation and clinical data anchored to the genome. Similarly, researchers with proteomics data seeking to integrate with other 'omics datasets can benefit from existing tools and databases that proide access to well-curated genome annotations. The UCSC Genome Browser and database provides a comprehensive resource of genome assemblies and annotations together with visualization and analysis tools. It has been in continuous use by scientists and students worldwide since its launch in the year 2001 to showcase the draft human genome assembly.

Two recent developments in the browser relevant to the proteomics community are the incorporation of the PeptideAtlas 2014 human build into the browser database and track display, and the CPTAC cancer proteomics data hub. These datasets are now publicly available for visualization, download, and data mining using UCSC tools at the UCSC Genome Browser website, http://genome.ucsc.edu.
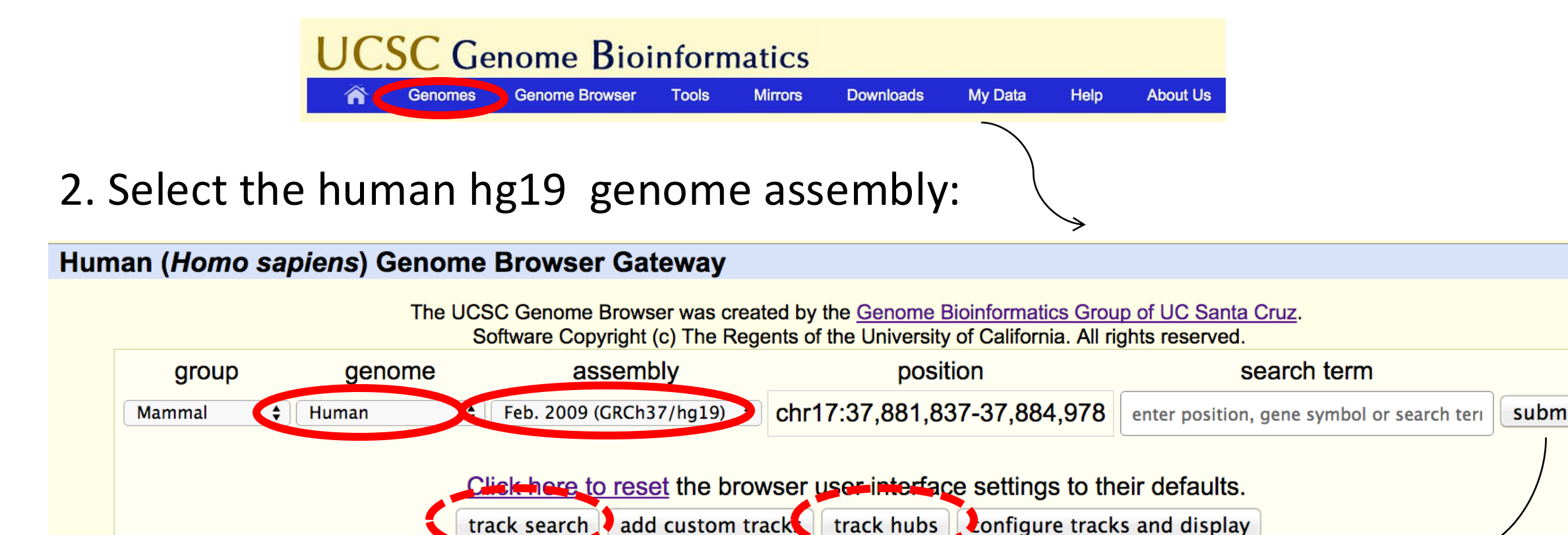
PeptideAtlas browser track

PeptideAtlas collects raw mass spectrometry proteomics datasets from laboratories around the world and reprocesses them in a uniform bioinformatics workflow using the Trans-Proteomic Pipeline . This track displays peptide identifications from the PeptideAtlas August 2014 (Build 433) Human build. This build, based on 971 samples containing 420,607,360 spectra, identified 1,021,823 distinct peptides, covering 15,136 canonical proteins.

http://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=peptideAtlas
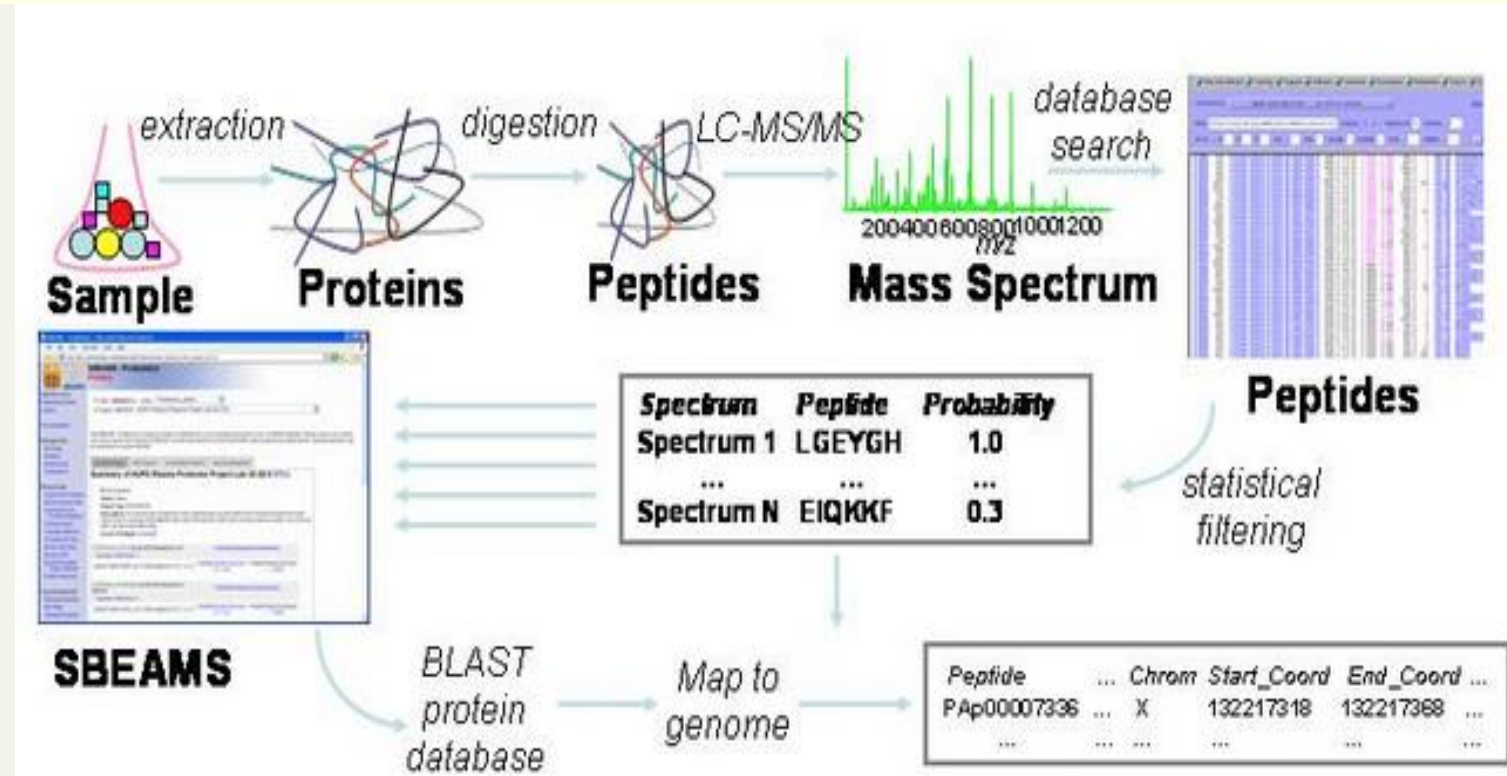
Cancer Proteomics (CPTAC) data hub

This track hub contains the peptides that were identified by Clinical Proteomic Tumor Analysis Consortium (CPTAC) in their deep mass spectrometry based characterization of the proteome content of breast, colorectal and ovarian biospecimens that were initially sequenced by The Cancer Genome Atlas. Please, address questions to info@fenyolab.org.

Global proteome measurements were performed on all tumor types to provide quantitative information on protein concentrations. For breast and ovarian tumors, quantitative measurements were also done on the phosphoproteome to provide more detailed information on cell signaling.
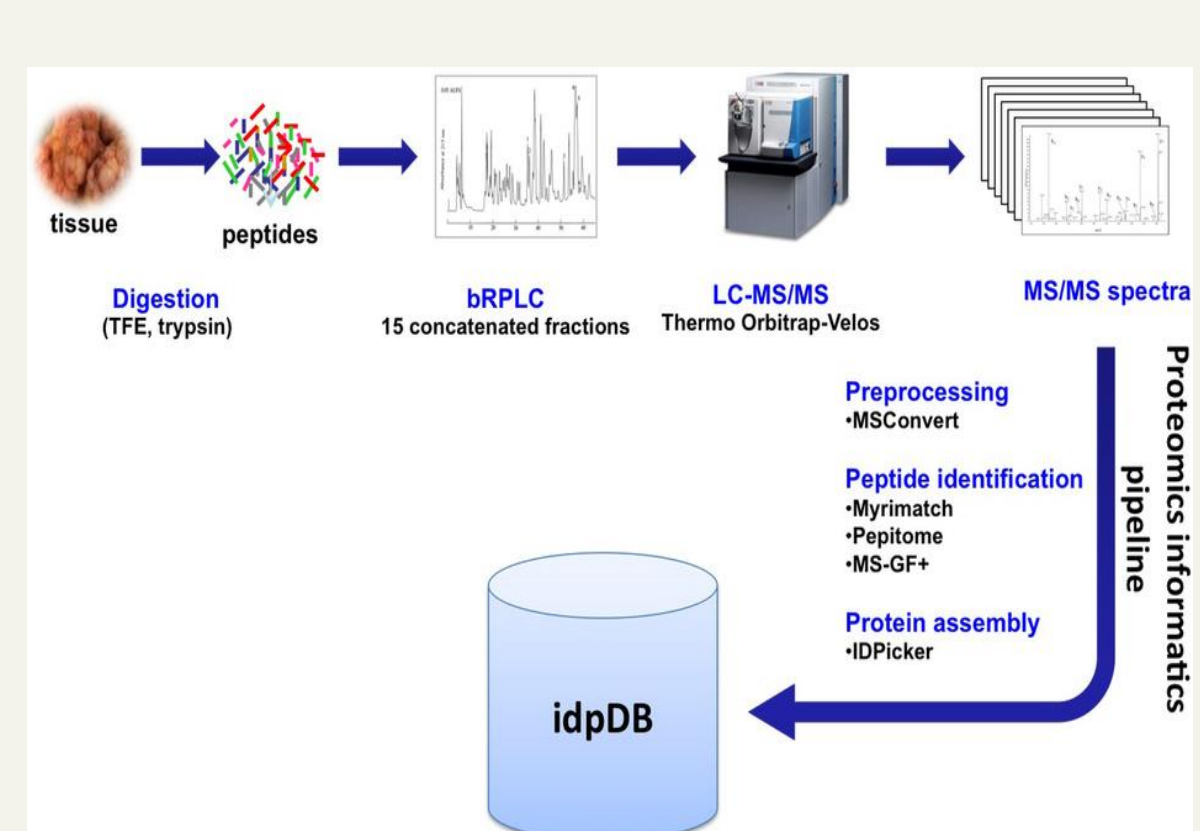
http://openslice.fenyolab.org/tracks/CPTAC/cptac/v1/hg19/CPTAC_v1.html

Display your own data in the browser

The UCSC browser provides multiple mechanisms for visualizing your own data in genomic context alongside datasets and annotations from scientists world-wide:

- Create a track data hub, and share with your collaborators
- Create a public track data hub, and share with the world
- Create confidential data tracks and hubs with privacy protection using a "Genome Browser in a Box" (GBiB)

Genome browser view of a 68kbp genomic region on human chromosome 17 where 3 cancer-associated genes (ERBB2, MIEN1, and GRB7) are colocated, along with a 4th gene (IKZF3; a tissue-specific (lymphocyte differentiation) transcription factor). Exons are denoted by tall blocks in the annotated gene tracks, with confirmatory evidence in the conservation signal produced using the UCSC 100-vertebrate genome sequence alignment, RNA-seq data from the ENCODE project, and PeptideAtlas peptide identifications. The cancer proteomics (CPTAC) tracks shown here show proteome and phosphoproteome signal in breast tumor samples from the TCGA (The Cancer Genome Atlas) project, and provide links to MRM assays. Clicking on an assay provides MS detail from CPTAC. Related tracks from the COSMIC cancer mutation database and ENCODE breast cancer cell line RNA-seq round out the view.
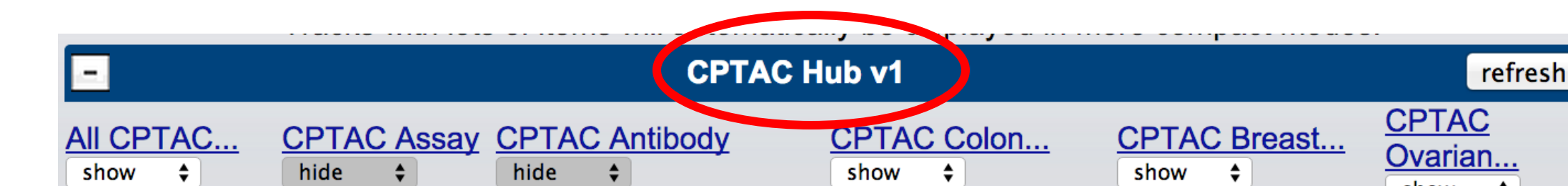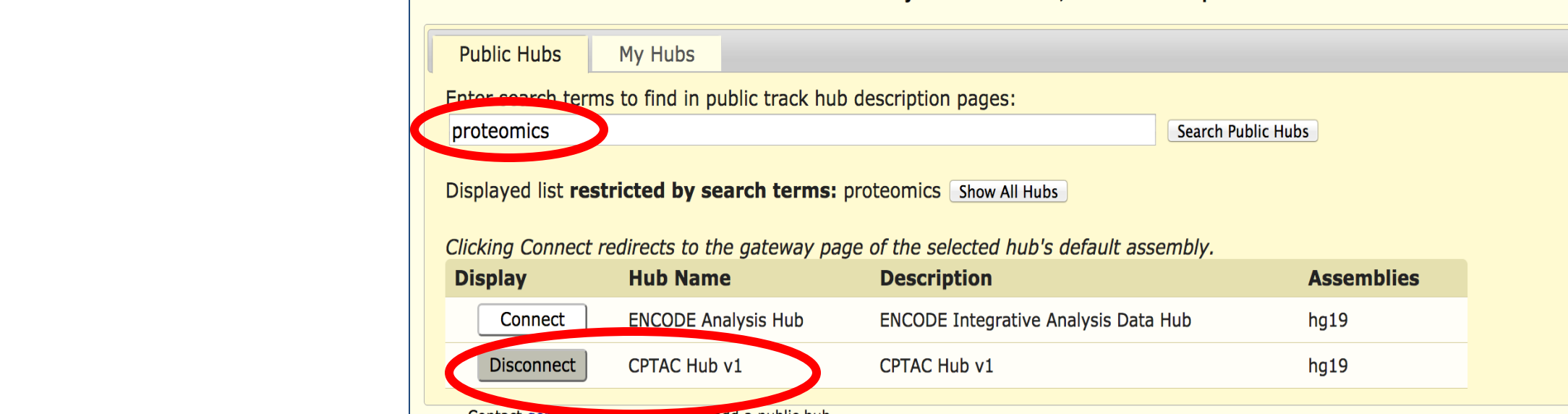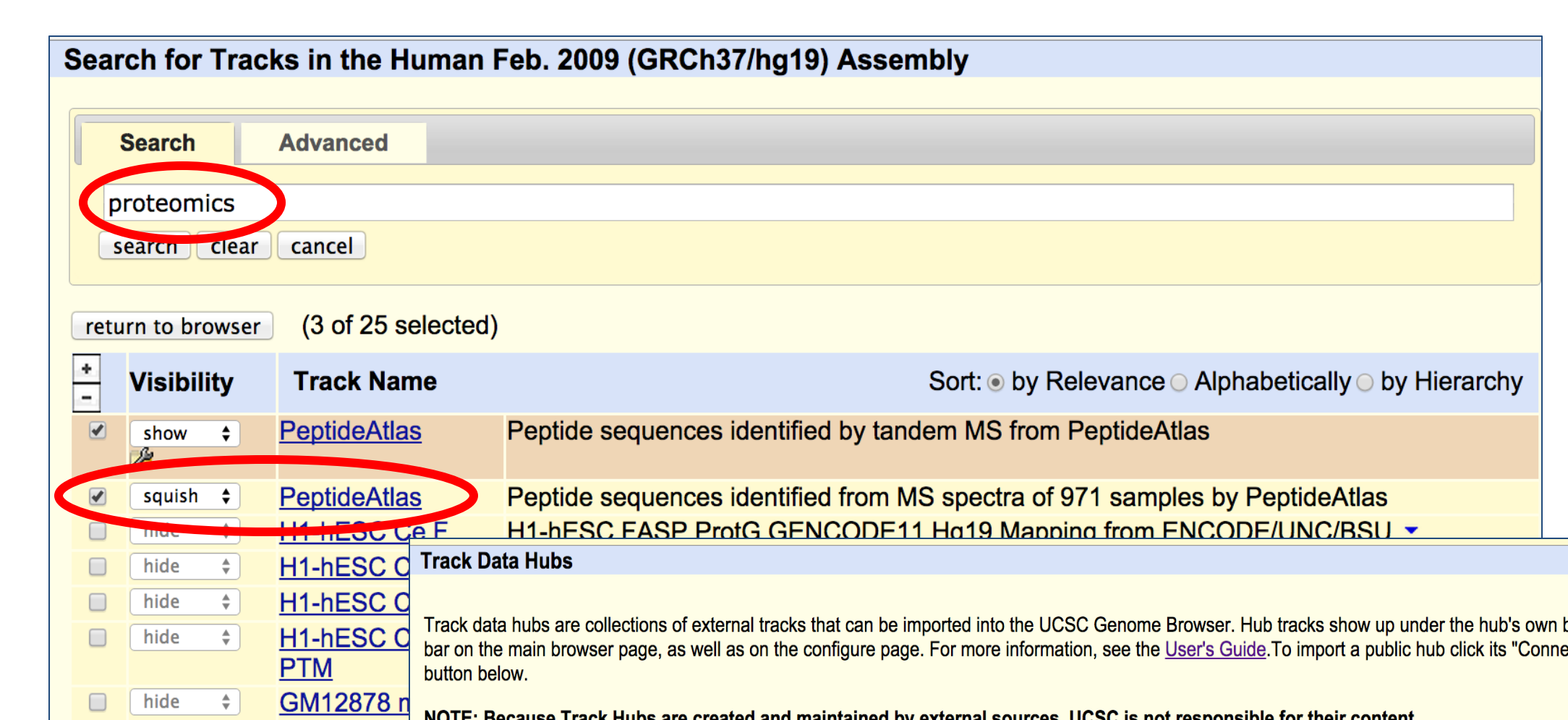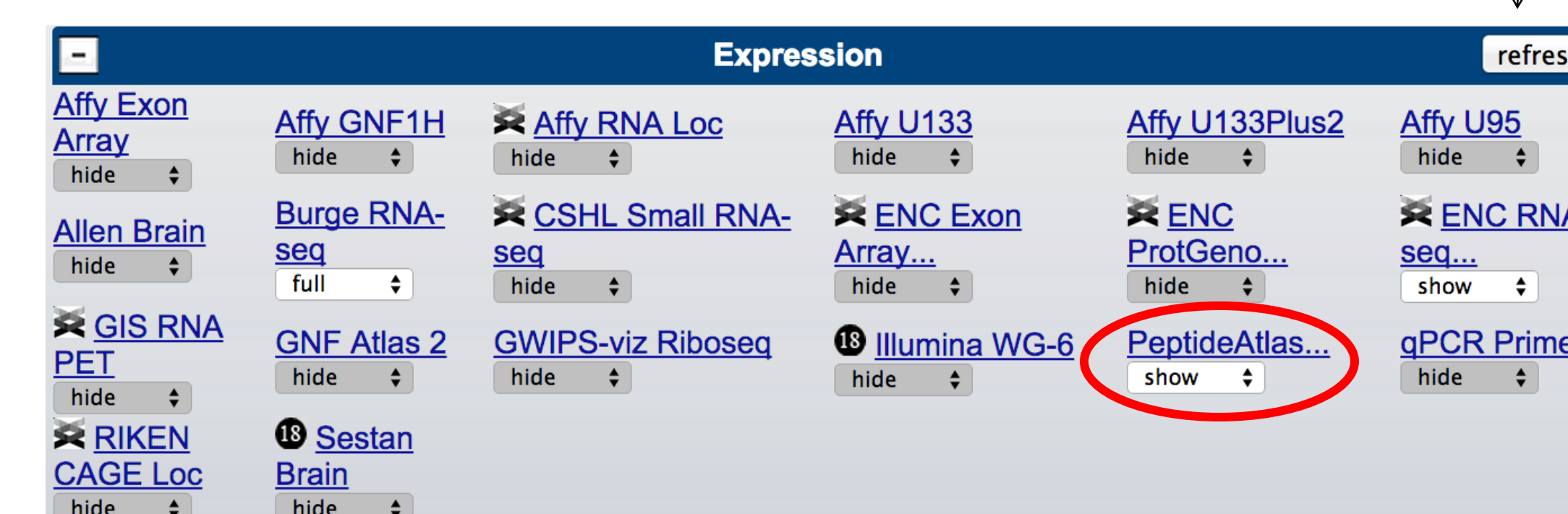

Genome browser view of a 1300bp genomic region on human chromosome 1 showing alternative splicing of the nuclear lamina protein, LMNA. Multiple isoforms annotated by GENCODE/Ensembl are shown together with PeptideAtlas peptide identifications (highlighted) that provide evidence for the exon skipping in transcript ENST00000347559. Clicking on a peptide in this track produces a detail page, with a link to MS and sample detail at PeptideAtlas.

Locate proteomics data

1. At genome.ucsc.edu, navigate to the genome selection page:

2. Select the human hg19 genome assembly:

3. Select tracks from the menu, or use the search features

Learn more

* Genome Browser help pages:
  http://genome.ucsc.edu/goldenPath/help

* Training information, videos, and user's guides:
  http://genome.ucsc.edu/training

* Recent reference:
  M Speir, AS Zweig, KR Rosenbloom, et. al.,
  The UCSC Genome Browser database: 2016 Update
  Submitted: Nucleic Acids Res (2016 Database Issue)
  BioRxiv preprint: http://dx.doi.org/10.1101/027037

* Find us on: ⓕ GenomeBrowser
               🐦 @GenomeBrowser
               ▶ UCSC Genome Browser

* To view a wiki page with this poster and additional information online, scan the QR code (above) or visit:
  http://genomewiki.cse.ucsc.edu/index.php/Hupo2015Poster

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute — genomics.ucsc.edu

Institute for Systems Biology — PeptideAtlas — www.peptideatlas.org

computational proteomics — fenyolab.org

NATIONAL CANCER INSTITUTE — Office of CANCER CLINICAL PROTEOMICS Research — proteomics.cancer.gov