UCSC Genome Browser

UNIVERSITY OF CALIFORNIA SANTA CRUZ | Genomics Institute

GTEx resources and new ‘barChart’ and ‘interaction’ track displays inspired by GTEx

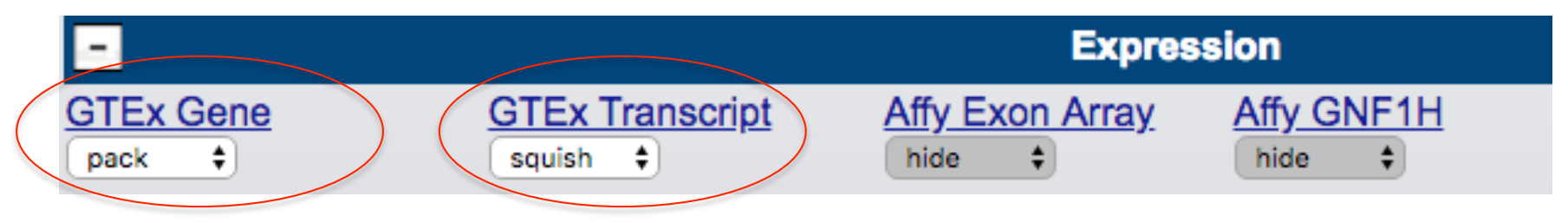
Kate Rosenbloom, Galt Barber, Jonathan Caspar, Hiram Clawson, Max Haeussler, Angie Hinrichs, Christopher Lee, Jairo Navarro,
Brian Raney, John Vivian, Brian Lee, Ann Zweig, David Haussler, Robert M Kuhn, and Jim Kent

UCSC Genome Browser Group, UC Santa Cruz Genomics Institute, University of California Santa Cruz

https://genome.ucsc.edu

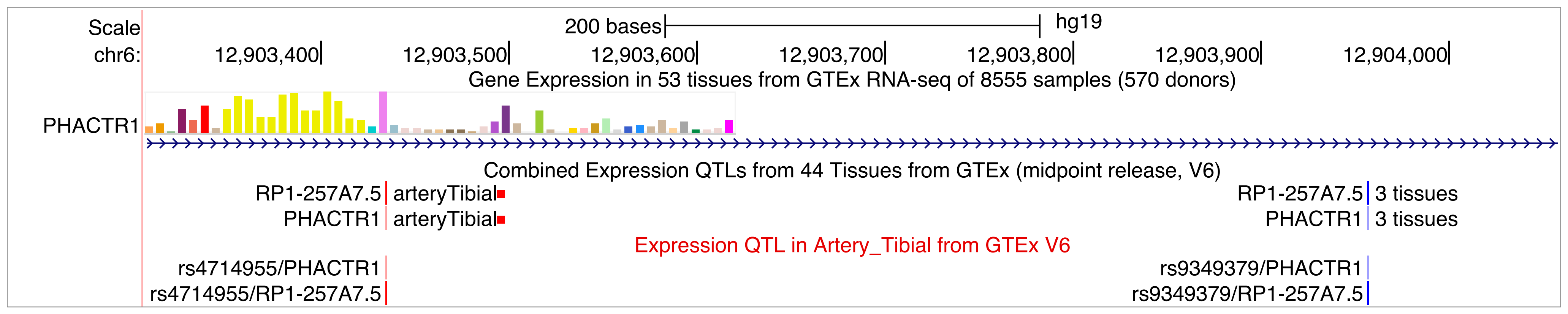
Inspired by the comprehensive and high-quality gene expression and variant datasets released by the *NIH Genotype-Tissue Expression (GTEx) projec*t, we have created new visualization and data mining resources in the UCSC Genome Browser based on the GTEx midpoint release (V6). Recognizing that the new displays developed for GTEx would be useful for user community display of their own data, we generalized the data management and display to create two new track types – *barChart* and *interact.*

GTEx tracks



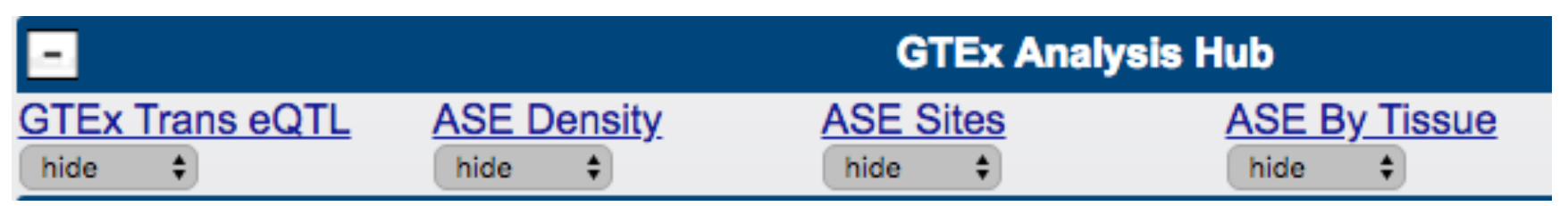
Gene and transcript expression in 53 tissues from GTEx RNA-seq of 8555 samples (570 donors)

Expression QTLs in 44 tissues from GTEx

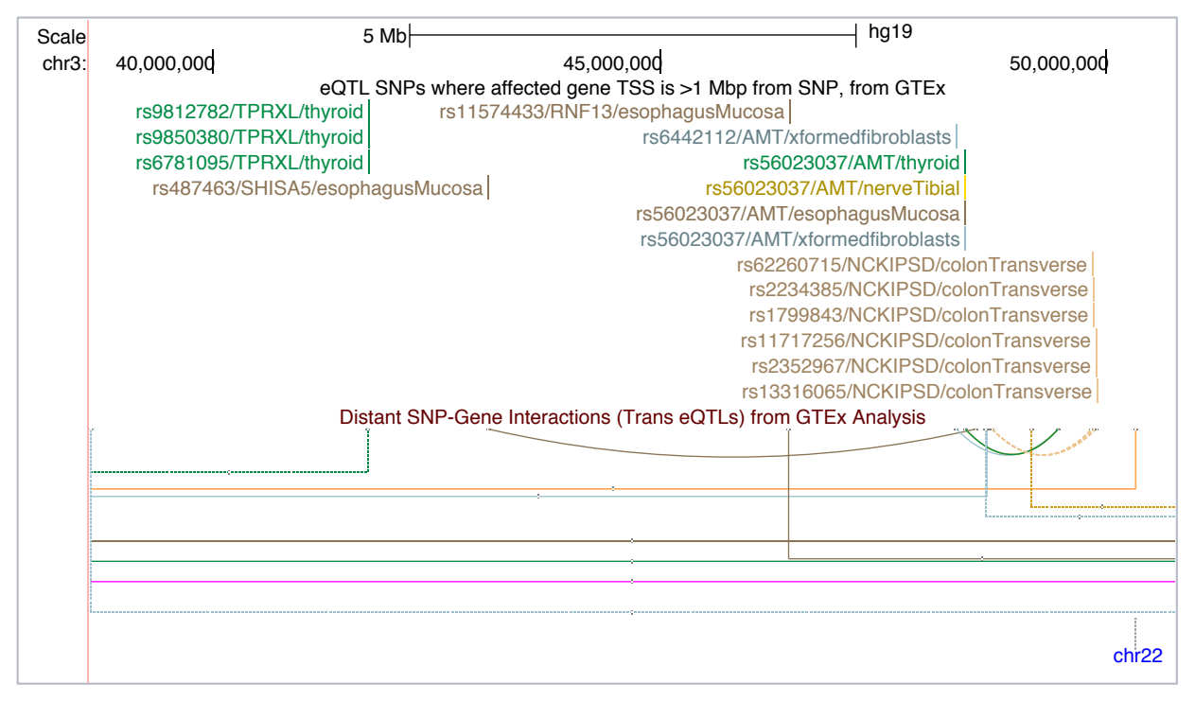


Intronic region of the PHACTR1 locus, where GTEx eQTL analysis reported in Nature October 2017 identifies two SNP’s affecting gene expression levels in artery -- one upregulating (colored red), and the other down-regulating (colored blue). The color intensity displayed reflects the effect strength.

Public GTEx track hubs

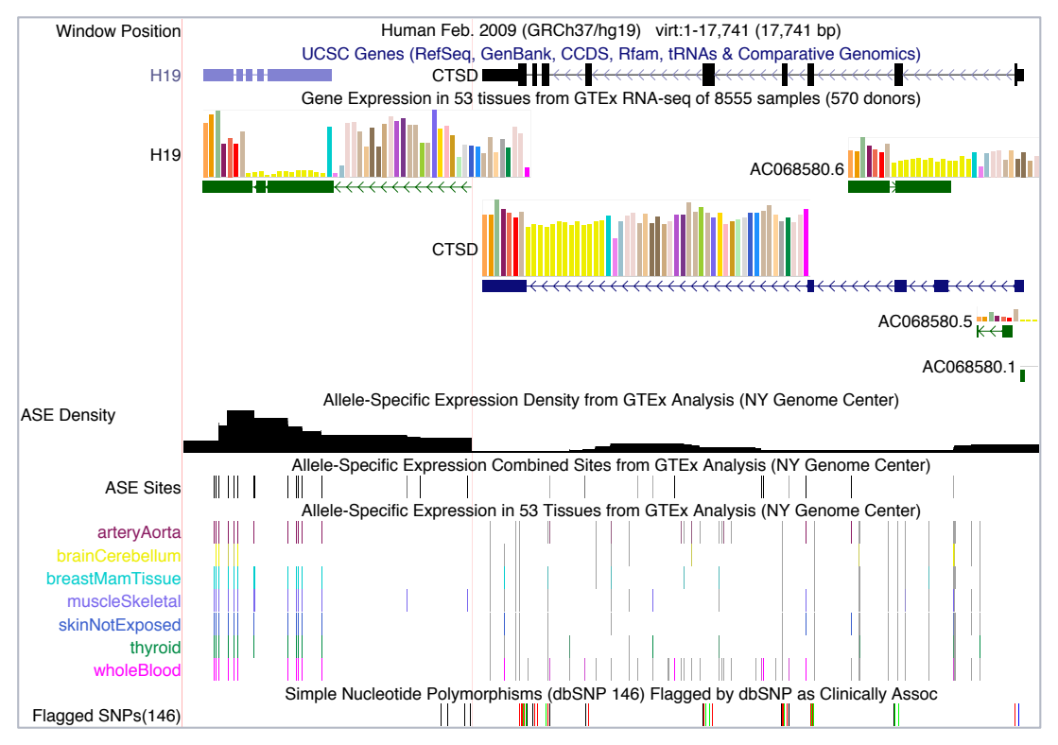


Distant SNP-gene interactions



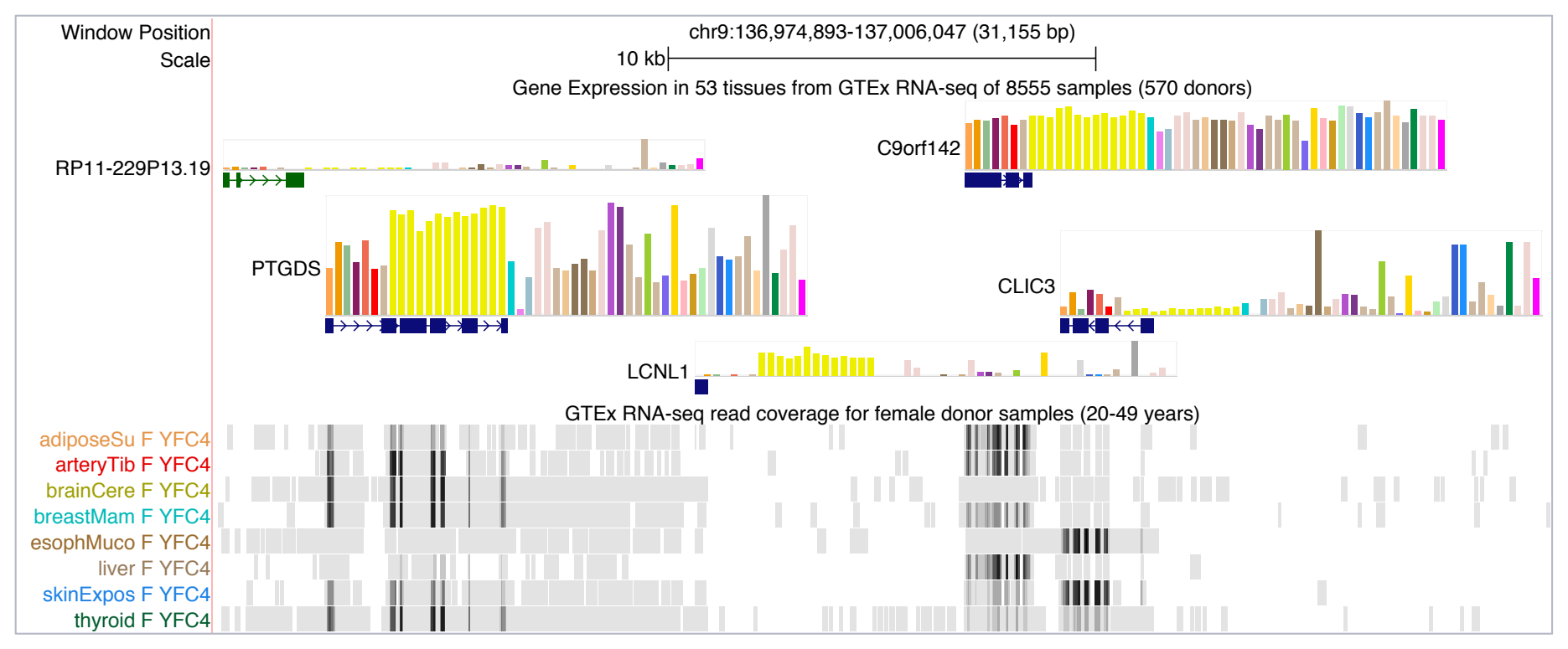
12 Mbp region where GTEx trans-eQTL analysis identifies one interchromosomal interaction, 4 with both SNP and gene in the region (curved display), and multiple longer interactions. These tracks, featuring the new paired region interaction display, are under development for release in spring 2018.

Allele-specific expression



ASE annotation from GTEx analysts at the NY Genome Center showing allelic imbalance at the H19 locus, a known imprinted gene.

RNA-seq read coverage from GTEx/UCSC Toil



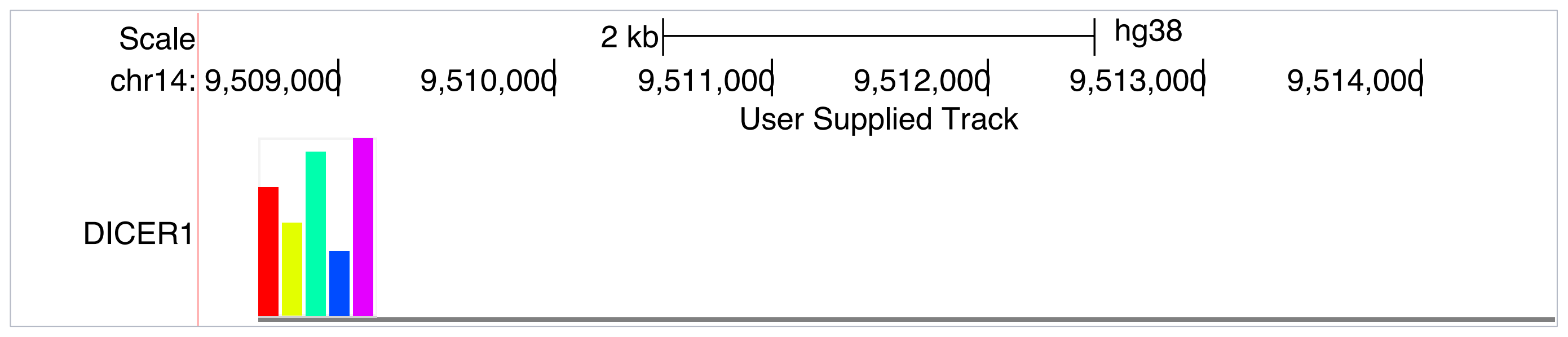
Genomic region with 6 transcriptionally active regions demonstrating tissue-specific expression, indicated by the gene-level annotation from the GTEx LDACC as well as the individual donor RNA-seq signal computed independently at UCSC, using the Toil pipeline.

barChart track type

The barChart track format displays a graph of category values over genomic regions. This format is useful for displaying gene expression and other datasets comparing a set of variables. The main display is designed to present summary data from large sample sets, An optional data matrix file with sample values will enable display of the data distribution as a boxplot on a details page.

```
track type=barChart barChartBars="blood lung liver heart muscle" visibility=pack

# chrom   start       end         name     score strand  name2             expcount   expscores
chr14     9508627     95158010    DICER1   900     -     ENSG000005100697  5          30.5,11.6,80.2,4.9,120.5
```



Acknowledgements

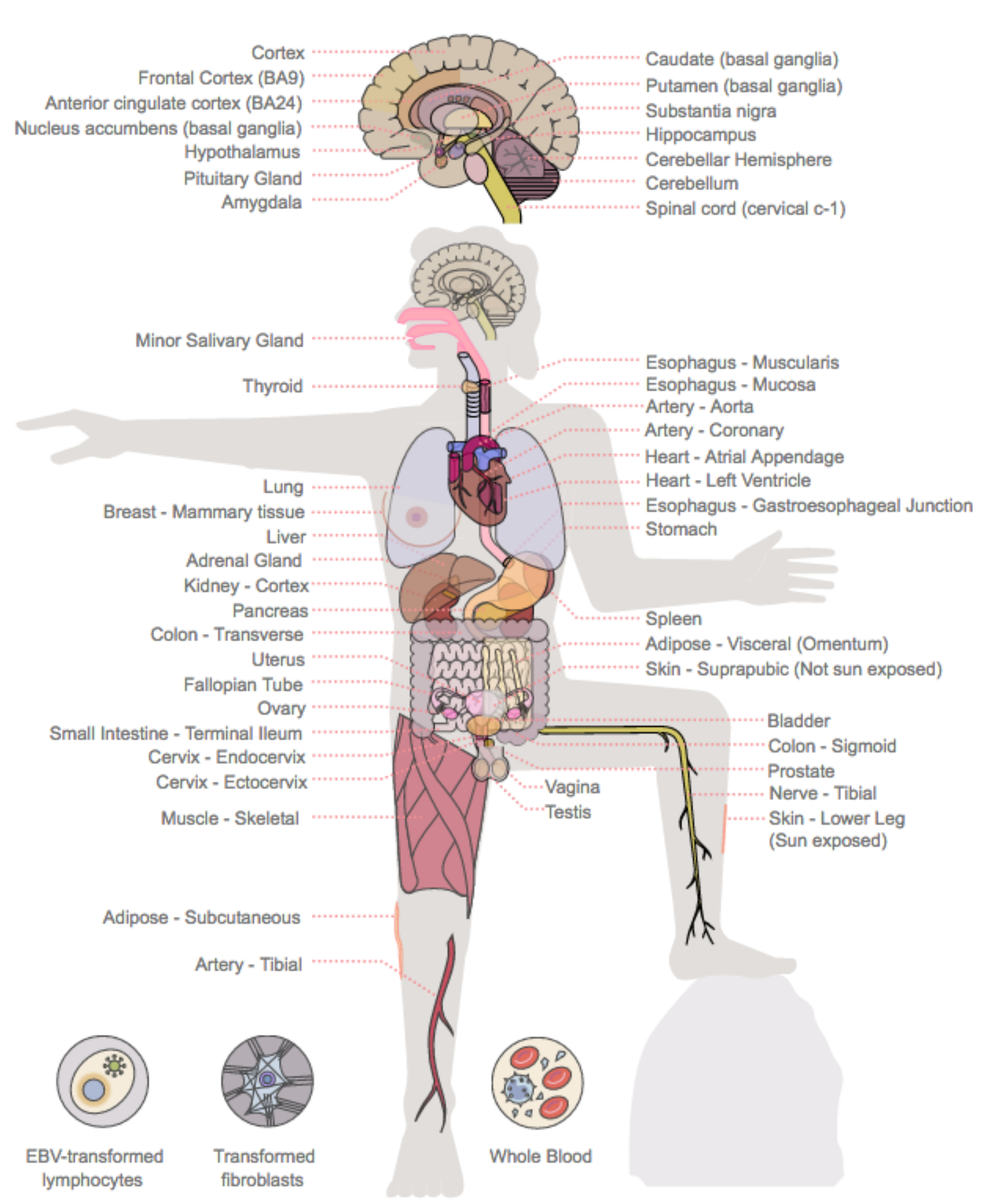
Thank you to the GTEx investigators, analysts, and the GTEx Laboratory, Data Analysis and Coordinating Center, for guidance during development of these resources. In particular, thanks to Tuuli Lappalainen and Stephane Castel (NY Genome Center), Christopher Brown (University of Pennsylvania), Farhad Hormozdiari (Harvard), Barbara Engelhart (Princeton), Alexis Battle (Johns Hopkins), and Ayelet Segre, Francois Aguet and Jared Nedzel (LDACC and GTEx portal teams, Broad Institute). This work was funded by NHGRI award 4U41HG002371 to the UCSC Center for Genomic Science.

GTEx Body Map

Tissue colors


- ■ Adipose - Subcutaneous
- ■ Adipose - Visceral (Omentum)
- ■ Adrenal Gland
- ■ Artery - Aorta
- ■ Artery - Coronary
- ■ Artery - Tibial
- ■ Bladder
- ■ Brain - Amygdala
- ■ Brain - Anterior cingulate cortex (BA24)
- ■ Brain - Caudate (basal ganglia)
- ■ Brain - Cerebellar Hemisphere
- ■ Brain - Cerebellum
- ■ Brain - Cortex
- ■ Brain - Frontal Cortex (BA9)
- ■ Brain - Hippocampus
- ■ Brain - Hypothalamus
- ■ Brain - Nucleus accumbens (basal ganglia)
- ■ Brain - Putamen (basal ganglia)
- ■ Brain - Spinal cord (cervical c-1)
- ■ Brain - Substantia nigra
- ■ Breast - Mammary Tissue
- ■ Cells - EBV-transformed lymphocytes
- ■ Cells - Transformed fibroblasts
- ■ Cervix - Ectocervix
- ■ Cervix - Endocervix
- ■ Colon - Sigmoid
- ■ Colon - Transverse
- ■ Esophagus - Gastroesophageal Juncti
- ■ Esophagus - Mucosa
- ■ Esophagus - Muscularis
- ■ Fallopian Tube
- ■ Heart - Atrial Appendage
- ■ Heart - Left Ventricle
- ■ Kidney - Cortex
- ■ Liver
- ■ Lung
- ■ Minor Salivary Gland
- ■ Muscle - Skeletal
- ■ Nerve - Tibial
- ■ Ovary
- ■ Pancreas
- ■ Pituitary
- ■ Prostate
- ■ Skin - Not Sun Exposed (Suprapubic)
- ■ Skin - Sun Exposed (Lower leg)
- ■ Small Intestine - Terminal Ileum
- ■ Spleen
- ■ Stomach
- ■ Testis
- ■ Thyroid
- ■ Uterus
- ■ Vagina
- ■ Whole Blood

Genotype Tissue Expression (GTEx) Project Body Map



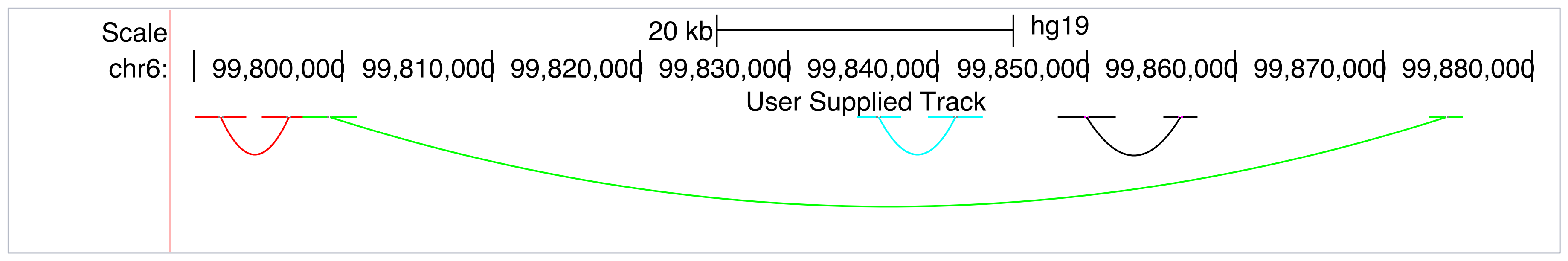
Credit: jwestdesign https://genome.ucsc.edu/gtexBodyMap.html

As a companion to the GTEx data resources, we developed an interactive human anatomy graphic of the 52 GTEx tissues. This graphic provides a user-friendly tissue legend for GTEx . It is available for tissue selection on the configuration page of several GTEx tracks, and will be released as an independent page available for linking by outside resources.

Interact track type

The interact track type was designed to display functional element interactions (e.g. SNP-Gene), but is also suitable for low-density chromatin interaction data (e.g. ChiaPet)

```
track type=interact visibility=full
# Interaction                                                         Source                                              Target
# chrom start     end       name    score value  exp     color      chrom  start     end       name strand  chrom  start     end       name   strand
chr6    99790007  99798261  inter1   233   4.0   brain   255,0.0    chr6   99790007  99793559  rs1    .     chr6   99794572  99798261  gene1   +
chr6    99797231  99875440  inter2   175   3.2   blood   0,255,0    chr6   99797231  99801068  rs2    .     chr6   99873105  99875440  gene2   +
chr6    99834530  99843149  inter3   175   3.1   liver   0,0,255    chr6   99834530  99837680  rs3    .     chr6   99839323  99843149  gene3   -
chr6    99848067  99857519  inter4   175   3.5   kidney  0          chr6   99848067  99852034  rs4    .     chr6   99855098  99857519  gene4   +
```



Interactions on the same chromosome are displayed as arcs or half-rectangles connecting the interaction endpoints (arcs are drawn if both endpoints are in the display window). The height of the arc/rectangle is based on the length of the interaction. Interchromosomal interactions are indicated by a single endpoint with a vertical extension.

More information

Genome Browser public mailing list: genome@soe.ucsc.edu

Genome Browser tutorials and videos:
https://genome.ucsc.edu/training/index.html