

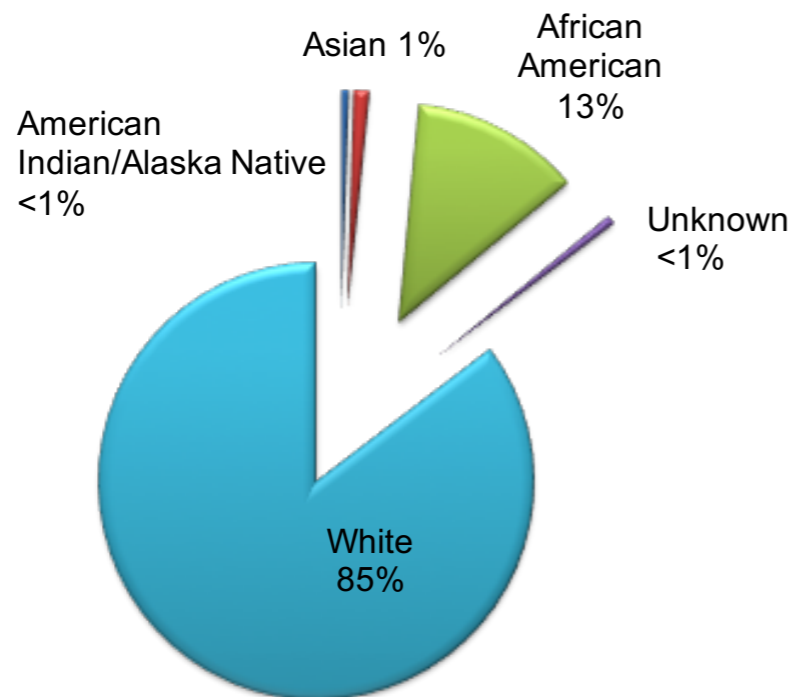
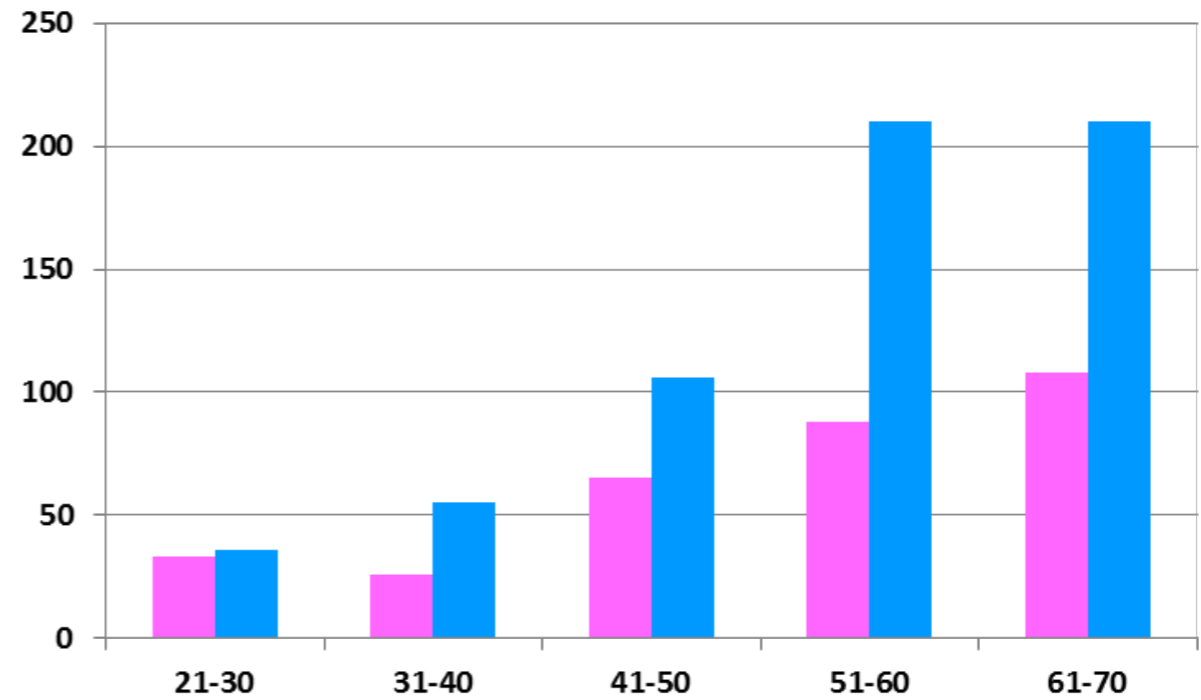
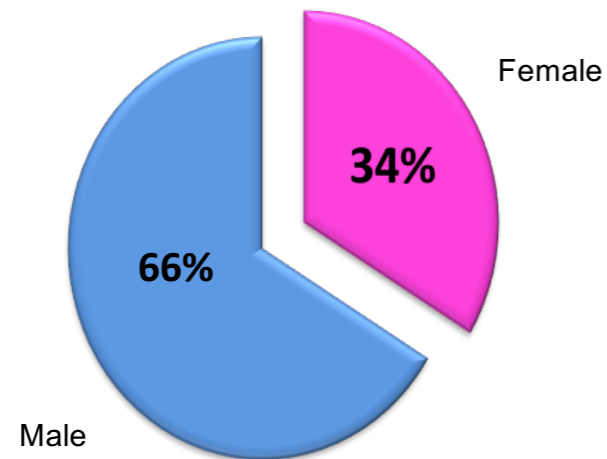
2016 GTEx mtg Notes

2016-07-11

Project overview

- **960** donors: *up to 53* tissue per donor / **25** tissue per donor *in average*
- DNA-seq **each donor** (WGS and WES, 30X and 100X, respectively)
- RNA-seq on all tissues: > **25K**
- Clinical / histopathological information for each donor / tissue
- Enhanced GTEx (*eGTEx*):
 - Protein quantifications (x2)
 - Methylation (x2)
 - Histone modifications
 - DNase-seq
 - mmPCR-seq (deep ASE)
 - Somatic DNA seq (deep exome seq)
 - Analysis of telomere structure
- *Single-cell projects funded* and ongoing with GTEx bank samples (within next year): J Eberwine (UPenn), K Zhang (UCSD), A Regev (Broad)
- GTEx collections for ENCODE, *ENTEx*: 4 samples, all tissues but brain
- *The GTEx project is ending next year, no plans to continue in any direction.*
There is probably going to be one last community meeting next year

Donor Demographics



- Figures from S Volpi's (NHGRI) "NIH overview" slides

Data production update: current snapshot

- Donor collection *complete*: **960** donors, **426** of which are brain donors*
 - *~**200** of them will be available as part of V7 + previous releases
- Tissue processing (~**880**)
- RNA-seq (**16K**)
- V6 (latest release): **450** donors, ~**7.4K** RNA-seq*
 - *numbers from GTEx portal: **524** subjects w/ SNP-chip, **148** w/ WGS, **450** w/ RNA-seq, **8.5K** RNA-seq experiments
- Midpoint AWG *manuscripts* in prep: **Dec. 2016**
- V7 is “*currently being released*” (late summer / early Fall): **635** donors / ~**13K** tissues
- Raw data (WES, WGS, RNA-seq) — dbGaP
- eQTL — GTEx portal
- (According to GTEx portal:) V5, V6 releases are under 8-27-14 NIH GDS Policy: “*once data is released, there are no restrictions on use or publication*”

V7

- Changes in *V7* compared to *V6*:
 - Genotyping: microarrays → WGS / WES
 - RNA seq alignment: Tophat → STAR
 - Gene expression: new collapsed gene model
 - Isoform quantification: FluxCapacitor → RSEM
 - eQTL discover: MatrixEQTL → FastQTL
- Core data:
 - Expression: read counts for genes, transcripts, etc; normalized expression for genes, transcripts; coverage tracks
 - eQTL; gene-level summary; significant variant-gene pairs; all variant-gene pairs; expression matrices; covariates
- Additional core data:
 - Splicing QTLs
 - ASE
 - Multi-tissue eQTL

V6p & V8

- *V6p*
 - Released in ~2 weeks on GTEx portal
 - Updated GENCODE (v19) annotations
 - eQTLs: FastQTL (instead of Matrix eQTL)
 - Includes eQTLs on chr. X

- *V8* (likely, to be released in early 2017) planned changes:
 - Shift to hg28/GRCh38 and latest GENCODE release: realignments / quantification
 - Re-evaluation of isoform quantification methods
 - Small RNA-seq pipeline: “we will be doing small-RNA sequencing on all samples, and will be releasing the data on all samples as well Though we’re still fine-tuning the methods”, K Arlie
 - Pipelines will be made available via FireCloud and Docker images

Key contacts (**DC spoke w/individual in person**):

Francois Aguet (LDACC Broad)

Kristin Ardlie

Max Haeussler (UCSC; Genome Browser Engineer)

Su Koester (NIMH)

Jared Nedzel (GTEx, Broad)

Kate Rosenbloom (UCSC)

Ayellet Segre (LDACC Broad)

Cassandra Trowbridge (LDACC Broad)

Simona Volpis (NHGRI)

Daniel Zerbino (ensembl)

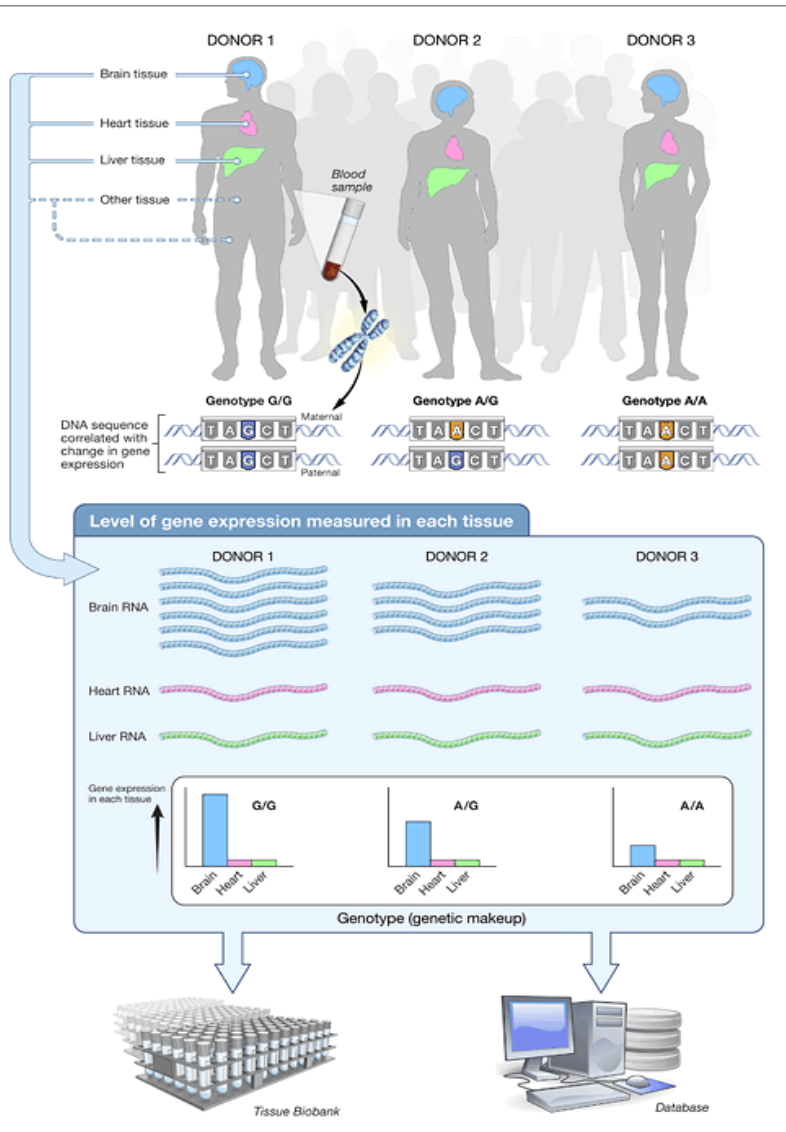
3rd GTEx Community Meeting
JULY 11TH, 2016
Stanford University

GTEx PROJECT
COMMUNITY
MEETING



GTE_x = Genotype-Tissue Expression

NIH Common Fund (commonfund.nih.gov/gtex)



GTE_x GOAL:

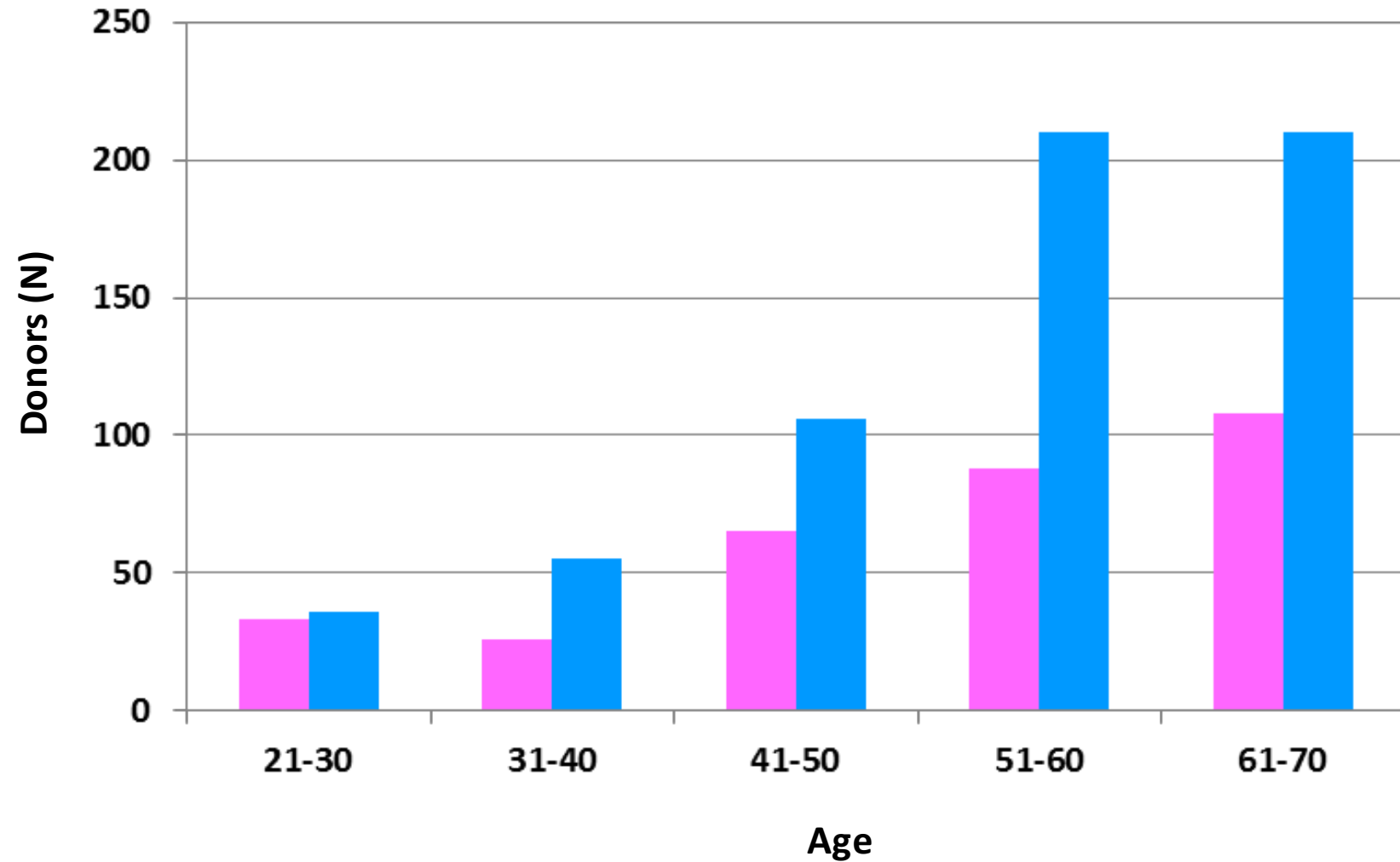
- to help unravel the complex interplay between genetic variation and gene expression across a wide range of non-diseased human tissues.
 - Atlas of gene expression & eQTLs
 - Biobank of tissues, DNA, RNA

by end 2017:

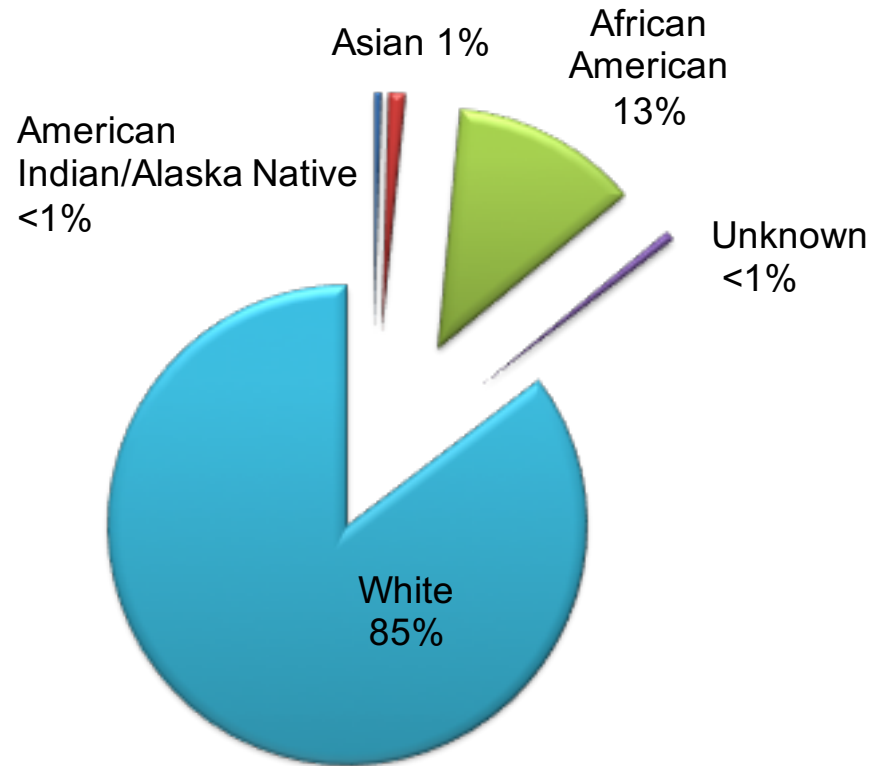
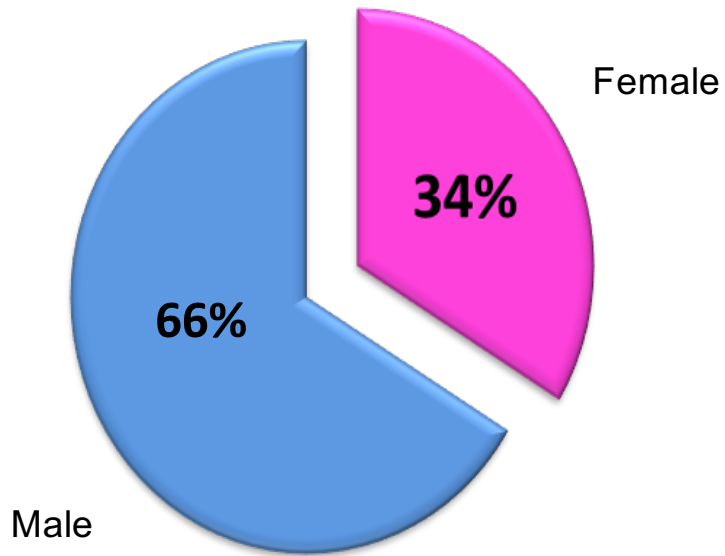
- ~960 Postmortem Donors
- WES & WGS
- RNA-Seq of ~30 tissues/donor (>20,000 tissues)
- Beyond Gene Exp



Donor Demographics



Donors - Sex and Race



Cause of Death

Cause of Death	21 - 40	41 - 70
Stroke	18%	28%
MI	5%	33%
Trauma - Blunt Injury, MVA, Falls, GSW	42%	8%

Death classification

Hardy Scale	overall	
0	53%	← Low PMI
1	4%	} Brain donors
2	25%	
3	6%	
4	12%	

0. Ventilator Case

1. Violent and fast death Deaths due to accident etc; terminal phase (TP) < 10 min
2. Fast death of natural causes; sudden unexpected deaths; TP < 1 hr
3. Intermediate death; $1 < TP < 24$ hrs; patients who were ill but death was unexpected
4. Slow death; TP > 1 day, deaths that are not unexpected

GTEEx resources

- GTEEx Portal: an open access database of GTEEx summary data: <http://www.gtexportal.org>

The screenshot shows the GTEEx Portal website. At the top is the logo "GTEEx Portal" with a DNA helix icon. Below the logo is a navigation bar with links: "GTEEx", "Datasets", "Gene Association", "eQTL Browser", "Biobank", "Documentation", and "Contact". There is also a search bar for "Search Gene or SNP ID..." and "Login" and "Register" buttons. A banner below the navigation bar features a background image of a person and a DNA helix, with text: "2016-03-16 Registration Open for 2016 GTEEx Community Meeting Read More >>". Below the banner are three main content sections: "Current Release" with "Latest Version: V6 dbGaP Accession phs000424.v6.p1" and a "Dataset Summary Statistics Report" with a bar chart; "Genetic Association" with "Single Tissue eQTLs" and a search bar for "Search eQTL by gene or SNP ID", and "eQTL IGV Browser" with a genomic track visualization; and "Transcriptome" with a search bar for "Search expression by gene ID...", "Top 100 Expressed Genes in a Tissue (e.g. Blood)", and "Gene Expression in Tissues" with a bar chart.

Jared Nedzel for the GTEEx portal
Daniel Zerbino for ENSEMBLE

GTEx resources cont.

dbGaP: controlled access of comprehensive GTEx clinical and raw sequencing data:

<http://www.ncbi.nlm.nih.gov/gap>



Common Fund (CF) Genotype-Tissue Expression Project (GTEx)

dbGaP Study Accession: phs000424.v6.p1

▸ [Study version history](#)

[Show BioProject list](#)

[Study](#) [Variables](#) [Documents](#) [Analyses](#) [Datasets](#) [Molecular Data](#)

Jump to: [Authorized Access](#) | [Attribution](#) | [Authorized Requests](#)

Study Description

Lay Description

The aim of the Genotype-Tissue Expression (GTEx) Project is to increase our understanding of how changes in

Important Links and Information

- Request access via [Authorized Access](#)
 - [Instructions](#) for requestors

Search Within This Study

Search for:

Sample Access

<http://www.gtexportal.org/home/samplesPage>

Home Analysis Datasets **Samples** Documentation News Help

Search Gene Expression

Gene Id...



Search eQTLs

Gene or SNP Id...



eQTL Genome Browser

Gene or SNP Id...



Latest Release

[V4 \(dbGaP phs000424.v4.p1\) >>](#)

GTEX Sample Request Forms

Download these documents and email completed requests to: nhgrigtex@mail.nih.gov. A Material Transfer Agreement (MTA) is required and needs to be in place before delivery of samples. Please do not copy any other email addresses with your submission.

Description	Form	Version
GTEX Biospecimen Access Requests	GTEX Biospecimens Access Requests 2015_05_07.docx	20150507
GTEX Biospecimens Access Policy	GTEX Biospecimens Access Policy 2015_05_07.docx	20150507
GTEX Material Transfer Agreement	GTEX NIH MTA v20150317.docx	20150317

Sample Search

Sorry, only logged in users have full access including the ability to search the samples.

Available Biospecimens

- PAXgene fixed, frozen tissue; PAXgene Fixed, Paraffin Embedded Tissue; RNA; DNA
- Flash frozen brain
- Lymphoblastoid and fibroblast cell lines

Sample Access cont.

<https://specimens.cancer.gov/>



A screenshot of the Specimen Resource Locator website homepage. The page has a dark blue header with navigation links: Home, Search, Biospecimen Resources, Information Resources, FAQ, Updates, Contact, and Login. Below the header is a large banner with the text "Specimen Resource Locator" and "A SERVICE OF THE NATIONAL CANCER INSTITUTE". The main content area is light gray and contains a section titled "About the Specimen Resource Locator" with two paragraphs of text and a green button that says "Click here to start searching". To the right of this section is a dark gray box titled "Quick Links" containing four buttons: "Search Resources", "How to Add a Collection", "Contact the Expediter", and "Follow us on Twitter".

The Specimen Resource Locator (SRL) is a biospecimen resource database designed to help researchers locate resources that may have samples needed for their investigational use.

The specimens come from non-commercial, either NCI or non-NCI-funded resources.

GTEx biospecimen collections SOPs

<http://biospecimens.cancer.gov/resources/sops/library.asp>

The screenshot shows the National Cancer Institute (NCI) website for the Biorepositories and Biospecimen Research Branch (BBRB). The page is titled "GTEx Standard Operating Procedures Library" and is part of the "Public Resources" section. It includes a navigation menu with options like Home, About BBRB, Programs, Best Practices, News and Events, Public Resources, and Patient Corner. The main content area lists three SOPs under the heading "A. Enrollment and Informed Consent":

1. BBRB-PM-0003-F1 GTEx Informed Consent Verification, Site 1
2. BBRB-PM-0003-F2 GTEx Informed Consent Verification, Site 2
3. BBRB-PM-0003-F4 GTEx Donor Eligibility Criteria Form

The page also features a search bar, a sidebar with additional resources, and a "Standard Operating Procedures (SOPs)" section with links to Introduction, The NIH GTEx Project, Reasons Behind caHUB SOP Release, and Important Notes on SOPs. The page is last updated on 07/17/15.

SOPs cover various operations including ethical and regulatory practices, biospecimen collections, data collection, shipping kits and checklists, and pathology review.

GTEx histological image viewer

http://biospecimens.cancer.gov/resources/tissue_image_library.asp

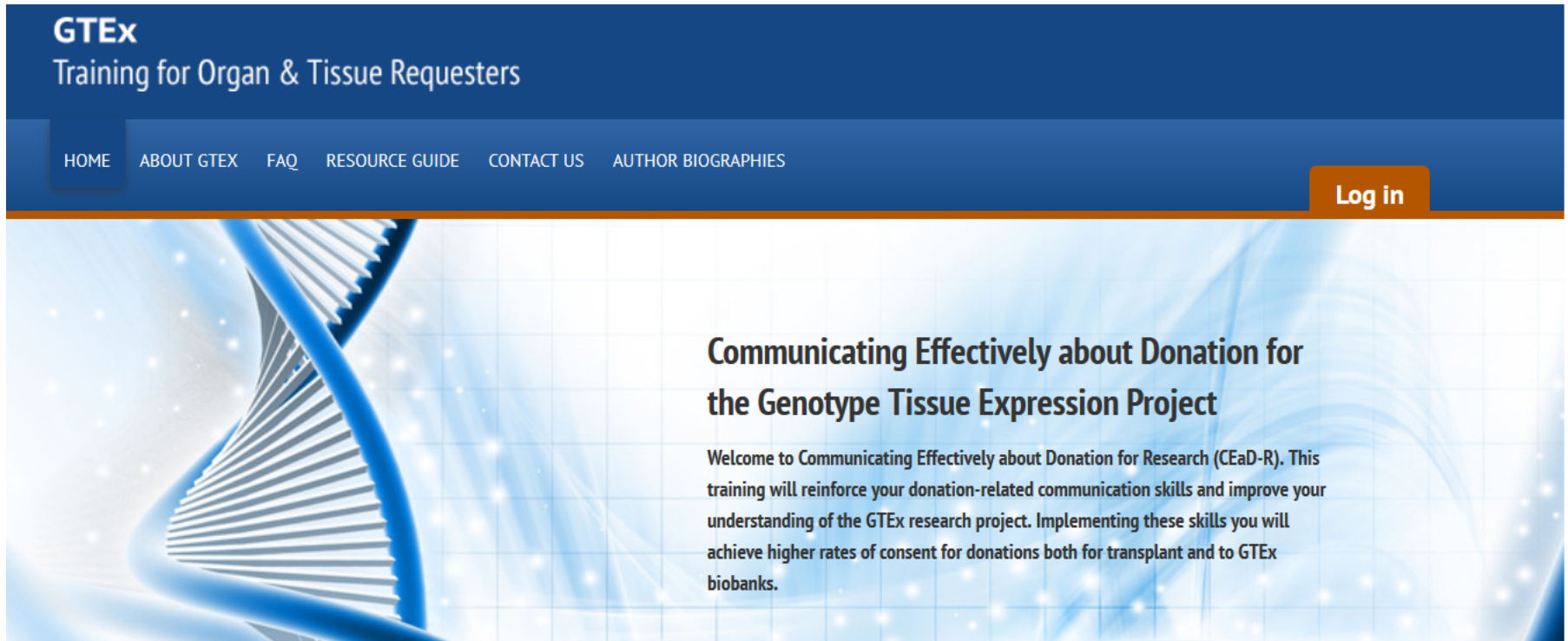


The screenshot shows the National Cancer Institute's GTEx Histological Images search interface. At the top, the National Cancer Institute logo and name are on the left, and the text "at the National Institutes of Health | www.cancer.gov" is on the right. Below this is a navigation bar with four main sections: BRD (Biospecimen Research Database), BBRB (Biorepositories and Biospecimen Research Branch), CDP (Cancer Diagnosis Program), and DCTD (Division of Cancer Treatment and Diagnosis). A "? help" link is also present. Below the navigation bar is a search bar with the text "GTEx Histological Images" and a "Search" label. The search bar contains a text input field and a "Go" button. Below the search bar is a "Hints" section with the text: "Please see [Terms](#) tab for list of tissues, and other fields for searching. See the online [help](#) for additional information on searching and using the Image Viewer." Below the hints is a section titled "All Field Search" with a bullet point: "■ Entering text without a specific field (see below), will search all fields for the words entered, i.e. *cortex* will find both kidney – cortex and brain – cortex from tissue".

There are numerous search options for a specific field search (tissue type, autolysis score, gender, acceptability, etc). No software is required, and the images can be viewed with zooming capability.

Training videos for consenting personnel

<http://gtextraining.org/>



The screenshot shows the top portion of the GTEx website. The header is dark blue with the text "GTEx Training for Organ & Tissue Requesters" in white. Below the header is a navigation bar with links: "HOME", "ABOUT GTEx", "FAQ", "RESOURCE GUIDE", "CONTACT US", and "AUTHOR BIOGRAPHIES". A "Log in" button is located on the right side of the navigation bar. The main content area features a large blue graphic of a DNA double helix on the left. To the right of the graphic, the text reads: "Communicating Effectively about Donation for the Genotype Tissue Expression Project". Below this title, a paragraph states: "Welcome to Communicating Effectively about Donation for Research (CEaD-R). This training will reinforce your donation-related communication skills and improve your understanding of the GTEx research project. Implementing these skills you will achieve higher rates of consent for donations both for transplant and to GTEx biobanks."

GTEx
Training for Organ & Tissue Requesters

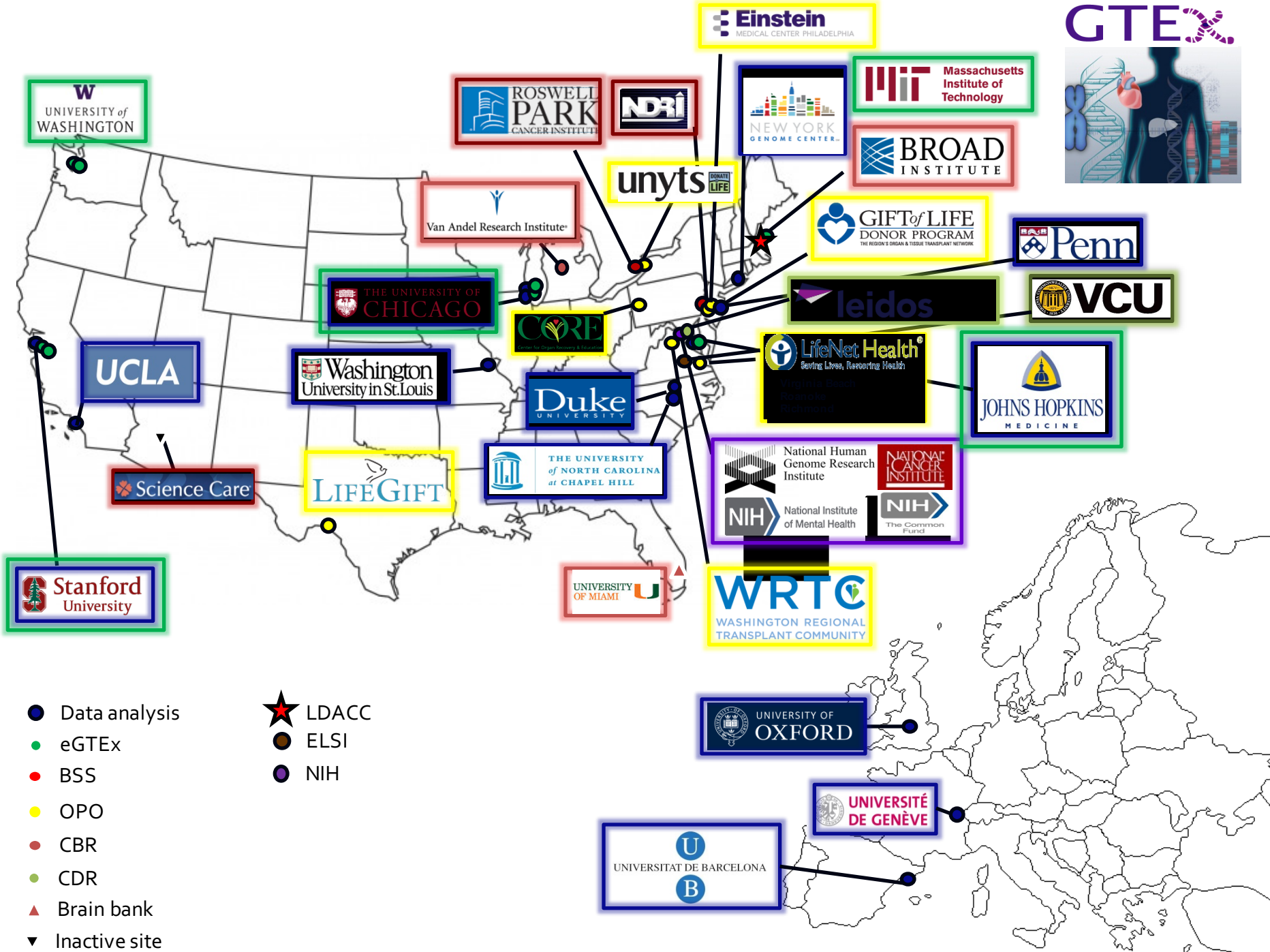
HOME ABOUT GTEx FAQ RESOURCE GUIDE CONTACT US AUTHOR BIOGRAPHIES

Log in

Communicating Effectively about Donation for the Genotype Tissue Expression Project

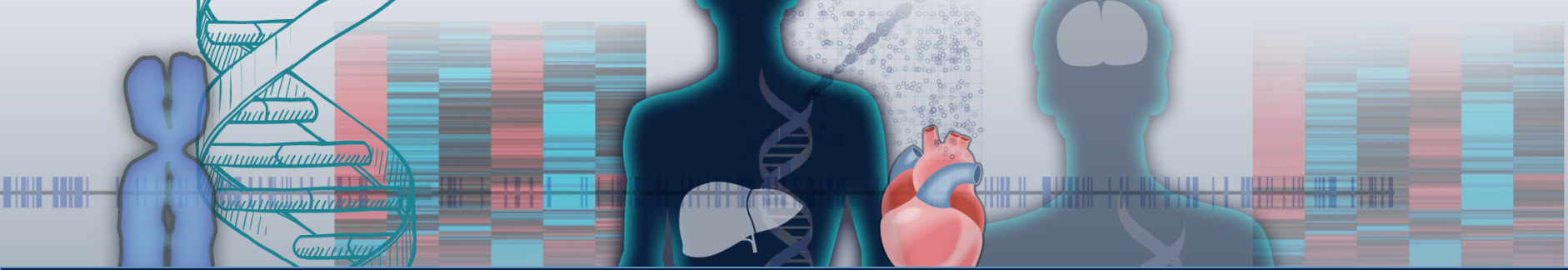
Welcome to Communicating Effectively about Donation for Research (CEaD-R). This training will reinforce your donation-related communication skills and improve your understanding of the GTEx research project. Implementing these skills you will achieve higher rates of consent for donations both for transplant and to GTEx biobanks.

Designed to help requesters communicate effectively about donation for GTEx.



Thank you!





The GTEx LDACC Project Update

Kristin Ardlie, Ph.D.
GTEx LDACC, Broad Institute

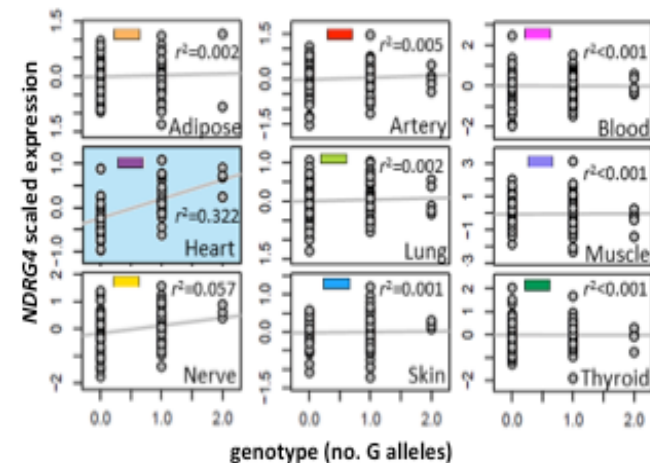
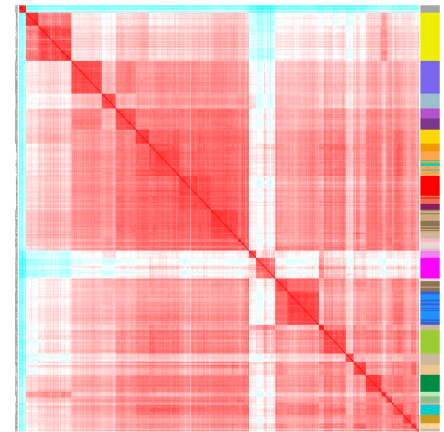
July 11, 2016



GTEx Project Goals

Characterize the regulatory architecture of human genome by understanding the role of genetic variation on gene expression variation across a wide range of non-diseased human tissues.

- Create an atlas of human tissue gene expression
- Comprehensive resource database of *cis*- and *trans*-eQTLs to enable studies of role of genetic variation on gene regulation across tissues; interpret GWAS studies



Scope - Primary Data Types

- ❑ 960 post-mortem donors
 - up to 53 tissues/donor (45 main sites)
- ❑ DNA sequence each donor
 - Whole genome (WGS) and whole exome (WES)
- ❑ RNA-sequencing on >25,000 tissues (~25 average/donor)
- ❑ Associated clinical and histopathological information
- ❑ Enhanced GTEx (eGTEx)
 - Protein quantifications (x 2)
 - Methylation (x2)
 - Histone modifications
 - DNase-seq
 - mmPCR-seq (deep ASE)
 - Somatic DNA seq (deep exome seq)
 - Analysis of telomere structure

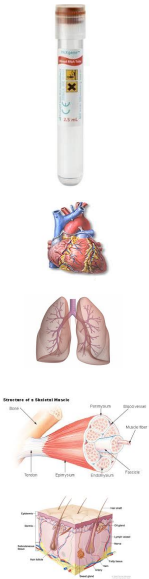
Multi-tissue AND multi-individual



Individual variation

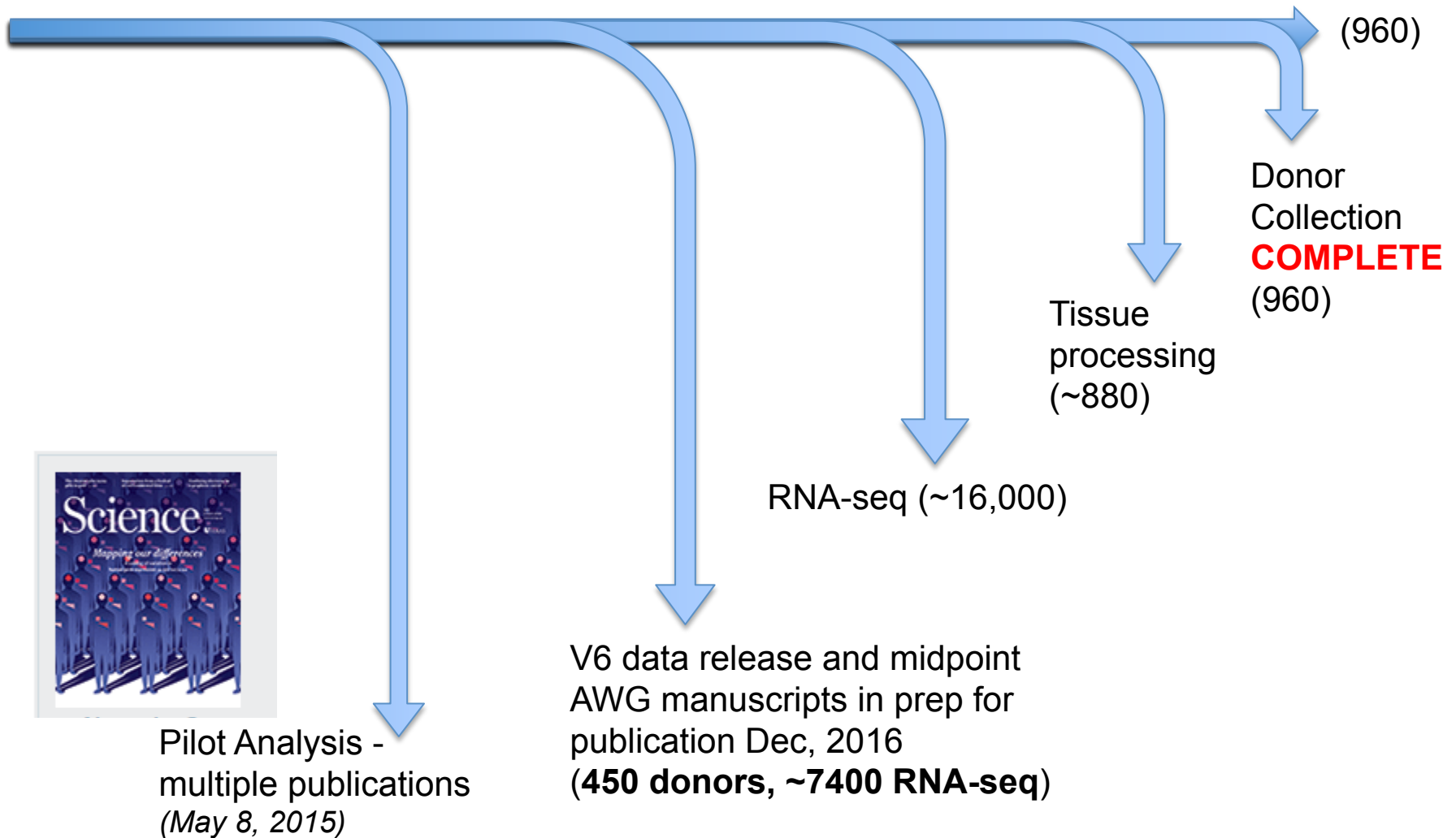
Within tissue -
across
population.

Tissue types

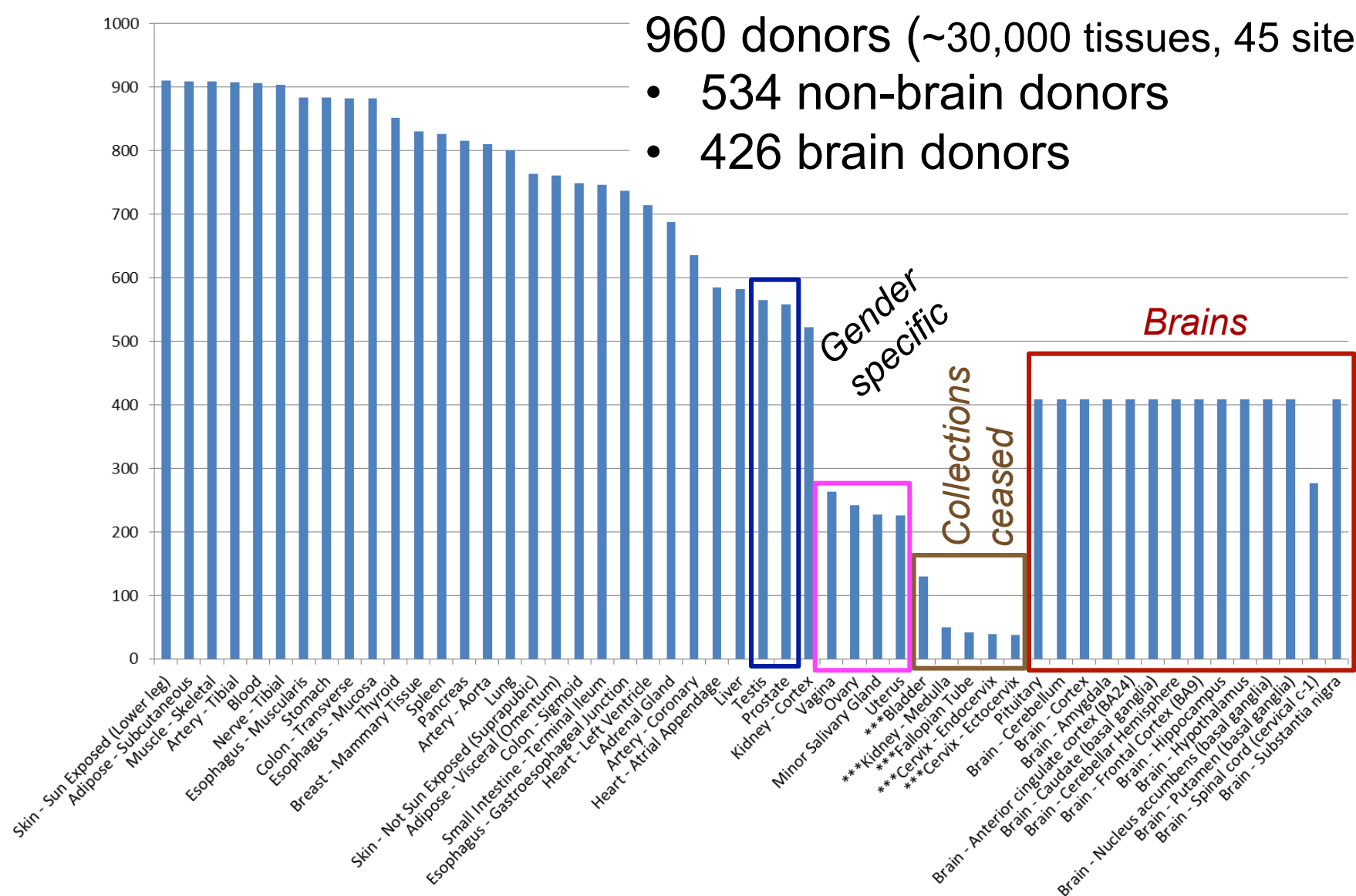


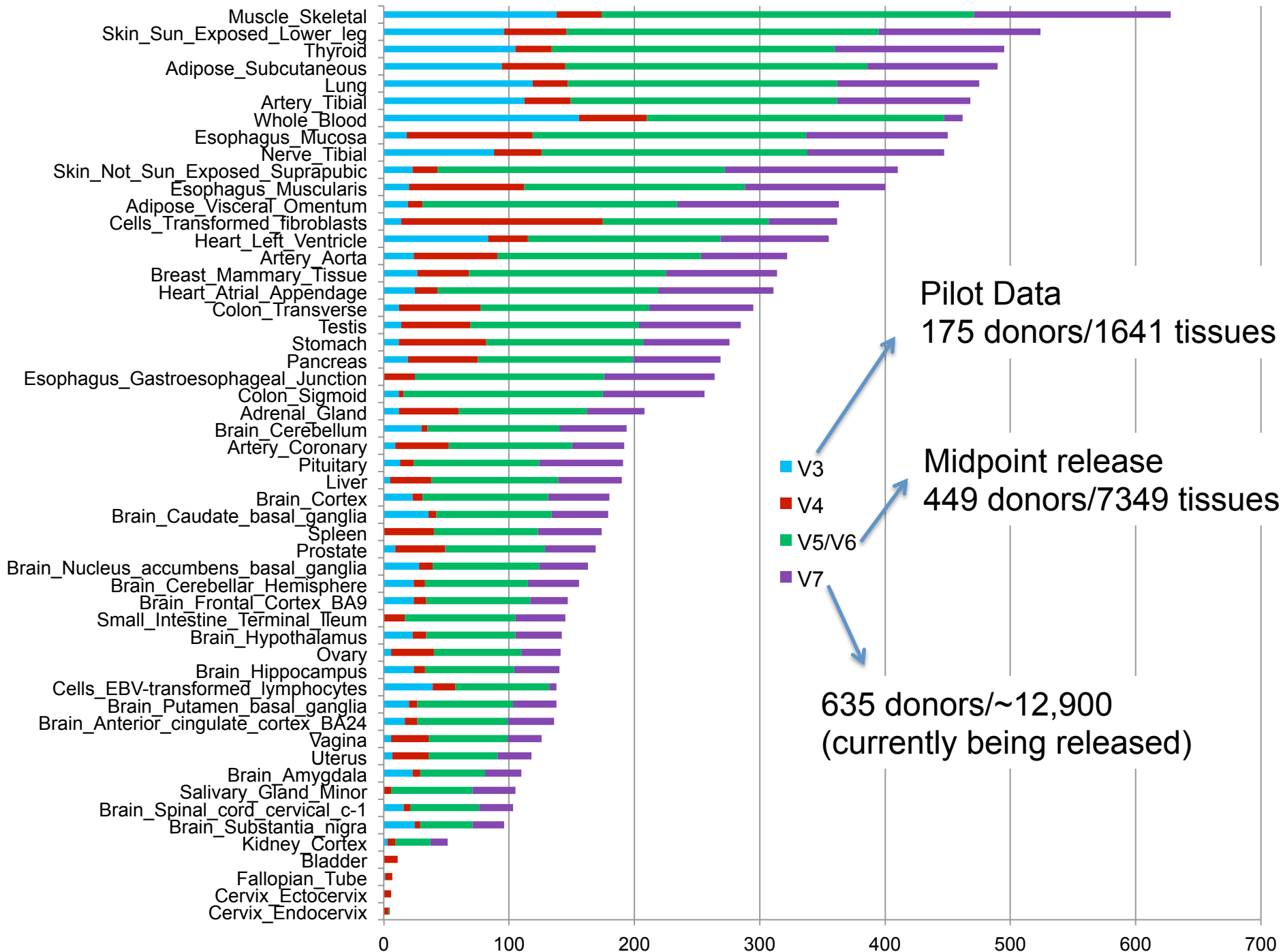
Within individuals –
across tissues.

GTEx Project Timeline



Tissue collection variable by site



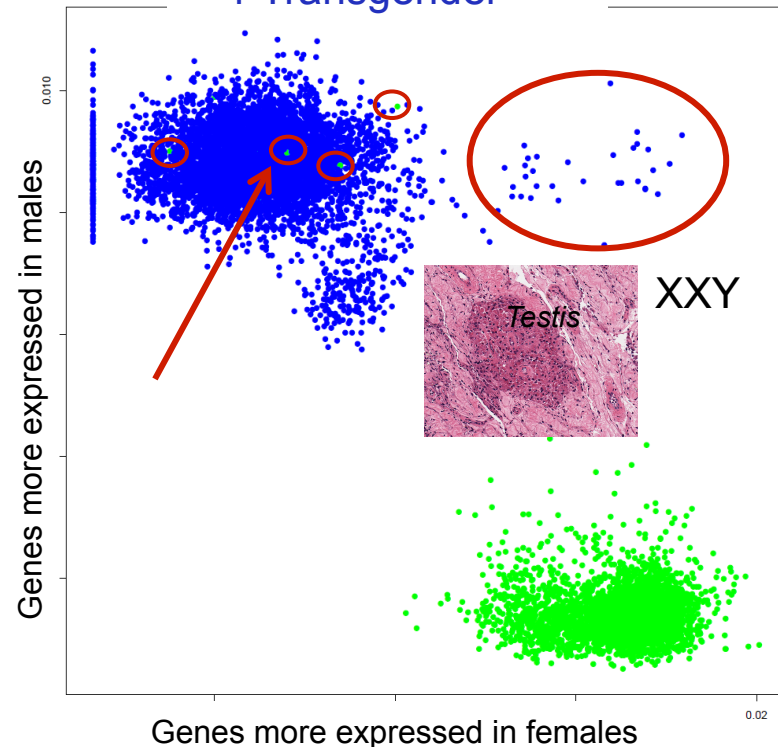


Data Release Categories

Raw Data

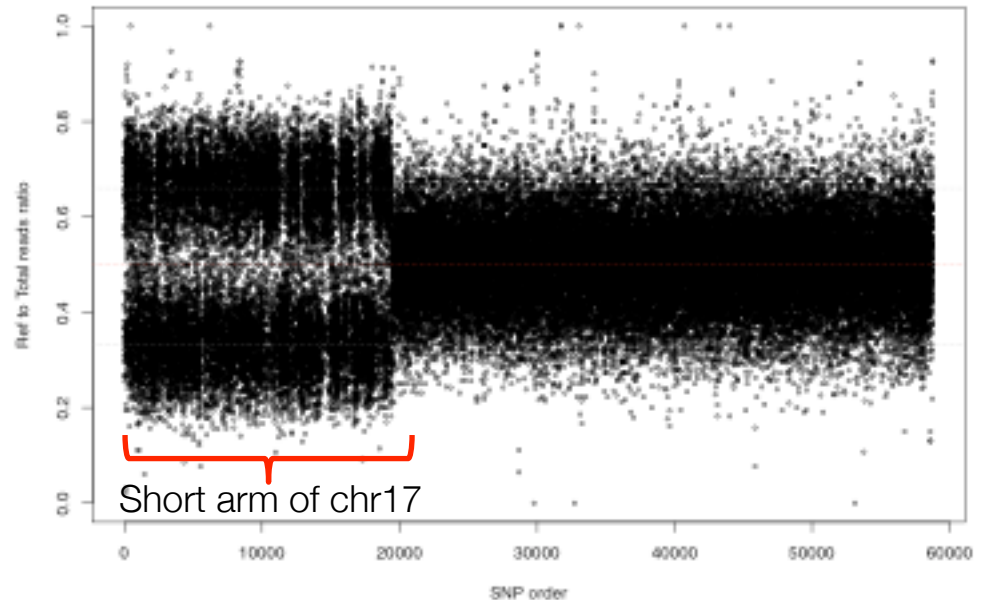
All raw data (WEX, WGS, RNA-seq) passing QC are released to dbGaP. Includes duplicates and samples EXCLUDED from analysis freezes (e.g. phenotypic, clinical exclusions)

- 4 Klinefelter (XXY)
- 1 Transgender



Large Chromosomal Abnormalities

- 2 Trisomy 21
- 1 17p duplication



Data Release Categories

Raw Data

All raw data (WEX, WGS, RNA-seq) passing QC are released to dbGaP. Includes duplicates and samples EXCLUDED from analysis freezes (e.g. phenotypic exclusions)

Analysis Freeze – RNA-seq/Expression

Available on GTEx Portal

EXCLUDES Duplicates and tissue expression outliers

Analysis Freeze - eQTL

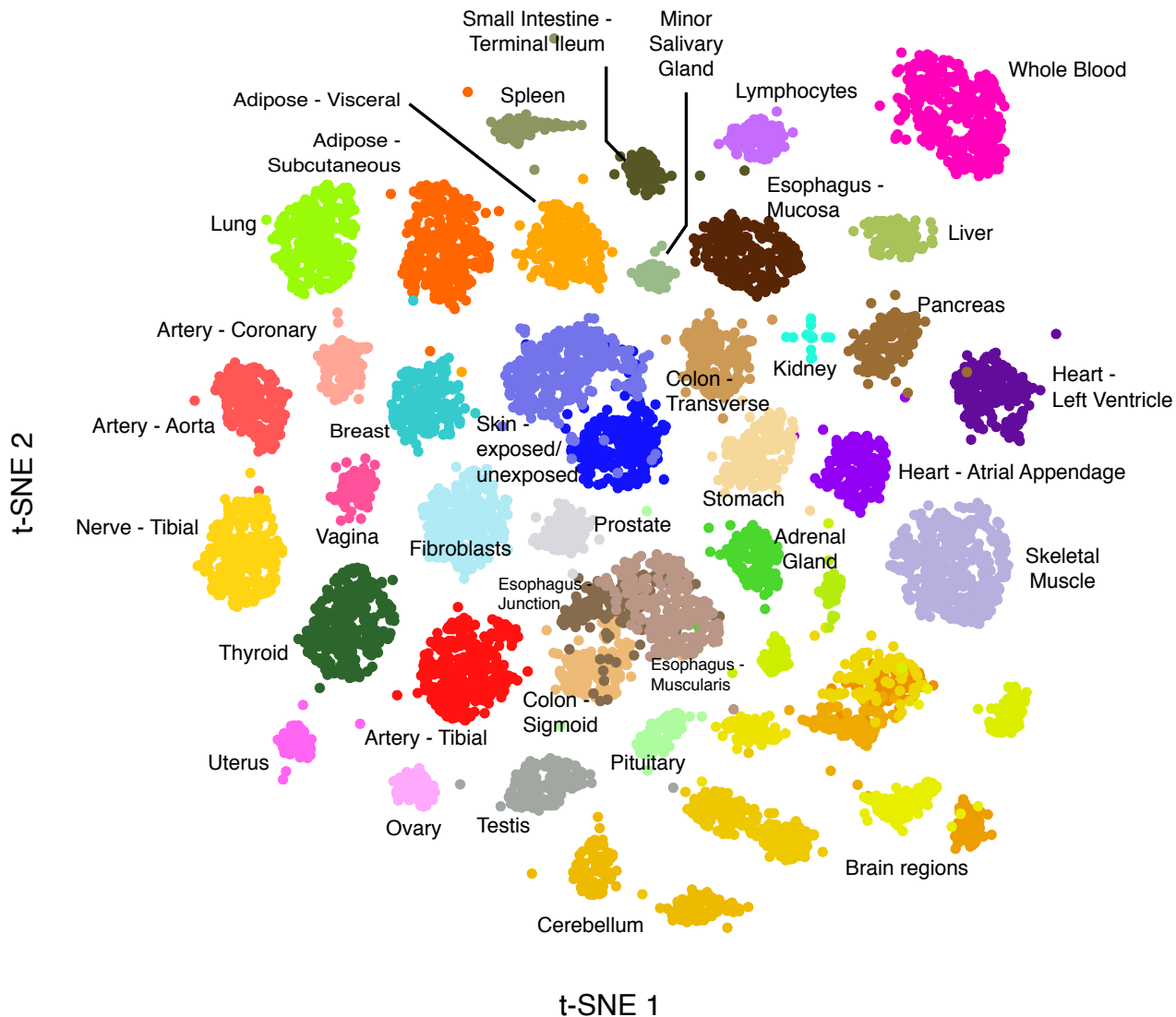
Available on GTEx Portal

Requires both genotype and expression data

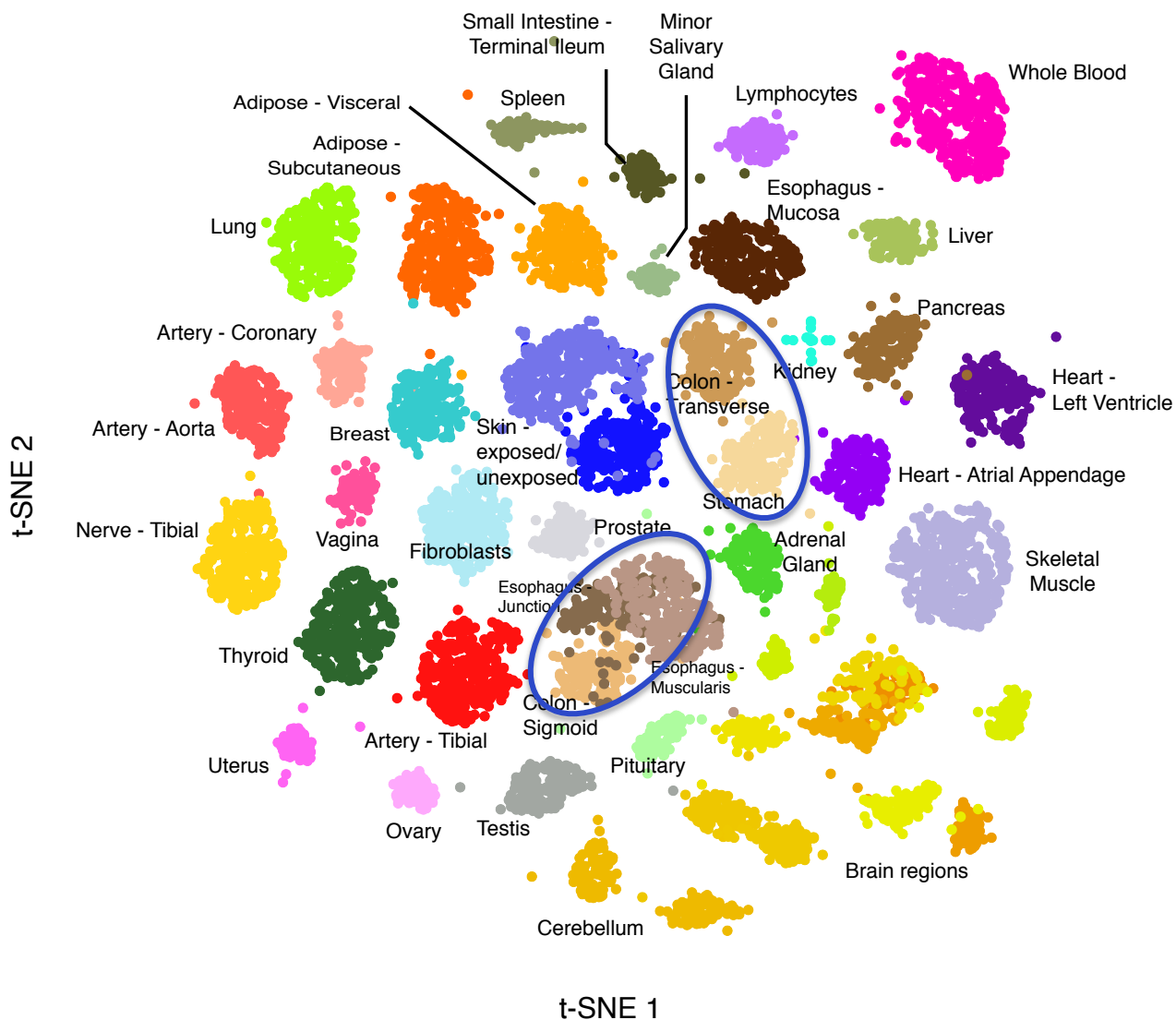
EXCLUDES Expression Data from samples without matching donor genotype data



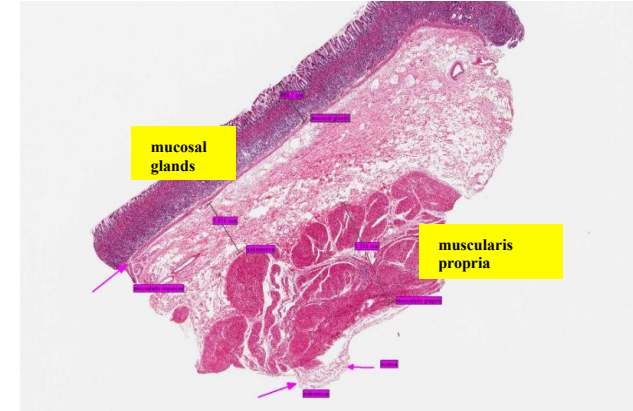
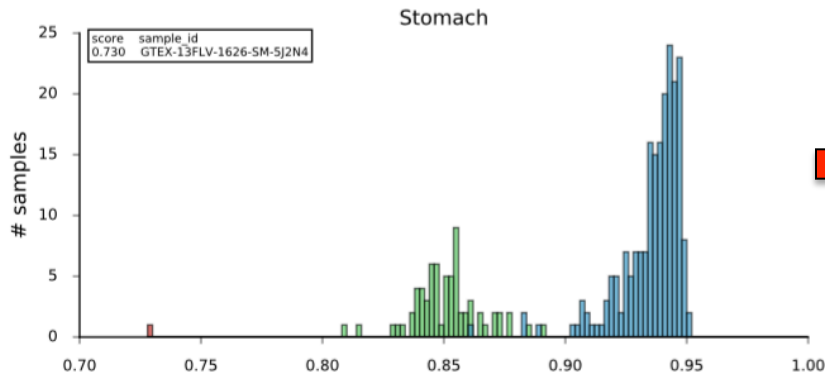
Production Data 1 - GTEx Transcriptome



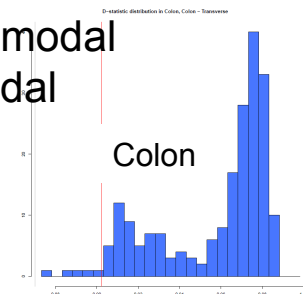
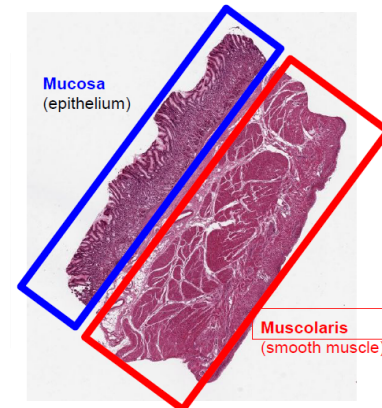
Production Data 1 - GTEx Transcriptome



Sample Heterogeneity - LCM



LCM – 30 tissues,
90 samples



1. Stomach – COMPLETE ✓

2. Colon, transverse – also bimodal

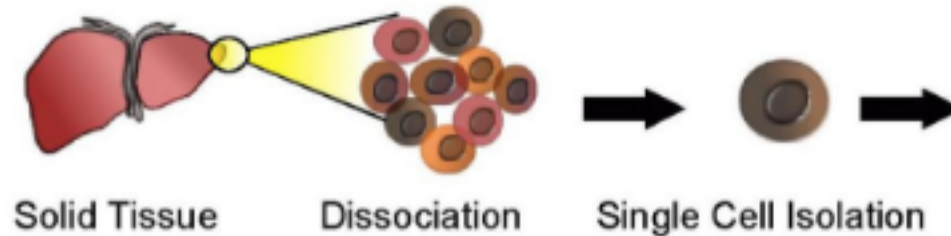
3. Terminal ileum – also bimodal

4. Pancreas

5. Skin

6. Liver

Sample Heterogeneity – Single Cell

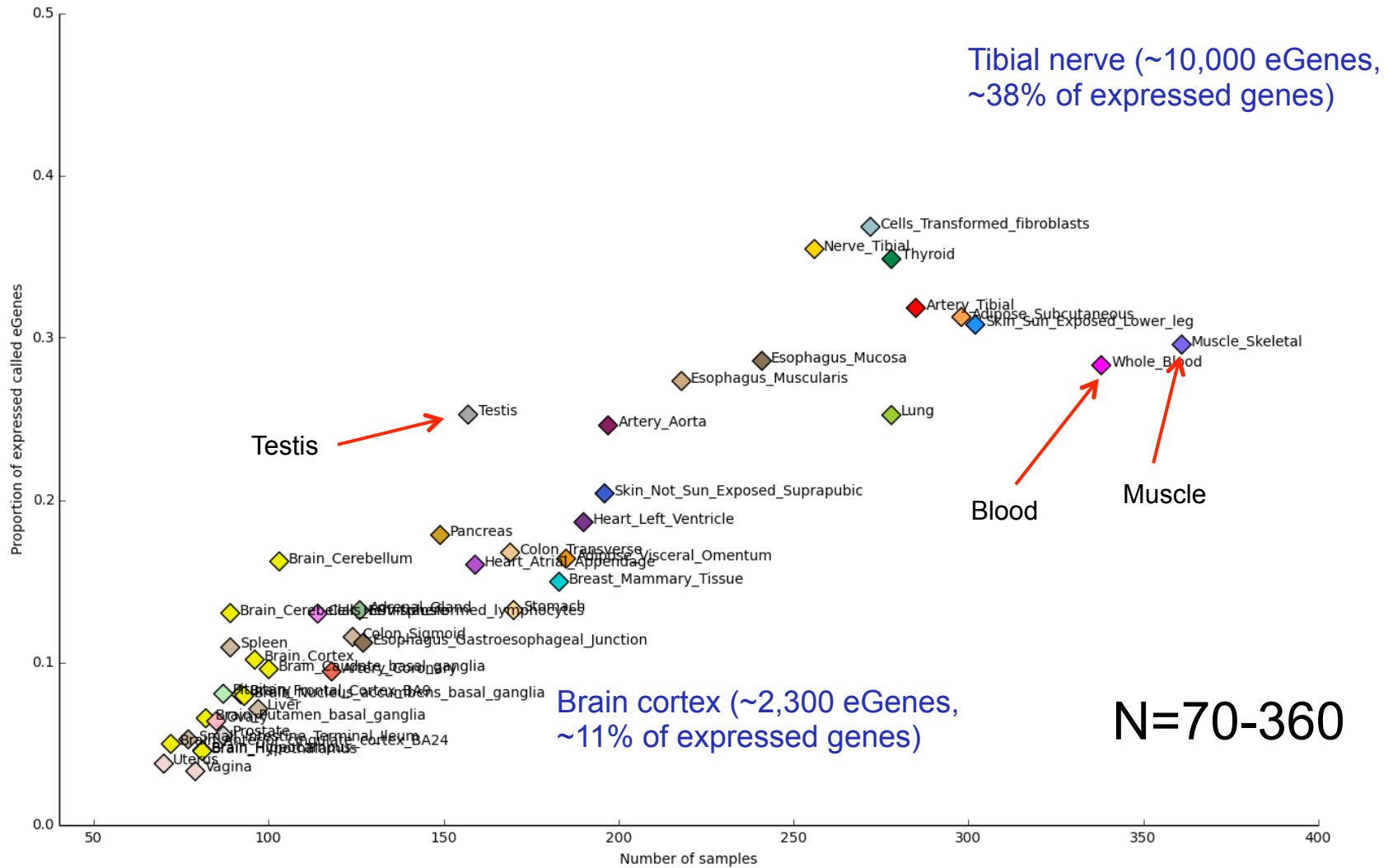


SEVERAL SINGLE CELL PILOT PROJECTS ONGOING WITH EXISTING GTEx BANKED SAMPLES:

1. James Eberwine – U. Penn
2. Kun Zhang – UC San Diego
3. Aviv Regev – Broad Institute

Production Data 2 - *cis*-eQTLs (eGenes)

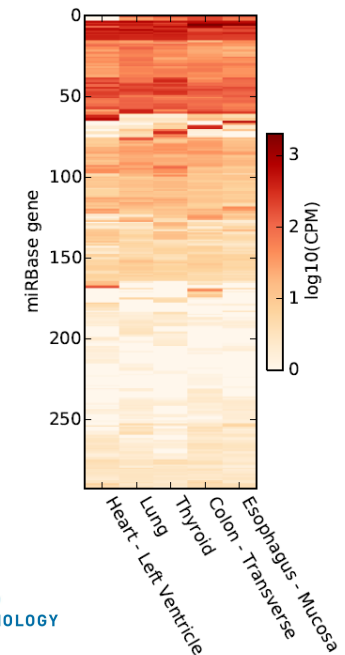
Proportion of expressed called eGenes



Samples

Ongoing and Additional Data

- Core Data production still ongoing (with new analysis methods and results from the AWG being added)
- Analysis pipeline being made more readily accessible
- Long read RNA-seq – PacBio
- Small RNA-seq – all samples
 1. Sequencing-based – for discovery
 2. High throughput, automated
 3. Low cost



ENCODE and GTEx – “ENTEx”

	Donor 1	Donor 2	Donor 3	Donor 4	GTEx #'s
Adipose - Subcutaneous	7.4	7.2	6.8	7.6	490
Adipose - Visceral (Omentum)	8.2	8.2	8.4	7.9	363
Adrenal Gland	8.7	7.3	9.5	9.1	196
Artery - Aorta	6.4	8.0	8.1	6.9	322
Artery - Coronary			8.5	7.7	183
Artery - Tibial	7.6	6.6	7.8	6.8	468
Breast - Mammary Tissue	7.0	8.2	6.9	7.3	314
Colon - Sigmoid	8.2	8.5	8.1	7.0	244
Colon - Transverse	7.4	7.5	7.8	7.1	285
Esophagus - Gastroesophageal Junction	7.0	8.1	8.5	7.8	264
Esophagus - Mucosa	8.8	9.8	9.9	9.4	450
Esophagus - Muscularis	Unacceptable	7.7	Unacceptable	3.5	402
Heart - Atrial Appendage			8.7	7.4	311
Heart - Left Ventricle			8.5	7.4	355
Liver			9.6		185
Lung	8.7	Unacceptable	8.2	6.0	475
Muscle - Skeletal	6.9	8.5	8.4	8.2	628
Nerve - Tibial	7.9	7.3	7.3	4.9	447
Ovary			8.3	7.7	135
Pancreas	8.2	7.1	8.0	7.1	269
Prostate	7.7	7.9			160
Skin - Not Sun Exposed (Suprapubic)	6.3	5.8	7.7	5.6	410
Skin - Sun Exposed (Lower leg)	8.6	8.5	8.0	7.8	524
Small Intestine - Terminal Ileum	Unacceptable	Unacceptable	Unacceptable	5.3	145
Spleen	7.8	7.5	7.6	8.4	174
Stomach	9.1	8.2	6.4	5.7	264
Testis	7.6	6.9			271
Thyroid	7.8	8.0	7.2	6.8	484
Uterus			8.4	3.5	111
Vagina			7.2	7.8	120

4 donors:

- GTEx collections for ENCODE
- 2 M + 2F
- All tissues, not brain

ENTEx Data available @ ENCODE

ENCODE Data Encyclopedia Materials & Methods Help

Experiment Matrix

Click or enter search terms to filter the experiments included in the matrix.

Assay

RNA-seq	71	Assay category	Transcription	119	Target of assay	control	39	Date released	May, 2016	111	Available data	fastq	226
ChIP-seq	63	DNA binding	63	histone	14	February, 2016	53	bam	199				
small RNA-seq	42	DNA accessibility	38	histone modification	14	June, 2016	36	bigWig	138				
ATAC-seq	19	Genotyping	4	transcription factor	11	July, 2016	11	tsv	119				
DNase-seq	19	RNA binding	2	chromatin remodeller	9	March, 2016	9	bed narrowPeak	19				

[+ See more...](#) [+ See more...](#) [+ See more...](#) [+ See more...](#)

Organism

Homo sapiens 226

Biosample type

tissue 226

Organ

large intestine	43
thyroid gland	20
esophagus	16
adrenal gland	14
stomach	14

[+ See more...](#)

Project

ENCODE 226

Genome assembly (visualization)

GRCh38	119
hg19	20

Audit category:

insufficient read depth 1

Audit category:

low read depth	23
antibody eligible via exemption	12
mild to moderate bottlenecking	12
moderate library complexity	4
inconsistent control read length	3

[+ See more...](#)

ASSAY

226 results

Clear Filters

BIOSAMPLE

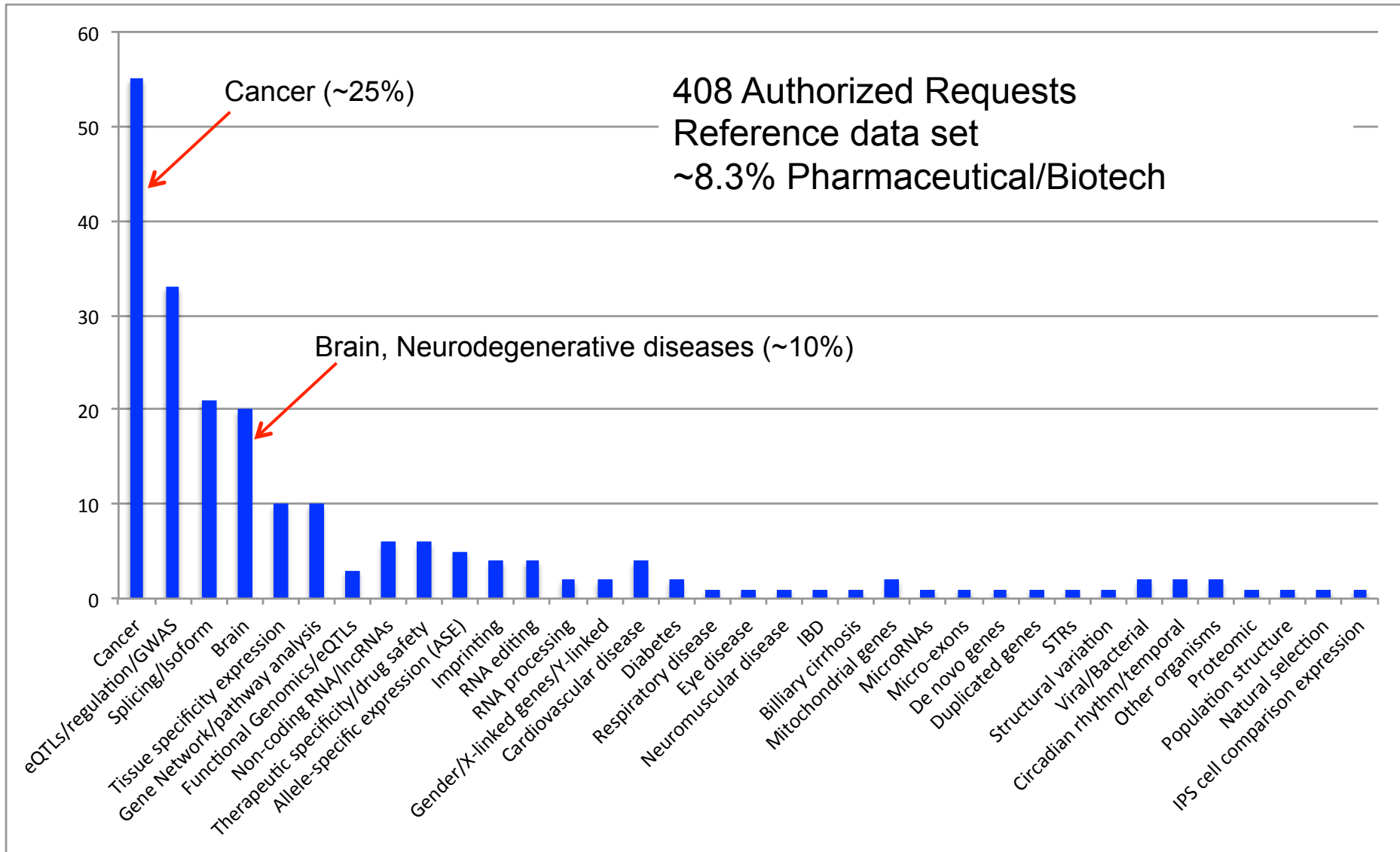
BIOSAMPLE	RNA-seq	ChIP-seq	small RNA-seq	ATAC-seq	DNase-seq	RAMPAGE	genotyping HTS	eCLIP
tissue								
transverse colon	2	7	4	4	3		4	
thyroid gland	2	13	1		2	2		
sigmoid colon	4	6	4	4	1			
adrenal gland	2	3	2		3	2	2	
esophagogastric junction	4	4	4	2				
stomach	8	4	2					
spleen	4	4	4	1				
adipose tissue	4	4	1	1				
mammary gland	3	4		3				
esophagus mucosa	4	4		4	1			
skeletal muscle tissue	4	3			1	1		
pancreas	2	2	2	2				
suprapubic skin	4	4						
muscle layer of esophagus	4	3						
adipose tissue of omentum	3	1	1					
tibial nerve	4	1						
aorta	2	1				1		
female gonad	2	1		1				
ileum	4							
upper lobe of left lung	4							
prostate		2		1				
right lobe of liver	2			1				
testis		2		1				
uterus		1		2				
heart left ventricle	2							
lower leg skin	2							
right cardiac atrium	2							
omentum	1							
tibial artery				1				
vagina		1						

[Download](#) [Visualize -](#)

Enhanced GTEx (eGTEx)

- Many of these same assays (e.g. methylation, ChIP-seq, DNase-seq) plus others, also being produced by same groups across larger number of GTEx donors

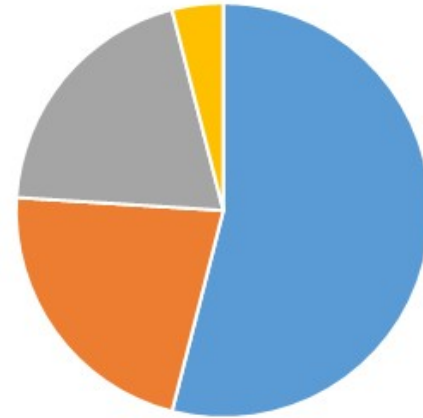
Data Access dbGaP (raw data)



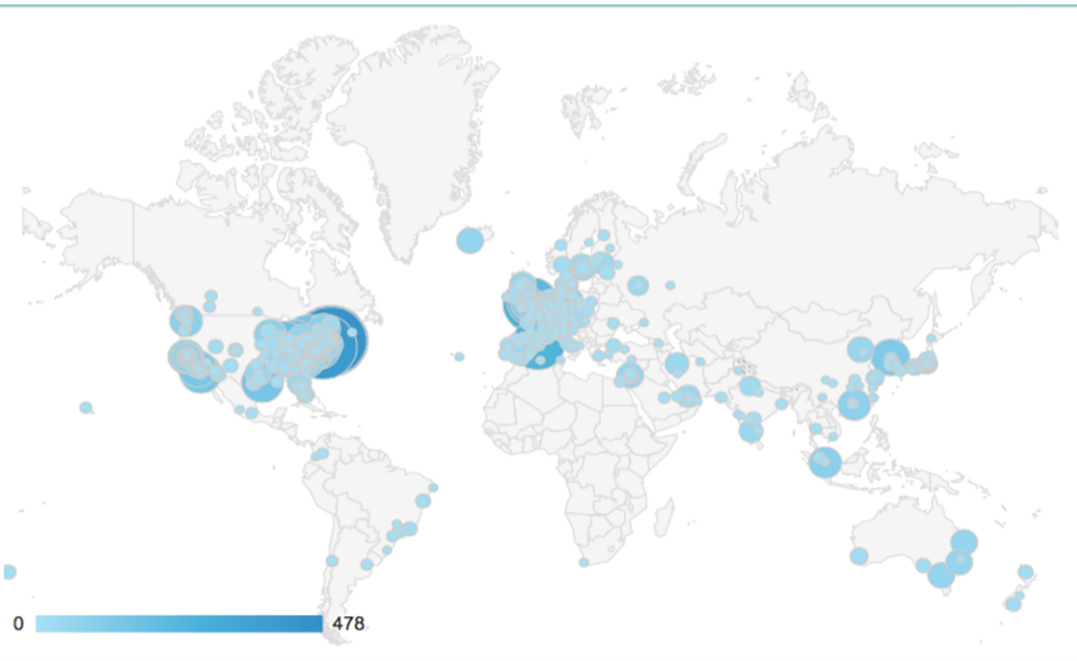
Data Access GTEEx Portal

www.gtexportal.org

Most data access via GTEEx Portal:
7,500 registered users
47,000 unique IP addresses
~8,000 unique visitors per month.
~53% Academic, 25% Biotech/Pharma



■ Academic network ■ Biotech/Pharma network ■ Private network ■ Unavailable



GTEEx Data also now integrated in to:

- The Protein Atlas
- UCSC browser
- Ensemble browser
- GeneCards

eQTL Plot Problems Resolved

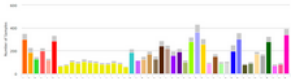
2015-05-27

[Read More >>](#)

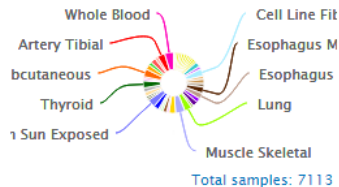
Current Release and Data Summary

Latest Release: V6 (dbGaP Accession phs000424.v6.p1)

Summary Statistics



Browse eQTL Tissues



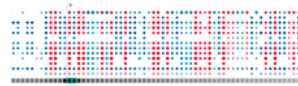
Genetic Association

Single Tissue eQTLs

eQTL IGV Browser



Gene eQTL Visualizer



Test Your Own eQTLs

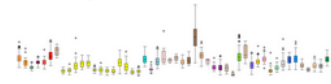
Analysis Results in Older Releases

- Multi-tissue eQTLs (v4 & v3)
- Splice QTLs (v3)
- Protein-truncating Variants (v3)
- Genomic Imprinting (v3)

Transcriptome

Top 100 Expressed Genes in a Tissue (e.g. Blood)

Gene Expression in Tissues



Exon and Isoform Expression



Links

- #### Documentation
- About GTEX
 - Publication Policy
 - Consortium Members

- #### External Links
- dbGaP
 - NIH Common Fund
 - NHGRI

News

GTEX Portal Performance Issues

2015-09-22

[Read More >>](#)

Normalized Expression Matrices and Covariates Released

2015-06-26

[Read More >>](#)

GTEX Portal Maintenance Outage June 14, 2015

2015-06-04

[Read More >>](#)

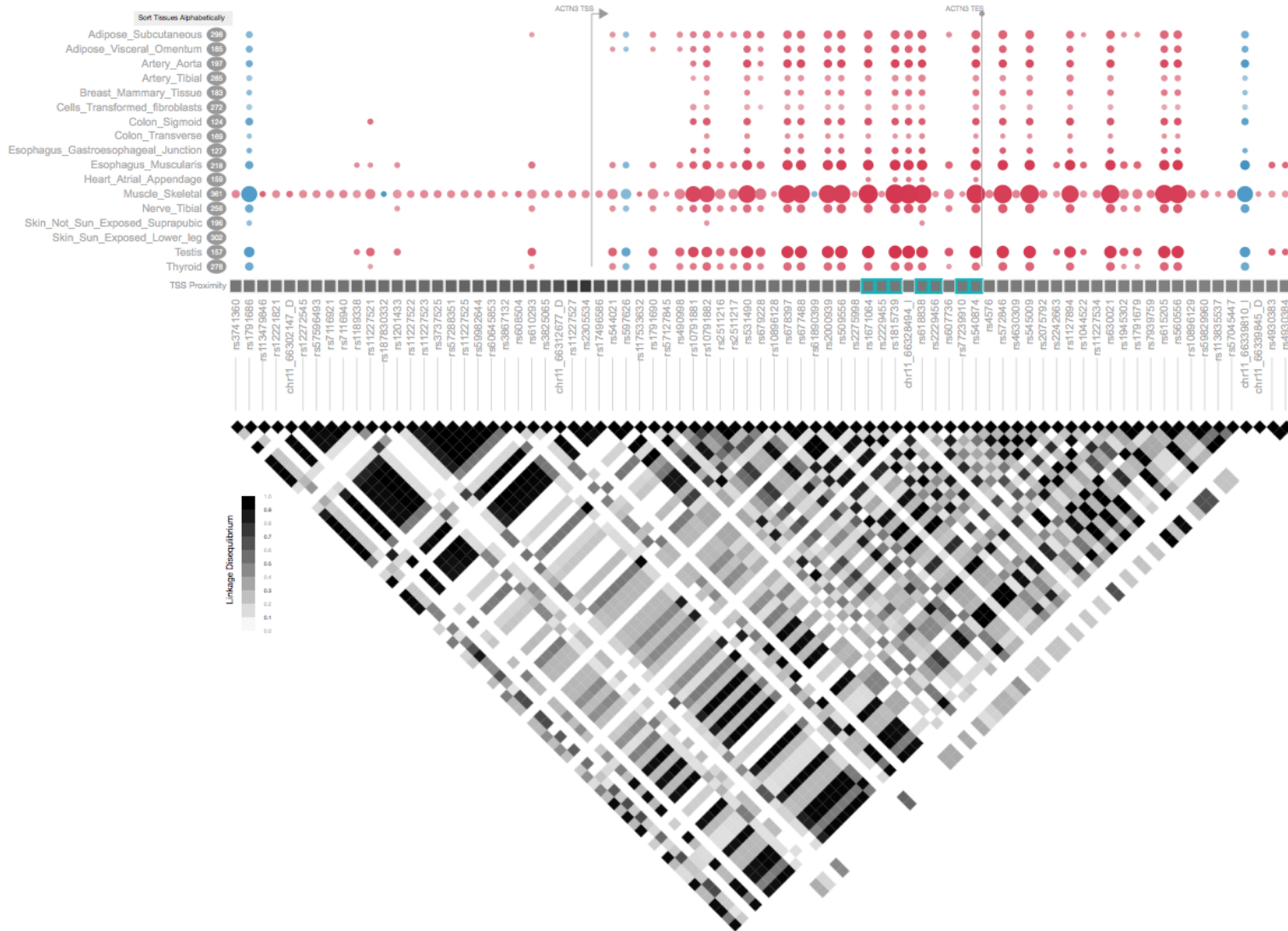
eQTL Plot Problems Resolved

2015-05-27

[Read More >>](#)

See Live Demo in Lobby
And also Ensemble live
demo in Lobby

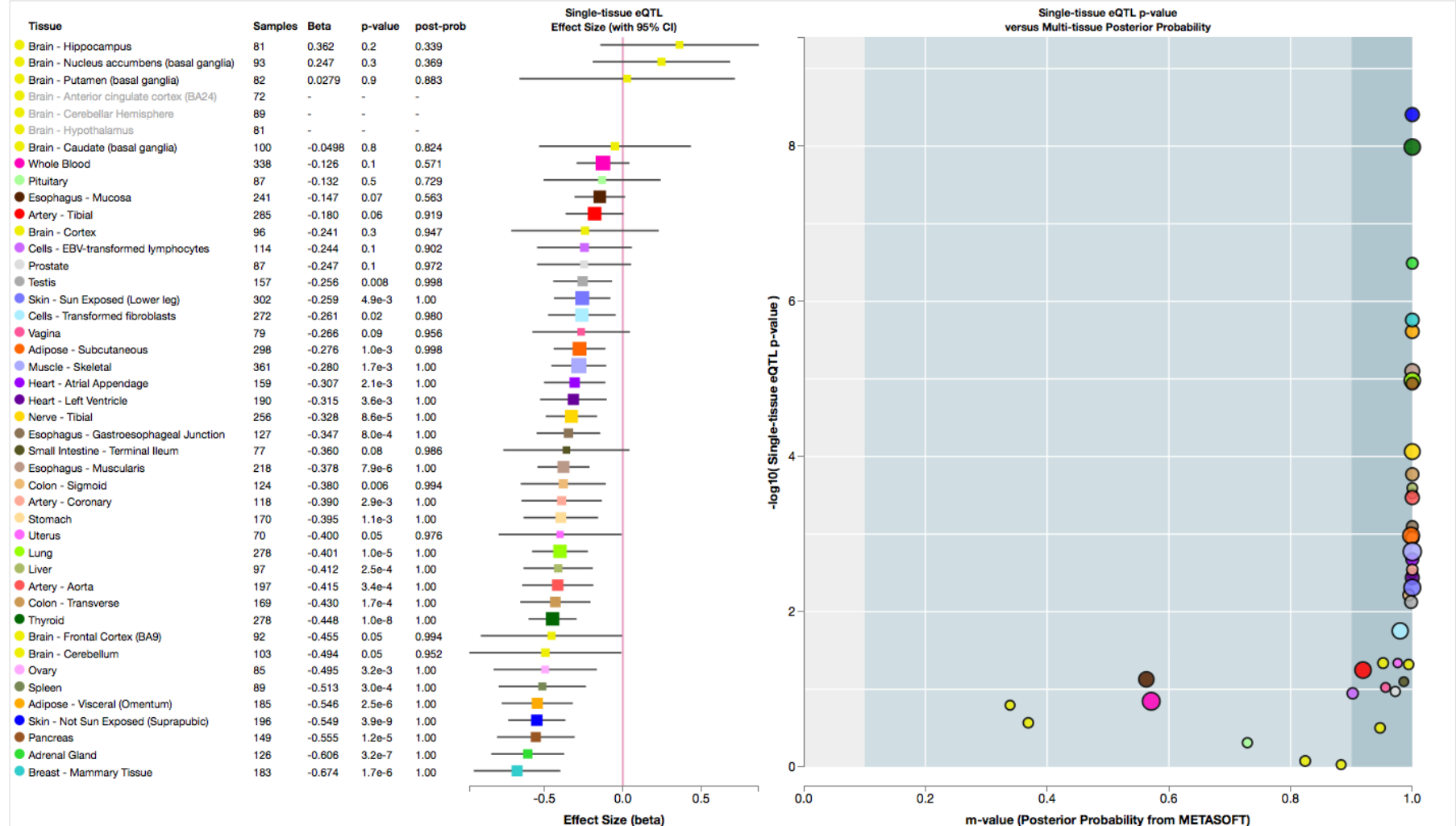
New Portal Features – LD track



New Portal Features – Single & Multi-tissue eQTL comparison

Multi-tissue eQTL Comparison

ENSG00000224956.5 RP11-206L10.1 and rs12096608 eQTL (Meta Analysis RE2 P-Value: 1.37159e-76)



New Portal Features - Histology

Subject

Adipose Tissue

- Adipose - Subcutaneous
- Adipose - Visceral (Omentum)

Adrenal Gland

- Adrenal Gland

Blood

- Cells - EBV-transformed lymphocytes
- Whole Blood

Blood Vessel

- Artery - Aorta
- Artery - Coronary
- Artery - Tibial

Brain

- 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

- Breast - Mammary Tissue

Colon

- Colon - Sigmoid
- Colon - Transverse

Esophagus

- Esophagus - Gastroesophageal Junction
- Esophagus - Mucosa
- Esophagus - Muscularis

Heart

- Heart - Atrial Appendage
- Heart - Left Ventricle

Kidney

- Kidney - Cortex

Liver

- Liver

Lung

- Lung

Muscle

- Muscle - Skeletal

Nerve

- Nerve - Tibial

Ovary

- Ovary

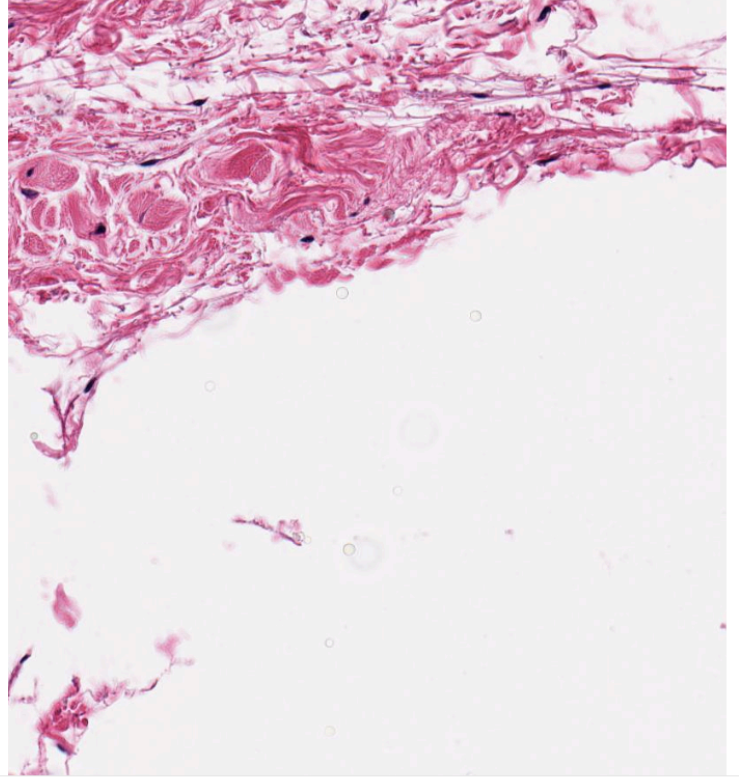

Pancreas

- Pancreas

Pituitary

- Pituitary

Data Slide



Slide details

Tissue:	Adipose - Subcutaneous
Gender:	male
Age Bracket:	60-69
Sample ID:	GTEX-N7MS-0326-SM-4E3K2
Hardy Scale:	2
Pathology Notes:	OK for analysis

Poster #3 (Meier)
Also #4 (Qi)

GTEEx Key Contacts

Scientific:

Kristin Ardlie, LDACC Broad: kardlie@broadinstitute.org

Francois Aguet, LDACC Broad: francois@broadinstitute.org

Ayellet Segre, LDACC Broad: asegre@broadinsitute.org

Casandra Trowbridge, LDACC Broad: ctrowbri@broadinstitute.org

Program:

Simona Volpis, NHGRI volpis@mail.nih.gov

Su Koester, NIMH koesters@mail.nih.gov

Data:

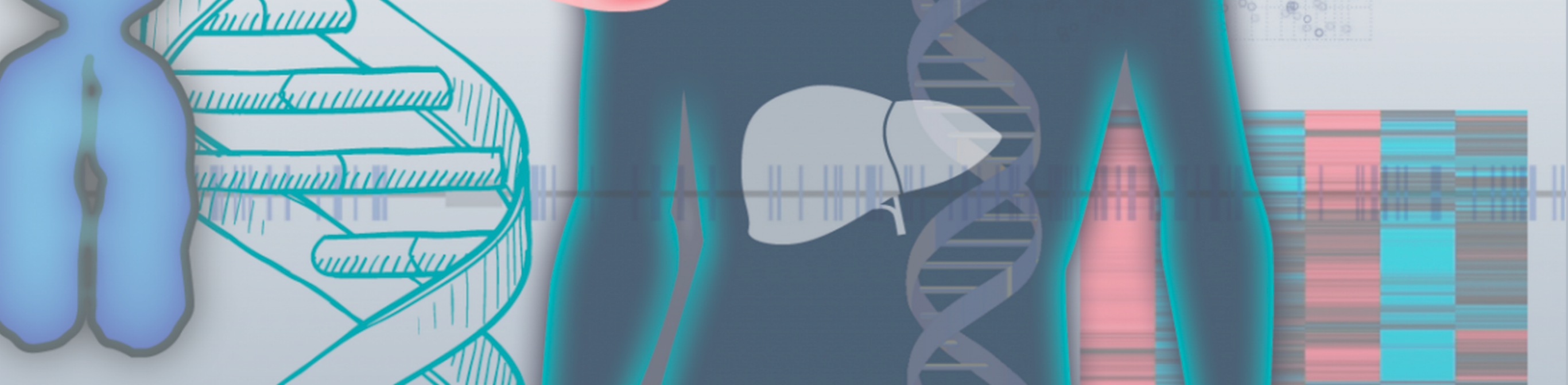
dbGaP (raw data):

[www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?
study_id=phs000424.v1.p1](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v1.p1)

GTEEx Portal: www.gtexportal.org

Sample Access: www.gtexportal.org/home/samplesPage





The v7 data release

GTEEx Community Meeting :: 07/11/2016

François Aguet, Broad Institute

Outline

- Summary of changes from v6
- Summary of raw and derived data produced for v7
- Summary of benchmarking results for alignment & isoform quantification
- New pipelines:
 - WGS/WES sample and variant QC
 - RNA-seq alignment, quantification, and QC
 - eQTL discovery
- Planned changes and additions for v8 release

Summary of changes from v6

- **Genotyping:** microarrays => WGS/WES
- **RNA-seq alignment:** TopHat 1.4 => STAR 2.4.2a
- **Gene expression:** new collapsed gene model
- **Isoform quantification:** FluxCapacitor => RSEM
- **eQTL discovery:** MatrixEQTL => FastQTL

Core derived data

Expression

- Read counts for genes, transcripts, exons, junctions
- Normalized expression for genes, transcripts (TPM)
- Coverage tracks (bigWig)

eQTL

- Gene-level summary: best variant, q-value, etc.
- Significant variant-gene pairs
- All variant-gene pairs
- Expression matrices (BED format); normalized + TPM
- Covariates

All derived data will be available on the GTEx Portal (<http://gtexportal.org/>)

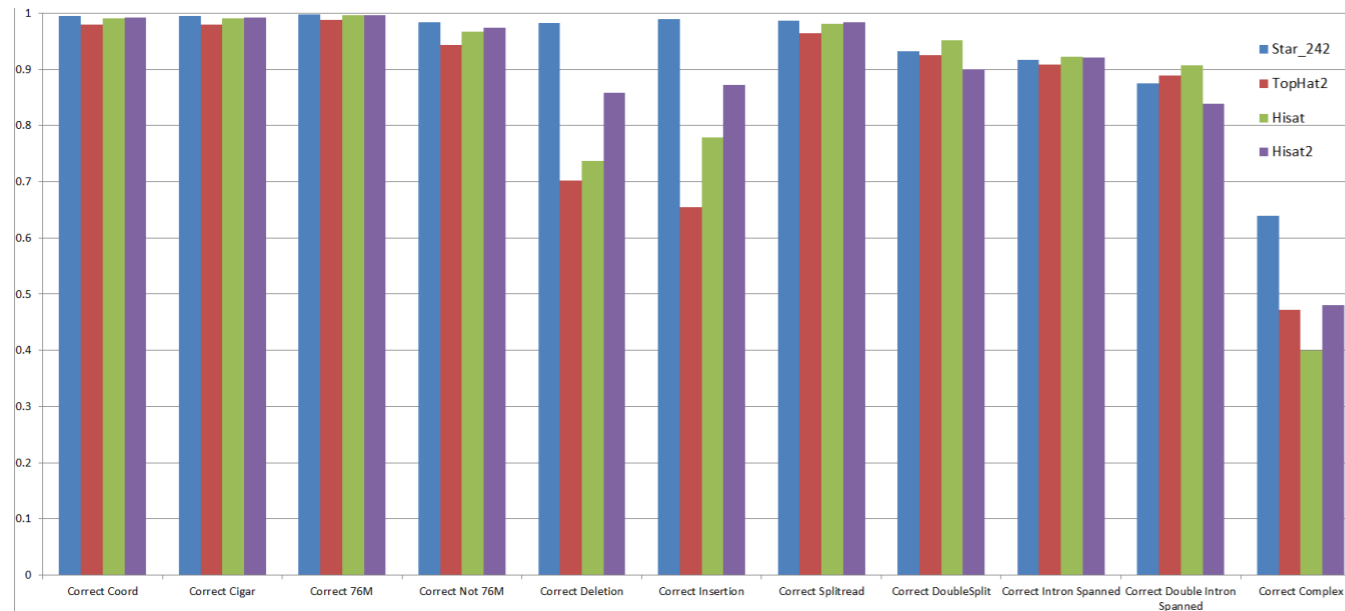
Additional derived data

- Splicing QTL
 - Altrans [Ongen & Dermitzakis, 2015]
 - sQTLseekeR [Monlong et al., 2014]
- Allele-specific expression [Castel et al., 2015; van de Geijn et al., 2015]
- Multi-tissue eQTL
 - Metasoft [Han & Eskin, 2012]

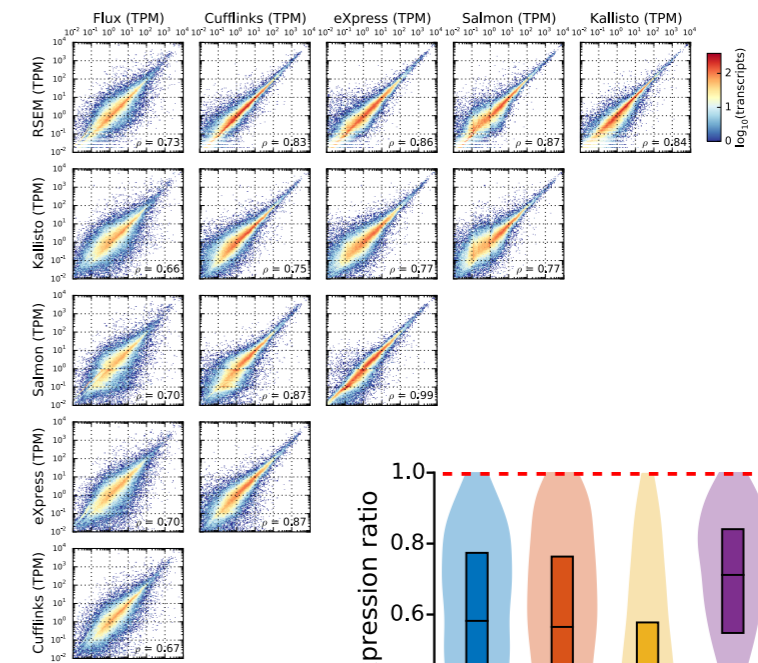
All derived data will be available on the GTEx Portal (<http://gtexportal.org/>)

Benchmarking

Spliced transcript alignment



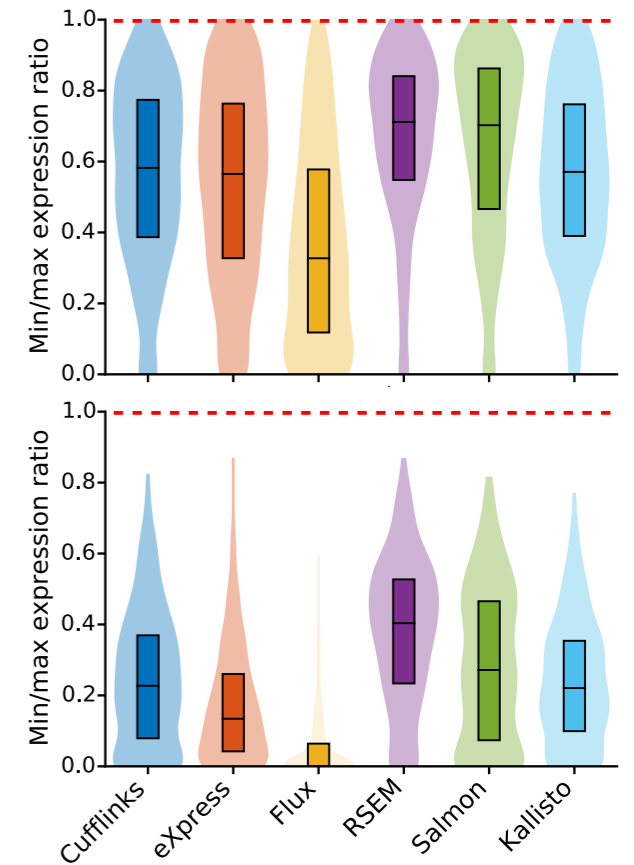
Transcript isoform expression estimation



Tim Sullivan, Broad Institute

	References
TopHat	Trapnell et al. 2009
TopHat2	Kim et al. 2013
STAR	Dobin, 2013
HISAT2	Kim et al., 2015

	Input alignment	References
Cufflinks	Genome	Trapnell et al. 2010 Trapnell et al. 2013
FluxCapacitor	Genome	Montgomery et al., 2010
RSEM	Transcriptome	Li et al. 2010 Li & Dewey 2011
eXpress	Transcriptome	Roberts et al. 2011 Roberts & Pachter 2012
Sailfish/ Salmon	Transcriptome / Raw reads	Patro et al. 2014 Patro et al., 2016
Kallisto	Raw reads	Bray et al. 2015



Gene-level expression quantification

- Quantification based on collapsed annotation (GENCODE v19)
 - Exclude exons from transcripts annotated as *retained_intron* or *read_through*
- GTEx RNA-seq protocol is unstranded
 - Exclude exon domains shared by overlapping genes
- Effect on eQTL discovery:
~10-15% more eGenes discovered vs. gene-level quantification from RSEM

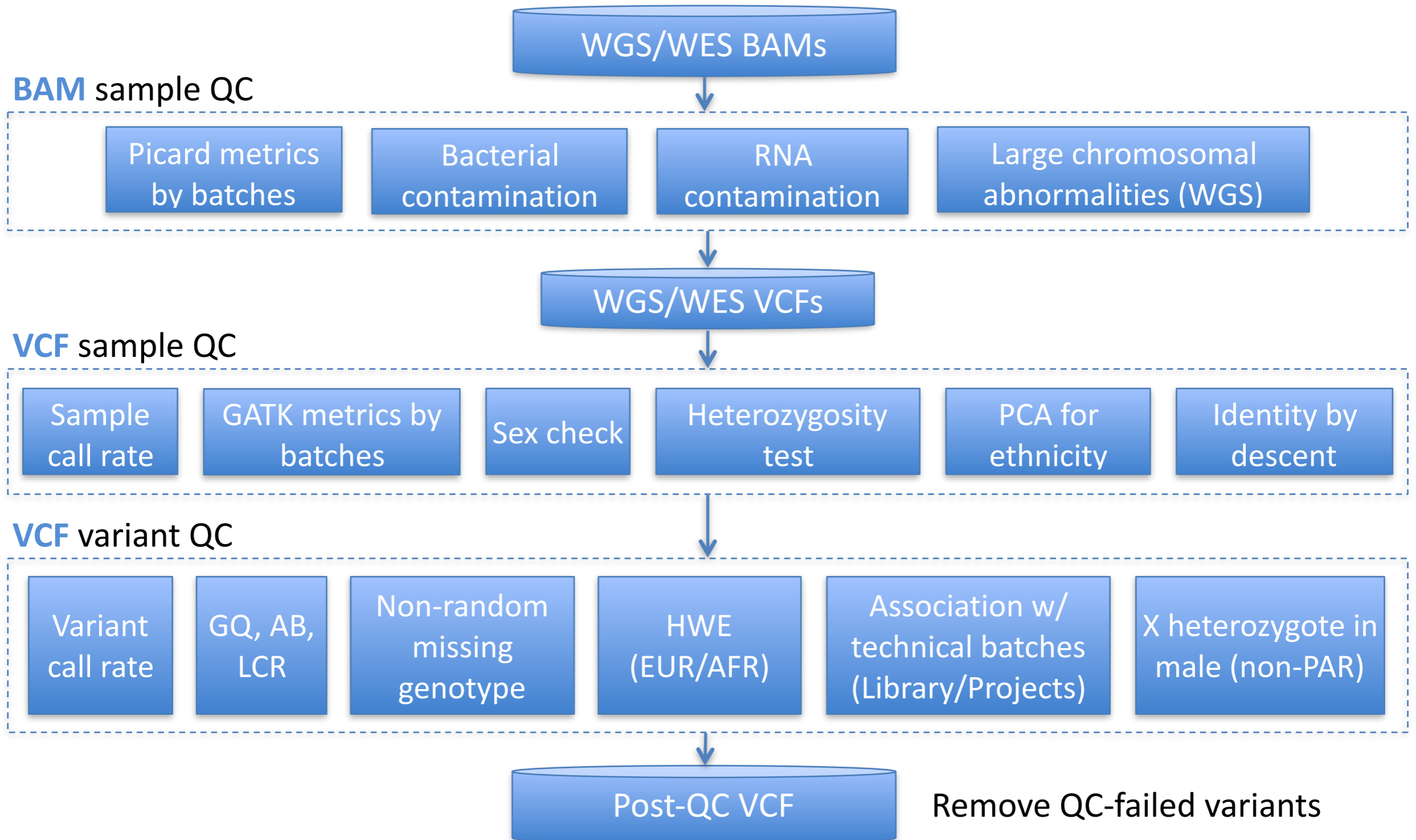
v6p release

- **Update of derived data only** (hosted on GTEx Portal)
 - Gene expression: read counts + RPKM GCT files.
 - eQTL: FastQTL instead of MatrixEQTL, otherwise identical. Includes chr. X eQTL.

Genotyping data and pipeline

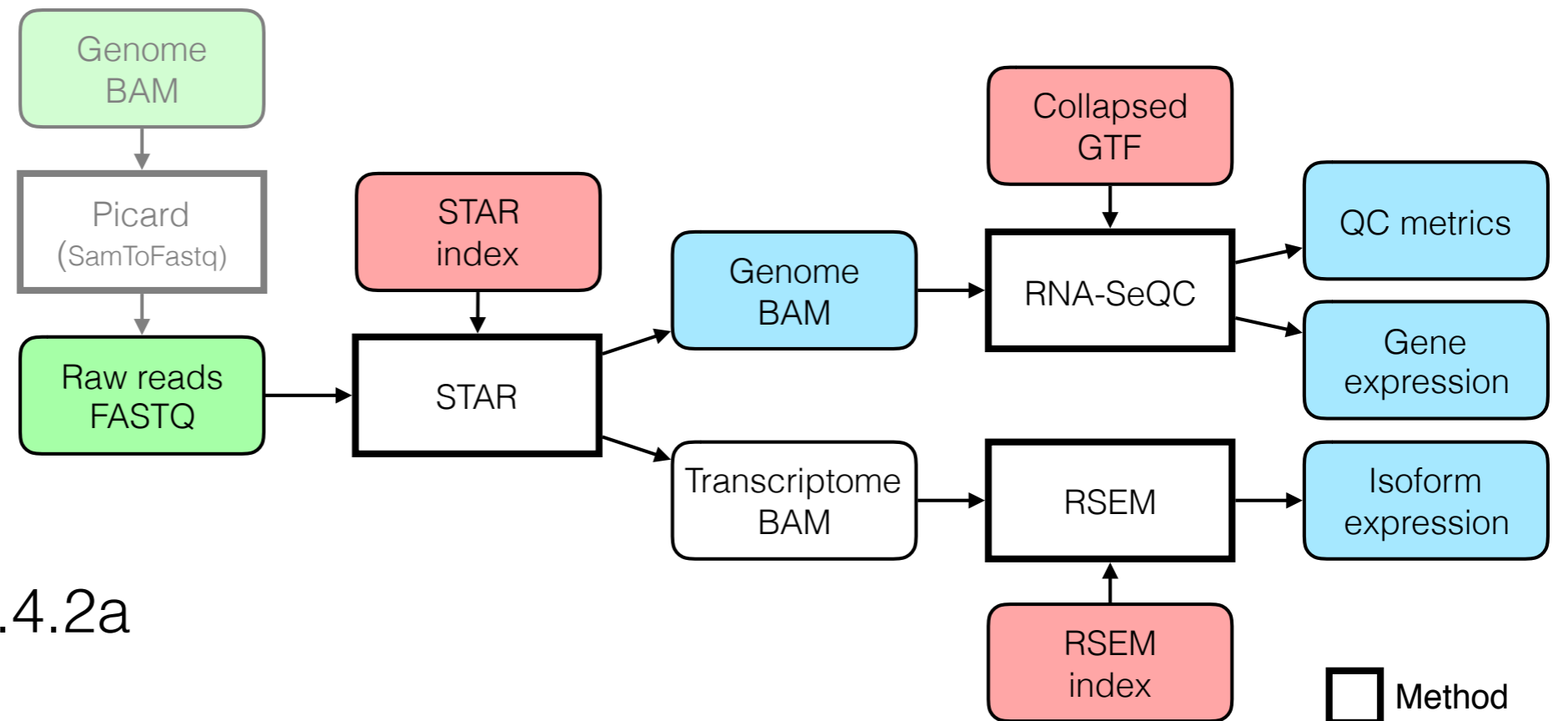
	WES	WGS
# donors	603	652
# donors (post sample QC for eQTL analysis)	603	635
Sequencing coverage	100x	30x
Alignment	BWA	BWA-MEM
Joint variant calling	HaplotypeCaller v3.4 (GATK)	HaplotypeCaller v3.4 (GATK)
Variant QC	-	GATK, Hail, Custom code
Functional and LoF annotations	Ensembl's Variant Effect Predictor + LOFTEE	Ensembl's Variant Effect Predictor + LOFTEE
Phasing of SNPs and indels	Local (in sequence read)	Local and long range with SHAPEIT
Structural variant calling	-	GenomeSTRiP, LUMPY (merged call set)

Overview of WGS/WES QC pipeline



See poster #20 (Li et al.)

RNA-seq alignment and quantification

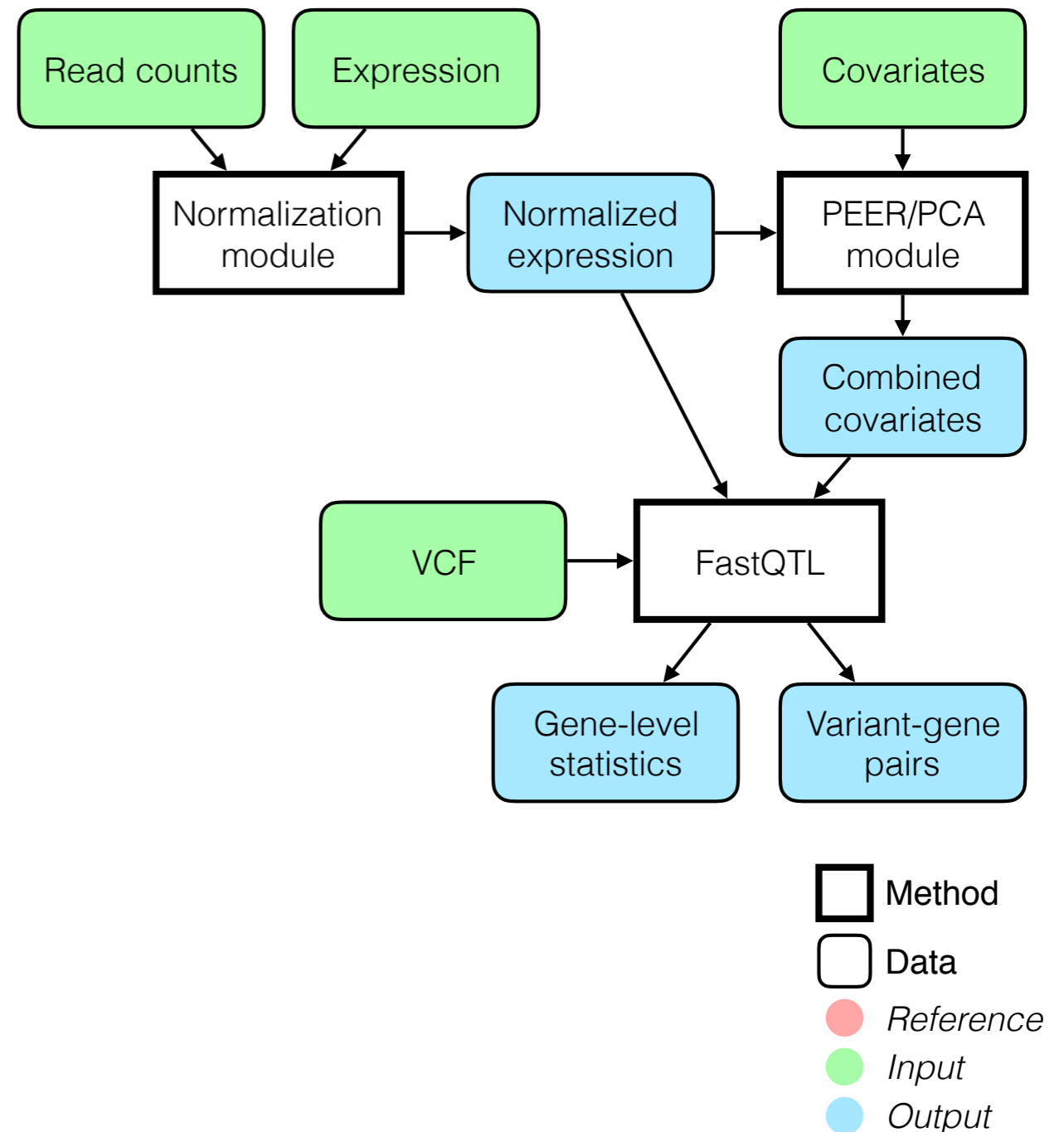


- Aligner: STAR v2.4.2a
- Gene expression: RNA-SeQC v1.1.9
- Transcript expression: RSEM v1.2.22
- QC metrics: RNA-SeQC v1.1.9

STAR: Dobin et al., *Bioinformatics*, 2013
RSEM: Li et al., *Bioinformatics*, 2010
RNA-SeQC: DeLuca et al., *Bioinformatics*, 2012

eQTL discovery

- QTL mapper: FastQTL
- Covariate correction:
 - PEER factors
 - Explicit covariates: Genotype PCs, gender
- *cis* window: $\pm 1\text{Mb}$
- $\text{MAF} \geq 0.01$ and ≥ 10 samples containing minor allele



Public release of pipelines on FireCloud

- Cloud-based genomics analysis platform developed at the Broad Institute: <http://firecloud.org>
 - Part of the NCI Cloud Pilot initiative; currently hosts TCGA data.
- Several GTEx pipelines already implemented (RNA-seq and eQTL); public release is imminent.
 - Also available as Docker images.



The screenshot shows the FireCloud web interface for a workspace named 'broad-firecloud-gtex/gtex_eqtl_test_0616'. The workspace is in a 'Complete' state. The interface includes a 'Workspace Owner' section with the email 'francois@broadinstitute.org' and a 'Created By' section with the same email and the date 'June 11, 2016 12:44 AM'. The 'Description' section is empty. The 'Workspace Attributes' section lists several attributes and their values, all pointing to 'gs://firecloud-gtex-projec': 'annotation_gtf', 'variant_lookup', 'genotype_pcs', 'vcf_index', 'explicit_covariates', and 'vcf'. The 'Analysis Submissions' section shows '8 Submissions' and '8 Done'. The interface also includes a 'Summary' tab and a 'Method Repository' link.

Outlook: planned changes/additions for v8 release

- Realignment/quantification to hg38/GRCh38 (+ latest GENCODE release) using FireCloud
- Re-evaluation of isoform quantification methods
- Small RNA-seq pipeline
(alignment, QC, quantification)
- FireCloud will facilitate collaborating on pipelines
(Docker-based).
Let us know if you're interested in contributing!

Acknowledgments



LDACC

K. Ardlie, G. Getz, A. Segrè, T. Sullivan, X. Li

E. Gelfand, C. Trowbridge

D. MacArthur, M. Kellis, J. Hirschhorn

Genomics Platform

GTEEx Portal

J. Nedzel, K. Huang, K. Hadley,

S. Meier, M. Noble



The GTEEx Project Consortium

Benchmarking Subgroup

eQTL Subgroup

Transcriptome Subgroup

Gender Subgroup



The Common
Fund

***Donors and
their families***