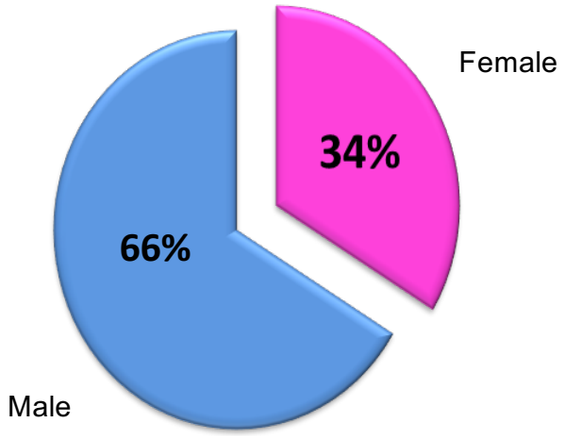
2016 GTEx mtg Notes

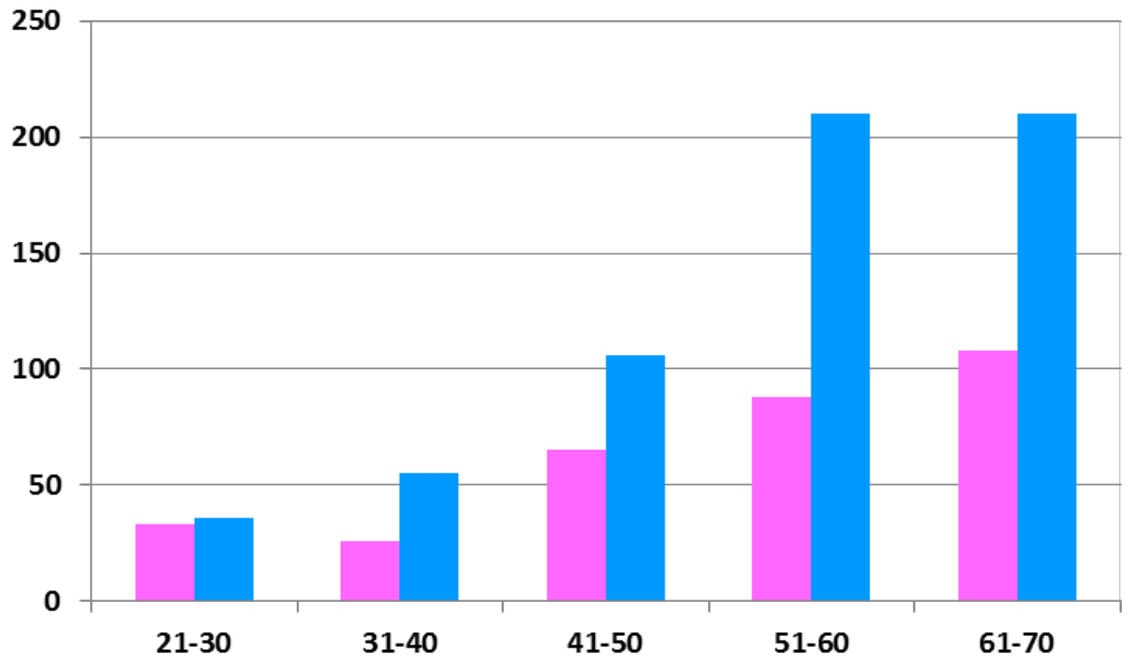
2016-07-11

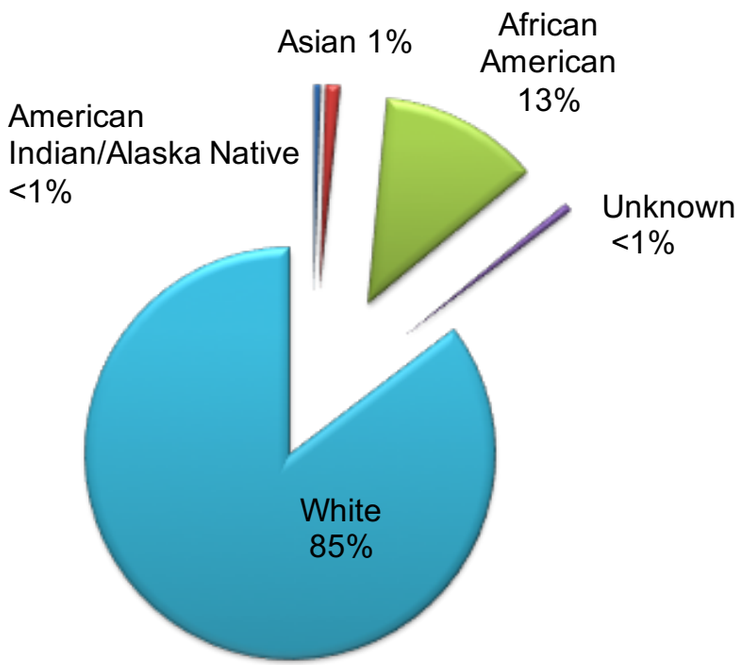
Project overview

- **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 samles, 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**):

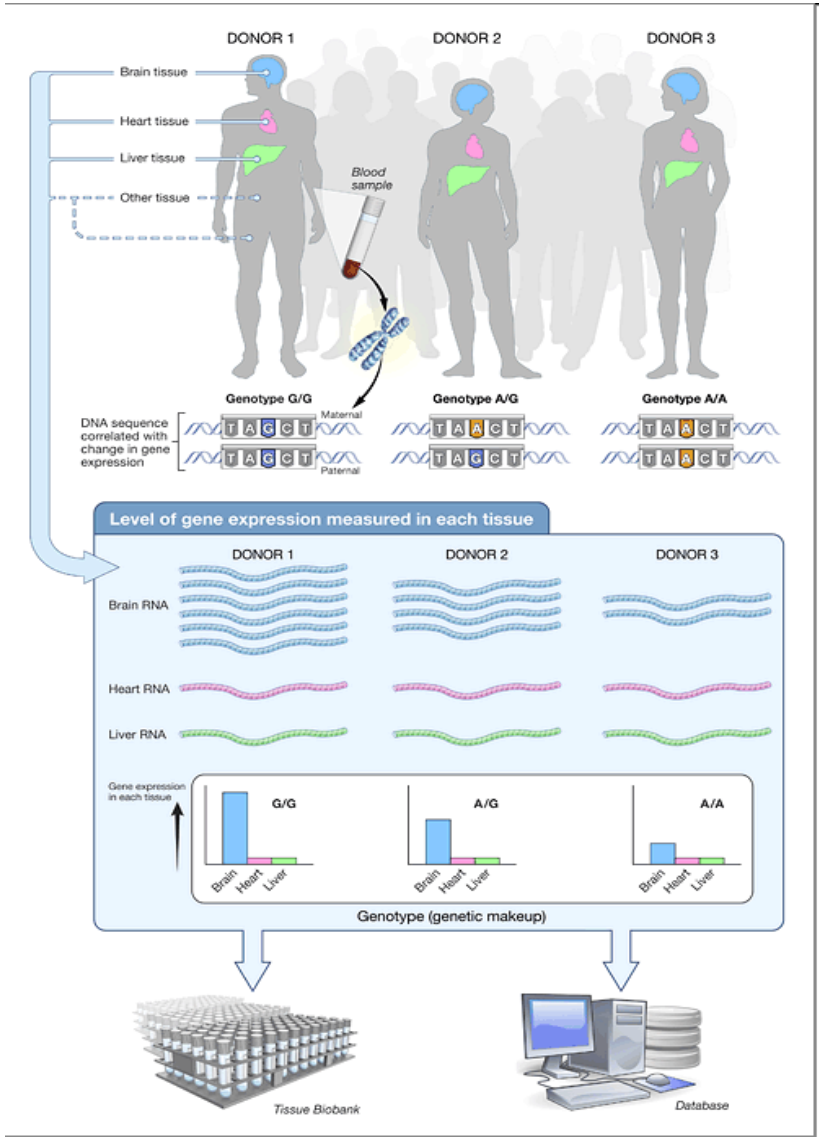
Francois Aguet (LDACC Broad)
Kristin Ardlie
Max Haeussler (UCSC; Genome Browser Engineer)
Su Koester (NIMH)
Jared Nedzel (GTEx, Broad)
Kate Rosenbloom (UCSC)
Ayellet Segre (LDACC Broad)
Casandra Trowbridge (LDACC Broad)
Simona Volpis (NHGRI)
Daniel Zerbino (ensembl)

3rd GTEx Community Meeting

JULY 11TH, 2016

Stanford University

GTEx = Genotype-Tissue Expression



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

GTEx 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:

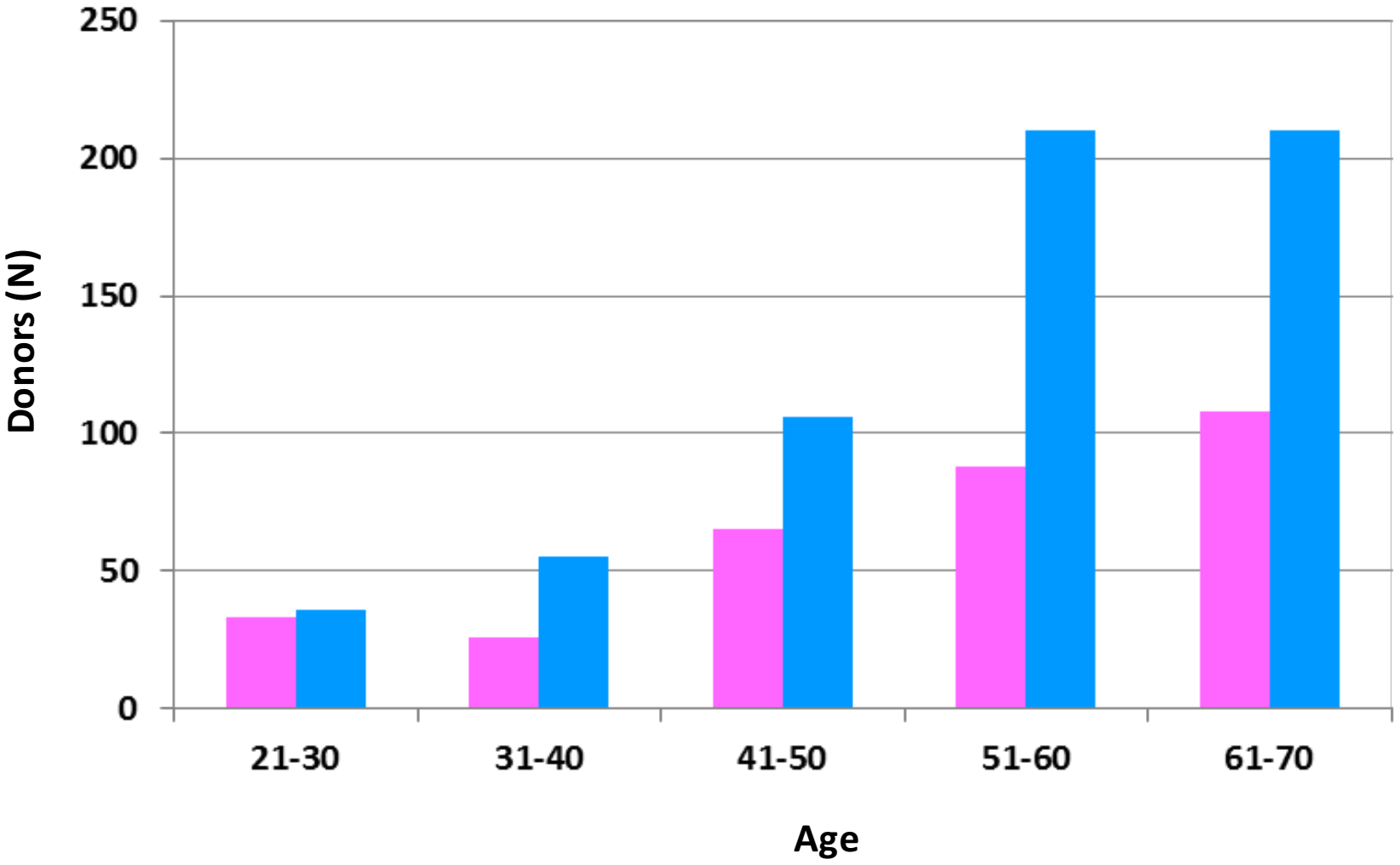
- ~960 Postmortem Donors
- WES & WGS
- RNA-Seq of ~30 tissues/donor (>20,000 tissues)
- Beyond Gene Exp

NIH The Common Fund

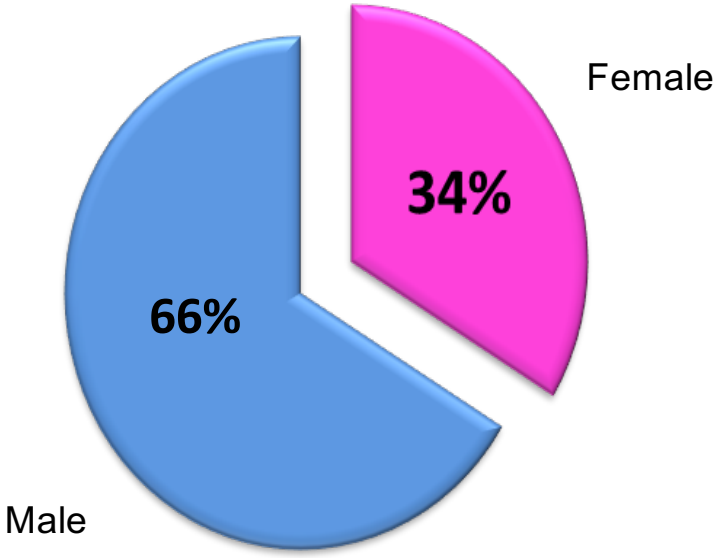
National Human Genome Research Institute

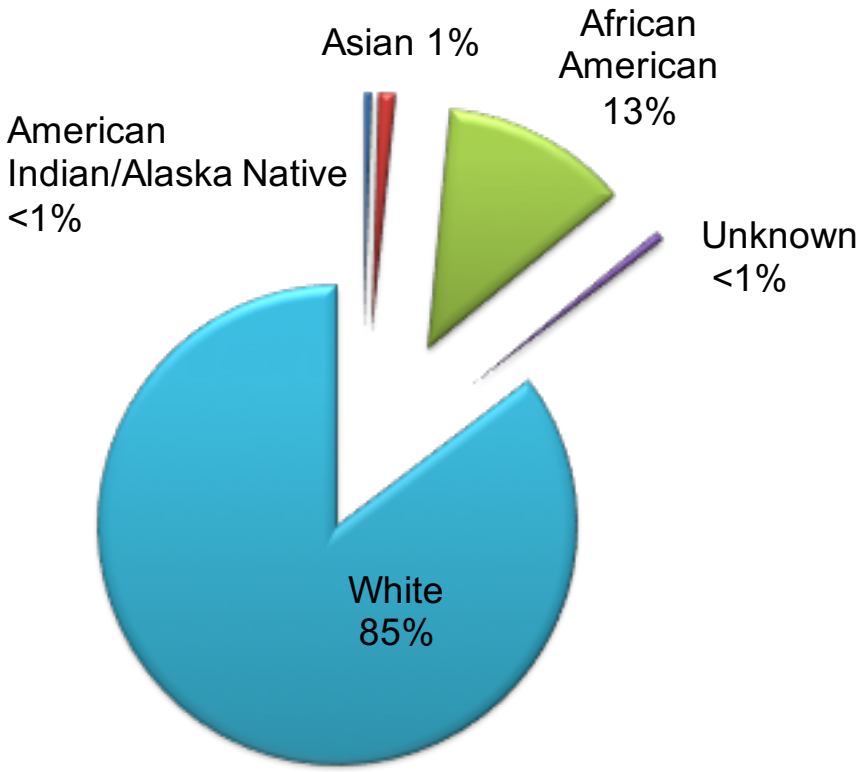
NIH National Institute of Mental Health

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%
1	4%
2	25%
3	6%
4	12%

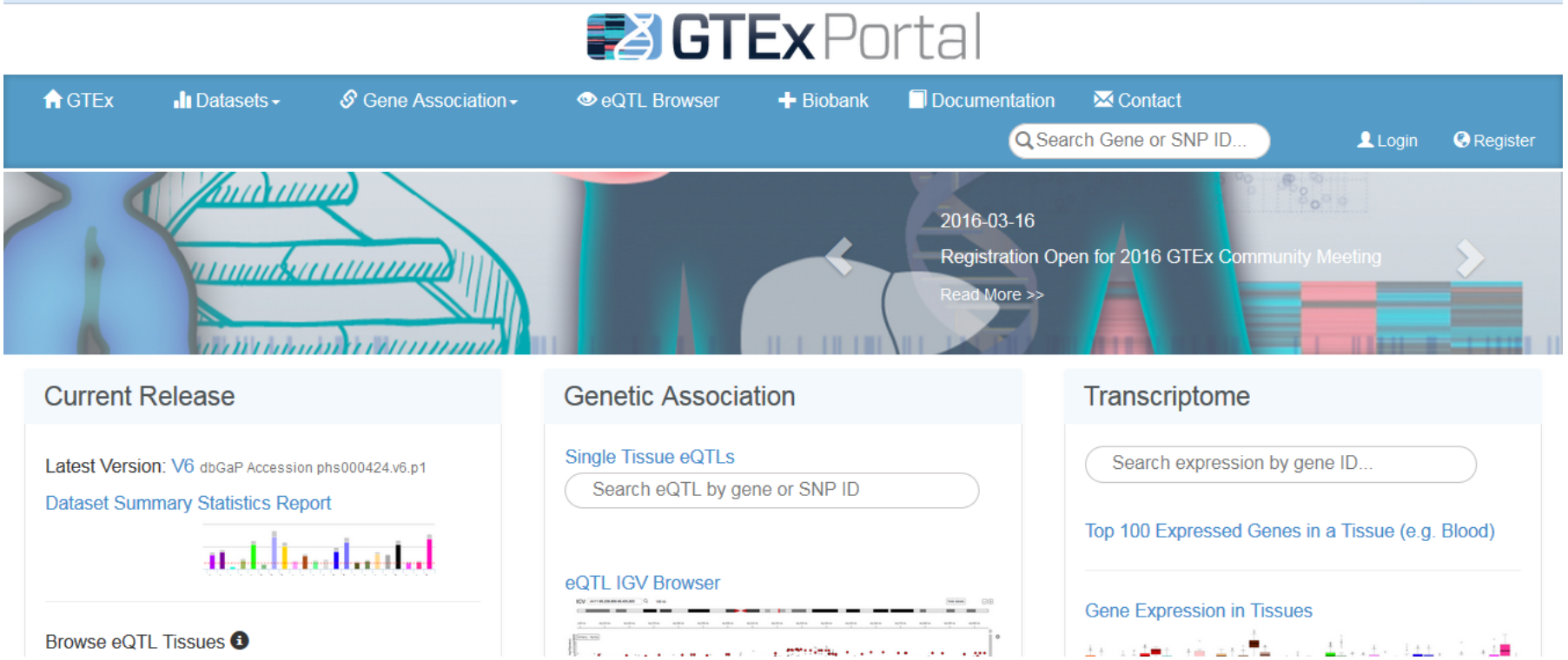
Low PMI

Brain donors

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

GTEx resources

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



Jared Nedzel for the GTEx portal

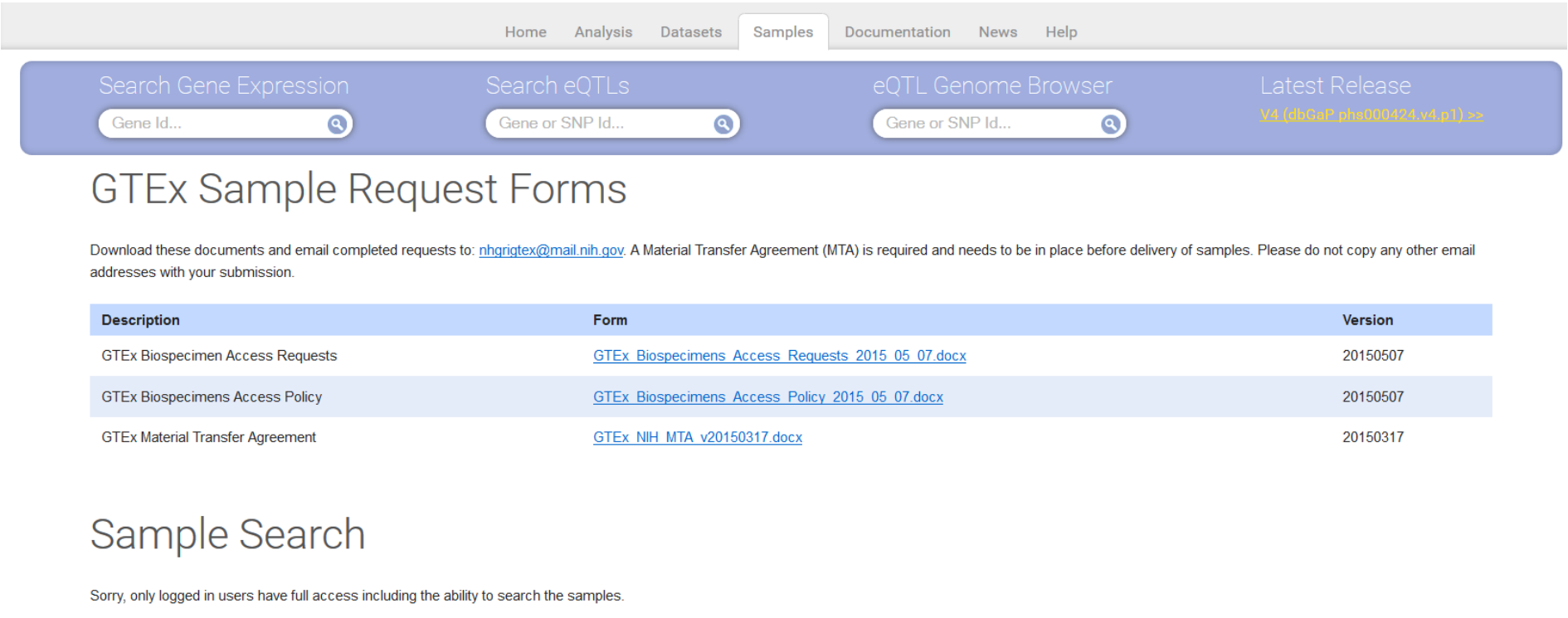
Daniel Zerbino for ENSEMBLE

GTEx resources cont.

dbGaP: controlled access of comprehensive GTEx clinical and raw sequencing data: http://www.ncbi.nlm.nih.gov/gap

Sample Access

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

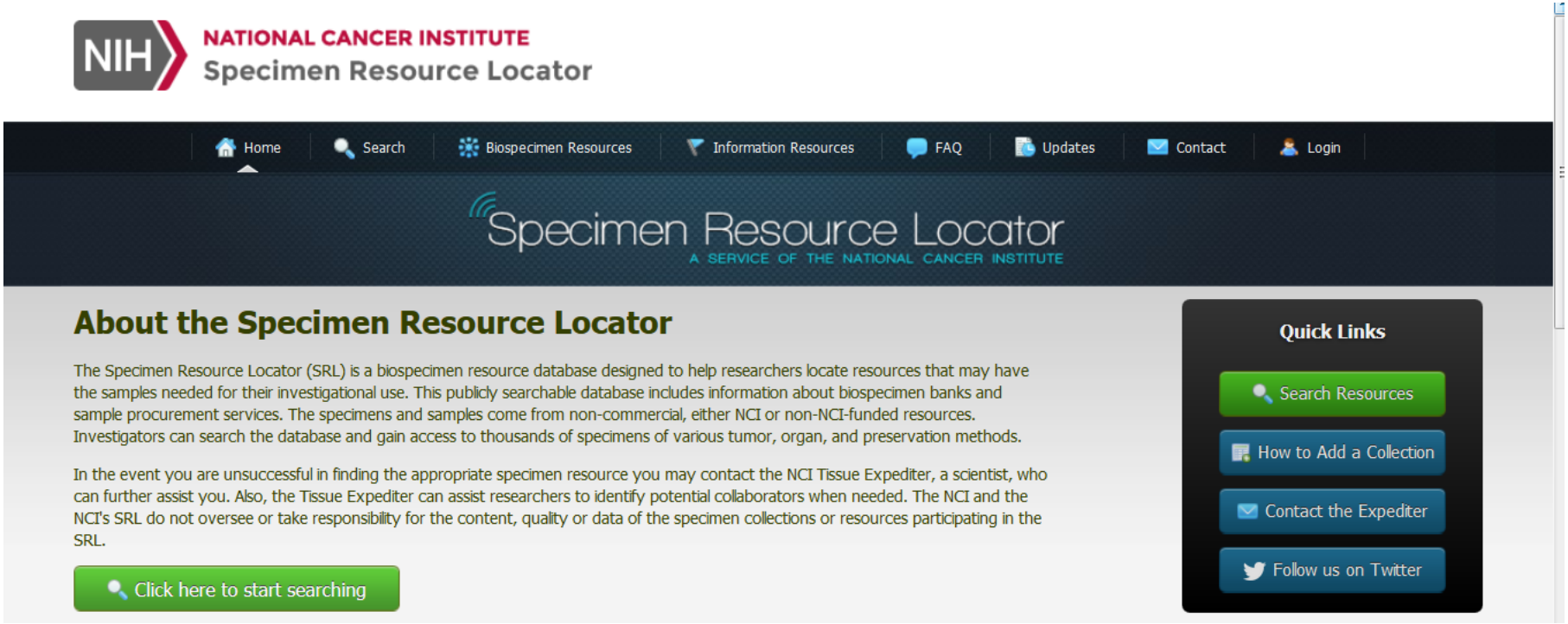


Available Biospecimens

- 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/



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.

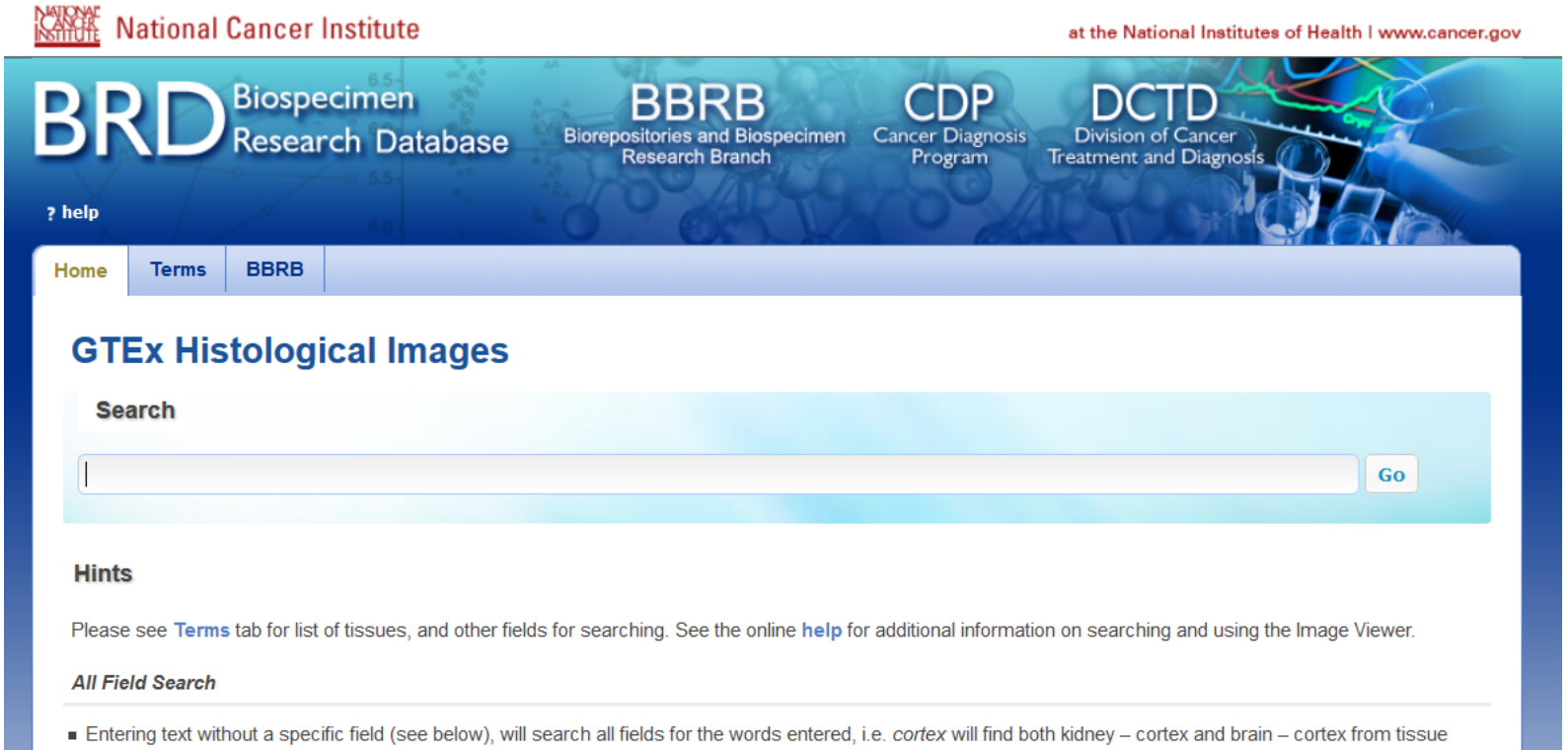
GTEx biospecimen collections SOPs

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

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

GTEx histological image viewer

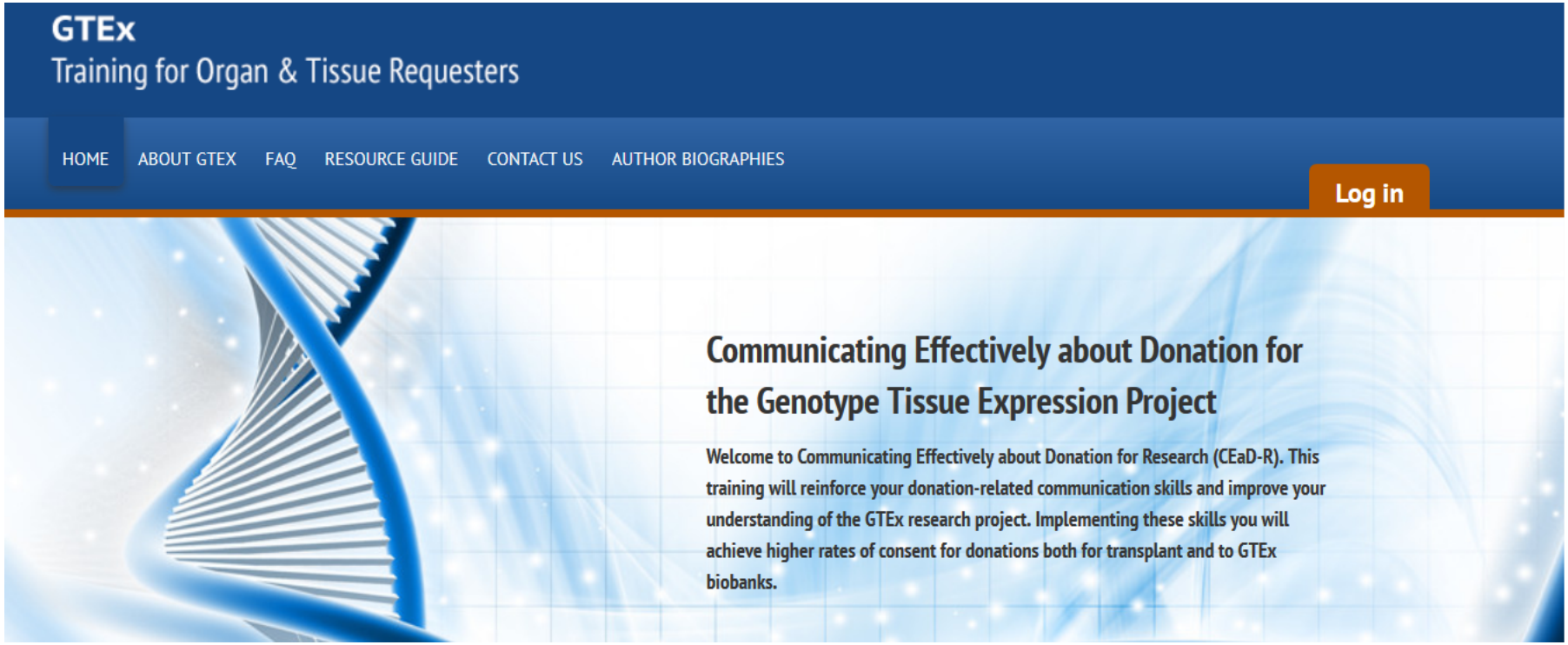
http://biospecimens.cancer.gov/resources/tissue_image_library.asp



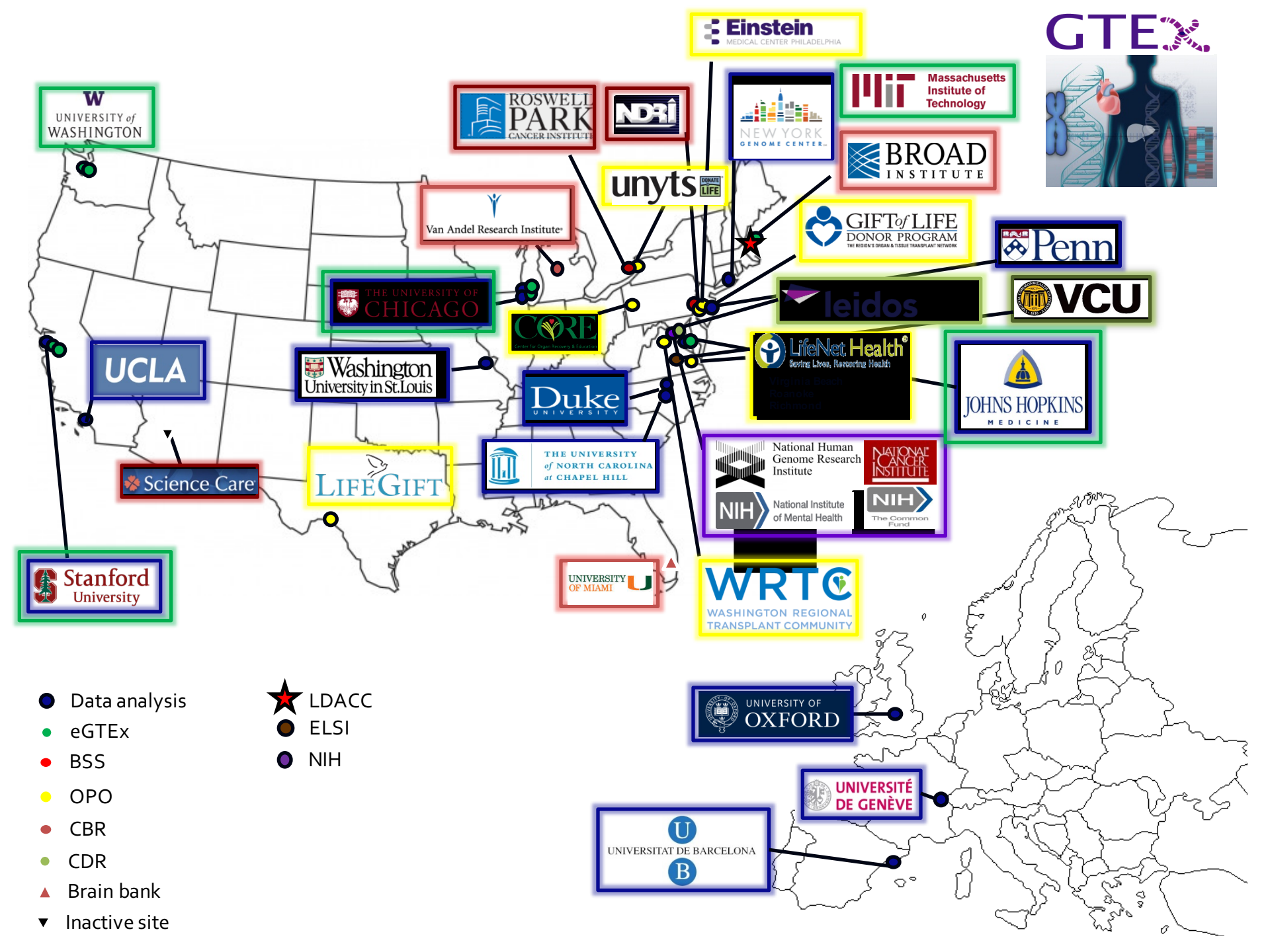
There are numerous search options for a specific field search (tissue type, autolysis score, gender, acceptability, etc). No software is required, and the images can be viewed with zooming capability.

Training videos for consenting personnel

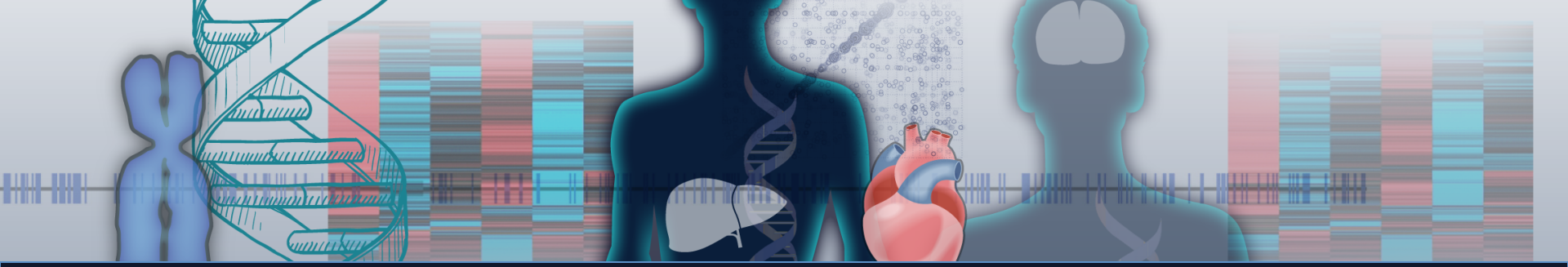
http://gtextraining.org/



Designed to help requesters communicate effectively about donation for GTEx.


GTEx
Einstein
MEDICAL CENTER PHILADELPHIA
Massachusetts Institute of Technology
UNIVERSITY of WASHINGTON
ROSWELL PARK CANCER INSTITUTE
NDRI
NEW YORK GENOME CENTER
BROAD INSTITUTE
unyts
Van Andel Research Institute
GIFT of LIFE DONOR PROGRAM
Penn
THE UNIVERSITY OF CHICAGO
leidos
VCU
CORE
UCLA
Washington University in St. Louis
LifeNet Health
Saving Lives, Restoring Health
JOHNS HOPKINS MEDICINE
Duke UNIVERSITY
Science Care
LIFEGIFT
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
National Human Genome Research Institute
NATIONAL CANCER INSTITUTE
NIH
National Institute of Mental Health
NIH The Common Fund
Stanford University
UNIVERSITY OF MIAMI
WRTC
WASHINGTON REGIONAL TRANSPLANT COMMUNITY
UNIVERSITY OF OXFORD
UNIVERSITÉ DE GENÈVE
UNIVERSITAT DE BARCELONA
Data analysis
eGTEx
BSS
OPO
CBR
CDR
Brain bank
Inactive site
LDACC
ELSI
NIH

Thank you!



The GTEx LDACC Project Update

Kristin Ardlie, Ph.D.

GTEx LDACC, Broad Institute

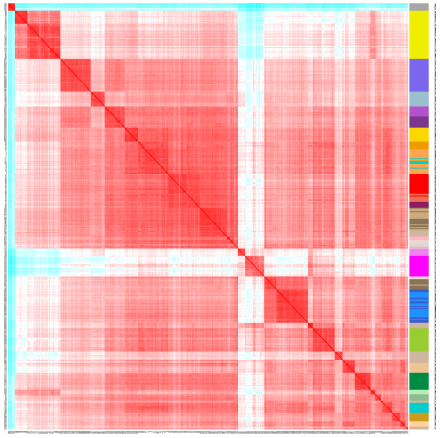
July 11, 2016

GTEx

BROAD INSTITUTE

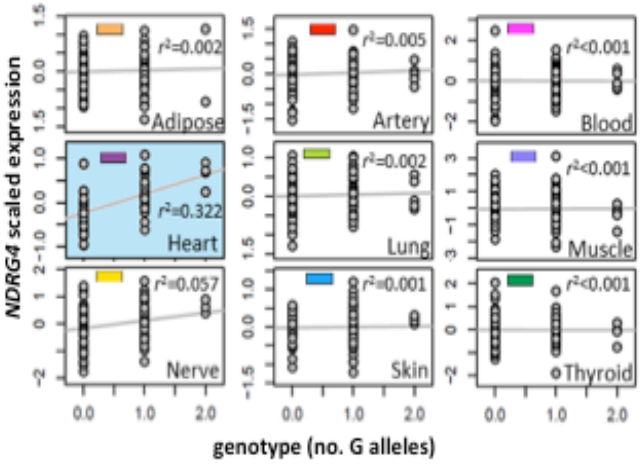
GTEx Project Goals

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



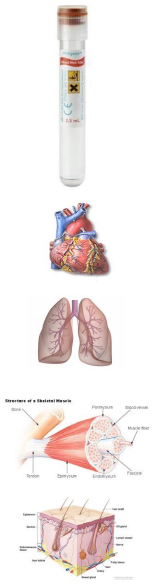
Scope - Primary Data Types

- 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



Tissue types

Within tissue - across population.

Within individuals – across tissues.

GTEx Project Timeline

(960)

Donor Collection **COMPLETE** (960)

Tissue processing (~880)

RNA-seq (~16,000)

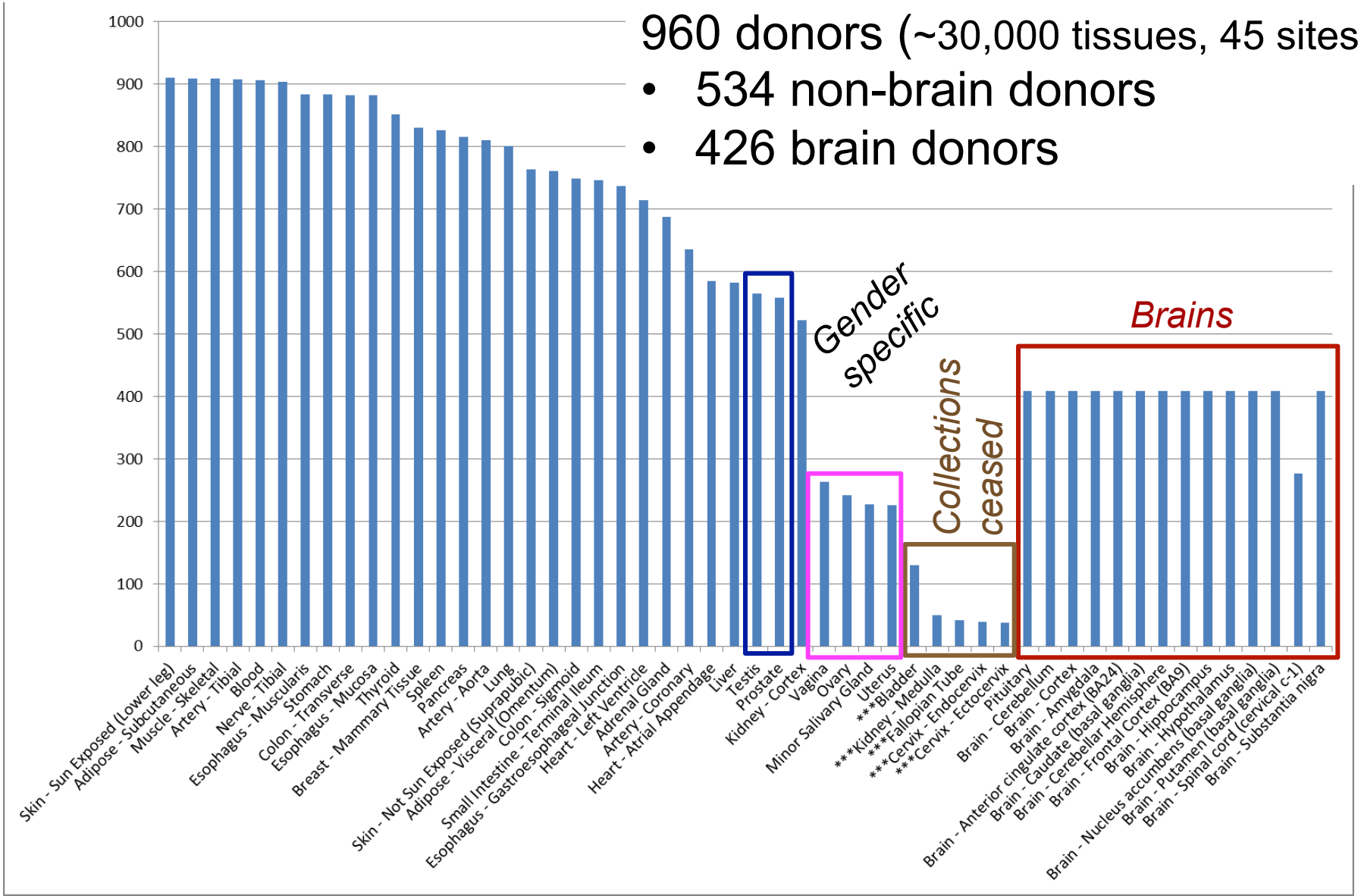
V6 data release and midpoint AWG manuscripts in prep for publication Dec, 2016 (**450 donors, ~7400 RNA-seq**)

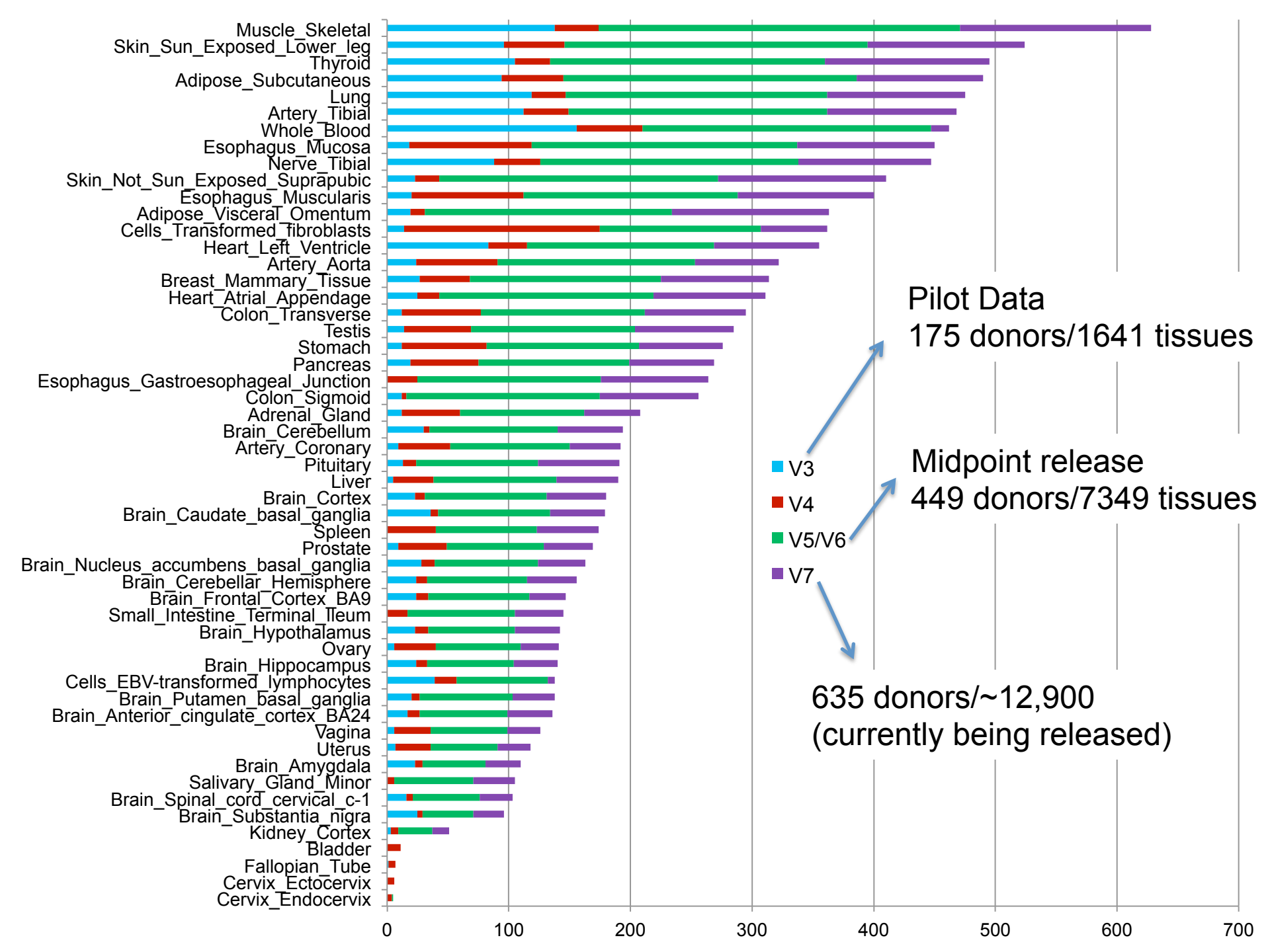
Pilot Analysis - multiple publications *(May 8, 2015)*

Tissue collection variable by site

960 donors (~30,000 tissues, 45 sites)

- 534 non-brain donors
- 426 brain donors



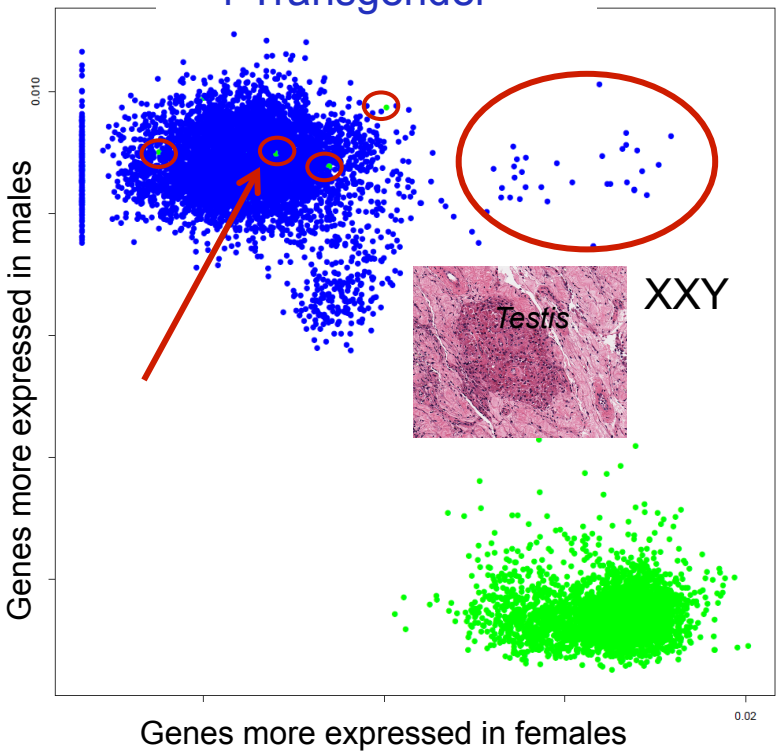

Muscle_Skeletal
Skin_Sun_Exposed_Lower_leg
Thyroid
Adipose_Subcutaneous
Lung
Artery_Tibial
Whole_Blood
Esophagus_Mucosa
Nerve_Tibial
Skin_Not_Sun_Exposed_Suprapubic
Esophagus_Muscularis
Adipose_Visceral_Omentum
Cells_Transformed_fibroblasts
Heart_Left_Ventricle
Artery_Aorta
Breast_Mammary_Tissue
Heart_Atrial_Appendage
Colon_Transverse
Testis
Stomach
Pancreas
Esophagus_Gastroesophageal_Junction
Colon_Sigmoid
Adrenal_Gland
Brain_Cerebellum
Artery_Coronary
Pituitary
Liver
Brain_Cortex
Brain_Caudate_basal_ganglia
Spleen
Prostate
Brain_Nucleus_accumbens_basal_ganglia
Brain_Cerebellar_Hemisphere
Brain_Frontal_Cortex_BA9
Small_Intestine_Terminal_Ileum
Brain_Hypothalamus
Ovary
Brain_Hippocampus
Cells_EBV-transformed_lymphocytes
Brain_Putamen_basal_ganglia
Brain_Anterior_cingulate_cortex_BA24
Vagina
Uterus
Brain_Amygdala
Salivary_Gland_Minor
Brain_Spinal_cord_cervical_c-1
Brain_Substantia_nigra
Kidney_Cortex
Bladder
Fallopian_Tube
Cervix_Ectocervix
Cervix_Endocervix
0
100
200
300
400
500
600
700
V3
V4
V5/V6
V7
Pilot Data
175 donors/1641 tissues
Midpoint release
449 donors/7349 tissues
635 donors/~12,900
(currently being released)

Data Release Categories

Raw Data

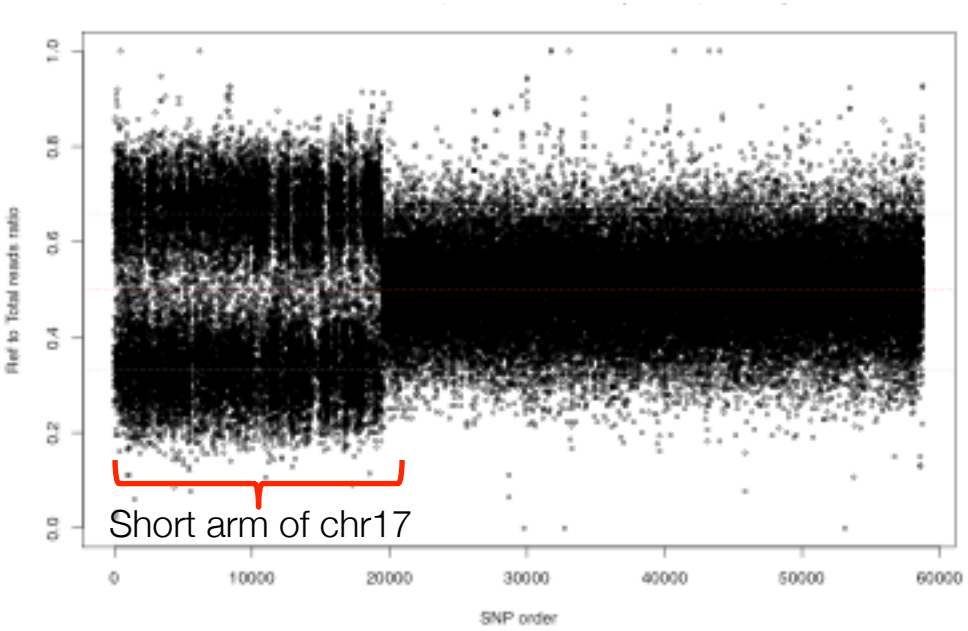
All raw data (WEX, WGS, RNA-seq) passing QC are released to dbGaP. Includes duplicates and samples EXCLUDED from analysis freezes (e.g. phenotypic, clinical exclusions)

- 4 Kleinfelter (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

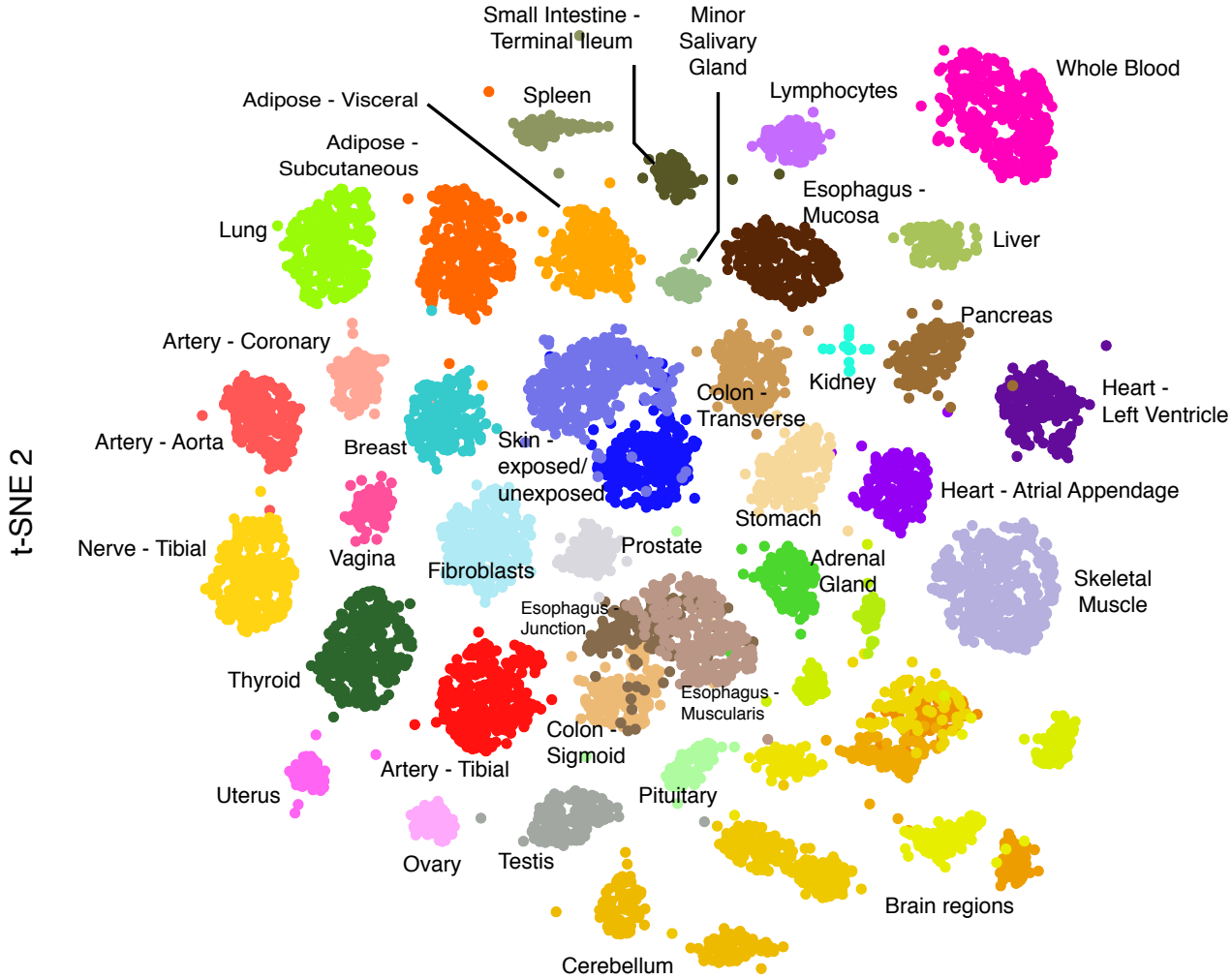
Analysis Freeze - eQTL

Available on GTEx Portal

Requires both genotype and expression data

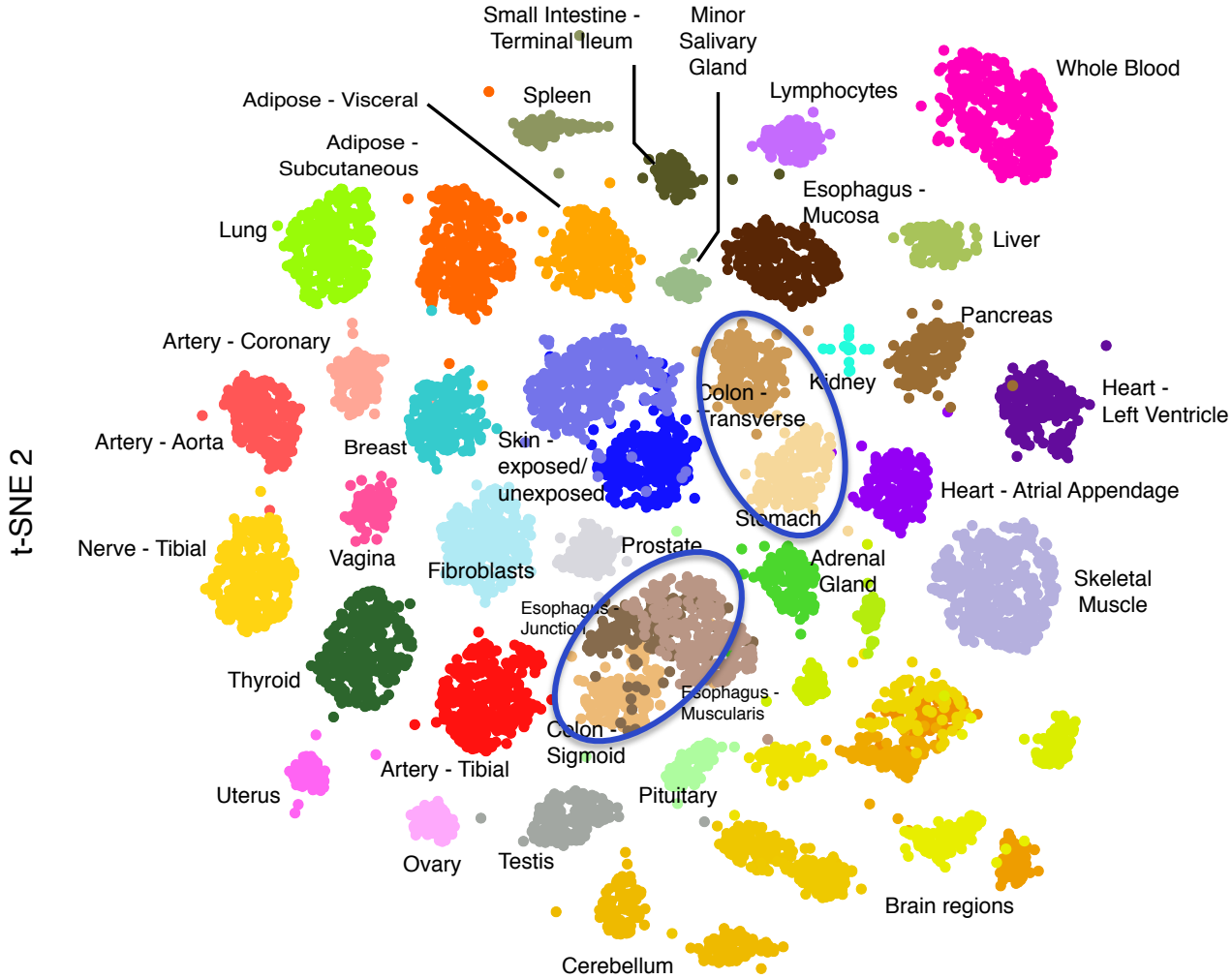
EXCLUDES Expression Data from samples without matching donor genotype data

Production Data 1 - GTEx Transcriptome



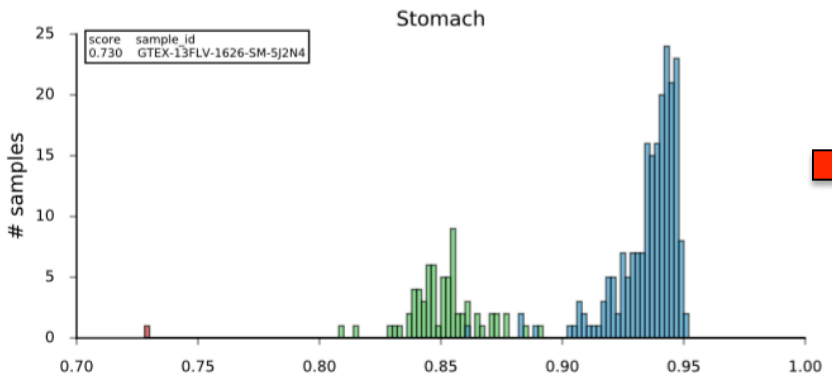
François Aguet

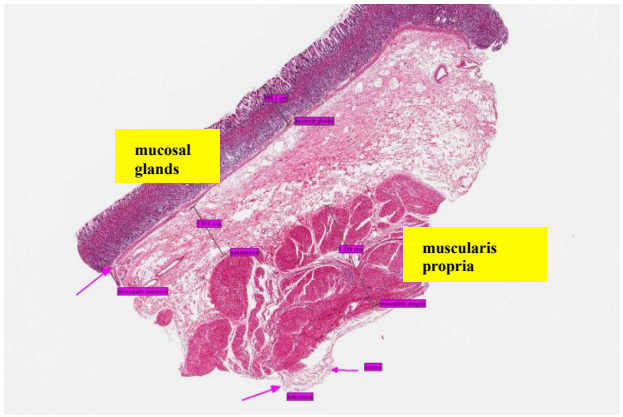
Production Data 1 - GTEx Transcriptome



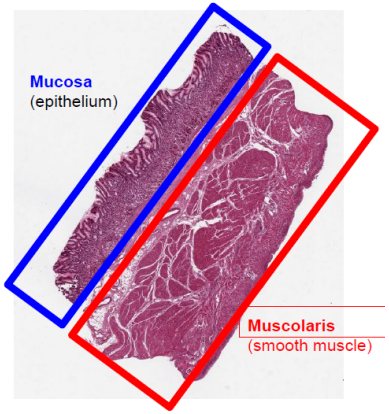
François Aguet

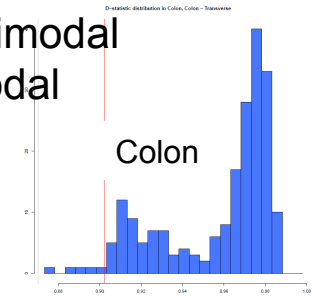
Sample Heterogeneity - LCM





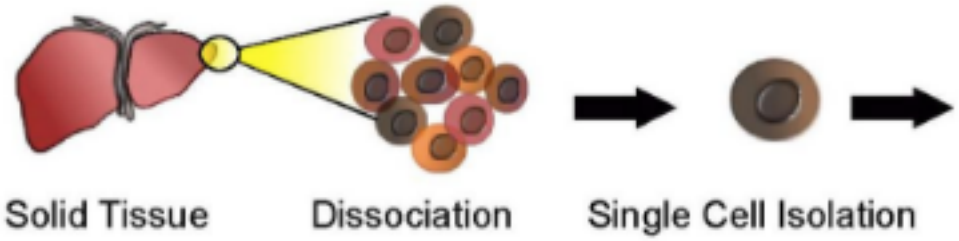
LCM – 30 tissues,
90 samples





1. Stomach – COMPLETE ✔
2. Colon, transverse – also bimodal
3. Terminal ileum – also bimodal
4. Pancreas
5. Skin
6. Liver

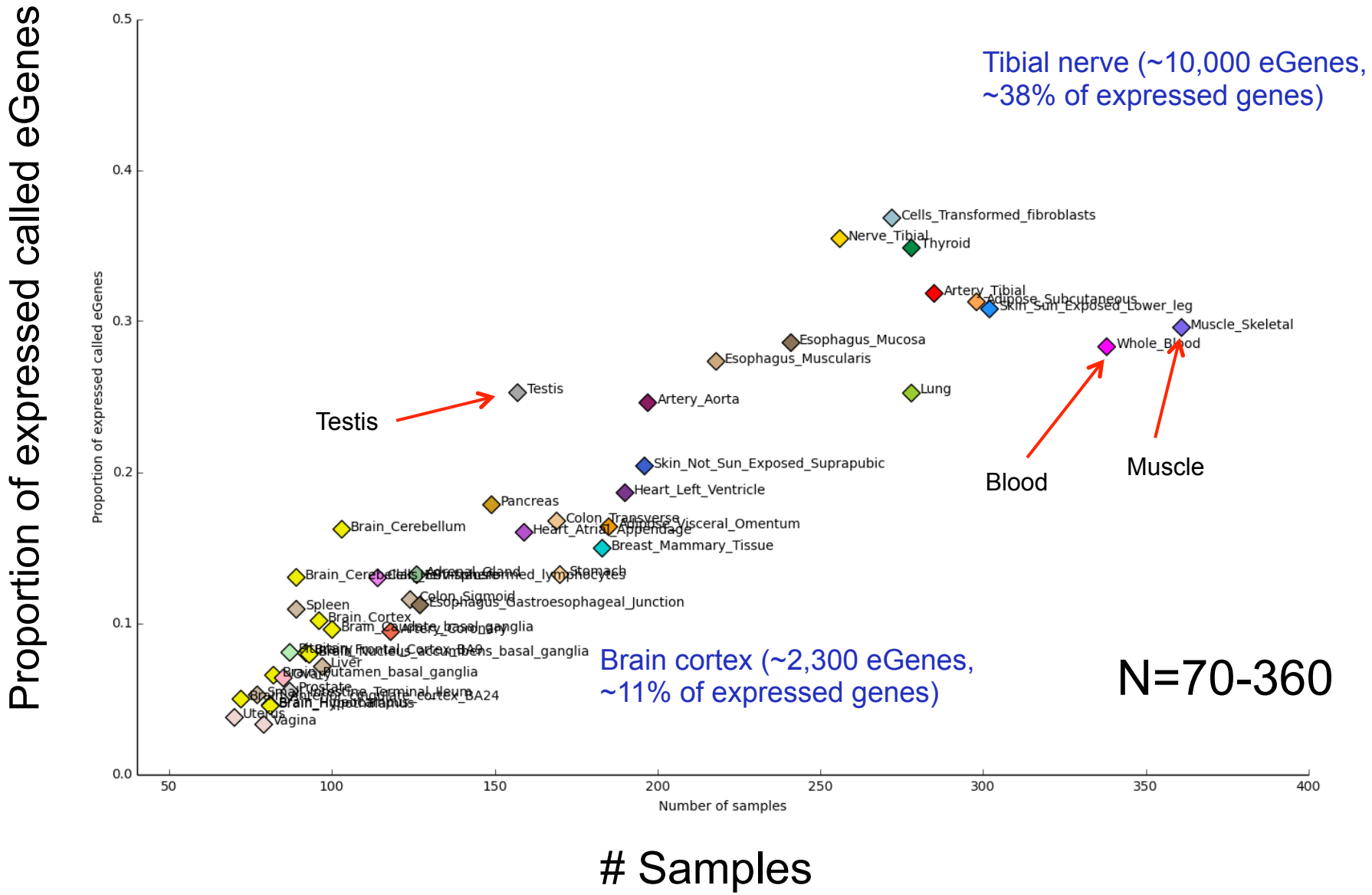
Sample Heterogeneity – Single Cell



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

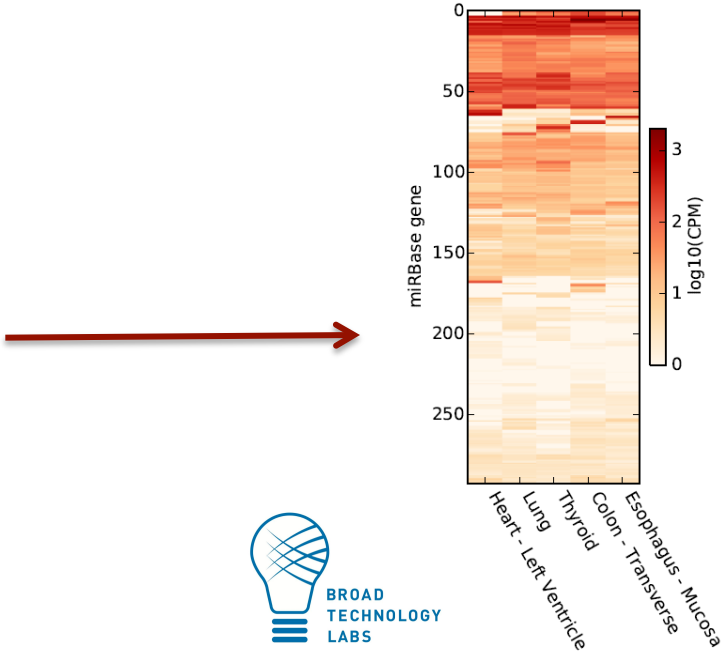
Production Data 2 - *cis*-eQTLs (eGenes)



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



BROAD TECHNOLOGY LABS

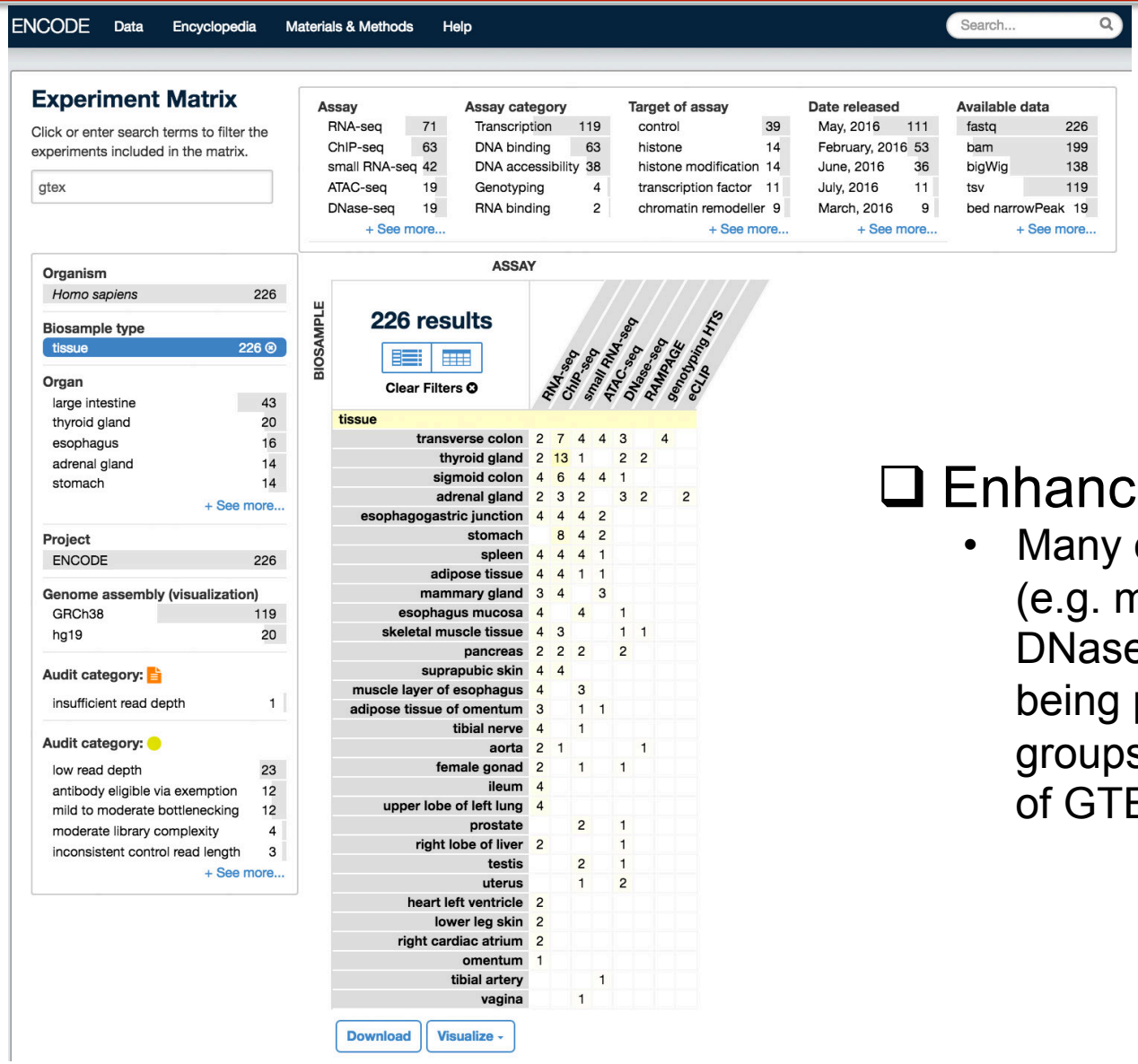
ENCODE and GTEx – "ENTEx"

	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:

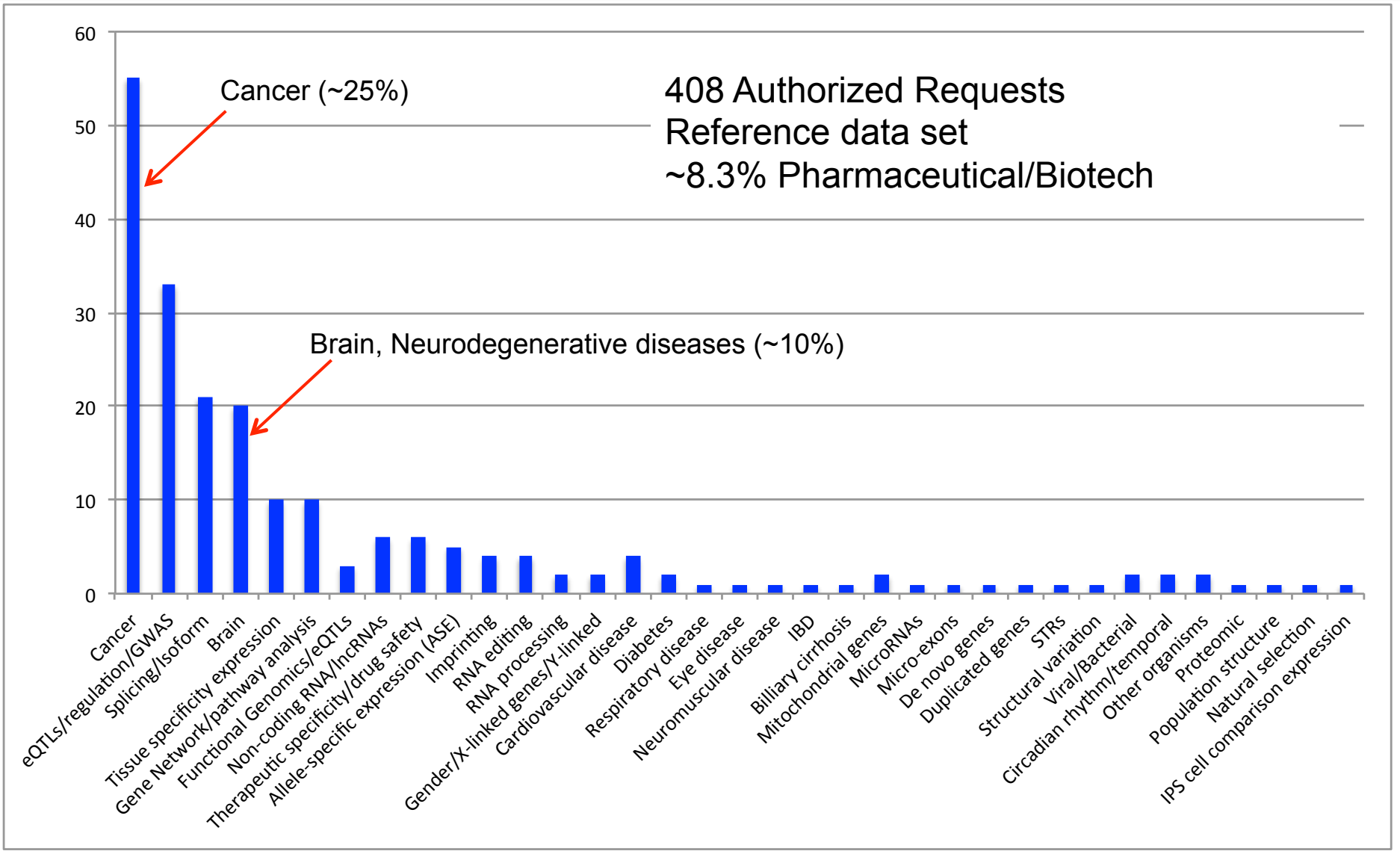
- GTEx collections for ENCODE
- 2 M + 2F
- All tissues, not brain

ENTEx Data available @ ENCODE



- 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

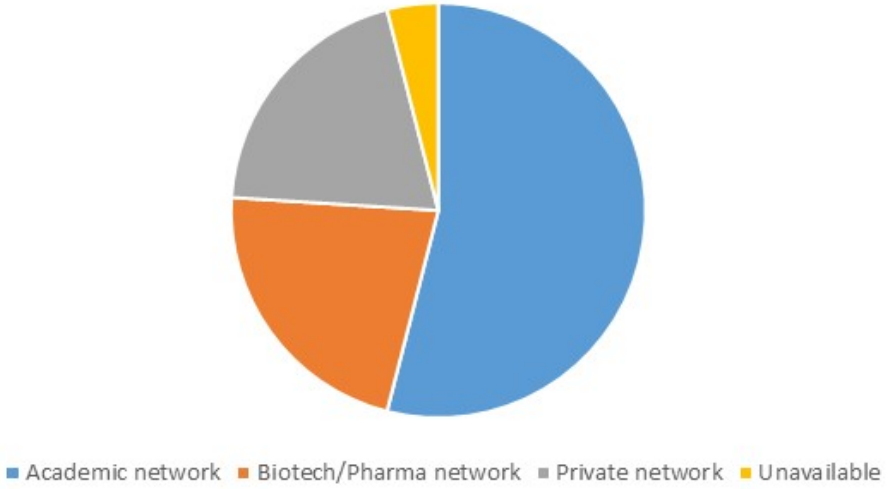
Data Access dbGaP (raw data)

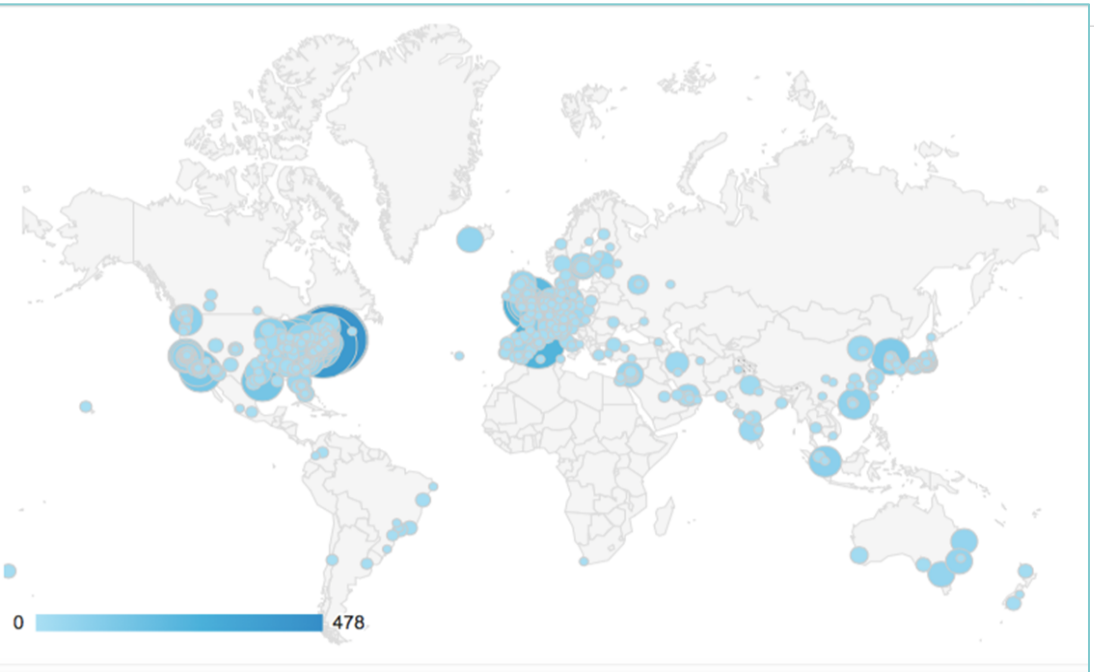


Data Access GTEx Portal

www.gtexportal.org

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





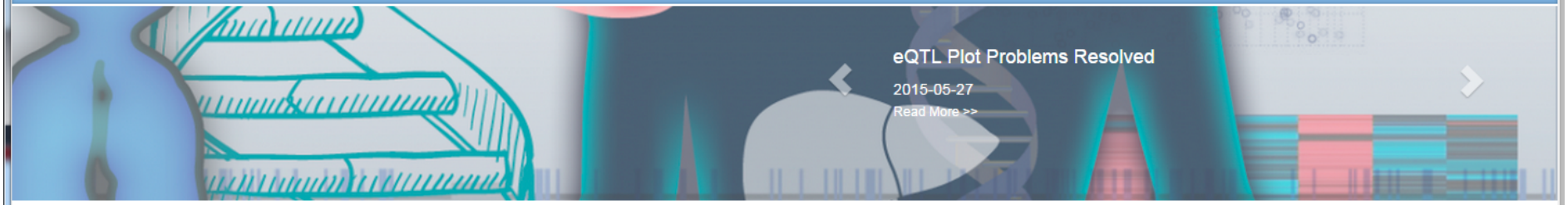
<u>GTEx Data also now integrated in to</u>:

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

www.gtexportal.org

GTEx Portal

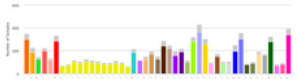
GTEx | Datasets | Gene Association | eQTL Browser | Biobank | Documentation | Contact | Search Gene or SNP ID... | Login | Register



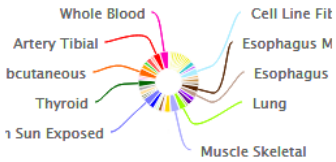
Current Release and Data Summary

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

- Summary Statistics



- Browse eQTL Tissues



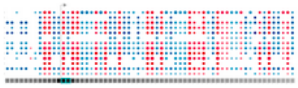
Genetic Association

- Single Tissue eQTLs

Search Gene or SNP

- eQTL IGV Browser

- Gene eQTL Visualizer



Search Gene...

- Test Your Own eQTLs

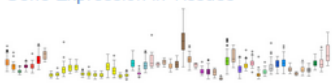
Analysis Results in Older Releases

- Multi-tissue eQTLs (V4 & V3)
- Splice QTLs (V3)
- Protein-truncating Variants (V3)
- Genomic Imprinting (V3)

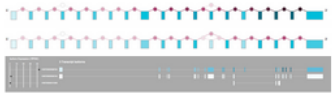
Transcriptome

Search Gene...

- Top 100 Expressed Genes in a Tissue (e.g. Blood)
- Gene Expression in Tissues



- Exon and Isoform Expression



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

Links

Documentation
- About GTEx
- Publication Policy
- Consortium Members

External Links
- dbGaP
- NIH Common Fund
- NHGRI

News

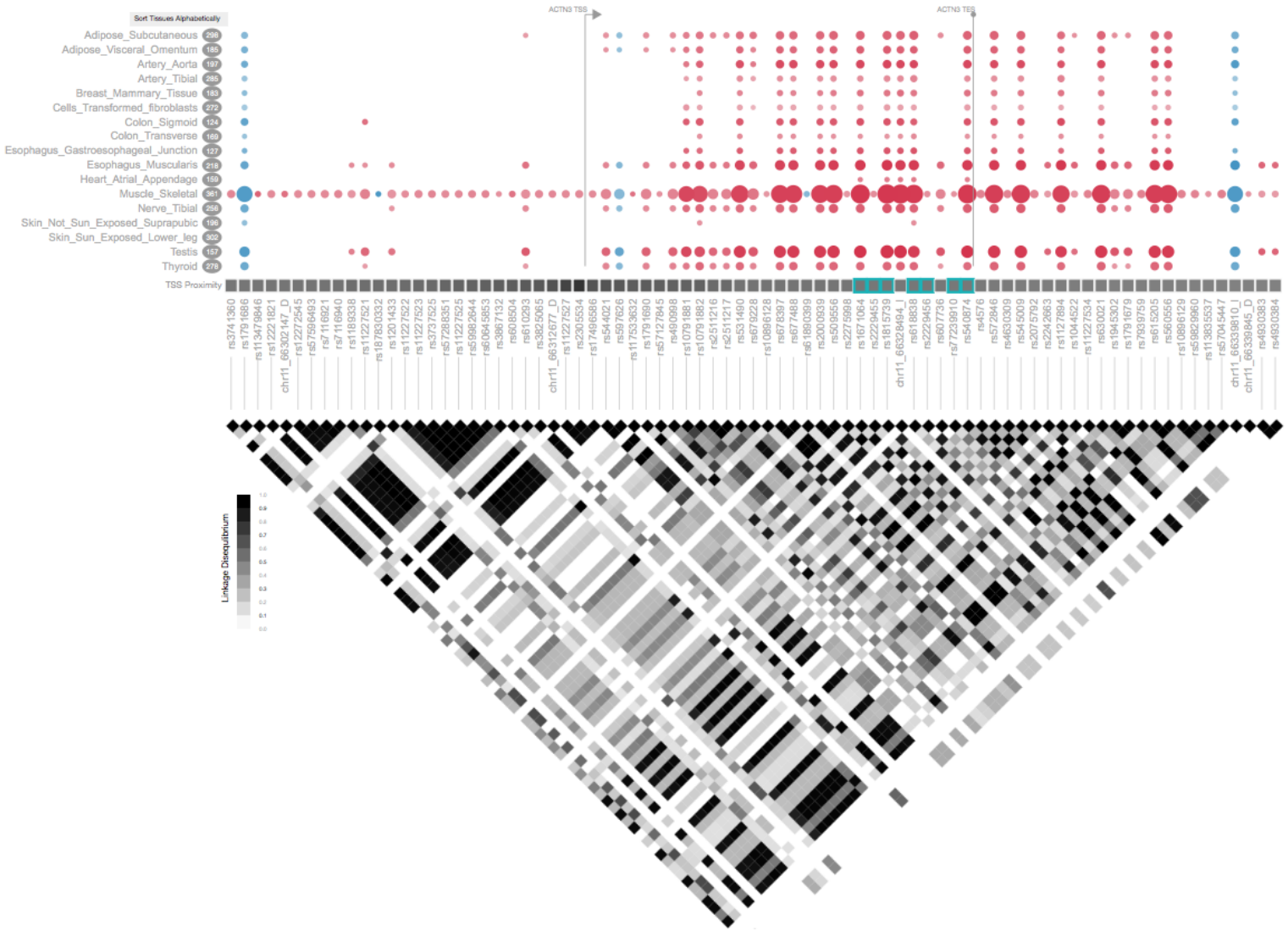
GTEx Portal Performance Issues
2015-09-22
Read More >>

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 >>

New Portal Features – LD track

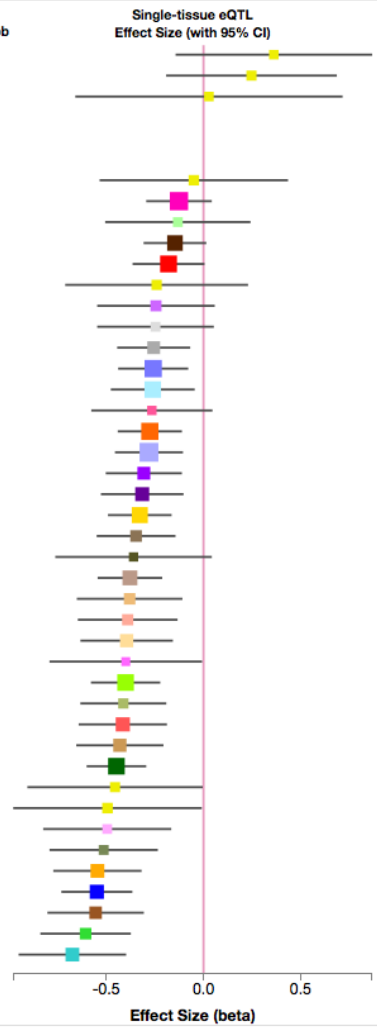


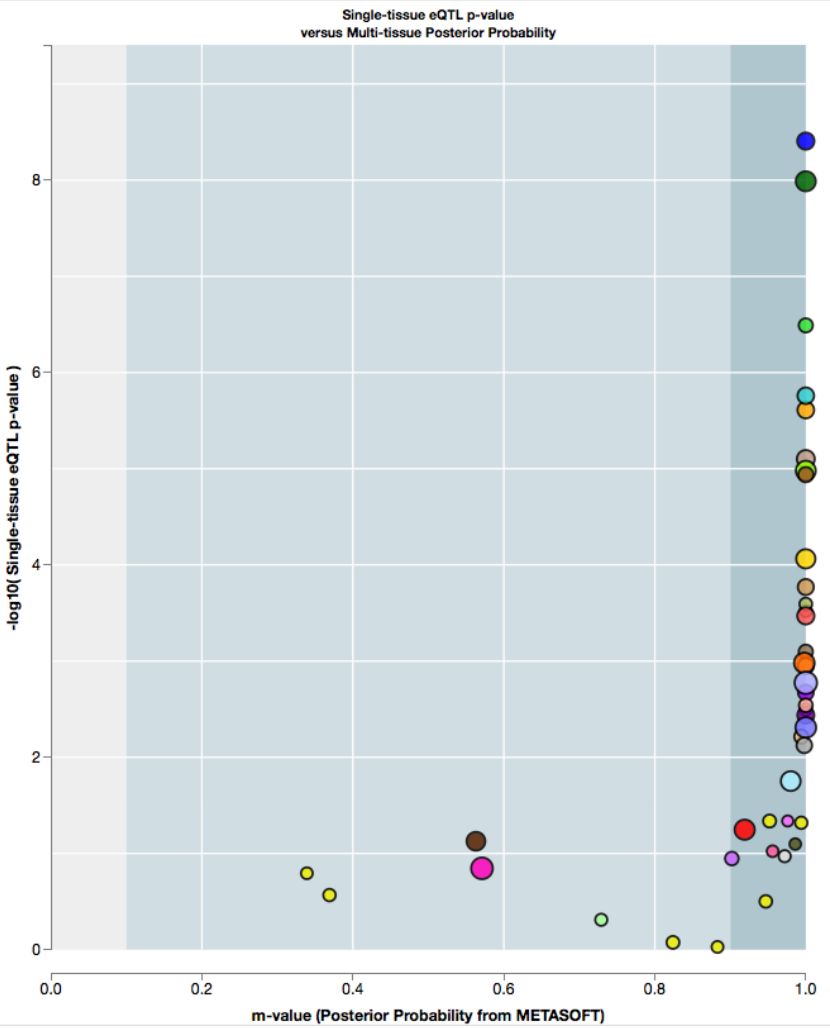
New Portal Features – Single & Multi-tissue eQTL comparison

Multi-tissue eQTL Comparison

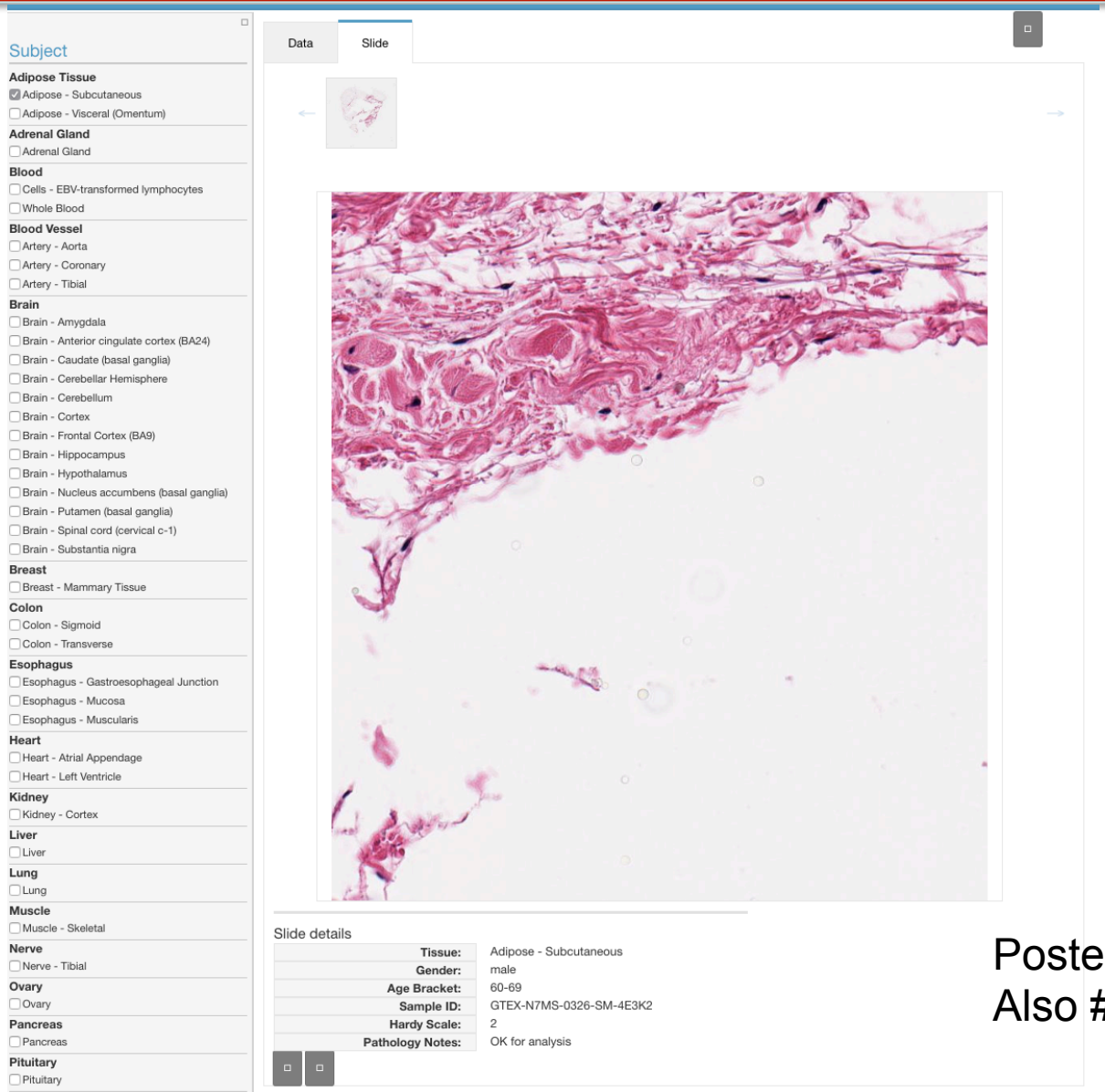
ENSG00000224956.5 RP11-206L10.1 and rs12096608 eQTL (Meta Analysis RE2 P-Value: 1.37159e-76)

Tissue	Samples	Beta	p-value	post-prob
Brain - Hippocampus	81	0.362	0.2	0.339
Brain - Nucleus accumbens (basal ganglia)	93	0.247	0.3	0.369
Brain - Putamen (basal ganglia)	82	0.0279	0.9	0.883
Brain - Anterior cingulate cortex (BA24)	72	-	-	-
Brain - Cerebellar Hemisphere	89	-	-	-
Brain - Hypothalamus	81	-	-	-
Brain - Caudate (basal ganglia)	100	-0.0498	0.8	0.824
Whole Blood	338	-0.126	0.1	0.571
Pituitary	87	-0.132	0.5	0.729
Esophagus - Mucosa	241	-0.147	0.07	0.563
Artery - Tibial	285	-0.180	0.06	0.919
Brain - Cortex	96	-0.241	0.3	0.947
Cells - EBV-transformed lymphocytes	114	-0.244	0.1	0.902
Prostate	87	-0.247	0.1	0.972
Testis	157	-0.256	0.008	0.998
Skin - Sun Exposed (Lower leg)	302	-0.259	4.9e-3	1.00
Cells - Transformed fibroblasts	272	-0.261	0.02	0.980
Vagina	79	-0.266	0.09	0.956
Adipose - Subcutaneous	298	-0.276	1.0e-3	0.998
Muscle - Skeletal	361	-0.280	1.7e-3	1.00
Heart - Atrial Appendage	159	-0.307	2.1e-3	1.00
Heart - Left Ventricle	190	-0.315	3.6e-3	1.00
Nerve - Tibial	256	-0.328	8.6e-5	1.00
Esophagus - Gastroesophageal Junction	127	-0.347	8.0e-4	1.00
Small Intestine - Terminal Ileum	77	-0.360	0.08	0.986
Esophagus - Muscularis	218	-0.378	7.9e-6	1.00
Colon - Sigmoid	124	-0.380	0.006	0.994
Artery - Coronary	118	-0.390	2.9e-3	1.00
Stomach	170	-0.395	1.1e-3	1.00
Uterus	70	-0.400	0.05	0.976
Lung	278	-0.401	1.0e-5	1.00
Liver	97	-0.412	2.5e-4	1.00
Artery - Aorta	197	-0.415	3.4e-4	1.00
Colon - Transverse	169	-0.430	1.7e-4	1.00
Thyroid	278	-0.448	1.0e-8	1.00
Brain - Frontal Cortex (BA9)	92	-0.455	0.05	0.994
Brain - Cerebellum	103	-0.494	0.05	0.952
Ovary	85	-0.495	3.2e-3	1.00
Spleen	89	-0.513	3.0e-4	1.00
Adipose - Visceral (Omentum)	185	-0.546	2.5e-6	1.00
Skin - Not Sun Exposed (Suprapubic)	196	-0.549	3.9e-9	1.00
Pancreas	149	-0.555	1.2e-5	1.00
Adrenal Gland	126	-0.606	3.2e-7	1.00
Breast - Mammary Tissue	183	-0.674	1.7e-6	1.00





New Portal Features - Histology



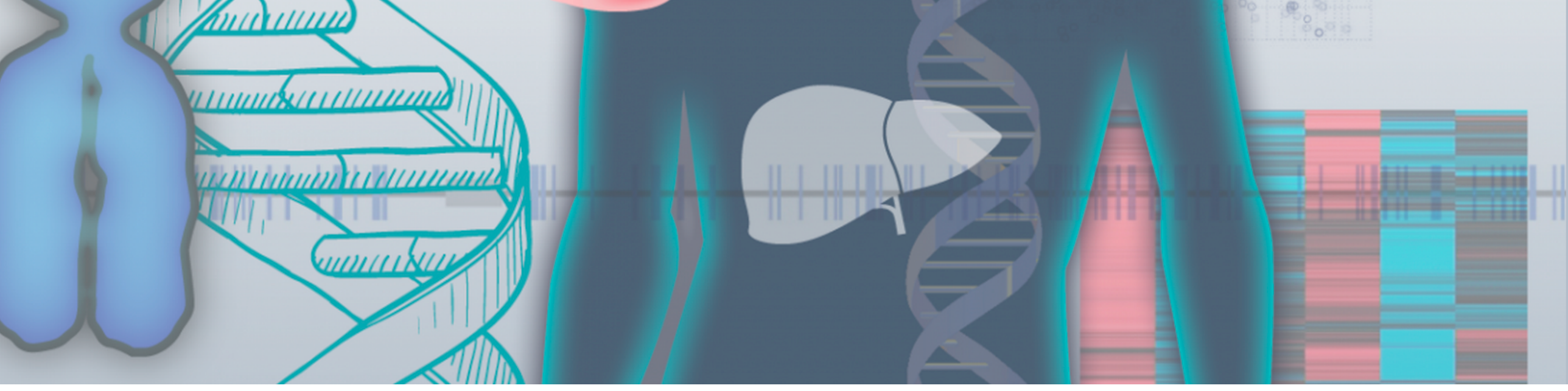
Poster #3 (Meier)
Also #4 (Qi)

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



The v7 data release

GTEx Community Meeting :: 07/11/2016

François Aguet, Broad Institute

BROAD INSTITUTE

GTEx

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]

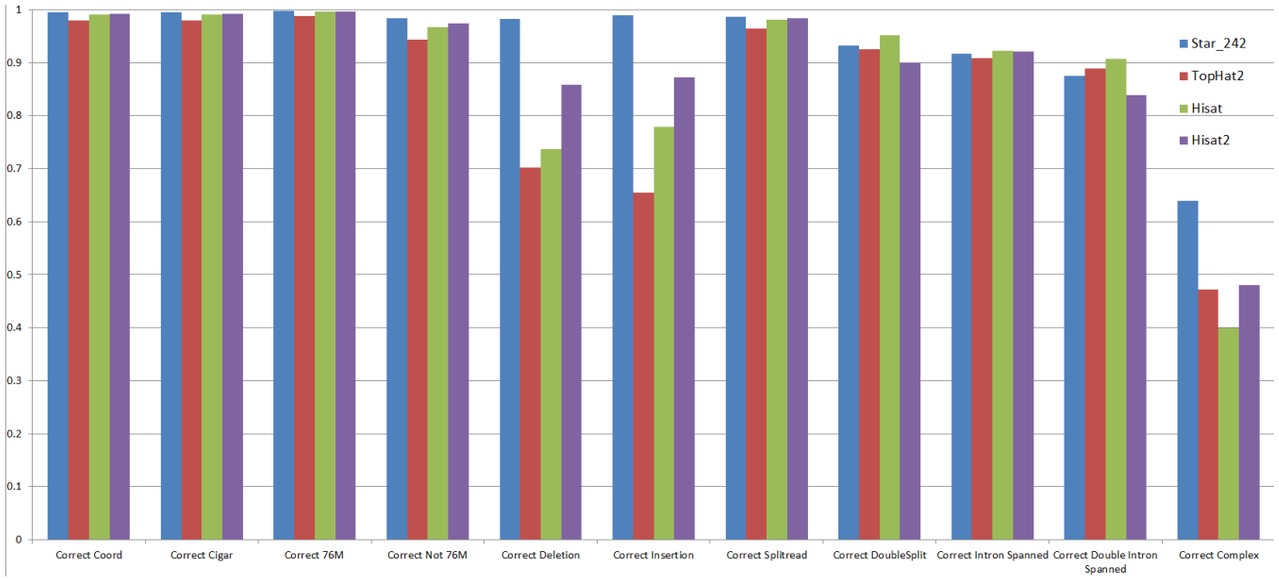
- Allele-specific expression
  [Castel et al., 2015; van de Geijn et al., 2015]

- Multi-tissue eQTL
  - Metasoft [Han & Eskin, 2012]

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

Benchmarking

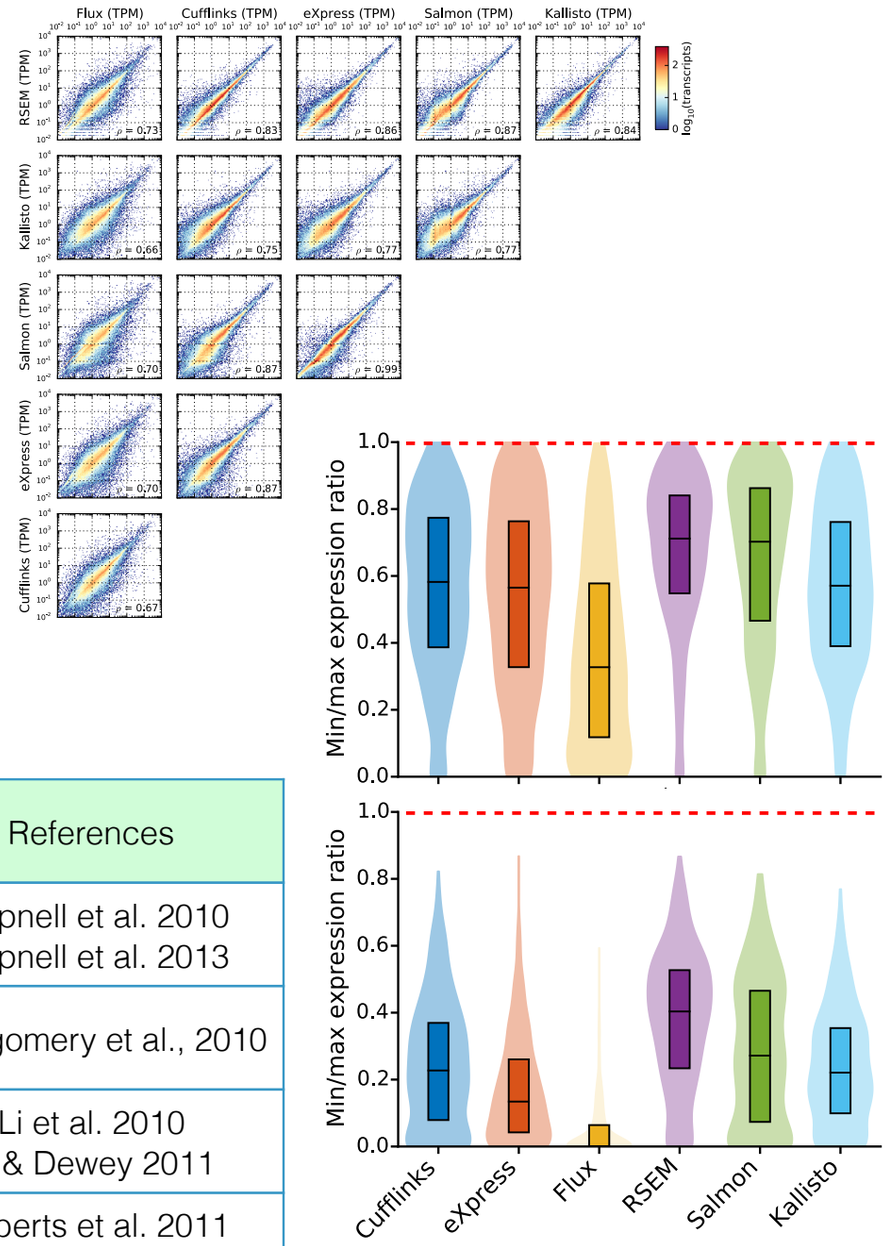
Spliced transcript alignment



Tim Sullivan, Broad Institute

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

Transcript isoform expression estimation



	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

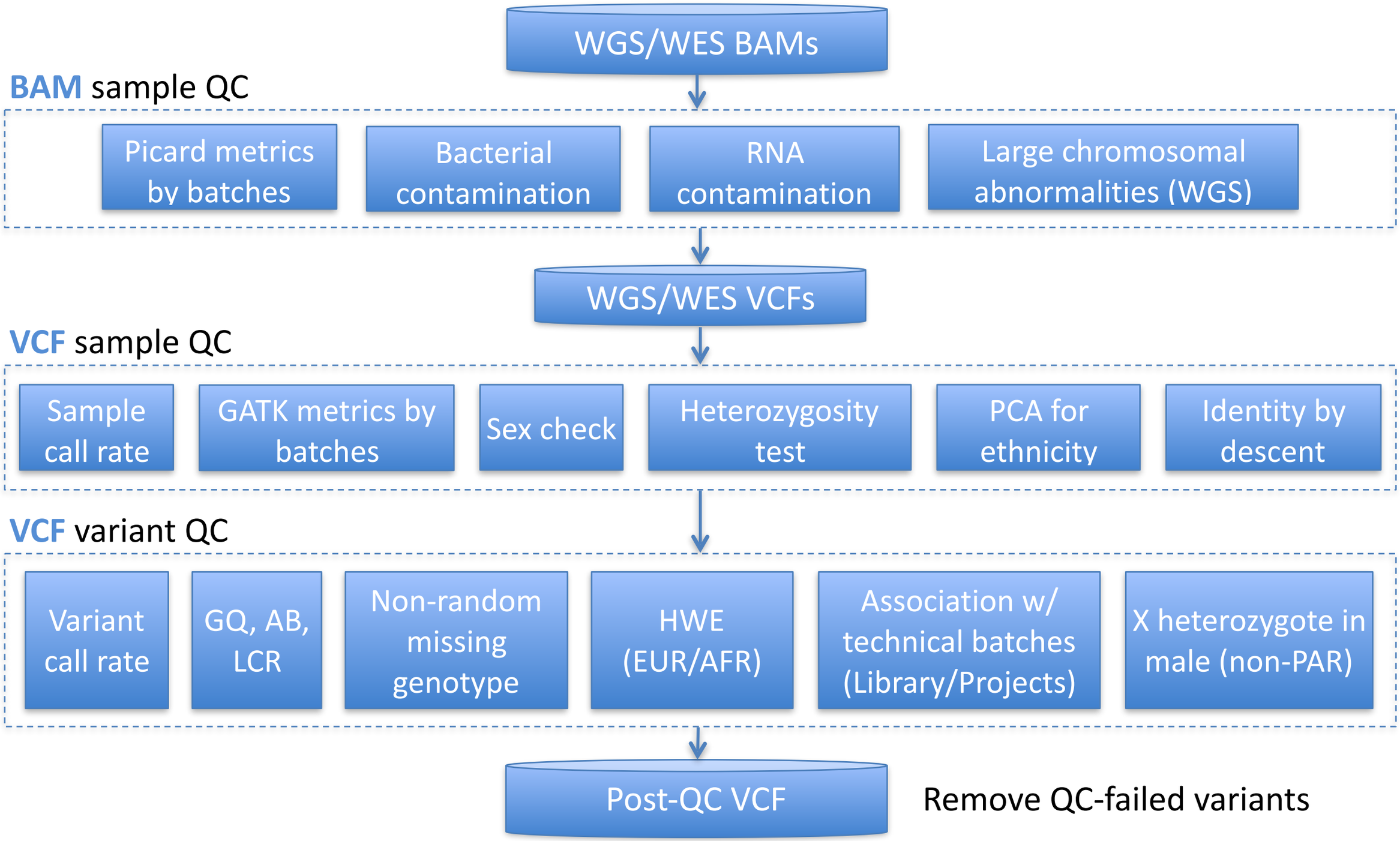
v6p release

- **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

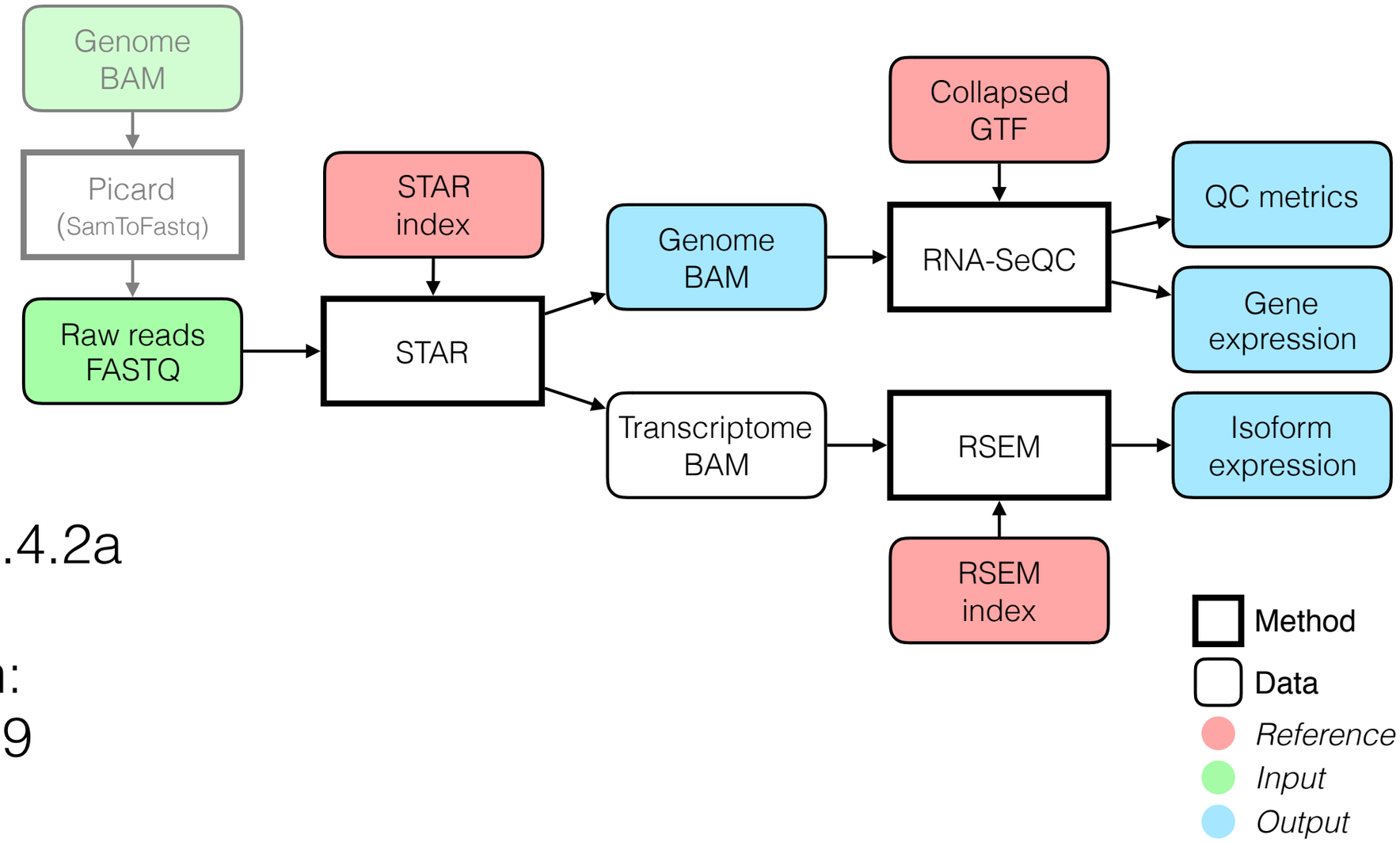
	WES	WGS
# donors	603	652
# donors (post sample QC for eQTL analysis)	603	635
Sequencing coverage	100x	30x
Alignment	BWA	BWA-MEM
Joint variant calling	HaplotypeCaller v3.4 (GATK)	HaplotypeCaller v3.4 (GATK)
Variant QC	-	GATK, Hail, Custom code
Functional and LoF annotations	Ensembl's Variant Effect Predictor + 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

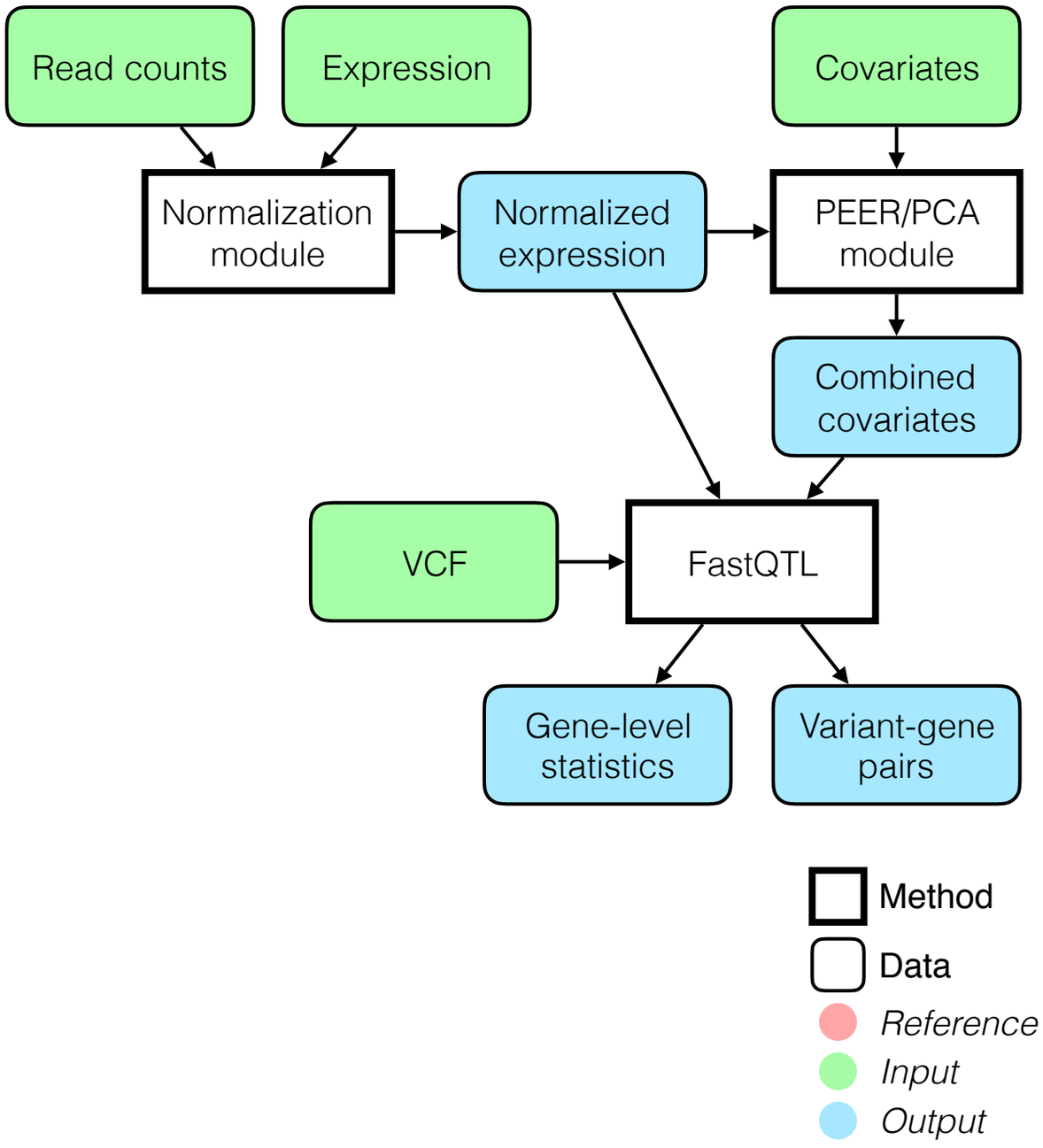


- Aligner: STAR v2.4.2a
- Gene expression:
  RNA-SeQC v1.1.9
- Transcript expression:
  RSEM v1.2.22
- QC metrics:
  RNA-SeQC v1.1.9

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

eQTL discovery

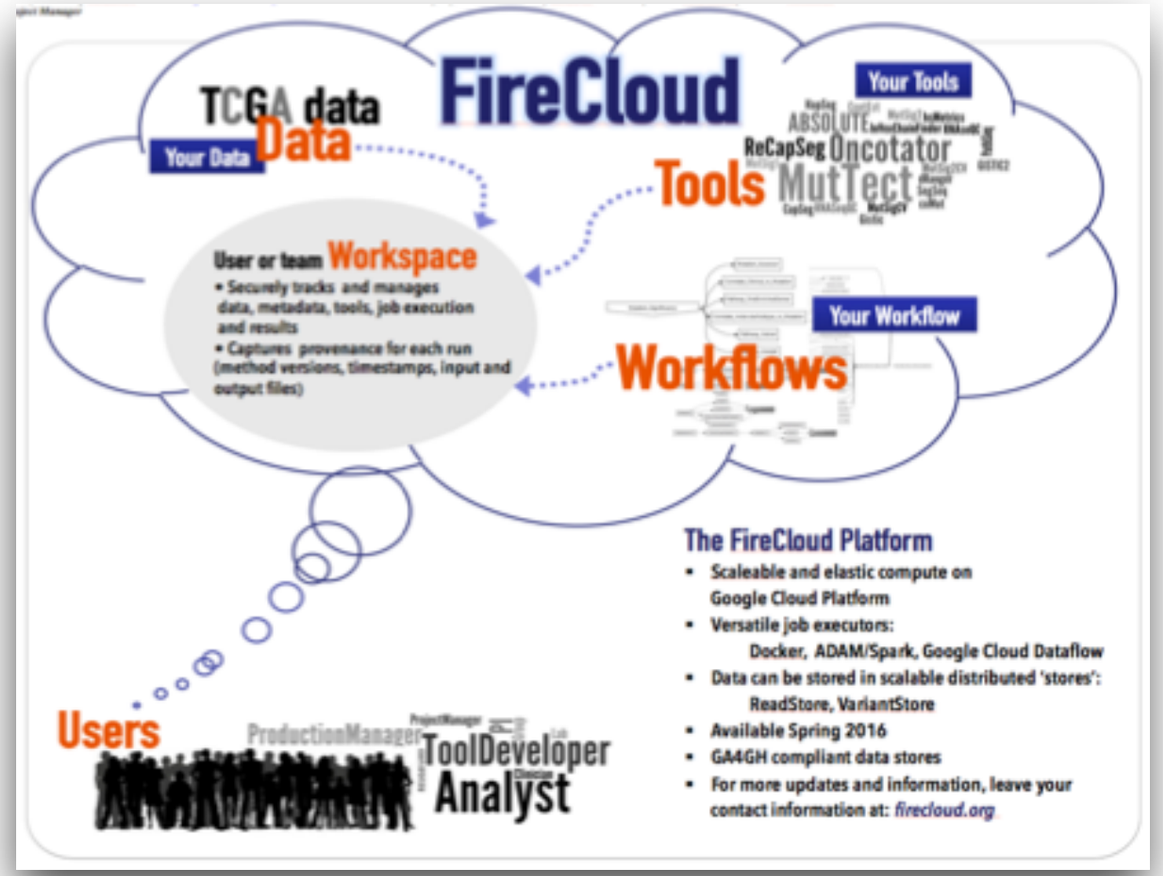
- QTL mapper: FastQTL
- Covariate correction:
  - PEER factors
  - Explicit covariates:
    Genotype PCs, gender
- *cis* window: ±1Mb
- MAF ≥ 0.01 and ≥10 samples containing minor allele

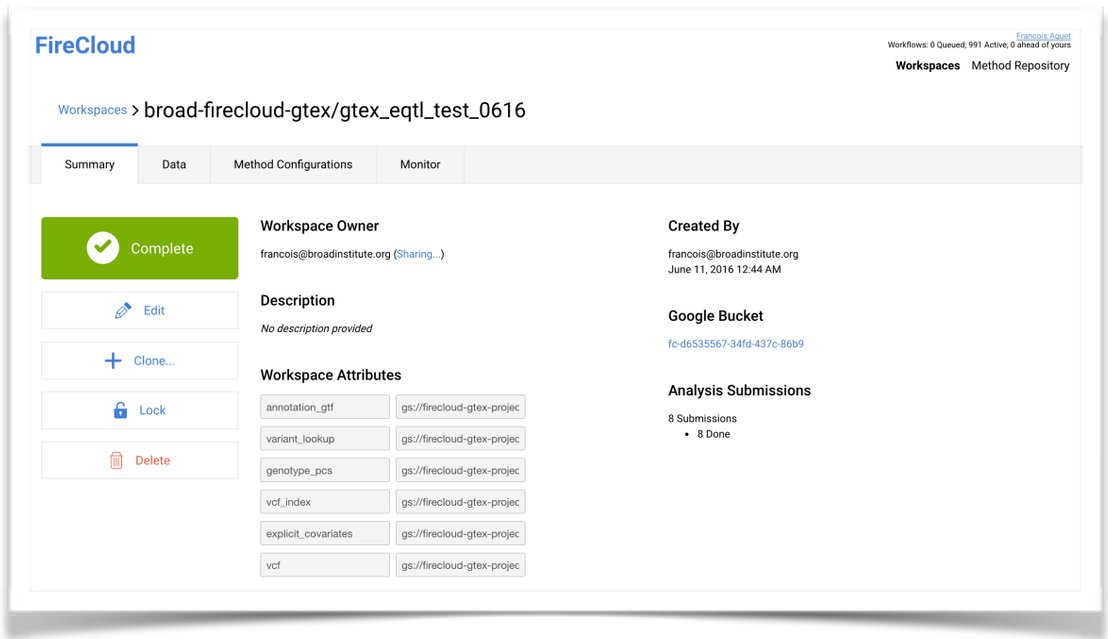


FastQTL: Ongen et al., *Bioinformatics*, 2016
PEER: Stegle et al., *PLoS Comp Bio*, 2010

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.





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

BROAD INSTITUTE

LDACC

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

E. Gelfand, C. Trowbridge

D. MacArthur, M. Kellis, J. Hirschhorn

Genomics Platform

GTEx Portal

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

S. Meier, M. Noble

GTEx

The GTEx Project Consortium

Benchmarking Subgroup

eQTL Subgroup

Transcriptome Subgroup

Gender Subgroup

NIH

The Common Fund

Donors and their families