

Genotype-Tissue Expression (GTEx)

2012 BRN Symposium

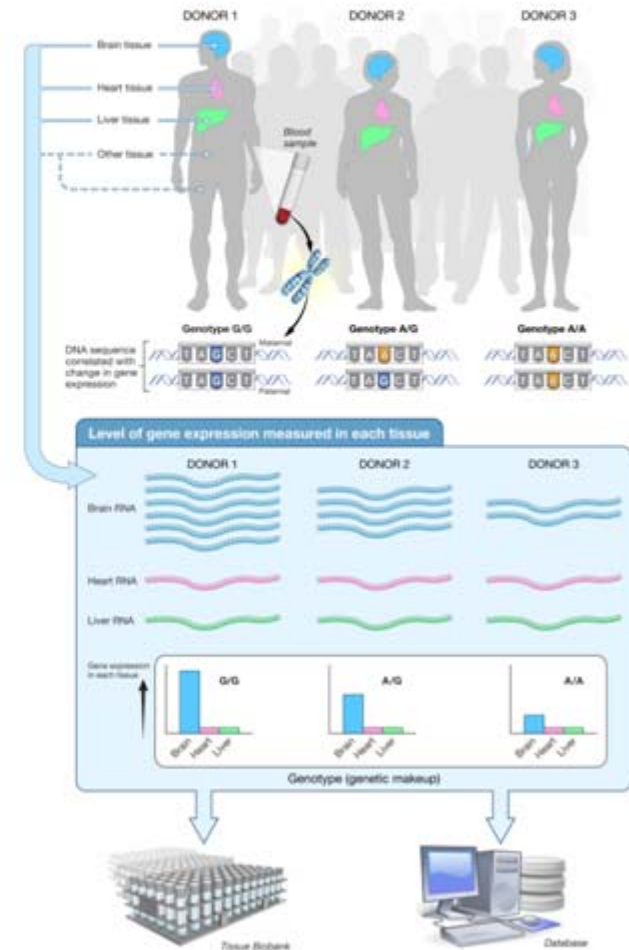
February 23, 2012

NHGRI: Jeff Struewing, Simona Volpi and Cathy Ng

NIMH: Susan Koester and Roger Little

GTE_x – Goal

- Establish a resource database and tissue bank in which to study the relationship between genetic variation and gene expression in reference/non-diseased human tissues



GTE_x – Overview

- NIH Common Fund Project
 - “NIH Roadmap”
- June 2008 Workshop
- NHGRI/NIMH/NCI & other ICs developed proposal
- Early 2009 – approved by NIH Director as 2-year Pilot
 - IC Director concerns about feasibility of enrolling post-mortem donors soon after death and obtaining high quality nucleic acids

Genotype-Tissue Expression (GTEx)

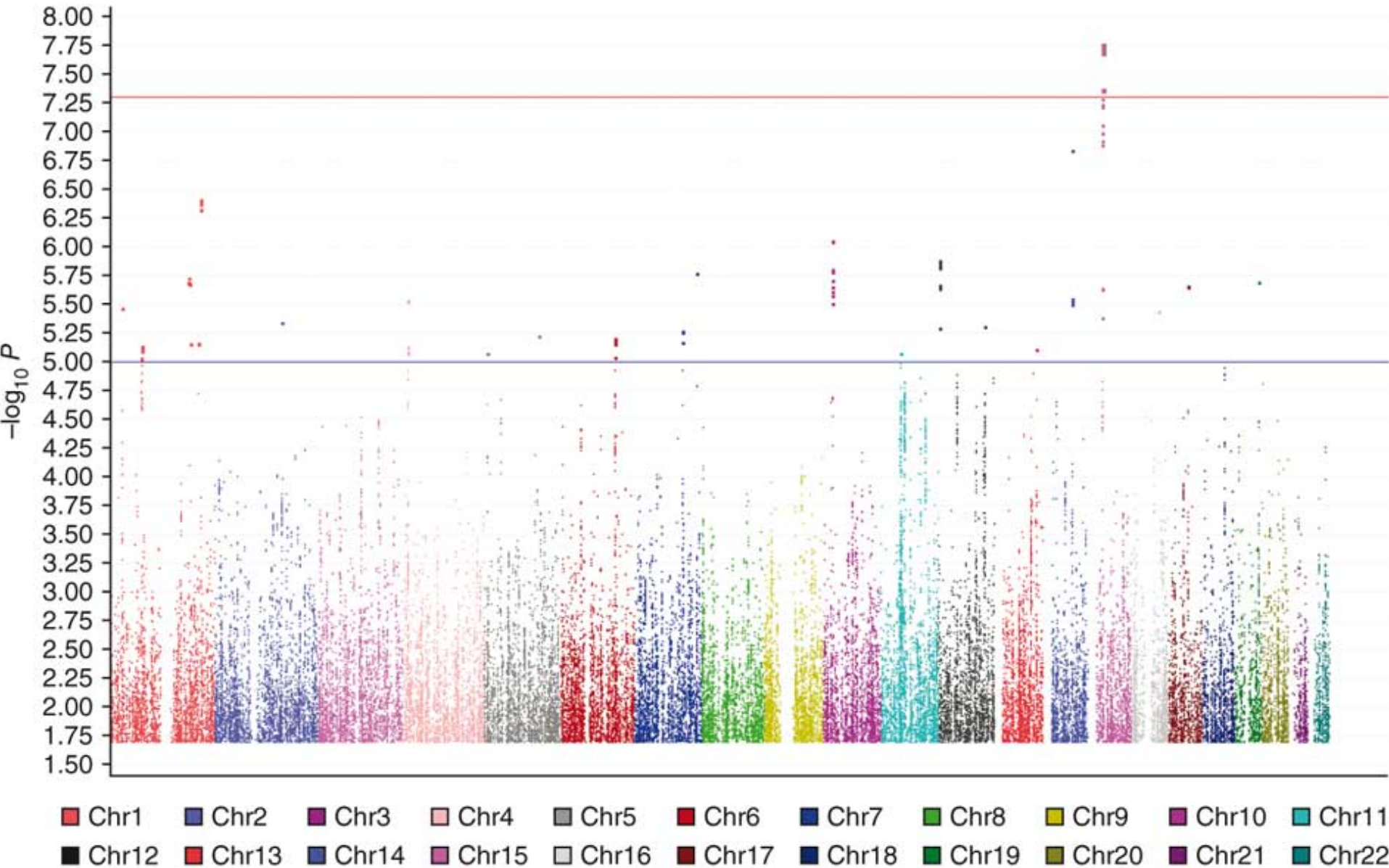
- NIH Common Fund (commonfund.nih.gov/gtex)
- Co-chairs – Eric Green, NHGRI & Tom Insel, NIMH
- Lead ICs (Program Officers)
 - ▣ NHGRI (Jeff Struewing, Simona Volpi, Gary Temple)
 - ▣ NIMH (Su Koester, Roger Little)
 - ▣ NCI (Carolyn Compton, Jim Vaught)
 - ▣ NHLBI (Weiniu Gan)
- Active participation from many other ICs

Agenda



- Scientific background
- Overview of GTE_x
- Implementation of GTE_x
- Where we are

Manhattan Plot



Nature Genetics 42 , 897–901 (2010) A genome-wide association study identifies a susceptibility locus for refractive errors and myopia at 15q14

Challenges in Functional Analysis of a GWAS Finding

- Most associations do not involve changes in the coding region of protein-coding genes
- Multiple 'candidate' genes may lie within the associated interval
- No annotated genes may be near the associated region

LETTERS

Genetic variants regulating *ORMDL3* expression contribute to the risk of childhood asthma

Miriam F. Moffatt^{1*}, Michael Kabesch^{2*}, Liming Liang^{3*}, Anna L. Dixon⁴, David Strachan⁵, Simon Heath⁶, Martin Depner², Andrea von Berg⁷, Albrecht Bufe⁸, Ernst Rietschel⁹, Andrea Heinzmann¹⁰, Burkard Simma¹¹, Thomas Frischer¹², Saffron A. G. Willis-Owen¹, Kenny C. C. Wong¹, Thomas Illig¹³, Christian Vogelberg¹⁴, Stephan K. Weiland¹⁵, Erika von Mutius², Gonçalo R. Abecasis³, Martin Farrall⁴, Ivo G. Gut⁶, G. Mark Lathrop⁶ & William O. C. Cookson¹

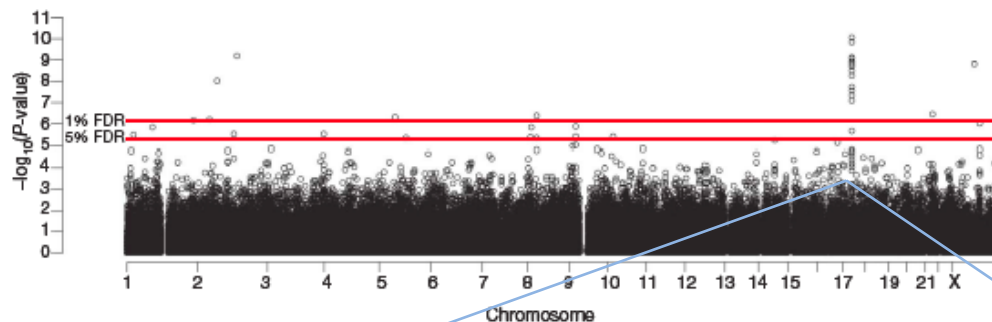
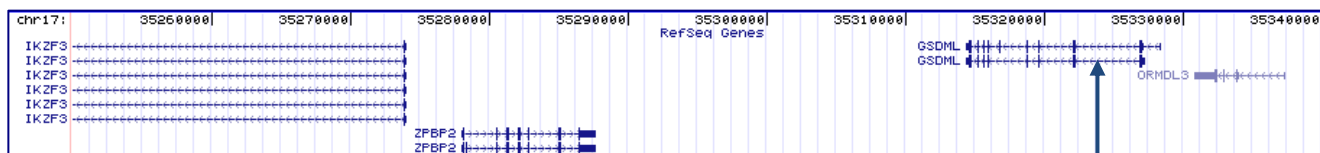
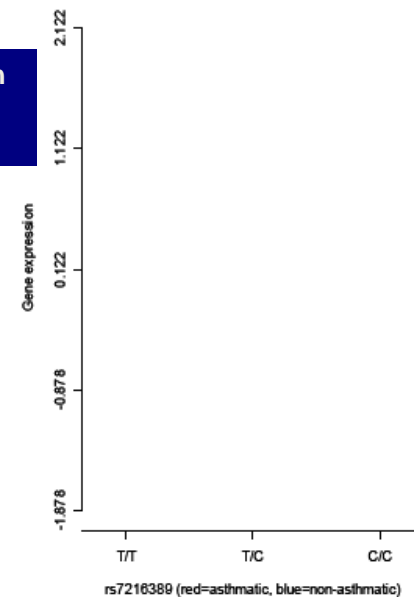


Figure 2 | Genome-wide association of 317,447 SNPs and asthma in 994 asthmatic children and 1,243 non-asthmatic children. Position in the



rs7216389

ORMDL3 expression by genotype



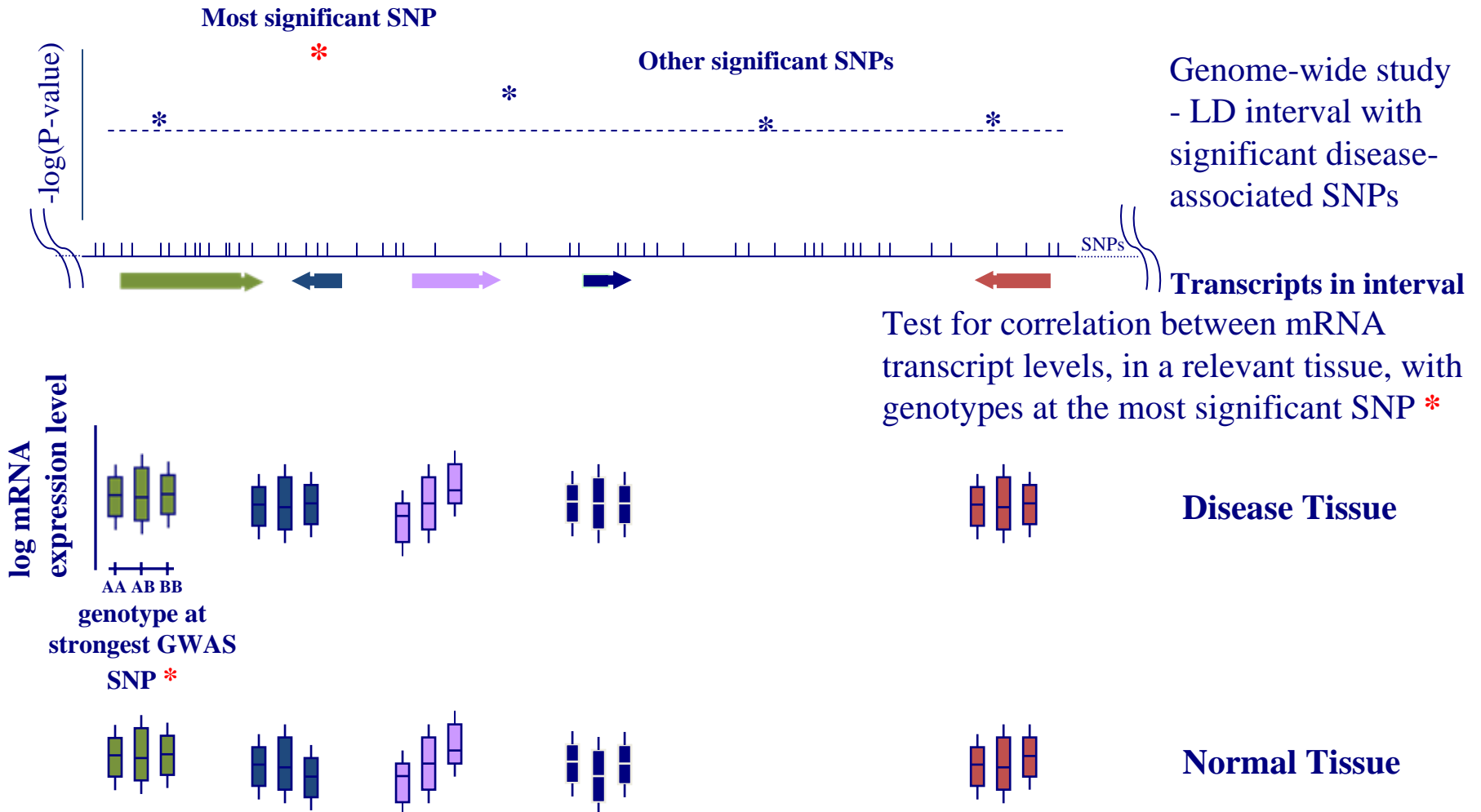
rs7216389 (red=asthmatic, blue=non-asthmatic)

lymphoblastoid
mRNA expression

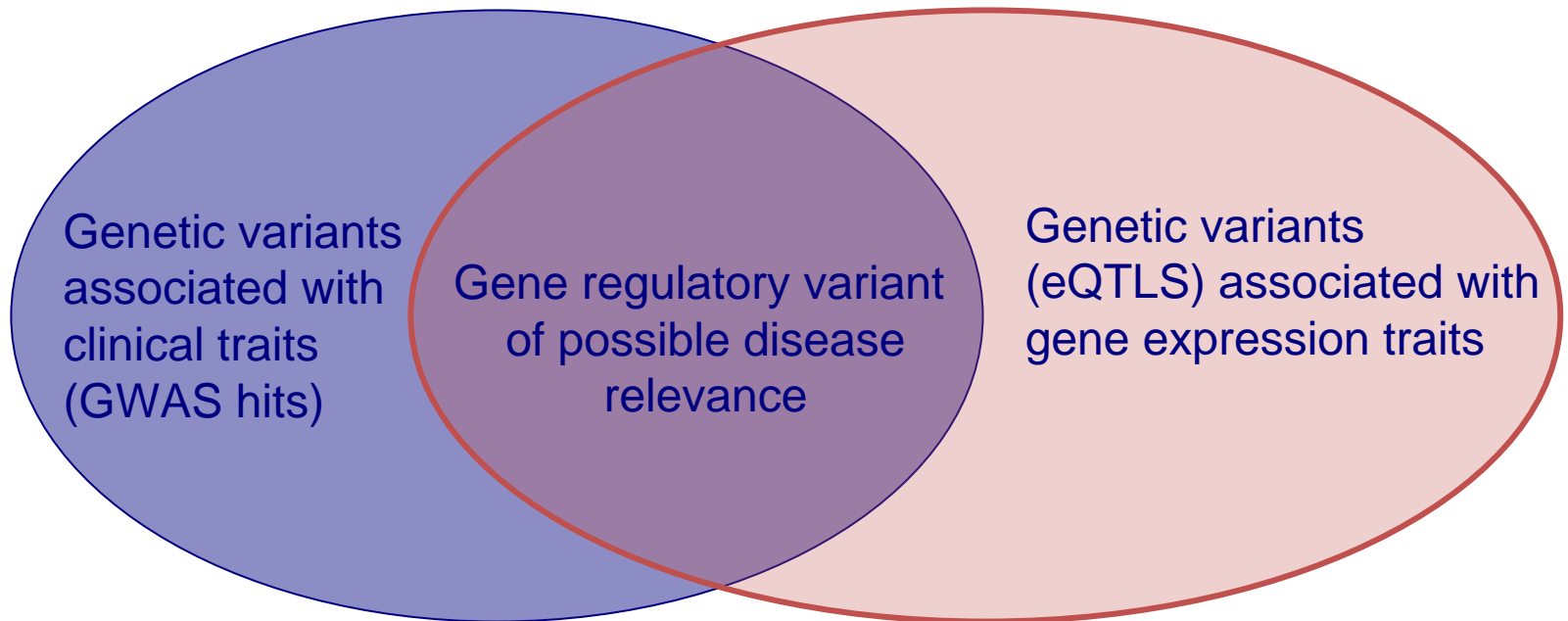
Expression Quantitative Trait Locus (eQTL)

- Gene expression (transcript levels) can be treated as heritable quantitative traits.
- Genomic sites associated with such expression traits are called expression quantitative trait loci or **eQTLs**.
- An **eSNP** is a specific DNA sequence variant that is associated with a change in the level of expression of a given gene.
- **eQTLs** can influence expression traits in *cis* or in *trans*.

Following up GWAS finding with expression analysis



eQTLs and Human Disease Genetics



- Provides way to infer causal mechanisms of GWAS loci.
- Natural genetic variation as perturbations of gene regulation.
- Beyond GWAS

Agenda

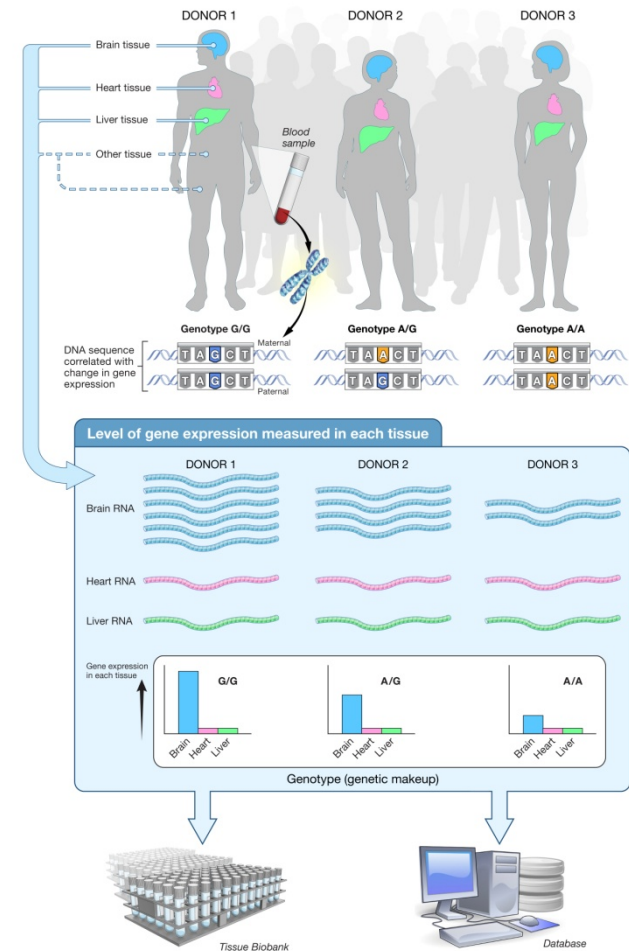


- Scientific background
- **Overview of GTEx**
- Implementation of GTEx
- Where we are

GTEx – Overview (I)

- Goal – Establish a resource database and tissue bank in which to study the relationship between genetic variation and gene expression in reference/non-diseased human tissues

□ <http://commonfund.nih.gov/GTEx>



GTEx – Overview (II)

- Common Fund 2-year feasibility pilot project
- Specific aims:
 - Enroll 160 post-mortem donors, unselected for disease status
 - Obtain high-quality RNA from 30+ tissues per donor
 - Parallel series of surgery donors (4-5 tissues)
 - Measure gene expression and high-density genotypes – identify *cis* eQTLs

GTEx – Pilot Milestones



- Enrollment of 10 donors/month
- RNA samples with $RIN > 6$ for at least 70% of 12 tissues
- Identify *cis*-eQTLs for 4-10% of transcripts
- Scale up: additional 750 donors

Autopsy Donors

GTE_x

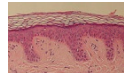
Surgery Donors

Goal n=160

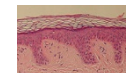
Goal n=100

From each donor

From each donor



SNP Chip & lymphoblastoid cell line



Goal – 30+ tissues
(Total n ~ 5,000 tissues)

Goal – 5 tissues
(Total n ~ 500 tissues)

NIH-supported brain bank

For each tissue

fibroblast cell line

For each tissue

Fix in PAXGene Tissue
(0.2 – 0.5 gram aliquots)

Fix in PAXGene Tissue
(0.2 – 0.5 gram aliquots)

BSS ↑



CBR

Storage

One aliquot

One aliquot fixed for histopathologic review

One aliquot

Storage

LDACC ↓

Homogenize, extract RNA, measure RNA quality

High-quality RNA (Goal > 70%)

RNA-Seq to 20-30 million reads + expression array an overlapping subset
76 bp paired-end Illumina HiSeq2000

BSS – Biospecimen Source Site
CBR – Comprehensive Biospecimen Resource
LDACC – Laboratory, Data Analysis, and Coordinating Center
v1/20/12

GTEx – Target Tissues

Paxgene Preserved Tissues

- ❑ Adipose
- ❑ Adrenal Gland
- ❑ Artery – Aorta
- ❑ Artery – Coronary
- ❑ Artery – Tibial
- ❑ Bladder
- ❑ Breast - Mammary Tissue
- ❑ Cervix – Ecto & Endocervix
- ❑ Colon
- ❑ Esophagus – Mucosa & Muscularis
- ❑ Fallopian Tube
- ❑ Heart
- ❑ Kidney – Cortex & Medulla
- ❑ Liver
- ❑ Lung
- ❑ Muscle – Skeletal
- ❑ Nerve – Tibial
- ❑ Ovary
- ❑ Pancreas

- ❑ Pituitary
- ❑ Prostate
- ❑ Skin
- ❑ Spleen
- ❑ Stomach
- ❑ Testis
- ❑ Thyroid
- ❑ Uterus
- ❑ Vagina
- ❑ Brain – Cerebellum
- ❑ Brain – Cortex

- Fibroblast Cell Line
- Blood
- Lymphoblastoid Cell Line

Fresh Frozen Brains

- ❑ Brain – Cerebellum
- ❑ Brain – Cortex
- ❑ Hippocampus
- ❑ Substantia nigra
- ❑ Anterior cingulate cortex (BA24)
- ❑ Amygdala
- ❑ Caudate (basal ganglia)
- ❑ Nucleus accumbens (basal ganglia)
- ❑ Putamen (basal ganglia)
- ❑ Hypothalamus
- ❑ Spinal cord (cervical c-1)

Agenda



- Scientific background
- Overview of GTE_x
- **Implementation of GTE_x**
- Where we are

Biospecimen Source Sites

- National Disease Research Interchange
 - Organ Procurement Organizations
 - LifeNet Health
 - Gift of Life Donor Program
 - Surgery & autopsy
 - Drexel University College of Medicine
 - Albert Einstein College of Medicine
 - Virginia Commonwealth University
- Roswell Park Cancer Institute
 - Organ Procurement Organization
 - Upstate New York Transplant Service
- ELSI study
 - Virginia Commonwealth University

caHUB The Cancer Human Biobank

Van Andel Research Institute
Comprehensive Biospecimen Resource (CBR)
Biorepository Operations & Pathology Review



SAIC-Frederick
Comprehensive Data Resource (CDR)
caHUB Data Coordinating Center

NCI/SAIC-Frederick sub-contractors



Laboratory, Data Analysis & Coordinating Center



- Project coordination
- Nucleic acid extractions
- Genotyping
- Gene expression (array & RNA-Seq)
- Statistical analysis

GTEx Database

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

- eQTL display
- Controlled access to individual-level data




Statistical Methods Development (R01s)

- Cox/Nicolae (Univ Chicago)
- Dermitzakis/Guigo/Koller/McCarthy (Univ Geneva)
- Liu (Harvard)
- Pritchard (Univ Chicago)
- Rusyn/Nobel/Wright (UNC-Chapel Hill)

Brain Bank

- Receive whole brains
- Dissect 11 regions for expression analysis



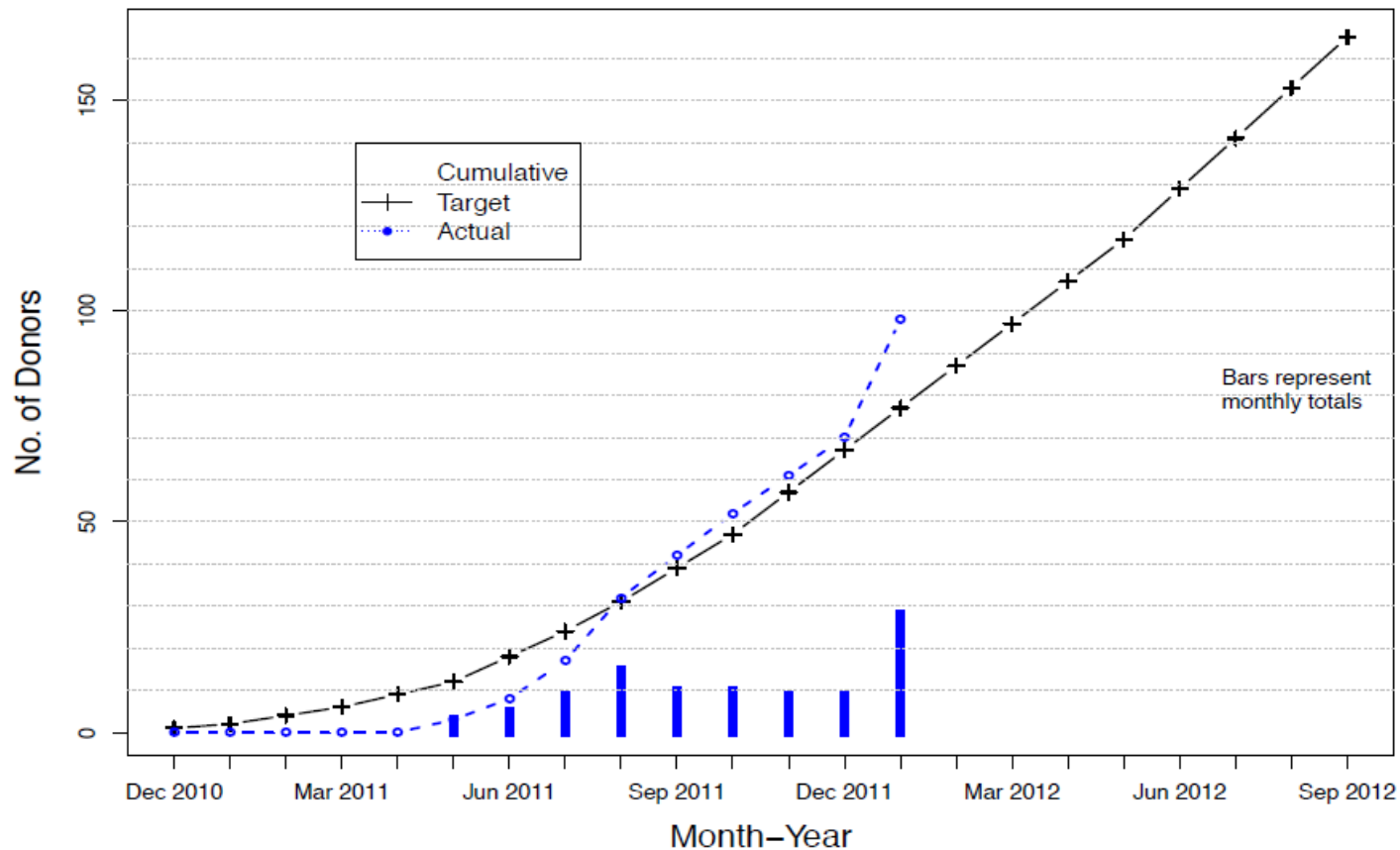
Agenda



- Scientific background
- Overview of GTEx
- Implementation of GTEx
- **Where we are**

GTEx – Enrollment

- May 2011 – February 4, 2012: 100 donors total
- Overall average of ~11 donors/month



GTE_x – RNA quality

- ❑ **RNA samples with RNA Integrity Number (RIN) ≥ 6 for at least 70% of 12 tissues**
 - ❑ Based on analysis for the first 57 donors (average of 26 tissues collected):
 - ❑ 62% (909/1457) samples have RIN>6
 - ❑ Over 70% of 12 tissues have RIN>6
 - ❑ Biospecimens collected as of January 31, 2012:
 - ❑ 2300 Paxgene Tissue Aliquots
 - ❑ 440 Frozen Brain Tissue Samples (Brain Bank)

GTE_x – eQTLs

- ❑ **Identify *cis*-eQTLs for 4-10% of transcripts**

- ❑ Samples size thus far is too small to calculate *cis*-eQTLs. However, genotyping and expression profiling are underway. The expected analysis by March 2012 includes:

- ❑ Genotyping (5M Illumina): 48 donors
 - ❑ Expression Arrays (Affymetrix): 480 samples
 - ❑ RNA-Seq (Illumina HiSeq2000): 290 tissues

Acknowledgments

<u>National Disease Research Interchange (NDRI)</u>	<u>SAIC-Frederick</u>
John Lonsdale, Jeff Thomas and Team	Greg Korzeniewski and Team
<u>Roswell Park Cancer Institute</u>	<u>Van Andel Research Institute</u>
Barbara Foster and Team	Scott Jewel and Team
<u>Broad Institute</u>	<u>Univ Miami Brain Bank</u>
Kristin Ardlie, Wendy Winckler and Team	Deborah Mash and Team
<u>Virginia Commonwealth University</u>	
Laura Siminoff and Team	