National Human Genome Research Institute

The Genotype-Tissue Expression (GTEx) Project

Cathy Ng[1], Gary Temple[1], Simona Volpi[1], A. Roger Little[2], Susan Koester[2], Carolyn Compton[3] and Jeffery P. Struewing[1]

[1]Division of Extramural Research, NHGRI, [2]NIMH, [3]Office of Biorespositories and Biospecimen Research, NCI

http://nihroadmap.nih.gov/GTEx/

Understanding the role of variation in the human genome is crucial to elucidating genetic contributions to human health and disease. Despite the results of genome-wide association studies (GWAS) documenting strong statistical associations between genetic variation and human traits, the functional role for most of these variants is largely unexplained. Nearly 90% of these GWAS-implicated sites lie outside of protein-coding sequences, suggesting that these variants might regulate gene expression.

The Genotype-Tissue Expression (GTEx) project was launched in 2010 as a two-year $25 million pilot sponsored by the NIH Common Fund with the goal of assessing the feasibility of collecting high-quality RNA from multiple tissues from healthy donors. The pilot project will collect and analyze RNA levels in 50+ human tissues from 160 low post-mortem interval donors and 4-6 tissues from 160 surgical controls that have been characterized for germline genetic variation through dense genotyping. The resulting data will be made available to the scientific community through an NCBI-supported GTEx database. Ultimately, GTEx will promote studies on the relationship between human genetic variation and gene expression.

By treating RNA expression levels as quantitative traits, expression quantitative trait loci (eQTLs) will be identified as sites containing genetic variation that correlate with changes in RNA expression. Such eQTLs have been associated with 4%-12% of expressed human genes, and with common complex human diseases, including obesity, atherosclerosis, type 2 diabetes, Crohn's disease, and asthma. Additionally, few studies have examined the tissue specificity of eQTLs. The GTEx project will thus serve as a resource database and tissue bank for many future studies, especially for understanding the functional basis of inherited susceptibility to disease.

Program Initiatives

Initiative 1: Donor Recruitment & Tissue Collection

Establish 3-4 Biospecimen Source Sites (BSS) to provide human tissues for biomedical research. The pilot will also address the ethical, legal and social implications raised in GTEx by conducting qualitative interviews with donors and relatives at one BSS to assess the consent process and impact of participation.

Initiative 2: Laboratory, Data Analysis and Coordinating Center (LDACC)

The LDACC coordinates the overall GTEx project and will:

- Track and receive samples and de-identified patient data
- Perform nucleic acid extraction and purification, genotyping, and RNA quantification by microarray and next-generation sequencing
- Perform statistical analyses to identify eQTLs
- Manage and integrate all phenotypic and molecular data into GTEx database at NCBI
- Organize major GTEx activities, including working group meetings

Initiative 3: Database and Genome Browsing Integration

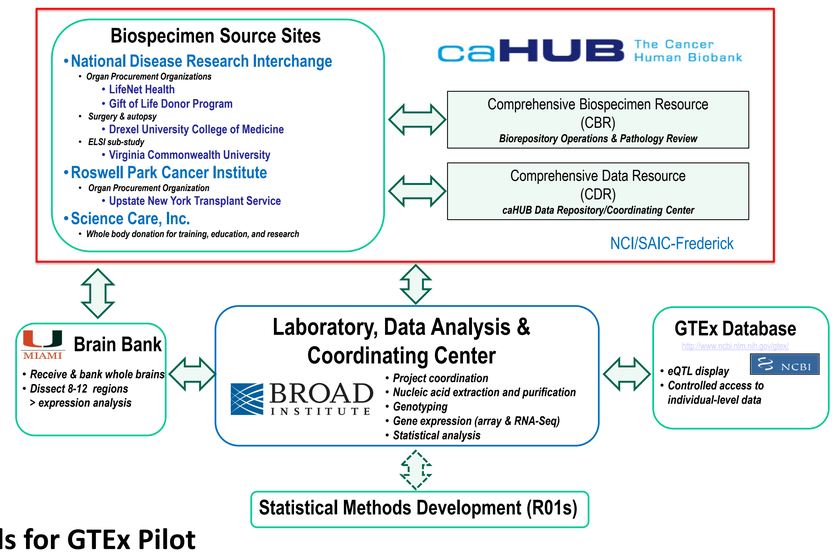
Establish a working database with the following functions:

- Receive, harmonize, and distribute raw genotype and tissue-specific gene expression data
- Provide a user-friendly interface to display pre-computed eQTL correlation statistics
- Integrate and display eQTL data in two genome browsers outside of NIH

Initiative 4: Development of eQTL Data Analysis Methods

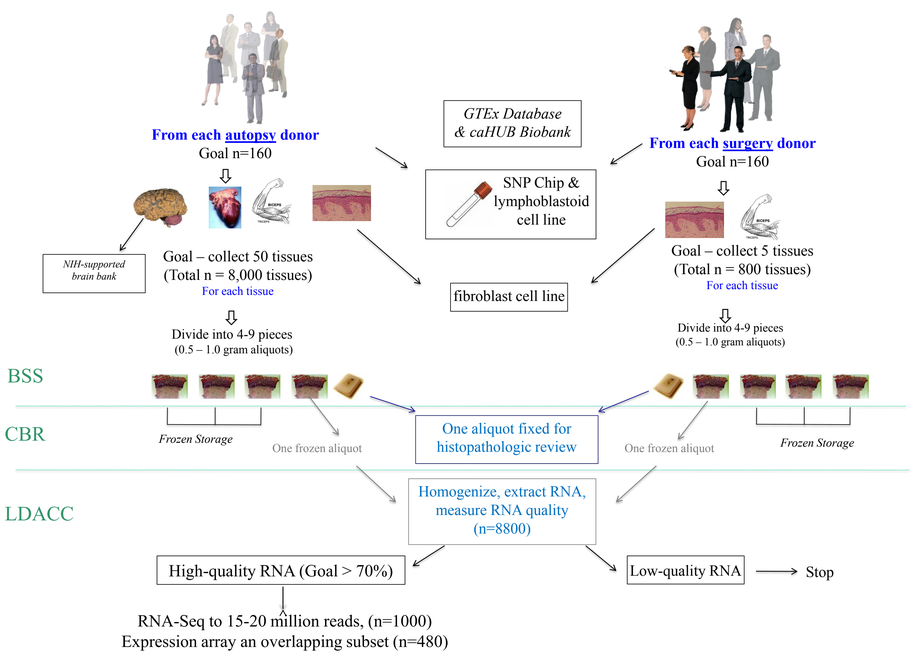
Establish methods for computational analysis of multiple tissues/donors to efficiently identify *cis*- and *trans*-eQTLs as well as the optimal number of donors and samples to collect in a scale-up project.

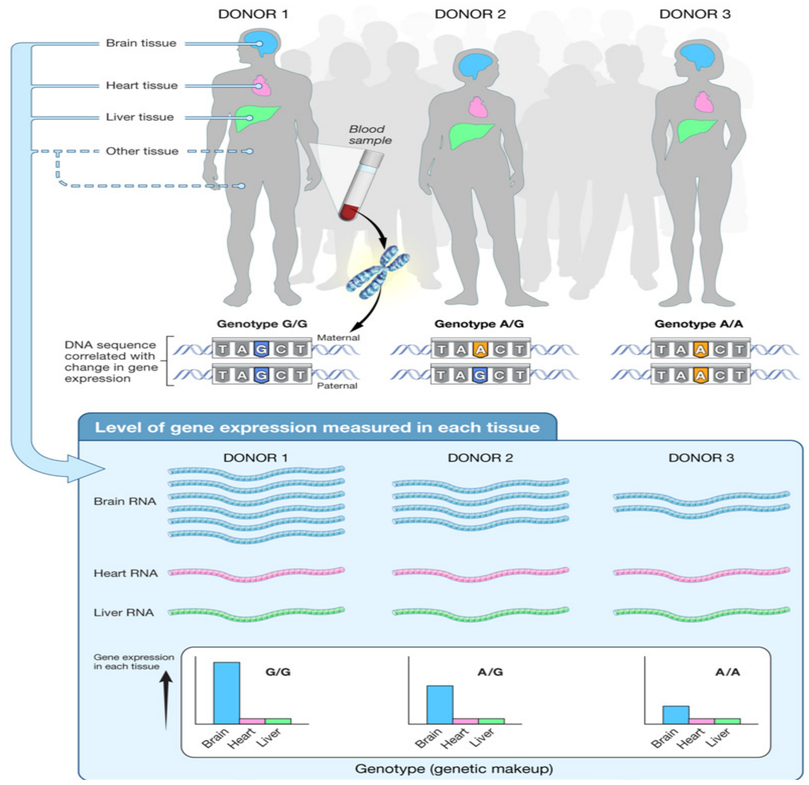
Overview of Infrastructure

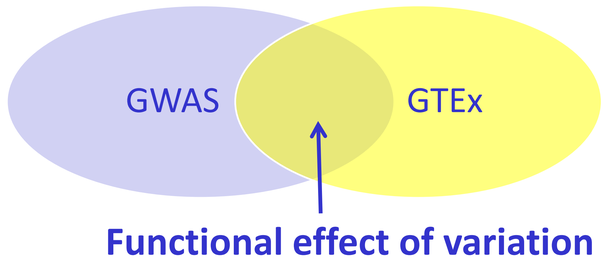


Goals for GTEx Pilot

- Obtain a total of 8,000 post-mortem samples (50 tissues in each one of 160 post-mortem donors), and 800 surgical samples (5 tissues in each one of 160 surgical donors)
- Collect high-quality RNA from at least 70% of at least 12 tissues (RIN ≥ 6)
- Enroll at least 10 post-mortem donors each month at pilot's conclusion
- Identify eQTLs in at least three tissues for at least 4% of expressed transcripts







Recent and Upcoming Progress

GTEx is holding its second, bi-annual working group meeting with principal investigators and project officers on June 2011 in the Washington D.C. area. Trial recruitments and efforts to establish data pipelines are underway.

An external scientific panel has been established to provide input to the GTEx project and assess criteria to scale up GTEx beyond the pilot project. Members include:

- **Eric Schadt, Ph.D**. (chair), chief scientific officer, Pacific Biosciences, Menlo Park, Calif.
- **Kevin K. Brown, M.D**., professor and vice chairman, Dept. of Medicine, Nat'l Jewish Health, Denver
- **Vivian G. Cheung, M.D**., professor, pediatrics and genetics, University of Pennsylvania School of Medicine and Howard Hughes Medical Institute, Philadelphia
- **Ross Hardison, Ph.D**., professor, biochemistry and molecular biology, Pennsylvania State University, University Park
- **Allan Jones, Ph.D**., chief executive officer, Allen Institute for Brain Science, Seattle
- **Rebecca Pentz, Ph.D**., professor of research ethics, dept. of hematology-oncology, Emory University School of Medicine, Atlanta
- **David L. Rimm, M.D**., Ph.D., professor, pathology, Yale University School of Medicine, New Haven, Conn.

Benefits of GTEx

In summary, GTEx will provide the research community with the following:

- A public resource of comprehensive RNA expression levels in multiple tissues from donors who are also densely genotyped
- Widely available data for the scientific community to evaluate whether a GWAS disease-associated variant is correlated with expression in a relevant tissue
- Powerful approach to identifying control sequences and networks, even without GWAS results
- Tissue archive for future studies and enable downstream studies of regulatory non-coding RNAs, chromatin modifications, proteomics, etc.