Pharmacogenomics Research Network RNA-seq Project:

Transcriptome Analysis of Pharmacogenes in Human Tissues

Courtney E. French[1], Aparna Chhibber[2], Eric R. Gamazon[3], Sook Wah Yee[2], Xiang Qin[4], Elizabeth Theusch[5], Amy Webb[6], Scott T. Weiss[7,8], Marisa W. Medina[5], Erin G. Schuetz[9], Alfred L. George, Jr.[10], Ronald M. Krauss[5], Christine Q. Simmons[10], Steven E. Scherer[4], Nancy J. Cox[3], Kathleen M. Giacomini[2], and Steven E. Brenner[1*]

[1]University of California, Berkeley, CA. [2]University of California, San Francisco, CA. [3]University of Chicago, Chicago, IL. [4]Baylor College of Medicine, Houston, TX. [5]Children's Hospital Oakland Research Institue, Oakland, CA. [6]Ohio State University, Columbus, Ohio. [7]Brigham and Women's Hospital, Boston, MA. [8]Harvard Medical School, Boston, MA. [9]St. Jude Children's Research Hospital, Memphis, TN. [10]Vanderbilt University, Nashville, TN. *brenner@compbio.berkeley.edu

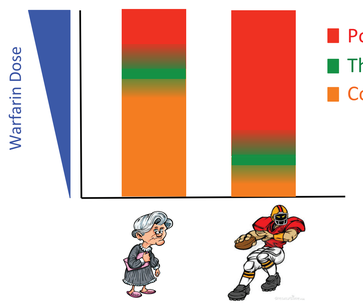
Download PDF

What is pharmacogenomics?

People vary in their response to therapeutic drugs

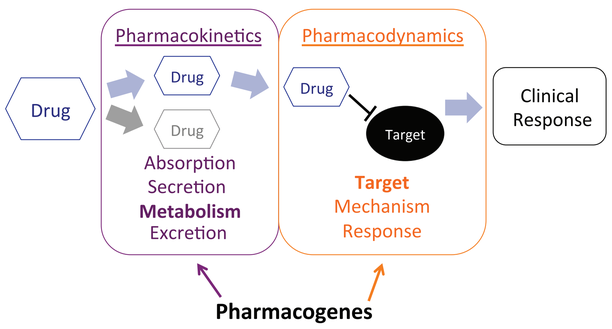
For example, **warfarin**:

- Anti-coagulant prescribed to individuals at risk for blood clots
- Narrow range for therapeutic dose
- Wide range of inter-individual variability in response



Warfarin response is known to be affected by variation in CYP2C9 and VKORC1, as well as age and weight.

These differential drug responses are due to both **human genetic variation** and **environmental factors**. Pharmacogenomics aims to determine the genetic basis of these variable responses in order to improve the efficiency of medications and prevent adverse effects. Of particular interest are **pharmacogenes** - those genes involved in drug response either by effecting the drug's activity (pharmacokinetics) or by being the target or downstream effector of the drug (pharmacodynamics).



Natural variation in pharmacogenes can affect an individual's response to a particular drug at a specific dose.

Aims

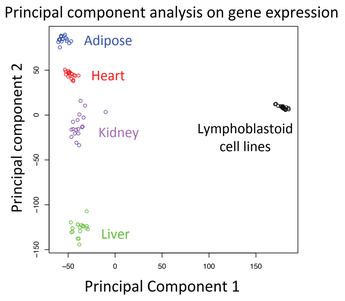
- Determine the expression and alternative splicing profiles of pharmacogenes in human tissues of pharmacological interest
- Investigate the variability of expression and splicing between individuals
- Create a database of this information for use by the greater pharmacogenomics community

RNA-seq of multiple human tissues

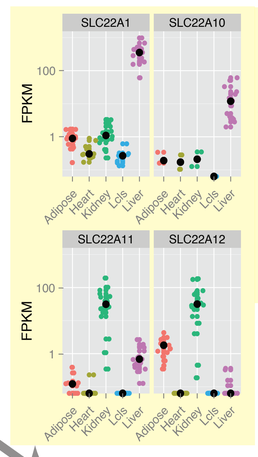
19 Liver samples | 18 Kidney samples | 25 Heart samples

25 Adipose samples | 45 Lymphoblastoid cell lines

To control for variabilty in sequencing depth between the tissues, all samples were randomly subsampled down to 20 million reads for all gene expression and splicing analysis. For most analyses, including this PCA, only 18 samples were used per tissue to control for the different numbers of samples sequenced for each.

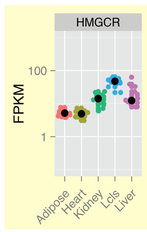


High variability in pharmacogene expression between individuals and tissues



SLC family of transporters
SLC transporters play an important role in drug absorption into cells. They can be highly tissue-specific and, for example, expressed very highly only in liver or only in kidney.

Even within a tissue, expression may vary almost 1000-fold between individuals, as for *SLC22A12*.

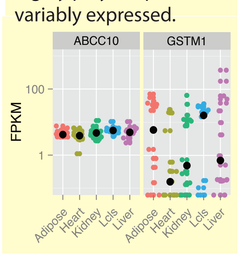


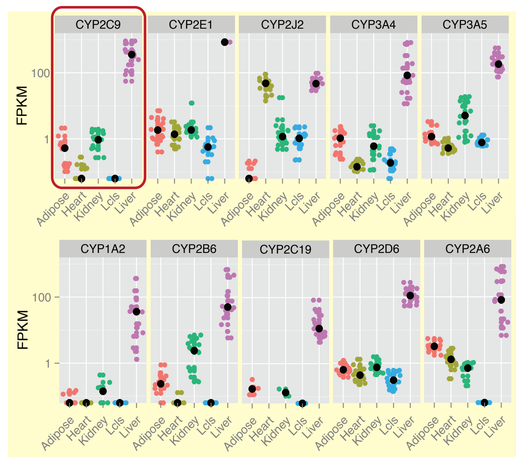
HMGCR is a target of statin cholesterol drugs that is relatively tightly regulated transcriptionally within a tissue, but is variable in alternative splicing between individuals.

- Adipose samples
- Heart samples
- Kidney samples
- Lymphoblastoid cell lines
- Liver samples
- Median tissue expression

FPKM: Gene expression level (fragments per kilobase per million reads)

ABCC10 encodes a transporter and is an examples of tight regulation across tissues and individuals. In contrast, ***GSTM1*** encodes a drug metabolizer that is highly polymorphic and variably expressed.

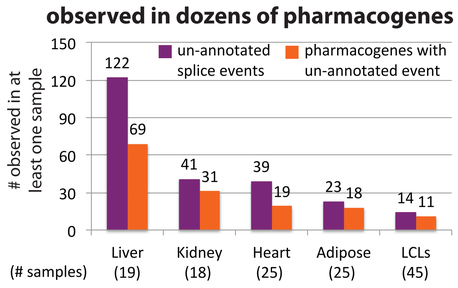




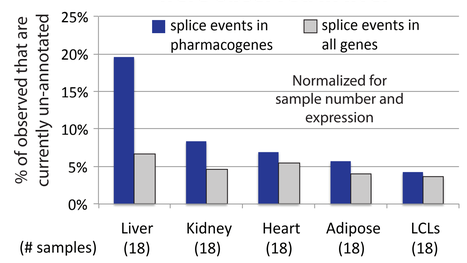
Cytochrome P450 genes
Many CYP genes metabolize drugs; they are highly expressed in liver and their expression may vary over 100-fold between individuals, potentially leading to differential drug responses. For example, *CYP2C9* metabolizes warfarin and variation in its activity is known to affect warfarin dose responses.

Hundreds of un-annotated and tissue-specific splice events were discovered in pharmacogenes

Evidence of novel alternative splicing was observed in dozens of pharmacogenes

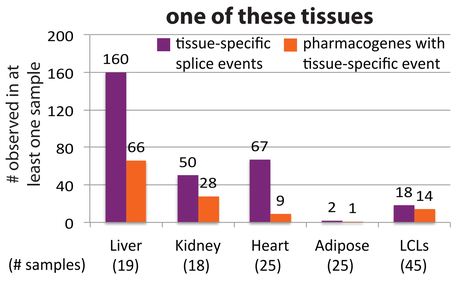


Most un-annotated events were observed in liver

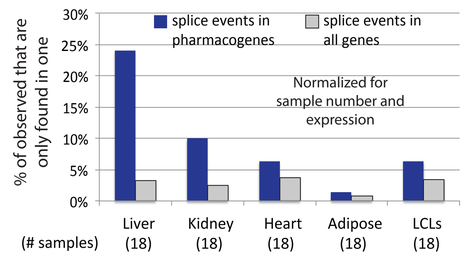


Observed splice events must have a read coverage of ≥ 5 reads/100bp. The annotations used were GENCODE v12 and Ensembl.

Many pharmacogenes have alternative splice events that were observed in only one of these tissues

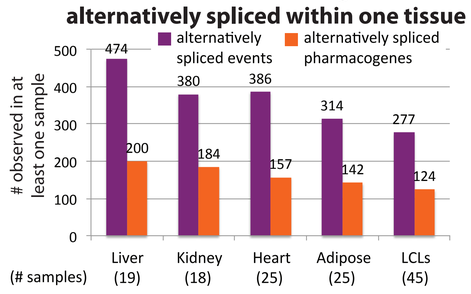


Liver had the most 'tissue-specific' events and adipose had the least

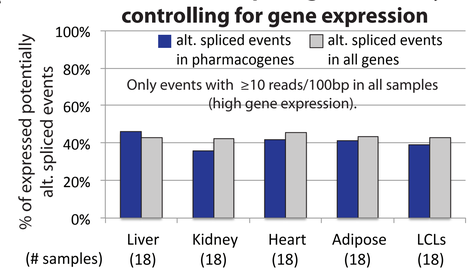


Observed splice events must have a read coverage of ≥ 5 reads/100bp. Tissue-specific splice events must have a read coverage of 0 in all samples of the other tissues.

Hundreds of pharmacogenes are alternatively spliced within one tissue

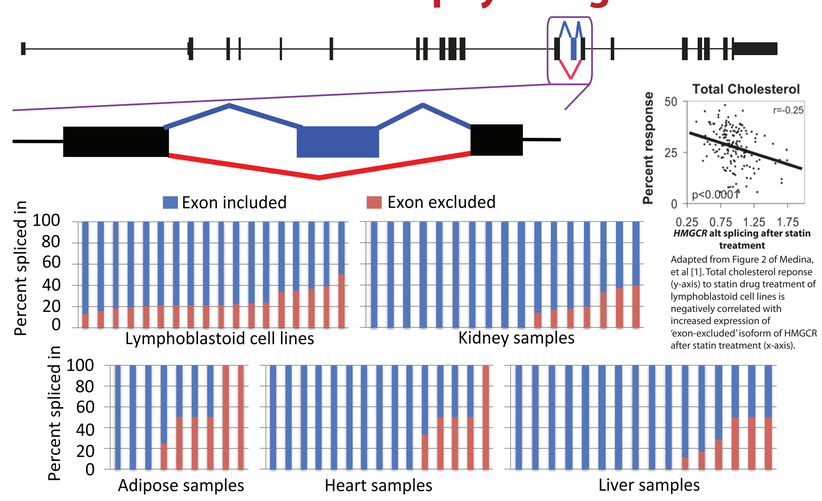


Little difference between tissues in amount of alternative splicing after strictly controlling for gene expression



Alternatively spliced events produce at least two splice forms with PSI ≥ 5 and read coverage ≥ 1 reads/100bp.

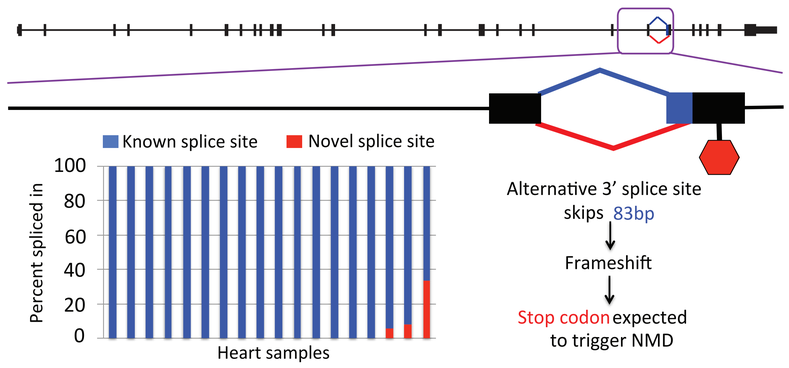
Alternative splicing in HMGCR correlated with drug response in LCLs is variable between individuals in physiological tissues



Adapted from Figure 2 of Medina, et al [1]. Total cholesterol reponse (y-axis) to statin drug treatment of lymphoblastoid cell lines is negatively correlated with increased expression of 'exon-excluded' isoform of HMGCR after statin treatment (x-axis).

Exclusion of exon 13 in the statin-binding domain of HMGCR has been correlated with reduced response to statin treatment for lowering cholesterol in lymphoblastoid cell lines [1]. We find that this splicing event is variable between individuals in all tissues tested, providing support that it is important in differential responses to statin cholesterol drugs. Samples without sufficiant read coverage to calculate percent spliced values are not shown.

Novel 3' splice site in SCN5A introduces stop codon in heart samples



SCN5A encodes a sodium channel important in maintaining normal heart rhythm and implicated in congenital, potentially fatal, long-QT syndrome (LQTS) [2]. Novel alternative splicing of exon 23 was observed in three heart samples and creates a premature termination codon predicted to trigger nonsense-mediated mRNA decay and thus affect the amount of protein produced.

Summary

- Pharmacogenes are expressed and spliced differentially between the different tissues
- Pharmacogenes can also be 1000-fold variable in expression between individuals
- Dozens of currently un-annotated splicing events in pharmacogenes were discovered
- Expression and splicing data for all genes and samples will be made available in a user-friendly database: Pharmaco-SeqDB

REFERENCES:

1. Medina MW, Gao F, Ruan W, Rotter JI, Krauss RM. Alternative splicing of 3-hydroxy-3-methylglutaryl coenzyme A reductase is associated with plasma low-density lipoprotein cholesterol response to simvastatin. *Circulation.* 2008. 118:355-62
2. Murphy LL, Moon-Grady AJ, Cuneo BF, Wakai RT, Yu S, Kunic JD, Benson DW, George AL Jr. Developmentally regulated SCN5A splice variant potentiates dysfunction of a novel mutation associated with severe fetal arrhythmia. *Heart Rhythm.* 2012. 9:590-7
3. Trapnell C, Pachter L, Salzberg SL. Tophat: discovering splice junctions with RNA-seq. *Bioinformatics.* 2009. 25:1105-11 (**Tophat for read alignment, http://ccb.jhu.edu/software/tophat/**)
4. Trapnell, et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology.* 2010. 28:511-5 (**Cufflinks for gene expression analysis, http://cufflinks.cbcb.umd.edu/**)
5. Brooks AN, et al. Conservation of an RNA regulatory map between Drosphila and mammals. *Genome Research.* 2011. 21:193-202 (**JuncBASE for alternative splicing analysis, http://compbio.berkeley.edu/proj/juncbase/Home.html**)

ACKNOWLEDGEMENTS
This study is supported by the NIH Pharmacogenomics Research Network (PGRN) - RNA-Seq Network-wide Project. We are grateful to the PGRN groups (http://pgrn.org/), especially PMT for the kidney samples, PAAR4Kids for the liver samples, PAT for the heart samples, PARC for the adipose and LCL samples, and BCM-HGSC for the RNA-seq.