Integrating Gene Expression Analysis into Genome-Wide Association Studies

Combining multiple forms of data in a single analysis enables more informative characterization of complex traits.

INTRODUCTION

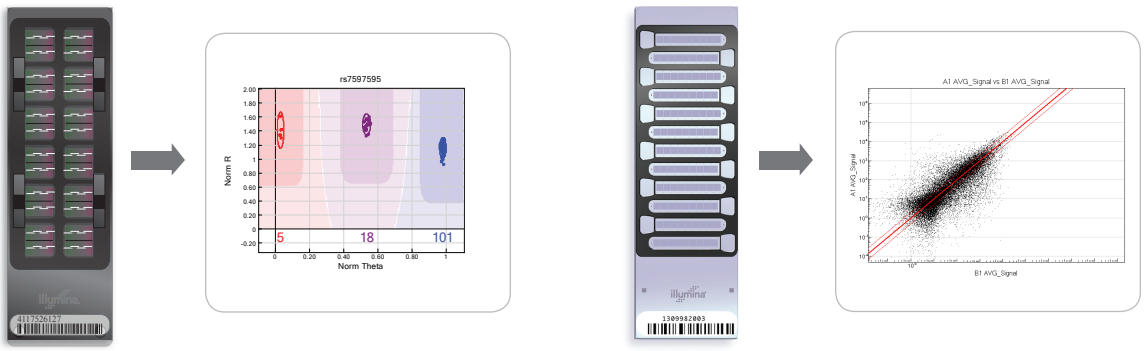
The recent availability of high-throughput, whole-genome genotyping arrays such as Illumina's Infinium® HD DNA Analysis BeadChips has enabled researchers to efficiently screen for associations between variations in the genome and phenotypes of interest. To date, hundreds of genome-wide association studies (GWAS) have identified quantitative trait loci (QTL) underlying many common complex diseases, demonstrating the utility of this approach for dissecting the genetic basis of polygenic traits[1–3].

Although GWAS can effectively map loci contributing to phenotypes of interest, they offer limited insight as to the causative genetic variation or the mechanism by which it confers its effect. As a result, the research community is recognizing the utility of integrating multiple forms of data into a single analysis. For example, the wealth of information that whole-genome expression profiling is capable of uncovering has been well-documented[4], but the potential value that genome-wide expression screens may add to clinical trait-based GWAS is beginning to be realized and is cause for growing excitement in the genetics field[5–8]. In part for this reason, Illumina has developed the HumanHT-12 Gene Expression BeadChip, a 12-sample array that presents a low cost solution for enhancing GWAS with whole-genome expression analysis. This document highlights some of the potential benefits of incorporating whole-genome gene expression data into a clinical trait-based GWAS.

EQTL ANALYSIS: GENE EXPRESSION AS A QUANTITATIVE TRAIT

In a traditional GWAS, the trait being investigated is associated with a region in the genome. This is also the case with eQTL (expression QTL) analysis, which treats mRNA abundance as a trait in a GWAS. An eQTL screen identifies loci, or eQTL, that may contribute directly or indirectly to expression levels. Expression QTL can be divided broadly

FIGURE 1: ILLUMINA SOLUTIONS FOR INTEGRATED ANALYSIS

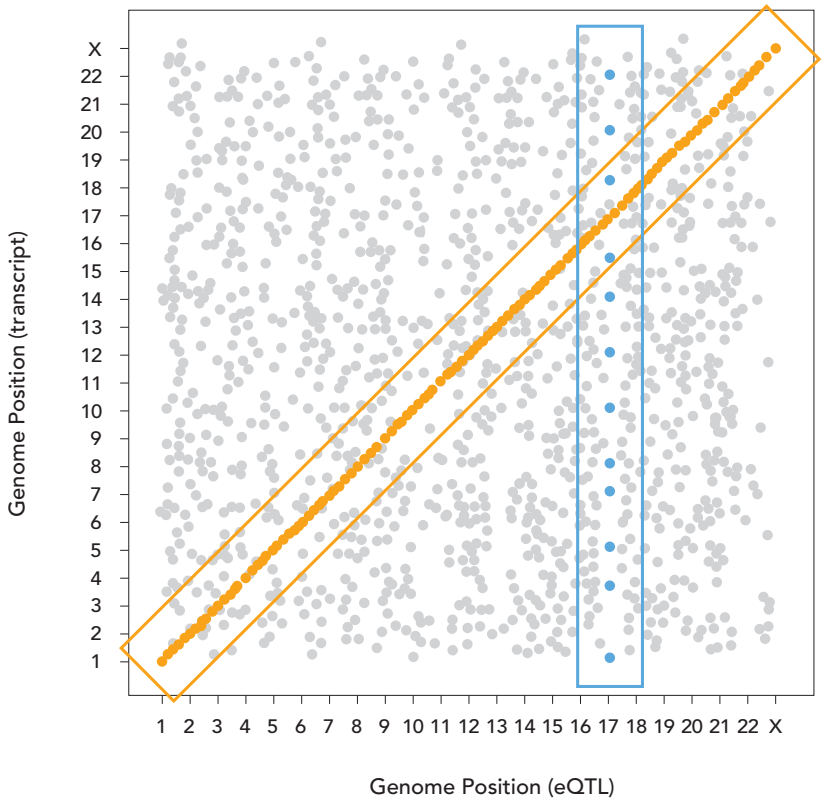


Infinium HD DNA Analysis BeadChip

HumanHT-12 Gene Expression BeadChip

The 12-sample HumanHT-12 Gene Expression BeadChip (right) targets more than 48,000 transcripts in the RefSeq database (Build 36.2, Release 22). This multi-sample whole-genome expression BeadChip matches the throughput of Illumina's Infinium HD DNA Analysis BeadChip (left) product line. Illumina's user-friendly analysis software permits evaluation of both genotype and expression data.

FIGURE 2: EQTL ANALYSIS DATA



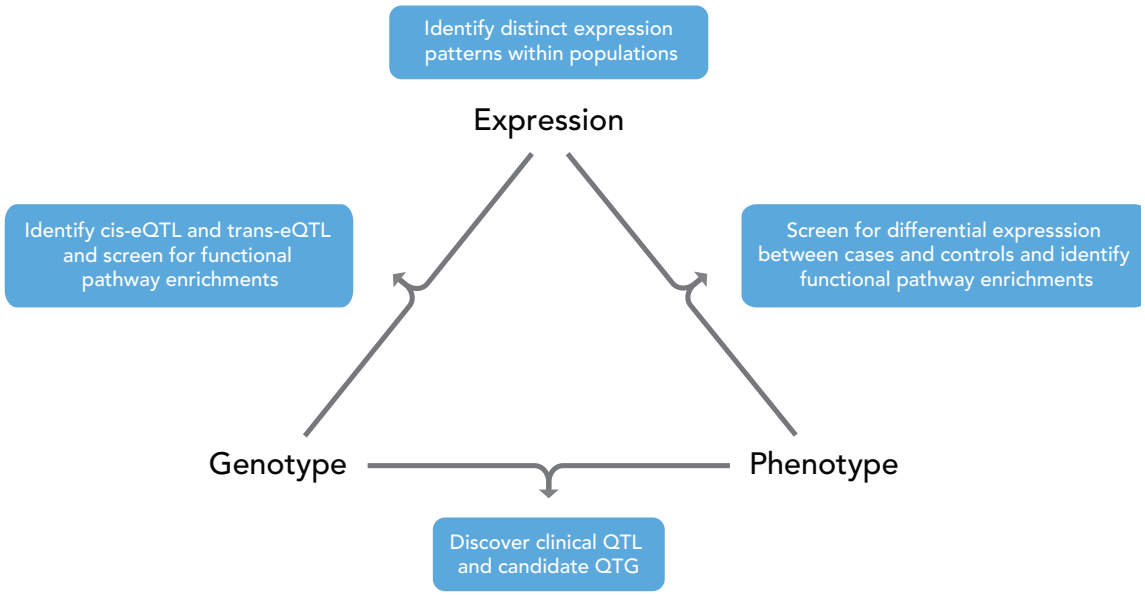
This graph illustrates observations from eQTL analysis that have been reported in the current literature. The diagonal band indicates cis-eQTL. These eQTL are detected when the locus that affects mRNA abundance overlaps the location of the affected gene. The horizontal band represents a trans-band or "eQTL hot spot," which suggests that expression of multiple genes map to the same single-nucleotide polymorphism (SNP).

into two categories based on their position relative to the gene whose expression is being measured. If the eQTL and gene positions overlap, the eQTL is considered to be cis-acting and the gene it overlaps to be cis-regulated (*Figure 2*). If the eQTL and gene location are non-overlapping, the eQTL is considered to be trans-acting. Cis-eQTL are usually believed to result from a variant in a regulatory region of the gene that affects its level of abundance. The causal mechanisms behind trans-eQTL can be considerably more variable. For instance, a trans-eQTL could reflect a variant that affects the abundance or activity of a transcription factor. Alternatively, a trans-eQTL could result from a variation in a component of a signaling cascade that ultimately affects the abundance of the mRNA being measured (*Figure 2*).

CIS-EQTL AND CANDIDATE GENE FILTERING

Although the process of mapping clinical QTL has become very efficient, the transition from QTL to the identification and validation of the quantitative trait genes (QTGs) containing the causative genetic variant has been severely rate-limiting[9,10]. While informatics tools such as SNP and functional pathway databases have utility in filtering QTG candidates, biological evidence supporting these data is often lacking. Adding expression signatures to a GWAS may serve to bridge this gap by enabling the detection of cis-regulated QTG within clinical QTL[6,8,11]. Cis-eQTL suggest a relationship between a local genetic variant and phenotypic variation that is specific to a unique QTG candidate and could be capable of driving the clinical trait difference. If a cis-regulated QTG candidate were supported by informatics-based evidence

FIGURE 3: POTENTIAL OUTCOMES OF COMBINING GENE EXPRESSION AND GENOTYPE DATA



such as known function relative to the clinical trait of interest or the presence of candidate causative SNPs, the data would present a compelling case for prioritizing the QTG candidate for further experimental validation.

DIFFERENTIAL EXPRESSION, TRANS-EQTL, AND FUNCTIONAL PATHWAY ANALYSIS

Adding expression analysis to a GWAS may also be useful in the identification of differential expression patterns across study populations. Screening for functional category enrichment among genes that are differentially regulated between case and control populations has been used to identify pathways that may be involved in conferring traits of interest[6,12]. In some cases, subsets of these genes might be under coordinated regulatory control, such that a single genetic variant is responsible for the differential mRNA abundance detected across them. When integrated with genotype data in the context of a GWAS, such gene sets may map to eQTL trans-bands, or " eQTL hot spots," that overlap a QTL for the clinical trait[6]. In such instances, any genes in the co-regulated set that are not annotated in the enriched functional category may be excellent candidates for downstream validation as novel members of the functional pathway and modulators of the clinical trait.

Comparing gene expression signatures within clinical trait populations may provide value to GWAS as well. For example, although the phenotype exhibited by the study population may appear uniform across affected individuals, its genetic basis may not be. The detection of discernible expression profiles within a sample population may be useful for identifying distinct subphenotype groups. These subgroups could then be analyzed independently, potentially reducing the level of noise and enabling the detection of clinical QTL that might otherwise be missed[6,8].

SUMMARY

Combining expression and genotype data sheds greater light on the biological context of QTL. It enables the generation of better-informed hypotheses and provides additional filtering tools for prioritizing candidate QTGs. The rapidly growing number of examples highlighting the importance of interplay among genotype, DNA methylation status, miRNA, and mRNA abundance in determining the incidence, nature, and severity of clinical phenotypes serves as an indication of the broad range of experiments that future studies will employ. Just as methods used in GWAS can be applied to expression data to generate eQTL results, they can also be applied to data from methylation,

miRNA, and similar experiments to create a more complete characterization of the molecular basis of complex traits. Clearly, an effective tool for characterizing a multifaceted trait is a multifaceted analysis. To that end, Illumina is building a growing portfolio of products that facilitate cost-effective and efficient data integration for the future of genetic discovery.

REFERENCES

(1) Pearson TA, Manolio TA (2008) How to Interpret a Genome-wide Assocation Study. JAMA 299(11):1335–1344.

(2) McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Genet 9:356–369.

(3) http://www.illumina.com/pagesnrn.ilmn?ID=89

(4) Segal E, Friedman N, Kaminski N, Regev A, Koller D (2005) From signatures to models: understanding cancer using microarrays. Nat Genet 37:S38–S45.

(5) Dermitzakis ET (2008) From gene expression to disease risk. Nat Genet 40:492–493.

(6) Drake TA, Schadt EE, Lusis AJ (2006) Integrating genetic and gene expression data: application to cardiovascular and metabolic traits in mice. Mamm Genome 17(6):466–79.

(7) Gilad Y, Rifkin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet 24(8):408–15.

(8) Schadt EE, Monks SA, Friend SH (2003) A new paradigm for drug discovery: integrating clinical, genetic, genomic and molecular phenotype data to identify drug targets. Biochem Soc Trans 31(2):437–43.

(9) DiPetrillo K, Wang X, Stylianou IM, Paigen B. (2005) Trends Genet. Bioinformatics toolbox for narrowing rodent quantitative trait loci. 21(12):683–92.

(10) Korstanje R and Paigen B (2002) From QTL to gene: the harvest begins. Nat Genet 31:235–236.

(11) Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, et al. (2005) Genome-wide associations of gene expression variation in humans. PLoS Genet 1(6):e78.

(12) Kloth JN, Gorter A, Fleuren GJ, Oosting J, Uljee S et al. (2008) Elevated expression of SerpinA1 and SerpinA3 in HLA-positive cervical carcinoma. J Pathol 215(3):222–30.

ADDITIONAL INFORMATION

For more information about Illumina Gene Expression and Infinium HD DNA Analysis BeadChips, please visit www.illumina.com or contact us at the address below.

Illumina, Inc.
Customer Solutions
9885 Towne Centre Drive
San Diego, CA 92121-1975
1.800.809.4566 (toll free)
1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com

FOR RESEARCH USE ONLY

Pub. No. 470-2008-008 Current as of 05 August 2008