Published in final edited form as:

Methods Mol Biol. 2013 ; 1015: 213–224. doi:10.1007/978-1-62703-435-7_14.

SCAN: A Systems Biology Approach to Pharmacogenomic Discovery

Eric R. Gamazon, R. Stephanie Huang, and Nancy J. Cox

Abstract

Genome-wide association (GWA) studies have identified thousands of genetic variants that contribute to disease and pharmacologic traits. More recently, high-throughput sequencing studies promise to provide a more complete catalog of genetic variants with roles in human phenotypic variation. Yet, characterizing the influence of functional variants on genes, RNAs, proteins, and ultimately disease or pharmacologic traits is a critical challenge for a vast majority of the implicated susceptibility loci. Here we describe SCAN, a bioinformatics resource we have developed to elucidate the functional consequences of genetic variants identified by genome-wide scans. In particular, this public resource implements a systems biology approach to pharmacogenomic discovery.

Keywords

eQTLs; Pharmacogenomics; Expression profiling; Transcriptome; SNP function; Genetic variation

1 Introduction

Genome-wide association (GWA) studies have provided molecular medicine and modern biology with increasingly complex datasets— in scope and diversity—and have enabled high-throughput genomic analyses of a broad spectrum of disease traits and pharmacologic outcomes. The massive amounts of data that characterize such studies generate an unprecedented volume of results whose exploration is facilitated, and often only feasible, through the use of versatile databases and computational tools. Indeed, our greater ability to conduct high-throughput surveys of human genetic variation has led to new computational challenges in the analysis and prioritization of findings aimed at the identification and characterization of genetic loci that predict phenotype.

Besides the deluge of data emerging from GWA studies, there is perhaps a greater challenge facing GWA as an approach to complex traits genetics. Although GWA studies have led to some notable successes in the identification of genetic variation influencing complex traits [1], the primary aim of such studies—a comprehensive account of the genetic etiology of disease risk or drug response—continues to challenge our best efforts. It is widely appreciated that GWA findings on their own do not necessarily enhance our ability to discern the underlying biological function of genetic signals. Furthermore, despite the growing list of reproducible GWA studies results, substantial heritability remains to be accounted for [2].

To meet some of these computational and analytic challenges, our group has devoted substantial resources to developing a publicly available genomic database, SCAN (<http://www.scandb.org>) [3], which serves results of transcriptome studies in HapMap lymphoblastoid cell lines (LCLs), and includes information on common copy number variants as well. Our preliminary studies with SCAN have shown that SNPs associated with complex traits are more likely to be expression quantitative trait loci (eQTLs) than minor allele-frequency-matched SNPs drawn from high-density SNP genotyping platforms [4]. This finding has been robustly observed across a wide range of definitions for trait-associated SNPs and eQTLs, and across a broad range of human phenotypes [1, 5]. SCAN provides a framework for utilizing the transcriptome and implements a functional annotation pipeline to expand on genetic association studies.

In this chapter, we describe the feature set of a bioinformatics resource that has proven useful for studies in systems-based pharmacogenomics [6, 7]. We motivate SCAN's particular variant annotation approach to the prioritization of results from the flood of data from GWA studies. Finally, we present various in silico experiments to demonstrate the utility of the particular approach to the integration of genomic datasets that has been implemented in SCAN.

2 Features

2.1 Genetic Variation: SNPs and CNVs

The cell lines from the International HapMap Project [8] have been the most frequently used for in vitro studies of gene expression and pharmacologic phenotypes [6, 9, 10]. The cell lines are Epstein-Barr virus transformed lymphocytes derived from apparently healthy individuals of different ancestry. The HapMap LCLs have been extensively utilized as a model for pharmacogenomic studies, most notably on cellular toxicities of anticancer agents. HapMap samples in 30 CEU (Utah residents with ancestry from Northern and Western Europe) trios and in 30 YRI (Yoruba in Ibadan, Nigeria) trios comprised the initial genotype dataset in SCAN. Studies that have been conducted on these samples have demonstrated that gene expression and drug response phenotypes such as cytotoxicity are heritable and include an appreciable genetic component. Extensive genotypic data are available on these samples from the International HapMap Project (<http://www.hapmap.org>) and, more recently, the 1000 Genomes Project (<http://www.1000Genomes.org>) [11]. The latter initiative promises to enable explorations of the role of rare variants in complex traits genetics by providing a more comprehensive catalog of human genetic variation. SCAN is poised to contribute to the development of methods for testing associations with complex traits, including drug response [12], using data from sequencing studies.

The first release of SCAN utilized the HapMap SNPs as the primary unit of analysis. With the recent release of an extensive catalog of copy number variants (CNVs) assayed in the HapMap cell lines from array-based technologies [13] and from population-based sequencing studies [14], SCAN has been expanded to include information on these genetic variants [15]. In a recent pharmacogenomic study of these CNVs, we have been able to show that, for an array of functionally diverse chemotherapeutics, the top CNV associations

with cytotoxicity are independent of known SNP associations and have thus not been interrogated by previous SNP-based GWA studies of these drugs [16].

2.2 Integration of Functional Annotations of Genetic Variation

Given the proliferation of databases and resources for genomic studies, it was felt that there was a need for consolidating publicly available *functional* annotations for loci emerging from genome-wide studies. We were particularly interested in developing novel tools relevant to large-scale, whole-genome studies aimed at characterizing susceptibility loci for complex traits. Functional classification of variants is one of the primary challenges of such studies for the purposes of prioritization for follow-up studies and of providing mechanistic hypotheses for observed associations between SNPs and disease or drug response. A database to assist in the identification of the functional consequences or potential biological impact of genetic variants from genome-wide scanning was a primary aim in the development of SCAN.

A priori information on how genetic variation may impact biological processes or molecular function may aid in the interpretation of results from GWA studies. Indeed, GWA analyses restricted to variants with increased probabilities of association may be utilized to reduce the number of tests performed and increase the power to detect susceptibility loci. Of course, many genetic variants have no known functional effect on disease or therapeutic outcome in humans. Furthermore, determining the influence of genetic variation on genes, RNAs, or proteins, and ultimately disease or drug response is a challenge of enormous complexity. SNPs located in coding regions may be *silent* or *synonymous*, resulting in no change in the gene product. A *missense* variant is clearly a useful marker for disease. For example, a *missense* polymorphism causes a change in amino acid from a glutamine in a normal hemoglobin gene to a valine in a sickle cell hemoglobin gene; as a result, a homozygote individual for the sickle cell allele develops sickle cell anemia. On the other hand, a sickle cell homozygote individual or a heterozygote individual at the hemoglobin gene locus may develop better resistance to malaria. A *nonsense* variant results in a premature termination codon. This type of polymorphism is responsible for at least some forms of such diseases as Duchenne Muscular Dystrophy (characterized by a damaged *dystrophin* gene) and Cystic Fibrosis (caused by mutations in the *CFTR* gene). On the other hand, nonsense-mediated decay (NMD) provides a cellular machinery for detecting nonsense mutations and preventing the expression of an aberrant protein. A *frameshift* interferes with the triplet nature of gene expression and causes a change in the reading frame of the codons (thus, a change in the resulting translation). A *frameshift* mutation in *NOD2* has been shown to be associated with Crohn's disease [17, 18]. *NOD2* is an outstanding candidate for inflammatory bowel disease (IBD), as tumor necrosis factor signaling and nuclear factor (NF) κ B activation in mononuclear cells are critical for IBD pathophysiology. Yet the *frameshift* in *NOD2* was found to be associated solely with Crohn's disease, and not with ulcerative colitis. SCAN annotates a coding variant with information on how it exerts its functional effect on its host gene by altering the gene product.

An SNP may have other effects on biological molecules. A transcript may contain variation in a region that is not translated. An SNP in a 5' untranslated (5' UTR) region may alter a

binding site for a protein, thus affecting mRNA stability; alternatively, it may promote or inhibit the initiation of translation. A recent study has shown that 5' UTR SNPs in nuclear transcripts encoding both mitochondrial and secreted proteins may influence gene regulation at the level of mRNA export [19]. Similarly, an SNP in a 3' untranslated (3' UTR) region may affect the binding site of a microRNA (miRNA), a posttranscriptional regulator that binds to the target transcript to induce translational repression or gene silencing. Dysregulation of miRNAs has been associated with higher tumor proliferation in human epithelial ovarian cancer [20] and with the etiologies of cardiovascular diseases [21] and of a psychiatric pheno-type [22]. SCAN annotates noncoding SNPs with information on location in the genome relative to the most proximal transcript.

2.3 Transcript Regulation

In contrast to conventional approaches to functional classification, SCAN provides an annotation pipeline to characterize genetic variation with high-throughput molecular phenotypes (e.g., gene expression traits). GWA studies have discovered numerous reproducible associations between common variants and complex human phenotypes, but only a small proportion can be attributed to protein- altering variants. Indeed a substantial proportion of the discovered associations are from noncoding regions; thus, these variants have been hypothesized to alter the expression levels of one or more target genes.

Variation in gene expression is an important feature of human phenotypic variation. In recent years, we have witnessed some transformative advances in the assay of gene expression on a genome-wide scale, which have enabled studies of transcript variation at an unprecedented resolution [23]. These technological advances are likely to contribute further to our understanding of the patterns of modulation of important pharmacokinetic and pharmacodynamic genes, the characterization of pharmacogenetically relevant co-regulated genes, and the elucidation of biological networks.

Studies conducted in LCLs have shown extensive differences in the genomic regulation of gene expression within and among ethnic populations [24]. Of enormous importance to the development of SCAN is the availability of large-scale datasets on the genetic regulation of gene expression. The assimilation of genomic data and high-throughput gene expression data for the identification of regulatory variation is an important approach to SNP functional annotation implemented in SCAN. Studies in our group [25, 26] and others [27, 28] have mapped gene expression variation to particular genomic loci known as expression quantitative trait loci or eQTLs. SCAN uses summary results of SNP associations to transcriptional expression to functionally characterize polymorphisms.

2.4 Multi-locus Linkage Disequilibrium

Genotype–phenotype correlations from GWA studies often span many correlated variants across multiple genes. SCAN annotates an SNP with the set of genes (located in a region spanning 500 kb, centered at the SNP) that contain variants in linkage disequilibrium (LD) with the given SNP. Furthermore, despite the enormous advances in genotyping platforms utilized in GWA studies, it is likely that the causative variants are not genotyped. SCAN provides a framework for interpreting results from GWA studies by assessing, for a given

set of SNPs, the coverage of high-throughput genotyping platforms relative to a reference panel (*see* Fig. 1 for an example). Multilocus LD, which is calculated using haplotype frequencies, provides a way to estimate how much the genotyped variants capture the available information at a locus; multilocus LD can then be used in the choice of genotyping platform for a candidate gene.

3 Implementation

SCAN provides a set of query tools (*see* Fig. 2) using the following interfaces, which are all available in batch query mode:

1. An SNP Query that returns physical and functional annotation, host and flanking genes, and genes whose expression levels are predicted by the variant, at a user-specified *P*-value threshold.
2. A Gene Query that retrieves all variants within and up to a user-specified distance (in kilobases) of the gene, maps the gene to its genomic coordinates relative to the reference assembly and returns the list of local (*cis* -) and distant (*trans* -acting) regulators of the gene. The eQTLs located within the 4 mb of a gene are classified as *cis* -acting; other eQTLs (including those on other chromosomes) are defined as *trans* -acting. Clicking on the gene symbol in the result output provides additional annotation including nomenclature, gene type (e.g., protein-coding), and whether the gene is expressed in the various tissues (e.g., LCL).
3. A Genomic Region Query that returns the list of variants in the specified genomic region, the list of all genes located with the region and all genes predicted to be regulated by the SNPs within the region, at a user-specified *P*-value threshold.
4. A CNV Query that returns the CNV's genomic coordinates, the copy number genotype, genes overlapping the CNV as well as flanking genes, and genes predicted to be regulated by the CNV at a user-specified *P*-value threshold.

For each query, the strength of the association between the eQTL (either *cis* - or *trans* -acting) and the target gene is provided. SCAN reports unadjusted *P*-values since the appropriate multiple testing correction method may vary with the study context (in particular, the number of statistical tests performed in a given study).

4 Applications

In this section, we illustrate the use of SCAN by conducting several *in silico* experiments aimed at exploiting the functional annotation system. We pursue particular research questions (to extend the results of published studies) and, in the process, define the protocols used to conduct these *in silico* experiments.

4.1 eQTLs for Known Asthma Susceptibility Gene

A GWA study of childhood onset asthma [29] implicated *ORMDL3* as a susceptibility gene. Multiple SNPs located on chromosome 17q21 have been found to be strongly and reproducibly associated with risk of disease [30]. *ORMDL3* is a gene that encodes a trans-membrane protein anchored in the endoplasmic reticulum. Particularly, genetic variants

regulating *ORMDL3* expression in *cis* were hypothesized to be determinants of disease susceptibility [29]. We sought to identify regulators of this gene in LCLs to validate the initial findings.

Protocol:

- a. In SCAN's Gene Query tool, enter the gene of interest (*ORMDL3*). SCAN queries can also be done in batch mode by entering or uploading a list, but we restrict the present analysis to this particular gene.
- b. Select “include gene start, end, and chromosome” option.
- c. Select “include SNPs that predict expression with *p* -value less than” and enter 0.0001 as SNP eQTL *p* -value threshold.
- d. Select “include CNVs that predict expression with *p* -value less than” and enter 0.01 as CNV eQTL *p* -value threshold.
- e. Select “Restrict to eQTLs on the same chromosome”.

Results:

The experiment we just conducted replicates many of the eQTLs reported for *ORMDL3* in the original study (*see* Fig. 3), including rs9303277 on *IKZF3* ($p = 4 \times 10^{-10}$), rs2290400 on *GSDMB* ($p = 4 \times 10^{-10}$), rs7216389 on *GSDMB* ($p = 4 \times 10^{-10}$), and a SNP rs4795405 ($p = 3 \times 10^{-6}$) outside of a gene. The first 3 of these SNPs are the top eQTL associations, in samples of European ancestry (CEU), for *ORMDL3* in SCAN. The recent study from the 1000 Genomes Project [11] has identified an SNP rs11078928, a variant in *GSDMB* , to be in strong LD with various SNPs near *ORMDL3* , suggesting that *GSDMB* may be the causative gene in the observed associations.

We set a loose *p* -value threshold since local (*cis*) associations imply a reduced multiple testing burden and, in the case of CNV eQTLs, their identification is exploratory. Going beyond the original study, we identified 2 CNVs on the same chromosome (namely, CNVR7003.1 and CNVR7095.3) that predict *ORMDL3* expression. These are outstanding candidate loci for follow-up studies on asthma susceptibility.

4.2 Functional Annotation of Chemotherapeutic Susceptibility Associated SNPs

The identification of patients who are likely to experience adverse events from a particular chemotherapeutic treatment is an important step in the individualization of cancer chemotherapeutics. A recent study [31] has applied an in vitro genome-wide cell- based model system to identify pharmacogenetic variants that can serve as germ-line genetic biomarkers of carboplatin susceptibility in head and neck cancer (HNC) patients. Two SNPs, rs2551038 and rs6870861, identified in a cell-based model were found to be associated with overall response to carboplatin-based therapy in HNC patients. We sought mechanistic hypotheses for the published SNP associations with chemoresponse.

Protocol:

- a. In SCAN's SNP Query tool, enter the SNPs of interest (rs2551038 and rs6870861). The SNPs should be white space-or comma-delimited.
- b. Select “include SNP info” option.
- c. Select “include host gene and SNP function” option.
- d. Select “include left- and right-flanking genes” option.
- e. Select “include genes that SNP predicts expression for with p -value less than” option and enter 0.0001 as SNP eQTL p -value threshold.

Results:

This protocol identified rs2551038 as an intronic SNP within *HINT1*, a gene that has been shown to act as a tumor suppressor and to exert a pro-apoptotic function [32]. On the other hand, rs6870861 is located outside of a gene, but is flanked on the right by *HINT1*. The two SNPs are regulating the expression of at least ten genes, mostly in *trans*, including *TLR4* ($p = 5 \times 10^{-7}$), the most significantly regulated gene (by both variants). Interestingly, a previous study has shown a link between *TLR4* signaling pathway, inflammation, tumor growth, and chemoresistance [33].

4.3 Identification of a CNV at a BMI-Associated Gene Locus

A polymorphism at *NEGR1* has been shown to be a reproducible association with body mass index at the locus; in particular, a 45 kb deletion at the locus was proposed as a candidate causal polymorphism [34]. We sought to identify copy number polymorphisms in this genomic region and determine their role on genome-wide gene expression.

Protocol:

- a. In SCAN's Gene Query, enter the gene (*NEGR1*) and select “include gene start, end, and chromosome” to identify the gene's genomic coordinates. The gene is located on chromosome 1 (71641212–72520864).
- b. In SCAN's Region query, enter a region (1:71611212:72550864) that spans 30 kb upstream and downstream of the gene.
- c. Select “include SNPs” option to list all SNPs in the specified region.
- d. Select “include genes” option to list all genes located in the region.
- e. Select “include CNVs” option to list all CNVs in the specified region.
- f. Click on the “CNVR216.1” link to retrieve the genomic boundaries and the genotypes of the deletion polymorphism.
- g. In SCAN's CNV Query, enter “CNVR216.1”.
- h. Select “include genes” that CNV predicts expression for with p -values less than option and enter 0.0005 as CNV eQTL p -value threshold.

Results:

This experiment identified a copy number polymorphism (CNVR216.1) that is located downstream from the *NEGR1* gene. The variant is a deletion with copy number genotypes 0, 1, and 2. Furthermore, at the given eQTL threshold, the CNV is associated with the expression of several target genes, including *HDAC5* and *FLJ42957*. Remarkably, *HDAC5* is one of several NAD⁺ dependent histone deacetylases (induction of which has been linked to calorie restriction [35]) found to be down-regulated in the fat tissue of obese co-twins [36]. Redoing the same experiment (using the larger region 1:71611212:72600864) yields a second deletion polymorphism CNVR217.1 of length 45 kb at this locus (the hypothesized causal variant). CNVR217.1 is also an eQTL for several target genes, including the diabetes gene *RALGPS2* [37].

5 Conclusion

We have described the current feature set of a bioinformatics database that facilitates the functional characterization of genetic variation and the identification of potential mechanistic underpinnings of associations from GWA studies. To date, SCAN has received over 800,000 queries. A catalog of eQTLs identified in primary human tissues (including liver, adipose, muscle, and cerebellum) is now being assimilated. Ongoing research efforts (e.g., rare variants from the 1000 Genomes Project, genetic regulation of protein levels and studies of epigenetic regulation of gene expression) promise to expand the scope and relevance of SCAN.

Acknowledgments

This work was funded through Pharmacogenomics of Anticancer Agents Research (PAAR; U01 GM61393), ENDGAME (ENhancing Development of Genome- wide Association Methods) initiative (U01 HL084715), the Genotype-Tissue Expression project (GTEx) (R01 MH090937), Rare Variants and Complex Human Phenotypes (U01HG005773), and the University of Chicago DRTC (Diabetes Research and Training Center; P60 DK20595).

References

1. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. 2009; 106:9362–9367. [PubMed: 19474294]
2. Manolio TA, et al. Finding the missing heritability of complex diseases. *Nature*. 2009; 461:747–753. [PubMed: 19812666]
3. Gamazon ER, et al. SCAN: SNP and copy number annotation. *Bioinformatics*. 2010; 26:259–262. [PubMed: 19933162]
4. Nicolae DL, et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010; 6:e1000888. [PubMed: 20369019]
5. Gamazon ER, et al. Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci. *Proc Natl Acad Sci USA*. 2010; 107:9287–9292. [PubMed: 20442332]
6. Welsh M, et al. Pharmacogenomic discovery using cell-based models. *Pharmacol Rev*. 2009; 61:413–429. [PubMed: 20038569]
7. Gamazon ER, et al. PACdb: a database for cell-based pharmacogenomics. *Pharmacogenet Genomics*. 2010; 20:269–273. [PubMed: 20216476]
8. International HapMap Consortium. The International HapMap Project. *Nature*. 2003; 426:789–796. [PubMed: 14685227]

9. Huang RS, et al. A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. *Proc Natl Acad Sci USA*. 2007; 104:9758–9763. [PubMed: 17537913]
10. Huang RS, et al. Genetic variants associated with carboplatin-induced cytotoxicity in cell lines derived from Africans. *Mol Cancer Ther*. 2008; 7:3038–3046. [PubMed: 18765826]
11. Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–1073. [PubMed: 20981092]
12. Gamazon ER, et al. A pharmacogene database enhanced by the 1000 Genomes Project. *Pharmacogenet Genomics*. 2009; 19:829–832. [PubMed: 19745786]
13. Conrad DF, et al. Origins and functional impact of copy number variation in the human genome. *Nature*. 2010; 464:704–712. [PubMed: 19812545]
14. Mills RE, et al. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011; 470:59–65. [PubMed: 21293372]
15. Gamazon ER, Nicolae DL, Cox NJ. A study of CNVs as trait-associated polymorphisms and as expression quantitative trait loci. *PLoS Genet*. 2011; 7:e1001292. [PubMed: 21304891]
16. Gamazon ER, Huang RS, et al. Copy number polymorphisms and anticancer pharmacogenomics. *Genome Biol*. 2011; 12:R46. [PubMed: 21609475]
17. Ogura Y, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*. 2001; 411:603–606. [PubMed: 11385577]
18. Hampe J, et al. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet*. 2001; 357:1925–1928. [PubMed: 11425413]
19. Cenik C. Genome analysis reveals interplay between 5'UTR introns and nuclear mRNA export for secretory and mitochondrial genes. *PLoS Genet*. 2011; 7:e1001366. [PubMed: 21533221]
20. Zhang L, et al. Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer. *Proc Natl Acad Sci USA*. 2008; 105:7004–7009. [PubMed: 18458333]
21. Yang B, et al. The muscle-specific microRNA miR-1 regulates cardiac arrhythmogenic potential by targeting GJA1 and KCNJ2. *Nat Med*. 2007; 13:486–491. [PubMed: 17401374]
22. Hansen T, et al. Brain expressed microRNAs implicated in schizophrenia etiology. *PLoS One*. 2007; 2:e873. [PubMed: 17849003]
23. Pickrell JK, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*. 2010; 464:768–772. [PubMed: 20220758]
24. Spielman RS, et al. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet*. 2007; 39:226–231. [PubMed: 17206142]
25. Zhang W, et al. Identification of common genetic variants that account for transcript isoform variation between human populations. *Hum Genet*. 2009; 125:81–93. [PubMed: 19052777]
26. Duan S, et al. Genetic architecture of transcript-level variation in humans. *Am J Hum Genet*. 2008; 82:1101–1113. [PubMed: 18439551]
27. Stranger BE, et al. Population genomics of human gene expression. *Nat Genet*. 2007; 39:1217–1224. [PubMed: 17873874]
28. Storey JD, et al. Gene-expression variation within and among human populations. *Am J Hum Genet*. 2007; 80:502–509. [PubMed: 17273971]
29. Moffatt MF, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*. 2007; 448:470–473. [PubMed: 17611496]
30. Bouzigon E, et al. Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med*. 2008; 359:1985–1994. [PubMed: 18923164]
31. Ziliak D, et al. Germline polymorphisms discovered via a cell-based, genome-wide approach predict platinum response in head and neck cancers. *Transl Res*. 2011; 157:265–272. [PubMed: 21497773]
32. Weiske J, Huber O. The histidine triad protein Hint1 triggers apoptosis independent of its enzymatic activity. *J Biol Chem*. 2006; 281:27356–27366. [PubMed: 16835243]
33. Kelly MG, et al. TLR-4 signaling promotes tumor growth and paclitaxel chemoresistance in ovarian cancer. *Cancer Res*. 2006; 66:3859–3868. [PubMed: 16585214]
34. Willer CJ, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet*. 2009; 41:25–34. [PubMed: 19079261]

35. Yang T, Sauve AA. NAD metabolism and sirtuins: metabolic regulation of protein deacetylation in stress and toxicity. *AAPS J.* 2006; 8:E632–E643. [PubMed: 17233528]
36. Pietilainen KH, et al. Global transcript profiles of fat in monozygotic twins discordant for BMI: pathways behind acquired obesity. *PLoS Med.* 2008; 5:e51. [PubMed: 18336063]
37. Hayes MG, et al. Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. *Diabetes.* 2007; 56:3033–3044. [PubMed: 17846124]

Mid Distribution of CAPN10 on Select Platforms

Platform	Population	Average Multi-locus LD	Q1 Multi-locus LD	Median Multi-locus LD	Q3 Multi-locus LD	No. of HapMap SNPs (MAF > .05)	No. of HapMap SNPs (MAF ≤ .05)
Affy6	CEU	0.874	0.747	0.995	1	26	2
Affy6	YRI	0.452	0.264	0.434	0.526	16	12
ill1M	CEU	0.894	0.729	1	1	26	2
ill1M	YRI	0.545	0.279	0.478	0.747	16	12

Fig. 1.

Platform coverage. SCAN enables the interpretation of results from GWA studies by assessing, for a given set of SNPs, the coverage of high-throughput genotyping platforms relative to a reference panel

SCAN
SNP and CNV Annotation Database

Home Gene SNP Region CNV LD Annotations

Enter a gene or a set of genes:

or choose a file with a list of genes:

☐ include gene start, end, and chromosome

☐ include SNPs inside gene and up to kb from the gene

☐ include SNPs that predict expression with p-value less than and frequency greater than

☐ include CNVs that predict expression with p-value less than for population

☐ Restrict to eQTLs on the same chromosome

[sample input file](#)

Fig. 2. Functional annotation. SCAN provides a set of interfaces for annotating SNPs, genes, genomic regions and CNVs with high-throughput molecular phenotypes (e.g., gene expression)

Fig. 3. Novel susceptibility loci identified from SCAN's annotation pipeline. SCAN replicates many of the eQTLs reported for *ORMDL3* in the original GWA study (Moffatt, M.F., et al., 2007) of childhood onset asthma. In addition to the reported loci, we identified 2 CNVs on the same chromosome (namely, CNVR7003.1 and CNVR7095.3) that predict *ORMDL3* expression. These are outstanding candidate loci for follow-up studies on asthma susceptibility