World Journal of Diabetes

Online Submissions: http://www.wjgnet.com/esps/
bpgoffice@wjgnet.com
doi:10.4239/wjd.v5.i2.97

World J Diabetes 2014 April 15; 5(2): 97-114
ISSN 1948-9358 (online)

TOPIC HIGHLIGHT

WJD 5th Anniversary Special Issues (2): Type 2 diabetes

Expression quantitative trait analyses to identify causal genetic variants for type 2 diabetes susceptibility

Swapan Kumar Das, Neeraj Kumar Sharma

Swapan Kumar Das, Neeraj Kumar Sharma, Section on Endocrinology and Metabolism, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC 27157, United States

Author contributions: Das SK performed the literature review, integrated key scientific concepts, and wrote the paper; Sharma NK performed the literature and data mining and reviewed the manuscript.

Supported by National Institutes of Health (NIH/NIDDK), Nos. R01 DK090111 and DK039311

Correspondence to: Swapan K Das, PhD, Section on Endocrinology and Metabolism, Department of Internal Medicine, Wake Forest School of Medicine, Medical Center Boulevard, NRC Building # E159, Winston-Salem, NC 27157, United States. sdas@wakehealth.edu
Telephone: +1-336-7136057 Fax: +1-336-7137200

Received: November 27, 2013 Revised: February 21, 2014
Accepted: March 13, 2014
Published online: April 15, 2014

Abstract

Type 2 diabetes (T2D) is a common metabolic disorder which is caused by multiple genetic perturbations affecting different biological pathways. Identifying genetic factors modulating the susceptibility of this complex heterogeneous metabolic phenotype in different ethnic and racial groups remains challenging. Despite recent success, the functional role of the T2D susceptibility variants implicated by genome-wide association studies (GWAS) remains largely unknown. Genetic dissection of transcript abundance or expression quantitative trait (eQTL) analysis unravels the genomic architecture of regulatory variants. Availability of eQTL information from tissues relevant for glucose homeostasis in humans opens a new avenue to prioritize GWAS-implicated variants that may be involved in triggering a causal chain of events leading to T2D. In this article, we review the progress made in the field of eQTL research and knowledge gained from those studies in understanding transcription regulatory mechanisms in human subjects. We highlight several novel approaches that can integrate eQTL analysis with multiple layers of biological information to identify ethnic-specific causal variants and gene-environment interactions relevant to T2D pathogenesis. Finally, we discuss how the eQTL analysis mediated search for "missing heritability" may lead us to novel biological and molecular mechanisms involved in susceptibility to T2D.

Key words: Type 2 diabetes; Single nucleotide polymorphisms; Expression quantitative trait locus; Expression regulatory SNPs; Gene-environment interaction; Genome-wide association study

Core tip: Identification of genetic variants that modulate the susceptibility to disease and elucidating their function at the molecular level is a major focus of type 2 diabetes (T2D) research. This article highlights the utility of expression quantitative trait analysis in discovering regulatory variants that increase susceptibility to T2D by modulating the expression of transcripts in tissues important for glucose homeostasis.

Das SK, Sharma NK. Expression quantitative trait analyses to identify causal genetic variants for type 2 diabetes susceptibility. *World J Diabetes* 2014; 5(2): 97-114 Available from: URL: http://www.wjgnet.com/1948-9358/full/v5/i2/97.htm DOI: http://dx.doi.org/10.4239/wjd.v5.i2.97

GENETIC DISSECTION OF TYPE 2 DIABETES SUSCEPTIBILITY

Diabetes is one of the most prevalent metabolic disor-

ders, characterized by elevated levels of plasma glucose, and is responsible for significant mortality and morbidity in human populations worldwide[1]. The latest estimate from the International Diabetes Federation indicates a global prevalence rate of 8.4% in adults and 382 million cases of diabetes in 2013[2]. It is one of the common diseases with a well-accepted genetic contribution[3]. Type 2 diabetes (T2D), a late onset subtype of diabetes, results from a derangement in the complex interplay of multiple physiological processes known to be involved in systemic glucose homeostasis. These processes include peripheral glucose uptake in muscle, secretion of hormones and incretins form pancreas and intestine, secretion of cytokines/adipokines from adipose tissue, hepatic glucose production, and neuro-endocrine regulation by central nervous system[4,5]. However, the relative contribution of these processes to T2D pathogenesis is debated. Based on this knowledge on intertwined and complex physiological processes it can be anticipated that T2D is a heterogeneous conglomeration of phenotypes, caused by multiple genetic perturbations and affecting different biological pathways. Predictably, deciphering the genetic etiology of T2D has remained challenging.

Until the last decade, searching for an association between T2D and sequence variants of selected candidate genes was the mainstay of research for finding genetic susceptibility factors. Based on available technology in those studies, researchers selected candidate genes either from loci detected by genome-wide linkage analyses or based on known physiological functions. In our earlier reviews, we discussed the knowledge gained from such studies in detail[6,7]. Success from those endeavors was very limited. However, this approach has identified genetic variants in the *TCF7L2* gene, to date is the best replicated and strongest (relative risk approximately 1.4) genetic susceptibility factor for T2D[8], but its role is still controversial[9-11].

In the middle of the last decade, a transformative change took place in the field of genetics of complex disease research. Advances in high-throughput genotyping technology, availability of the complete human genome sequence, a dense catalogue of common genetic variants, and a population-specific linkage disequilibrium map of these variants lead to the implementation of genome-wide association studies (GWAS), which interrogate the entire genome to identify common genetic variants (minor allele frequency $\geqslant 0.05$) associated with a disease[12]. GWAS have yielded unprecedented success in identifying well-replicated susceptibility loci for T2D, glucose homeostasis traits, obesity, and related metabolic phenotypes[3,13-15]. Nevertheless, these successes come with significant caveats. Based on the most recent analyses, the 63 T2DM-associated loci discovered so far in Caucasian populations together account only for 5.7% of the liability-scale variance in disease susceptibility, and sibling relative risk (λ_s) attributed jointly by these variants is 1.104[13]. Moreover, few of the T2D loci identified primarily in European- or Asian-derived populations are convincingly replicated in African American, Native American, and Hispanic populations, all of whom have a higher prevalence of T2D than Caucasians[14,16]. These GWAS-identified loci do not appear to explain the well-established roles for adipose, muscle, and liver in diabetes pathogenesis[17], and few of these loci have been linked to a molecular mechanism. Several investigators have attempted to implicate function to T2D-associated loci based on their proximity to a gene, assuming that the associated single nucleotide polymorphisms (SNP) alters the function of a nearby gene[18]. Some have drawn enthusiastic conclusions about the role of these variants exclusively in insulin secretion[19]. However, proof of such an assumption is lacking. Given the small effect on T2D susceptibility and the statistical noise inherent in performing 10^6 or more tests, exclusive reliance on larger T2D GWAS alone is unlikely to identify the source of undefined T2D susceptibility (often referred to as "missing heritability"[20]).

EXPRESSION QUANTITATIVE TRAITS: MOLECULAR ENDOPHENOTYPES

One of the major findings from the T2D GWAS is that most of the trait-associated SNPs are located in intronic, intergenic, or other non-coding regions of the genome[3,21]. Further fine mapping analysis also failed to find any coding or other variants that would provide a molecular biological explanation of the elevated disease risk attributed by these loci.

The abundance of a transcript is a quantitative trait. Studies in human populations showed a wide, heritable variation of transcript levels among individuals, and thus lead to the concept of "expression quantitative trait loci" (eQTL)[22,23]. The heritability of eQTLs has been replicated in multiple human tissue or cell types, with approximately 30% of eQTLs having $h^2 > 0.3$, and an estimated 58%-85% being heritable[24-28]. The abundance of a transcript can be directly modified by polymorphisms in non-coding regulatory elements. Many SNPs are associated with quantitative transcript levels and are considered as expression regulatory SNPs (eSNPs). eSNPs close to the transcription start sites (TSS) of the eQTLs are named "*cis*" or "local" eSNPs , whereas eSNPs located $> \pm 500$ kb from the TSS or on a different chromosome are considered "*trans*" or "distal" eSNPs[22,29]. Similar to a published study[30], here we will refer to eQTLs as the transcripts rather than SNP-transcript pair, and eSNPs as the genetic variants (SNPs) associated with the expression profile of a transcript.

Based on this knowledge, many laboratories (including ours) hypothesized that GWAS-associated non-coding variants are eSNPs and can modulate T2D susceptibility by altering transcript levels (or splicing). This concept is based on the "central dogma" of gene expression and presents a causal model of genetic susceptibility (Figure 1). In this model, transcript abundance is considered as an intermediate phenotype between genetic loci (DNA sequence variants) and subclinical (*e.g.*, insulin resistance)

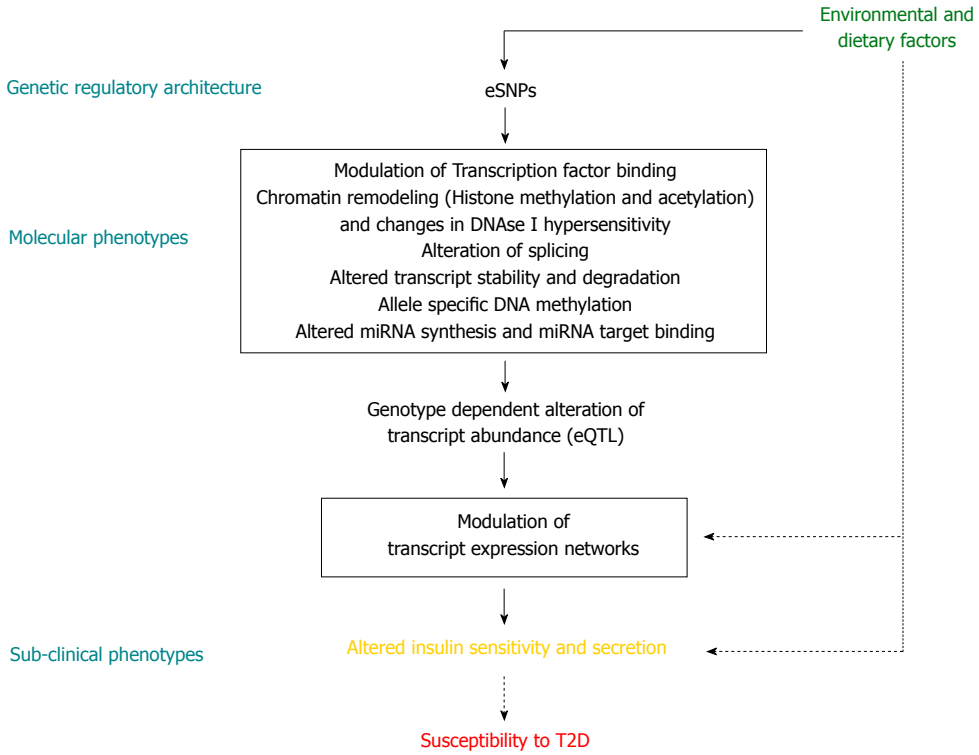

Figure 1 A causal model of genetic susceptibility. Genetic regulatory architecture modulates molecular phenotypes in interaction with environmental factors and alters disease susceptibility. eSNP: Expression regulatory SNP; eQTL: Expression quantitative trait loci; T2D: Type 2 diabetes.

or clinical (*e.g.*, T2D) phenotypes. Since transcript abundance is a proximal molecular endophenotype affected by genetic variants, it is likely to be a less heterogeneous phenotype (compared to complex clinical phenotypes like those of T2D), and thus more amenable to genetic mapping methods due to superior statistical power.

EQTL MAPPING

Study designs and analytical frameworks for eQTL mapping are similar to those for mapping any other quantitative traits [*e.g.*, body mass index (BMI), fastin glucose, glycosylated hemoglobin]. However, genetic analysis of human phenotypes including QTLs carries a unique set of problems[29]. In general, eQTL analyses integrate genome-wide expression (in tissues or cells) and genotype data in multiple individuals (related or unrelated). These analyses use linkage- or association-based statistical genetic methods to map regulatory regions and genetic variants that may explain individual variations in transcript expression. Microarray- or RNA-seq[31-33] based methods are used to generate large numbers of quantitative transcript phenotypes. Therefore, the number of statistical tests involved in eQTL mapping studies is significantly higher than in traditional QTL analysis[34]. A detailed discussion on methods used in eQTL analysis is beyond the scope of this article, and we refer our readers to other reviews on this specific subject[29,34-36].

Published eQTL studies have implemented linkage analysis by using 400-2000 microsatellite makers[24,26] to localize regulatory intervals, whereas other studies have genotyped large numbers of common SNPs (> 100000) to discover the eSNPs[25,28,37] associated with eQTLs. With the advancement of genomic technology, we can now simultaneously genotype more than 4.5 million SNPs or can have a whole genome sequence for each individual included in an eQTL study by highly multiplexed "next generation" sequencing methods[38]. These advances pose additional statistical and computational challenges, and will require appropriate correction and adjustment of significance thresholds for the massive number of independent tests performed (and hypotheses tested) to control false discovery. The power to detect eSNPs depends on their effects (average difference in the transcript abundance between genotypes, scaled by the standard deviation of the transcript abundance within genotype classes) and allele frequency[34]. Consequently, detection of eSNPs with a lower effect allele frequency and a lower effect size will require a larger sample size.

One interesting observation from published eQTL studies is that most of the strong eSNPs are located near the TSS with no discernable trend in the 5' or 3' direction[28,39,40]. As a result, most studies consider SNPs within close proximity of the TSS (± 500 kb window) as *cis*-eSNPs. Since the genomic context of most eQTL transcripts are known, statistical adjustment for the actual number of SNPs tested within 500 kb will be more appropriate for *cis*-eSNP discovery. Any SNP outside the *cis*-region is tested as a *trans*-eSNP for a transcript. The molecular biological basis of trans-regulation is less studied;

current information suggests that the variants that affect transcription factors, miRNAs, or long-range chromatin interaction may act as trans-eSNPs. To identify *trans*-eSNPs, the number of tests needed is far greater, and the tests require more stringent significance threshold criteria and a larger sample size. Thus, use of a false discovery rate based on a permutation analysis to correct for multiple testing[34], and considering the correlation among transcript levels and highly correlated SNP structures, are useful approaches to identify this biologically important class of regulatory SNPs.

Several heterogeneous sources of variability hidden in the data may lead to both spurious eSNPs and missed associations in eQTL analyses if not properly addressed. Statistical models that correct for hidden structures within the sample (such as race, admixture, and family relatedness), artifacts in expression data (including batch effects and probe bias), environmental influences, and other known and unknown factors are required to improve sensitivity and interpretability of eQTL analyses[41]. Methods that showed significant usefulness in tackling these confounding factors include Bayesian approaches developed by Stegle *et al*[42] (implemented in probabilistic estimation of expression residuals or probabilistic estimation of expression residuals software), linear mixed-effects model-based approaches developed by Listgarten *et al*[43] (implemented in LMM-EH-PS or Linear Mixed Model-Expression Heterogeneity-Population Structure software), surrogate variable analysis, and inter-sample correlation emended approaches[44,45].

The heavy computational burden involved in eQTL analyses sometimes forces researchers to restrict their analysis to a small subset of selected transcripts and SNPs. Improvement of computational algorithms, parallelization of programs by efficient scripting, and utilization of efficient processing hardware are among many approaches needed to improve scalability and computational efficiency required for eQTL analyses. Implementation of these approaches will enhance discovery by increasing the capacity to utilize the complete data set[46,47].

EQTLS AND DISEASE GENE MAPPING

Molecular and cell biological experiments in model organisms and cells have significantly advanced our understanding about the role of non-coding DNA sequences in genetic regulation, transcriptional circuitry, the transcriptional apparatus, and chromatin regulation. This work has led to new insights into the complex mechanisms involved in dysregulation of gene expression in various human diseases[48]. Recent genome-wide studies in human cells by different international consortia [including ENCyclopedia Of DNA elements (ENCODE)][49] further have improved our mechanistic understanding of the role of DNA sequence variants in quantitative modulation of gene expression[50-52]. eQTL studies have been extensively used to identify genetic regulators involved in natural variation of gene expression[28,37,39] and to understand tissue-specific architecture of genetic regulatory mechanisms[24,30,53-59].

However, an intriguing application of eQTL mapping is the use of eSNP data to interpret disease or disease-related phenotypic association signals, and thereby elucidate specific biological mechanisms underlying the increased genetic risk attributed by the DNA sequence variants. Identification of genetic variants simultaneously associated with disease and eQTLs (in relevant tissue) significantly facilitates identification of potential causal genes. Discovery of genetic variants in *ORMDL3* as a susceptibility factor for childhood-onset asthma[60] and *VNN1* variants that influence high-density lipoprotein cholesterol concentrations[26] are two early examples of the successful implementation of eQTL mapping in disease gene hunting. The review by Cookson *et al*[36] offer a more detailed discussion on those success stories.

Several recent studies have integrated GWAS and eQTL analyses (data generated in different sets of subjects) and have used the overlap of two signals as a tool to interpret GWAS findings. Although this work is a good starting point, we need to be cautious about using the overlap of two statistical signals (eSNP and the disease phenotype-associated SNP/phSNP). Careful thought is required before making a claim of identifying a disease-causing variant. Montgomery and Dermitzakis (2011) described three situations[41] when a coincidence of eQTL and disease phenotype GWA signal may distract from identification of causal variants: (1) eSNP and phSNP are in the same linkage disequilibrium (LD) block but are two different SNPs. This is not considered as exact overlap, and they may tag different causal variants; (2) eSNPs and phSNPs are the same but SNP density differs between the eQTL and GWAS data. Lack of proper resolution in one or both studies may be misleading and will not elucidate the correct functional SNP; and (3) eSNPs may have a pleiotropic effect and may regulate the expression of "gene Y" in "tissue 1", but the same eSNP may regulate the expression of "gene X" in "tissue 2". Thus, if the eQTL study is done in "tissue 1" (a "surrogate" tissue) but not in "tissue 2" (the "disease-relevant" tissue in which the true causal effect is manifested), then despite the overlap of eSNPs and phSNPs, we will incorrectly link "gene Y" to the disease phenotype.

In general, eSNPs that are universal have a stronger effect, but a significant proportion of eSNPs show tissue-specific effects[30,53,54]. However, it is difficult to select "relevant" tissue, or the relevant tissue may not be accessible from human subjects for analysis for many complex diseases. Ongoing efforts of international consortia, including GTEx, to develop multi-tissue eQTL databases (Table 1) is a significant step forward in addressing this limitation[61-64].

Many investigators have developed statistical approaches to formally test the overlap of GWAS and eQTL signals to distinguish accidental colocalization from true sharing of causal variants. The regulatory trait concordance method designed by Nica *et al*[65] accounts for local LD structure and integrates eQTL and GWAS results to reveal the subset of association signals due to

Table 1 Selected expression quantitative trait loci databases

Database	Website (URL)	Cell/tissue type	Project
eQTL Browser	http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/	LCL, liver, brain, fibroblast, T-cell	17 projects
Genvar	http://www.sanger.ac.uk/resources/software/genevar/	Adipose, LCL Skin fibroblast from healthy female twins	MuTHER
		LCL from 8 populations	Hapamap3
		Fibroblast, LCL and T-cell from umbilical cord	GenCord
GTEx eQTL Browser	http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi	Multiple tissues including liver, brain regions, LCL	GTEx
PACdb	http://www.pacdb.org/	Gene-drug or GXD eSNPs from LCL model	Dolan and Cox lab
SGR Database	http://systems.genetics.ucla.edu/	22 mouse and several human datasets. Data includes aortic endothelial and smooth muscle, adipose, brain, liver, macrophages and muscle tissue Includes GXE eSNP data from cell experiments	Lusis lab
SCAN	http://www.scandb.org/newinterface/about.html	CEU and YRI LCLs from HapMap	Cox Lab
seeQTL	http://www.bios.unc.edu/research/genomic_software/seeQTL/	HapMap LCLs	

SGR: Systems genetics resource; eQTL: Expression quantitative trait loci; T2D: Type 2 diabetes; eSNP: Expression regulatory SNP; LCL: Lymphoblastoid cell lines; GXD: Gene-by-drug interaction; GXE: Gene-by-environment interaction; CEU: HapMap caucasian from CEPH collection; YRI: HapMap African from Yuroba, Nigeria.

cis- or *trans*-eQTLs. He *et al*[66] (2013) developed an algorithm named "Sherlock" based on a Bayesian statistical framework to identify potential gene-disease associations by matching genetic signatures of expression (collective information of *cis*- and *trans*-eSNPs) of a gene to that of the disease phenotype by using GWAS data of the disease and the eQTL data of related tissue. These novel approaches are likely to expand our ability to harvest new insights from genetic association studies for disease phenotypes.

T2D-ASSOCIATED VARIANTS ARE ESNPS IN TISSUES IMPORTANT FOR GLUCOSE HOMEOSTASIS

Genome-wide eQTL analyses in transformed lymphocytes (lymphoblastoid cell lines, or LCLs) provided the first evidence that SNPs associated with complex diseases phenotypes are more likely to be eSNPs than minor allele frequency-matched SNPs randomly selected from high-throughput GWAS genotyping platforms. Nicolae *et al*[67] (2010) utilized an Affymetrix GeneChip Human exome 1.0 ST array to generate exon-level expression data of LCLs from 87 Caucasian (CEU) and 89 African (YRI) subjects from the HapMap project. They performed a quantitative-trait transmission disequilibrium test to identify eSNPs from 2 million genotyped SNPs. A study by Nica *et al*[65] (2010) utilized an Illumina Sentrix WG-6-V2 whole-genome expression array to generate total transcript-level expression data of LCLs from 109 unrelated CEU subjects (from the HapMap 3 project) and performed Spearman rank correlation analysis to identify eSNPs from 1186075 genotyped SNPs. Key findings from these studies[65] include: (1) SNPs reproducibly associated with complex human traits are likely to be eSNPs; (2) Enrichment of complex trait GWAS-implicated SNPs are more evident among *cis*-eSNPs but not among *trans*-eSNPs; and (3) eSNPs discovered in LCLs are more strongly enriched for SNPs associated with immunity-related conditions (*e.g.*, Crohn's disease, type 1 diabetes, rheumatoid arthritis), but such enrichment was not observed for metabolic disorders (*e.g.*, T2D and coronary artery disease). These studies indicate that eQTL studies using surrogate tissue samples may be helpful for some diseases. However, understanding the functional role of T2D-associated SNPs will probably require expansion of eQTL studies into tissues more relevant for T2D pathophysiology. These studies also had significantly lower power to identify *trans*-eQTLs due to comparatively small sample sizes, and will require reevaluation of the role of *trans*-eSNPs in larger sample sets.

Zhong *et al*[68] (2010) used genetics of gene expression (GGE) analysis in tissues from two cohorts of human subjects (Cohort 1: liver-specific GGE cohort with post mortem liver samples from 427 subjects; Cohort 2: liver, subcutaneous adipose and omental adipose from 922 subjects who had Roux-en-Y gastric bypass surgery). They identified 18785 unique eSNPs in the combined set of data. They found 2189, 2286, and 1999 eSNPs specific to liver, omental adipose, and subcutaneous adipose, respectively. However, they also noticed that 72% of *cis*-eSNPs identified in liver, 79% of those found in omental adipose and 80.5% from subcutaneous tissue were also found in the other two tissues. Given the metabolic relevance of these tissues, they further interrogated data from three large-scale T2D GWAS datasets to test whether the set of eSNPs were more likely to be associated with T2D compared to randomly selected SNPs. These tissue eSNPs showed a significant enrichment of T2D-associated SNPs. For example, in the DIAGRAM (DIAbetes Genetics Replication and Meta-analysis) GWAS data set, 7.34% of the eSNPs showed a significant association with T2D ($P < 0.05$) compared to an average of 6.12% SNPs in the random sets, representing a modest 1.20-fold enrichment for SNPs in the eSNP (or SNP in LD at $r^2 > 0.89$) set over the random sets (p-enrichment = 1.33×10^{-9})[68]. In that study, omental adipose tissue eSNPs also showed

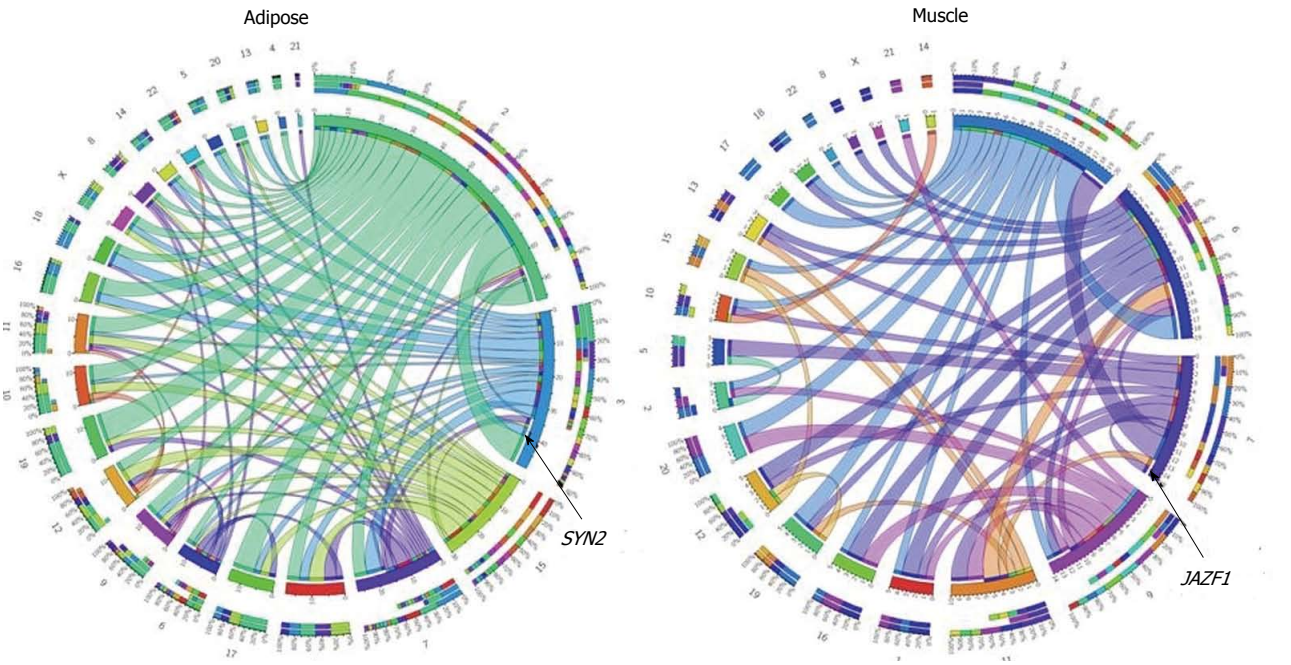

Figure 2 Type 2 diabetes or glucose homeostasis traits associated variants are expression regulatory SNP. We tested *Cis* and *Trans* regulatory role of 68 SNPs that showed reproducible associations with T2D or Glucose homeostasis traits[72]. At a threshold of $P < 0.0001$, 25 and 19 of these SNPs in adipose and muscle, respectively, showed association with expression of a *cis*- or trans-transcript. This figure represents a CIRCOS plot of eQTL and eSNP chromosomal location relationships, indicating the predominance of *trans*-regulation among 183 and 62 significant ($P < 0.0001$) eQTL-eSNPs associations in adipose and muscle respectively. Rare *cis*-regulation (*SYN2* in adipose and *JAZF1* in muscle) is marked. eSNP: Expression regulatory SNP; eQTL: Expression quantitative trait loci; T2D: Type 2 diabetes.

further significant enrichment when restricted to adipose expression network genes differentially expressed with T2D. Thus, these studies support the notion that T2D-associated SNPs may modulate expression of transcripts in tissues relevant for glucose homeostasis.

Fu *et al*[53] (2012) analyzed eQTLs in blood ($n = 1240$) and other tissues (liver, $n = 62$; muscle, $n = 62$; subcutaneous adipose, $n = 83$; and visceral adipose, $n = 77$); out of 1954 SNPs associated with complex disease traits from a GWAS catalogue, 907 were *cis*-eSNPs. However, 28.7% of these trait-associated *cis*-eSNPs showed a tissue-specific (in blood versus other tissue) and discordant effect on gene expression. The discordant effect includes tissue-specific regulation, alternative regulation by different eSNPs, different effect size and, in a few cases, opposite allelic direction. The study also showed that SNPs associated with complex traits are more likely ($P = 2.6 \times 10^{-10}$) to exert a tissue- specific effect on gene expression[53]. No comparisons were made between other tissues due to small sample size. This study indicates that use of tissues in eQTL analysis may have implications for inferring transcriptional effects of SNPs, especially for the complex disease susceptibility variants.

This work also emphasizes the importance of investigating disease-relevant tissue for characterizing functional effects of T2D and other disease-associated variants. However, it is difficult to determine "relevant tissue" even for diseases with known pathophysiology. T2D is clearly of polygenic etiology, and relevant tissue could be distinct for genes involved. Moreover, gene expression is regulated by environmental (*e.g.*, diet), epigenetic, and other unknown factors, and eQTL discovery from tissue samples may be affected by the physiological state of the donors[41]. For example, profound hyperglycemia and dyslipidemia observed in T2D subjects will modulate and even may mask primary causal changes in genetic regulatory networks. Thus, multi-tissue eQTL analysis in physiologically characterized individuals could be a safe option to scrutinize the circularity of cause and effect in genetic regulatory signals, and holds the promise to offer insights into the novel mechanisms driving genetic susceptibility to T2D.

Most initial eQTL studies seeking to identify a regulatory role for T2D-associated SNPs have focused on *cis*-eQTLs. However, studies by Voight *et al*[69] (in adipose, $n = 603$; and blood, $n = 745$ subjects) and our laboratory (in adipose and muscle of 168 non-diabetic subjects who were physiologically evaluated) showed that only a few top T2D GWAS-identified signals can be explained as *cis*-eQTLs, and T2D-associated non-coding SNPs are less likely to regulate expression of the closest gene[70]. Results were similar in an eQTL analysis that used human islet cells from 63 cadaver donors[71]. A genome- wide study by our laboratory[72] in adipose and muscle tissue of 62 subjects (31 insulin- resistant and 31 insulin-sensitive subjects matched for BMI) showed that at a less stringent threshold ($P < 0.0001$), among 68 well-replicated T2D/glucose homeostasis-associated SNPs, 25 and 19 of them were eSNPs in adipose and muscle, respectively (Figure 2). However, after stringent (Bonferroni) correction, only SNP rs13081389 was a *cis*-eSNP for the *SYN2* gene in adipose ($P < 4.7 \times 10^{-8}$, 15507 expressed transcripts were

tested in adipose). Interestingly, these 68 SNPs showed significant enrichment for *trans*-eSNPs in adipose and muscle, but not in LCLs[72]. Many of these *trans*-eSNPs show associations with expression of ⩾ 10 transcripts and may be a "master regulator". Expanding this search for the top 1000 T2D-associated SNPs from a Wellcome Trust Case Control study also confirmed the *trans*/distal regulatory SNPs[72]. We also showed that replicated T2D- and glucose homeostasis-associated SNPs are enriched for *trans*-eQTLs for transcripts differentially expressed between insulin-resistant and insulin-sensitive people[72]. A recent eQTL study using a large cohort of blood samples also supported the *trans*-regulatory role of 233 complex trait-associated SNPs[73]. Thus, the genetic regulatory architecture of T2D is complex, tissue-specific, and likely extends beyond the *cis*-regulatory mechanism.

EQTL ANALYSIS FOR PRIORITIZING T2D-ASSOCIATED VARIANTS TO IDENTIFY NOVEL CANDIDATE GENES

The multiple testing corrections utilized in genome-wide statistical analyses allow detection of only the strongest effects and penalize weaker associations that may be biologically meaningful[74]. Investigators have implemented several approaches to prioritize T2D association signals from large GWAS datasets to identify biological mechanisms responsible for genetic predisposition. One common approach includes selection of genes close to T2D GWAS-implicated SNPs and shows differential expression in T2D subjects compared to normoglycemic subjects (or in animal models of T2D). This approach is based on the idea that T2D-associated variants may modulate the expression of nearby genes in tissues important for glucose homeostasis. Parikh *et al*[75] used publicly available expression microarray data from different tissues (pancreas, adipose, muscle, and liver from T2D patients and rat models of T2D) to prioritize among the 275 genes located near 1170 T2D GWAS-implicated SNPs. A recent study by Taneera *et al*[71] used expression profiling of human pancreatic islet cells for functional prioritization of genes in the vicinity of 47 T2D-associated SNPs. However, available data from several human tissue eQTL analyses indicate that only a few T2D-associated SNP act as *cis*-eSNPs, and no enrichment of differentially expressed genes was observed around T2D GWAS-implicated variants[72]. Thus, a logical alternative for prioritizing T2D-associated variants is to utilize a reverse genetics approach and restrict the genetic search space to the subset of variants that are eSNPs in relevant tissues. These eSNPs are statistically associated with expression of transcript and thus have a strong possibility of being a "key driver" in perturbing gene-expression regulatory networks.

Selecting the genes based on eSNPs among those also associated with T2D in large GWAS datasets will prioritize genes with a significantly high chance of being causally involved with susceptibility to T2D, and thus may be helpful in identifying additional genetic susceptibility loci from GWAS datasets. A genome-wide analysis of adipose tissue transcriptomes from 62 insulin-resistant and -sensitive subjects identified 172 differentially expressed transcripts[76]. We checked adipose eQTL data from the MuTHER study[55] to find eSNPs of these differentially expressed transcripts. We further mined the DIAGRAM GWA meta-analysis results[13] for association of these eSNPs with T2D. This analysis[77] identified that the strongest *cis*-eSNP (rs11037579, $P = 4.21 \times 10^{-6}$) for the *HSD17B12* in adipose tissue was also associated with T2D [$P = 3.80 \times 10^{-4}$, OR = 1.06 (95%CI: 1.03-1.1)]. Individuals carrying the T2D risk allele T for the intronic SNP rs11037579 had lower expression of *HSD17B12* in adipose tissue. This result corroborates the finding that *HSD17B12* expression is downregulated in the adipose tissue of insulin-resistant subjects. The *HSD17B12* gene codes a bifunctional enzyme involved in the biosynthesis of estradiol and the elongation of very long chain fatty acids. Several variants within ± 500 kb of this gene are eSNPs (including a 3'UTR SNP rs1061810) in adipose, LCL, and other tissues, and show an association with T2D (although below the genome-wide threshold) (Figure 3). Further functional studies will be required to identify true causal SNPs. However, this integrative approach demonstrates the validity of such an approach in prioritizing novel T2D susceptibility loci. In fact, two recent integrative genomic studies showed that eSNPs for *PFKM* (SNP rs11168327) gene in muscle and *ARAP1* (SNP rs11603334) gene in pancreatic beta cell are associated with T2D[78,79].

EQTL AND BIOLOGICAL NETWORK ANALYSIS TO IDENTIFY ETHNIC-SPECIFIC GENES FOR T2D:

Age-standardized prevalence of T2D varies among ethnic and racial groups[14,80]. T2D is almost twice as prevalent in adult non-Hispanic African Americans (14.9%) in the United States compared to European Americans (7.6%)[81]. Yet only a few of the associated T2D-loci - identified primarily in European- or Asian-derived populations - are replicated in African American, Hispanics, and Native Americans[14,16,82-84]. Intriguingly, studies have identified distinctive physiologic features of glucose homeostasis in African Americans and Hispanics[85-87]. Compared to non-Hispanic Caucasians matched on age, gender, and BMI, African Americans are more insulin-resistant (lower S_I), but show a greater acute insulin response to intravenous glucose (AIR_G) and a higher disposition index ($DI = S_I \times AIR_G$). A genetic basis for these physiological differences seems likely, but remains unidentified.

Published studies of expression across ethnic groups (mostly restricted to lymphocytes or HapMap LCLs) showed distinct ethnic-specific expression[37,88-90]. Zhang *et al*[90] (2008) reported differential expression of up to 67% of transcripts between LCLs from subjects of

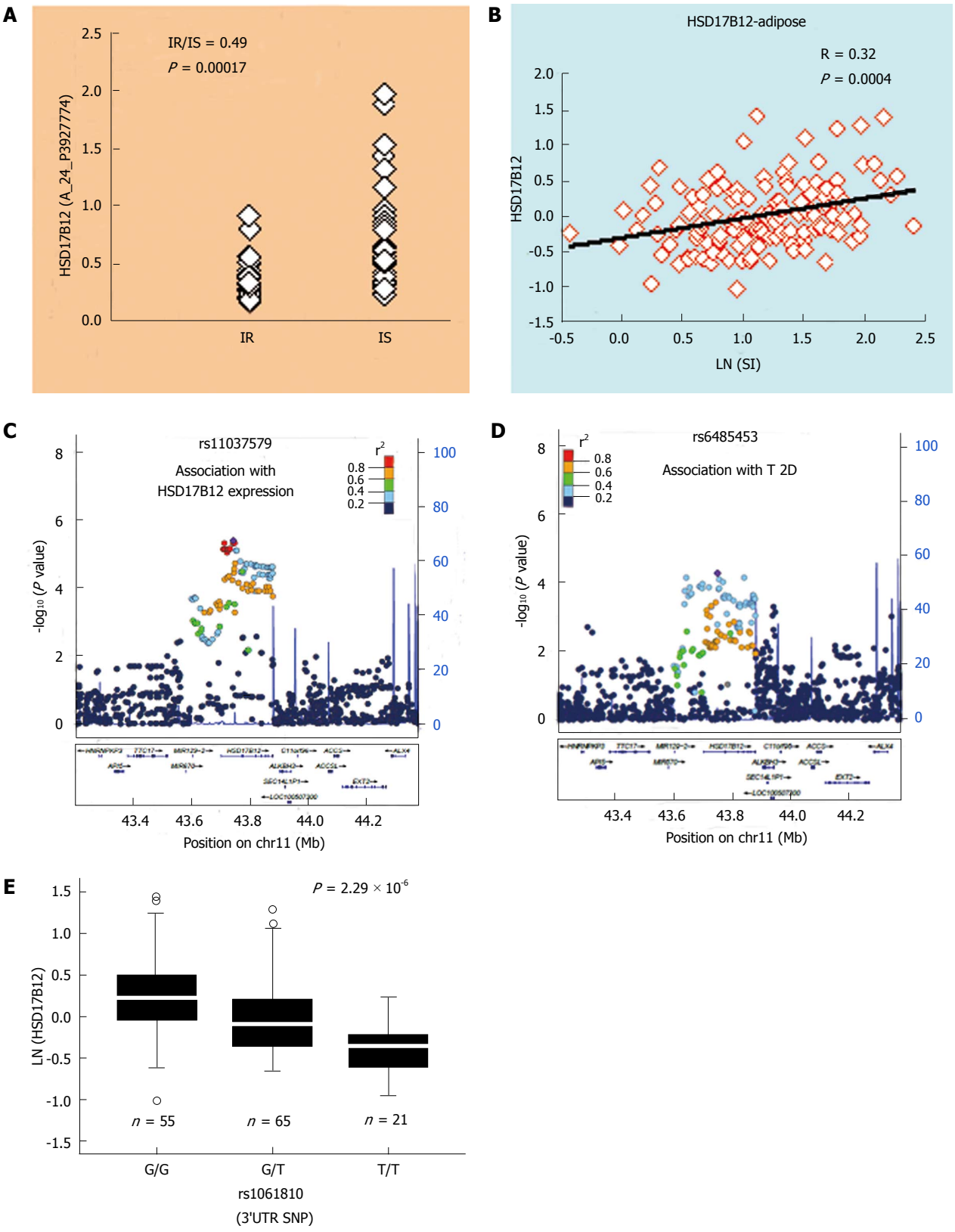

Figure 3 Prioritizing type 2 diabetes-associated variants by expression quantitative trait loci analysis: An example. *HSD17B12* is one of 172 genes differentially expressed in adipose tissue of insulin-resistant (IR, $n = 31$) *vs* insulin-sensitive (IS, $n = 31$) subjects in a genome-wide study (A) by Elbein *et al*[76]. Its expression in subcutaneous adipose of non-diabetic subjects ($n = 141$) also shows a significant correlation (B) with insulin sensitivity (SI). Strongest *cis*-eSNP for adipose tissue (C) expression of HSD17B12 (in adipose eQTL from the MuTHER project)[55] is also associated with T2D (D) in a large GWAS meta-analysis (in DIAGRAM.v3 data from 12171 T2D and 56862 controls)[13]. This locus also includes a 3'UTR SNP rs1061810 that shows association (E) with T2D and expression of HSD17B12 (in qRT-PCR analysis in adipose tissue from 141 non-diabetic subjects). eSNP: Expression regulatory SNP; eQTL: Expression quantitative trait loci; T2D: Type 2 diabetes.

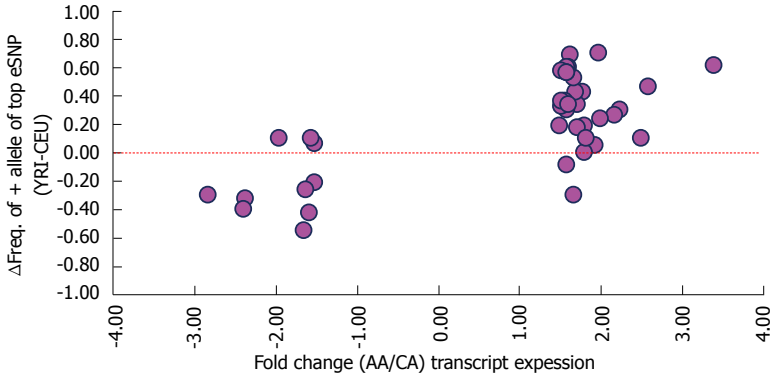

Figure 4 Population differences in expression of transcripts in adipose tissue is accounted for by the effect allele frequency difference of expression regulatory SNPs among racial groups. X axis: Fold change in average expression of 41 transcripts between African-American (AA, *n* = 37) and Caucasian (CA, *n* = 99) subjects. Y axis: Differences in strongest eSNP allele frequency of these transcripts between HapMap subjects of Caucasian (CEU) and African (YRI) ancestry for alleles associated with higher expression. *e*SNP: Expression regulatory SNP.

European (CEU) and African (YRI) descent, with enrichment of ribosome biogenesis, antimicrobial response and cell-cell adhesion. Spielman *et al*[37] (2007) attributed the 1097 genes that differed between CEU and Asian (CHB) LCL samples to eSNP frequency. Our comparison of genome-wide expression profiles (using an Agilent 44K expression array) from adipose and muscle tissue of non-diabetic Caucasians (n = 40) and African Americans (n = 22) identified transcripts associated with insulin sensitivity (S_I), many of which (*e.g.*, *CLIC6*, *HSD11B1*, *SERPINA3*, *THBS1*, *TMEM135* and *TNMD* in Adipose) show distinct ethnic-specific expression[76].

Comparison of adipose tissue expression data between Caucasians and African Americans in a larger cohort (using an Illumina –HT12.V4 array for 99 Caucasians and 37 African Americans) identified 117 differentially expressed (fold change $\geqslant$ 1.5 and false discovery rate $\leqslant$ 5%) transcripts[91]. By mining adipose tissue eQTL data from the MuTHER project[55], we found that about 35% of these differentially expressed transcripts are strongly modulated ($P < 1 \times 10^{-5}$) by *cis*-eSNPs in adipose tissue. In line with the findings by Spielman *et al*[37] (2007) in LCL, we also found that in adipose tissue, the degree of differential expression (fold change African Americans/Caucasians) shows strong concordance with the difference in the effect allele frequency of top *cis*-eSNPs (Figure 4) between HapMap African (YRI) and Caucasian (CEU) subjects.

These studies suggest that the distinct genetic architecture of eSNPs determines the ethnic-specific expression profile in tissues important for glucose homeostasis. Ethnic-specific derangements of gene expression networks in tissues involved in glucose homeostasis may explain distinctive physiologic effects, including differences in insulin action and secretion between ethnic and racial groups. Perturbation of gene expression networks associated with early pathophysiologic events (including insulin resistance) is driven by regulatory variants (eSNPs). The distinct genetic architecture of these variants (including linkage disequilibrium and allele frequency) may determine their ethnic-specific (or predominant) effect on expression and T2D susceptibility. Thus, integration of genome-wide expression analysis and eQTL analyses may be a useful approach to identify the primary genetic factors for ethnic-specific susceptibility to T2D.

Expression of transcripts involved in the same biological function tend to be co-regulated by similar factors (genetic or environmental) and can be identified as distinct network modules, where genes within a module are more highly interconnected (correlated) with each other than genes in other modules. Statistical approaches like weighted gene co-expression network analysis (WGCNA software package developed in "R" programming environment implements this analytical method) are useful for identifying modular structures of the co-expression networks[92,93] in tissues important for glucose homeostasis. Evaluation of the correlation of each module eigengene with the S_I and other T2D-related metabolic phenotypes, and determination of the preservation of these modules between ethnic groups based on observed network density and connectivity, will identify molecular processes or molecular interaction structures associated with phenotypes that undergo ethnic-specific reconfiguration by genetic or non-genetic causal regulators.

Several recently developed statistical metrics[94,95], including modular differential connectivity, offer powerful tools to identify the modules with significant ethnic-specific changes in interaction strength. The eSNPs are causal variants (or in linkage disequilibrium with causal variants) that regulate the expression level of neighboring (or distal) genes. Thus, eSNPs serve as a primary source of natural perturbation to infer causal relationships among and between genes in gene-expression networks[96]. The distinct allelic architecture of these SNPs may determine ethnic-specific modular differential connectivity. Genes with eSNPs can be considered as "parent nodes" in expression networks. This information is used as a "structure prior" in the network reconstruction analysis to orient the edges of the networks. Reconstructing ethnic-specific networks by utilizing different causality modeling methods, including Bayesian network reconstruction approaches, may identify key causal regulators of these networks[97,98]. Thus, a multiscale biological network analysis that utilizes eQTL information to distinguish causal from correlated disease effects is a novel approach to understand how causal regulators propagate their effects in mediating ethnic-specific susceptibility to disease.

A similar approach was used recently to identify genetic factors in animal models of diabetes and other complex human diseases, including Alzheimer's disease[95].

A study by Zhong *et al*[68] (2010) in adipose tissue of C57BL/6-ob/ob × BTBR-ob/ob mice F2 progeny identified a strong causal subnetwork for T2D traits (called the "purple" module, enriched for genes involved in plasma glucose and insulin levels). They found that 37 eSNPs of genes in this module showed significant association with T2D in a GWAS report. Through additional prioritization steps and subsequent function validation studies, they identified mallic enzyme (*ME1*) as a key causal gene in this T2D subnetwork. A strong *cis*-eSNP of *ME1* was associated with T2D. Future applications of such integrative genomic strategies in T2D or related disorders in human populations may prove insightful.

EQTL ANALYSES TO IDENTIFY GENE-ENVIRONMENT INTERACTIONS RELEVANT FOR T2D

As discussed above, GWAS have identified DNA sequence variants in the susceptibility to T2D, but these variants account for only a part of the estimated heritability[13,14]. Interactions between sequence variants and environmental stimuli are a logical step in better understanding the development of T2D. Thus, some of the missing heritability for T2D susceptibility may be explained by studies of the interaction between environmental factors and genetic variants or gene-environment (GXE) interactions[99]. Modeling GXE interactions in clinical or epidemiological settings is challenging and costly, due to few validated tools for assessing exposure (including dietary exposure), the need for large sample sizes, and the heterogeneity of exposures in populations[100-103]. Environmental factors usually influence insulin resistance and T2D risk over long periods of time; thus, accurate assessment of long-term exposure is needed to identify GXE interactions. A recent series of studies by Patel *et al*[104-106] utilized data resources from the National Health and Nutritional Examination Survey and integrated GWAS and environment-wide association studies to identify environmental factors, genetic factors, and GXE interactions involved in T2D susceptibility. However, they noted several significant limitations of such epidemiological approaches in adequately addressing influence of genetic variations on differences in environmental response in human populations.

Environmental factors, including diet and derived metabolites, can influence phenotypes by modulating gene expression in several ways. Variations in responses to environmental factors among individuals, and how these responses predispose to metabolic and other disorders, have been recognized[107]. Genetic variants modulate the environmental factor-mediated transcriptional response, which in turn dictates cellular response and may explain variability in metabolic responses to those factors[99]. Such dependency on external conditions or GXE interactions has been reported for genetic effects on gene expression in different organisms[108-110]. Transcripts responsive to environmental perturbation factors may manifest as eQTLs and are modulated by *cis*- and *trans*-eSNPs. A subset of these eSNPs associated with T2D, obesity, and/or glucose homeostasis traits may thus exhibit distinctive patterns of GXE eSNPs. Thus, identifying environmental factors that modulate insulin sensitivity and other early pathophysiological manifestations of T2D and its integration into eQTL analyses will further improve the power to construct causal gene regulatory networks involved in T2D susceptibility.

A few recent studies implemented a novel "cellular genomics" approach[111] to elucidate genetic controls on GXE interactions, critical to understanding the pathophysiology of complex diseases. In this novel paradigm, researchers analyzed the molecular consequences of genetic variants to assess interactions with environmental factors *via* quantification of processes (like gene expression) in cells from human subjects grown in uniform culture conditions. This concept is illustrated in Figure 5. Utilizing transformed lymphocytes, the studies examined genetic control in response to radiation, chemotherapeutic drugs, and hormones (glucocorticoids)[112-114]. Two similar studies in primary human cells mapped genetic regulators responding to growth factors (BMP-2), hormones (dexamethasone), cytokines (prostaglandin E2 in human osteoblasts), and oxidized low density lipoprotein (in human aortic endothelial cells)[115,116]. Despite the encouraging success of these studies, no studies so far have evaluated GXE interactions with a cellular genomic model relevant to T2D and related metabolic disorders. Although this model may miss some whole organism-level complexity[117] of T2D pathogenesis (which involves multiple tissues), it does represent an innovative approach by going from cellular to organismal phenotype analysis for identification of function of genetic variants involved in T2D susceptibility. Mapping GXE eSNPs for function-based prioritization of T2D and related metabolic disease-associated SNPs is a critical step towards designing efficacious strategies to reduce the public health burden of common metabolic disorders triggered by increased exposure to dietary and other environmental factors.

EQTL AND PHARMACOGENOMIC STUDIES FOR T2D

Several classes of anti-diabetic medications are used for the treatment of T2D[118]. Pharmacogenomic studies reviewing the role of genetic variants on drug responses (including adverse drug reactions) have yielded significant findings, including novel disease mechanisms for several complex diseases[119]. But a similar success for T2D has not been achieved[5,120]. Pharmacological interventions using peroxisomal proliferator activated receptor gamma (PPARγ) agonists like pioglitazone improve insulin sensitivity and can reduce the risk of progression to T2D[121]. However, approximately 25% of patients do not respond adequately to PPARγ agonists[122]. Genome-wide transcriptomic analysis by our laboratory showed significant

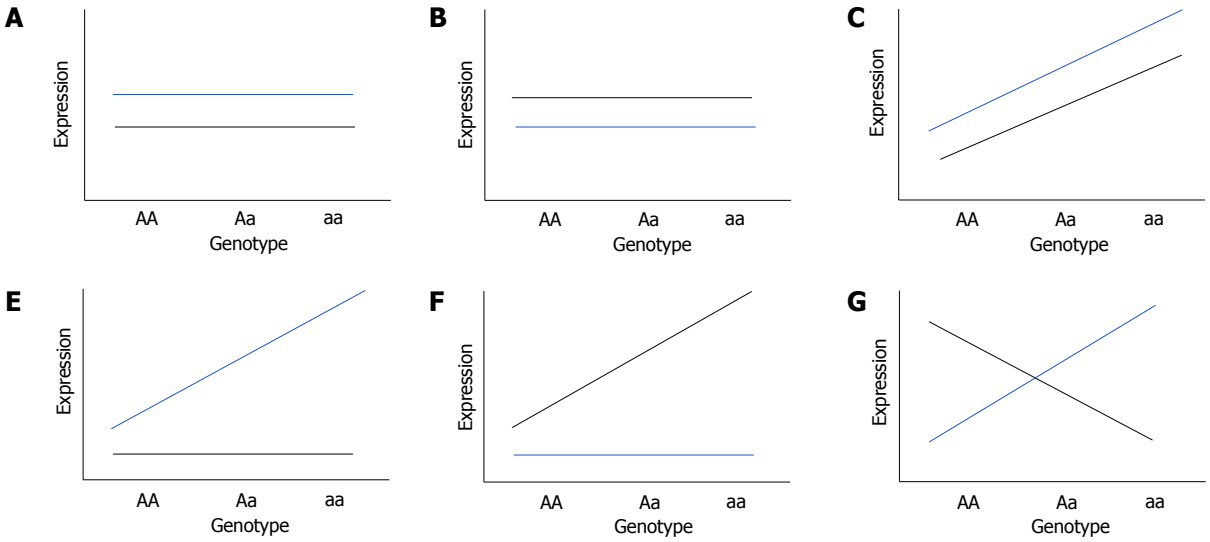

Figure 5 Types of gene-by-environment interactions in cellular genomic models to study gene-by-environment expression quantitative trait locis. Cells from a cohort of subjects are grown in pairs under uniform *in vitro* treated and untreated conditions to study environment-dependent or -independent effects of genotype on expression of transcripts (a quantitative trait). β1 and β2 are genotype effects on transcript expression under treated and control conditions, respectively. Different models of gene-by-environment (GXE) includes Null model: β1 = β2 = 0 (A and B); No-interaction eQTL model: β1 = β2 ≠ 0 (C); Treated-only expression quantitative trait loci model: β1 ≠ 0 and β2 = 0 (D); Control-only eQTL model: β1 = 0 and β2 ≠ 0 (E); and General interaction eQTL model: β1 ≠ 0 and β2 ≠ 0 but β1 ≠ β2 (F). Black line indicates expression in cells under control condition (untreated) while blue line indicates expression in environmental/dietary factor treated cells.

inter-individual variability in gene-expression response after pioglitazone treatment in people with impaired glucose tolerance[123]. However, little is known about the genetic architecture of variation in pioglitazone-mediated transcriptional response in human populations. Identifying the genetic variations that interact with pharmacological treatments like PPARγ agonists is of high clinical interest. eSNPs may modulate the expression of key transcripts in response to anti-diabetic drugs in target tissues and can explain the interindividual variability in treatment outcome[124,125]. Identifying genetic (and epigenetic) variants that modulate the pharmacological treatment-mediated transcriptional response, which in turn dictates the treatment outcome in T2D, is an open area of research. A novel approach that systematically characterizes the set of eSNPs involved in anti-diabetic medicine-mediated transcriptional modulation (gene-drug interaction eSNPs, or GXD eSNPs) in tissues relevant to glucose homeostasis will be useful in stratifying populations in efficacy studies, to improve the quality of clinical decision-making and treatment options for T2D.

FINDING EQTLS: END FOR A NEW BEGINNING

eQTL analyses provide statistical evidence for genotype-dependent variations in transcript abundance and should be considered a starting point for investigating the effects of DNA polymorphisms at the molecular level[34]. Transcript abundance depends on a dynamic relationship between transcript synthesis, stability, and degradation[48]. Thus, DNA polymorphisms may affect transcript abundance by several known and unknown mechanisms. Studies in human subjects have shown that sequence-specific regulation of mRNA expression is mediated by several molecular mechanisms, including allelic variability in transcription factor binding, chromatin remodeling, changes in DNase I hypersensitivity by histone methylation and acetylation, interaction between chromatin segments, alteration of splicing, sequence-dependent allele-specific DNA methylation, alteration of miRNA synthesis, and miRNA target binding[50-52,126-130]. GWAS-implicated variants for complex diseases are enriched in non-coding functional domains of the genome, including sequences involved in chromatin remodeling[131-133]. Many transcripts that show strong co-expression and *cis*-eSNPs for one transcript may appear as trans-eSNPs for a co-regulated transcript located in other chromosomes. Thus, a functional role of prioritized *cis*- and *trans*-eSNPs needs to be validated by appropriate molecular experiments to distinguish causal from correlative effects[134-136]. Studies have used allelic expression imbalance analysis, electrophoretic mobility shift assays, and transient transfection based luciferase reporter assays[56,137-141] to identify the molecular effects of genetic variants (*cis*-eSNPs) on gene expression; however, high-throughput methods are needed to validate in parallel the large number of findings from genomic studies[134,135,142]. Several novel high-throughput methods, including massively parallel reporter assays and massively parallel functional dissection, are now available to show evidence of causality for regulatory variants[143-146]. Functional relevance of the candidate eQTL transcripts in T2D pathophysiology also need to be validated by demonstrating their effects upon experimental up- or down-regulation in *in vitro* or *in vivo* experimental models[147,148].

In summary, many factors (including genetic, epigenetic and environmental factors) affect susceptibility to

T2D. Instead of investigating different sources of variation in isolation, an integrative functional omics paradigm that traces early molecular changes through layers of biological information, including eQTLs, promises to be a useful approach[136]. Such an approach will promote optimal understanding of the etiology of T2D and lead to the identification of ethnic-specific primary causal variants. Ultimately, the knowledge gained from studies using these approaches can be used to build better classifiers of T2D risk than those based on DNA sequence variants alone.

ACKNOWLEDGMENTS

We thank Karen Klein (Translational Science Institute, Wake Forest University Health Sciences) and Ethel Kouba (Internal Medicine, Endocrinology) for critical reading and editing of our manuscript. We dedicate this article to the memory of Dr. Steven C. Elbein.

REFERENCES

1 **Boyle JP**, Thompson TJ, Gregg EW, Barker LE, Williamson DF. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Popul Health Metr* 2010; **8**: 29 [PMID: 20969750 DOI: 10.1186/1478-7954-8-29]

2 **International Diabetes Federation IDF**. IDF Diabetes Atlas, 6th Edition. IDF Diabetes Atlas 2013. Available from: URL: http://www.idf.org/sites/default/files/EN_6E_Atlas_Full_0.pdf

3 **Groop L**, Pociot F. Genetics of diabetes--are we missing the genes or the disease? *Mol Cell Endocrinol* 2014; **382**: 726-739 [PMID: 23587769 DOI: 10.1016/j.mce.2013.04.002]

4 **Doria A**, Patti ME, Kahn CR. The emerging genetic architecture of type 2 diabetes. *Cell Metab* 2008; **8**: 186-200 [PMID: 18762020 DOI: 10.1016/j.cmet.2008.08.006]

5 **Johnson AM**, Olefsky JM. The origins and drivers of insulin resistance. *Cell* 2013; **152**: 673-684 [PMID: 23415219 DOI: 10.1016/j.cell.2013.01.041]

6 **Das SK**, Elbein SC. The search for type 2 diabetes susceptibility loci: the chromosome 1q story. *Curr Diab Rep* 2007; **7**: 154-164 [PMID: 17425920]

7 **Das SK**, Elbein SC. The Genetic Basis of Type 2 Diabetes. *Cellscience* 2006; **2**: 100-131 [PMID: 16892160]

8 **Grant SF**, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, Styrkarsdottir U, Magnusson KP, Walters GB, Palsdottir E, Jonsdottir T, Gudmundsdottir T, Gylfason A, Saemundsdottir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdottir U, Gulcher JR, Kong A, Stefansson K. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006; **38**: 320-323 [PMID: 16415884 DOI: 10.1038/ng1732]

9 **Grant SF**. Understanding the elusive mechanism of action of TCF7L2 in metabolism. *Diabetes* 2012; **61**: 2657-2658 [PMID: 23093653 DOI: 10.2337/db12-0891]

10 **Boj SF**, van Es JH, Huch M, Li VS, José A, Hatzis P, Mokry M, Haegebarth A, van den Born M, Chambon P, Voshol P, Dor Y, Cuppen E, Fillat C, Clevers H. Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell* 2012; **151**: 1595-1607 [PMID: 23260145 DOI: 10.1016/j.cell.2012.10.053]

11 **McCarthy MI**, Rorsman P, Gloyn AL. TCF7L2 and diabetes: a tale of two tissues, and of two species. *Cell Metab* 2013; **17**: 157-159 [PMID: 23395164 DOI: 10.1016/j.cmet.2013.01.011]

12 **Visscher PM**, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012; **90**: 7-24 [PMID: 22243964 DOI: 10.1016/j.ajhg.2011.11.029]

13 **Morris AP,** Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, Kumar A, Lagou V, Langenberg C, Luan J, Lindgren CM, Muller-Nurasyid M, Pechlivanis S, Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD, Dimas AS, Loos RJ, Vedantam S, Chen H, Florez JC, Fox C, Liu CT, Rybin D, Couper DJ, Kao WH, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JR, Platou CG, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stancakova A, Stirrups K, Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, Burtt N, Carey J, Charpentier G, Crenshaw AT, Doney AS, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutskov K, Langford C, Leander K, Lindholm E, Lobbens S, Mannisto S, Mirza G, Muhleisen TW, Musk B, Parkin M, Rallidis L, Saramies J, Sennblad B, Shah S, Sigurethsson G, Silveira A, Steinbach G, Thorand B, Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvanen AC, Eriksson JG, Peltonen L, Nothen MM, Balkau B, Palmer CN, Lyssenko V, Tuomi T, Isomaa B, Hunter DJ, Qi L, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njolstad I, Pedersen NL, Khaw KT, Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyovalti E, Saltevo J, Laakso M, Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, Hveem K, Cauchi S, Froguel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jockel KH, Moebus S, Peters A, Illig T, de Faire U, Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012; **44**: 981-990 [PMID: 22885922 DOI: 10.1038/ng.2383]

14 **Saxena R,** Elbers CC, Guo Y, Peter I, Gaunt TR, Mega JL, Lanktree MB, Tare A, Castillo BA, Li YR, Johnson T, Bruinenberg M, Gilbert-Diamond D, Rajagopalan R, Voight BF, Balasubramanyam A, Barnard J, Bauer F, Baumert J, Bhangale T, Bohm BO, Braund PS, Burton PR, Chandrupatla HR, Clarke R, Cooper-DeHoff RM, Crook ED, Davey-Smith G, Day IN, de Boer A, de Groot MC, Drenos F, Ferguson J, Fox CS, Furlong CE, Gibson Q, Gieger C, Gilhuijs-Pederson LA, Glessner JT, Goel A, Gong Y, Grant SF, Grobbee DE, Hastie C, Humphries SE, Kim CE, Kivimaki M, Kleber M, Meisinger C, Kumari M, Langaee TY, Lawlor DA, Li M, Lobmeyer MT, Maitland-van der Zee AH, Meijs MF, Molony CM, Morrow DA, Murugesan G, Musani SK, Nelson CP, Newhouse SJ, O'Connell JR, Padmanabhan S, Palmen J, Patel SR, Pepine CJ, Pettinger M, Price TS, Rafelt S, Ranchalis J, Rasheed A, Rosenthal E, Ruczinski I, Shah S, Shen H, Silbernagel G, Smith EN, Spijkerman AW, Stanton A, Steffes MW, Thorand B, Trip M, van der HP, van der AD, van Iperen EP, van Setten J, Vliet-Ostaptchouk JV, Verweij N, Wolffenbuttel BH, Young T, Zafarmand MH, Zmuda JM, Boehnke M, Altshuler D, McCarthy M, Kao WH, Pankow JS, Cappola TP, Sever P, Poulter N, Caulfield M, Dominiczak A, Shields DC, Bhatt

DL, Zhang L, Curtis SP, Danesh J, Casas JP, van der Schouw YT, Onland-Moret NC, Doevendans PA, Dorn GW, Farrall M, FitzGerald GA, Hamsten A, Hegele R, Hingorani AD, Hofker MH, Huggins GS, Illig T, Jarvik GP, Johnson JA, Klungel OH, Knowler WC, Koenig W, Marz W, Meigs JB, Melander O, Munroe PB, Mitchell BD, Bielinski SJ, Rader DJ, Reilly MP, Rich SS, Rotter JI, Saleheen D, Samani NJ, Schadt EE, Shuldiner AR, Silverstein R, Kottke-Marchant K, Talmud PJ, Watkins H, Asselbergs FW, de Bakker PI, McCaffery J, Wijmenga C, Sabatine MS, Wilson JG, Reiner A, Bowden DW, Hakonarson H, Siscovick DS, Keating BJ. Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. *Am J Hum Genet* 2012; **90**:410-425 [PMID: 22325160 DOI: 10.1016/j.ajhg.2011.12.022]

15 **Torres JM**, Cox NJ, Philipson LH. Genome wide association studies for diabetes: perspective on results and challenges. *Pediatr Diabetes* 2013; **14**: 90-96 [PMID: 23350725 DOI: 10.1111/pedi.12015]

16 **Ng MC**, Saxena R, Li J, Palmer ND, Dimitrov L, Xu J, Rasmussen-Torvik LJ, Zmuda JM, Siscovick DS, Patel SR, Crook ED, Sims M, Chen YD, Bertoni AG, Li M, Grant SF, Dupuis J, Meigs JB, Psaty BM, Pankow JS, Langefeld CD, Freedman BI, Rotter JI, Wilson JG, Bowden DW. Transferability and fine mapping of type 2 diabetes loci in African Americans: the Candidate Gene Association Resource Plus Study. *Diabetes* 2013; **62**: 965-976 [PMID: 23193183 DOI: 10.2337/db12-0266]

17 **Watanabe RM**. The genetics of insulin resistance: Where's Waldo? *Curr Diab Rep* 2010; **10**: 476-484 [PMID: 20820957 DOI: 10.1007/s11892-010-0143-1]

18 **Perry JR**, McCarthy MI, Hattersley AT, Zeggini E, Weedon MN, Frayling TM. Interrogating type 2 diabetes genome-wide association data using a biological pathway-based approach. *Diabetes* 2009; **58**: 1463-1467 [PMID: 19252133 DOI: 10.2337/db08-1378]

19 **Florez JC**. Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? *Diabetologia* 2008; **51**: 1100-1110 [PMID: 18504548 DOI: 10.1007/s00125-008-1025-9]

20 **Manolio TA**, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747-753 [PMID: 19812666 DOI: 10.1038/nature08494]

21 **Hindorff LA**, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009; **106**: 9362-9367 [PMID: 19474294 DOI: 10.1073/pnas.0903103106]

22 **Rockman MV**, Kruglyak L. Genetics of global gene expression. *Nat Rev Genet* 2006; **7**: 862-872 [PMID: 17047685 DOI: 10.1038/nrg1964]

23 **Gibson G**, Weir B. The quantitative genetics of transcription. *Trends Genet* 2005; **21**: 616-623 [PMID: 16154229 DOI: 10.1016/j.tig.2005.08.010]

24 **Emilsson V**, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, Mouy M, Steinthorsdottir V, Eiriksdottir GH, Bjornsdottir G, Reynisdottir I, Gudbjartsson D, Helgadottir A, Jonasdottir A, Jonasdottir A, Styrkarsdottir U, Gretarsdottir S, Magnusson KP, Stefansson H, Fossdal R, Kristjansson K, Gislason HG, Stefansson T, Leifsson BG, Thorsteinsdottir U, Lamb JR, Gulcher JR, Reitman ML, Kong A, Schadt EE, Stefansson K. Genetics of gene expression and its effect on disease. *Nature* 2008; **452**: 423-428 [PMID: 18344981 DOI: 10.1038/nature06758]

25 **Dixon AL**, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrall M, Lathrop GM, Abecasis GR, Cookson WO. A genome-wide association study of global gene expression. *Nat Genet* 2007; **39**: 1202-1207 [PMID: 17873877 DOI: 10.1038/ng2109]

26 **Göring HH**, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JB, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet* 2007; **39**: 1208-1216 [PMID: 17873875 DOI: 10.1038/ng2119]

27 **Chen Y**, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardson A, Castellini LW, Wang S, Champy MF, Zhang B, Emilsson V, Doss S, Ghazalpour A, Horvath S, Drake TA, Lusis AJ, Schadt EE. Variations in DNA elucidate molecular networks that cause disease. *Nature* 2008; **452**: 429-435 [PMID: 18344982 DOI: 10.1038/nature06757]

28 **Stranger BE**, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavaré S, Deloukas P, Dermitzakis ET. Population genomics of human gene expression. *Nat Genet* 2007; **39**: 1217-1224 [PMID: 17873874 DOI: 10.1038/ng2142]

29 **Cheung VG**, Spielman RS. Genetics of human gene expression: mapping DNA variants that influence gene expression. *Nat Rev Genet* 2009; **10**: 595-604 [PMID: 19636342 DOI: 10.1038/nrg2630]

30 **Dobrin R**, Greenawalt DM, Hu G, Kemp DM, Kaplan LM, Schadt EE, Emilsson V. Dissecting cis regulation of gene expression in human metabolic tissues. *PLoS One* 2011; **6**: e23480 [PMID: 21912597 DOI: 10.1371/journal.pone.0023480]

31 **Montgomery SB**, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 2010; **464**: 773-777 [PMID: 20220756 DOI: 10.1038/nature08903]

32 **Battle A**, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, Haudenschild CD, Beckman KB, Shi J, Mei R, Urban AE, Montgomery SB, Levinson DF, Koller D. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res* 2014; **24**: 14-24 [PMID: 24092820 DOI: 10.1101/gr.155192]

33 **Lappalainen T**, Sammeth M, Friedländer MR, 't Hoen PA, Monlong J, Rivas MA, Gonzàlez-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, Greger L, van Iterson M, Almlöf J, Ribeca P, Pulyakhina I, Esser D, Giger T, Tikhonov A, Sultan M, Bertier G, MacArthur DG, Lek M, Lizano E, Buermans HP, Padioleau I, Schwarzmayr T, Karlberg O, Ongen H, Kilpinen H, Beltran S, Gut M, Kahlem K, Amstislavskiy V, Stegle O, Pirinen M, Montgomery SB, Donnelly P, McCarthy MI, Flicek P, Strom TM, Lehrach H, Schreiber S, Sudbrak R, Carracedo A, Antonarakis SE, Häsler R, Syvänen AC, van Ommen GJ, Brazma A, Meitinger T, Rosenstiel P, Guigó R, Gut IG, Estivill X, Dermitzakis ET. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013; **501**: 506-511 [PMID: 24037378 DOI: 10.1038/nature12531]

34 **Mackay TF**, Stone EA, Ayroles JF. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 2009; **10**: 565-577 [PMID: 19584810 DOI: 10.1038/nrg2612]

35 **Almasy L**, Blangero J. Human QTL linkage mapping. *Genetica* 2009; **136**: 333-340 [PMID: 18668207 DOI: 10.1007/s10709-008-9305-3]

36 **Cookson W**, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. *Nat Rev Genet* 2009; **10**: 184-194 [PMID: 19223927 DOI: 10.1038/nrg2537]

37 **Spielman RS**, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* 2007; **39**: 226-231 [PMID: 17206142 DOI: 10.1038/ng1955]

38 **Montgomery SB**, Lappalainen T, Gutierrez-Arcelus M, Dermitzakis ET. Rare and common regulatory variation in population-scale sequenced human genomes. *PLoS Genet* 2011; **7**: e1002144 [PMID: 21811411 DOI: 10.1371/journal.pgen.1002144]

39 **Stranger BE**, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavaré S, Deloukas P, Dermitzakis ET. Genome-wide associations of gene expression variation in humans. *PLoS Genet* 2005; **1**: e78 [PMID: 16362079 DOI: 10.1371/journal.pgen.0010078]

40 **Veyrieras JB**, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, Pritchard JK. High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet* 2008; **4**: e1000214 [PMID: 18846210 DOI: 10.1371/journal.pgen.1000214]

41 **Montgomery SB**, Dermitzakis ET. From expression QTLs to personalized transcriptomics. *Nat Rev Genet* 2011; **12**: 277-282 [PMID: 21386863 DOI: 10.1038/nrg2969]

42 **Stegle O**, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* 2012; **7**: 500-507 [PMID: 22343431 DOI: 10.1038/nprot.2011.457]

43 **Listgarten J**, Kadie C, Schadt EE, Heckerman D. Correction for hidden confounders in the genetic analysis of gene expression. *Proc Natl Acad Sci USA* 2010; **107**: 16465-16470 [PMID: 20810919 DOI: 10.1073/pnas.1002425107]

44 **Kang HM**, Ye C, Eskin E. Accurate discovery of expression quantitative trait loci under confounding from spurious and genuine regulatory hotspots. *Genetics* 2008; **180**: 1909-1925 [PMID: 18791227 DOI: 10.1534/genetics.108.094201]

45 **Leek JT**, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 2007; **3**: 1724-1735 [PMID: 17907809 DOI: 10.1371/journal.pgen.0030161]

46 **Shabalin AA**. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012; **28**: 1353-1358 [PMID: 22492648 DOI: 10.1093/bioinformatics/bts163]

47 **Wright FA**, Shabalin AA, Rusyn I. Computational tools for discovery and interpretation of expression quantitative trait loci. *Pharmacogenomics* 2012; **13**: 343-352 [PMID: 22304583 DOI: 10.2217/pgs.11.185]

48 **Lee TI**, Young RA. Transcriptional regulation and its misregulation in disease. *Cell* 2013; **152**: 1237-1251 [PMID: 23498934 DOI: 10.1016/j.cell.2013.02.014]

49 **Bernstein BE**, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57-74 [PMID: 22955616 DOI: 10.1038/nature11247]

50 **Kilpinen H**, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, Migliavacca E, Wiederkehr M, Gutierrez-Arcelus M, Panousis NI, Yurovsky A, Lappalainen T, Romano-Palumbo L, Planchon A, Bielser D, Bryois J, Padioleau I, Udin G, Thurnheer S, Hacker D, Core LJ, Lis JT, Hernandez N, Reymond A, Deplancke B, Dermitzakis ET. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* 2013; **342**: 744-747 [PMID: 24136355 DOI: 10.1126/science.1242463]

51 **Kasowski M**, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM, Habegger L, Rozowsky J, Shi M, Urban AE, Hong MY, Karczewski KJ, Huber W, Weissman SM, Gerstein MB, Korbel JO, Snyder M. Variation in transcription factor binding among humans. *Science* 2010; **328**: 232-235 [PMID: 20299548 DOI: 10.1126/science.1183621]

52 **McVicker G**, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, Lewellen N, Myrthil M, Gilad Y, Pritchard JK. Identification of genetic variants that affect histone modifications in human cells. *Science* 2013; **342**: 747-749 [PMID: 24136359 DOI: 10.1126/science.1242429]

53 **Fu J**, Wolfs MG, Deelen P, Westra HJ, Fehrmann RS, Te Meerman GJ, Buurman WA, Rensen SS, Groen HJ, Weersma RK, van den Berg LH, Veldink J, Ophoff RA, Snieder H, van Heel D, Jansen RC, Hofker MH, Wijmenga C, Franke L. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet* 2012; **8**: e1002431 [PMID: 22275870 DOI: 10.1371/journal.pgen.1002431]

54 **Greenawalt DM**, Dobrin R, Chudin E, Hatoum IJ, Suver C, Beaulaurier J, Zhang B, Castro V, Zhu J, Sieberts SK, Wang S, Molony C, Heymsfield SB, Kemp DM, Reitman ML, Lum PY, Schadt EE, Kaplan LM. A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res* 2011; **21**: 1008-1016 [PMID: 21602305 DOI: 10.1101/gr.112821.110]

55 **Grundberg E**, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, Nisbett J, Sekowska M, Wilk A, Shin SY, Glass D, Travers M, Min JL, Ring S, Ho K, Thorleifsson G, Kong A, Thorsteindottir U, Ainali C, Dimas AS, Hassanali N, Ingle C, Knowles D, Krestyaninova M, Lowe CE, Di Meglio P, Montgomery SB, Parts L, Potter S, Surdulescu G, Tsaprouni L, Tsoka S, Bataille V, Durbin R, Nestle FO, O'Rahilly S, Soranzo N, Lindgren CM, Zondervan KT, Ahmadi KR, Schadt EE, Stefansson K, Smith GD, McCarthy MI, Deloukas P, Dermitzakis ET, Spector TD. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* 2012; **44**: 1084-1089 [PMID: 22941192 DOI: 10.1038/ng.2394]

56 **Innocenti F**, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, Mirkov S, Ramirez J, Liu W, Lin YS, Moloney C, Aldred SF, Trinklein ND, Schuetz E, Nickerson DA, Thummel KE, Rieder MJ, Rettie AE, Ratain MJ, Cox NJ, Brown CD. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet* 2011; **7**: e1002078 [PMID: 21637794 DOI: 10.1371/journal.pgen.1002078]

57 **Nica AC**, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, Travers M, Potter S, Grundberg E, Small K, Hedman AK, Bataille V, Tzenova Bell J, Surdulescu G, Dimas AS, Ingle C, Nestle FO, di Meglio P, Min JL, Wilk A, Hammond CJ, Hassanali N, Yang TP, Montgomery SB, O'Rahilly S, Lindgren CM, Zondervan KT, Soranzo N, Barroso I, Durbin R, Ahmadi K, Deloukas P, McCarthy MI, Dermitzakis ET, Spector TD. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet* 2011; **7**: e1002003 [PMID: 21304890 DOI: 10.1371/journal.pgen.1002003]

58 **Petretto E**, Bottolo L, Langley SR, Heinig M, McDermott-Roe C, Sarwar R, Pravenec M, Hübner N, Aitman TJ, Cook SA, Richardson S. New insights into the genetic control of gene expression using a Bayesian multi-tissue approach. *PLoS Comput Biol* 2010; **6**: e1000737 [PMID: 20386736 DOI: 10.1371/journal.pcbi.1000737]

59 **Schadt EE**, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 2008; **6**: e107 [PMID: 18462017 DOI: 10.1371/journal.pbio.0060107]

60 **Moffatt MF**, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E, Heinzmann A, Simma B, Frischer T, Willis-Owen SA, Wong KC, Illig T, Vogelberg C, Weiland SK, von Mutius E, Abecasis GR, Farrall M, Gut IG, Lathrop GM, Cookson WO. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007; **448**: 470-473 [PMID: 17611496 DOI: 10.1038/nature06014]

61 The Genotype-Tissue Expression (GTEx) project. *Nat Gen-*

et 2013; **45**: 580-585 [PMID: 23715323 DOI: 10.1038/ng.2653]
62 **Gamazon ER**, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, Dolan ME, Cox NJ. SCAN: SNP and copy number annotation. *Bioinformatics* 2010; **26**: 259-262 [PMID: 19933162 DOI: 10.1093/bioinformatics/btp644]
63 **Xia K**, Shabalin AA, Huang S, Madar V, Zhou YH, Wang W, Zou F, Sun W, Sullivan PF, Wright FA. seeQTL: a searchable database for human eQTLs. *Bioinformatics* 2012; **28**: 451-452 [PMID: 22171328 DOI: 10.1093/bioinformatics/btr678]
64 **Yang TP**, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, Deloukas P, Dermitzakis ET. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* 2010; **26**: 2474-2476 [PMID: 20702402 DOI: 10.1093/bioinformatics/btq452]
65 **Nica AC**, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet* 2010; **6**: e1000895 [PMID: 20369022 DOI: 10.1371/journal.pgen.1000895]
66 **He X**, Fuller CK, Song Y, Meng Q, Zhang B, Yang X, Li H. Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am J Hum Genet* 2013; **92**: 667-680 [PMID: 23643380 DOI: 10.1016/j.ajhg.2013.03.022]
67 **Nicolae DL**, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 2010; **6**: e1000888 [PMID: 20369019 DOI: 10.1371/journal.pgen.1000888]
68 **Zhong H**, Beaulaurier J, Lum PY, Molony C, Yang X, Macneil DJ, Weingarth DT, Zhang B, Greenawalt D, Dobrin R, Hao K, Woo S, Fabre-Suver C, Qian S, Tota MR, Keller MP, Kendziorski CM, Yandell BS, Castro V, Attie AD, Kaplan LM, Schadt EE. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet* 2010; **6**: e1000932 [PMID: 20463879 DOI: 10.1371/journal.pgen.1000932]
69 **Voight BF**, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson BK, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, van Dam RM, van Haeften TW, van Herpt T, Vliet-Ostaptchouk JV, Walters GB, Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllensten U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Mohlke KL, Morris AD, Palmer CN, Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Wareham NJ, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB, Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC, Frayling TM, Groop L, Sladek R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefansson K, Altshuler D, Boehnke M, McCarthy MI. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010; **42**: 579-589 [PMID: 20581827 DOI: 10.1038/ng.609]
70 **Sharma NK**, Langberg KA, Mondal AK, Elbein SC, Das SK. Type 2 diabetes (T2D) associated polymorphisms regulate expression of adjacent transcripts in transformed lymphocytes, adipose, and muscle from Caucasian and African-American subjects. *J Clin Endocrinol Metab* 2011; **96**: E394-E403 [PMID: 21084393 DOI: 10.1210/jc.2010-1754]
71 **Taneera J**, Lang S, Sharma A, Fadista J, Zhou Y, Ahlqvist E, Jonsson A, Lyssenko V, Vikman P, Hansson O, Parikh H, Korsgren O, Soni A, Krus U, Zhang E, Jing XJ, Esguerra JL, Wollheim CB, Salehi A, Rosengren A, Renström E, Groop L. A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. *Cell Metab* 2012; **16**: 122-134 [PMID: 22768844 DOI: 10.1016/j.cmet.2012.06.006]
72 **Elbein SC**, Gamazon ER, Das SK, Rasouli N, Kern PA, Cox NJ. Genetic risk factors for type 2 diabetes: a trans-regulatory genetic architecture? *Am J Hum Genet* 2012; **91**: 466-477 [PMID: 22958899 DOI: 10.1016/j.ajhg.2012.08.002]
73 **Westra HJ**, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, Zhernakova A, Zhernakova DV, Veldink JH, Van den Berg LH, Karjalainen J, Withoff S, Uitterlinden AG, Hofman A, Rivadeneira F, 't Hoen PA, Reinmaa E, Fischer K, Nelis M, Milani L, Melzer D, Ferrucci L, Singleton AB, Hernandez DG, Nalls MA, Homuth G, Nauck M, Radke D, Völker U, Perola M, Salomaa V, Brody J, Suchy-Dicey A, Gharib SA, Enquobahrie DA, Lumley T, Montgomery GW, Makino S, Prokisch H, Herder C, Roden M, Grallert H, Meitinger T, Strauch K, Li Y, Jansen RC, Visscher PM, Knight JC, Psaty BM, Ripatti S, Teumer A, Frayling TM, Metspalu A, van Meurs JB, Franke L. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45**: 1238-1243 [PMID: 24013639 DOI: 10.1038/ng.2756]
74 **Naukkarinen J**, Surakka I, Pietiläinen KH, Rissanen A, Salomaa V, Ripatti S, Yki-Järvinen H, van Duijn CM, Wichmann HE, Kaprio J, Taskinen MR, Peltonen L. Use of genome-wide expression data to mine the "Gray Zone" of GWA studies leads to novel candidate obesity genes. *PLoS Genet* 2010; **6**: e1000976 [PMID: 20532202 DOI: 10.1371/journal.pgen.1000976]
75 **Parikh H**, Lyssenko V, Groop LC. Prioritizing genes for follow-up from genome wide association studies using information on gene expression in tissues relevant for type 2 diabetes mellitus. *BMC Med Genomics* 2009; **2**: 72 [PMID: 20043853 DOI: 10.1186/1755-8794-2-72]
76 **Elbein SC**, Kern PA, Rasouli N, Yao-Borengasser A, Sharma NK, Das SK. Global gene expression profiles of subcutaneous adipose and muscle from glucose-tolerant, insulin-sensitive, and insulin-resistant individuals matched for BMI. *Diabetes* 2011; **60**: 1019-1029 [PMID: 21266331 DOI: 10.2337/db10-1270]
77 **Sharma NK**, Langberg KA, Das SK. Association of Functional SNPs in the HSD17B12 Gene with T2D and Glucose Homeostasis: Results from an Integrative Genomic Analysis(Abstract: 1523P, American Diabetes association 72nd Annual Scientific Sessions, Philadelphia, 2012). *Diabetes* 2012; **61**: A396 [DOI: 10.2337/db12-1329-1552]
78 **Kulzer JR**, Stitzel ML, Morken MA, Huyghe JR, Fuchsberger C, Kuusisto J, Laakso M, Boehnke M, Collins FS, Mohlke KL. A Common Functional Regulatory Variant at a Type 2 Diabetes Locus Upregulates ARAP1 Expression in the Pancreatic Beta Cell. *Am J Hum Genet* 2014; **94**: 186-197 [PMID: 24439111 DOI: 10.1016/j.ajhg.2013.12.011]
79 **Keildson S**, Fadista J, Ladenvall C, Hedman AK, Elgzyri T, Small KS, Grundberg E, Nica AC, Glass D, Richards JB, Barrett A, Nisbet J, Zheng HF, Rönn T, Ström K, Eriksson KF, Prokopenko I, Spector TD, Dermitzakis ET, Deloukas P, McCarthy MI, Rung J, Groop L, Franks PW, Lindgren CM,

Hansson O. Expression of phosphofructokinase in skeletal muscle is influenced by genetic variation and associated with insulin sensitivity. *Diabetes* 2014; **63**: 1154-1165 [PMID: 24306210 DOI: 10.2337/db13-1301]
80 **Diamond J**. The double puzzle of diabetes. *Nature* 2003; **423**: 599-602 [PMID: 12789325 DOI: 10.1038/423599a]
81 **National Diabetes Information Clearinghouse**. National Diabetes Statistics, 2011. Available from: URL: http://diabetes.niddk.nih.gov/dm/pubs/statistics/ . 12-6-2011.
82 **Haiman CA**, Fesinmeyer MD, Spencer KL, Buzková P, Voruganti VS, Wan P, Haessler J, Franceschini N, Monroe KR, Howard BV, Jackson RD, Florez JC, Kolonel LN, Buyske S, Goodloe RJ, Liu S, Manson JE, Meigs JB, Waters K, Mukamal KJ, Pendergrass SA, Shrader P, Wilkens LR, Hindorff LA, Ambite JL, North KE, Peters U, Crawford DC, Le Marchand L, Pankow JS. Consistent directions of effect for established type 2 diabetes risk variants across populations: the population architecture using Genomics and Epidemiology (PAGE) Consortium. *Diabetes* 2012; **61**: 1642-1647 [PMID: 22474029 DOI: 10.2337/db11-1296]
83 **Lewis JP**, Palmer ND, Hicks PJ, Sale MM, Langefeld CD, Freedman BI, Divers J, Bowden DW. Association analysis in african americans of European-derived type 2 diabetes single nucleotide polymorphisms from whole-genome association studies. *Diabetes* 2008; **57**: 2220-2225 [PMID: 18443202 DOI: 10.2337/db07-1319]
84 **Waters KM**, Stram DO, Hassanein MT, Le Marchand L, Wilkens LR, Maskarinec G, Monroe KR, Kolonel LN, Altshuler D, Henderson BE, Haiman CA. Consistent association of type 2 diabetes risk variants found in europeans in diverse racial and ethnic groups. *PLoS Genet* 2010; **6**: [PMID: 20865176 DOI: 10.1371/journal.pgen.1001078]
85 **Haffner SM**, D'Agostino R, Saad MF, Rewers M, Mykkänen L, Selby J, Howard G, Savage PJ, Hamman RF, Wagenknecht LE. Increased insulin resistance and insulin secretion in nondiabetic African-Americans and Hispanics compared with non-Hispanic whites. The Insulin Resistance Atherosclerosis Study. *Diabetes* 1996; **45**: 742-748 [PMID: 8635647]
86 **Rasouli N**, Spencer HJ, Rashidi AA, Elbein SC. Impact of family history of diabetes and ethnicity on -cell function in obese, glucose-tolerant individuals. *J Clin Endocrinol Metab* 2007; **92**: 4656-4663 [PMID: 17878257 DOI: 10.1210/jc.2007-0919]
87 **Kodama K**, Tojjar D, Yamada S, Toda K, Patel CJ, Butte AJ. Ethnic differences in the relationship between insulin sensitivity and insulin response: a systematic review and meta-analysis. *Diabetes Care* 2013; **36**: 1789-1796 [PMID: 23704681 DOI: 10.2337/dc12-1235]
88 **Duan S**, Huang RS, Zhang W, Bleibel WK, Roe CA, Clark TA, Chen TX, Schweitzer AC, Blume JE, Cox NJ, Dolan ME. Genetic architecture of transcript-level variation in humans. *Am J Hum Genet* 2008; **82**: 1101-1113 [PMID: 18439551 DOI: 10.1016/j.ajhg.2008.03.006]
89 **Stranger BE**, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, Price A, Raj T, Nisbett J, Nica AC, Beazley C, Durbin R, Deloukas P, Dermitzakis ET. Patterns of cis regulatory variation in diverse human populations. *PLoS Genet* 2012; **8**: e1002639 [PMID: 22532805 DOI: 10.1371/journal.pgen.1002639]
90 **Zhang W**, Duan S, Kistner EO, Bleibel WK, Huang RS, Clark TA, Chen TX, Schweitzer AC, Blume JE, Cox NJ, Dolan ME. Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet* 2008; **82**: 631-640 [PMID: 18313023 DOI: 10.1016/j.ajhg.2007.12.015]
91 **Das SK**, Sharma NK, Hasstedt SJ, Mondal AK, Ma L, Langberg KA, Elbein SC. An integrative genomics approach identifies activation of thioredoxin/thioredoxin reductase-1-mediated oxidative stress defense pathway and inhibition of angiogenesis in obese nondiabetic human subjects. *J Clin Endocrinol Metab* 2011; **96**: E1308-E1313 [PMID: 21593104 DOI: 10.1210/jc.2011-0101]
92 **Allen JD**, Xie Y, Chen M, Girard L, Xiao G. Comparing statistical methods for constructing large scale gene networks. *PLoS One* 2012; **7**: e29348 [PMID: 22272232 DOI: 10.1371/journal.pone.0029348]
93 **Zhang B**, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005; **4**: Article17 [PMID: 16646834 DOI: 10.2202/1544-6115.1128]
94 **Langfelder P**, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? *PLoS Comput Biol* 2011; **7**: e1001057 [PMID: 21283776 DOI: 10.1371/journal.pcbi.1001057]
95 **Zhang B**, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA, Zhang C, Xie T, Tran L, Dobrin R, Fluder E, Clurman B, Melquist S, Narayanan M, Suver C, Shah H, Mahajan M, Gillis T, Mysore J, MacDonald ME, Lamb JR, Bennett DA, Molony C, Stone DJ, Gudnason V, Myers AJ, Schadt EE, Neumann H, Zhu J, Emilsson V. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* 2013; **153**: 707-720 [PMID: 23622250 DOI: 10.1016/j.cell.2013.03.030]
96 **Kang HP**, Yang X, Chen R, Zhang B, Corona E, Schadt EE, Butte AJ. Integration of disease-specific single nucleotide polymorphisms, expression quantitative trait loci and coexpression networks reveal novel candidate genes for type 2 diabetes. *Diabetologia* 2012; **55**: 2205-2213 [PMID: 22584726 DOI: 10.1007/s00125-012-2568-3]
97 **Schadt EE**, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, Sachs A, Lusis AJ. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 2005; **37**: 710-717 [PMID: 15965475 DOI: 10.1038/ng1589]
98 **Zhu J**, Wiener MC, Zhang C, Fridman A, Minch E, Lum PY, Sachs JR, Schadt EE. Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS Comput Biol* 2007; **3**: e69 [PMID: 17432931 DOI: 10.1371/journal.pcbi.0030069]
99 **Schadt EE**, Björkegren JL. NEW: network-enabled wisdom in biology, medicine, and health care. *Sci Transl Med* 2012; **4**: 115rv1 [PMID: 22218693 DOI: 10.1126/scitranslmed.3002132]
100 **Manolio TA**, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. *Nat Rev Genet* 2006; **7**: 812-820 [PMID: 16983377 DOI: 10.1038/nrg1919]
101 **Thomas D**. Gene--environment-wide association studies: emerging approaches. *Nat Rev Genet* 2010; **11**: 259-272 [PMID: 20212493 DOI: 10.1038/nrg2764]
102 **Tucker KL**, Smith CE, Lai CQ, Ordovas JM. Quantifying diet for nutrigenomic studies. *Annu Rev Nutr* 2013; **33**: 349-371 [PMID: 23642200 DOI: 10.1146/annurev-nutr-072610-145203]
103 **Ober C**, Vercelli D. Gene-environment interactions in human disease: nuisance or opportunity? *Trends Genet* 2011; **27**: 107-115 [PMID: 21216485 DOI: 10.1016/j.tig.2010.12.004]
104 **Patel CJ**, Chen R, Kodama K, Ioannidis JP, Butte AJ. Systematic identification of interaction effects between genome- and environment-wide associations in type 2 diabetes mellitus. *Hum Genet* 2013; **132**: 495-508 [PMID: 23334806 DOI: 10.1007/s00439-012-1258-z]
105 **Patel CJ**, Chen R, Butte AJ. Data-driven integration of epidemiological and toxicological data to select candidate interacting genes and environmental factors in association with disease. *Bioinformatics* 2012; **28**: i121-i126 [PMID: 22689751 DOI: 10.1093/bioinformatics/bts229]
106 **Patel CJ**, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PLoS One* 2010; **5**: e10746 [PMID: 20505766 DOI: 10.1371/journal.pone.0010746]
107 **Smith CE**, Ngwa J, Tanaka T, Qi Q, Wojczynski MK, Lemai-

tre RN, Anderson JS, Manichaikul A, Mikkilä V, van Rooij FJ, Ye Z, Bandinelli S, Frazier-Wood AC, Houston DK, Hu F, Langenberg C, McKeown NM, Mozaffarian D, North KE, Viikari J, Zillikens MC, Djoussé L, Hofman A, Kähönen M, Kabagambe EK, Loos RJ, Saylor GB, Forouhi NG, Liu Y, Mukamal KJ, Chen YD, Tsai MY, Uitterlinden AG, Raitakari O, van Duijn CM, Arnett DK, Borecki IB, Cupples LA, Ferrucci L, Kritchevsky SB, Lehtimäki T, Qi L, Rotter JI, Siscovick DS, Wareham NJ, Witteman JC, Ordovás JM, Nettleton JA. Lipoprotein receptor-related protein 1 variants and dietary fatty acids: meta-analysis of European origin and African American studies. *Int J Obes* (Lond) 2013; **37**: 1211-1220 [PMID: 23357958 DOI: 10.1038/ijo.2012.215]

108 **Smith EN**, Kruglyak L. Gene-environment interaction in yeast gene expression. *PLoS Biol* 2008; **6**: e83 [PMID: 18416601 DOI: 10.1371/journal.pbio.0060083]

109 **Gerke J**, Lorenz K, Ramnarine S, Cohen B. Gene-environment interactions at nucleotide resolution. *PLoS Genet* 2010; **6**: e1001144 [PMID: 20941394 DOI: 10.1371/journal.pgen.1001144]

110 **Gagneur J**, Stegle O, Zhu C, Jakob P, Tekkedil MM, Aiyar RS, Schuon AK, Pe'er D, Steinmetz LM. Genotype-environment interactions reveal causal pathways that mediate genetic effects on phenotype. *PLoS Genet* 2013; **9**: e1003803 [PMID: 24068968 DOI: 10.1371/journal.pgen.1003803]

111 **Dermitzakis ET**. Cellular genomics for complex traits *Nat Rev Genet* 2012; **13**: 215-220 [PMID: 22330769 DOI: 10.1038/nrg3115]

112 **Huang RS**, Duan S, Shukla SJ, Kistner EO, Clark TA, Chen TX, Schweitzer AC, Blume JE, Dolan ME. Identification of genetic variants contributing to cisplatin-induced cytotoxicity by use of a genomewide approach. *Am J Hum Genet* 2007; **81**: 427-437 [PMID: 17701890 DOI: 10.1086/519850]

113 **Maranville JC**, Luca F, Richards AL, Wen X, Witonsky DB, Baxter S, Stephens M, Di Rienzo A. Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes. *PLoS Genet* 2011; **7**: e1002162 [PMID: 21750684 DOI: 10.1371/journal.pgen.1002162]

114 **Smirnov DA**, Morley M, Shin E, Spielman RS, Cheung VG. Genetic analysis of radiation-induced changes in human gene expression. *Nature* 2009; **459**: 587-591 [PMID: 19349959 DOI: 10.1038/nature07940]

115 **Grundberg E**, Adoue V, Kwan T, Ge B, Duan QL, Lam KC, Koka V, Kindmark A, Weiss ST, Tantisira K, Mallmin H, Raby BA, Nilsson O, Pastinen T. Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. *PLoS Genet* 2011; **7**: e1001279 [PMID: 21283786 DOI: 10.1371/journal.pgen.1001279]

116 **Romanoski CE**, Lee S, Kim MJ, Ingram-Drake L, Plaisier CL, Yordanova R, Tilford C, Guan B, He A, Gargalovic PS, Kirchgessner TG, Berliner JA, Lusis AJ. Systems genetics analysis of gene-by-environment interactions in human cells. *Am J Hum Genet* 2010; **86**: 399-410 [PMID: 20170901 DOI: 10.1016/j.ajhg.2010.02.002]

117 **Inselman AL**, Hansen DK, Lee HY, Nakamura N, Ning B, Monteiro JP, Varma V, Kaput J. Assessment of research models for testing gene-environment interactions. *Eur J Pharmacol* 2011; **668** Suppl 1: S108-S116 [PMID: 21816149 DOI: 10.1016/j.ejphar.2011.05.084]

118 **Huang C**, Florez JC. Pharmacogenetics in type 2 diabetes: potential implications for clinical practice. *Genome Med* 2011; **3**: 76 [PMID: 22126607 DOI: 10.1186/gm292]

119 **Giacomini KM**, Yee SW, Ratain MJ, Weinshilboum RM, Kamatani N, Nakamura Y. Pharmacogenomics and patient care: one size does not fit all. *Sci Transl Med* 2012; **4**: 153ps18 [PMID: 23019654 DOI: 10.1126/scitranslmed.3003471]

120 **Manolopoulos VG**, Ragia G, Tavridou A. Pharmacogenomics of oral antidiabetic medications: current data and pharmacoepigenomic perspective. *Pharmacogenomics* 2011; **12**: 1161-1191 [PMID: 21843065 DOI: 10.2217/pgs.11.65]

121 **DeFronzo RA**, Tripathy D, Schwenke DC, Banerji M, Bray GA, Buchanan TA, Clement SC, Henry RR, Hodis HN, Kitabchi AE, Mack WJ, Mudaliar S, Ratner RE, Williams K, Stentz FB, Musi N, Reaven PD. Pioglitazone for diabetes prevention in impaired glucose tolerance. *N Engl J Med* 2011; **364**: 1104-1115 [PMID: 21428766 DOI: 10.1056/NEJMoa1010949]

122 **Igarashi M**, Jimbu Y, Kimura M, Hirata A, Yamaguchi H, Tominaga M. Effect of pioglitazone on atherogenic outcomes in type 2 diabetic patients: a comparison of responders and non-responders. *Diabetes Res Clin Pract* 2007; **77**: 389-398 [PMID: 17275945 DOI: 10.1016/j.diabres.2006.12.022]

123 **Rasouli N**, Kern PA, Elbein SC, Sharma NK, Das SK. Improved insulin sensitivity after treatment with PPARγ and PPARα ligands is mediated by genetically modulated transcripts. *Pharmacogenet Genomics* 2012; **22**: 484-497 [PMID: 22437669 DOI: 10.1097/FPC.0b013e328352a72e]

124 **Kasarskis A**, Yang X, Schadt E. Integrative genomics strategies to elucidate the complexity of drug response. *Pharmacogenomics* 2011; **12**: 1695-1715 [PMID: 22118053 DOI: 10.2217/pgs.11.115]

125 **Wang L**. Pharmacogenomics: a systems approach. *Wiley Interdiscip Rev Syst Biol Med* 2010; **2**: 3-22 [PMID: 20836007 DOI: 10.1002/wsbm.42]

126 **Gaffney DJ**, Veyrieras JB, Degner JF, Pique-Regi R, Pai AA, Crawford GE, Stephens M, Gilad Y, Pritchard JK. Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol* 2012; **13**: R7 [PMID: 22293038 DOI: 10.1186/gb-2012-13-1-r7]

127 **Gonzàlez-Porta M**, Calvo M, Sammeth M, Guigó R. Estimation of alternative splicing variability in human populations. *Genome Res* 2012; **22**: 528-538 [PMID: 22113879 DOI: 10.1101/gr.121947.111]

128 **Kerkel K**, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, Li K, Murty VV, Schupf N, Vilain E, Morris M, Haghighi F, Tycko B. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet* 2008; **40**: 904-908 [PMID: 18568024 DOI: 10.1038/ng.174]

129 **Gamazon ER**, Ziliak D, Im HK, LaCroix B, Park DS, Cox NJ, Huang RS. Genetic architecture of microRNA expression: implications for the transcriptome and complex traits. *Am J Hum Genet* 2012; **90**: 1046-1063 [PMID: 22658545 DOI: 10.1016/j.ajhg.2012.04.023]

130 **Sanyal A**, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature* 2012; **489**: 109-113 [PMID: 22955621 DOI: 10.1038/nature11279]

131 **Maurano MT**, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012; **337**: 1190-1195 [PMID: 22955828 DOI: 10.1126/science.1222794]

132 **Trynka G**, Sandor C, Han B, Xu H, Stranger BE, Liu XS, Raychaudhuri S. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* 2013; **45**: 124-130 [PMID: 23263488 DOI: 10.1038/ng.2504]

133 **Pasquali L**, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman I, Tena JJ, Morán I, Gómez-Marín C, van de Bunt M, Ponsa-Cobas J, Castro N, Nammo T, Cebola I, García-Hurtado J, Maestro MA, Pattou F, Piemonti L, Berney T, Gloyn AL, Ravassard P, Skarmeta JL, Müller F, McCarthy MI, Ferrer J. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* 2014; **46**: 136-143 [PMID: 24413736 DOI: 10.1038/ng.2870]

134 **Ward LD**, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol* 2012; **30**:

1095-1106 [PMID: 23138309 DOI: 10.1038/nbt.2422]
135 **Edwards SL**, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet* 2013; **93**: 779-797 [PMID: 24210251 DOI: 10.1016/j.ajhg.2013.10.012]
136 **Chakravarti A**, Clark AG, Mootha VK. Distilling pathophysiology from complex disease genetics. *Cell* 2013; **155**: 21-26 [PMID: 24074858 DOI: 10.1016/j.cell.2013.09.001]
137 **Pound LD**, Sarkar SA, Cauchi S, Wang Y, Oeser JK, Lee CE, Froguel P, Hutton JC, O'Brien RM. Characterization of the human SLC30A8 promoter and intronic enhancer. *J Mol Endocrinol* 2011; **47**: 251-259 [PMID: 21798992 DOI: 10.1530/JME-11-0055]
138 **Cauchi S**, Del Guerra S, Choquet H, D'Aleo V, Groves CJ, Lupi R, McCarthy MI, Froguel P, Marchetti P. Meta-analysis and functional effects of the SLC30A8 rs13266634 polymorphism on isolated human pancreatic islets. *Mol Genet Metab* 2010; **100**: 77-82 [PMID: 20138556 DOI: 10.1016/j.ymgme.2010.01.001]
139 **Pang DX**, Smith AJ, Humphries SE. Functional analysis of TCF7L2 genetic variants associated with type 2 diabetes. *Nutr Metab Cardiovasc Dis* 2013; **23**: 550-556 [PMID: 22402060 DOI: 10.1016/j.numecd.2011.12.012]
140 **Mondal AK**, Sharma NK, Elbein SC, Das SK. Allelic expression imbalance screening of genes in chromosome 1q21-24 region to identify functional variants for Type 2 diabetes susceptibility. *Physiol Genomics* 2013; **45**: 509-520 [PMID: 23673729 DOI: 10.1152/physiolgenomics.00048.2013]
141 **Mondal AK**, Das SK, Baldini G, Chu WS, Sharma NK, Hackney OG, Zhao J, Grant SF, Elbein SC. Genotype and tissue-specific effects on alternative splicing of the transcription factor 7-like 2 gene in humans. *J Clin Endocrinol Metab* 2010; **95**: 1450-1457 [PMID: 20097709 DOI: 10.1210/jc.2009-2064]
142 **Peters DT**, Musunuru K. Functional evaluation of genetic variation in complex human traits. *Hum Mol Genet* 2012; **21**: R18-R23 [PMID: 22936690 DOI: 10.1093/hmg/dds363]
143 **Melnikov A**, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan CG, Kinney JB, Kellis M, Lander ES, Mikkelsen TS. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* 2012; **30**: 271-277 [PMID: 22371084 DOI: 10.1038/nbt.2137]
144 **Patwardhan RP**, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM, Ahituv N, Pennacchio LA, Shendure J. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat Biotechnol* 2012; **30**: 265-270 [PMID: 22371081 DOI: 10.1038/nbt.2136]
145 **Smith RP**, Taher L, Patwardhan RP, Kim MJ, Inoue F, Shendure J, Ovcharenko I, Ahituv N. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat Genet* 2013; **45**: 1021-1028 [PMID: 23892608 DOI: 10.1038/ng.2713]
146 **Kheradpour P**, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* 2013; **23**: 800-811 [PMID: 23512712 DOI: 10.1101/gr.144899.112]
147 **Cox RD**, Church CD. Mouse models and the interpretation of human GWAS in type 2 diabetes and obesity. *Dis Model Mech* 2011; **4**: 155-164 [PMID: 21324932 DOI: 10.1242/dmm.000414]
148 **Cheung VG**, Nayak RR, Wang IX, Elwyn S, Cousins SM, Morley M, Spielman RS. Polymorphic cis- and trans-regulation of human gene expression. *PLoS Biol* 2010; **8**: [PMID: 20856902 DOI: 10.1371/journal.pbio.1000480]

P- Reviewers: Cai H, Guneli EM, Harwood HJ, Mansour AA, Vorobjova T **S- Editor**: Wen LL **L- Editor**: A **E- Editor**: Wu HL

百世登
Baishideng®

Published by **Baishideng Publishing Group Co., Limited**
Flat C, 23/F., Lucky Plaza,
315-321 Lockhart Road, Wan Chai, Hong Kong, China
Fax: +852-65557188
Telephone: +852-31779906
E-mail: bpgoffice@wjgnet.com
http://www.wjgnet.com