Minireview

Genetics and functional genomics of type 2 diabetes mellitus
Ayo Toye and Dominique Gauguier

Address: The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.

Correspondence: Dominique Gauguier. E-mail: gdomi@well.ox.ac.uk

Abstract

Genome-wide studies of transcription in the skeletal muscle of type 2 diabetic patients have identified coordinated changes in the expression of genes involved in oxidative phosphorylation, and have underlined the central role of the oxidative-phosphorylation regulator, PCG1α. These findings help unravel the complex pathogenesis and inheritance of polygenic type 2 diabetes mellitus.

Type 2 diabetes mellitus is a heterogeneous clinical entity, characterized by high blood sugar levels or hyperglycaemia, arising from a deteriorated tissue response to the biological effects of insulin (insulin resistance) [1] and impaired glucose-induced insulin secretion [2]. It is a classical example of multifactorial disorder: the etiology of type 2 diabetes combines both genetic and environmental factors. The 'westernized lifestyle' is a likely cause of the recent rapid rise in disease incidence, which is projected to double worldwide by the year 2025 (of 150 million diabetics globally, approximately 90% live in western countries). The management of type 2 diabetes and its complications accounts for a significant share of national annual healthcare spending in all western countries. Together, these facts have galvanized international efforts to research the causes of type 2 diabetes mellitus and to develop effective preventive and therapeutic tools and strategies for managing the disease.

Identifying causes and consequences of hyperglycaemia

International genetic and genomic projects have delivered a wide range of tools and resources for genome-wide investigations to help the search for genetic factors involved in type 2 diabetes mellitus and have contributed to recent progress in our knowledge of the etiology of the disease (for the remainder of this article, when we say 'diabetes' we mean type 2 diabetes mellitus). Genetic markers and dense genetic maps allow extensive searches for gene variants that co-segregate with the disease. Although many genetic loci have been described that are linked to either diabetes or associated pathophysiological markers [3], *CALPAIN10* remains the only known gene underlying polygenic type 2 diabetes mellitus [4].

The vast amount of information derived from the annotation of the human genome and the genomes of model organisms provide opportunities for using new strategies to identify genes underlying diabetes. Knowledge of the human transcriptome and proteome provides unparalleled power for searching comprehensively for the causes and consequences of the disease, and diagnostic and therapeutic targets and agents relevant to it. In theory it is possible to characterize gene-expression patterns (in terms of transcription, translation and post-translational changes) in any cell type and to generate a broad functional picture of the molecular mechanisms involved in health and disease. Differences in gene expression between diabetic patients and controls are indicative of a relationship between a specific gene or series of genes and the disease in a specific context.

The transcriptome forms the crucial intermediate between the genetic-code template and the protein products that constitute the structural components and operational machinery of organisms. The technology required to array large numbers of transcripts or oligonucleotides and detect differences in the amount of individual transcripts is obviously

more straightforward than that required for high-throughput genome-wide protein screening. Fast and efficient genome-wide transcription profiling, using microarrays, can reveal an overall picture of genes potentially involved in the causes and consequences of a disease, and this technology has been successfully used in various fields of biomedical research.

The diabetic muscle transcriptome

Skeletal muscle constitutes about 40% of mammalian body mass and is, with fat and liver tissues, a key site of insulin stimulated glucose disposal [5]. Impaired insulin-mediated glucose disposal by muscle is a key feature of diabetes [6] and skeletal muscles are, therefore, key tissue targets for gene-expression studies of diabetes. Two recent landmark studies [7,8] have used microarray analyses to highlight the central role in the human diabetic-muscle phenotype of two key regulators of oxidative phosphorylation: the peroxisome proliferation-activated receptor-γ (PPARγ) coactivator 1 (PGC1α) and nuclear respiratory factor (NRF1). These studies were based on determination of the expression levels of large sets of 7,129 [8] and 22,000 genes [7], respectively, in muscle biopsies of diabetic patients, prediabetics and controls.

Novel analytical methods used to exploit the microarray data generated in the two studies clearly underline the importance of considering modest but coordinated changes in genome-wide transcription patterns, rather than individual changes that meet an arbitrary threshold. None of the genes examined in either study was differentially expressed between the different groups tested (diabetic patients, prediabetics and controls) following robust statistical correction for false discovery during multiple hypothesis testing. Both studies examined differential covariation - between test subject groups or classes - of sets of genes (within gene group covariation). Genes were labeled by ontology classification, then ranked by differential expression, and this sequential procedure facilitated the detection of pathway-dependent co-regulated genes. Mootha *et al.* [7] computed a Kolmogorov-Smirnov test statistic (termed Gene Sets Enrichment Approach, or GSEA) for each gene, which was then evaluated against an empirically determined (permutation test) threshold for declaring association between a gene set and diabetes. In contrast, Patti *et al.* [8] examined only genes that were differentially expressed before correction for multiple testing ($p < 0.05$).

Remarkably, both studies found that the transcription levels of a class of genes involved in oxidative phosphorylation mechanisms were consistently lower in diabetics than in controls, even following robust correction for multiple testing, and both therefore suggested that the downregulation of mitochondrial oxidative phosphorylation genes may precede diabetes onset. The results of real-time quantitative PCR demonstrated that the expression of NRF1, which regulates the expression of nuclear-encoded mitochondrial

genes, was decreased by 29% in diabetics compared to non-diabetics [8]. Both studies [7,8] also found that PGC1α expression was significantly lower (22-36%) in diabetics than in other test groups, and Patti *et al.* [8] showed that PGC1β expression was also significantly decreased (46%) in diabetics compared to non-diabetic subjects.

Importance of genetic and genomic expression studies

Poor replicability of significant genetic linkage and association to a disease is often attributed to a number of factors, including, for example, ethnic stratification, population-specific linkage disequilibrium between markers and causal variants, and gene-by-gene and gene-by-environment interactions. The possible importance of some of these factors has been addressed in an attempt to replicate previously published associations in 16 candidate genes for type 2 diabetes mellitus. The gene encoding PPARγ was the only one for which the study could confirm an association between common variants and a significant increase in diabetes risk [9]. The existence in a non-diabetic population of disease susceptibility gene variants that are not expressed phenotypically, for example because of the late onset of diabetes, is also a factor that reduces the power of genetic-linkage and association analyses for this disease. Remarkably, although Mootha *et al.* [7] and Patti *et al.* [8] independently investigated patient samples from different populations (Mexican-Americans and Caucasians, respectively) using different microarrays and different data-analysis tools, the results from both studies converged at the same gene pathway. Using biological samples from patients selected for homogenous clinical phenotypes may have minimized the effect of genetic variability on gene transcription. It was, however, reported that reduced oxidative phosphorylation gene transcription was also found in 'prediabetics' [8].

Population size is also a key factor affecting the detection of genetic linkage and association. Results from association studies in type 1 diabetes mellitus provided support for the requirement of very large population and family datasets to improve the statistical robustness and reliability of genetic-association studies of complex traits [10]. On the other hand, screening a very large number of single nucleotide polymorphisms (SNPs) in a relatively modest panel of patients was sufficient to identify a functional sequence variant associated with susceptibility to myocardial infarction [11]. It is becoming apparent that, in contrast to traditional genetic association studies, gene-expression profiling studies can provide robust and functionally valid results with samples from a very limited number of well-selected patients and controls [7,8].

But in contrast to gene-expression profiling, genetic-linkage mapping provides statistical evidence for the involvement of chromosomal regions co-segregating with a disease or

disease-associated phenotypes. It is essentially a strategy to define etiologically important factors underlying multifactorial disorders, in which potential biases are either minimized or well defined. Unless all pathophysiological variables associated with diabetes - glucose intolerance, altered insulin secretion and action, obesity, dyslipidemia and high blood pressure - and its complications are comprehensively characterized in both patients and controls of a large cohort, results from linkage analysis can only identify chromosomal regions involved in the control of one phenotype. This strategy removes a number of the constraints and limitations of functional studies, including for example tissue- and developmental-stage-specific gene expression. Most importantly, linkage analysis can point to sequence elements, including pseudogenes [12], non polyadenylated RNAs and non-coding sequences [13] that may play an important unrecognized role in complex diseases but cannot be directly tested by current microarray-based transcription-profiling technologies.

Altered gene expression in diabetes - cause or consequence?

Determining the causes and consequences of a complex disease is the most challenging issue that genetic and functional genomic studies currently attempt to address [14]. It is pertinent to consider conceptually how decreased oxidative phosphorylation might result in diabetes. A large body of existing scientific literature supports the hypothesis of Patti *et al.* [8]: decreased expression of PGC1α coupled with reduced expression of NRF1 and PPARγ eventually result in decreased oxidative phosphorylation and lipid oxidation, accumulation of lipid in skeletal muscle and ultimately diabetes. Many organs, including endocrine pancreas, liver and fat, are important in maintaining glucose homeostasis and energy balance. One must therefore consider whether differential transcription of components of the oxidative-phosphorylation pathway in diabetes is limited to muscle and how it relates to the whole-organism functional picture of diabetes. For example, ATP generated from oxidative phosphorylation is required for normal glucose-stimulated insulin release. The downregulation of oxidative-phosphorylation genes seen in diabetic muscle would be expected to be mirrored in pancreatic islets and may account for a lack of insulin related to beta-cell dysfunction in advanced diabetes [15].

In conclusion, transcriptomics is well advanced for technological reasons and will deliver vast amounts of information on the complex regulation of tissue- and time-specific gene expression in disease onset and progression. One of the important perspectives of genome-wide transcriptomic studies lies in the possibility to integrate genes and expressed sequence tags (ESTs) of unknown or poorly defined roles into functional pathways, thus improving functional genome annotation. Although GSEA is a significant advance for interpreting microarray data, it relies on our current understanding of the regulation of metabolic and hormonal networks. Results from the investigation of the diabetic muscle transcriptome [7,8] suggest that gene-transcription profiling on its own may have a limited impact on the enrichment of system biology maps. Gene transcription, however, represents only a single dimension in functional genome maps. In particular, as exemplified by Epstein *et al.* [16], post-translational modification is a crucial mechanism for gene regulation that is not necessarily discernible by studying the transcriptome. Metabonomics and metabolomics are also genome-wide technologies that will enrich genome annotation in health and disease [17,18]. Global integrative analyses of multidimensional genomic parameters exploiting the concept of a biological atlas advanced by Vidal *et al.* [19] probably represents the most optimal use of resources to isolate variables that are relevant to complex diseases in the whole organism, and to describe their interplay.

References
1. Saltiel AR, Kahn CR: **Insulin signalling and the regulation of glucose and lipid metabolism.** *Nature* 2001, **414**:799-806.
2. Bell GI, Polonsky KS: **Diabetes mellitus and genetically programmed defects in beta-cell function.** *Nature* 2001, **414**:788-791.
3. Hanson RL, Knowler WC: **Quantitative trait linkage studies of diabetes-related traits.** *Curr Diab Rep* 2003, **3**:176-183.
4. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, *et al.*: **Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes Mellitus.** *Nat Genet* 2000, **26**:163-175.
5. Yang X, Pratley RE, Tokraks S, Bogardus C, Permana PA: **Microarray profiling of skeletal muscle tissues from equally obese, non-diabetic insulin-sensitive and insulin-resistant Pima Indians.** *Diabetologia* 2002, **45**:1584-1593.
6. Yechoor VK, Patti ME, Saccone R, Kahn CR: **Coordinated patterns of gene expression for substrate and energy metabolism in skeletal muscle of diabetic mice.** *Proc Natl Acad Sci USA* 2002, **99**:10587-10592.
7. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, *et al.*: **PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.** *Nat Genet* 2003, **34**:267-273.
8. Patti ME, Butte AJ, Crunkhorn S, Cusi K, Berria R, Kashyap S, Miyazaki Y, Kohane I, Costello M, Saccone R, *et al.*: **Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1.** *Proc Natl Acad Sci USA* 2003, **100**:8466-8471.
9. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, *et al.*: **The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes.** *Nat Genet* 2000, **26**:76-80.
10. Dahlman I, Eaves IA, Kosoy R, Morrison VA, Heward J, Gough SC, Allahabadia A, Franklyn JA, Tuomilehto J, Tuomilehto-Wolf E, *et al.*: **Parameters for reliable results in genetic association studies in common disease.** *Nat Genet* 2002, **30**:149-150.
11. Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Hiroshi, Sato H, Sato H, Hori M, *et al.*: **Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction.** *Nat Genet* 2002, **32**:650-654.

12. Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, Takahashi S, Yagami K, Wynshaw-Boris A, Yoshiki A: **An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene.** *Nature* 2003, **423:**91-96.

13. Eddy SR: **Non-coding RNA genes and the modern RNA world.** *Nat Rev Genet* 2001, **2:**919-929.

14. Page GP, George V, Go RC, Page PZ, Allison DB: **"Are We There Yet?": deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits.** *Am J Hum Genet* 2003, **73:**711-719.

15. Weir GC, Laybutt DR, Kaneto H, Bonner-Weir S, Sharma A: **Beta-cell adaptation and decompensation during the progression of diabetes.** *Diabetes* 2001, **50:**S154-S159.

16. Epstein AC, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole DR, Mukherji M, Metzen E, Wilson MI, Dhanda A, *et al.*: **C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation.** *Cell* 2001, **107:**43-54.

17. Nicholson JK, Wilson ID: **Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism.** *Nat Rev Drug Discov* 2003, **2:**668-676.

18. Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HW, Clarke S, Schofield PM, McKilligin E, Mosedale DE, Grainger DJ: **Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics.** *Nat Med* 2002, **8:**1439-1444.

19. Vidal M: **A biological atlas of functional maps.** *Cell* 2001, **104:**333-339.