Technologies for Diabetes Genomics

Patricia W. Mueller, Ph.D.

Abstract

The genetic risk for diabetes largely depends on the type of diabetes and the penetrance and severity of the effect of the contributing genes. This ranges from the high-risk mutations of neonatal diabetes and maturity-onset diabetes of the young to the lower, but still significant, risk conferred by common human leukocyte antigen alleles in type 1 diabetes to the still-lower risk conferred by the common variants associated with type 2 diabetes. There are many new molecular technologies, each with their own set of methodological issues, that have been used for genome-wide association studies and that can be used for determining the genetic risk for these various types of diabetes. These technologies include whole genome single nucleotide polymorphism microarrays, high-throughput polymorphism analyzers, next-generation sequencers, and copy-number variant technologies.

J Diabetes Sci Technol 2009;3(4):735-738

Introduction

Recent advances in molecular technology have resulted in considerable research into the genetics of different forms of diabetes and the role that genetics plays in predicting diabetes. The genetic risk for diabetes largely depends on the type of diabetes and the penetrance and severity of the effect of the contributing genes. This ranges from the higher penetrant mutations of neonatal diabetes and maturity-onset diabetes of the young (MODY) to the lower, but still significant, risk conferred by common human leukocyte antigen (HLA) alleles in type 1 diabetes to the still-lower risk conferred by the common variants associated with type 2 diabetes. The following is a brief review of the different types of genetic risk for these types of diabetes and an overview of the many new molecular technologies, some of which have been used to discover these risk factors, others of which can be used to measure them on a more routine basis, and technologies designed to measure less understood types of genomic variation such as copy-number variants that are being used in current investigations of disease.

Genetic Risk in Different Types of Diabetes

Mutations in three genes cause permanent neonatal diabetes in 50% of patients diagnosed with diabetes before 6 months of age. These children usually do not have the autoantibodies predictive of type 1 diabetes. Specific mutations in the islet potassium channel Kir6.2 encoded by KCNJ11, the sulfonylurea receptor (SUR1) encoded by ABCC8, and the insulin gene are highly

penetrant for this form of diabetes.[1] Different mutations within a gene can also cause different effects; for example, loss of function mutations in KCNJ11 and ABCC8 can cause over-secretion of insulin while gain of function mutations can cause the opposite effect.[2] In addition to being more predictive, this genetic information has treatment implications since children whose diabetes is caused by mutations in the genes encoding for Kir6.2 and SUR1 achieve better glycemic control when treated with sulfonylurea compounds than when treated with insulin. Because of these treatment implications, it is increasingly common for children who develop diabetes before 6 months of age to be tested for these mutations.

A relatively common form of monogenic diabetes is MODY, and this form of diabetes has significant diagnostic overlap with other forms of diabetes seen in young adults.[3] Almost 90% of MODY cases result from mutations in glucokinase (encoded by GCK) and transcription factors HNF-1α (encoded by TCF1 or HNF1A), HNF-1β (encoded by TCF2 or HNF1B), and HNF-4α (encoded by HNF4A). Since the different molecular forms of MODY result in different clinical phenotypes, molecular diagnostics can contribute considerably to defining the prognosis and therapy of this form of diabetes as well.

For type 1 diabetes, about half of the total risk is genetic and about half of that genetic risk is in the HLA region on the short arm of chromosome 6. Other genes, such as the 5′ region of the insulin gene, PTPN22, CD25, and IL-2 have been associated with type 1 diabetes in genome-wide association studies as well, but their contribution to risk is small. The highest associations of type 1 diabetes with genetic risk occur in the HLA class II region at DRB1 and DQB1, but genome scan associations extend considerable distances from these genes. The DRB1*03 and *04 alleles that are most highly associated with type 1 diabetes are also common in the population; therefore, the predictive value of these HLA alleles is low. However, it has been suggested that extending these haplotypes into HLA class I sites and further, up to lengths of 9 million base pairs, can greatly improve predictive algorithms for the general population.[4] Currently, a strategy for identifying those at risk of developing type 1 diabetes before the onset of clinical symptoms is being studied in research settings. This approach is to screen for high-risk HLA haplotypes in newborns and then follow those children for the development of the autoantibodies predictive of type 1 diabetes. The three most commonly used autoantigens for these studies are glutamic acid decarboxylase 65, the intracellular region of islet cell antigen-2 (IA-2ic), and

insulin. A fourth autoantigen, zinc transporter 8, has recently been introduced and evaluated.[5] As the number of these different types of autoantibodies increase in an individual, the risk of developing type 1 diabetes also increases. Therefore, multiple autoantibody positivity, also referred to as epitope spreading, raises the risk for developing the clinical symptoms of type 1 diabetes.

For type 2 diabetes, various genome-wide studies have identified 17 to 18 associated genomic loci, but no major locus equivalent to the risk conferred by the HLA region for type 1 diabetes has been found.[6–8] Interestingly, many of these loci implicate pancreatic beta-cell function in the pathogenesis of type 2 diabetes, and only one is clearly associated with insulin resistance. Of these, the largest effect size is just over 1.4 for the transcription factor, TCF7L2. The next largest effect sizes are in the range of 1.21 to 1.25 for KCNJ11, the peroxisome proliferator-activated receptor γ, and a kinase inhibitor involved in islet development, CDKN2A/B. While these genes offer opportunities for the development of therapeutic interventions and some are already drug targets, their evaluation as predictors of type 2 diabetes has not resulted in significant increases in the area under the receiver operating characteristic (ROC) curve for the plot of sensitivity versus (1-specificity) when compared to simple clinical algorithms. One study by Lango and colleagues compared a predictive model including age, body mass index (BMI), and sex that had an area under the ROC curve of 0.78, with the same model plus genetic risk variants, and the discriminative accuracy was only marginally increased to 0.80.[8] Another study by Wilson and associates used a clinical model consisting of parental diabetes, obesity, BMI, hypertension, low levels of high-density lipoprotein cholesterol, elevated triglyceride levels, and impaired fasting glucose that resulted in an area under the ROC curve of 0.85.[9]

Because genetic factors are most predictive in highly penetrant diseases, areas of growing application for genomic technologies are newborn screening and diagnostic medicine. It is also likely that there will be future applications for HLA analysis in the area of autoimmune disease prediction.

Genomic Technologies—Different Types for Different Purposes

There is considerable normal variability in the human genome, including single nucleotide polymorphisms (SNPs), copy-number variation (generally larger deletions and duplications), small deletions and insertions, pseudogenes,

gene rearrangements, short and long tandem repeats, and methylation, which is the source of imprinting that controls whether the maternal or paternal gene is expressed. There is also an explosion of new genomic technologies underway to measure these different types of variability, and even newer technologies are under development. The Centers for Disease Control and Prevention conducts studies with, develops methods for, and evaluates many of these technologies for a variety of public health applications.

These new technologies can be classified by use. Whole genome association SNP microarrays, such as Illumina Infinium and the Affymetrix arrays, with up to 1 million SNPs, are used for genomic searches for disease associations. These microarrays provide an efficient method of genotyping large numbers of predefined SNPs throughout the genome to determine the association of chromosomal regions with specific diseases.

High-throughput polymorphism analyzers, such as the Illumina Golden Gate system, the Sequenome, the AB SNPlex and Biotrove OpenArray real-time quantitative polymerase chain reaction (PCR) system, the Luminex, and the Third Wave system are used for analyzing multiple SNPs or SNP sets defined by the user or defined in analytical reagents provided by manufacturers. These platforms are appropriate for genotyping SNPs in large numbers of people and provide a means of multiplexing different polymorphisms or arraying singleplex assays to genotype them more efficiently and at lower cost than individual assays.

Sanger sequencing and pyrosequencing are current methods that can be used for focused resequencing in gene regions. In addition, resequencing microarrays are being used increasingly for mutation detection in gene regions with many mutations that are specific to small numbers of people, such as individual families.

Next-generation sequencers, such as the Illumina Solexa, the AB Solid, and the Roche 454 are newer approaches to sequencing that operate on different principles, depending on the manufacturer, and they have a much higher throughput. Their goal is to make whole genome sequencing more affordable and accessible. These sequencers may also be useful for focused gene region resequencing, but optimizing for this application is still in progress.

Copy-number variant technologies, such as the Multiple Ligation PCR Amplification probe sets, Agilent microarrays, NimbleGen microarrays, and Affymetrix and Illumina microarrays are methods that range from disease-specific probe sets to detect deletions and duplications that may not be detected by sequencing to arrays that can assess deletions and duplications in large genomic regions.

Genomic Technologies—Methodological Considerations

With each of these types of methodology comes an important set of method development and validation issues that are important for accurate results and appropriate troubleshooting of problems. For whole genome association studies using SNP microarrays, there is a need for stringent quality constraints to prevent artifactual associations between genomic loci and disease due to genotyping errors and lack of optimization of genotype calling software.

Important considerations for high-throughput polymorphism analysis are (1) allele dropout in some cases due to unknown SNPs that can destabilize primers and probes, (2) DNA isolation procedures that yield DNA of poor quality that might include interfering substances such as heparin or hemoglobin, (3) assay primer and probe interference when assays are multiplexed in one reaction vessel, and (4) software design and settings, especially when hardware from one manufacturer is used with reagents or software from another. For SNP microarrays, as well as other high-throughput SNP analyzers, the use of whole genome amplification with low DNA concentration samples can introduce allele ratio biases.

For Sanger sequencing, important considerations are appropriate design and optimized temperature selection to prevent allele dropout and nonspecific binding, strategies for dealing with frame shifts, the realization that sequencing may miss heterozygous deletions because of missing primer binding sites on only one chromosome, and appropriate method comparisons for validation. For microarray resequencing, sufficient resolution with sufficient oligonucleotide coverage and adequate oligonucleotide repetition to detect the mutations of interest accurately are needed.

For next-generation sequencers, some of the challenges are (1) primer design and DNA amplification strategies compatible with the different sequencing principles and approaches, (2) read length variation among the next-generation sequencers and incomplete reassembly of the sequence, and (3) the need for sequencing designs

that will work for both normal DNA and DNA with disease mutations, including complex chimerical gene rearrangements, deletions, and duplications.

Issues related to the new methods for detecting copy-number variation include (1) probe design with the appropriate resolution for exons and desired intron sequences, (2) destabilization of probes caused by unknown SNPs, (3) DNA quality and quantity requirements compatible with blood spots for newborn screening applications, and (4) validation using available transformed cell lines that may have chromosomal abnormalities and aneuploidy after repeated regrowth.

Other common issues include (1) availability of appropriate materials for validation, (2) DNA yield and concentration that can be limiting in newborn screening applications, (3) ease of operation, (4) whether to contract a measurement out or do it "in-house," and (5) the cost per sample for the desired sample throughput.

Summary

In summary, the ability to predict disease from genetic information in different forms of diabetes generally increases with the penetrance of the disease and the effect size of the variant and generally decreases with increasing numbers of genes of small effect. For diseases intermediate in this spectrum, such as type 1 diabetes and other autoimmune diseases, predictive strategies can be developed that take advantage of both information on genetic risk and the autoantibodies that are predictive of these diseases.

There are many new types of genomic technologies to measure the considerable variability in the human genome, each with their own set of methodological issues, and the types of diabetes for which genotyping will provide the most benefit are being clarified. When sequencing whole human genomes becomes common, the challenge for the future will be to distinguish between extensive normal genetic variability and the variability that causes or significantly predisposes to disease.

References:

1. Støy J, Greeley SA, Paz VP, Ye H, Pastore AN, Skowron KB, Lipton RB, Cogen FR, Bell GI, Philipson LH, United States Neonatal Diabetes Working Group. Diagnosis and treatment of neonatal diabetes: a United States experience. Pediatr Diabetes. 2008;9(5):450–9.

2. Flanagan SE, Ellard S. Identification of mutations in the Kir6.2 subunit of the K(ATP) channel. Methods Mol Biol. 2008;491:235–45.

3. Owen KR, McCarthy MI. Genetics of type 2 diabetes. Curr Opin Genet Dev. 2007;17(3):239–44.

4. Baschal EE, Eisenbarth GS. Extreme genetic risk for type 1A diabetes in the post-genome era. J Autoimmun. 2008;31(1):1–6.

5. Wenzlau JM, Juhl K, Yu L, Moua O, Sarkar SA, Gottlieb P, Rewers M, Eisenbarth GS, Jensen J, Davidson HW, Hutton JC. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. Proc Natl Acad Sci U S A. 2007;104(43):17040–5.

6. Florez JC. Clinical review: the genetics of type 2 diabetes: a realistic appraisal in 2008. J Clin Endocrinol Metab. 2008;93(12):4633–42.

7. Ridderstråle M, Groop L. Genetic dissection of type 2 diabetes. Mol Cell Endocrinol. 2009;297(1-2):10–7.

8. Lango H, UK Type 2 Diabetes Genetics Consortium, Plamer CN, Morris AD, Zeggini E, Hattersley AT, McCarthy MI, Frayling TM, Weedon MN. Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes. 2008;57(11):3129–35.

9. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agnostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: The Framingham Offspring Study. Arch Intern Med. 2007;167(10):1068–74.