



Precision Medicine in Diabetes: Is It Time?

Diabetes Care 2016;39:1085–1088 | DOI: 10.2337/dc16-0586

Jose C. Florez

THE CHALLENGE: HETEROGENEITY OF DIABETES

Diabetes is a heterogenous disease. Its very definition, anchored by thresholds for hyperglycemia, rests on the final common event on which disparate pathological processes converge. Although the traditional classification into type 1 and type 2 diabetes has proven useful in differentiating distinct pathophysiological mechanisms with clear therapeutic implications, it remains insufficient in explaining the wide variety of clinical manifestations of this disease. For example, we see lean members of specific ethnic groups with antibody-negative, nonketotic diabetes; we treat patients with childhood-onset, antibody-positive diabetes who become insulin resistant as they age; we do not fully understand why some patients progress rapidly to microvascular and/or macrovascular complications or require aggressive escalation of therapy; and we cannot predict the rate of β -cell failure, the degree of weight loss required to normalize glycemia, or the type of medication best suited for a given patient. Attempts at capturing greater granularity have resulted in the creation of new entities, such as latent autoimmune diabetes in the adult to describe autoimmune diabetes with onset after age 30 (1), and ketosis-prone diabetes to describe antibody-negative diabetes with a ketotic onset but only transient insulin requirements (2,3); however, absent a full understanding of the mechanism and its implications, these subtypes emerge mired in controversy or fail to penetrate clinical care (4,5).

Evidently, despite extensive epidemiological and physiological characterization, we have fallen short in cataloging risk factors, identifying triggering events, elucidating pathophysiological pathways, outlining prognostic course, selecting effective therapies, and predicting complications. Specific patients continue to represent diagnostic challenges, and our approach to therapy continues to be based on population averages. Recent thoughtful attempts at tailoring type 2 diabetes therapy to individual patients have focused on personal, social, and economic considerations rather than on a lucid molecular understanding of the underlying disease process (6).

THE OPPORTUNITY: RECENT TECHNICAL AND ANALYTICAL ADVANCES

We are entering a new era in which technological developments coupled with expanded computational power and increased statistical sophistication have enabled the global query of discrete biological axes. Manufacturing advances introduced arrays that could measure mRNA transcript levels or DNA single nucleotide variants comprehensively in a single experiment. A deeper knowledge of the patterns of human genetics variation has allowed for the explosion of genome-wide association studies. Remarkable efficiency gains in high-throughput next-generation sequencing have seeded whole-exome and whole-genome sequencing studies as well as ancillary explorations of the transcriptome (RNA-seq), transcription factor binding sites (ChIP-seq), open chromatin (ATAC-seq), the epigenome, or the microbiome. Mass spectroscopy can be applied to the study of small metabolites or proteins in organic fluids, including posttranslational modifications. In this manner, multiple dimensions of the molecular architecture of biological systems can be interrogated with respect to native and perturbed metabolic states (7).

This technological progress has been accompanied by concomitant enhancements in bioinformatic and analytical tools, often shared publicly in the precompetitive space. Health care systems have digitalized clinical information and increasingly made the electronic medical record available for clinical investigation (8). Large private and even

Diabetes Unit and Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; Metabolism Program and Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; and Department of Medicine, Harvard Medical School, Boston, MA

Corresponding author: Jose C. Florez, jcflorez@partners.org.

Received 17 March 2016 and accepted 23 March 2016.

© 2016 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered.

national biobanks have been created to streamline this research function, and both funding bodies and scientific journals have required data sharing in central repositories as a condition for research support or publication.

We therefore live in the midst of a revolution of big data across all domains of the human experience, ranging from molecular to societal dimensions. We practice medicine and conduct research within an unprecedented whirlwind of data, spanning from populations to the individual. Soon, it will be possible to capture the metabolic state of a single patient at the molecular and cellular level with great precision through multiple time points in his or her development (9). Nonetheless, whether this emerging trove of information will deepen our understanding of health and disease and lead to improved outcomes in public health remains to be seen. An outstanding but crucial challenge to the field is the ability to integrate these disparate data sources in a manner that informs a holistic view of an organism such that synergy begets understanding.

GENOMICS IN DIABETES: EARLY SUCCESSES AND TEMPERED EXPECTATIONS

So where are we today on the generation and interpretation of big data with respect to diabetes? Are we producing the body of knowledge that can be applied to the individual patient in an effort to enhance precision in diagnostics and therapeutics?

Of all the potential data axes, genomics has seen the fastest progress to date, in part because of the exquisite precision now achieved in genetic measurements, the spectacular expansion in our ability to capture genetic data in high-throughput environments, the stability of genetic biomarkers through the lifetime of an individual, and the unidirectional vector from genetic exposure to phenotype. That is, the genetic method of inquiry is unique in ensuring that the exposure of interest (genotype) precedes phenotype and is not in turn affected by the disease process or its treatment. In addition, given that big data explorations involve thousands of variables and allow for a deluge of statistical tests, investigators have to be mindful of the statistical penalties incurred by multiple hypothesis testing and the preemptory need for

replication: At present and for various reasons, reproducibility is more tractable for genetic associations than for other data types.

While oncology pioneered the introduction of personalized molecular diagnosis and tailored treatment based on genetics, diabetes care has witnessed tremendous progress as well. Meta-analyses of genome-wide association studies and ongoing comprehensive sequencing experiments have yielded a plethora of genetic associations from an agnostic vantage point, which can open unsuspected windows into the pathophysiology of type 1 (10) and type 2 (11,12) diabetes. Whereas these genomic explorations have explained only a small fraction of the genetic contribution to the phenotype, in conjunction with physiological measures, they can be used to improve our nosology of the disease, and we can begin to characterize the clusters that may define specific subtypes (13).

In one illustration of this approach, investigators have begun to use genetic information to define a subtype of type 2 diabetes that resembles lipodystrophy, a clinical syndrome of insulin resistance caused by ectopic fat accumulation. A genetic risk score of all insulin resistance-associated variants was refined through a phenotypic clustering approach and yielded a subset of variants that were highly associated with higher triglycerides, lower HDL cholesterol, and greater hepatic steatosis, despite reduced adiposity; more significantly, the leaner individuals who carried a heavier genetic burden were more likely to develop type 2 diabetes or coronary artery disease (14). Thus, genetic knowledge can help define physiological subgroups of patients commonly classified under the broad umbrella of type 2 diabetes and select them for targeted preventive or therapeutic interventions.

On the other hand, except for *HLA* haplotypes in type 1 diabetes, the modest genetic effects observed thus far limit their widespread use as predictive tools. Clinical risk factors that are easily measured in the clinic, such as fasting glucose or BMI, contain the bulk of the predictive information for type 2 diabetes outcomes, especially when they begin to deviate from the metabolically normal state. In this context, the conjunction of genetic and metabolic

biomarkers adds only marginal predictive information in the short to medium term (15). Such predictors might be more useful in younger individuals who have not yet manifested metabolic deterioration (16).

One potentially useful application of these genetic predictors involves the combination of allele effects at multiple loci through genetic risk scores and their deployment in specific clinical scenarios where traditional phenotypic definitions do not fit a precise diagnosis. Because *HLA* haplotypes are highly predictive of type 1 diabetes, a genetic risk score for type 1 diabetes can help in the diagnosis of autoimmune diabetes in an obese young adult or exclude type 1 diabetes in a lean young adult who may have monogenic diabetes instead (17). Because the prevalence of type 2 diabetes is much higher than that of type 1 diabetes, tests for type 2 diabetes may achieve higher predictive value at lower sensitivities or specificities.

Although the use of genetic or other molecular biomarkers for disease prediction may be applicable to only a narrow niche, it might still be used for stratification of the population into subgroups who might benefit more from specific public health interventions, particularly where resources are limited. In this sense, genetic risk scores can predict which groups are at higher risk for type 2 diabetes; more significantly, intensive lifestyle modification is highly effective regardless of genetic burden, suggesting that preventive interventions might be deployed preferentially to people who harbor the highest risk, all other risk factors being equal (18,19).

The modest genetic effects observed for disease outcomes might be partially attributed to selection pressure by which deleterious variants with strong effects might not have been tolerated to rise to high frequencies through human evolution. These pressures might be less evident for pharmacological effects because the overwhelming majority of drugs have been introduced into human populations very recently. Thus, it might be possible to observe variants with stronger effects on pharmacological response. This is indeed the case for rare forms of diabetes, where the nature of the genetic defect is known and it can be overcome with targeted drugs. Thus, patients with type 3 maturity-onset

diabetes of the young are more sensitive to sulfonylureas than to metformin (20), and sulfonylureas are the drug of choice for neonatal diabetes due to mutations in the *ABCC8* or *KCNJ11* genes that encode the subunits of the sulfonylurea receptor and its associated potassium channel (21,22). In type 2 diabetes, it would be particularly useful to identify nonresponders to metformin a priori because a sizable number of patients do not achieve adequate glycemic control with metformin despite its universal adoption as a first-line agent in the treatment of type 2 diabetes (23). Genome-wide approaches combined with the electronic medical record have been launched (24), and preliminary reports of metformin ineffectiveness in preventing diabetes for certain variant carriers await confirmation (25).

INTEGRATION WITH OTHER DATA SOURCES

The use of big data to subphenotype diabetes is not confined to molecular biomarkers, however. Vast information captured in the electronic medical record can be used as well. A dedicated multidisciplinary team of bioinformaticians and

clinicians, working in a setting that enables this type of research, can leverage the mass of data points accrued in the course of patient care to begin to illuminate pathophysiology. This has been recently attempted at Mount Sinai Hospital in New York, where 11,210 patients with type 2 diabetes were studied for similarities through a topological approach that created data-driven patient-patient networks. Three separate clusters were identified with specific characteristics and outcomes; genetic polymorphisms associated with each cluster also demonstrated associations with the corresponding disease entities at the gene level (26). Although this represents a significant first step, three key questions emerge: 1) Where and how can these analyses be replicated, since the ability to refute a hypothesis is a crucial tenet of the scientific method; 2) Can the information gained be interpreted for physiological insight; and 3) How do these results support clinical decision making?

Indeed, one of the hardest challenges concerns the integration of disparate sources of information in a way that is

reproducible, interpretable, and actionable (Table 1). One example of such an effort related to diabetes has been recently published (27). By using a machine-learning algorithm, investigators in Israel created an algorithm that incorporates data on dietary habits, blood parameters, physical activity, anthropometrics, and gut microbiota to predict postprandial glycemic responses. Predictions were validated in an independent cohort, and application of these results to a new prospective clinical trial achieved improved glycemic outcomes. As results of this type accumulate, being able to devise simple and scalable clinical tools should allow for further testing and refinement, as well as lead to their wider penetration into patient care.

ACTION POINTS FOR THE FUTURE

In the end, a number of elements must coalesce for big data to deliver on precision medicine. First, appropriate data sources must be available at scale, whether it be from research cohorts, health care systems, or dedicated biobanks. Second, the method of data collection must be robust, with technologies that extract viable measurements with

Table 1—Challenges and progress in the implementation of precision medicine in diabetes

Challenge	Progress
The information captured on a given biological axis in an individual is often incomplete.	Comprehensive omics methods have been developed to query each biological axis in a global manner.
The information captured on a given biological axis in an individual is often static.	Longitudinal cohorts are being assembled where data exist along multiple time points (e.g., electronic medical records, biobanks).
The information captured on a given biological axis in an individual is often imprecise.	Technical accuracy and analytical quality control are improving the precision of biological measurements.
The enormous quantity of information available lends itself to data dredging and spurious findings.	Investigators, journals, and funding bodies are embracing the application of rigorous statistical standards to account for multiple hypothesis testing.
Effect sizes are too modest to be detected.	Sample sizes continue to increase through cross-institutional and international collaborations, as well as through the data sharing required by journals and funders.
The multiple dimensions of biological data typically reside in silos.	Methods and techniques for data integration across multiple dimensions are being designed, developed, and tested.
Tissues of relevance to diabetes are difficult to access.	Tissue banks, discarded clinical samples, dedicated cohort collections, commercial cell repositories, and newly developed inducible pluripotent stem cells are being harnessed to establish appropriate data sets.
Results from big data are seldom reproducible.	Investigators are beginning to share protocols, whereas top-tier journals require data sharing and higher burdens of proof.
Results from big data are seldom interpretable.	Physician-scientists, physiologists, and basic biologists are being drawn to collaborate on, and even generate, pertinent data sets.
Results from big data are seldom clinically useful.	Clinicians who are conversant in these methods are beginning to derive actionable physiological insight.
To leverage big data in precision medicine requires a multidisciplinary approach.	Most successful teams now involve bioinformaticians, computer scientists, statisticians, biologists, laboratory experts, physiologists, and clinicians.
The workforce is unprepared to assimilate, let alone act upon, big data.	Training venues, educational material, and discussion forums are beginning to proliferate as public interest mounts.

sufficient precision and accuracy at high throughput. Third, a multidisciplinary team with complementary expertise in bioinformatics, statistics, software engineering, quality control, biological sample processing, and clinical medicine must be engaged in a collaborative framework, preferably across jurisdictional and institutional boundaries. Fourth, expert panels that collect and assimilate the information generated by these efforts must be able to distill it into actionable interventions for patient care. Finally, a body of educators must be convened to train both their peers and the next generation of health care practitioners to disseminate and implement these interventions, which can then be iteratively evaluated for public health impact.

Such a vision is being realized in multiple settings, with varying degrees of completeness and penetration. In type 2 diabetes genomics, the initial steps in constructing such an edifice have been taken by the Accelerating Medicines Partnership, which is designing and building the Type 2 Diabetes Knowledge Portal (www.type2diabetesgenetics.org) with support from the National Institutes of Health, philanthropy, and the pharmaceutical industry. The diabetes field is ripe, and the unmet clinical need is clear. The critical question is whether we, as a community of investigators and practitioners who care deeply for our patients, are ready to meet the challenge.

Funding. J.C.F. is supported by a Massachusetts General Hospital Scholars Award.

Duality of Interest. J.C.F. has received a consulting honorarium from Sanofi. No other potential conflicts of interest relevant to this article were reported.

Prior Presentation. Parts of this article were presented at the 76th Scientific Sessions of the American Diabetes Association, New Orleans, LA, 10–14 June 2016.

References

- Tuomi T, Carlsson A, Li H, et al. Clinical and genetic characteristics of type 2 diabetes with and without GAD antibodies. *Diabetes* 1999;48:150–157
- Banerji MA, Chaiken RL, Huey H, et al. GAD antibody negative NIDDM in adult black subjects with diabetic ketoacidosis and increased frequency of human leukocyte antigen DR3 and DR4. *Flatbush diabetes*. *Diabetes* 1994;43:741–745
- Haaland WC, Scaduto DJ, Maldonado MR, et al. A-beta-subtype of ketosis-prone diabetes is not predominantly a monogenic diabetic syndrome. *Diabetes Care* 2009;32:873–877
- Gale EA. Latent autoimmune diabetes in adults: a guide for the perplexed. *Diabetologia* 2005;48:2195–2199
- Rolandsson O, Palmer JP. Latent autoimmune diabetes in adults (LADA) is dead: long live autoimmune diabetes! *Diabetologia* 2010;53:1250–1253
- Inzucchi SE, Bergenstal RM, Buse JB, et al. Management of hyperglycemia in type 2 diabetes, 2015: a patient-centered approach: update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* 2015;38:140–149
- Collins FS. Exceptional opportunities in medical science: a view from the National Institutes of Health. *JAMA* 2015;313:131–132
- McCarty CA, Chisholm RL, Chute CG, et al.; eMERGE Team. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics* 2011;4:13
- McCarthy MI. Genomic medicine at the heart of diabetes management. *Diabetologia* 2015;58:1725–1729
- Pociot F, Akolkar B, Concannon P, et al. Genetics of type 1 diabetes: what's next? *Diabetes* 2010;59:1561–1571
- Billings LK, Florez JC. The genetics of type 2 diabetes: what have we learned from GWAS? *Ann N Y Acad Sci* 2010;1212:59–77
- McCarthy MI. Genomics, type 2 diabetes, and obesity. *N Engl J Med* 2010;363:2339–2350
- Dimas AS, Lagou V, Barker A, et al.; MAGIC Investigators. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* 2014;63:2158–2171
- Yaghootkar H, Scott RA, White CC, et al. Genetic evidence for a normal-weight “metabolically obese” phenotype linking insulin resistance, hypertension, coronary artery disease, and type 2 diabetes. *Diabetes* 2014;63:4369–4377
- Walford GA, Porneala BC, Dauriz M, et al. Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes. *Diabetes Care* 2014;37:2508–2514
- de Miguel-Yanes JM, Shrader P, Pencina MJ, et al.; MAGIC Investigators; DIAGRAM+ Investigators. Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. *Diabetes Care* 2011;34:121–125
- Oram RA, Patel K, Hill A, et al. A type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. *Diabetes Care* 2016;39:337–344
- Hivert MF, Jablonski KA, Perreault L, et al.; DIAGRAM Consortium; Diabetes Prevention Program Research Group. Updated genetic score based on 34 confirmed type 2 diabetes Loci is associated with diabetes incidence and regression to normoglycemia in the diabetes prevention program. *Diabetes* 2011;60:1340–1348
- Hivert MF, Christophi CA, Franks PW, et al.; Diabetes Prevention Program Research Group. Lifestyle and metformin ameliorate insulin sensitivity independently of the genetic burden of established insulin resistance variants in Diabetes Prevention Program participants. *Diabetes* 2016;65:520–526
- Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT. Genetic cause of hyperglycaemia and response to treatment in diabetes. *Lancet* 2003;362:1275–1281
- Pearson ER, Flechtner I, Njølstad PR, et al.; Neonatal Diabetes International Collaborative Group. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N Engl J Med* 2006;355:467–477
- Babenko AP, Polak M, Cavé H, et al. Activating mutations in the *ABCC8* gene in neonatal diabetes mellitus. *N Engl J Med* 2006;355:456–466
- Kahn SE, Haffner SM, Heise MA, et al.; ADOPT Study Group. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. *N Engl J Med* 2006;355:2427–2443
- Zhou K, Bellenguez C, Spencer CC, et al.; GoDARTS and UKPDS Diabetes Pharmacogenetics Study Group; Wellcome Trust Case Control Consortium 2; MAGIC Investigators. Common variants near *ATM* are associated with glycemic response to metformin in type 2 diabetes. *Nat Genet* 2011;43:117–120
- Jablonski KA, McAteer JB, de Bakker PI, et al.; Diabetes Prevention Program Research Group. Common variants in 40 genes assessed for diabetes incidence and response to metformin and lifestyle intervention in the Diabetes Prevention Program. *Diabetes* 2010;59:2672–2681
- Li L, Cheng WY, Glicksberg BS, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci Transl Med* 2015;7:311ra174
- Zeevi D, Korem T, Zmora N, et al. Personalized nutrition by prediction of glycemic responses. *Cell* 2015;163:1079–1094