Perspective: a systems approach to diabetes research

Martin Kussmann1,2,3,*, Melissa J. Morine4,5, Jörg Hager1, Bernhard Sunderegger1 and Jim Kaput1,6

1 Nestlé Institute of Health Sciences SA, Lausanne, Switzerland
2 Faculty of Life Sciences, Ecole Polytechnique Fedérale, Lausanne, Switzerland
3 Faculty of Science, Aarhus University, Aarhus, Denmark
4 The Microsoft Research - University of Trento Centre for Computational and Systems Biology, Rovereto, Italy
5 Department of Mathematics, University of Trento, Trento, Italy
6 Endocrinology, Diabetes and Metabolism Division, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland

*Correspondence: Martin Kussmann, Nestlé Institute of Health Sciences SA, Campus EPFL, Quartier de l’innovation, Bâtiment H, CH 1015 Lausanne, Switzerland
Correspondence: martin.kussmann@nestle.com

We review here the status of human type 2 diabetes studies from a genetic, epidemiological, and clinical (intervention) perspective. Most studies limit analyses to one or a few omic technologies providing data of components of physiological processes. Since all chronic diseases are multifactorial and arise from complex interactions between genetic makeup and environment, type 2 diabetes mellitus (T2DM) is a collection of sub-phenotypes resulting in high fasting glucose. The underlying gene-environment interactions that produce these classes of T2DM are imperfectly characterized. Based on assessments of the complexity of T2DM, we propose a systems biology approach to advance the understanding of origin, onset, development, prevention, and treatment of this complex disease. This systems-based strategy is based on new study design principles and the integrated application of omics technologies: we pursue longitudinal studies in which each subject is analyzed at both homeostasis and after (healthy and safe) challenges. Each enrolled subject functions thereby as their own case and control and this design avoids assigning the subjects a priori to case and control groups based on limited phenotyping.

Analyses at different time points along this longitudinal investigation are performed with an integrated application of omics technologies: we pursue longitudinal studies in which each subject is analyzed at both homeostasis and after (healthy and safe) challenges. Each enrolled subject functions thereby as their own case and control and this design avoids assigning the subjects a priori to case and control groups based on limited phenotyping.

www.frontiernl.org

October 2013 | Volume 4 | Article 205 | #1

INTRODUCTION

Type 2 diabetes mellitus (T2DM) is a complex disease with epidemic proportions. The International Diabetes Federation estimates a total of 371 million type 2 diabetes worldwide1. China and India alone account for more than 150 million cases. Hence, there is a public health, economic, scientific, and ethical call for a proactive and preventive approach to the individual and public health burden caused by diabetes and its co-morbidities. Scientifically based preventive approaches should complement the reactive, pharmaceutical approach of management and treatment.

Type 2 diabetes mellitus is a multi-factorial condition that can already occur during gestation and has variable onset, severity, and outcome in juveniles, adults, and the elderly. Genetic (predisposition), epigenetic (developmental programming), and environmental factors (diet and physical activity) contribute to T2DM, with epigenetics being of far only being suggested from an epidemiological and animal model perspective. The complexity of the T2DM phenotype has challenged the fragmented scientific, clinical, and regulatory perspective. Most studies limit analyses to one or a few omic technologies providing data of components of physiological processes. Since all chronic diseases are multifactorial and arise from complex interactions between genetic makeup and environment, type 2 diabetes mellitus (T2DM) is a collection of sub-phenotypes resulting in high fasting glucose. The underlying gene-environment interactions that produce these classes of T2DM are imperfectly characterized. Based on assessments of the complexity of T2DM, we propose a systems biology approach to advance the understanding of origin, onset, development, prevention, and treatment of this complex disease. This systems-based strategy is based on new study design principles and the integrated application of omics technologies: we pursue longitudinal studies in which each subject is analyzed at both homeostasis and after (healthy and safe) challenges. Each enrolled subject functions thereby as their own case and control and this design avoids assigning the subjects a priori to case and control groups based on limited phenotyping.

We review here the status of human type 2 diabetes studies from a genetic, epidemiological, and clinical (intervention) perspective. Most studies limit analyses to one or a few omic technologies providing data of components of physiological processes. Since all chronic diseases are multifactorial and arise from complex interactions between genetic makeup and environment, type 2 diabetes mellitus (T2DM) is a collection of sub-phenotypes resulting in high fasting glucose. The underlying gene-environment interactions that produce these classes of T2DM are imperfectly characterized. Based on assessments of the complexity of T2DM, we propose a systems biology approach to advance the understanding of origin, onset, development, prevention, and treatment of this complex disease. This systems-based strategy is based on new study design principles and the integrated application of omics technologies: we pursue longitudinal studies in which each subject is analyzed at both homeostasis and after (healthy and safe) challenges. Each enrolled subject functions thereby as their own case and control and this design avoids assigning the subjects a priori to case and control groups based on limited phenotyping.

Although systems biology is considered an emerging paradigm for biological research, its roots can be traced back to von Bertalanffy (1951). Systems research typically refers to integrating combinations of high-dimensional (epi)genomic, transcriptomic, proteomic, or metabonomic data. Conceptual experimental approaches have been proposed for obesity, diabetes, and cardiovascular disease (MacLellan et al., 2012; Zhao et al., 2012; Meng et al., 2013). However, only a few multi-scale experimental results have been reported (Morine et al., 2010, 2011; Lau et al., 2012). Many of these examples analyze the physiological system – that is, the system is defined as processes occurring within the body. However, environmental factors such as nutrition, activity, rest/sleep, exposures to toxins and stress, and psychological factors, are known to influence an organism’s physiology, too.

Our definition of a “multi-scale systems” approach means deep characterization of subjects at omic and clinical level, and importantly – as many environmental factors as possible, and integration of these data (Figure 1) in interaction networks (including both functional interactions and statistical correlations). “Longitudinal” means that every study subject is assessed multiple times and ideally with challenges to homeostasis (van Ommen et al., 2009). These combined approaches differ from the standard methodology, in which single homeostatic “snapshots” are taken.
FIGURE 1 | Systems approach to analyses of complex traits. Systems research typically focuses on data measured (objects with dashed lines) from accessible tissues (adipose, muscle, occasionally other biopsies) and body fluids. Hence, many systems studies are analyzing a small subset of, e.g., the total metabolite complement. Since the measured molecules (proteins, RNAs, metabolites, etc.) are influenced by the expression of genomic information, DNA must also be characterized: the level and relationships between biomarkers are context-specific and the context here is the genes inherited from parents. DNA methylation, microRNAs, and chromatin modifying enzymes also change the expression of genetic information and can be measured in some tissues. These regulators are in turn regulated by environmental signals of nutrients and bioactives in food, and by a large number of physiological factors (sleep, ability to handle stress, joy and personal activity) and public environmental conditions (pollutants, allergens, allelochemicals from food, etc.). Many of these factors are not measured in systems research but they directly or indirectly can influence expression of genetic information. These metadata need to be captured in future experiments to understand the context of the physiological measurements and outcomes. The challenge of systems research is integrating these multi-scale data sets, which is still in its infancy.

between a priori defined groups of subjects, which are assumed to be matched for age, gender, genetic, socio-economic, or other confounders. These studies are furthermore often technology-driven, assuming that improvements in high-throughput genomics, transcriptomics, proteomics, or metabolomics suffice to correlate molecular signatures with disease phenotype. This goal has, however, been difficult to achieve, particularly in complex diseases, due in part to molecular and clinical heterogeneity between “cases.”

We do not conduct our human clinical studies by assigning subjects to groups based on a limited number of parameters. Rather, we propose a longitudinal study design, in which the same study subject is repeatedly monitored over time and individuals with like response are assigned to metabolic response groups only after multidimensional characterization. This approach uses each subject as their own control, challenging the classical case–control study design. Such an individual-focused longitudinal approach centers on analysis of trajectories of each measured variable over time.
Moreover, instead of studying a phenotype at homeostasis only, we expose subjects to metabolic challenges (e.g., oral glucose tolerance, high-fat meals or diets, physical exercise, or cognitive tests) and probe their metabolic elasticity, i.e., how rapidly or slowly they return to homeostasis after such safe, acute challenges. These repeated challenges and their omic read-outs of the body’s reactions can enable better definition of an individual’s health status and, potentially, their susceptibility to disease. A case in point: women who develop gestational diabetes during the stress of pregnancy are at a five-fold greater risk for developing T2DM; hence, metabolic responses to homeostatic challenges in healthy subjects may identify susceptibilities to later disease, which would derail their healthy trajectory before symptoms occur. These insights can in turn reveal early biomarkers for emerging metabolic disease and signatures for metabolic groups.

The integration of the comprehensive analyses of an individual’s genetic predisposition, epigenetic programming, and reaction at all omic levels may pose problems of too high dimensionality. However, we take a systems approach rather than a biochemical compound-class approach in the sense of completely mapping out key pathways and networks in diabetes (i.e., identifying and quantifying all enzymes, nutrients, and metabolites in those pathways). We term this “systems omics” as opposed to “classical” proteomics/metabonomics/lipidomics.

HUMAN EP(IGENETIC) INDIVIDUALITY WITH REGARD TO T2DM

Many genetics-oriented studies of diabetes have not taken into account any information on an individual’s environment, such as nutrition, physical activity, or lifestyle; typically, genome-wide association studies (GWAS) in cohorts derived from diabetic and healthy populations have identified multiple genetic, population-attributable (fraction) risk loci. These markers are usually single-nucleotide polymorphisms (SNPs), which – in combination – are expected to predict disease risk in individuals. However, they explain only a fraction of the disease risk, partly because of the mis-perceived definition of “population-attributable fraction” (PAF), which is the decrease in disease incidence if the allele were not present in that population (Rockhill et al., 1998; Kaput, 2008). This decrease cannot be generalized to other populations. Myles et al., for example, showed that allele frequencies of a subset of 25 GWAS-identified risk alleles were not consistent with contributions to disease incidence in Africans or Asians (Myles et al., 2008). Others have shown that different pathways that alter glucose levels (i.e., insulin sensitivity, gluconeogenesis, insulin secretion) differ between ancestral groups (Missio and Vlţă, 2004). This is important from a systems perspective, as it suggests that a given genotype can arise from distinct perturbations of system function, and this puts into question the assumption of many GWAS studies that all cases are identical from a patho-physiological viewpoint.

In addition, human populations are not at equilibrium with exponential population growth occurring only during the last 150 years. The consequence of this rapid expansion is that more recent and rare variants exist than assumed by the common variant/common disease model, which is the theoretical basis of the GWAS experimental designs (Cargill and Daley, 2000; Hirschhorn and Daly, 2005; Iyengar and Elston, 2007; Fu et al., 2013). Hence, the sample size for a representative proportion of the population roughly equals the effective population size.

Gene–diet interactions have been analyzed in humans, but typically in a single-gene versus single-nutrient design, where a complex phenotype (such as response to nutrient) is thought to be caused by a single variant in one gene out of a total of ~20,000 human genes. Although selected gene–diet associations have been replicated in multiple populations, e.g., in the study “Association between the APOA2 promoter polymorphism and body weight in Mediterranean and Asian populations: replication of a gene-saturated fat interaction” (Corella et al., 2011), the effect size of these associations is known to explain only a small percentage of the overall phenotype. Alternatively, classical nutritional epidemiology studies associating a nutrient to a phenotype were performed in clinical cohorts or in entire (sub)populations, and these studies neglected the genetic contributions to the phenotype. This design is based on the assumption of finding the statistical main effect of the diet or lifestyle.

It is well established that genetic factors contribute significantly to T2DM. Estimates of heritability for T2D range from 50% to 80% and are much higher than for T1D. However, these estimates rely on the assumption that environmental exposures are equal for those that are affected and those that are not. This may not be true, even when comparing monzygotic to dizygotic twins (Loos et al., 2001; Ollikainen et al., 2010). Studies of maturity onset diabetes of the young (MODY), a highly penetrant monogenic form of type 2 diabetes, have proven that single gene mutations can lead to T2DM (Johansson et al., 2012; Mitchell, 2012), although the age of onset, severity, outcome, and age of death demonstrate the importance of modifying genes. Many SNPs affect gene expression or protein function more subtly than loss-of-function mutations. While not explaining most of the estimated heritability, GWAS have provided new insights into the etiology of T2DM. One striking and unexpected outcome of genetic studies of both MODY and late onset T2DM is that most genes identified so far are involved in insulin secretion. A caveat though is that over 90% of GWAS were conducted in Europeans and the findings cannot be generalized to other populations. Myles et al., for example, showed that allele frequencies of a subset of 25 GWAS-identified risk alleles were not consistent with contributions to disease incidence in Africans or Asians (Myles et al., 2008). These insights are at a five-fold greater risk for developing T2DM; hence, metabolic responses to homeostatic challenges in healthy subjects may identify susceptibilities to later disease, which would derail their healthy trajectory before symptoms occur. These insights can in turn reveal early biomarkers for emerging metabolic disease and signatures for metabolic groups.

The integration of the comprehensive analyses of an individual’s genetic predisposition, epigenetic programming, and reaction at all omic levels may pose problems of too high dimensionality. However, we take a systems approach rather than a biochemical compound-class approach in the sense of completely mapping out key pathways and networks in diabetes (i.e., identifying and quantifying all enzymes, nutrients, and metabolites in those pathways). We term this “systems omics” as opposed to “classical” proteomics/metabonomics/lipidomics.

**HUMAN EP(IGENETIC) INDIVIDUALITY WITH REGARD TO T2DM**

Many genetics-oriented studies of diabetes have not taken into account any information on an individual’s environment, such as nutrition, physical activity, or lifestyle; typically, genome-wide association studies (GWAS) in cohorts derived from diabetic and healthy populations have identified multiple genetic, population-attributable (fraction) risk loci. These markers are usually single-nucleotide polymorphisms (SNPs), which – in combination – are expected to predict disease risk in individuals. However, they explain only a fraction of the disease risk, partly because of the mis-perceived definition of “population-attributable fraction” (PAF), which is the decrease in disease incidence if the allele were not present in that population (Rockhill et al., 1998; Kaput, 2008). Since GWAS is determined by the average occurrence of a SNP between cases and controls, the risk allele is actually a PAF and not the personal or individual risk factor. In fact, family history is a better predictor of diabetes risk than SNPs identified by GWAS (Miegs et al., 2008).

A number of authors have identified limitations in the GWAS designs (McCarthy et al., 2008; Goldstein, 2009; Mansolillo et al., 2009; Neel and Goldstein, 2010):

- Poorly defined phenotypes.
- Indirect scoring of causal SNPs by distant gene markers.
- Challenge to find rare alleles.
- Insufficient consideration of gene × gene interactions (epistasis).
- Insufficient consideration of gene × environment interactions.
- Copy number variants.
- Epigenetic factors.

In addition, human populations are at a five-fold greater risk for developing T2DM; hence, metabolic responses to homeostatic challenges in healthy subjects may identify susceptibilities to later disease, which would derail their healthy trajectory before symptoms occur. These insights can in turn reveal early biomarkers for emerging metabolic disease and signatures for metabolic groups.

The integration of the comprehensive analyses of an individual’s genetic predisposition, epigenetic programming, and reaction at all omic levels may pose problems of too high dimensionality. However, we take a systems approach rather than a biochemical compound-class approach in the sense of completely mapping out key pathways and networks in diabetes (i.e., identifying and quantifying all enzymes, nutrients, and metabolites in those pathways). We term this “systems omics” as opposed to “classical” proteomics/metabonomics/lipidomics.
Nevertheless, the fact that so far most genes identified by GWAS are involved in insulin secretion may shed new light on the role of obesity and insulin resistance in T2DM etiology. Indeed, one may speculate that T2DM (like type 1) is fundamentally a β-cell disease and obesity-related insulin resistance is a secondary risk factor, which may push “at-risk” individuals across the disease threshold. The knowledge gap is whether genetic variants involved in insulin secretion independently increase the risk of T2DM in addition to the physiological effects caused by high body mass index (BMI).

If we want to prevent T2DM, we need to disentangle the causal variations and environmental factors. To do this we need new approaches to studying genetic variation that do not rely on a design, in which type 2 diabetics with different causal variants are compared to controls, because individuals classified as T2DM and control have highly variable phenotypes. The future and the success of genetic studies of T2DM will not be solved by ever bigger association studies, but rather in well-defined studies enrolling carefully phenotyped subjects, who should be exposed to a (dietary or other) external challenge to reveal their individual responses. New sequencing technologies are expected to identify additional or other external challenge to reveal their individual responses. New sequencing technologies are expected to identify additional rare variants with somewhat higher effect sizes. In addition to this revised study design, data acquisition, processing, and interpretation also need to be revisited: all molecular and clinical data need to be regarded as a single input into a systems approach to eventuall unravel those mechanisms that are the most important risk factors for a given group of individuals. Only then can we design effective interventions, either nutritional, life-style or drug-based, to prevent or at least delay the development of this disease.

**HUMAN STUDIES ON T2DM – TODAY AND TOMORROW**

Many, if not most, human studies have been designed in a case–control design with the most important consideration being the power of the experiment to detect the null hypothesis (Bader, 2001; Longmate, 2001; Lee and Whitmore, 2002; Teo, 2008; Visscher et al., 2008). An unintended consequence has been an increasing misclassification: adding subjects to case or control groups based on simplistic phenotypic characterization. For example, most population-based studies use fasting plasma glucose levels or other homeostatic measures associated with T2DM. However, elevated fasting glucose levels could be produced by decreased insulin production, increased gluconeogenesis, insulin insensitivity, or increased glucose uptake in the intestine (Kaput et al., 2007; Malecki and Skupien, 2008). Individuals differ in these pathways and may have different genetic variants that contribute to these glucose abnormalities. Grouping all subtypes of T2D therefore decreases the signal and increases noise in these case–control analyses. The failure to account for environment–gene interactions also confounded GWAS analyses.

Two related approaches may improve study designs and both require more comprehensive phenotyping before analyzing genetic differences. The first approach is to compare genetic makeup of individuals at the extremes of a given phenotype (Ahituv et al., 2007; Kaput, 2008; Khor and Goh, 2010). This strategy may work for highly differentiated phenotypes, for example, tall versus short stature (Letteri, 2009) or the subset of metabolic phenotypes that can be clearly discriminated by homeostatic measurements. The second approach uses short-term (acute) or long-term (chronic) challenges to homeostatic systems or, ideally, a combination of them. The oral glucose tolerance test is used for assessing glucose regulation and a prime example of such acute challenges. Other nutritional and functional challenges (e.g., exercise) may also be used for assessing metabolic robustness (van Ommen et al., 2009). Results from such challenge studies using extensive omics, similar to the approach described in this article, have been published (Wisper et al., 2009; Heinzenmann et al., 2012) and confirm the large intra-individual variability observed at homeostasis. However, small sample sizes limited the ability to cluster individuals with similar responses.

O’Sullivan and coworkers, for example, could group individuals based on biomarker patterns following long-term vitamin D supplementation (O’Sullivan et al., 2011). We propose here to combine acute challenges of homeostasis, such as oral glucose tolerance, with longitudinal analyses, ideally following defined nutritional or activity interventions which alter physical condition, such as weight loss or improved cardiovascular health. The response to single homeostatic challenges is influenced by the current physiology of the individual: to illustrate, O2 consumption measured at homeostasis or in acute challenges (e.g., upon a single exercise bout) will likely differ before and after months of athletic training. So, which time point and physiology accurately measures an individual’s aerobic capacity robustness?

The study design and execution suggested here, including extensive omics at homeostasis and after acute challenges, measured longitudinally, increases both cost and complexity of assessing health and disease. The current model of funding and conducting biomedical research was designed in the pre-genomic era. The average research grant in the USA was US$ 482,276 in 2012 and most grants are 3 year awards, with resources too small to conduct multi-scale scientific research3. While smaller scale science funding must be maintained, the complexity of biological processes requires complex analysis that can only be done through extensive collaborations in consortia. Many examples of these consortia now exist: the Human Genome (Lander et al., 2001), the International HapMap (International HapMap Consortium, 2003), the ENCODE (Birney et al., 2007), and the 1000Genomes (The 1000 Genomes Project Consortium) projects (organized in the United States); and many of the Framework projects in Europe such as NuGO4, Food4Me5. The paradigm of funding and conducting biomedical research is changing.

Human studies of systemic, metabolic conditions like T2DM must not be restricted to the assessment of the human host only: humans carry ∼100 times more bacterial than human cells in and on their body, colonizing mucosal surfaces (e.g., gut, oral cavity, vagina) and the skin (Uwins et al., 2012). The intestinal microbiota are the most complex of these bacterial communities (Dimitrov, 2011) and centrally involved in host metabolic (Harris et al., 2012) and immune health (Hakansson and Molin, 2011). We therefore proposed in an earlier perspective published in this journal an “extended nutrigenomics” approach encompassing the

---

3 http://www.nugo.org
4 http://www.food4me.org
5 http://report.nih.gov/findings.aspx/index.cfm
host, microbial and food genomes to better understand gene-diet interactions in humans (Kussmann and Van Bladeren, 2011). Today’s insights into host-microbiome relationships derive from large-scale (earlier 16 sRNA sequencing, nowadays deep sequencing) studies and mainly reflect associations between a microbiota population census on the one hand and a host condition on the other hand. It has for example been shown that human subjects with different metabolic conditions (e.g., diabetes, Larsen et al., 2010 or obesity, Turnbaugh et al., 2006) have different gut microbiota. Next steps have recently been taken toward establishing a more causal (Knapp et al., 2015) than purely associative relationship (Cox et al., 2013): the biochemical activities of such gut microbial populations have been probed by metabonomic means and so-called “core” human microorganisms have been characterized with metabolic functions generalizable between individuals (Huse et al., 2012; Li et al., 2013). This may pave the way for tailoring diets to metabolic groups of consumers and patients to induce or maintain a favorable microbiome (Nicholson et al., 2005; Wu et al., 2011). Possible causal relationships between gut microbes and host health need to be considered from both perspectives: microbes possibly altering host phenotype and host factors possibly determining microbial composition. Yokota et al. (2012) took it from the host end asking whether bile acid is a determinant of the gut microbiota on a high-fat diet and discussing how the identification of host factors determining the gut microbiota can contribute to understanding the causality between changes in gut microbiota and disease development. Vrieze et al. (2015) even took the ultimate step from causality to possible treatment considering fecal transplantation as a clinical therapy for restoring intestinal microbial balance in human disease.

OMICS MONITORING OF HUMAN STUDIES ON T2DM – TODAY AND TOMORROW

Omic studies in nutrition and diabetes have typically been performed in a technology-driven, rather than a technology-rooted manner. While the advent of high-throughput and comprehensive genomics and other omic platforms have advanced biological knowledge, the emphasis on single technology-driven nutritional and biomedical research projects still limits assessing health and disease conditions and trajectories: there are a number of investigations that attempt to find either genetic (Billings and Florez, 2007; Gluckman et al., 2009; Low et al., 2011). Epigenetic investigations that attempt to find either genetic (Billings and Florez, 2007; Gluckman et al., 2009; Low et al., 2011) or metabonomic (Bao et al., 2010), or epigenetic (Pinney and Simmons, 2010), or proteomic (Rao et al., 2009; Riaz et al., 2010a,b) or metabonomic (Bao et al., 2009; Hsu et al., 2009; Griffin et al., 2011) biomarkers for T2DM. Our suggestion is that cross-platform approaches are needed to truly interrogate physiological processes.

While genetic markers can inform on inborn predisposition and susceptibility, epigenomics can reveal those genetic marks altered by the environment, particularly diet (Waterland and Jirtle, 2003, 2004; Gluckman et al., 2009; Low et al., 2011). Epigenetic modifications increasingly appear to deliver the molecular basis for the intuitive observation that environment shapes phenotype during lifespan and even across generations (Morgan and Whitelaw, 2008; Morris, 2009). However, epigenetic marks do not only affect DNA bases (methylation) but also the “DNA packaging” proteins such as histones (Jenuwein and Allis, 2001), other chromosomal components that regulate transcription, and small RNAs (Bei et al., 2007; Choudhuri, 2011). Despite recent developments of proteomics in deciphering the histone code (Trelle and Jensen, 2007; Sidoli et al., 2012) and the epigenetic interplay between DNA- and histone modifications, there are few studies that integrate proteomics and DNA sequencing to generate a more complete picture of epigenetic marks that affect gene expression and phenotype after environmental exposure (Fukas et al., 2003; Johnson et al., 2007).

A body of pure proteome approaches is emerging that analyzes energy metabolism-relevant tissues across different conditions (pancreas, Chen et al., 2007ab; Kim et al., 2008; muscle, Wang et al., 2010; Giebelstein et al., 2012; adipose tissue, Ahmed et al., 2010; Perez-Perez et al., 2012; pancreatic beta cells, Song et al., 2009; Maris et al., 2011; and mitochondria, Forner et al., 2006; Chen et al., 2012). These tissues were typically isolated from normal (control) and diabetic (case) human subjects, mouse strains or genetic mouse models. Potential protein/peptide biomarker signatures are being associated by analyzing blood or urine, in order to facilitate less invasive sampling.

Metabonomic studies in T2DM can be classified in those based on (proton) nuclear magnetic resonance (1H-NMR) spectroscopy (Rezzi et al., 2007; Lindon and Nicholson, 2008; Collino et al., 2009), and those deploying mass spectrometry (MS; Toyokôka, 2008; Kettse et al., 2009; Muchur and Rea, 2012). NMR-enabled metabonomic profiling in the context of diabetes was typically done on human urine (Kussmann et al., 2006; Faber et al., 2007; Maher et al., 2009; Zhao et al., 2010). The intrinsic NMR advantages like high-throughput, robustness, and minimal sample preparation are now being complemented by the higher sensitivity and greater structure elucidation power of MS (Lenz et al., 2005; Kussmann et al., 2008). However, investigations integrating metabolomics with proteomics or even genetics or epigenetics have remained the exceptions (Holmes et al., 2001; Vlasi et al., 2007; Hardiman, 2011).

Current proteomic discovery platforms typically span a dynamic range of 10^6 to 10^8 whereas the protein concentration range in human blood plasma is estimated to be 10^6 (Lescuyer et al., 2007; Surinova et al., 2011). Similarly, current metabolomic workflows cover a range of lipophilic and hydrophilic metabolites (Whiteley et al., 2012); carbohydrates (Soo and Hui, 2010), amino acids and ketones (Suhre et al., 2010); and the complexity of lipid biochemistry has resulted in lipidomics as a specialized new omics discipline (Eising et al., 2009; Shevchenko and Simons, 2010). These proteomic and metabonomic studies are typically designed from a biochemical, compound-class perspective.

Complementary to the aforementioned study designs we propose here “systems omics,” i.e., to join forces of transcriptomics, proteomics, metabolomics, and lipidomics to completely map out – that is identify and quantify all (micro-)nutrients, metabolic enzymes and their substrates and products – in a given biochemical pathway. We thereby aim at deeply characterizing the dynamics of those pathways that have emerged as consistently (de-)regulated in a specific health condition and under relevant environmental – such as dietary – influences.

The value of a systems approach was recently demonstrated by Jain et al. (2013), who described the molecular architecture of T2DM pathophysiology using multi-omics data, combined
with physical and genetic interaction networks. Their work was largely motivated by the observation that, despite the symptomatic complexity of T2DM (incl. alteration of immune function, oxidative stress, and nutrient metabolism), the majority of GWAS to date have failed to highlight these processes as genetic determinants of T2DM onset/progression. Starting with publically available T2DM GWAS data, the authors demonstrated that the only significantly enriched pathway with T2DM SNPs was that of MODP. By extending these T2DM SNPs to their first-degree neighbors in physical and genetic interaction networks (creating a "T2DM interactome"), and incorporating gene expression data from multiple tissues in healthy and diabetic individuals, they identified a range of pathways with known relevance to T2DM. In particular, the TGFβ signaling pathway contained multiple genes present in the GWA, interactome, and transcriptionomic data sets. Jain et al. (2013) showed that a multi-omics approach can reveal disease-related processes that may not be evident in a single omics data set, and can also aid biological interpretation of GWA data.

Based on the components identified in key T2DM-related pathways, we propose that MS assays for all nutrients, enzymes, and metabolites in these contexts be developed and applied. These assays are based on single- and multiple reaction monitoring (SRM and MRM) of target molecules, so called MS-based enzyme-linked immunosorbent assays (ELISAs; Maiolica et al., 2012), which use sequencing of specific peptides that are unique surrogates for their parent protein. For measuring metabolites in this "systems omics" strategy, intact mass-based metabolite identification and quantification (the latter requiring either isotope-labeled or unlabeled internal standards that are either chemically identical or highly similar to the target analytes) is of high interest. Concomitantly, high-resolution and high-mass accuracy MS have become the preferred technology platforms for global, quantitative metabolite profiling. Targeted SRM and MRM methods that can filter, detect, identify, and quantify selected metabolites against complex biochemical matrices are powerful platforms in clinical metabolomics, particularly in systems omics studies proposed here.

Omics integration is one way to address the spatial and temporal complexity of type 2 diabetes, being a multi-organ disease with multiple and interrelated contributing dysfunctions, rather than "just" a beta cell disorder. A systems, low-grade, chronic inflammation drives metabolic deteriorations in several organs in T2DM. This raises the question of both meaningful and feasible sampling of body tissues and fluids and how one could extrapolate from intra-cellular networks to a systemic multi-organ disease. Repeated sampling of tissue biopsies during a longitudinal study is not feasible, because key organs for metabolic function like pancreas, liver or gut are not only accessible by highly invasive means. We are therefore restricted to the periphery for sampling, i.e., blood, urine, and stool. Blood and urine omics have become standard approaches to assess protein, peptide (Crosley et al., 2009), lipid (Song et al., 2013), and metabolite (Guy et al., 2009) profiles. These profiles become more meaningful when sampled in the same individual over time and when metabolic/physical challenges are compared to the resting condition, as we propose here. Moreover, biomolecular blood profiles can inform about homeostasis (or deviation thereof) whereas urine profiles can indicate the body’s measures (e.g., secretion) to achieve or restore this homeostasis. White and red blood cells have unique, cell-specific metabolisms and may inform about metabolic or immune status if analyzed in conjunction with hematological analysis of cell types (platelets, eosinophils, other white blood cell types; Crosley et al., 2009). Stool samples have become the non-invasive choice to access the gut microbiota (see previous section; Flores et al., 2012a,b).

Hence, only repeated, peripheral, non-invasive sampling over time and across challenges is both meaningful and feasible to generate systems-level insights into a multi-organ disease like T2DM in humans.

**RECONSTRUCTING THE SYSTEM OF T2DM WITH NETWORK ANALYSIS**

With exponential reduction in cost of data generation, processing and managing, it has become more common to measure multiple types of omics data in a given study. However, translation of these omics data sets into concrete knowledge of biological systems is contingent on informative methods for multi-omics analysis. The simplest approach to analysis of, for instance, a transcriptomic and proteomic data set, would be to generate lists of genes and proteins that respond to a given perturbation or correlate with a given phenotype, then attempt to build a biological interpretation post hoc. However, this analytical approach does not capture the functional interactions within and between the different classes of molecular species, which ultimately coalesce to form the biological system of interest.

Biological network analysis offers a natural framework for analysis and interpretation of omics data sets. This said, pathway analysis is currently more widely adopted than network analysis, owing in part to strong methodological development and implementation, as well as straightforward biological interpretability. The limitation of such pathway analysis is that the pathway models, by design, do not capture the intersection of these pathways that create the entire system. While global interaction networks inherently capture these intersections, analysis of these networks is uniquely challenging due to their considerable size. The I2D database, for example, contains 173,338 human protein interactions as of February 2013. A practical consideration in the analysis of such data sets is the presence of false-positive – which are pervasive in online interaction databases – or otherwise irrelevant interactions. Therefore, utilization of a global interaction data set in the context of an omic analysis often requires manual curation to remove, possibly, interactions inferred by homology, or interactions that do not physically occur in a tissue of interest.

Once such a framework network is identified, the next challenge is to identify, which sub-regions (i.e., collections of nodes) significantly respond to a given perturbation, or correlate with a given phenotype. A range of algorithms have been developed for this purpose, including heuristic (Liu et al., 2007) and exact (Qiu et al., 2009) community detection approaches, as well as network path tracing (Morine et al., 2011).
Similar to (and possibly driven by) the to-date trend of generating and analyzing a single omic data type, molecular interaction networks (as stored in interaction databases such as MetaCyc, IntAct, etc.) commonly represent a single type of interaction – for instance, metabolic, protein-protein, or regulatory interaction. However, it is intuitive that these classes of interactions do not occur in isolation: metabolic reactions produce metabolites, which may regulate signaling cascades, which in turn may activate transcription regulation. Reconstruction, curation, and analysis of such multi-level networks will be instrumental in understanding complex, multi-factorial diseases such as T2DM from a systems standpoint.

A notable example of this type of multi-level network is the Biological Expression Language (BEL) framework, which includes a knowledge base of curated, directed interactions and a dedicated language (i.e., a syntax and semantics). The BEL framework contains information-rich interactions within and between different classes of biological components. Each interaction is categorized by type (e.g., binding, reaction, translocation, etc.) and includes a line of “evidence” text parsed from the associated publication, describing the context of the interaction. As an example, Figure 2 illustrates the BEL network using a list of T2DM-related molecular processes (based on a node keyword search for “insulin secretion,” “insulin receptor signaling,” “glucose transport,” “glucose import,” or “glucose homeostasis”) as seed nodes, and then extending to the first-degree neighbors. Apart from demonstrating the complexity of interactions related to T2DM, this sub-network illustrates the range of biological molecule types that are causally associated with the diabetic phenotype.

Knowledge bases such as BEL demand a great deal of effort to be curated and would therefore benefit from more scientific community contributions in terms of reporting identified molecular interactions. Public transcriptomic databases have benefited considerably by the common editorial requirement for deposition of transcriptomic data into databases such as Gene Expression Omnibus (GEO) and ArrayExpress. A similar

Abbreviations in the network nodes are as follows: CHEBI, Chemicals of Biological Interest names (http://www.ebi.ac.uk/chebi/); EGID, Enzyme Gene IDs (http://www.ncbi.nlm.nih.gov/gene); GO, Gene Ontology names (http://www.geneontology.org/); HGNC, Human Genome Nomenclature Committee (http://www.genenames.org/); MESH, Medical Subject Heading Disease names (http://www.nlm.nih.gov/pubs/factsheets/mesh.html); MESHCL, Medical Subject Heading Cellular Structure (http://bioportal.bioontology.org/ontologies/46835?p=terms&conceptid=D022085); MG, Mouse Genome Information gene symbols (http://www.informatics.jax.org/); NCH, Human Molecular Complex names (http://www.ncbi.nlm.nih.gov/books/NBK7576/); PFH, Human Protein Family names (http://mordred.bioc.cam.ac.uk/˜fpspd/Database/nombres.html); PFM, Mouse Protein Family names (http://www.sanger.ac.uk/resources/database/pfam.html); RGD, Rat Genome Database gene symbols (http://rgd.mcw.edu/); SPAC, SwissProt accession numbers (http://www.uniprot.org).
We suggest here an integrated systems approach of longitudinal multi-omics challenge studies to reveal early molecular signs of diabetes. We argue that this strategy will identify better biomarkers and diagnostics that would in turn enable personalized intervention through tailored diets prior to onset of chronic disease. These interventions would consist of (micro)nutrients and other functional ingredients that would be used for disease prevention rather than disease management or cure. Having defined this as one of the key objectives in diabetes research, we acknowledge at the same time the challenge of completely mapping out a complex human tissue or body fluid at proteomic or metabolomic level.

**REFERENCES**

The 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65. doi: 10.1038/nature11515

Ahituv, N., Kasneci, N., Schadt, W., U察awo, A., Martin, J., Hubert, S., et al. (2007). Medical sequencing at the extremes of human body mass. Am. J. Hum. Genet. 80, 779–791. doi:10.1086/515471

Ahmed, M., Neville, M. J., Edelman, M. J., Keeler, B. M., and Karp, F. (2010). Proteomic analysis of human adipose tissue after repositioning treatment shows coordinated changes to promote glucose uptake. Obesity (Silver Spring) 18, 27–34. doi:10.1038/oby.2009.208

Badger, J. S. (2001). The relative hypoglycemic and insulinotropic effects of different meal plans on hypoglycemia in healthy individuals. Diabetologia 44, 20–27. doi:10.1007/s001250000649

Birney, E., Stamatoyannopoulos, J., Andrews, B., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., et al. (2007). An integrated map of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816. doi:10.1038/nature05874

Boursot, P. (2002). Quantitative proteomic analysis of platelet, peripheral blood mononuclear cell, plasma, urine and saliva proteins. Genome Biol. 4, 95–102. doi:10.1186/gb-2003-4-2-95

Chen, R., Bresnall, T. A., Pan, S., Cooke, K., Morey, K. W., Lana, J., et al. (2007a). Quantitative proteomics analysis reveals that proteins differentially expressed in chronic pancreatitis are also frequently involved in pancreatic cancer. Mol. Cell. Proteomics 6, 1535–1542. doi:10.1074/mcp.M600377-MCP210

Chen, B., Pan, S., Cooke, K. M., Morey, K. W., Brenner, M. P., Goodlett, D. R., et al. (2007b). Comparison of pancreatic protein extracts from cancer versus pancreas using quantitative proteomic analysis. Proteomics 7, 70–79. doi:10.1002/pmic.200600041

Chen, X., Wei, S., and Yang, F. (2012). Mitochondria in the pathogenesis of diabetes: a proteomic view. Proteomics 12, 448–465. doi:10.1002/pmic.201100245

Choudhuri, S. (2011). From Waddington’s epigenetic landscape to small noncoding RNA: some important milestones in the history of epigenetics research. Environ. Microbiol. 13, 252–274. doi:10.1111/j.1462-2920.2011.02380.x

Collins, S., Martin, F. F., Kochhar, S., and Reznik, S. (2007). Monitoring healthy metabolic trajectories with nutritional genomics. Nutrigenetics 1, 101–110. doi:10.3904/nig3010101

Cordua, T., Tai, E. S., Sofi, F., Vecchi, S. K., Cikola, O., Seton-Peirce, M., et al. (2011). Association between the API2AZ promoter polymorphism and obesity and glucose homeostasis in European and Asian populations: replication of a gene-environmental interaction. Int. J. Obes. (Lond.) 35, 665–675. doi:10.1038/ijo.2010.107

Danford, K. A., Mühling, N., Pollen, A. C., Bowenman, F. G., Heim, C., Mulholland, F., et al. (2009). Variation in protein levels obtained from human blood cells and biofluids for platelet, peripheral blood mononuclear cell, plasma, urine and saliva samples. Genet. Nutr. 4, 95–102. doi:10.1186/1475-2840-4-12

Dimitrov, D. V. (2011). The human gut microbiome genetics of the host–microbe interactions. OMICS 15, 419–435. doi:10.1089/omi.2011.0109

Eising, G. C., symposium, J. L., Zendejas-Balbuena, V., Duchoslav, E., Elkouzi, K., Klemm, M. W., et al. (2009). Global analysis of the yeast liposome by quantitative shotgun mass spectrometry. Proc. Natl. Acad. Sci. U.S.A. 106, 2156–2164. doi:10.1073/pnas.0810701106

Fahim, J. H., Malhotra, D., Tohi, H., Maher, A. D., Creekfield, D., Holmes, E., et al. (2007). Metabolomics in diabetes research. J. Diabetes Sci. Technol. 1, 549–557. doi:10.1530/JS-01-120

Flores, R., Shi, J., Gail, M. H., Ravell, J., and Goedeker, J. (2012a). Assessment of the human fecal microbiota I: Measurement and reproducibility of selection enzyme activities. J. Clin. Invest. 122, 188–193. doi:10.1172/JCI52958

Forner, F., Foster, L. J., Campmans, S., Valka, G., and Mann, M. (2006). Quantitative proteomic comparison of rat mitochondria from mouse, rat, and human. Cell. Proteomics 5, 608–619. doi:10.1016/j.cplett.2005.11.034

Fu, W., O’Connor, T. D., Jun, G., Kang, H. M., Abecevic, G., Leal, S. M., et al. (2011). Analysis of 6,515 exomes reveals the recent origin of the human APOA2 promoter polymorphism. Proc. Natl. Acad. Sci. U.S.A. 108, 2222–2226. doi:10.1073/pnas.1011190108

Fukui, H., Hard, P. D., Wolf, D., Nan, X., Red, A. P., and Konturicz, T. (2003). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone modification. J. Biol. Chem. 278, 4055–4060. doi:10.1074/jbc.M110126200

Hallerstein, J., Poehlmann, G., Huykend, K., Scheldinger, W., Dietrich, J. W., Levin, K., et al. (2012). The proteomic signature of insulin-resistant human skeletal muscle reveals increased glycolytic and decreased mitochondrial enzymes. Diabetologia 55, 1114–1127. doi:10.1007/s00125-012-2456-x

Gluckman, P. D., Hanson, M. A., Bukkens, T., Lee, P. M., and Baddele, A. S. (2009). Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. Nat. Rev. Endocr. Metab. 5, 401–408. doi:10.1038/nrendo.2009.102

Goldstein, D. B. (2009). Common genetic variation and human traits. Nat. Rev. 10, 1066–1068. doi:10.1038/nr.2009.264

Griffis, J. L., Atkinson, H. J., Steckel, C., and Salim, B. M. (2011). A metaphase description of most of the data in “A metabolomic comparison of urinary changes in type 2 diabetes in obese, morbid and healthy volunteers.” J. Proteome Res. 10.1111/j.1365-2362.2012.02660.x

Gay, P. A., Tavami, I., Brusa, S. J., Ramadani, Z., and Kochhar, S. (2008). Global metabolic profiling analysis on human urine by UPLC-TOFMS issues and method validation in nutritional metabolomics. J. Chromatogr. B Analyt. Methods Biomed. Life Sci. 872, 255–261. doi:10.1016/j.jchromb.2008.04.034

Hakulinen, A., and Möllig, G. (2011). Gut microbiota and inflammation. Nutrients 3, 648–660. doi:10.3390/nu3030657

Harbison, C. R., Antoch, Zohib, Raimondi, Giuseppe, J. W., Levin, K., et al. (2012). The genetic basis of metabolic individuality in humans. Nature genetics 42, 1507–1518. doi:10.1038/ng.2699

Harrison, K., Kasner, A., Moyer, G., and Chrousos, C. J. (2012). Is the gut microbiota a new factor contributing to obesity and its metabolic disorders? J. Obes. 2012, 879151. doi:10.1155/2012/879151

Hausmann, S. M., Mitchell, C. F., Reznik, S., Kochhar, S., Lindon, J. C., Holmes, E., et al. (2012). Stability and robustness of human metabolic phenotypes in response to sequential

"fgene-04-00205" — 2013/10/14 — 12:08 — page 8 — #8
Kussmann et al. Systems approach to diabetes research

www.frontiersin.org October 2013 | Volume 4 | Article 206 | 9

"fgene-04-00205" — 2013/10/14 — 12:08 — page 9 — #9

MacLellan, W. R., Wang, Y., and Lane, A. J. (2012). Systems-based approaches to cardiovascular disease. Nat. Rev. Cardiol. 9, 172–184. doi: 10.1038/nrcardio.2011.208

Malek, A. D., Lindsell, J. C., and Nicholson, K. J. (2009). 1H NMR-based metabolomics for investigating diabetes. Future Med. Chem. 7, 777–747. doi: 10.1055/s-0035-152098

Marinac, A., Jung, M. A., Erdhardt, L. and Aebischer, R. (2012). Targeted proteome investigation via selected reaction monitoring mass spectrometry. J. Proteomics 75, 3405–3513. doi: 10.1016/j.jprot.2012.04.048

Malacova, M., and Slupphaug, J. (2008). Problems in differential diagnosis of diabetes types. Pol. Arch. Med. Wewn. 110, 475–480.

Mancini, T. A., Colles, F. S., Cox, N. J., Goldin, D. B., Hindorff, L. A., Hunning, D. J., et al. (2009). Finding the missing heritability of complex diseases. Nature 461, 747–753. doi: 10.1038/nature08494

Marti, M., Wilkens, E., Coop, M., D’Hertog, W., Canha, D. A., Kerr, H., et al. (2011). Oleic-induced beta cell dysfunction and apoptosis: a proteomic approach to glucoketoponemia by an unsaturated fatty acid. J. Proteome Res. 10, 3572–3585. doi: 10.1021/pr100720t

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, B. D., Little, J. I., et al. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. 9, 356–369. doi: 10.1038/nrg2344

Matsuyama, K., Wakioka, K., Nakajima, T., Saiki, K., Morita, Y., Hino, K., et al. (2008). Genotype score in addition to common risk factors for prediction of type 2 diabetes. N. Engl. J. Med. 359, 2483–2496. doi: 10.1056/NEJMoa0804742

Meng, Q., Mahalik, V. P., Tak, H., and Yang, X. (2013). Systems biology approaches and applications in obesity, diabetes, and cardiovascular diseases. Curr. Cardiovasc. Risk Rep. 7, 75–7. doi: 10.1007/s12196-013-0238-y

Mishra, R. J., and Rao, S. L. (2012). Applications of mass spectrometry to metabolomics and metabolome analysis: detection of biomarkers of aging and age-related diseases. Mass Spectrom. Rev. 31, 76–95. doi: 10.1002/mas.20230

Mizrahi, A., and Vikram, N. K. (2004). Ironmed nontype 2 diabetes (metabolic syndrome) and obesity in Asian Indians: evidence and implications. Nahrung 48, 482–481. doi: 10.1002/1521-410X.2004.01.020
Kussmann et al. Systems approach to diabetes research
Morgan, D. K., and Whitelaw, E. (2008). The case for transgen- 

tritional epigenetic inheritance in humans. Memory. Genome 8, 319– 

207. doi: 10.1016/j.magen.2008.03.002
Morin, M. J., Tierney, A. C., van O'Sullivan, A., Balducci, D., Para-

Ollikainen, M., Smith, K. R., Joo, Nicholson, J. K., Holmes, E., and 

Need, A. C., and Goldstein, D. B. (2010). Metabolic profiling of the 

human neonatal epigenome. Biochim. Biophys. Acta 1805, 4176–4188. 

doi: 10.1016/j.bbadis.2010.06.025
Myles, S., Davison, D., Barrett, J., Wilson, I. D. (2005). Gut microor-

organisms, mammalian metabolism and personalized health care. 

Nutrigenomics October 2013 | Volume 4 | Article 205

"fgene-04-00205" — 2013/10/14 — 12:08 — page 10 — #10
Kussmann et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108. doi: 10.1126/science.1208344

Yokota, A., Fukiya, S., Islam, K. B., Ooka, T., Opara, Y., Harahai, Y., et al. (2012). Bile acid a determinant of the gut microbiota on a high-fat diet? *Gut Microbes* 3, 455–459. doi: 10.4161/gmic.21256

Zhao, L., Nicholson, J. K., Lu, A., Wang, Z., Tang, H., Helman, E., et al. (2012). Targeting the human genome-microbiome axis for drug discovery: inspirations from global systems biology and traditional Chinese medicine. *J. Proteome Res.* 11, 3509–3519. doi: 10.1021/pr3005629

Zhao, L. C., Zhang, X. D., Liu, S. X., Gao, W. C., Wang, H. X., and Lin, D. H. (2010). A metabonomic comparison of urinary changes in Zucker and GK rats. *J. Biomed. Biotechnol.* 2010, 431894. doi: 10.1155/2010/431894

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 April 2013; paper pending published: 06 May 2013, accepted: 24 September 2013, published online: 16 October 2013.

Citation: Kussmann M, Morine MJ, Hager J, Sonderegger B and Kaput J (2013) Perspective: a systems approach to diabetes research. *Front. Genet.* 4:205. doi: 10.3389/fgene.2013.00205

This article was submitted to Nutrigenomics, a section of the journal *Frontiers in Genetics*. Copyright © 2013 Kussmann, Morine, Hager, Sonderegger and Kaput. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.