Molecular Genetics and Metabolism 121 (2017) 43–50

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism

journal homepage: www.elsevier.com/locate/ymgme

Regular Article

Type-2 diabetes-associated variants with cross-trait relevance: Post-GWAs strategies for biological function interpretation

Francesca Frau [a], Daniel Crowther [a], Hartmut Ruetten [b], Karla V. Allebrandt [a,*]

[a] Department of Translational Informatics, R&D Translational Med. and Early Development, Sanofi-Aventis, Deutschland GmbH, Industriepark Höchst, D-65926 Frankfurt am Main, Germany
[b] R&D Translational Med. and Early Clinical, Sanofi-Aventis, Deutschland GmbH, Industriepark Höchst, D-65926 Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 19 December 2016
Received in revised form 13 March 2017
Accepted 13 March 2017
Available online 22 March 2017

Keywords:
Type 2 diabetes genes
Post-GWA study strategy
Systems genomics
Network medicine

ABSTRACT

Genome-wide association studies (GWAs) for type 2 diabetes (T2D) have been successful in identifying many loci with robust association signals. Nevertheless, there is a clear need for post-GWAs strategies to understand mechanism of action and clinical relevance of these variants. The association of several comorbidities with T2D suggests a common etiology for these phenotypes and complicates the management of the disease. In this study, we focused on the genetics underlying these relationships, using systems genomics to identify genetic variation associated with T2D and 12 other traits. GWAs studies summary statistics for pairwise comparisons were obtained for glycemic traits, obesity, coronary artery disease, and lipids from large consortia GWAs meta-analyses. We used a network medicine approach to leverage experimental information about the identified genes and variants with cross traits effects for biological function interpretation. We identified a set of 38 genetic variants with cross traits effects that point to a main network of genes that should be relevant for T2D and its comorbidities. We prioritized the T2D associated genes based on the number of traits they showed association with and the experimental evidence showing their relation to the disease etiology. In this study, we demonstrated how systems genomics and network medicine approaches can shed light into GWAs discoveries, translating findings into a more therapeutically relevant context.

1. Introduction

Common type 2 diabetes (T2D) is a complex multifactorial disease, modulated by gene-gene and gene-environmental interactions (for review see [1]. More than 80 loci are now described to be associated with T2D [2], but mechanistic understanding has only been achieved for a couple of these genes. Identifying cross traits effects of T2D associated variants can contribute to explain epidemiological links between these phenotypes.

Diseases with shared etiology are being shown to be associated with common genetic variants of pleiotropic effects. The *HLA* region, for instance, is well known for being associated with several autoimmune diseases including type 1 diabetes, multiple sclerosis and rheumatoid arthritis [3]. *AKT1* is associated with ovarian cancer, breast cancer, and colorectal cancer [4,5]. *ITGA1* variants are associated with T2D, fasting insulin, beta-cell function by homeostasis model assessment, and 2-h post–oral glucose tolerance test and insulin levels. *ITGA1* is also involved in liver fibrosis, insulin secretion, and bone healing [6].

To investigate the biology underlying the cross traits associations that could be involved in the pathophysiology linking T2D to the other investigated traits, we used systems genetics and network medicine approaches. This included several post-GWAs strategies (Fig. 1) as the: (i) identification of T2D associated variants with cross traits relevance using summary statistics from 13 GWAs meta-analyses. We analyzed T2D associated variants [2] versus associated variants from 4 different main domains: **glycemic traits** (2 h glucose, HbA1C, fasting glucose, fasting insulin [7–9]), **obesity** [10], **lipids** (low-density lipoprotein-LDL; Triglycerides-TG; total cholesterol-TC, high-density lipoprotein-HDL, [11]) and **coronary artery disease** (CAD [12]); (ii) eQTLs annotation, (iii) human islet gene expression data in T2D versus healthy controls and, (iv) pathway enrichment analyses and overconnectivity of identified cross trait relevant GWAs signals. Using the results of these approaches we draw hypotheses about gene relevance for disease etiology. Our analyses revealed a number of genes with cross traits effects, which we annotated with several analytical tools with the aim to leverage information on their functional relevance for T2D and related comorbidities. This manuscript summarizes these results and shows a pipeline of methods that are useful in predicting gene biological relevance from GWAs results.

Abbreviations: T2D, type 2 diabetes; GWA, Genome-wide association; LD, Linkage disequilibrium; eQTL, Expression quantitative trait locus; SNP, Single nucleotide polymorphism; LDL, low-density lipoprotein cholesterol; TG, Triglycerides; TC, Total cholesterol, and HDL, high-density lipoprotein cholesterol.

* Corresponding author.
E-mail address: karlaviviani.allebrandt@sanofi.com (K.V. Allebrandt).

http://dx.doi.org/10.1016/j.ymgme.2017.03.004

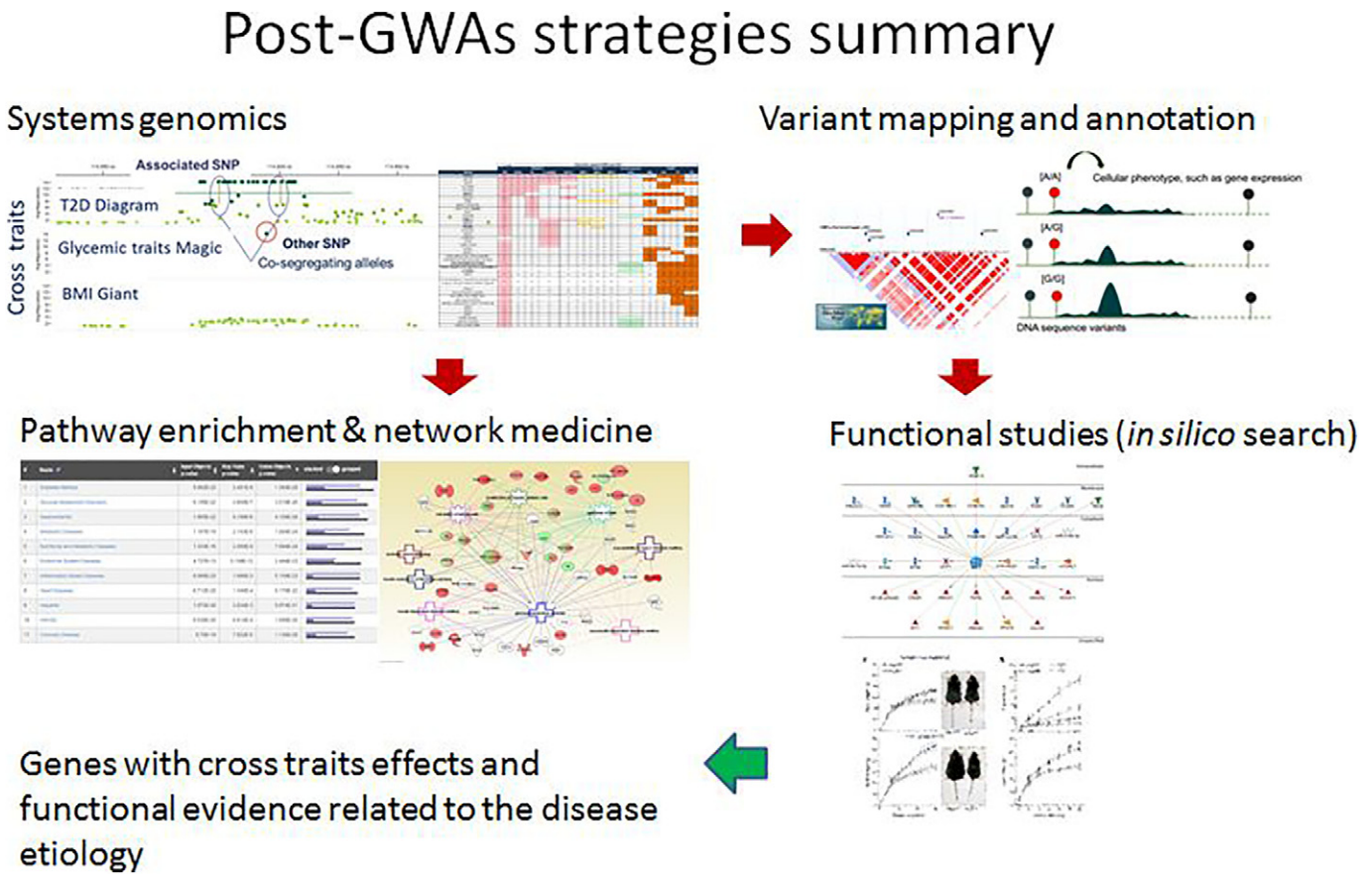

Fig. 1. Study schema. We used a systems genomics approach for the identification of T2D associated variants with cross traits relevance using summary statistics from 13 GWAs meta-analyses for T2D [2] versus 4 main domains: glycemic traits (2 h glucose, HbA1C, fasting glucose, fasting insulin [7–8]), obesity [9], lipids (low-density lipoprotein-LDL; Triglycerides-TG; total cholesterol-TC, high-density lipoprotein-HDL, [10]) and coronary artery disease (CAD [11]). This approach was followed by fine mapping and functional annotation of individual SNPs with eQTLs. To understand biological relevance, we conducted functional network and pathway enrichment analyses with the identified cross trait relevant GWAs signals.To leverage further functional information for biological interpretation of the GWAs signals, we looked into protein interactions and at human islet gene expression data in T2D versus healthy controls.

2. Methods

The methodology workflow used is summarized in Electronic Supplementary Material (ESM) Fig. 1.

GWA studies datasets

Published summary statistics from GWAs analyses included in this study were obtained as open source data from the following consortia: DIAbetes Genetics Replication and Meta-analysis (DIAGRAM), the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC), The Genetic Investigation of ANthropometric Traits (GIANT), Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIOGRAM), and Global Lipids Genetics Consortium (GLGC). These were studies for 13 traits, T2D [2], fasting glucose (FG), fasting insulin (FI), 2 h glucose (2hglu) [7], HbA1c [8], obesity class I (BMI-1 ≥ 30), obesity class II (BMI-2 ≥ 35), obesity class III (BMI-3 ≥ 40 kg/m2) [10], CAD [12], TC, HDL, LDL, and TG [11]. These were European cohorts with sample size between 42,854 to 196,475 (ESM text and ESM Table 1). Given that linkage disequilibrium (LD) varies between ethnic populations, we selected results from European cohorts and used only a Central European reference panel to calculate LD between variants. We used summary statistics containing study-by-study pre-filtered variants with Hardy–Weinberg equilibrium (HWE) p-value $> 10^{-7}$, individual study filter for the individual and SNP call rate varied from a minimum of 0.90 to 0.99, depending of the study included in the meta-analyses. A filter for MAF < 0.01 was included in the command line when running the cross traits analyses with the summary statistics files. Study URLs containing the individual study protocols from the data that was downloaded for analyses are cited at the end of the manuscript.

2.2. Identification of regions with cross traits association and main connectivity network

To identify variants with cross traits effects we applied a clumping procedure using PLINK [13], based on empirical estimates of linkage disequilibrium (LD) between single nucleotide polymorphisms (SNPs) (Fig. 1, ESM Table 2). Variants associated with T2D at a p-value $< 10^{-4}$ (–clump-p1 0.0001) in the DIAGRAM meta-analyses were used as reference (–clump-index-first) to identify corresponding variants associated with other investigated traits at a p-value $\leq 10^{-3}$ (–clump-p2 0.001) and to do not exclude any potential significant cross trait effect. We used HapMap 2 (CEU release 23a) as the reference dataset to estimate the LD between variants; most of the datasets used had been imputed for this reference panel. The clumping analysis was performed using LD $r^2 > 0.8$ (–clump-r2 0.8) to clump variants to an index SNP within a range of 250 kb (–clump-kb 250). This clumping strategy did not account for the direction of association with disease as variants may have an opposite effect in different traits. Only clumps containing clumped SNPs with p2-significant results in more than one result file (–clump-replicate) were considered. To identify clumped genomic regions corresponding to genes we used the –clump-range function with a gene list (hg18). We used the obtained gene names (ESM Table 3) to identify transcripts in HumanMine. We created a hierarchical cluster matrix by Euclidian distance method and the average linkage cluster algorithm, plotting genes versus traits that their respective SNPs were