



# Pancreatic Islet Protein Complexes and Their Dysregulation in Type 2 Diabetes

Helle Krogh Pedersen<sup>1†</sup>, Valborg Gudmundsdottir<sup>1†</sup> and Søren Brunak<sup>1,2\*</sup>

<sup>1</sup> Department of Bio and Health Informatics, Technical University of Denmark, Kgs Lyngby, Denmark, <sup>2</sup> Disease Systems Biology, Faculty of Health and Medical Sciences, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

## OPEN ACCESS

### Edited by:

Rosalba Giugno,  
University of Verona, Italy

### Reviewed by:

Christophe Magnan,  
Paris Diderot University, France  
Piero Marchetti,  
University of Pisa, Italy

### \*Correspondence:

Søren Brunak  
soren.brunak@cpr.ku.dk

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Bioinformatics and Computational  
Biology,  
a section of the journal  
Frontiers in Genetics

Received: 01 February 2017

Accepted: 27 March 2017

Published: 20 April 2017

### Citation:

Pedersen HK, Gudmundsdottir V and  
Brunak S (2017) Pancreatic Islet  
Protein Complexes and Their  
Dysregulation in Type 2 Diabetes.  
Front. Genet. 8:43.  
doi: 10.3389/fgene.2017.00043

Type 2 diabetes (T2D) is a complex disease that involves multiple genes. Numerous risk loci have already been associated with T2D, although many susceptibility genes remain to be identified given heritability estimates. Systems biology approaches hold potential for discovering novel T2D genes by considering their biological context, such as tissue-specific protein interaction partners. Pancreatic islets are a key T2D tissue and many of the known genetic risk variants lead to impaired islet function, hence a better understanding of the islet-specific dysregulation in the disease-state is essential to unveil the full potential of person-specific profiles. Here we identify 3,692 overlapping pancreatic islet protein complexes (containing 10,805 genes) by integrating islet gene and protein expression data with protein interactions. We found 24 of these complexes to be significantly enriched for genes associated with diabetic phenotypes through heterogeneous evidence sources, including genetic variation, methylation, and gene expression in islets. The analysis specifically revealed ten T2D candidate genes with probable roles in islets (*ANPEP*, *HADH*, *FAM105A*, *PDLIM4*, *PDLIM5*, *MAP2K4*, *PPP2R5E*, *SNX13*, *GNAS*, and *FRS2*), of which the last six are novel in the context of T2D and the data that went into the analysis. Fifteen of the twenty-four complexes were further enriched for combined genetic associations with glycemic traits, exemplifying how perturbation of protein complexes by multiple small effects can give rise to diabetic phenotypes. The complex nature of T2D ultimately prompts an understanding of the individual patients at the network biology level. We present the foundation for such work by exposing a subset of the global interactome that is dysregulated in T2D and consequently provides a good starting point when evaluating an individual's alterations at the genome, transcriptome, or proteome level in relation to T2D in clinical settings.

**Keywords:** diabetes, data integration, protein complexes, tissue specificity, pancreatic islets, patient network biology

## INTRODUCTION

Diabetes is a multi-tissue metabolic disease caused by defects in insulin action, insulin secretion, or both, resulting in hyperglycemia. The heritability of type 2 diabetes (T2D) has been estimated to range from 25 to 80% (Prasad and Groop, 2015). Despite that more than 120 T2D risk loci have been identified so far (Prasad and Groop, 2015) their combined effect explains only a fraction of the

heritability. The unexplained heritability of complex traits is expected to mainly reside in a large number of common and rare variants across the human genome (Yang et al., 2015). Identifying the remaining variants involved in T2D through traditional single-variant association analyses will require greatly increased sample sizes compared to current studies for improving statistical power (Morris et al., 2012). Integrative systems biology approaches hold the promise to facilitate this process by considering gene products in the context of cellular networks rather than in isolation, thus improving power through the use of existing biological knowledge.

Genome-wide analyses, such as genome-wide association studies (GWAS) and studies of differential expression or methylation, often rank thousands of genes for phenotype associations. Integrating such data is a powerful way to identify genes important in the disease pathogenesis that are not identifiable in any single dataset but become evident when considering the different evidence sources collectively (Kodama et al., 2012; Pers et al., 2013). Combining such integrative evidence with protein complexes provides additional insight into the biological context and has the potential to reveal novel therapeutic targets (Lage et al., 2012).

The subset of protein complexes active in a given tissue is restricted by the tissue-specific proteome, which is important to consider because disease-associated genes have a tendency to exhibit tissue-specific gene expression in affected tissues (Lage et al., 2008). Previous studies have shown that disease-gene prioritization is improved when using tissue-specific networks compared to tissue-naïve protein interaction networks (Magger et al., 2012; Ganegoda et al., 2014). Consequently, considering disease associated genes in the appropriate context is a promising avenue for making further inroads into disease understanding (Gross and Ideker, 2015). Such tissue-specific analyses are now enabled by the increasing amount of large-scale tissue and cell type specific data sets (Lonsdale et al., 2013; Kim et al., 2014; Uhlén et al., 2015), making it possible to disentangle or deconvolute tissue and cell type-specific processes.

A key diabetes tissue is the islet of Langerhans, which plays an important role in diabetes pathology. Islets are scattered around in the pancreas where they only constitute 1–2% of the total organ mass. They consist of a number of different highly specialized endocrine cell-types with the insulin-producing beta-cells and glucagon-producing alpha-cells being of the highest relevance to diabetes (Danielsson et al., 2014). Utilizing tissue-specific data, one major aim of this study was to create a pancreatic and beta-cell specific resource of protein complexes to serve as an integration scaffold in this and future studies. Previous work on tissue-specific protein interaction networks did either not include human pancreatic islets (Guan et al., 2012; Barshir et al., 2013; Basha et al., 2015) or were restricted to tissue-specific gene expression data (Bossi and Lehner, 2009; Magger et al., 2012; Greene et al., 2015). By focusing on the pancreatic islet, we supplement these resources by integrating high-confidence physical protein interaction network data with islet-specific gene expression data from both microarray and RNAseq studies, as well as protein expression from immunohistochemistry-based protein profiling.

Another major aim of the study was to identify a set of islet protein complexes that are likely dysregulated or dysfunctioning in T2D. To investigate this, we searched for complexes that were enriched for genes implicated in diabetic phenotypes through heterogeneous sources of evidence, ranging from genetic variation to methylation and gene expression in islets. The resulting complexes thus represent functional units whose perturbation can give rise to a diabetic phenotype and at the same time provide insight into the genetic heterogeneity that contributes to the pathogenesis of T2D in pancreatic islets.

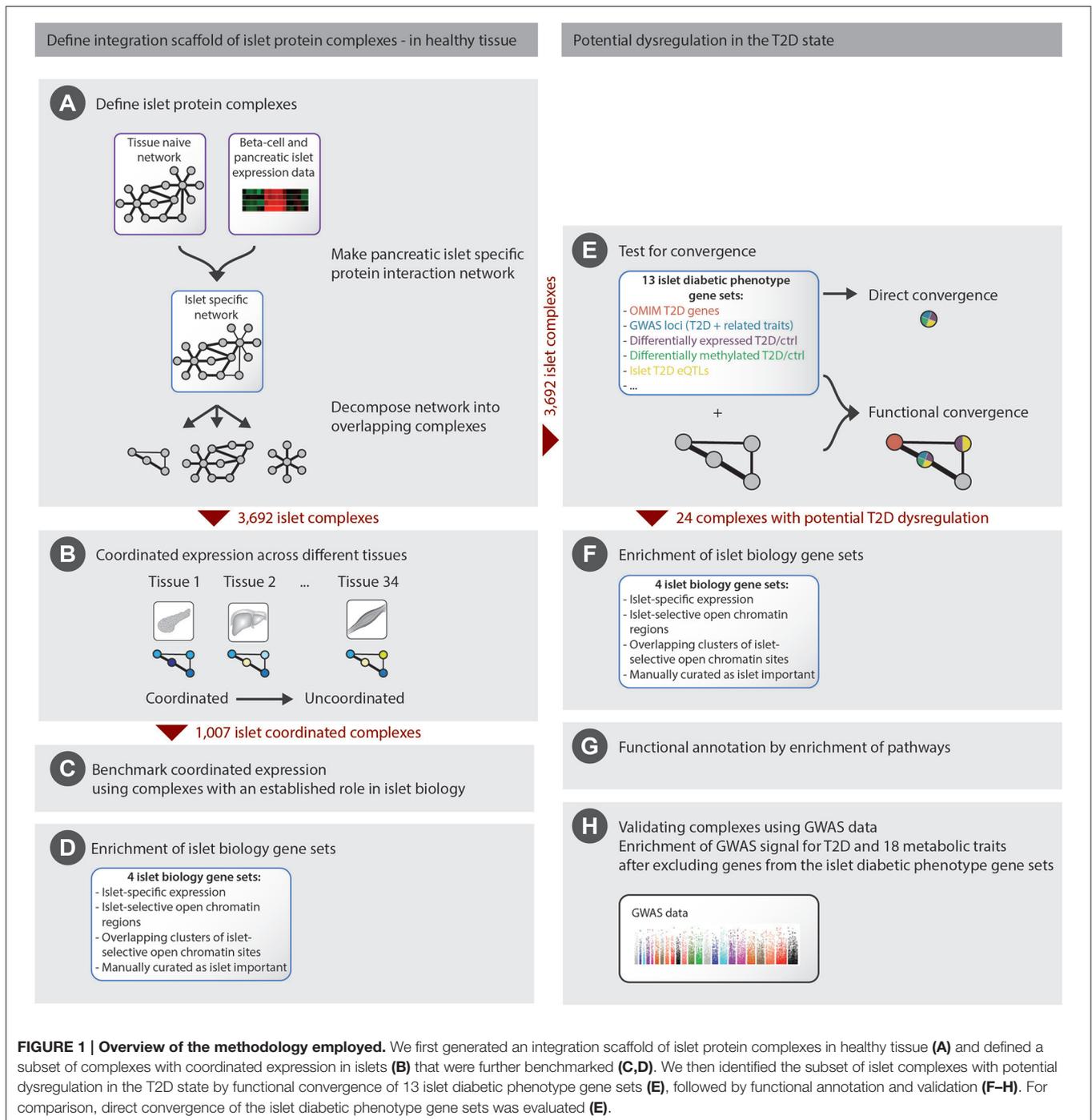
## RESULTS

### Defining a Catalog of 3,692 Islet Protein Complexes

We generated an islet-specific protein interaction network using gene and protein expression data combined with high-confidence protein interactions (see Section Methods and **Figure 1A**). This network was further decomposed into 3,692 overlapping protein complexes (10,805 genes) using the two complementary methods, ClusterOne (Nepusz et al., 2012) and spoke-hub, focusing on high internal connectivity and hub-topology, respectively (see Section Methods for details). We specifically chose network decomposition algorithms that allow for overlapping complexes as many proteins participate in multiple processes, making it difficult to decide on a single partition that closely reflects biological reality. These complexes, ranging in size from 6 to 50 proteins, captured different regions in network topology space, some being sparsely connected whereas others showed complete internal connectivity with all nodes having a physical interaction with all other nodes (Supplementary Table 1). This set of complexes represents a catalog of islet protein complexes and their constituents.

### Coordinated Expression of Islet Protein Complexes

Tissue-specific coordination of gene expression among members of a protein complex may indicate an important function of the complex in the respective tissue (Han et al., 2004; Taylor et al., 2009; Börnigen et al., 2013). To investigate the status of the islet complexes, we calculated the degree of coordinated expression of each of the 3,692 complexes across a range of 34 tissues as the normalized average Pearson correlation coefficient of interacting proteins, using data from the GTEx consortium (Ardlie et al., 2015) and the study by Nica et al. (2013) (see Section Methods for details and **Figure 1B**). To evaluate the importance of coordinated expression for islet relevant complexes, we defined a set of 76 benchmarking islet complexes, each constituted by 10% of genes known to be of major importance for islet function and identity (Pasquali et al., 2014; **Figure 1C**). These benchmarking complexes had significantly higher coordinated expression in either islets, beta-cells, or non-beta islet cells compared to the background distribution of all other complexes (MWU,  $P = 9.6 \times 10^{-4}$ , Supplementary Figure 1). These results suggest that coordinated



islet gene expression of protein complex members can indicate an important role in islet biology. We therefore defined a subset of 1,007 islet-coordinated complexes where at least one of the islet tissue components (whole islets, beta, or non-beta cells) was among the three highest ranked across the 34 tissues tested (see Section Methods). Moreover, the 1,007 complexes were enriched (MWU,  $P = 2.8 \times 10^{-4}$ ) for genes residing in islet regulatory regions defined as having islet-selective open chromatin in the

transcription start site or gene-body (Supplementary Table 2; Figure 1D).

While these 1,007 complexes are of special interest in the context of islet function, previous work related to the cell cycle (de Lichtenberg et al., 2005) has illustrated that protein complexes can be functional even though not fully coordinated due to sophisticated, temporal regulation. We therefore included all 3,692 complexes in the further analyses on T2D dysregulation.

## Limited *Direct* Overlap of Islet Diabetes Gene Sets

Having a catalog of 3,692 islet relevant protein complexes we next turned to investigate which of those were most likely to be implicated in T2D (Figure 1E). The underlying hypothesis is that complexes exhibiting pronounced convergence of genes originating from different evidence sources related to diabetes are likely to play a role in the disease.

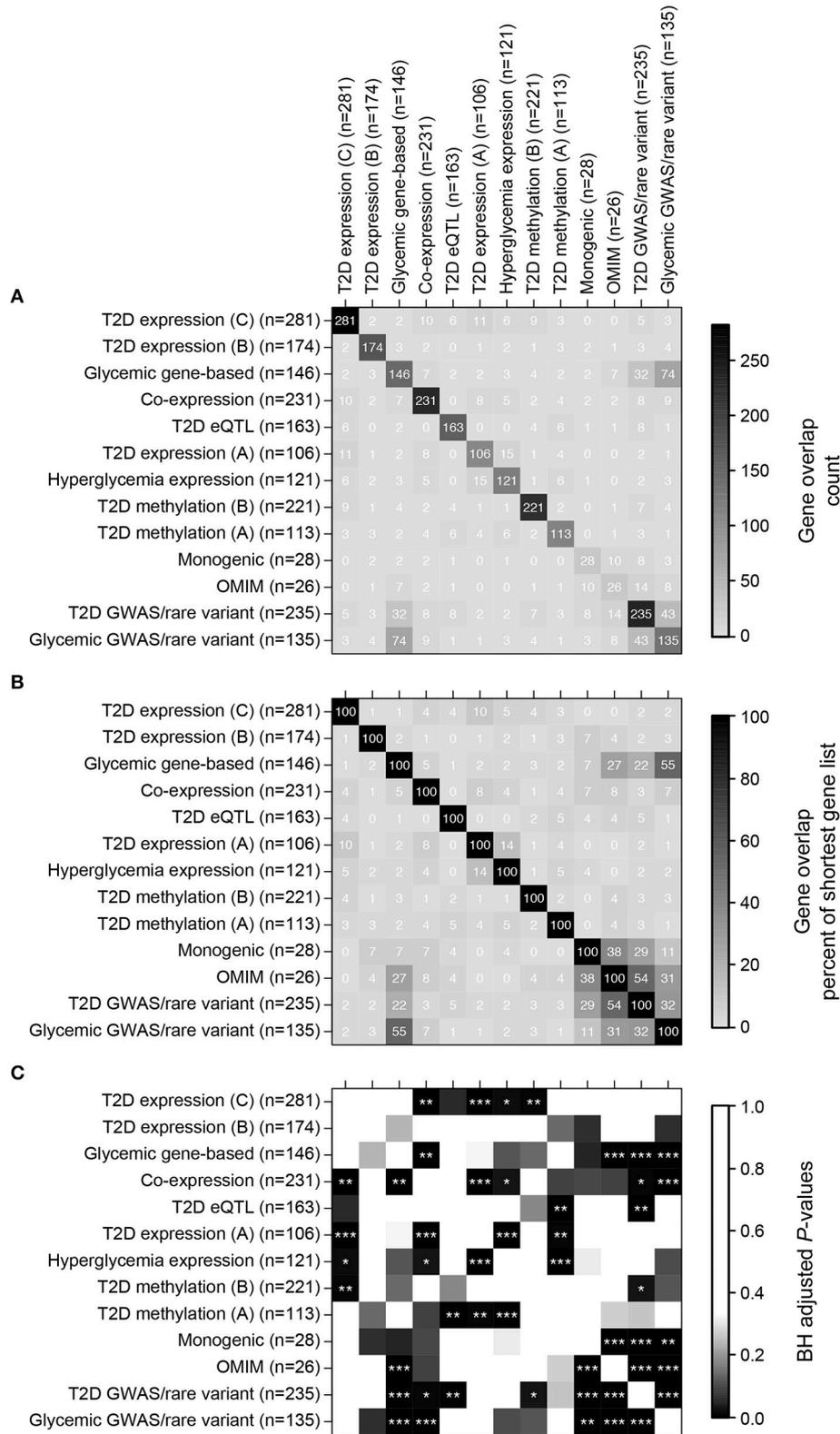
We thus compiled 13 sets of genes associated with T2D, monogenic forms of diabetes and related metabolic phenotypes

(Table 1), hereafter termed islet diabetic phenotype gene sets. Despite all gene sets being related to diabetes, they generally showed surprisingly little direct overlap, although many pairwise overlaps were still larger than expected by chance (Figure 2). The largest overlaps, ranging from 11 to 55% relative to the size of the shortest list, were observed between gene sets based on genetic variation (Monogenic, OMIM, T2D GWAS/rare variant, Glycemic GWAS/rare variant, and Glycemic gene-based), which is to some extent expected as many genes causing monogenic forms of diabetes also harbor variants associated with T2D and

**TABLE 1 | Description of the thirteen islet diabetic phenotype gene sets and the four islet biology related gene sets.**

Name	Description	References	# Genes (# genes in network)
<b>ISLET DIABETIC PHENOTYPE GENE SETS</b>			
<b>GWAS LOCI AND RARE VARIANT GENES</b>			
T2D GWAS/rare variant	Genes in the vicinity of T2D GWAS SNPs, using a boundary of 110 kb upstream and 40 kb downstream of each gene, as well as genes harboring rare variants associated with T2D.	Morris et al., 2012; Albrechtsen et al., 2013; Flannick et al., 2014; Mahajan et al., 2014; Steinthorsdottir et al., 2014; Wessel et al., 2015	235 (162)
Glycemic GWAS/rare variant	Genes in the vicinity of GWAS SNPs (FG, BMI-adjusted FG, 2 h Glu, BMI-adjusted 2 h Glu, insulinogenic index, disposition index, proinsulin), using a boundary of 110 kb upstream and 40 kb downstream of each gene, as well as genes harboring rare variants associated with FG, proinsulin, or insulinogenic index.	Strawbridge et al., 2011; Scott et al., 2012; Huyghe et al., 2013	135 (107)
<b>GWAS GENES (GENE-BASED TEST)</b>			
Glycemic gene-based	Genes associated with FG, 2 h Glu, or proinsulin using a gene-based analysis.	Scott et al., 2012; Huyghe et al., 2013	146 (130)
<b>OMIM T2D GENES</b>			
OMIM	Genes associated with "Diabetes mellitus, noninsulin-dependent; NIDDM" in the OMIM database (accession #125853)		26 (24)
<b>MONOGENIC DIABETES GENES</b>			
Monogenic	MODY and other monogenic diabetes genes.	McCarthy, 2010; Scott et al., 2012	28 (28)
<b>ISLET eQTL GENES For 47 T2D SNPs (CIS AND TRANS)</b>			
T2D eQTL	Five cis and 176 trans eQTLs in islets, based on 47 SNPs associated with T2D.	Taneera et al., 2012	163 (129)
<b>GENES DIFFERENTIALLY METHYLATED IN ISLETS (T2D vs. CTRL)</b>			
T2D methylation (A)	Genes in differentially methylated regions that are also differentially expressed.	Dayeh et al., 2014	113 (88)
T2D methylation (B)	Genes in differentially methylated regions.	Volkmar et al., 2012	221 (169)
<b>GENES CO-EXPRESSED WITH 2+ T2D GENES</b>			
Co-expression	Genes that are co-expressed in islets with 2 or more of 48 T2D genes.	Taneera et al., 2012	231 (197)
<b>GENES DIFFERENTIALLY EXPRESSED IN ISLETS (T2D OR HYPERGLYCEMIC vs. CTRL)</b>			
Hyperglycemia expression	Differentially expressed genes in islets, in hyperglycemic vs. normoglycemic individuals.	Taneera et al., 2012	121 (109)
T2D expression (A)	Differentially expressed genes in islets, in T2D patients vs. controls.	Taneera et al., 2012	106 (90)
T2D expression (B)	Differentially expressed genes in islets, in T2D patients vs. controls.	Dominguez et al., 2011	174 (150)
T2D expression (C)	Differentially expressed genes in beta-cells, in T2D patients vs. controls.	Marselli et al., 2010	281 (237)
<b>ISLET BIOLOGY GENE SETS</b>			
Islet specific	Top 30 islet specific genes.	Morán et al., 2012	33 (28)
Open chromatin	Genes with islet-selective (compared to five non-islet cell lines) open chromatin in the transcription start sites or gene-body.	Gaulton et al., 2010	319 (226)
Open chromatin clusters	Genes overlapping <i>clusters</i> of islet-selective open chromatin sites.	Gaulton et al., 2010	1,512 (1,340)
Islet biology	Sixty-seven genes curated as important for islet cell identity and function, Supplementary Table 2.	Pasquali et al., 2014	67 (57)

2 h Glu, 2 hour glucose; BMI, body mass index; FG, fasting glucose; T2D, type 2 diabetes.



**FIGURE 2 | Direct overlap of the thirteen islet diabetic phenotype gene sets. (A)** Overlap in terms of gene counts. **(B)** Overlap in terms of percent overlap relative to the size of the shortest gene sets. **(C)** BH-adjusted *P*-values for testing significance of overlap (hypergeometric test using all 22,766 genes as background), stars are as follows: \*\*\**P* ≤ 0.001, \*\**P* ≤ 0.01, \**P* ≤ 0.05.

glycemic traits (Bonnefond and Froguel, 2015). Twenty genes were found to be part of four or more of the 13 gene sets (Supplementary Table 3), many of which are well-known T2D susceptibility genes while others are less well-established in the context of diabetes, some of those examples are highlighted in **Box 1**.

## Complexes Showing *Functional* Overlap of Islet Diabetes Gene Sets

We next investigated if the 13 islet diabetic phenotype gene sets functionally converged on any of the 3,692 islet protein complexes, by calculating the combined enrichment for the 13 gene sets for each complex (**Figure 1E**). We found that the 1,007 complexes with coordinated expression in islets were enriched for small *P*-values (MWU,  $P = 1.66 \times 10^{-5}$ ) and we furthermore observed significant convergence of the islet diabetic phenotype gene sets in 24 complexes (9 coordinated, 15 un-coordinated) after adjusting for multiple hypothesis testing (BH adjusted  $P < 0.05$ ; Supplementary Table 4, Data Sheet 1). All of these 24 complexes contained one or more gene supported by genetic evidence (GWAS, rare variants or monogenic forms of diabetes), suggesting that the majority are likely to play a causal role in the development of T2D (Supplementary Table 4).

The 24 complexes were additionally enriched for genes in all four islet biology gene sets (Supplementary Table 2; **Figure 1F**), suggesting an important role in pancreatic islets. The complexes largely showed limited gene-overlap (Supplementary Figure 2), which indicates that they span different parts of the islet interactome.

We next investigated the biological functions of the 24 diabetic phenotype associated complexes (**Figure 1G**), and found that the complexes segregate into functional distinctive groups based on their pathway enrichment patterns (**Figure 3**). A number of these groups were characterized by molecular processes well-known to be dysregulated in diabetic islets—such as potassium channels, glucokinase, incretin signaling, and Wnt signaling—while others were enriched for processes

less established in the islet pathogenesis of T2D, such as insulin-, interleukin-, and ephrin-signaling, cell and adherens junctions and neurotransmitter release. Interestingly, seven of the 24 complexes contained one or more target of FDA-approved drugs, many of which are not anti-diabetic agents (Data Sheet 1).

## Leveraging the Complexes to Propose Novel T2D Genes

The 294 genes constituting the 24 complexes are all interesting in the context of diabetes (Supplementary Table 5). Obviously, many of them already have an established role in T2D. By contrast, the subset of 217 genes that were not part of any of the 13 islet diabetic phenotype gene sets comprise an interesting set for further prioritization. In particular, we identified six genes (*MAP2K4*, *PDLIM5*, *PPP2R5E*, *SNX13*, *GNAS*, and *FRS2*) of high interest as novel T2D associated genes, as they all have additional support for being of relevance for islet biology or function from the islet biology gene sets and furthermore SNPs in the vicinity of these genes are associated with T2D or glycemic traits with  $P < 1 \times 10^{-4}$  (**Table 2**).

Interestingly, after our analysis was completed, a targeted study of variants in the *PDLIM5* gene reported an association with T2D (rs11097432,  $P = 1.07 \times 10^{-3}$ ; Owusu et al., 2017). Additional support for the prioritized genes emerges from the recent wave of single-cell transcriptomics studies of human islets that were published after our analysis was finished (Segerstolpe et al., 2016; Wang et al., 2016; Xin et al., 2016; Lawlor et al., 2017). Remarkably, *GNAS* is among the 11 genes showing consistent differential expression in diabetic cell types (compared to non-diabetic) with same direction of effect in beta-cells (higher in T2D) in the first three studies and, furthermore, one (of 41 genes) found by both Lawlor et al. and Segerstolpe et al. with same direction of effect in alpha-cells (lower in T2D; Lawlor et al., 2017). In addition, Xin et al. (2016), reports *GNAS* to be abundant in all four major islet endocrine cell types (alpha, beta, delta, PP) in both non-diabetic and T2D donors (but not

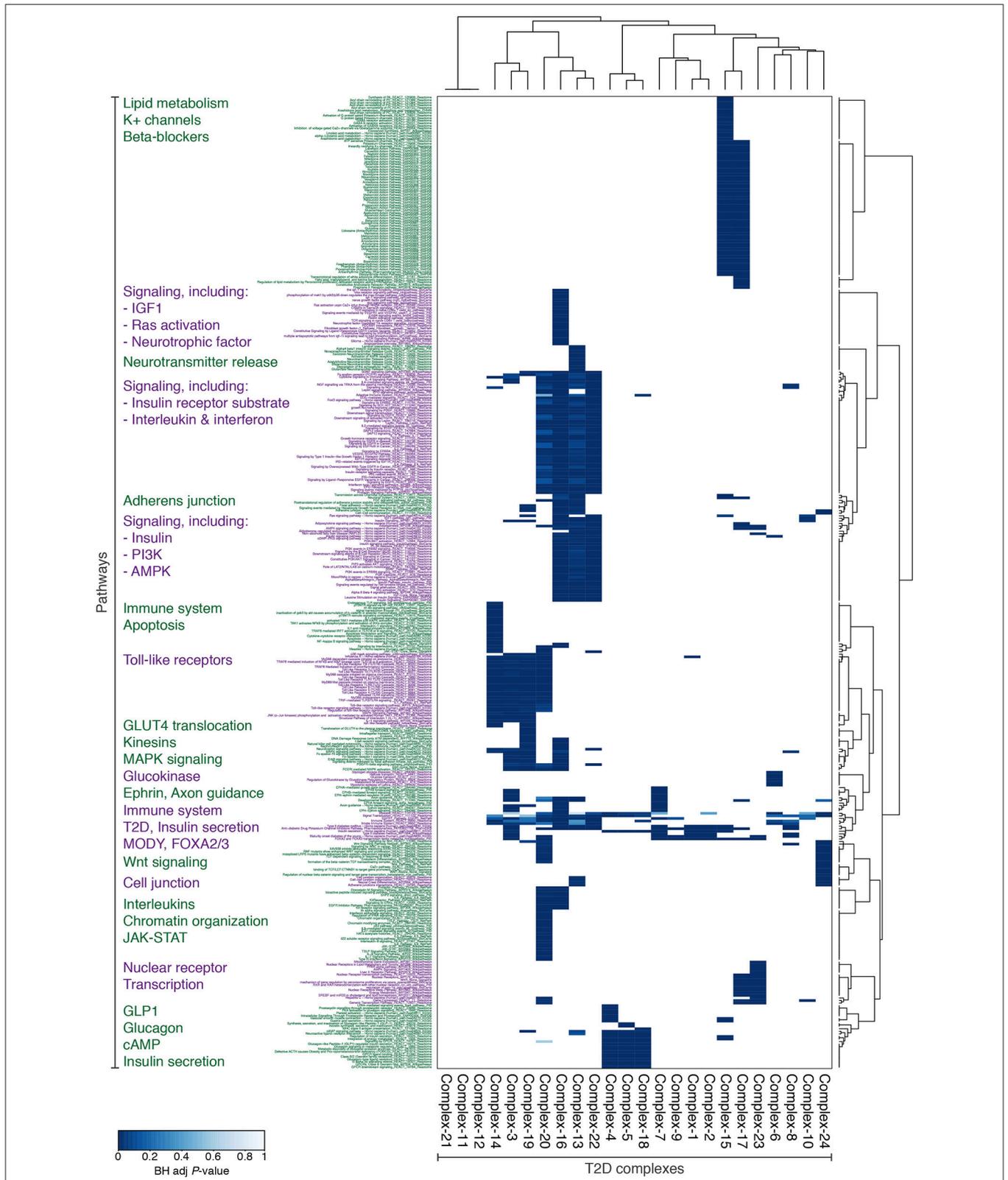
### BOX 1 | T2D CANDIDATE GENES PRIORITISED BY DIRECT CONVERGENCE.

The following genes were supported by four or more of the thirteen islet diabetic phenotype evidence sources, many across different levels of molecular regulation, but have not been strongly established in the context of T2D.

**The alanyl (membrane) aminopeptidase (*ANPEP*)** gene resides in a locus on chromosome 15 containing variants associated with T2D in South Asian individuals (Kooner et al., 2011) and its expression levels are furthermore associated with the T2D associated SNP rs560887 (*G6PC2* locus on chromosome 2), thus, representing a trans-eQTL (Taneera et al., 2012). In addition, the *ANPEP* gene promoter is located in a region that is hypomethylated in T2D islets (Volkmar et al., 2012), and finally the gene itself is differentially expressed in T2D beta-cells (Marselli et al., 2010). Collectively, these heterogeneous data types indicate together a plausible role of *ANPEP* in the pathogenesis of T2D in pancreatic islets. Supporting our observation, this gene has been proposed as the causal gene in this GWAS locus through a study of allelic expression profiling (Locke et al., 2015). A variant in this gene is associated with the levels of a peptide derived from the C3 complement protein that plays a role in the innate immune system (Shin et al., 2014).

**Hydroxyacyl-CoA dehydrogenase (*HADH*)** was differentially expressed in islets in three independent data sets comparing T2D patients and controls, as well as being co-expressed in islets with two or more T2D candidate genes. Mutations in *HADH* are known to cause familial hyperinsulinism (Glaser, 2013), which motivated a targeted study of common variants in the gene that however did not find any association with T2D (van Hove et al., 2006). Yet, our observations suggest that the expression of the gene is affected in pancreatic islets in T2D and that it may play a role in the disease.

**The islet expression of Family with sequence similarity 105, member A (*FAM105A*) and PDZ and LIM domain 4 (*PDLIM4*)** was associated with both T2D (Marselli et al., 2010; Taneera et al., 2012) and hyperglycemia (Taneera et al., 2012). *FAM105A* was furthermore coexpressed with the T2D genes *SLC30A8*, *G6PC2* and *KCNJ11* (Taneera et al., 2012) while *PDLIM4* resides in a region of the genome that was differentially methylated in islets when comparing T2D patients and controls (Dayeh et al., 2014). A variant upstream of *PDLIM4* (rs7727038) shows a nominal association ( $P = 5.2 \times 10^{-5}$ ) with fasting glucose in the MAGIC consortium (Dupuis et al., 2010). Both of these genes encode for proteins with relatively unknown functions.



**FIGURE 3 | The 24 complexes with potential T2D dysregulation are enriched for diverse and relevant functions.** Subset of Consensus PathDB-pathways, for which at least one protein complex is enriched with BH-adjusted  $P < 0.001$ . The pathways and complexes are clustered with Ward's hierarchical clustering using an asymmetric binary similarity measure.

**TABLE 2 | Plausible novel T2D genes prioritized from the complexes with potential T2D dysregulation.**

Gene symbol	Gene name	# Islet diabetic phenotype gene sets	# Islet biology gene sets	Minimum <i>P</i> -value for associated SNPs	Corresponding GWAS trait
MAP2K4	Mitogen-activated protein kinase kinase 4	0	1	$7.83 \times 10^{-6}$ (rs929441)	AUCIns/AUCGluc
PDLIM5	PDZ and LIM domain 5	0	1	$9.87 \times 10^{-5}$ (rs17021900)	Fasting glucose
PPP2R5E	Protein phosphatase 2, regulatory subunit B, epsilon isoform	0	1	$7.05 \times 10^{-5}$ (rs10151995)	Fasting glucose
SNX13	Sorting nexin 13	0	1	$4.02 \times 10^{-6}$ (rs2723517)	HbA1c
GNAS	GNAS complex locus	0	1	$4.73 \times 10^{-5}$ (rs6026565)	Fasting glucose, Manning
FRS2	Fibroblast growth factor receptor substrate 2	0	1	$9.76 \times 10^{-6}$ (rs12425398)	Fasting glucose, Manning

Genes are prioritized if they, besides being part of a protein complex showing potential T2D dysregulation, are part of at least one of the four islet biology gene sets and harbor at least one SNP with  $P < 1 \times 10^{-4}$  in one or more of the 19 GWAS described in Supplementary Table 6. Only the best SNP *P*-value and corresponding GWAS trait are shown. To focus on novel T2D genes, genes in any of the 13 islet diabetic phenotype gene sets are excluded.

significantly differentially expressed). *SNX13* also exhibits cell type specific differential expression in T2D, being lower in delta cells of diabetic donors (fold change =  $-13.02$ , FDR =  $4.93 \times 10^{-2}$ ; Xin et al., 2016). Whole islet gene expression (profiled with microarrays and RNA-seq) is further nominally associated with lower HbA1c levels for both *GNAS* ( $P = 2.14 \times 10^{-3}$ , FDR =  $4.30 \times 10^{-2}$ ) and *SNX13* ( $P = 1.61 \times 10^{-2}$ , FDR =  $1.02 \times 10^{-1}$ ; Fadista et al., 2014). In mice, disruption of the G protein  $\alpha$ -subunit (one of the *GNAS* gene products) maternal (but not paternal) allele leads to severe obesity, hypertriglyceridemia, impaired glucose tolerance and insulin resistance (Xie et al., 2008). Together, these observations add support for the genes being important for shaping the diabetic phenotype in one or more islet cell types.

Both *MAP2K4* and *GNAS* are known to be involved in pancreatic cancer [Cancer Gene Census (Forbes et al., 2017) and Intogen (Gonzalez-Perez et al., 2013) databases]. The *MAP2K4* gene encodes the mitogen-activated kinase kinase (MKK)4, which constitutes a part of the apoptotic-effect mediating MEKK1-MKK4-JNK pathway (Xia et al., 1995) and is inhibited in pancreatic beta-cells by the glucagon-like peptide-1 analog extending-4, resulting in protection from palmitate-induced apoptosis (Natalicchio et al., 2013). *MAP2K4* is furthermore a proposed tumor suppressor gene and is significantly under-expressed in metastatic compared to benign pancreatic endocrine tumors (or islet cell tumors; Couvelard et al., 2006). These results point to an important role of *MAP2K4* in the survival of pancreatic islet cells, which is a process central to the etiology of both diabetes and pancreatic carcinomas. Further studies of the potential dual role of *MAP2K4* and *GNAS* might help elucidating the molecular basis for the complex bidirectional relationship observed between diabetes and pancreatic cancer (Li, 2012).

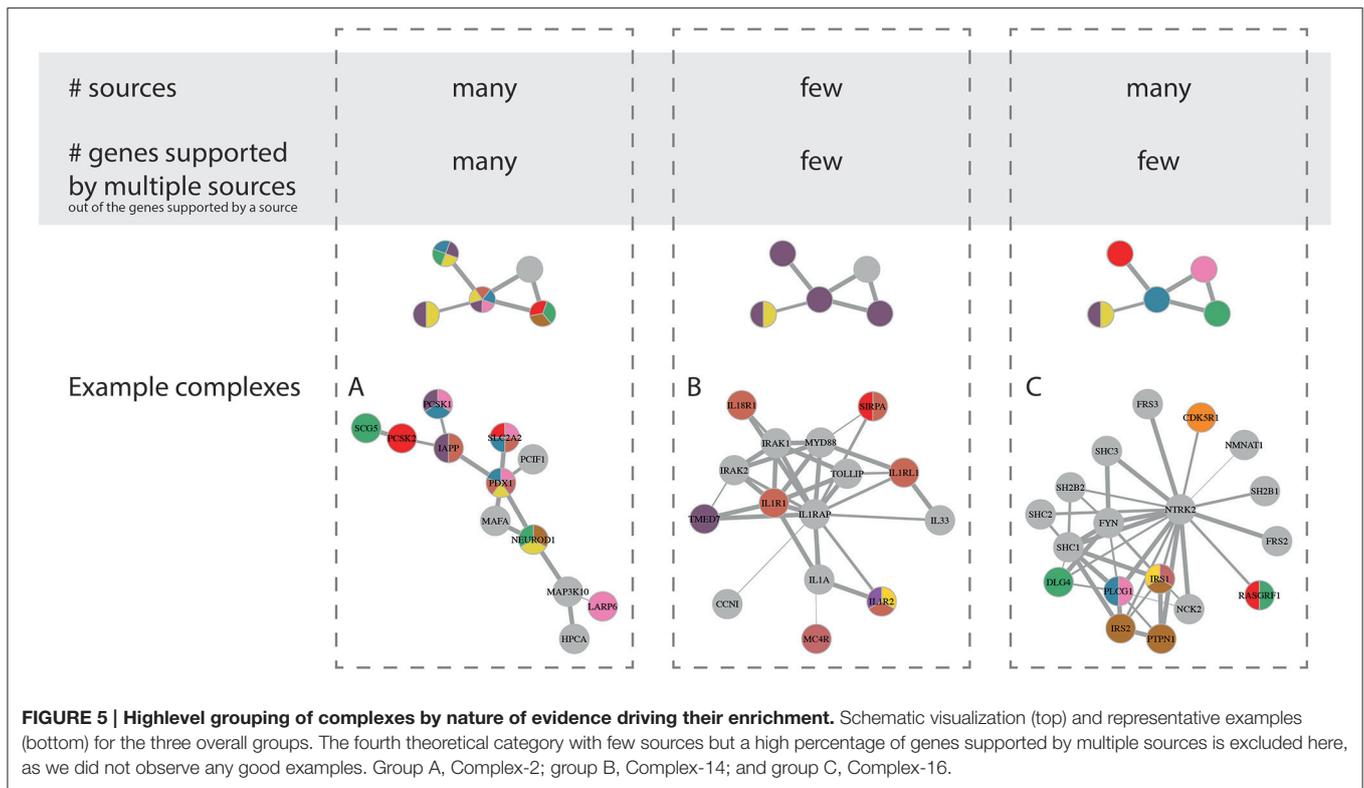
## Verification of Potential T2D Dysregulation of Complexes Using GWAS Data

As the 24 complexes were enriched for genes associated with diabetes and glycemic traits (input genes), it is likely that their

disruption gives rise to these phenotypes. Thus, the remaining (non-input) genes in the complexes have a high likelihood of also contributing to the same traits. We tested this hypothesis by investigating the enrichment of GWAS signals for T2D and glycemic traits from the DIAGRAM and MAGIC consortiums in each of the 24 complexes (Figure 1H). Using the Meta-Analysis Gene-set Enrichment of variant Associations (MAGENTA) tool to test the enrichment, we identified 30 significant ( $P < 0.05$ ) complex-trait combinations, spanning 15 complexes and 13 traits, of which 25 remained significant after excluding any genes that were used as input in the corresponding gene sets used for discovery of the complexes (Supplementary Figure 3). The last definition was applied to avoid any circularity, as the different GWAS might be the source of the association leading to the gene being in the islet diabetic phenotype gene sets that were used to define the 24 complexes with potential T2D dysregulation. These results indicate that the non-input genes in the complexes indeed harbor variants that are associated with the same phenotypes, although not so strongly that they could be discovered by the GWAS analysis alone.

We further investigated if the GWAS enrichment within the complexes was driven by many genes in loci with modest associations converging in the same functional context or mainly by one or a few genes with low *P*-values (minimum *P*-value for the SNPs mapping to the respective genes). We therefore repeated the analysis after excluding all genes with genome-wide significant *P*-values ( $P < 5 \times 10^{-8}$ ) in the respective GWAS and found that the enrichment for 27 out of 30 complex-trait combinations remained significant (Figure 4, Group 1). This suggests that the majority of the complexes represent examples where many small effects collectively perturb their function, leading to a molecular phenotype that gives rise to disturbed glucose homeostasis. All of the three complex-trait combinations that became non-significant (Figure 4, Group 2) contained one or more gene with a genome-wide significant signal ( $P < 5 \times 10^{-8}$ ), indicating that these genes were the main driver of the enrichment.





number of genes directly involved in insulin transcription and secretion, such as the insulin regulating transcription factors *PDX1* and *MAFA*, *PCSK1* and *PCSK2*, which are known to localize with insulin in islets, *IAPP*, which is co-secreted with insulin and *SCG5*, which is a marker of insulin secreting tumors. Interestingly, the *LARP6* gene in the complex was included in the islet diabetic phenotype gene sets because of its proximity to the fasting proinsulin associated SNP rs1549318 (Strawbridge et al., 2011). Its presence in the complex suggests that *LARP6* may play an important role in beta-cell function and insulin secretion. In line with the function of the genes in the complex, the overall complex was enriched for genetic associations with HOMA-B based on MAGIC data.

Complex-14 is an example from group (B), where the enrichment was driven by genes mainly supported by the same gene set (5/7 genes), namely the “Hyperglycemia expression” data. The additional supporting gene sets were mainly from gene expression or methylation sources, while it only contained one gene (*MC4R*) supported by genetic evidence that was only weakly connected to the remainder of the complex. Furthermore, no enrichment was found for low SNP *P*-values in MAGIC and DIAGRAM data. This might as such be an example of a complex that is rather involved in a response to the diabetic state in the islets than playing a causal role. This is fitting with it being mainly composed of interleukins and toll-like receptors and enriched for inflammatory response and apoptosis pathways that have a clear relevance to the beta-cell mass deterioration in T2D pathogenesis.

Finally, Complex-16 is an example of a complex where the enrichment was supported by multiple sources, but few consensus support genes. Such complexes are interesting because

they could not have been revealed using any data type alone, but constitute a functionally related group of genes that are identified by multiple types of diabetes-associated evidence. Complex-16 was strongly enriched for brain-derived neurotrophin (BDNF) signaling. BDNF has indeed been shown to affect the histological organization of beta and non-beta cells in the pancreatic islets (Yamanaka et al., 2006). This complex was furthermore enriched for GWAS signals for fasting glucose levels, fasting insulin levels and T2D.

## DISCUSSION

To harvest the power of data integration, we have brought together results from genetic studies of islet-relevant phenotypes and human islet studies spanning different levels of molecular regulation. We identify 24 protein complexes with strong supporting evidence for being implicated in diabetes pathogenesis in pancreatic islets and show how they are enriched for multiple modest effects of genetic variants associated with glycemic traits. Furthermore, we specifically prioritize ten candidate genes for T2D, of which six are novel, based on the investigation of either direct or functional convergence of the evidence sources. Additionally, we compose a set of 3,692 islet protein complexes that can serve as an integration scaffold for future studies.

By comparing the direct overlap between the heterogeneous islet diabetes-related gene sets we identified genes such as *ANPEP* and *HADH* that are currently not well-established as diabetes susceptibility genes but had consensus support across evidence sources. These observations highlight that such a straightforward

data integration approach is able to pinpoint potentially new disease genes. Apart from these few, but interesting, examples of genes that were part of multiple gene sets, the generally limited direct overlap between the gene sets emphasizes the necessity of integrative systems biology approaches focusing on functional entities rather than single genes for further understanding of the dysregulation and dysfunctioning occurring in diabetic islets.

Previous work on congenital heart disease (Lage et al., 2012) has shown similar results, where a limited overlap was observed between genes identified in different types of genetic studies whereas they converged significantly in protein networks related to heart development. Here we extended this approach to T2D, where we found the prioritized complexes to mainly be involved in signaling cascades, immune functions, apoptosis and cell-cell communication in addition to the expected insulin secretion pathway. We thus show that these particular molecular mechanisms are consistently supported by complementary types of molecular data from human islets to form a major component of the T2D etiology. These results reduce the many previously observed pathways related to T2D pathogenesis in human and animal islets from single omics studies to a set of highly credible pathways.

A previous systems genetics study of the T2D state in human islets (Taneera et al., 2012) identified a set of 20 genes that collectively explained a significant portion of HbA1c variation. Here we add to those results by combining multiple independent data sets to identify nine additional T2D candidate genes that likely play a role in pancreatic islets. Furthermore, we prioritized specific protein complexes and their associated pathways that provide biological insight into T2D pathogenesis.

The majority of the 24 protein complexes found in this study were enriched for modest GWAS signals, suggesting that multiple small effects collectively perturb the complexes and give rise to variation in glycemic phenotypes. We thus provide insight into the mechanisms by which common genetic variation translates into a disease phenotype, which supports that the multifactorial genetic architecture of complex traits is constituted by a large number of variants disrupting cellular networks (Schadt, 2009).

An advantage to investigating functional convergence on protein complexes is that not all genes in the complex need to have prior diabetes-related evidence for the complex to be significant. Consequently, this approach concurrently prioritizes genes without prior diabetes-related evidence, but whose products interact with other diabetes relevant proteins in the islet, such as the six T2D candidate genes highlighted in **Table 2**. Furthermore, complexes containing both genes from GWAS loci and genes supported by other evidence sources, provide support for the GWAS gene mediating the signal in that locus, such as *LARP6* in the complex Complex-2 that resides in a proinsulin associated GWAS locus. Lastly, the complexes provide a functional context for the disease genes. Many genes naturally participate in several functions, reflected by the overlap of many of the complexes. For such multifunctional genes, the approach outlined here prioritizes the subset of disease relevant complexes and thus the disease relevant functions.

A major goal for T2D and other common diseases is to identify causal pathways and network modules underlying

disease pathogenesis to enable precise risk prediction and development of new therapeutic strategies (McCarthy, 2015). Furthermore, such pathways and network modules need to be identified in a tissue-specific context (Gross and Ideker, 2015). Here we provide causal network modules for T2D in the form of tissue-specific protein complexes that provide more biological insight into the disease pathogenesis than disease genes in isolation and furthermore form a basis for integrating person-specific genetic, transcriptomic, or proteomic profiles in a clinical setting. Dissecting these complexes can moreover reveal new drug-targets, such as genes interacting with targets of currently used anti-diabetic medications, genes supported by multiple evidence sources or their more druggable interaction partners. Furthermore, complexes that contain targets of FDA-approved drugs may highlight opportunities for drug repurposing in the search for new diabetes treatments.

## METHODS

### Construction of a Pancreatic Islet-Specific Protein Interaction Network

Previous tissue-specific protein interaction networks mainly fall into three categories: node-removal, where interactions between proteins absent in the given tissue are excluded (Bossi and Lehner, 2009; Barshir et al., 2013; Basha et al., 2015), edge-reweight, where interactions between absent proteins are down-weighted (Magger et al., 2012), and data-driven Bayesian methodologies (Guan et al., 2012; Basha et al., 2015; Greene et al., 2015). Here we created both an edge-reweighted as well as a node-removal islet-specific protein interaction network, to accommodate downstream network analysis approaches that did or did not consider edge-weights, respectively.

The islet-specific protein interaction networks were constructed by pruning high confidence protein interaction from an updated version (2014) of the InWeb database (Lage et al., 2007; 14,536 proteins with 337,951 high-confidence interactions) using the data sets described in Supplementary Table 7. More specifically, for the node-removal protein interaction network, genes not passing the specified cutoffs in all of the data sets were considered less likely to be expressed in pancreatic islets and thus removed from the pruned islet network. For the edge-reweighted protein interaction network, lowly expressed genes were not removed but instead the confidence score of their interactions were down-weighted using the approach proposed by Magger et al. (2012):

$$w'_{ij} = w_{ij} * rw^n$$

where  $w_{ij}$  is the original edge weight between protein  $i$  and  $j$ ,  $n$  is the number of lowly expressed genes in the tissue constituting the interaction (i.e., {0,1,2}), and  $rw$  is the probability that a gene is expressed in the tissue even though it does not pass the cut-offs listed below, which was chosen to be 0.1.

If genes were not covered by any of the data sets—or in the case of the Human Protein Atlas (HPA) data, annotated with uncertainty—a benefit-of-the-doubt approach was applied where such genes were considered present.

## Included Data-Sources

Tissue-specific protein expression profiles based on immunohistochemistry using tissue microarrays were obtained from the HPA version 13, 11/6-2014, downloaded on 10/3-2015 from [www.proteinatlas.org](http://www.proteinatlas.org), with Ensembl version 75.37 (Uhlén et al., 2015). Proteins were categorized as present, absent, or uncertain based on the reliability and level of their expression value. Specifically, proteins with supportive expression values were categorized as absent if they were not detected and otherwise as present if they had low, medium or high expression values whereas proteins with uncertain expression values were categorized as uncertain.

Microarray gene expression data from the GNF Tissue Atlas (GNF) (GEO: GSE1133) was downloaded from BioGPS (<http://biogps.org/downloads/>; Su et al., 2004).

## Defining Topology-Based Complexes within the Network

Many different methods with different objective functions have been proposed for defining clusters of genes in protein interaction networks. Here we applied two complementary approaches; one aiming at identifying tightly connected genes, and one centered on spoke-hub complexes as often applied in previous work (Lage et al., 2007; Börnigen et al., 2013).

Strongly connected components in the edge-weighted islet-specific network were identified by ClusterONE, a non-partitioning graph decomposition algorithm (Nepusz et al., 2012), using a minimum density of 0.2, which is calculated as the average edge weight within the complex if missing edges are assumed to have a weight of zero, and a maximum overlap of 0.3 between two complexes before they were merged using the multi-merge option, and otherwise default parameters. ClusterONE uses the matching score as default for calculating the overlap between two complexes, which is defined as the intersection size squared, divided by the product of the sizes of the two complexes.

A three-step approach was applied to define spoke-hub-complexes. First, for each gene in the network a complex was defined by all its first order interaction partners. Next a topology filter was applied to prune complexes for interaction partners that tend to interact with many proteins in an unspecific way, due to either experimental artifacts or for biological reasons. In brief, genes were removed from the complex if <5% of its interaction partners were within the given complex. Lastly, overlapping clusters were merged using the same approach as for ClusterONE. Since this approach ignores edge-weights it was applied to the node-removal version of the islet-specific protein interaction network.

Finally, overlapping complexes resulting from the two approaches were merged using the same approach as before. Complexes with fewer than 6 or more than 50 nodes were discarded in the downstream analysis, resulting in 3,692 islet complexes. Diameter and average degree, clustering coefficient and betweenness-centrality were calculated for each complex using the igraph R-package (Csardi and Nepusz, 2006).

## Coordinated Expression of Protein Complexes

The TissueRanker approach (Börnigen et al., 2013) utilizes the assumption that a mutation in a hub-spoke complex is likely to have an affect in tissues where the proteins within the complex show high degree of coordinated expression and thus, that the degree of coordinated expression may aid in prioritizing tissues in which the complex is active and where deregulation of the complex could be detrimental. Here we extended the methodology to complexes with more complex topology. In brief the  $PCC.mean_c^t$  for complex  $c$  in tissue  $t$  is defined as the average pairwise Pearson correlation coefficient ( $PCC$ ) of gene expression (RPKM values) between any two interacting genes within the complex for the given tissue:

$$PCC_{xy}^t = \frac{\sum_{i=1}^{N_s} (x_i^t - \bar{x}^t) (y_i^t - \bar{y}^t)}{\sqrt{\sum_{i=1}^{N_s} (x_i^t - \bar{x}^t)^2} \sqrt{\sum_{i=1}^{N_s} (y_i^t - \bar{y}^t)^2}}$$

$$PCC.mean_c^t = \frac{\sum_{x=1}^{N_g} \sum_{y \in I_x} PCC_{xy}^t}{2 \cdot N_e}$$

where  $N_s$  is the number of samples for tissue  $t$ ,  $N_g$  is the number of genes in protein complex  $c$ ,  $N_e$  is the number of edges in protein complex  $c$ , and  $I_x$  is the interaction partners of gene  $x$  excluding any self-loops.

To alleviate any potential bias arising from different numbers of tissues samples (Börnigen et al., 2013) we further standardized the  $PCC.mean_c^t$  values within a tissue by first converting the average correlation coefficients to an approximately normal distribution using Fisher transformation:

$$z_c^t = \frac{1}{2} \ln \frac{1 + PCC.mean_c^t}{1 - PCC.mean_c^t}$$

$$CE_c^t = \frac{z_c^t - \mu^t}{\sigma^t}$$

The resulting  $z$ -scores are here referred to as coordinated expression (CE) and used to compare tissue relevance across tissues for a given complex.

RPKM values from RNAseq data for 31 tissues from the Genotype-Tissue Expression (GTEx) project were obtained through the database of Genotypes and Phenotypes (dbGaP) (study accession phs000424.v4.p1, version from 17/1-2014; Mailman et al., 2007). However, since the GTEx data does not include pancreatic islets, RNAseq data for whole islets, beta cells, and non-beta cells (from pancreatic islets; Nica et al., 2013) were combined with the GTEx data.

We defined 1,007 islet complexes with coordinated expression as the subset of the 3,692 islet complexes where at least one of the islet tissue components (whole islet, beta, and non-beta cells) was among the three tissues with highest coordinated expression level among the 34 included tissues.

## Compiling Islet Biology and Islet Diabetic Phenotype Related Gene Sets

We compiled a set of 13 complementary sets of genes associated with T2D, monogenic forms of diabetes and related metabolic

phenotypes (Table 1). These 13 gene sets are collectively referred to as islet diabetic phenotype gene sets and were chosen because of their relevance to the islet tissue.

We obtained GWAS SNPs and genes supported by gene-based tests for T2D (Morris et al., 2012; Mahajan et al., 2014), fasting glucose (Dupuis et al., 2010; Scott et al., 2012), 2 hour glucose (2 h glu) during an oral-glucose tolerance test (Dupuis et al., 2010; Scott et al., 2012), and proinsulin (Strawbridge et al., 2011). SNPs in GWAS loci were mapped to a gene if they fell within 110 kb upstream or 40 kb downstream of its transcription start and stop sites respectively, as these boundaries have been shown to capture the majority of *cis*-eQTLs associations (Veyrieras et al., 2008; Ardlie et al., 2015). We additionally included all genes that were reported in eQTL associations for the GWAS SNPs from the respective publications (Dupuis et al., 2010; Voight et al., 2010; Strawbridge et al., 2011; Morris et al., 2012; Scott et al., 2012; Mahajan et al., 2014). We also included genes harboring rare variants associated with either fasting glucose and T2D (Albrechtsen et al., 2013; Flannick et al., 2014; Steinthorsdottir et al., 2014; Wessel et al., 2015) or insulin processing and secretion (Huyghe et al., 2013). Genes associated with monogenic forms of diabetes were obtained from a literature review (McCarthy, 2010) and a curated list from a previous study (Morris et al., 2012).

Genes differentially expressed in islets were obtained from a study by Taneera et al. (2012). In addition, two other microarray datasets of beta-cell and islet gene expression, respectively, were obtained from the Gene Expression Omnibus database (accession IDs: GSE20966 and GSE25724) and differential gene expression between T2D patients and non-diabetic controls evaluated using the “limma” R package as implemented in the NCBI GEO2R tool. Genes with  $P < 0.001$  were included in the gene sets, except for the dataset by Dominguez et al. (2011) where a stricter cutoff of  $P < 0.0001$  was applied due to inflated significance values. We further included additional gene sets defined by the islet gene expression study from Taneera et al. (2012), namely genes that showed *cis*- or *trans*-eQTL associations with T2D associated SNPs and genes that were co-expressed with >2 T2D candidate genes. Finally, we included genes that were differentially methylated in islets in T2D patients compared to non-diabetic controls (Volkmar et al., 2012) or were both differentially methylated and differentially expressed (Dayeh et al., 2014).

We furthermore constructed four gene sets related to islet function, referred to as islet biology gene sets (Table 1). These sets included genes with islet-specific expression (Morán et al., 2012), genes in islet-selective open chromatin regions or genes overlapping clusters of islet-selective open chromatin sites (Gaulton et al., 2010) and genes manually curated as islet important (Pasquali et al., 2014).

Finally, we obtained a list of proteins that are targets of FDA approved drugs from the druggable human proteome (Uhlén et al., 2015).

The direct overlap of the gene sets was tested using a hypergeometric test with all 22,766 human genes as background.

## Functional Convergence Testing

To test the protein complexes for potential dysregulation in T2D, the degree of functional convergence of diabetes related genes was assessed. For each complex, the enrichment of each of the 13 islet diabetic phenotype gene sets was first calculated using a hypergeometric test and the corresponding  $P$ -values were next combined using Fisher's combined probability test ( $P_{combined}$ ).

The likelihood of observing a similar degree of functional convergence by chance was estimated for each complex by randomly sampling 100,000 sets of the same number of genes from the whole network. An empirical  $P$ -value ( $P_{emp}$ ) was then calculated by counting how many of these 100,000 random sets had a  $P_{combined} \leq$  the real case divided by the number of random sets ( $n = 100,000$ ).  $P_{emp}$  was adjusted for multiple hypotheses testing across complexes using a Benjamini–Hochberg correction and complexes with  $P_{emp.adj} < 0.05$  were considered significant and thus showing potential T2D dysregulation. Genes in the gene sets without any interaction partners were excluded from the test.

## Functional Annotation of Protein Complexes

We downloaded 3,906 biological pathways from ConsensusPathDB release 30 (Kamburov et al., 2013). Over-representation analysis of pathways was tested using a hypergeometric test. In short, all gene sets with at least two candidate genes were tested. The background was restricted to the subset of all genes within the protein interaction network that participate in at least one pathway and similarly, only input genes that were part of the background were included for testing.

## Testing for Enrichment of Diabetes-Related GWAS Signal

We further investigated whether the complexes with potential T2D dysregulation were enriched for association with T2D or glycemic traits in 19 different GWA-studies (Supplementary Table 6) using the MAGENTA method (Segrè et al., 2010).

The analysis was repeated using three definitions of complexes: (1) including all genes in the complexes, (2) excluding genes from the complex that had genome-wide significant  $P$ -values ( $P < 5 \times 10^{-8}$ ) in the respective GWAS, i.e., different genes are excluded from the complexes when testing enrichment in the different GWAS studies, and (3) excluding genes that were used for input in the corresponding gene set. For example, all fasting glucose associated genes were excluded from the complexes when testing for enrichment using “Fasting glucose” and “Fasting glucose, Manning” but not when testing for enrichment of e.g., “Fasting insulin”.

In the MAGENTA analysis we used the 95<sup>th</sup> percentile of all gene  $P$ -values as the  $P_{cutoff}$  and SNPs in GWAS loci were mapped to a gene if they fell within 110 kb upstream or 40 kb downstream of its transcription start and stop sites, respectively.

## Statistical Analysis and Visualization

Statistical analyses were performed in the statistical computing language R (R Core Team, 2014) and network visualizations were made in R using the igraph package (Csardi and Nepusz, 2006). Tissue depictions in figures were adapted from Stumvoll et al. (2010).

## AUTHOR CONTRIBUTIONS

HP, VG, and SB conceived the study and provided the initial design and data analysis framework. HP and VG performed the analysis and drafted the original manuscript. HP, VG, and SB contributed to the interpretation and corresponding text. All authors approved the version to be published.

## FUNDING

The Technical University of Denmark has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115317 (DIRECT), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies in kind contribution. The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, is supported financially by the Novo Nordisk Foundation (Grant agreement NNF14CC0001).

## REFERENCES

- Albrechtsen, A., Grarup, N., Li, Y., Sparsø, T., Tian, G., Cao, H., et al. (2013). Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* 56, 298–310. doi: 10.1007/s00125-012-2756-1
- Ardlie, K. G., Deluca, D. S., Segre, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., et al. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648–660. doi: 10.1126/science.1262110
- Barshir, R., Basha, O., Eluk, A., Smoly, I. Y., Lan, A., and Yeger-Lotem, E. (2013). The tissueNet database of human tissue protein-protein interactions. *Nucleic Acids Res.* 41, D841–D844. doi: 10.1093/nar/gks1198
- Basha, O., Flom, D., Barshir, R., Smoly, I., Tirman, S., and Yeger-Lotem, E. (2015). MyProteinNet: build up-to-date protein interaction networks for organisms, tissues and user-defined contexts. *Nucleic Acids Res.* 43, W258–W263. doi: 10.1093/nar/gkv515
- Bonnefond, A., and Froguel, P. (2015). Rare and common genetic events in type 2 diabetes: what should biologists know? *Cell Metab.* 21, 357–368. doi: 10.1016/j.cmet.2014.12.020
- Börnigen, D., Pers, T. H., Thorrez, L., Huttenhower, C., Moreau, Y., and Brunak S. (2013). Concordance of gene expression in human protein complexes reveals tissue specificity and pathology. *Nucleic Acids Res.* 41:e171. doi: 10.1093/nar/gkt661
- Bossi, A., and Lehner, B. (2009). Tissue specificity and the human protein interaction network. *Mol. Syst. Biol.* 5, 260. doi: 10.1038/msb.2009.17
- Couvelard, A., Hu, J., Steers, G., O'Toole, D., Sauvanet, A., Belghiti, J., et al. (2006). Identification of potential therapeutic targets by gene-expression profiling in pancreatic endocrine tumors. *Gastroenterology* 131, 1597–1610. doi: 10.1053/j.gastro.2006.09.007
- Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *Inter J. Complex Syst.* 1695.
- Danielsson, A., Pontén, F., Fagerberg, L., Hallström, B. M., Schwenk, J. M., Uhlen, M., et al. (2014). The human pancreas proteome defined by transcriptomics and antibody-based profiling. *PLoS ONE* 9:e115421. doi: 10.1371/journal.pone.0115421
- Dayeh, T., Volkov, P., Saló, S., Hall, E., Nilsson, E., Olsson, A. H., et al. (2014). Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet.* 10:e1004160. doi: 10.1371/journal.pgen.1004160

## ACKNOWLEDGMENTS

The authors wish to thank Kirstine GI Belling for critical comments to the manuscript and Jose MG Izarzugaza and Jessica Xin Hu for providing cancer gene lists. The Genotype-Tissue Expression (GTEx) dataset used for the analyses described in this manuscript was obtained from dbGaP at [www.ncbi.nlm.nih.gov/gap](http://www.ncbi.nlm.nih.gov/gap) through dbGaP accession number phs000424.v4.p1. Data on glycemic traits has been contributed by MAGIC investigators and have been downloaded from [www.magicinvestigators.org](http://www.magicinvestigators.org).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fgene.2017.00043/full#supplementary-material>

## Data Availability

The 24 complexes with potential T2D dysregulation are provided in supplementary Data Sheet 2 and further visualized in Data Sheet 1.

- de Lichtenberg, U., Jensen, L. J., Brunak, S., and Bork, P. (2005). Dynamic complex formation during the yeast cell cycle. *Science* 307, 724–727. doi: 10.1126/science.1105103
- Dominguez, V., Raimondi, C., Somanath, S., Bugliani, M., Loder, M. K., Edling, C. E., et al. (2011). Class II phosphoinositide 3-kinase regulates exocytosis of insulin granules in pancreatic  $\beta$  cells. *J. Biol. Chem.* 286, 4216–4225. doi: 10.1074/jbc.M110.200295
- Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., et al. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* 42, 105–116. doi: 10.1038/ng.520
- Fadista, J., Vikman, P., Ottosson, E., Guerra, I., Lou, J., and Taneera, J. (2014). Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 111, 13924–13929. doi: 10.1073/pnas.1402665111
- Flannick, J., Thorleifsson, G., Beer, N. L., Jacobs, S. B. R., Grarup, N., Burt, N. P., et al. (2014). Loss-of-function mutations in *SLC30A8* protect against type 2 diabetes. *Nat. Genet.* 46, 357–363. doi: 10.1038/ng.2915
- Forbes, S. A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 45, D777–D783. doi: 10.1093/nar/gkw1121
- Ganegoda, G., Wang, J., Wu, F.-X., and Li, M. (2014). Prediction of disease genes using tissue-specified gene-gene network. *BMC Syst. Biol.* 8(Suppl. 3):S3. doi: 10.1186/1752-0509-8-S3-S3
- Gaulton, K. J., Nammo, T., Pasquali, L., Simon, J. M., Giresi, P. G., Fogarty, M. P., et al. (2010). A map of open chromatin in human pancreatic islets. *Nat. Genet.* 42, 255–259. doi: 10.1038/ng.530
- Glaser, B. (2013). *Familial Hyperinsulinism*. Seattle, WA: University of Washington. Available online at: <http://www.ncbi.nlm.nih.gov/books/NBK1375/>
- Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M. P., Jene-Sanz, A., et al. (2013). IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* 10, 1081–1082. doi: 10.1038/nmeth.2642
- Greene, C. S., Krishnan, A., Wong, A. K., Ricciotti, E., Zelaya, R. A., Himmelstein, D. S., et al. (2015). Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* 47, 569–576. doi: 10.1038/ng.3259
- Gross, A. M., and Ideker, T. (2015). Molecular networks in context. *Nat. Biotechnol.* 33, 720–721. doi: 10.1038/nbt.3283
- Guan, Y., Gorenshyeyn, D., Burmeister, M., Wong, A. K., Schimenti, J. C., Handel, M. A., et al. (2012). Tissue-specific functional networks for

- prioritizing phenotype and disease genes. *PLoS Comput. Biol.* 8:e1002694. doi: 10.1371/journal.pcbi.1002694
- Han, J.-D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., et al. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430, 88–93. doi: 10.1038/nature02555
- Huyghe, J. R., Jackson, A. U., Fogarty, M. P., Buchkovich, M. L., Stančáková, A., Stringham, H. M., et al. (2013). Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.* 45, 197–201. doi: 10.1038/ng.2507
- Kamburov, A., Stelzl, U., Lehrach, H., and Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.* 41, D793–D800. doi: 10.1093/nar/gks1055
- Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., et al. (2014). A draft map of the human proteome. *Nature* 509, 575–581. doi: 10.1038/nature13302
- Kodama, K., Horikoshi, M., Toda, K., Yamada, S., Hara, K., Irie, J., et al. (2012). Expression-based genome-wide association study links the receptor *CD44* in adipose tissue with type 2 diabetes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7049–7054. doi: 10.1073/pnas.1114513109
- Kooner, J. S., Saleheen, D., Sim, X., Sehmi, J., Zhang, W., Frossard, P., et al. (2011). Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat. Genet.* 43, 984–989. doi: 10.1038/ng.921
- Lage, K., Greenway, S. C., Rosenfeld, J. A., Wakimoto, H., Gorham, J. M., Segrè, A. V., et al. (2012). Genetic and environmental risk factors in congenital heart disease functionally converge in protein networks driving heart development. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14035–14040. doi: 10.1073/pnas.1210730109
- Lage, K., Hansen, N. T., Karlberg, E. O., Eklund, A. C., Roque, F. S., Donahoe, P. K., et al. (2008). A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20870–20875. doi: 10.1073/pnas.0810772105
- Lage, K., Karlberg, E. O., Størling, Z. M., Olason, P. I., Pedersen, A. G., Rigina, O., et al. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* 25, 309–316. doi: 10.1038/nbt1295
- Lawlor, N., George, J., Bolisetty, M., Kursawe, R., Sun, L., Sivakamasundari, V., et al. (2017). Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res.* 27, 208–222. doi: 10.1101/gr.212720.116
- Li, D. (2012). Diabetes and pancreatic cancer. *Mol. Carcinog.* 51, 64–74. doi: 10.1002/mc.20771
- Locke, J. M., Hysenaj, G., Wood, A. R., Weedon, M. N., and Harries, L. W. (2015). Targeted allelic expression profiling in human islets identifies cis-regulatory effects for multiple variants identified by type 2 diabetes genome-wide association studies. *Diabetes* 64, 1484–1491. doi: 10.2337/db14-0957
- Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., et al. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 45, 580–585. doi: 10.1038/ng.2653
- Magger, O., Waldman, Y. Y., Ruppin, E., and Sharan, R. (2012). Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS Comput. Biol.* 8:e1002690. doi: 10.1371/journal.pcbi.1002690
- Mahajan, A., Go, M. J., Zhang, W., Below, J. E., Gaulton, K. J., Ferreira, T., et al. (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* 46, 234–244. doi: 10.1038/ng.2897
- Mailman, M., Feolo, M., Jin, Y., and Kimura, M. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* 39, 1181–1186. doi: 10.1038/ng1007-1181
- Marselli, L., Thorne, J., Dahiya, S., Sgroi, D. C., Sharma, A., Bonner-Weir, S., et al. (2010). Gene expression profiles of Beta-cell enriched tissue obtained by laser capture microdissection from subjects with type 2 diabetes. *PLoS ONE* 5:e11499. doi: 10.1371/journal.pone.0011499
- McCarthy, M. I. (2010). Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.* 363, 2339–2350. doi: 10.1056/NEJMra0906948
- McCarthy, M. I. (2015). Genomic medicine at the heart of diabetes management. *Diabetologia* 58, 1725–1729. doi: 10.1007/s00125-015-3588-6
- McCulloch, L. J., van de Bunt, M., Braun, M., Frayn, K. N., Clark, A., and Gloyn, A. L. (2011). GLUT2 (*SLC2A2*) is not the principal glucose transporter in human pancreatic beta cells: implications for understanding genetic association signals at this locus. *Mol. Genet. Metab.* 104, 648–653. doi: 10.1016/j.ymgme.2011.08.026
- Morán, I., Akerman, I., van de Bunt, M., Xie, R., Benazra, M., Nammo, T., et al. (2012). Human  $\beta$  cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab.* 16, 435–448. doi: 10.1016/j.cmet.2012.08.010
- Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segrè, A. V., Steinthorsdottir, V., et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* 44, 981–990. doi: 10.1038/ng.2383
- Natalicchio, A., Labarbuta, R., Tortosa, F., Biondi, G., Marrano, N., Peschechera, A., et al. (2013). Exendin-4 protects pancreatic beta cells from palmitate-induced apoptosis by interfering with GPR40 and the MKK4/7 stress kinase signalling pathway. *Diabetologia* 56, 2456–2466. doi: 10.1007/s00125-013-3028-4
- Nepusz, T., Yu, H., and Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods* 9, 471–472. doi: 10.1038/nmeth.1938
- Nica, A. C., Ongen, H., Irminger, J.-C., Bosco, D., Berney, T., Antonarakis, S. E., et al. (2013). Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Res.* 23, 1554–1562. doi: 10.1101/gr.150706.112
- Owusu, D., Pan, Y., Xie, C., Harirforoosh, S., and Wang, K.-S. (2017). Polymorphisms in *PDLIM5* gene are associated with alcohol dependence, type 2 diabetes, and hypertension. *J. Psychiatr. Res.* 84, 27–34. doi: 10.1016/j.jpsychires.2016.09.015
- Pasquali, L., Gaulton, K. J., Rodríguez-Seguí, S. A., Mularoni, L., Miguel-Escalada, I., Akerman, I., et al. (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* 46, 136–143. doi: 10.1038/ng.2870
- Pers, T. H., Dworzynski, P., Thomas, C. E., Lage, K., and Brunak, S. (2013). MetaRanker 2.0: a web server for prioritization of genetic variation data. *Nucleic Acids Res.* 41, W104–W108. doi: 10.1093/nar/gkt387
- Prasad, R., and Groop, L. (2015). Genetics of type 2 diabetes: pitfalls and possibilities. *Genes* 6, 87–123. doi: 10.3390/genes6010087
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Austria: R Foundation for Statistical Computing.
- Schadt, E. E. (2009). Molecular networks as sensors and drivers of common human diseases. *Nature* 461, 218–223. doi: 10.1038/nature08454
- Scott, R. A., Lagou, V., Welch, R. P., Wheeler, E., Montasser, M. E., Luan, J., et al. (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* 44, 991–1005. doi: 10.1038/ng.2385
- Segerstolpe, Å., Palasantza, A., Eliasson, P., Andersson, E.-M., Andréasson, A.-C., Sun, X., et al. (2016). Single-cell transcriptome profiling of human pancreatic islets in health and Type 2 diabetes. *Cell Metab.* 24, 593–607. doi: 10.1016/j.cmet.2016.08.020
- Segrè, A. V., Groop, L., Mootha, V. K., Daly, M. J., and Altshuler, D. (2010). Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* 6:e58. doi: 10.1371/journal.pgen.1001058
- Shin, S.-Y., Fauman, E. B., Petersen, A.-K., Krumsiek, J., Santos, R., Huang, J., et al. (2014). An atlas of genetic influences on human blood metabolites. *Nat. Genet.* 46, 543–550. doi: 10.1038/ng.2982
- Steinthorsdottir, V., Thorleifsson, G., Sulem, P., Helgason, H., Grarup, N., Sigurdsson, A., et al. (2014). Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat. Genet.* 46, 294–298. doi: 10.1038/ng.2882
- Strawbridge, R. J., Dupuis, J., Prokopenko, I., Barker, A., Ahlqvist, E., Rybin, D., et al. (2011). Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* 60, 2624–2634. doi: 10.2337/db11-0415

- Stumvoll, M., Goldstein, B. J., and van Haeften, T. W. (2010). Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 365, 1333–1346. doi: 10.1016/S0140-6736(05)61032-X
- Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6062–6067. doi: 10.1073/pnas.0400782101
- Taneera, J., Lang, S., Sharma, A., Fadista, J., Zhou, Y., Ahlqvist, E., et al. (2012). A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. *Cell Metab.* 16, 122–134. doi: 10.1016/j.cmet.2012.06.006
- Taylor, I. W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., et al. (2009). Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat. Biotechnol.* 27, 199–204. doi: 10.1038/nbt.1522
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Tissue-based map of the human proteome. *Science* 347:394. doi: 10.1126/science.1260419
- van Hove, E. C., Hansen, T., Dekker, J. M., Reiling, E., Nijpels, G., Jorgensen, T., et al. (2006). The *HADHSC* gene encoding short-chain L-3-hydroxyacyl-CoA dehydrogenase (SCHAD) and type 2 diabetes susceptibility: the Damage study. *Diabetes* 55, 3193–3196. doi: 10.2337/db06-0414
- Veyrieras, J.-B., Kudaravalli, S., Kim, S. Y., Dermitzakis, E. T., Gilad, Y., Stephens, M., et al. (2008). High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* 4:e1000214. doi: 10.1371/journal.pgen.1000214
- Voight, B. F., Scott, L. J., Steinthorsdottir, V., Morris, A. P., Dina, C., Welch, R. P., et al. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* 42, 579–589. doi: 10.1038/ng.609
- Volkmar, M., Dedeurwaerder, S., Cunha, D. A., Ndlovu, M. N., Defrance, M., Deplus, R., et al. (2012). DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. *EMBO J.* 31, 1405–1426. doi: 10.1038/emboj.2011.503
- Wang, Y. J., Schug, J., Won, K.-J., Liu, C., Naji, A., Avrahami, D., et al. (2016). Single-cell transcriptomics of the human endocrine pancreas. *Diabetes* 65, 3028–3038. doi: 10.2337/db16-0405
- Wessel, J., Chu, A. Y., Willems, S. M., Wang, S., Yaghootkar, H., Brody, J. A., et al. (2015). Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat. Commun.* 6, 5897. doi: 10.1038/ncomms6897
- Xia, Z., Dickens, M., Raingeaud, J., Davis, R. J., and Greenberg, M. E. (1995). Opposing effects of ERK and JNK-p38 MAP kinases on apoptosis. *Science* 270, 1326–1331.
- Xie, T., Chen, M., Gavrilova, O., Lai, E. W., Liu, J., and Weinstein, L. S. (2008). Severe obesity and insulin resistance due to deletion of the maternal *Gsa* allele is reversed by paternal deletion of the *Gsa* imprint control region. *Endocrinology* 149, 2443–2450. doi: 10.1210/en.2007-1458
- Xin, Y., Kim, J., Okamoto, H., Ni, M., Wei, Y., Adler, C., et al. (2016). RNA Sequencing of single human islet cells reveals Type 2 diabetes genes. *Cell Metab.* 24, 608–615. doi: 10.1016/j.cmet.2016.08.018
- Yamanaka, M., Itakura, Y., Inoue, T., Tsuchida, A., Nakagawa, T., Noguchi, H., et al. (2006). Protective effect of brain-derived neurotrophic factor on pancreatic islets in obese diabetic mice. *Metab. Clin. Exp.* 55, 1286–1292. doi: 10.1016/j.metabol.2006.04.017
- Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A. E., Lee, S. H., et al. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* 47, 1114–1120. doi: 10.1038/ng.3390

**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.

Copyright © 2017 Pedersen, Gudmundsdottir and Brunak. 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.