ARTICLE   OPEN

Cross-species gene expression analysis identifies a novel set of genes implicated in human insulin sensitivity

Rima Chaudhuri[1,2], Poh Sim Khoo[2], Katherine Tonks[2,3], Jagath R Junutula[4,10], Ganesh Kolumam[4], Zora Modrusan[4], Dorit Samocha-Bonet[2,5], Christopher C Meoli[2], Samantha Hocking[6,9], Daniel J Fazakerley[1,2], Jacqueline Stöckli[1,2], Kyle L Hoehn[7], Jerry R Greenfield[2,3,5], Jean Yee Hwa Yang[8] and David E James[1,2,9]

OBJECTIVE: Insulin resistance (IR) is one of the earliest predictors of type 2 diabetes. However, diagnosis of IR is limited. High fat fed mouse models provide key insights into IR. We hypothesized that early features of IR are associated with persistent changes in gene expression (GE) and endeavored to (a) develop novel methods for improving signal:noise in analysis of human GE using mouse models; (b) identify a GE motif that accurately diagnoses IR in humans; and (c) identify novel biology associated with IR in humans.
METHODS: We integrated human muscle GE data with longitudinal mouse GE data and developed an unbiased three-level cross-species analysis platform (single gene, gene set, and networks) to generate a gene expression motif (GEM) indicative of IR. A logistic regression classification model validated GEM in three independent human data sets ($n = 115$).
RESULTS: This GEM of 93 genes substantially improved diagnosis of IR compared with routine clinical measures across multiple independent data sets. Individuals misclassified by GEM possessed other metabolic features raising the possibility that they represent a separate metabolic subclass. The GEM was enriched in pathways previously implicated in insulin action and revealed novel associations between β-catenin and Jak1 and IR. Functional analyses using small molecule inhibitors showed an important role for these proteins in insulin action.
CONCLUSIONS: This study shows that systems approaches for identifying molecular signatures provides a powerful way to stratify individuals into discrete metabolic groups. Moreover, we speculate that the β-catenin pathway may represent a novel biomarker for IR in humans that warrant future investigation.

INTRODUCTION

Insulin resistance (IR) and obesity are major risk factors for a range of diseases including type 2 diabetes (T2D) and cardiovascular disease. However, some obese people are insulin sensitive as lean individuals,[1–3] whereas a subgroup of IR individuals are lean.[4,5] Hence, there is a need to improve diagnosis of IR because lifestyle changes can prevent or delay such disease onset. However, this is a challenge as IR persists for many years without obvious perturbation in glucose homeostasis and so IR individuals may not present to their physician before T2D or cardiovascular disease onset. This is compounded by the fact that current clinical measures such as body mass index, waist/hip ratio, and fasting glucose have limited diagnostic value.

Most accurate methods for assessing IR such as the hyperinsulinemic-euglycaemic clamp[6] or surrogate measures such as liver lipid analysis are not feasible in routine clinical use. Metabolite biomarkers that correlate with insulin sensitivity and T2D have been reported,[7–11] but their diagnostic value is unclear. Gene expression (GE) represents a potential alternative for measuring disease risk as it represents an integrated response of multiple system parameters. For example, where the functional consequence of a single-nucleotide polymorphism (SNP) is altered transcription, GE can accurately predict an individual's SNP genotype,[12] which may be the true determinant of disease risk.

Although gene expression has been explored as a biomarker for IR this has been limited by the inability to detect large numbers of differentially expressed genes.[13] This could be explained by a range of factors: IR could represent a multitude of diseases; GE could be too dynamic with an underlying rhythmicity that masks IR-specific changes; technical variance may exceed biological variance. This requires additional approaches to overcome these potential barriers. One approach to circumvent variation in human data is the use of cross-species data integration.[14–16] The rationale here is that model systems can be used to recapitulate human diseases with much less variance than is found in human data sets. Robust changes in GE in such model systems can then be tested as a positive training set in human data. This will reveal shared features between the model system and human disease and aid identification of human disease-associated signatures that would not have been isolated from analysis of human data alone.

High fat diet (HFD) fed rodents are widely used as models for human IR.[17–20] As in humans, exposure of rodents to HFD results in increased adiposity, glucose intolerance, hyperinsulinemia and

[1]Charles Perkins Centre, School of Molecular Bioscience, The University of Sydney, Sydney, NSW, Australia; [2]Diabetes and Obesity Program, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia; [3]Department of Endocrinology and Diabetes Centre, St Vincent's Hospital, Sydney, NSW, Australia; [4]Genentech Incorporated, South San Francisco, CA, USA; [5]Faculty of Medicine, University of New South Wales, Sydney, NSW, Australia; [6]Department of Endocrinology, Royal North Shore Hospital, Sydney, NSW, Australia; [7]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia; [8]School of Mathematics and Statistics, University of Sydney, Sydney, NSW, Australia and [9]School of Medicine, The University of Sydney, Sydney, NSW, Australia.
Correspondence: JYH Yang (jean.yang@sydney.edu.au) or DE James (david.james@sydney.edu.au)
[10]Current address: Cellerant Therapeutics, Inc., 1561 Industrial Rd, San Carlos, CA 94070, USA
Received 8 December 2014; revised 24 July 2015; accepted 24 August 2015

peripheral IR in muscle, adipose tissue, and liver.[17–24] This is comparable to phenotypes seen in prediabetic humans.

We have conducted a proof of principle study to evaluate the feasibility of developing a molecular profile using cross-species muscle GE analysis that can accurately diagnose IR in humans. We describe a gene expression motif (GEM) comprising 93 genes that substantially improves IR diagnosis when tested in 115 individuals across 3 separate studies. In addition, we identified pathways including β-catenin (Wnt Signaling) and Jak1 (Jak-Stat Signaling) to be up-regulated in muscle biopsies of insulin resistant individuals. Inhibition of these proteins in muscle cells showed significant defects in glucose uptake indicating an important role in muscle insulin action.

MATERIALS AND METHODS

The following sections are contained within the Supplementary Information.

Study design
Supplementary Methods 1.1.

Human subject and mouse phenotype studies
Supplementary Methods 1.2.1.

Microarray studies
Supplementary Methods 1.2.2-1.2.5.

Statistical analyses
Supplementary Methods 1.3-1.9.

Experimental validation
Supplementary Methods 1.10.

R-code for data analysis
(SI.zip: Main_Analysis_Chaudhuri_et_al.zip).

RESULTS

Our goal was to identify a GE signature that could accurately diagnose whole body insulin sensitivity. This is important because IR is one of the earliest determinants of T2D. We designed our study with the following innovative attributes: use of state of the art methods to define metabolic status including a hyperinsulinemic glucose clamp (ClampIS), the gold standard for measuring whole body insulin sensitivity, and liver lipid fat; while IR is associated with obesity many obese people are not IR and so we wanted to identify a signature that predicts IR independently of obesity. Thus, we studied individuals who were either obese insulin sensitive or matched for adiposity with IR; we sought a signature that predicted IR and not T2D *per se* because of the complications associated with T2D like hyperglycemia. Because of inherent variance in human GE data we first identified IR-associated genes and pathways in a mouse IR model and used these data in a weighted analysis of the human GE data (see Figure 1 for workflow summary and Supplementary Methods (SM) 1.1 for details).
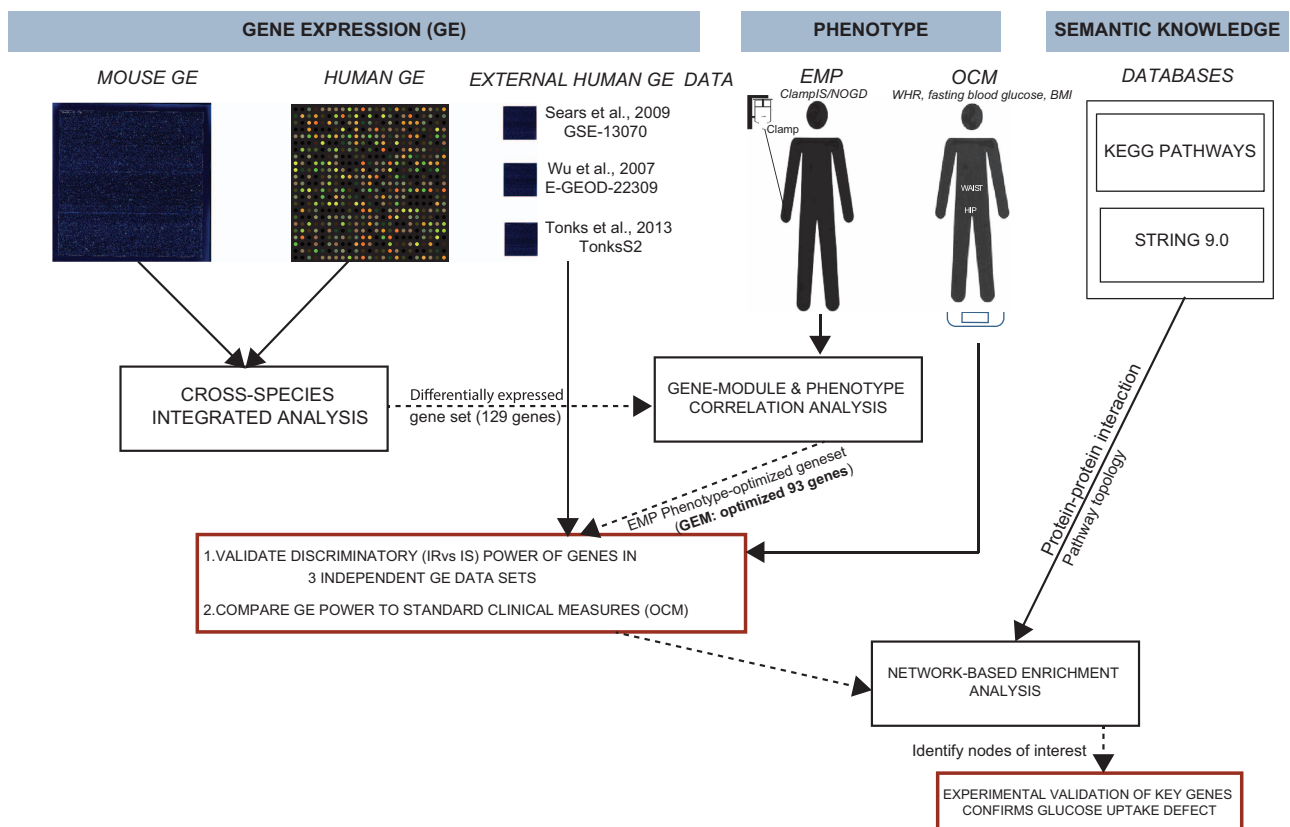

Figure 1. Study workflow. Diagram shows the gradual inclusion of different types of data sets (gene expression, human clinical data such as office-based clinical measures (OCM) and extensive metabolic phenotypic parameters (EMP), external gene expression data sets from publically available databases, biological pathways and protein–protein interaction networks) into the workflow at different stages of the novel analysis approach presented here and their respective outcomes.

Five days of HFD was sufficient to cause glucose intolerance in mice[24] (Supplementary Figure S1A and B). Muscle GE was measured after 0, 5 and 42 d of HFD feeding. IR develops rapidly in liver[24] and more slowly in muscle, reaching a plateau by 1 or 3 wks on HFD, respectively. Hence, this time course of HFD feeding was selected to enable identification of genes that do not just correlate with IR but that may also be causal. A cohort of 81

age-matched humans was selected to comprise four distinct groups: obese insulin sensitive (OIS), obese IR (OIR), T2D and lean.[25] Human GE data was derived from muscle biopsies of a subset of these individuals. OIS, OIR, T2D groups had higher adiposity, body mass index, waist/hip ratio and visceral adiposity than lean subjects.[25] Lean and OIS groups were more insulin sensitive than OIR and T2D groups, measured by the clamp or
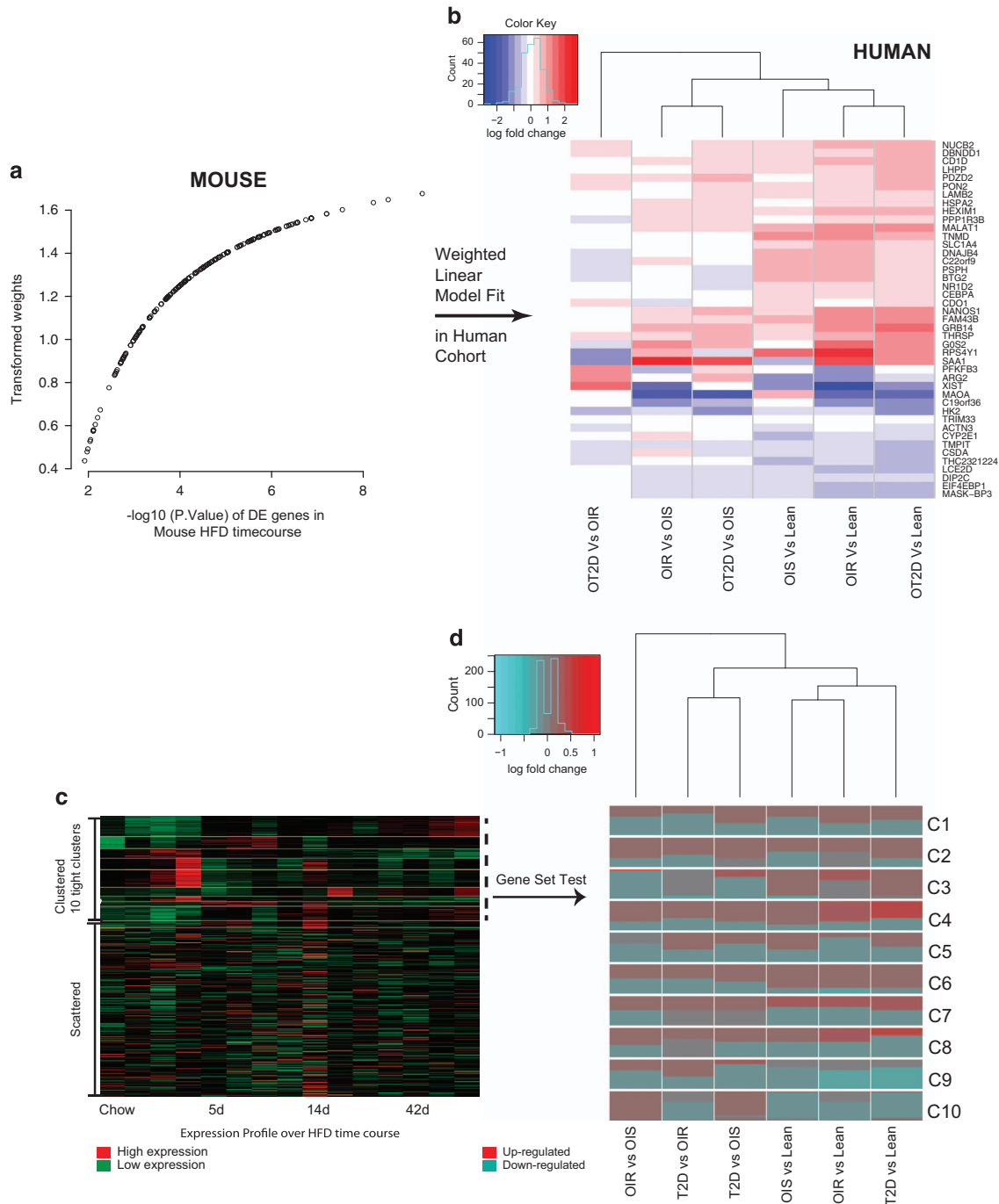

Figure 2. Single gene and gene-set cross-species analysis reveals 129 genes implicated in IR and/or obesity. Red indicates upregulation and blue indicates down-regulation of genes. (**a**) Transformation of statistical significance (*P* value) of genes from mouse DE analysis to standardized weights (**b**) the fold change (log2 scale) of genes across the different comparisons from the single gene-by-gene analysis in humans. Fold change $\geqslant 1.5$, significance of 0.05 and average gene expression > 7 was used for identifying DE genes from the human analysis. (**c**) Tight clustering of co-regulated genes in the mouse time course GE data identified 10 tight clusters (**d**) shows the proportion, magnitude, and direction of regulation of these 10 gene clusters from mouse GE data in humans across the six group comparisons. All heatmaps were generated using R package gplots (http://CRAN.R-project.org/package=gplots). DE, differentially expressed; GE, gene expression; IR; insulin resistance.

non-oxidative glucose disposal (NOGD), a semi quantitative estimation of glycogen synthesis, and fasting insulin (Supplementary Figure S1C–K and SM 1.2; in-silico regression analysis is able to reproduce this as shown in Supplementary Figure S2).

Cross-species integrated analysis identifies 129 genes implicated in IR

It has been previously reported that there is limited signal in GE data from human muscle from which to generate a signature of muscle IR.[13] Therefore, we set out to use a cross-species integration approach, using GE data from a mouse model of insulin resistance to guide analysis of the human GE data. A weighted analytic approach was performed to integrate human-mouse GE data, whereby differentially expressed mouse genes were assigned prior importance and tested for differential expression in human data. Two stages of analysis were implemented to identify differentially expressed genes under conditions of IR/T2D and obesity: (a) individual level analysis, through a standalone human GE analysis, a weighted linear model where significantly altered mouse genes are tested in human data and MA-plots in human GE. Supplementary Results (SR) 1.1A and

SR 1.2 provides details for all 43 genes altered at the single gene level in humans (Figure 2a, b); (b) group level assessment, such as Gene Set Test on human GE profiles using co-regulated gene clusters from mouse data (Figure 2c, d). Further details can be found in SR 1.1B. These analyses resulted in identification of 129 differentially expressed genes (43 from (a) and 86 unique genes from (b)) (Supplementary Data S1). Pathway analysis of this gene set revealed an enrichment of the insulin signaling pathway and other pathways linked to insulin action and IR, including PPAR signaling,[26,27] sphingolipid and ceramide metabolism,[28] amino acid metabolism,[11] and TCA cycle[29] (Supplementary Data S3).

Optimization

We performed weighted gene co-expression network analysis[30,31] to optimize these 129 genes by defining modules/clusters with similar expression patterns, and relating them to benchmark phenotypic traits for OIR and OIR individuals only, leading to identification of 12 distinct clusters (Figure 3a). We next considered how well these 12 modules correlated with metabolic traits: ClampIS/NOGD, which is quantitative measure of whole body insulin sensitivity, visceral adiposity and liver fat[32] correlate with insulin resistance, and Homeostatsis Model
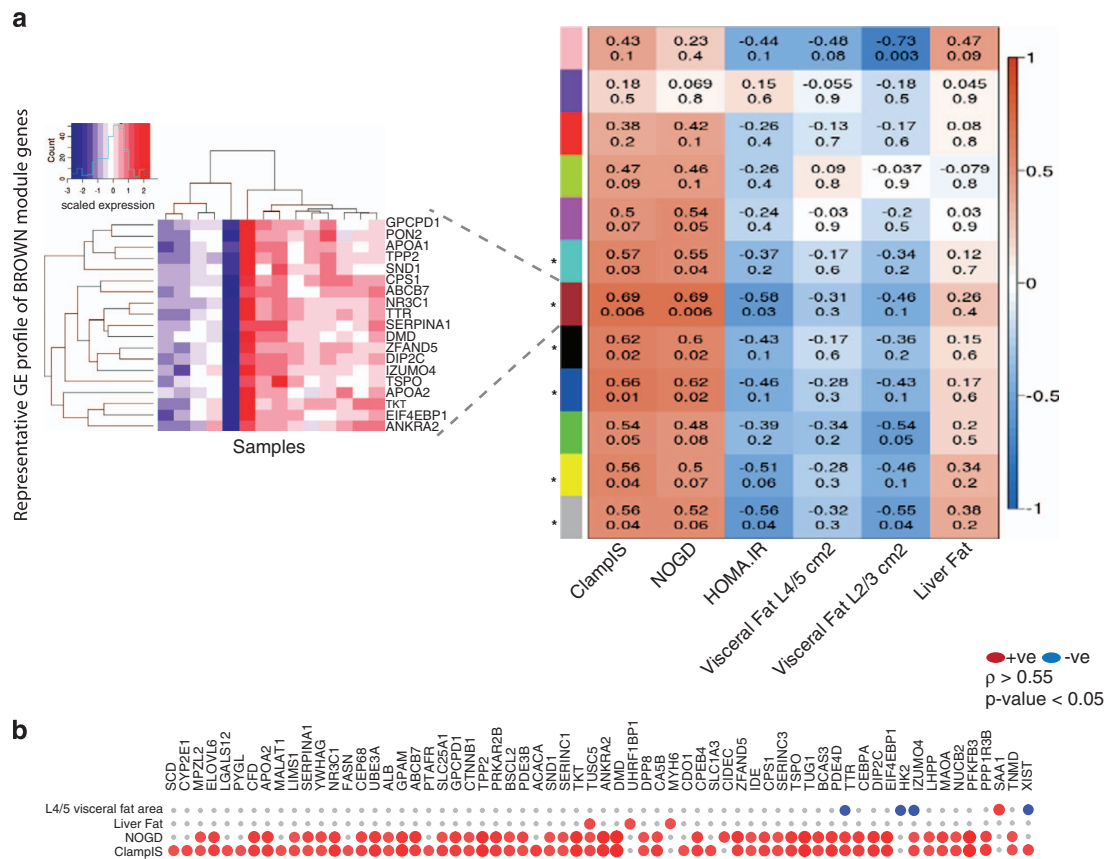

Figure 3. Gene-network and single gene level correlation analysis between clinical EMP measures and the gene sets obtained from our integrative analysis. Red indicates positive correlation and blue indicates negative correlation. (**a**) The 12 gene modules obtained from clustering the 129 genes using weighted gene co-expression network analysis are represented through the 12 colors that index the rows to the matrix (right). Correlation coefficient (value on top) and the P value significance (bottom value) of each of the 12 clusters with four clinical measures (ClampIS, NOGD, HOMA-IR, visceral fat (L2/3 and L4/5 cm^2) and liver fat) are shown. Six clusters (turquoise: 25 genes, brown: 19 genes, black: 10 genes, blue: 20 genes, yellow: 13 genes and gray: 12 genes) that significantly correlate (i.e., $\rho > 0.55$ and P value < 0.05) with the EMP measures such as ClampIS and NOGD are highlighted (*). The brown module is expanded to list the genes contained within this module and their GE pattern across the obese insulin sensitive and obese insulin resistant (OIS and OIR) subjects in the heatmap (left). (**b**) The magnitude of correlation is shown by the size of the circles of the 64 single genes within 129 DE gene set that reached significance ($\rho > 0.55$ and P value < 0.05) with the clinical measures of interest, as shown in the correlation plot generated using package corrplot in R. EMP, extensive metabolic phenotypic parameters; GE, gene expression.

Assessment - Insulin Resistance (HOMA-IR)[33] is a crude measure of insulin sensitivity. Six gene modules comprised of 90 genes were positively correlated with ClampIS/NOGD (* in Figure 3a). No module correlated with liver fat. We also calculated the individual gene-phenotype correlations for each gene; 64 were significantly associated with at least one clinical trait (Figure 3b, Supplementary Data File S1) and 48 correlated with whole body insulin sensitivity measured by clamp (details in SR 1.3).

Overall, we identified six candidate gene modules from the weighted gene co-expression network analysis and five individual genes (*DMD*, *MAOA*, *PFKFB3*, *XIST* and *SLC1A4*) from a linear regression analysis using ClampIS as response variable[34] (SR 1.3). This subset of genes was closely related to whole body insulin sensitivity for OIS and OIR subjects. This optimized set of 93 genes is referred to as GEM. The functional enrichment of GEM is shown in Supplementary Data S3.

Validation

To quantify the utility of the information contained within the GEM as a diagnostic for IR, we undertook a systematic validation approach to measure its performance as a signature to segregate individuals based on their whole body insulin sensitivity (see Supplementary Methods (SM) 1.7–1.8). The prediction accuracy of

Figure 4. Comparative evaluation of the diagnostic power of metabolic status by the gene sets obtained from our integration approach with routine clinical measures in internal and external GE data. The performance ratio of the three gene signatures (GEM, T2DKEGG, and genes from Väremo *et al.*) in three external GE data sets (*n* = 115) when compared with office-based clinical measures (fasting blood sugar level, waste hip ratio, and BMI), more sophisticated measures such as visceral fat, liver fat, and HOMA-IR and extensive metabolic phenotypic (EMP) measures of insulin sensitivity such as ClampIS and NOGD are shown. The median performance ratios reflect the performance of each gene set or clinical variables in classifying IR from IS individuals along with their confidence limits. The performance ratio is directly proportional to the classification accuracy of the features; value of 1 indicates random chance. BMI. body mass index; GE, gene expression; GEM, gene expression motif; IR; nsulin resistance; NOGD, non-oxidative glucose disposal.

whole body insulin sensitivity measured by clamp is 100%, as the OIS and the OIR individuals were segregated based on these criteria, and the random outcome is ~50%. Thus, the expected performance ratio (prediction accuracy/random) of the clamp in classifying OIR from OIS people is ~2 (i.e., 100/50), (Figure 4) whereas a ratio of 1 corresponds to random chance. In our simulations, whole body insulin sensitivity (ClampIS/NOGD) reached a median performance ratio of 1.94. The routine clinical measures (waist/hip ratio, blood sugar level and body mass index) had a median ratio of ~1.2, while the more sophisticated clinical measures liver density, visceral fat and HOMA-IR had a value of ~1.33. As the GEM was derived from the TonksS1 GE data we tested its prediction accuracy in TonksS2, an additional cohort of GE data from the same original study[25] comprised of a total of 16 individuals and the GEM prediction ratio was ~1.6, which was a considerable improvement over office-based clinical measures. To extend the validation of GEM we examined its ability to diagnose IR in 2 external GE data sets. The performance median ratio score of GEM in both external GE data sets was ~1.65. It is highly significant that GEM was highly diagnostic outside of the training dataset and that it outperformed alternate measures of insulin sensitivity (i.e., visceral fat and liver fat).

We next compared the performance of the GEM to two external gene signatures: (i) a publically available T2D gene set that was assembled by literature curation; and (ii) an external set of 12 T2D biomarkers from muscle recently reported by Väremo *et al.*,[35] GEM outperformed both of these signatures in distinguishing between OIR and OIS subjects (Figure 4). The median values for the diagnostic accuracy for all three gene signatures across the cross-validation rounds in the 3 GE data sets (Supplementary Table S1; external datasets described in Supplementary Data File S2) highlights the substantial difference between the gene-signature based median accuracy and random diagnosis in all data sets. There was good individual prediction concordance between GEM and Väremo *et al.*, whereas T2DKEGG predicted poorly in two out three GE data sets. There was only one common gene (HK2) between the 45 T2DKEGG gene set and the 93 GEM genes and pathway analysis revealed one overlapping pathway (the insulin signaling pathway). There were no genes in common between GEM and genes from Väremo *et al.*[35] Therefore, this validates GEM as an independent identifier of human IR.

To characterize the subject-specific classification response rate of GEM, we calculated the misclassification rate of all subjects in our internal data sets (TonksS1 and TonksS2) and refer to this as a GEM score (Supplementary Figure S3A). Subjects frequently misclassified by GEM (high GEM score) were of interest. These included B13 and D16 (TonksS1) and B6 and D17 (TonkS2). B13 and B6, were classified as OIR by clamp but OIS by GEM; D16 and D17 were reciprocally classified. B13 and B6 had the lowest visceral/subcutaneous fat ratios and highest NOGD in the OIR group, whereas the visceral/subcutaneous fat ratios for D16 and D17 were in the top 25% (Supplementary Figure S3D, E, G). Various studies have suggested that visceral fat is a more important contributor to IR than body mass index[36,37] and accumulation of subcutaneous fat provides protective effects against IR.[38–40] Recently, several new genetic loci were reported that links insulin biology and IR to body fat distribution.[41] The respiratory quotient during the clamp for D17 and D16 were the lowest in the OIS group, a measure of metabolic flexibility. Hence, despite being insulin sensitive, they displayed indications of metabolic impairment (Supplementary Figure S3F) (see SR 1.4 for other details). These findings suggest that GEM classification may take into account global metabolic dysfunction associated with IR, rather than just impaired whole body glucose clearance.
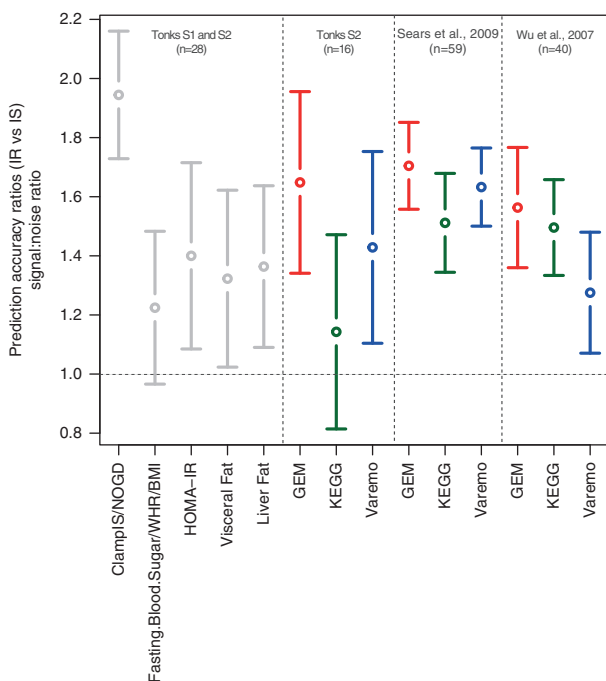
Network level analysis

Our original gene set of 129 genes and GEM enriched for several metabolic transcription factors, mitochondrial proteins and genes with SNPs associated with T2D/obesity or altered lipid and metabolite levels (SR 1.5). To further interrogate our gene sets we examined their connectivity with the insulin signaling pathway (ISP) proteins as this is the dominant driver of insulin action in muscle. A network-based enrichment analysis[42] identified genes such as JAK1, β-catenin (*CTNNB1*), 14-3-3-gamma (*YWHAG*), *SCD*, *PPP3R1*, *NR3C1*, *PFKFB3*, *GRB14*, *ACLY* and *RETN* (resistin) within the
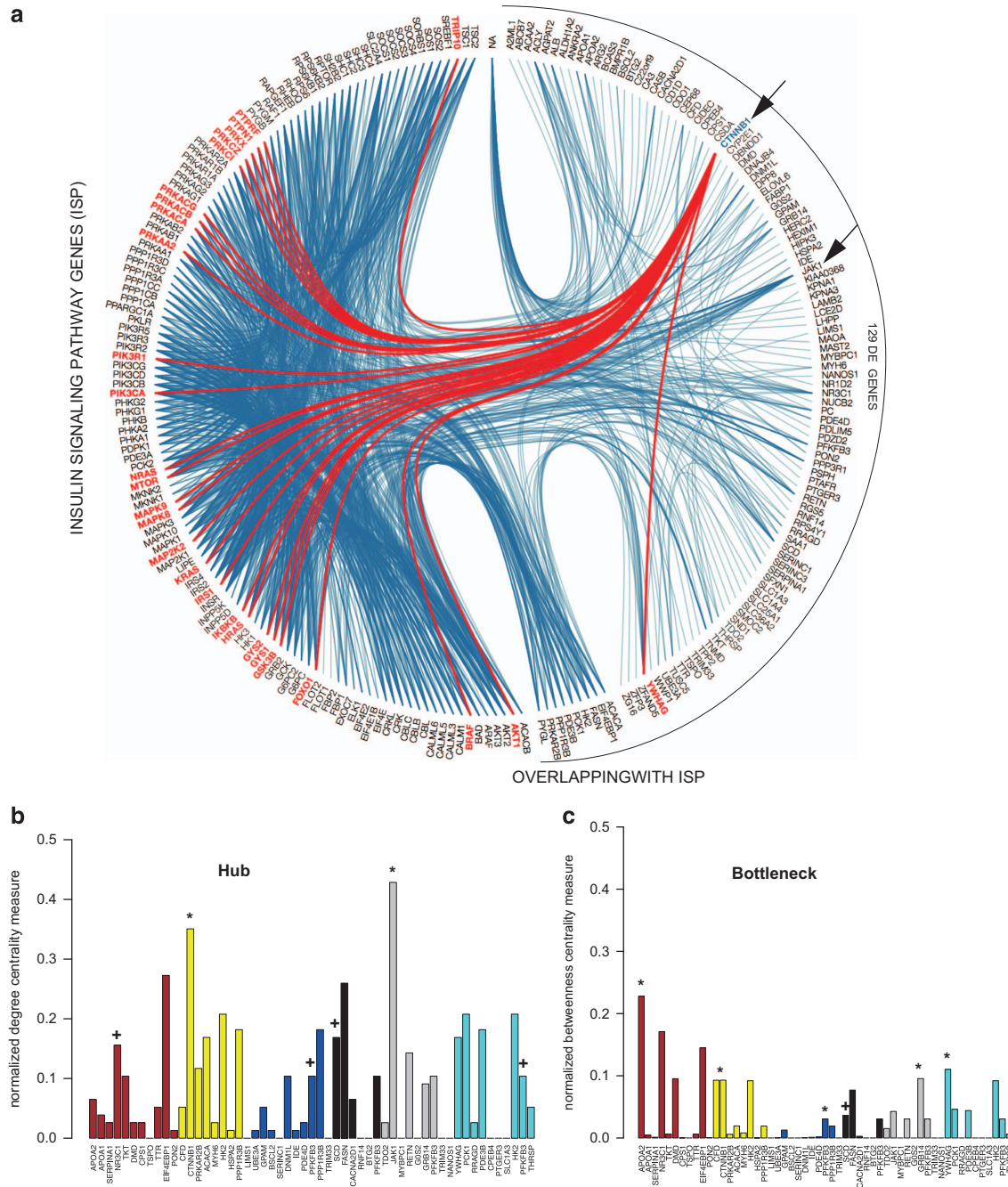

Figure 5. The protein–protein interaction wheel between the insulin signaling pathway (ISP) and 129 DE genes highlight β-catenin and Jak1 to be top connectors and network bottlenecks. (**a**) The right hand side of the wheel lists all members of the 129 DE gene set and the left hand side of the wheel shows members of the ISP. At the bottom, a small section marked 'OVERLAPPING WITH ISP' shows the nine genes within our DE gene set that overlap with the 137 ISP genes. Thick blue lines on the right hand side of the wheel are indicative of high connectivity (degree) with the proteins in the ISP. The diagram highlights the interactions of β-catenin (CTNNB1) on the right hand side with the ISP genes in red; both β-catenin and Jak1 are labeled with arrows to indicate their extent of connectivity to the ISP genes. (**b**) The genes that comprise each of the six significantly correlated gene modules (represented by their module colors from weighted gene co-expression network analysis analysis) with EMP measures are listed here. The bar plot quantitatively represents the degree of connectivity of each gene within these six GEM gene modules with the ISP. The hubs of each module are highlighted (*), if the hub is a member of ISP then the next ranking non-ISP member is marked with (+) (**c**) Bottleneck or between-ness distribution of the genes comprising the GEM modules. Similar to (**a**) top ranking bottlenecks within each module are marked with (*) and second ranking non-ISP members are marked with (+).

original 129 that potentially cross-talk with those comprising the ISP through a protein–protein interaction map (Figure 5a) (see SM 1.9 and SR 1.6).

Next, we focused on genes comprising the GEM. Only the hub proteins (degree) in yellow (β-catenin: member of Wnt signaling pathway) and gray (Jak1: member of Jak-Stat pathway) modules were not members of the ISP and hence novel nodes in insulin action (SM 1.9) (Figure 5b). Similarly, we identified five novel bottlenecks, β-catenin, 14-3-3-gamma, PFKFB3, GRB14, and APOA2 (Figure 5c). Although there is extensive evidence relating the above proteins to IR (SR 1.7); the link between Jak1 and β-catenin to IR is relatively unexplored. This prompted us to interrogate the role of these proteins in insulin action in muscle cell.

β-catenin and Jak1 have a role in insulin action

We tested in-silico to confirm if Jak1 and β-catenin were the bottlenecks of communication between their respective source pathways, Jak-Stat and Wnt signaling pathways, and the ISP (SM 1.9 and Supplementary Figure S4). Through between-ness distributions of the protein interaction network of Wnt signaling and the ISP, we determined β-catenin to be the top bottleneck of communication (Figure 6a). Similarly, Jak1 was one of the top five most important proteins enabling interactions between Jak-Stat pathway and the ISP, (Figure 6b).

We further interrogated the Wnt:ISP and Jak-Stat:ISP networks by detecting community structures within them through a hierarchical decomposition process (SM 1.9). We focused on the communities that contained β-catenin and Jak1, respectively, and obtained a map of their first-degree neighbors (Figure 6c, d). We observed rich connections between β-catenin and insulin signaling proteins (red circles) such as GSK3B, FOXO1, PTPN1, PIK3R1, PTPRF, and IKBKB. Similarly, through the first-degree neighbor map of Jak1, we found dense connections with key insulin signaling proteins such as INSR, IRS1/2, GRB2, SH2B2, RAF1, PRKCZ, and PIK3R1. Using gene-set tests, we also found the Wnt and Jak-Stat Signaling Pathways to be up-regulated in OIR compared with OIS individuals in both our internal GE data sets (TonksS1: Wnt *P* value 8.86e-06; Jak-Stat 0.1 and TonksS2: Wnt 0.1, Jak-Stat 0.006). Based on these observations we predicted that β-catenin and Jak1 likely have an important role in skeletal muscle insulin action. To test this, we examined the consequences of perturbing these nodes on insulin action in L6 muscle cells (Figure 6e, f). Treatment of L6 myotubes with pyrvinium pamoate, an inhibitor of β-catenin and Wnt signaling,[43] inhibited insulin stimulated 2-deoxyglucose uptake (Figure 6e). Similarly, treatment with the Jak1 inhibitor GPLG0634,[44] inhibited insulin stimulated 2-deoxyglucose uptake in a dose-dependent manner (Figure 6f). These effects were not due to direct inhibition of glucose transporter activity (Supplementary Figure S5 and SM 1.10). This highlights the power of such approaches for identifying not only novel disease signatures but also novel regulatory nodes that act upstream of well established processes.

DISCUSSION

IR is one of the earliest risk factors for metabolic disease yet clinically it often remains undiagnosed. Most methods for assessing IR are either expensive or too specialized, emphasizing the need for improved and simplistic approaches. Variation in human GE data can preclude identification of GE signatures that describe a phenotype. To circumvent this we devised a cross-species integration framework. Here robust GE changes in mice progressing to IR were used to augment GE differences in humans. This enabled identification of a 93 gene expression motif that diagnoses individuals with IR more accurately than commonly used clinical measures both in internal and external GE data sets

(Figure 4). One of the key regulatory nodes associated with this GEM was the β-catenin pathway.

A major confounder for IR diagnosis is obesity; while many IR people are obese this is not always the case. To circumvent this problem, in the present study we compared OIS with OIR enabling us to exclude a potential role for obesity *per se*. GEM was able to predict outliers within the obese groups that had been classified based on the hyperinsulinemic euglycemic clamp. For example, the GEM score identified two individuals who were classified as OIR by clamp and OIS by GEM, they had a very low visceral to subcutaneous adipose tissue ratio that is indicative of insulin sensitivity.[36–40] Such features can now act as a future predictor of the performance of GEM in patient-specific diagnosis i.e., individuals with alternate clinical features such as low visceral/ subcutaneous fat ratio and borderline NOGD are likely to be diagnosed as OIS instead of OIR by GEM. Hence, these individuals may represent a discrete IR subgroup(s). It will also be exciting to determine if these subgroups also segregate based on other features such as lifestyle or genetics. Overall, this suggests that multi-gene signatures like GEM may be diagnostic of combinatorial features making this approach potentially more clinically useful than single physiological measures.

The ability of GEM to identify subgroups of IR individuals also emphasizes the multifactorial/heterogeneous nature of IR, the clinical manifestations of which remain to be explored. It is important to point out that the GEM described here emanated from a cross-species integrated approach using HFD fed mice, however, it is well established that many other perturbations are known to trigger IR including exposure to dietary components other than fat, physical inactivity and steroids. Thus, it will be of interest to determine if GEMs driven by alternate mouse models can segregate different forms of IR in humans.

GEM arose from an unbiased systems biology approach and so it is of interest to explore the inherent biology within this signature. GEM possesses a number of genes and pathways previously implicated in IR and/or insulin action (SR 1.5). This validates this approach supporting the utility of GEM not just as a classifier but also as a discovery tool. Using network analysis we identified β-catenin and Jak1 as two potentially novel components of the IR nexus and a role for both of these pathways in insulin action was confirmed using functional analysis in muscle cells *in vitro*.

Jak1 has previously been implicated in adipocyte IR,[45] and Jak-Stat signaling has been proposed to be involved in the development of IR in cardiac myocytes;[46] to our knowledge this is not the case for β-catenin. The expression of β-catenin was highly correlated with the strongest phenotypic predictors of IR, such as whole body insulin sensitivity, at the single gene level (Figure 3b). In humans, abnormal Wnt signaling has been associated with early onset obesity and T2D.[47] Recently, it has been shown that loss of β-catenin renders mice resistant to HFD due to increased energy expenditure and insulin sensitivity due to hyperactivity.[48] Further, β-catenin binds to, and regulates the activity of both FOXOs and T-cell factor (TCFs) transcription factors.[49,50] In addition, SNPs in the *TCF7L2* gene, which is the transcription factor that regulates β-catenin, significantly increases the risk of developing T2D.[51,52] Intriguingly, TCF7L2 was originally thought to be a beta cell specific gene that controls insulin secretion.[52,53] However, subsequent studies have found that TCF7L2 principally functions as a regulator of insulin action in liver and possibly other tissues.[54,55] This is important, as overall most SNPs associated with T2D are thought to principally have a role in beta cell dysfunction and very few SNPs have been found in genes that control insulin action in tissue such as muscle. As we observed β-catenin to be a major regulatory node in a GEM associated with IR we speculate that this pathway may represent a major biomarker for defects in insulin action.
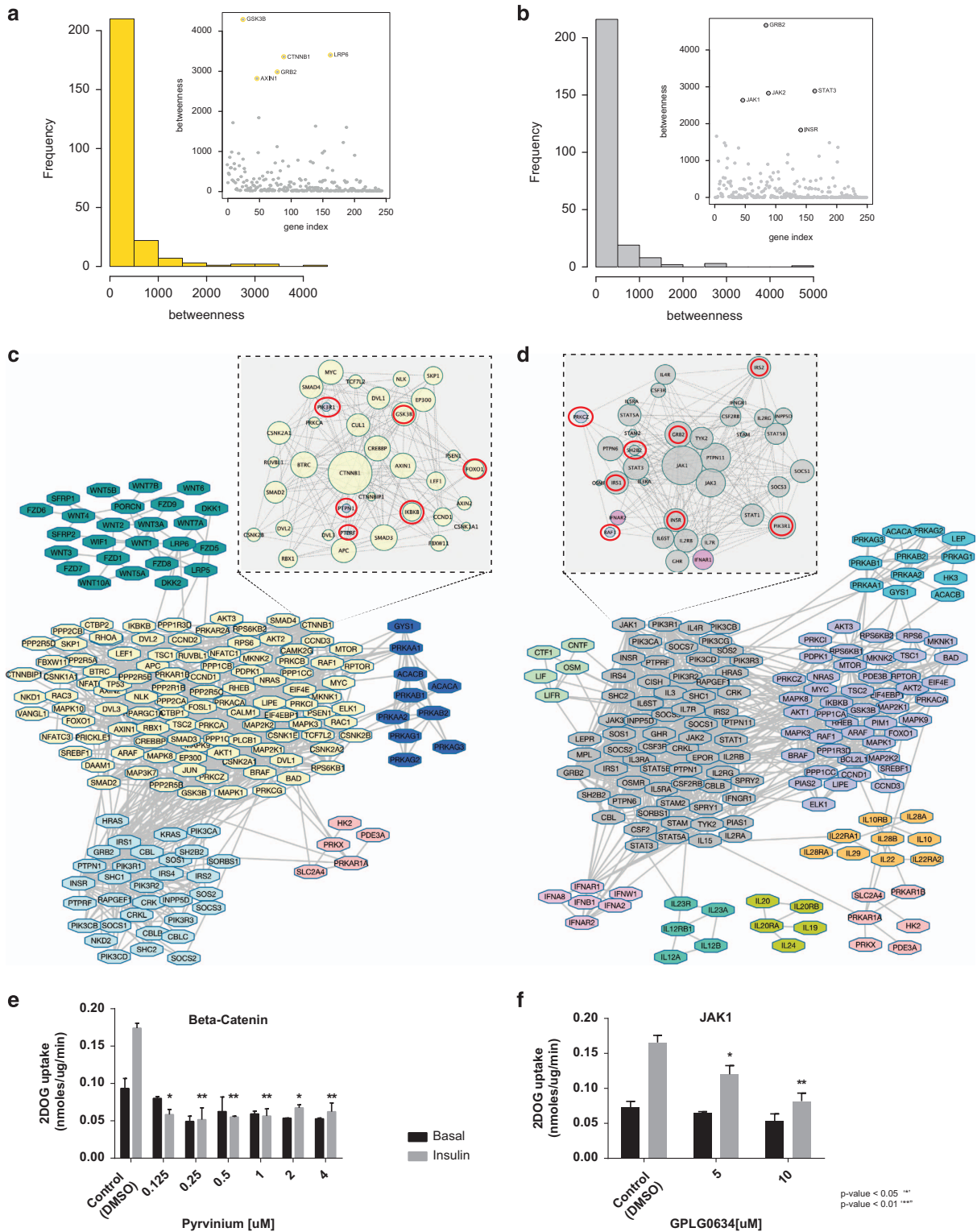
Figure 6. Detailed network maps of Wnt signaling and Jak-Stat signaling pathways with insulin signaling pathway reveals strong connections of β-catenin and Jak1 to selected insulin signaling proteins; experimental validation of these two proteins provides evidence of their role in muscle insulin action. (**a** and **b**) Shows the between-ness distributions of proteins in the Wnt signaling and insulin signaling pathway (ISP) in (**a**) and the distribution of Jak-Stat and ISP in (**b**). Both β-catenin and Jak1 are one of the top communication bottlenecks in these networks. (**c** and **d**) Edge between-ness derived graph communities (>5 members) in the Wnt:ISP and Jak-Stat:ISP networks are visualized through the different colors. The first-degree neighbors of β-catenin are expanded in (**c**) and Jak1 in (**d**), where the size of the nodes reflects the degree of connection and the red circles denote members of the ISP. (**e** and **f**) Shows the experimental validation of Jak1 and β-catenin in insulin stimulated glucose uptake assays. 2-deoxyglucose (2DOG) uptake into L6 muscle cells was measured in the absence or presence of insulin. 2DOG uptake in the absence or presence of Jak1 inhibitor GPLG0634 (**e**) or the β-catenin inhibitor pyrvinium (**f**).

In conclusion, systems biology approaches as described here involving cross-species data integration have enormous potential in preventive medicine. Obviously, longitudinal study designs will be required to evaluate the true worth of such approaches and their application to the clinical workplace will be required. The value of this GE motif needs to be ascertained in a more readily available clinical sample such as peripheral blood cells. Nevertheless, from this cross sectional analysis of a relatively modest number of subjects we delineated a multi-gene signature that encapsulates a complex array of biological pathways to diagnose IR with comparable accuracy to the best currently available physiological/clinical methods. This justifies more focus on systems biology approaches to tackle the cause and prevention of complex metabolic diseases.

DATA AVAILABILITY

The microarray data for the human and mouse study have been deposited in the Gene Expression Omnibus under identifiers GSE73034 and GSE73036, respectively (both part of super series reference GSE73037).

CONTRIBUTIONS

RC and JYHY and DEJ were responsible for study design and RC for all data analysis. PSK performed cellular experiments in L6 muscle cells. KLH, CCM and DJF performed animal experiments. JRJ, GK and ZM performed all microarray experiments. KT recruited volunteers, and ran the human clinical studies including phenotyping and collection of muscle. DSB was involved in human tissue collection and with JRG provided clinical expertise. JS, DJF and SH revised the article critically and provided important intellectual content. The manuscript was written by RC and DEJ and was revised and approved by all authors.

COMPETING INTERESTS

The authors declare no conflict of interest.

REFERENCES
1 Meigs JB, Wilson PWF, Fox CS, Vasan RS, Nathan DM, Sullivan LM *et al*. Body mass index, metabolic syndrome, and risk of type 2 diabetes or cardiovascular disease. *J Clin Endocrinol Metab* 2006; **91**: 2906–2912.
2 Ärnlöv J, Ingelsson E, Sundström J, Lind L. Impact of body mass index and the metabolic syndrome on the risk of cardiovascular disease and death in middle-aged men. *Circulation* 2010; **121**: 230–236.
3 Samocha-Bonet D, Chisholm DJ, Tonks K, Campbell LV, Greenfield JR. Insulin-sensitive obesity in humans a favorable fat phenotype? *Trends Endocrinol Metab* 2012; **23**: 116–124.
4 Conus F, Allison DB, Rabasa-Lhoret R, St-Onge M, St-Pierre DH, Tremblay-Lebeau A *et al*. Metabolic and Behavioral Characteristics of Metabolically Obese but Normal-Weight Women. *J Clin Endocrinol Metab* 2004; **89**: 5013–5020.
5 Conus F, Rabasa-Lhoret R, Péronnet F. Characteristics of metabolically obese normal-weight (MONW, subjects. *Appl Physiol Nutr Metab* 2007; **32**: 4–12.
6 DeFronzo RA, Tobin J, Andres R. Glucose clamp technique: a method for quantifying insulin secretion and resistance. *Am J Physiol* 1979; **237**: 214–223.
7 Newsholme P, Bender K, Kiely A, Brennan L. Amino acid metabolism, insulin secretion and diabetes. *Biochem Soc Trans* 2007; **35**: 1180–1186.
8 Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E *et al*. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011; **17**: 448–453.
9 Gall W, Beebe K, Lawton K, Adam K-P, Mitchell M, Nakhle PJ *et al*. α-Hydroxybutyrate Is an Early Biomarker of Insulin Resistance and Glucose Intolerance in a Nondiabetic Population. *PLoS ONE* 2010; **5**: e10883.
10 Huffman KM, Shah SH, Stevens RD, Bain JR, Muehlbauer M, Slentz CA *et al*. Relationships between circulating metabolic intermediates and insulin action in overweight to obese, inactive men and women. *Diabetes Care* 2009; **32**: 1678–1683.
11 Newgard CB, J An, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF *et al*. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* 2014; **9**: 311–326.
12 Schadt EE, Woo S, Hao K. Bayesian method to predict individual SNP genotypes from gene expression data. *Nat Genet.* 2012; **44**: 603–608.
13 Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J *et al*. PGC-1[alpha]-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003; **34**: 267–273.
14 Kristiansson E, Österlund T, Gunnarsson L, Arne G, Joakim Larsson DG, Nerman O. A novel method for cross-species gene expression analysis. *BMC Bioinformatics* 2013; **14**: 1–14.
15 Fang H, Tong W, Perkins R, Shi L, Hong H, Cao X *et al*. Bioinformatics approaches for cross-species liver cancer analysis based on microarray gene expression profiling. *BMC Bioinformatics* 2005; **6**: 1–10.
16 Lu Y, Huggins P, Bar-Joseph Z. Cross species analysis of microarray expression data. *Bioinformatics* 2009; **25**: 1476–1483.
17 Storlien LH, James DE, Burleigh KM, Chisolm DJ, Kraegen EW. Fat feeding causes widespread in vivo insulin resistance, decreased energy expenditure, and obesity in rats. *Am J Physiol* 1986; **251**: E576–E583.
18 Park S-Y, Cho Y-R, Kim H-J, Higashimori T, Danton C, Lee M-K *et al*. Unraveling the temporal pattern of diet-induced insulin resistance in individual organs and cardiac dysfunction in c57bl/6 Mice. *Diabetes* 2005; **54**: 3530–3540.
19 Panchal SK, Brown L. Rodent Models for Metabolic Syndrome Research. *J Biomed Biotechnol* 2011; **2011**: 351982.
20 Surwit RS, Kuhn CM, Cochrane C, McCubbin JA, Feinglos MN. Diet-induced type II diabetes in C57BL/6J mice. *Diabetes* 1988; **37**: 1163–1167.
21 Erdmann J, Kallabis B, Oppel U, Sypchenko O, Wagenpfeil S, Schusdziarra V. Development of hyperinsulinemia and insulin resistance during the early stage of weight gain. *Am J Physiol Endocrinol Metab* 2008; **294**: E568–E575.
22 Samocha-Bonet D, Campbell LV, Viardot A, Freund J, Tam CS, Greenfield JR *et al*. A family history of type 2 diabetes increases risk factors associated with overfeeding. *Diabetologia* 2010; **53**: 1700–1708.
23 Brøns C, Jensen CB, Storgaard H, Hiscock NJ, White A, Appel JS *et al*. Impact of short-term high-fat feeding on glucose and insulin metabolism in young healthy men. *J Physiol* 2009; **587**: 2387–2397.
24 Turner N, Kowalski GM, Leslie SJ, Risis S, Yang C, Lee-Young RS *et al*. Distinct patterns of tissue-specific lipid accumulation during the induction of insulin resistance in mice by high-fat feeding. *Diabetologia* 2013; **56**: 1638–1648.
25 Tonks KT, Ng Y, Miller S, Coster AC, Samocha-Bonet D, Iseli TJ *et al*. Impaired Akt phosphorylation in insulin-resistant human muscle is accompanied by selective and heterogeneous downstream defects. *Diabetologia* 2013; **56**: 875–885.
26 Finck BN, Bernal-Mizrachi C, Han DH, Coleman T, Sambandam N, LaRiviere LL *et al*. A potential link between muscle peroxisome proliferator- activated receptor-α signaling and obesity-related diabetes. *Cell Metab* 2005; **1**: 133–144.
27 Park KS, Ciaraldi TP, Abrams-Carter L, Mudaliar S, Nikoulina SE, Henry RR. PPAR-γ gene expression is elevated in skeletal muscle of obese and type ii diabetic subjects. *Diabetes* 1997; **46**: 1230–1234.
28 Chavez JA, Summers SA, Ceramide-Centric A. View of Insulin resistance. *Cell Metab* 2012; **15**: 585–594.
29 Newgard C. Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell Metab* 2012; **15**: 606–614.
30 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9**.
31 Langfelder P, Horvath S. Fast R functions for robust correlations and hierarchical clustering. *J Stat Softw* 2012; **46**: 1–17.
32 Naukkarinen J, Heinonen S, Hakkarainen A, Lundbom J, Vuolteenaho K, Saarinen L *et al*. Characterising metabolically healthy obesity in weight-discordant monozygotic twins. *Diabetologia* 2014; **57**: 167–176.
33 Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 1985; **28**: 412–419.
34 Friedman J, Hastie T, Robert T. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010; **33**: 1–22.
35 Väremo L, Scheele C, Broholm C, Mardinoglu A, Kampf C, Asplund A *et al*. Proteome- and transcriptome-driven reconstruction of the human myocyte

metabolic network and its use for identification of markers for diabetes. *Cell Rep* 2015; **11**: 921–933.

36 Taksali SE, Caprio S, Dziura J, Dufour S, Calí AMG, Goodman TR et al. High visceral and low abdominal subcutaneous fat stores in the obese adolescent: a determinant of an adverse metabolic phenotype. *Diabetes* 2008; **57**: 367–371.

37 Kim H, Bae S, Kim E, Kim C. Visceral-to-subcutaneous fat ratio using fat measurement CT is essential for the assessment of risk for diabetes, hypertension and early atherosclerosis. in: *Endocr. Abstr.* 2012; **29**: 1222.

38 Livingston EH. Lower body subcutaneous fat accumulation and diabetes mellitus risk. *Surg Obes Relat Dis* 2006; **2**: 362–368.

39 Sandeep S, Gokulakrishnan K, Velmurugan K, Deepa M, Mohan V. Visceral & subcutaneous abdominal fat in relation to insulin resistance & metabolic syndrome in non-diabetic south Indians. *Indian J Med Res* 2010; **131**: 629–635.

40 Jung C-H, Kim B-Y, Kim K-J, Jung S-H, Kim C-H, Kang S-K et al. Contribution of subcutaneous abdominal fat on ultrasonography to carotid atherosclerosis in patients with type 2 diabetes mellitus. *Cardiovasc Diabetol* 2014; **13**: 67.

41 Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 2015; **518**: 187–196.

42 Glaab E, Baudot A, Krasnogor N, Schneider R, Valencia A. EnrichNet: network-based gene set enrichment analysis. *Bioinformatics* 2012; **28**: i451–i457.

43 Thorne CA, Hanson AJ, Schneider J, Tahinci E, Orton D, Cselenyi CS et al. Small-molecule inhibition of Wnt signaling through activation of casein kinase 1α. *Nat Chem Biol* 2010; **6**: 829–836.

44 Van Rompaey L, Galien R, van der Aar EM, Clement-Lacroix P, Nelles L, Smets B et al. Preclinical characterization of GLPG0634, a selective inhibitor of JAK1, for the treatment of inflammatory diseases. *J Immunol* 2013; **191**: 3568–3577.

45 McGillicuddy FC, Chiquoine EH, Hinkle CC, Kim RJ, Shah R, Roche HM et al. Interferon γ attenuates insulin signaling, lipid storage, and differentiation in human adipocytes via activation of the JAK/STAT pathway. *J Biol Chem* 2009; **284**: 31936–31944.

46 Frias MA, Montessuit C. JAK-STAT signaling and myocardial glucose metabolism. *JAK-STAT* 2013; **2**: e26458.

47 Laudes M. Role of WNT signalling in the determination of human mesenchymal stem cells into preadipocytes. *J Mol Endocrinol* 2011; **46**: R65–R72.

48 Elghazi L, Gould AP, Weiss AJ, Barker DJ, Callaghan J, Opland D et al. Importance of β-catenin in glucose and energy homeostasis. *Sci. Rep* 2012; **2**, Epub 2012 Sep 25.

49 Essers M, de Vries-Smits L, Barker N, Polderman P, Burgering H, Korswagen BM. Functional interaction between beta-catenin and FOXO in oxidative stress signaling. *Science* 2005; **308**: 1181–1184.

50 Hoogeboom D, Essers MAG, Polderman PE, Voets E, Smits LMM, Burgering BMT. Interaction of FOXO with β-catenin inhibits β-catenin/T cell factor activity. *J Biol Chem* 2008; **283**: 9224–9230.

51 Welters HJ, Kulkarni RN. Wnt signaling: relevance to beta-cell biology and diabetes. *Trends Endocrinol Metab* 2008; **19**: 349–355.

52 Grant SFA, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J et al. Variant of transcription factor 7-like 2 (TCF7L2, gene confers risk of type 2 diabetes. *Nat Genet* 2006; **38**: 320–323.

53 Lyssenko V, Lupi R, Marchetti P, Del Guerra S, Orho-Melander M, Almgren P et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. *J Clin Invest* 2007; **117**: 2155.

54 Ip W, Shao W, Song Z, Chen Z, Wheeler MB, Jin T. Liver-Specific expression of dominant-negative transcription factor 7-like 2 causes progressive impairment in glucose homeostasis. *Diabetes* 2015; **64**: 1923–1932.

55 Boj SF, van Es JH, Huch M, Li VSW, José A, Hatzis P et al. Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell* 2015; **151**: 1595–1607.

Supplementary Information accompanies the paper on the *npj Systems Biology and Applications* website (http://www.nature.com/npjsba)