Transcription Factor Co-expression Networks of Adipose RNA-Seq Data Reveal Regulatory Mechanisms of Obesity

Ruta Skinkyte-Juskiene¹, Lisette J.A. Kogelman¹,² and Haja N. Kadarmideen¹,³,*

¹Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Gronnegårdsvej 7, 1870 Frederiksberg C, Denmark; ²Danish Headache Center, Department of Neurology, Glostrup Research Institute, Rigshospitalet Glostrup, Nordre Ringvej 69, 2600 Glostrup, Denmark; ³Section of Systems Genomics, Department of Bio and Health Informatics, Technical University of Denmark, Kemitorvet, Building 208, 2800 Kgs. Lyngby, Denmark

Abstract: Background: Transcription Factors (TFs) control actuation of genes in the genome and are key mediators of complex processes such as obesity. Master Regulators (MRS) are the genes at the top of a regulation hierarchy which regulate other genes.

Objective: To elucidate clusters of highly co-expressed TFs (modules), involved pathways, highly interconnected TFs (hub-TFs) and MRS leading to obesity and leanness, using porcine model for human obesity.

Methods: We identified 817 expressed TFs in RNA-Sequencing dataset representing extreme degrees of obesity (DO; lean, obese). We built a single Weighted Transcription Factor Co-expression Network (WTFCN) and TF sub-networks (based on the DO). Hub-TFs and MRS (using iRegulon) were identified in biologically relevant WTFCNs modules.

Results: Single WTFCN detected the Red module significantly associated with DO (P < 0.03). This module was enriched for regulation processes in the immune system, e.g.: Immune system process (Padj = 2.50E-06) and metabolic lifestyle disorders, e.g. Circadian rhythm - mammal pathway (Padj = 2.33E-11). Detected MR, hub-TF SP1 was involved in obesity, immunity and osteoporosis. Within the obese sub-network, the Red module suggested possible associations with immunity, e.g. TGF-beta signaling pathway (Padj = 1.73E-02) and osteoporosis, e.g. Osteoclast differentiation (Padj = 1.94E-02). Within the lean sub-network, the Magenta module displayed associations with type 2 diabetes, obesity and osteoporosis e.g. Notch signaling pathway (Padj = 2.40E-03), osteoporosis e.g. hub-TF VDR (a prime candidate gene for osteoporosis).

Conclusion: Our results provide insights into the regulatory network of TFs and biologically relevant hub TFs in obesity.

Keywords: Obesity, Transcription factors, WGCNA, Transcriptomics, RNA-Seq, Gene networks.

1. INTRODUCTION

Transcription Factors (TFs) are the proteins that bind to specific DNA sequences and control the rate of genetic information transcription from DNA to mRNA, and they perform this function alone or in a complex with other genes. Particular interest goes to interactions where a given TF regulates other TFs, or itself [1]. A Master Regulator (MR) is a gene that occupies the very top of a regulatory hierarchy and is not influenced by the regulatory influence of any other gene [2]. MRS code for the TFs, which in turn alter the expression of downstream genes in the pathway [3].

Eukaryotic cellular functions are highly connected through networks of transcriptional regulators that regulate other transcriptional regulators [4]. The combinatorial cross-regulation of hundreds of sequence-specific TFs defines a regulatory network that underlies cellular identity and function [1]. However, the structure of core regulatory networks and their component sub-networks are largely undefined.

Obesity is a major health concern that leads to a chronic inflammatory state, including diabetes and cardiovascular disease [5]. Undirected Weighted Gene Co-expression Network Analysis (WGCNA) [6] can be used to find clusters of highly correlated genes (modules) and relating them to external traits. WGCNA has shown its potential to unravel the gene regulatory architecture of complex traits and diseases, for example: Alzheimer's disease in humans [7] parasite resistance in sheep [8], muscling in sheep [9], wool develop-
ment in sheep [10], feed efficiency in cattle [11], muscle and meat properties in pigs [12] and obesity in a pig model [13]. WGCNA has been applied to RNA-Seq [13], to microarrays data, e.g. in the case of several human diseases [14, 15] and to porcine reproductive and respiratory syndrome virus studies [16].

The principle of WGCNA could be specifically targeted to TF co-expression networks to detect TF modules or the most promising hub-TFs that are potential regulators of the majority of the other TFs in a network. However, scale-free topology-based weighted co-expression networks of TFs involved in obesity (e.g. using the WGCNA method) have not been constructed yet. Such a TF network is expected to lead to a better understanding of the complex gene regulation of obesity development and subsequent pathogenesis and potentially lead to the detection of predictive genetic biomarkers for prevention and prediction of obesity and obesity-related diseases. For identification of drug targets and biomarker for diseases it is important to catalogue master regulatory genes (Master Regulators or MR). It is possible that some of the MRs may control other MRs, as has been found previously in, e.g. MRs: AphA and LuxR were found repressing each other's expression [17], and cell cycle transcriptional regulators regulated expression of other regulators [18]. So it is important to elucidate the complex networks of regulatory factors and pathways that are affected.

The main purpose of this research was to construct a WTFCN based on RNA-Seq data and to detect modules, pathways and novel regulator TFs governing regulatory processes in obesity and obesity-related diseases. In this research, we have used a publicly available dataset (GEO: GSE61271) containing adipose RNA-Seq data of a pig model for human obesity. To the best of our knowledge, this is the first study to use a network approach on TFs to detect regulatory genes, based on RNA-Seq transcriptomics for human obesity in an animal model.

2. MATERIALS AND METHODS

The complete picture of the WTFCN construction workflow is presented in Fig. (1).

2.1. Study Population

In this study, we used publicly available RNA-Seq data and obesity state (lean or medium or obese) from a porcine model for human obesity (Gene Expression Omnibus (GEO) database of NCBI with accession number: GSE61271). The description of the dataset has previously been published in [13]. In short, an F2 pig resource population was created by crossing Göttingen minipig boars with Duroc sows. In total, 36 pigs were selected for RNA-Sequing of subcutaneous adipose tissue based on a genetic obesity index. Those pigs were divided into groups representing three different DO: lean (n = 12), median (n = 12), and obese (n = 12). RNA-Seq was performed on the HiSeq2500 platform and obtained reads were aligned using STAR aligner using the Sus scrofa10.2.72 genome. Read counts were estimated using HTSeq, filtered based on low expression levels and then normalized using the voom() variance-stabilization function in the R-package Limma.

2.2. Identification of TFs

Transcription factors were identified from the PANTHER Classification System database [19]. Using the PANTHER protein class Ontology browser, we retrieved genes classified as “Transcription Factor” (PC00218) in Sus scrofa. In our RNA-Seq dataset, 817 TFs of the 1312 identified TFs (list of TFs extracted at 19.04.2016) were expressed (Supplementary Data Sheet 1).

2.3. Single Weighted Transcription Factor Co-expression Network Construction

Using the Weighted Gene Co-expression Network Analysis (WGCNA) [6] approach we constructed a single WTFCN of 817 TFs using all 36 pigs. WTFCN construction followed the same workflow as described in [13]. Summarized, the adjacency matrix was created by calculating Pearson’s correlations between the expression levels of the 817 TFs. To obtain a scale-free network, meaning it will contain a few nodes (TFs) that are highly connected with many other nodes, the adjacency matrix was raised to a power β (7) which was chosen based on the scale-free topology fitting index (R² > 0.8). The Topological Overlap Measure (TOM) was calculated representing the number of shared neighbors, and based on the dissimilarity TOM a TF clustering dendrogram was created using hierarchical clustering. TF modules, clusters of highly interconnected TFs, were detected as branches of the TF dendrogram using the cutreeHybrid algorithm and named according to a color. It is believed that highly interconnected genes act in similar biological pathways [20]. To identify biologically relevant modules, the Module Eigengene (ME) of each module was calculated as the first principal component of the module and checked using the proportion of variance in the module explained by the MEs. The MEs were correlated with the DO: groups of pigs were assigned to numerical values (lean was assigned a ‘1’, median a ‘2’ and obese a ‘3’) which were fitted as continuous variable to calculate correlations between the module expression and phenotype. This resulted in the Module-Trait Relationship (MTR), which were used to select potential interesting associations with obesity. Potential interesting modules were further analyzed by detecting TFs that possess a high connectivity (degree of connections with other TFs), the so-called hub-TFs. The hubs are of great interest due to the number of the connections they have, as potential malfunction or disruption of the expressed levels of the hub TF will likely affect the complete module where it is present [20]. Furthermore, the TF significance (correlation between the individual TF expression levels and DO, where DO was fitted as a continuous variable) and the module membership (MM; correlation between TF expression profile and the ME of a given module) were used in order to identify biologically and statistically plausible genes.

2.3.1. Functional Enrichment Analysis

Functional enrichment analysis was performed using several approaches. First, the GOSeq R Bioconductor package [21] was applied on selected modules. Before running GOSeq in selected modules, pig Ensembl Gene IDs (Sscrofa10.2) were replaced with human Ensembl Gene IDs (GRCh38.p3) using the Ensembl BioMart database [22]. The non-native Gene Identifier [21], containing a complete set of
Study population

Identification of Transcription Factors

Construction of single WTFCN, obese and lean sub-networks

Identification of significant WTFCNs modules

Ensembl IDs: Sscrofa10.2 → GRCh38.p3

Functional analysis

Intramodular analysis

Analysis in Cytoscape

Differently connected and expressed genes

Fig. (1). Workflow to build TF co-expression networks for different degrees of obesity using RNAseq data and further functional analyses.
transcription factors analogous to GRCh38.p3, was created and used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis [23], taking length bias into account [24]. Obtained p-values were adjusted for multiple-testing using the Benjamini-Hochberg (BH) correction. GO terms and KEGG pathways were called significant when the adjusted p-value was below 0.05. If GOseq analysis did not yield any significant GO and/or KEGG pathways results, WEB-based Gene SeT AnaLysis Toolkit (WebGestalt) [25] was used for GO and/or KEGG pathway analysis. Secondly, gene names of single TFs were extracted using the BioMart database GRCh38.p3.

2.4. Construction of Sub-networks

Using the same set of TFs and same network approach as described above, we constructed two sub-networks based on the DO: lean and obese. Resulting modules from the sub-networks were tested for Module Significance (MS). The MS is determined as the average of the TFs significance measure (the correlation between the expression values of TFs and DO) using both lean and obese pigs for all TFs in a given module. This measure is highly related to the correlation between module eigengene and the trait [20]. In order to compute MS for lean sub-network TFs expression values from lean and obese sub-networks were correlated with lean sub-network traits. In order to compute MS for obese sub-network TFs expression values from lean and obese sub-networks were correlated with obese sub-network traits. High MS means that the module is enriched with essential TFs. Functional and intra-modular analyses were performed on the modules with the strongest absolute MS.

2.4.1. Differentially Connected TFs in Sub-networks

Gene co-expression network analysis lends itself to identify entire groups of differentially regulated genes - a highly relevant endeavor in finding the underpinnings of complex traits that are polygenic in nature [26]. Differential network analysis concerns with identifying both differentially connected and differentially expressed genes [26]. In order to indicate which TFs were differently regulated in the networks we have compared difference in connectivity (DiffK) between the lean vs. obese sub-network, using the following formula:

$$\text{DiffK} = K_1(i) - K_2$$

where $K_1$ presents the connectivity of the lean sub-network and $K_2$ presents the connectivity of the obese sub-network. $K_1$ and $K_2$ were calculated by dividing each gene’s connectivity by the maximum sub-network connectivity, e.g.:

$$K_1(i) = \frac{k_1(i)}{\max(k)}$$

When DiffK was positive, it meant that the TFs were more highly connected in the lean sub-network than in the obese sub-network, while a negative DiffK meant that the TFs were more highly connected in the obese sub-network than in the lean sub-network. To facilitate the comparison between the connectivity measures of each network, standardization was performed in each network by dividing the TF connectivity by the maximum network connectivity. The difference between the connectivity values of two sub-networks was defined as difference in connectivity (DiffK).

DiffK values were distributed between -1 and 1. Genes were called differentially connected when the absolute value of DiffK was above 0.4. To measure differential TF expression between sub-networks the Student t-test was calculated, testing the difference between expression values of the lean and obese group. TFs were considered to be differentially expressed with an absolute value of the t-test statistic > 1.96.

2.5. Analysis with Cytoscape Software

We expanded the analysis by integrating WTFCN results into Cytoscape v3.3.0 [27]. Analysis and graphical representation was performed on significant modules of single WTFCN and on the modules having strongest absolute MS in the differentially co-expressed networks. Integration of genes expression and TF significance level into Cytoscape can open new perspectives in the analysis of biological networks, since the integration of topological analysis with expression data enhances the predictive power of the bioinformatics analysis [28]. NetworkAnalyzer, a Java plugin the Cytoscape platform [29], was used to perform topological analyses of the network in order to get an overview of the TFs’ (represented by nodes) activity and interactions (represented by edges) between them within the modules. The NetworkAnalyzer tutorial and Centralities tutorial [28] were used to describe node and edge attributes (Table S1). Integrating TFs expression data from the modules having the strongest absolute MS into Cytoscape explains functional relationship, interaction, dynamics, role in expression within the network.

Furthermore, the Cytoscape app iRegulon (version 1.3) [30] was used to predict MRs in biologically relevant WTFCNs modules. iRegulon is a computational method which identifies MRs and predicts direct target genes in a set of human, mouse and Drosophila genes. iRegulon uses more than 9,000 known position weight matrices from various sources and different species and link them to candidate-binding TFs using a “motif2TF” procedure. This allows to link motifs of TFs from other species to candidate human TFs. Predicted MRs are rated and grouped by an enrichment score, the NES score (Normalized Enrichment Score). Predicted MRs having highest NES scores and present in analyzed modules as TFs were considered as present MRs. Targets (nodes) of predicted MRs which were overlapping with targets (nodes) of present MRs were considered as present MRs targets. As mentioned above, WTFCN constructs an undirected network; however regulatory links were drawn on WTFCN modules according to present MRs and their targets, detected by iRegulon.

3. RESULTS AND DISCUSSION

3.1. Single WTFCN Analysis

For analysis of network topology, the soft thresholding power ($\beta = 6$) was chosen based on the criterion of scale-free topology. Using the 817 TFs in WTFCN we identified 18 modules, each containing at least 20 TFs (Fig. S1). The Red module (65 TFs) of single WTFCN (Red_single) had the highest correlation with obesity (MTR = 0.36, $P = 3.00E-02$) (Table 1) and was selected for further analysis. GOseq analysis of this module resulted in the detection of 22 GO categories (FDR < 0.05). Most of the GO terms indicated regulations in
the immunity system; e.g., immune system process (Padj = 2.50E-06), cytokine production (Padj = 1.03E-05), and interleukin-1 production (Padj = 1.83E-05). GOSeq did not detect any significant KEGG pathways, but WebGestalt functional enrichment analysis resulted in KEGG pathways related to immunity (e.g., TGF-beta signaling pathway, Padj = 7.06E-16), osteoporosis (e.g., Osteoclast differentiation, Padj = 1.00E-04) and obesity and diabetes (e.g., Circadian rhythm - mammal, Padj = 2.33E-11). It was found [31] that disruption of the circadian clockwork with the external environment (e.g., light-dark cycle) might have a role in the development of metabolic disorders (e.g., obesity and diabetes). In the Red_singles module we found two hub-TFs: IRF5 and SPI1. The highest ClosenessCentrality, Degree and Radiality in this module had TF ZYX, indicating its highest number of linked edges, fastest information spread within the network and the highest centrality by regulating other TFs. Highest Stress had TF MAFA, indicating its relevance in connecting regulatory molecules and possible involvement in cellular processes. iRegulon detected MR SPI1 (NES = 3.715). From 14 predicted targets four present targets (all upregulated TFs) were found in the module: MAFA, IRF5, ATF3 and ZNF710 (Fig. 3). As mentioned above, SPI1 was also one of the hub-TFs in this module. Therefore importance of SPI1 was confirmed as it is detected by two different methods. SPI1 was found to be regulating adipose differentiation related to protein in macrophages [32]. A study by [33] suggested SPI1 as a novel obesity therapeutic target, indicating that reduction in levels of SPI1 leads to an increase in lipid accumulation. In one of our previous studies, SPI1 was detected as an candidate regulator of a possessing strong associations between obesity, immunity and osteoporosis [13]. Moreover, its possible key role in the link between obesity and osteoporosis was shown. In a recent study, [34] found significant interaction between the long noncoding RNAs cyp2c91 with the SPI1 as regula-
tor gene linked to obesity. SPI1 can also regulate alternative splicing of target genes. Its target upregulated TF MAFA (also having highest Stress in Red_singles) was found to be dynamically regulated during adipogenesis [35]. Findings by [36] showed that MAFA expression in human adipocytes is upregulated during adipogenesis. MAFA was indicated as a regulator and a marker of adipose tissue inflammation, a process that subsequently causes insulin resistance. [37] found a SNP located close to MAFA associated with higher BMI, hypertension and diabetes, contributing to the risk of coronary artery disease and ischemic stroke. Another SPI1 target, ATF3, may play a role in adipocyte hypoxia-mediated mitochondrial dysfunction in obesity [38]. Furthermore, ATF3 polymorphism was found associated with atherosclerotic disease [39]. Hub-TF IRF5 (also regulated by SPI1) is well known to be of importance in immune system activity [40]. Moreover, a recent study by [41] found IRF5 expression negatively associated with insulin sensitivity and collagen deposition in visceral adipose tissue. Mice lacking IRF5 were healthy, and blocking of IRF5 was suggested as a possibility for stopping fat formation. [42] reported ZYX as aging-related and a potential biomarker for osteoporosis. [43] found the association of ZYX polymorphisms with carcass and meat quality traits. Functional annotation results, hub-TFs and MR of the Red_singles module from the single WTCFN analysis show its obvious biological role in immunity and obesity-related functions.

### 3.2. Sub-networks Analysis

Separate sub-networks were constructed on lean and obese DO groups. To obtain scale-free topology, a soft thresholding power of $\beta = 10$ was used in each sub-network. After WTCFN construction, 12 and 10 modules were

<table>
<thead>
<tr>
<th>Modules</th>
<th>Single WTCFN MTR (P value)</th>
<th>Obese Sub-network</th>
<th>Lean Sub-network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>-</td>
<td>-0.14</td>
<td>-0.20</td>
</tr>
<tr>
<td>Blue</td>
<td>-0.01 (1)</td>
<td>-0.04</td>
<td>0.002</td>
</tr>
<tr>
<td>Brown</td>
<td>0.035 (0.8)</td>
<td>-0.04</td>
<td>0.14</td>
</tr>
<tr>
<td>Green</td>
<td>-0.13 (0.4)</td>
<td>-0.05</td>
<td>-0.18</td>
</tr>
<tr>
<td>Greenyellow</td>
<td>-</td>
<td>-0.04</td>
<td>-</td>
</tr>
<tr>
<td>Magenta</td>
<td>-</td>
<td>-0.03</td>
<td>-0.26</td>
</tr>
<tr>
<td>Pink</td>
<td>-</td>
<td>-0.11</td>
<td>-0.14</td>
</tr>
<tr>
<td>Purple</td>
<td>-</td>
<td>-0.08</td>
<td>-0.13</td>
</tr>
<tr>
<td>Red</td>
<td>0.36 (0.03)</td>
<td>-0.15</td>
<td>0.08</td>
</tr>
<tr>
<td>Tan</td>
<td>-</td>
<td>-0.12</td>
<td>-</td>
</tr>
<tr>
<td>Turquoise</td>
<td>-0.3 (0.08)</td>
<td>-0.10</td>
<td>-0.03</td>
</tr>
<tr>
<td>Yellow</td>
<td>0.11 (0.5)</td>
<td>0.02</td>
<td>-0.22</td>
</tr>
</tbody>
</table>

(For interpretation of the references to color in this Table, the reader is referred to the web version of this paper.)
Fig. (2). Module significance of the obese and the lean sub-networks. Y axis represents correlation of TF significance, X axis - modules names. A) The Red obese module had the highest positive mean value (COR = 0.1477) in the obese vs. lean network and was enriched with obesity associated TFs, which indicated its significant difference in the obese sub-network. B) The Magenta lean module had the highest positive mean value (COR = 0.2598) in the obese vs. lean network and was enriched with leanness-associated TFs, which indicated its significant difference in the lean sub-network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Fig. (3). Visualization in Cytoscape of interactions between TFs, MR and hub-TFs in the Red single module of WTFCN. The module contains both directed and undirected edges, as it is compiled by combining analyses from two sources: WTFCN and Cytoscape. Size of the nodes correspond to kWithin value (the bigger the node, the higher the value), color of the nodes corresponds to TF expression level (bright red - highly expressed/upregulated nodes; bright green - lowly expression, downregulated nodes). Shape of the nodes corresponds to significance level: square shape (P < 0.05), circle shape (P > 0.05). Circular layouts of the module are grouped clockwise according to nodes parameter kWithin, starting from the bottom of the picture. Both single WTFCN hub-TFs (IRF5 and SPI1) have highest kWithin values and were found mostly expressed in the module. Edges stroke color and width of edges correspond to EdgeBetweenness. Higher value of EdgeBetweenness wider and bluer stroke. Master regulator, hub-TF SPI1 is colored yellow, its regulatory links - red arrows points to regulated targets: MAFB, IRF5, ATF3 and ZNF710. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)
transcription factors was found upregulated after a fat loss and morphogenesis in all multicellular organisms. [47] studied differentiating embryonic stem cells, found that depletion of CBX8 partially impairs the transcriptional activation of differentiating genes. HOXD3, HOXD4, and HOXC6 belong to the homeobox family genes, which encode highly conserved TFs playing an important role in morphogenesis in all multicellular organisms [40]. HOXD4 may be involved in the development of atherosclerosis [46]. HOXC6 was found upregulated after a fat loss [47]. Functional annotation results, hub-TFs and MR of the Red obese module from the obese sub-network analysis show its biological role in immunity and osteoporosis related functions.

The Magenta lean module (57 TFs) had the strongest absolute MS (COR = 0.26) of the TF significance in lean vs. obese network and enrichment with leanness-associated TFs (Fig. 2). The module displayed KEGG pathways in diabetes (e.g. Basal transcription factors, Padj = 4.20E-05) [48], type 2 diabetes, obesity and osteoporosis (e.g. Notch signaling pathway, Padj = 2.40E-03) [49, 50]. Hub-TF VDR (vitamin D receptor) was found after the intra-modular analysis. [51] indicated VDR association with obesity in type 2 diabetic subjects. [52] displayed that VDR-knockout mice had a slower growth rate and accumulated less fat mass than the wild type mice. [53] suggested VDR as Quantitative Trait Loci (QTL) for fat mass variation in young Chinese men. VDR is a prime candidate gene for osteoporosis [54]. iRegulon analysis did not detect any present MRs within the module. Highest Degree had LMO4 and CCNK, Stress - CREB3L4. A study by [55] identified LMO4 as a novel modulator of leptin function. In a study by [56]. CCNK may regulate activity of cyclin-dependent kinases (Cdks), where Cdks knockout mice had insulin deficient diabetes. CREB3L4 was found inhibiting adipogenesis in preadipocytes [57]. Functional annotation results, hub-TFs and MR of the Magenta lean module from the lean sub-network analysis show its biological role in diabetes, obesity and osteoporosis related functions.

3.2.1. Common Differentially Expressed TFs in Significant Modules of Sub-networks

TFs of the modules with the strongest absolute MS were compared if they are differentially expressed between lean and obese groups. Transcription factors present in the Red obese module were aligned to the TFs in the Magenta lean module. Common TFs found between the modules with their expression values are listed in Table 2. The Red obese module and the Magenta lean module had 11 TFs in common: RARG, RP11-644F5.10, PAX9, ZNF404, ZNF618, ZFP64, GTF2I, FHL2, NPLR, CBX8 (hub-TF) and PHF14. A subset of the common TFs were found differentially expressed between the modules. ZNF618 (Zinc Finger Protein 618) was downregulated in the Red obese module and upregulated in the Magenta lean module. Zink finger was found containing proteins function in various biological processes: gene transcription, translation, cytoskeleton organization, epithelial development, cell adhesion, chromatin remodeling, etc. [58]. ZFP64,
CBX8 and PHF14 were upregulated in the Red obese module and downregulated in the Magenta lean module. Two of these three TFs we found to have associations with the immune system: ZFP64 [59] and CBX8 [60].

Table 2. Common TFs comparing significant modules between lean and obese sub-networks.

<table>
<thead>
<tr>
<th>TFs in Common</th>
<th>Expression Values of Modules in Sub-networks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The Red obese module</td>
</tr>
<tr>
<td>RARG</td>
<td>0.79</td>
</tr>
<tr>
<td>BP11-644F5.10</td>
<td>0.61</td>
</tr>
<tr>
<td>PAX9</td>
<td>0.22</td>
</tr>
<tr>
<td>ZNF404</td>
<td>0.55</td>
</tr>
<tr>
<td>ZNF618</td>
<td>0.36</td>
</tr>
<tr>
<td>ZFP64</td>
<td>0.85</td>
</tr>
<tr>
<td>GTF21</td>
<td>0.61</td>
</tr>
<tr>
<td>FHL2</td>
<td>0.60</td>
</tr>
<tr>
<td>NPRL2</td>
<td>0.80</td>
</tr>
<tr>
<td>CBX8</td>
<td>0.92</td>
</tr>
<tr>
<td>PHF14</td>
<td>0.72</td>
</tr>
</tbody>
</table>

(For interpretation of the references to color in this Table, the reader is referred to the web version of this paper.)

3.2.2. Analysis of Differentially Connected TFs in Sub-networks

In the lean vs. obese network we detected 53 differentially connected TFs which were more highly connected in the obese sub-network than in the lean sub-network. Four of them, ZZZ3, GTF3C3, ZNF10 and STAT2, were also found to be differentially expressed (t-test > 1.96). In addition, we found 26 differentially connected TFs that were more highly connected in the lean than in the obese sub-network, two of which (NLRP3 and FHL1) were found to be differentially expressed (t-test > 1.96). STAT2 was described above. Very little is known about GTF3C3 and ZNF10; however, in recent genome-wide meta-analysis study [61] ZZZ3 was identified as a new locus for clinical classes of obesity. NLRP3, previously found as DE in [62], contributes to obesity-induced inflammation and insulin resistance [58] and was recently proposed as a new therapeutic target for diabetic complications [55]. Some of these results confirm the findings by [62].

3.2.3. Differences in the Construction of Different Regulatory Networks

WTFCN constructs an undirected network; however, using iRegulon we increased the confidence of our findings with respect to the hub-TFs. Recently, iRegulon in combination with Genomica software (a software tool for regulatory network construction) was used to build a regulatory network for insulin-mediated genes [63], but iRegulon in conjunction with WGCNA has never been used before. The use of different statistical bioinformatics methods and computational algorithms might lead to a difference in network construction and thereby altered results. For instance, Bayesian network integrates prior knowledge in a principled manner to increase the inference reliability [64, 65]. Higher performance of inference is achieved in simulated and real data [66]. Bayesian networks specify the joint probability distribution over all genes down to the conditional distributions. This network could be used to compute the probabilistic relationships between DO and trait. The Ordinary Differential Equation (ODE) models the regulatory system, but not directly infers the regulatory network. In ODE models, gene regulations are modeled by derivative equations which quantify the change rate of gene expression of one gene in the system as a function of expressions of all related genes that refer to its regulators. ODE model-based linear programming can also be used to quantify the rate of change of gene expression as a function of the expression of other genes [67]. The comparison of different network algorithms and different computing strategies in our data is beyond the scope of this original research article.

CONCLUSION

In this study, we utilized publicly available RNA-Seq data (along with other metadata) from a pig model for human obesity deposited at the NCBI's Gene Expression Omnibus that were accessible through GEO Series accession number GSE61271. Using this data, we constructed a novel Weighted TF Co-expression Networks (WTFCN) to elucidate the complex patterns of gene regulation and master regulators involved in obesity development. A single WTFCN was constructed using TF expression data of lean, obese and median pigs, and TF modules acting on the development of obesity were identified by weighted scale-free topology network methods. Modules having strongest absolute MS in lean and obese sub-networks represent more specific TFs acting in lean or obese animals. Integrating TFs’ expression data from these modules into Cytoscape explained functional relationship, interaction, dynamics and role in expression within the network. Our results obtained after WTFCNs and Cytoscape analysis were overlapping and supplemented each other and new regulatory links between MRs (SPI1, HOXD4) and their targets were drawn in WTFCNs, using iRegulon analysis. These findings support the idea of complex gene regulation, alternative splicing of target genes in different conditions. Our results suggested TFs that are potentially either causal or regulatory for degree of obesity. One of the most promising TFs in this research was SPI1, which was detected as a hub-TF in WTFCNA and as a MR in iRegulon. Such overlapping between two different methods confirms SPI1 being a true hub-TF. As described above, SPI1 is a strong regulator of obesity, immunity and osteoporosis. Furthermore, SPI1 acts in macrophages and the phagocytosis process. Its four upregulated targets (MAFB, IRF5, ATF3, and ZNF710) are known in regulation of adipogenesis, BMI, coronary artery diseases,
hypertension and diabetes. The above-mentioned evidence supporting our results indicates that SPI1 is a possible major regulator in obesity. However, experimental validation should be performed before use and for better understanding of the complex role of interactions of highlighted nodes. Promising MRs and TFs should be validated by independent experiments. To the best of our knowledge, this is the first study that extended the principle of WGCN and integrated it with Cytoscape in order to find hub-TFs, MRs or genes that are potential key regulators in obesity of the majority of TFs. Our results on TF networks for obesity are expected to lead to a better understanding of the complex TFs regulation of obesity development and provide insights into detection of TFs or drug targets, predictive genetic and biomarkers for prevention and prediction of obesity and obesity-related diseases.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

CONSENT FOR PUBLICATION

Not applicable.

CONFICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

LJAK was supported by a grant from the Danish Innovation Fund for the BioChild Project (grant number 0603-00457B and project website: www.biochild.ku.dk). HNK and RSJ thank the EU-FP7 Marie Curie Actions Grant (CIG-293511) and FeedOMICS project grant from DFF-FTP (grant number 4184-00268B) for partially supporting publication of this work and funding their time spent on this research.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher’s website along with the published article.

REFERENCES
