Systems Genetics Analysis of obesity in a Porcine Model using WISH Network Method

L.J.A. Kogelman[1], S.D. Pant[1], J. Karjalainen[2], Lude Franke[2], M. Fredholm[1], and *H. N. Kadarmideen[1] *(*Correspondence)*
[1]Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark; [2]Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands

ABSTRACT: High-throughput genotype data have been extensively used to investigate the biological and genetic background of complex traits, such as obesity. However, results explaining complete genetic variation are limited. Systems genetics approaches are increasingly used to elucidate more of the currently limited knowledge of complex traits. Weighted Interaction SNP Hub (WISH) network method is a novel tool which is able to elucidate underlying biology and capture potential causal variants. Here we studied obesity, using systems genetics approaches in an F2 pig population. We created an Obesity Index (OI) based on multi-trait selection indexes containing nine obesity-related phenotypes, to perform a GWAS and WISH network analysis. WISH detected several modules associated with OI. Functional annotation revealed several genes (e.g. NPC2) and pathways (e.g. the insulin signaling pathway) which were (in-)directly related to obesity. This study shows the potential for systems genetics analysis of high-throughput genotype data for obesity.

Introduction

High-throughput genotype data (HTG) are extensively used for complex traits and diseases in genome-wide association studies (GWAS) to find genes associated with the trait of interest. Unfortunately, to date the detected loci explain a limited amount of the genetic variation in many complex traits (Manolio et al. (2009)), as well as do not provide holistic understanding of the systems genetics. Obesity is a complex trait that is associated with various serious diseases (e.g. type 2 diabetes and various types of cancer). Because of the exponential rise in the prevalence of obesity, and its financial, social and welfare consequences, the urge for a decent biological and genetic knowledge is rising. Here, we use a porcine model for human obesity, since the pig has similar metabolic, digestive and cardiovascular features, and it resembles humans more than rodents (Spurlock and Gabler (2008)). We have created an F2 pig population, and demonstrated the potential of this population to study human obesity (Kogelman et al. (2013)).

In complex diseases, it has been indicated that gene *x* gene interactions may have a key role, which is not taken into account in GWAS (Cordell (2009)). Systems genetics approaches, which include the effect of interacting alleles and their functional/biological role, may reveal more of the biological and genetic background of complex traits and diseases (Kadarmideen et al. (2006)); Civelek and Lusis (2013); Do et al. (2014)). Several network approaches have been developed to understand the interaction of genes and their influence on complex diseases (Segal et al. (2003); Diez et al. (2010)). Recently, we published the Weighted Interaction SNP Hub (WISH) network method based on HTG, which makes it possible to identify clusters of highly interconnected SNPs (modules) that are related to disease risks or a phenotype (Kogelman and Kadarmideen (2014)). The WISH network method detects those modules using the interconnectedness between SNPs, based on genotype interactions. Functional annotation and pathway analysis of detected modules may lead to the identification of biologically relevant pathways underlying those phenotypes.

In the present study, we applied systems genetics approaches in the F2 pig resource population, to unravel pathways and detect potential causal obesity genes. This included a GWAS and WISH on OI; and pathway profiling and functional enrichment analyses on the resulting variants and SNP modules. These systems genetic approaches revealed novel biological and genetic determinants not otherwise identified only by GWAS.

Materials and Methods

Experimental design. We have created an F2 pig resource population (Kogelman et al. (2013)), by generating an intercross of two production breeds (Yorkshire and Duroc) and the Göttingen Minipig. While the production breeds are prone to leanness and muscle growth due to genetic selection, the minipigs are prone to obesity and share metabolic impairments seen in obese humans (Johansen et al. (2001)). The F2 population (454 pigs) was intensively phenotyped (e.g. weight, body confirmation, DXA scanning and slaughter characteristics) and genotyped using the Illumina Porcine 60K SNP Chip.

Obesity Index. Multi-trait selection indexes, based on estimated breeding values (EBVs) for several traits of interest, are intensively used in animal breeding to determine extreme phenotypes (Cameron (1997)). We created the Obesity Index (OI) to determine the degree of obesity of the F2 animals, by calculating selection index weights and combining the EBVs for nine different obesity-related traits (reported in Kogelman et al. (2013)) into one aggregate total merit index. Traits selected for the OI were: weight and abdominal circumference at slaughter, average daily gain, estimated fat mass and percentage of fat at DXA scanning, back fat thickness at position 1 and position 2, weight of leaf fat and weight of omental fat at slaughtering. The OI followed, as expected, a normal distribution among the F2 animals.

Genome-wide association analysis. A GWAS was performed using the R package GenABEL (Aulchenko et al. (2007)) to test associations between the SNPs and the OI. We ran the basic linear model without correcting for any effects, as the OI is constructed based on EBVs which are already corrected for the population structure and other fixed effects: $y = \mu + g + e$, where y = OI, μ = the phenotypic mean, g = the SNP genotypes (codes as 0,1, and 2) and e = the model errors. We calculated the Bonferroni corrected p-

values resulting in a suggestive association at $P_{adj} = 1.32E^{-6}$ (0.05/SNP number) and a highly significant association at at $P_{adj} = 2.64E^{-8}$ (0.001/SNP number).

Combined linkage disequilibrium linkage analysis. Findings of highly significant regions in the GWAS were confirmed using a combined linkage disequilibrium linkage analysis (LDLA) approach, using the information of the intercross pedigree (paper presented in this conference by Pant et al.). Statistical significance was calculated via a likelihood ratio test of the full model (OI regressed over the phenotypic mean and Identity by descent probabilities of chromosomal segments flanked by successive marker pairs) vs. the null model (containing only the phenotypic mean). The Identity by descent probabilities were estimated using a linkage disequilibrium multi-locus iterative peeling algorithm described in Meuwissen and Goddard (2010).

WISH network construction. We previously published and described the WISH network method (Kogelman and Kadarmideen (2014)). This method assumes that SNPs which are highly correlated with each other (clusters of SNPs, here termed modules), will work co-operatively in pathways. Because of computational limits the top 2500 SNPs were selected based on their genome-wide significance (GWAS) and connectivity (sum of correlations of a particular SNP with all other SNPs). The adjacency matrix was calculated as the Pearson's correlations between SNP genotypes, raised to the power of 5 to create a scale free network. A SNP dendrogram was created based on the topological overlap measure (TOM), representing the relatedness between SNPs, to identify clusters of highly interconnected SNPs (modules). The correlation between the eigenvalue of each module (1st principal component) and the phenotypes of interest (OI and other obesity related phenotypes) was used to create the Genome-wide Module Association Test (GMAT) for selection of potential biologically interesting modules.

Systems biology analyses. All significant GWAS top SNPs and identified SNP modules by the WISH network were further analyzed by detecting overrepresented gene ontology (GO) terms and pathways. Genes located at the detected SNPs were obtained using the NCBI2R R-package (available at http://cran.r-roject.org/web/packages/NCBI2R/), using a flanking distance of 20 kB to cover the promotional regions of genes.

For downstream analyses we used GOEAST (Gene Ontology Enrichment Analysis Software Toolkit, Zheng and Wang (2008)) and GeneNetwork (http://genenetwork.nl). In GOEAST, the Gene Batch tool was used to import the gene symbols, and following the hypergeometric statistical test, with the Yekutieli multiple-testing adjustment method, was performed to identify significantly overrepresented GO terms and corresponding pathways. GeneNetwork was used to identify overrepresented GO terms, KEGG pathways, phenotypes and tissues. The database is constructed using human, mouse and rate expression data. Gene functions were predicted against known pathways and gene sets in various biological databases. Overrepresentation of GO-terms and pathways was tested using the Mann-Whitney U test, and P-values were afterwards corrected for multiple testing using the Bonferroni correction.

Results and Discussion

Genome-wide association analysis. The GWAS results were presented in a Manhattan plot (Fig. 1), showing 1024 SNPs above the suggestive significance threshold ($P_{adj}=1.32E^{-6}$) and 404 SNPs above the highly significance threshold ($P_{adj}=2.64E^{-8}$).

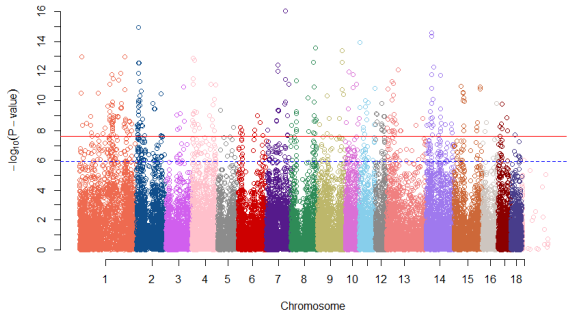

Figure 1. A Manhattan plot of GWA study single-locus P-values. The blue dash line indicates a suggestive significance threshold with adjusted Bonferroni correction at $P_{adj}=1.3E^{-6}$ and the red line indicates a highly significant threshold with adjusted Bonferroni correction at $P_{adj}=2.6E^{-8}$.

Three of the most significant SNPs are rs81396056 ($P=9.09E^{-17}$), rs81238148 ($P=1.13E^{-15}$) and rs81416774 ($P=2.56E^{-13}$), which are situated within the NPC2 gene, OR4D10 gene and CACNA1E gene, resp. The NPC2 gene is encoding a protein that plays a role in the regulation of cholesterol transport through the late endosomal/lysosomal system, affecting cholesterol homeostasis (Storch and Xu (2009)). The OR4D10 gene is an olfactory receptor gene, which are responsible for the perception of smell, through neuronal responses, and thereby affecting eating behavior (Nasser (2008)). The CACNA1E gene encodes a protein in a voltage dependent calcium channel. Variants in this gene have been associated with type 2 diabetes, insulin resistance, and impaired insulin secretion in non-diabetic subjects (Trombetta et al. (2012)). Highly significant GWAS results were confirmed by the LDLA approach.

GeneNetwork finds various cellular, transport related processes, as e.g. the GO Biological Processes (BP) *cell chemotaxis* ($P=6E^{-6}$), and the KEGG pathway *endocytosis* ($P=1E^{-6}$). The associated phenotypes (according to the Mouse Genome Informatics (MGI) database) show a clear share for inflammatory related phenotypes, e.g. *decreased inflammatory response* ($P=3E^{-5}$) and *decreased macrophage cell number* ($P=4E^{-5}$). GOEAST detected a clear enrichment for GO terms associated with the glucose/insulin metabolism in the BP category, e.g. *negative regulation of insulin secretion* ($P=2.39E^{-7}$) and *cellular response to glucose stimulus* ($P=2.94E^{-11}$). Other overrepresented GO terms in the BP category were *glycolysis* ($P=6.01E^{-7}$) and *skeletal muscle fiber development* ($P=1.91E^{-5}$).

WISH network construction. Using the WISH network method, we detected 17 modules of at least 50 SNPs per module. Based on the GMAT (Fig. 2) we selected two modules for downstream systems biology analyses.

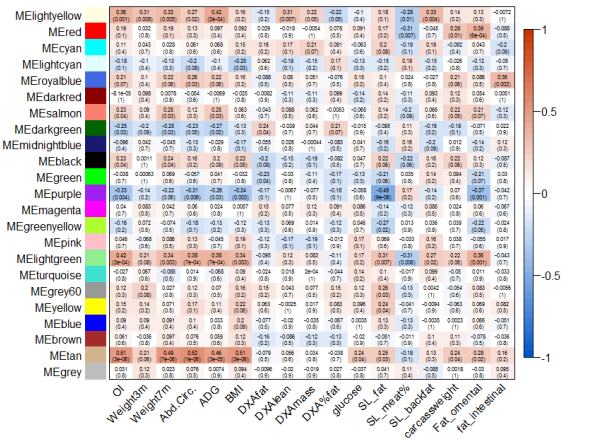

Figure 2. GMAT of modules detected using the WISH network method, representing the positive (blue) and negative (red) correlations with obesity related phenotypes.

The GeneNetwork workflow run on Tan module ($GMAT_{OI}$=0.51, 64 unique genes) revealed, various overrepresented phenotypes related to diabetes (i.e. *increased susceptibility to autoimmune diabetes*, P=4.13E^{-4}; and *abnormal pancreatic beta cell morphology*, P =6.58E^{-4}). Those may be resulting from the overrepresented BP GO term branched chain family amino acid metabolic process (P =6.98E^{-5}), as those amino acids are associated with the metabolic homeostasis (Wang and Guo (2013)).

The GeneNetwork workflow run on the Lightgreen module ($GMAT_{OI}$=0.42, 47 unique genes) revealed the highly overrepresented GO process purinergic receptor activity (P=7.21E^{-25}). Purinergic receptors have been implicated in several different functions as e.g., learning and memory, locomotor and feeding behavior, and sleep (Burnstock (2010)). One of the genes mainly causing the overrepresentation of this pathway is P2RX7, encoding the protein P2X purinoreceptor 7, which have been implicated in ATP-mediated cell death, control of receptor trafficking, and inflammation. Moreover, the P2X7 has been associated with diabetes, e.g. by its influence on the regulation of beta cells (Glas et al. (2009)). Furthermore, the BP category in GOEAST also shows the overrepresented GO term fructose 2,6-bisphosphate metabolic process (P=3.95E^{-20}), which activates the glycolysis pathway and inhibits the gluconeogenesis pathway, and consequently regulating glucose homeostasis. These results are also consistent with regulation of feed intake and weight gain in pigs (Do et al. (2014)).

Future studies will include extension of the pathway analyses using the DEPICT (Data-driven Expression-Prioritized Integration for Complex Traits) method (Pers et al. (2013)

Conclusion

This study shows the potential of systems genetics methods that involve network and pathway based approaches to jointly use phenotypes and high-throughput genotype data. Particularly WISH method moves GWAS well beyond its scope to profile pathways and build underlying genome-wide genetic interaction networks involved in complex traits. We have shown that such an approach could lead to elucidating pathways and complex gene-gene interaction networks involved in obesity. We detected many diverse genes (e.g. NPC2 and OR4D10), and consequently many different pathways (e.g. glycolysis pathway) were distinguished. This study, to the best of our knowledge, is the first systems genetics analyses on a large scale porcine model in which a wide range of obesity traits has been investigated. Consequently, it has revealed much of the complexity of obesity as a disease and provides valuable biological insights into human obesity.

Acknowledgements

The project is supported by a grant from the Ministry of Science and Technology to the "UNIK Project for Food Fitness and Pharma for Health", funding from the Danish Council for Strategic Research to BioChild Project (www.biochild.ku.dk), and from a Ph.D. stipend awarded to Lisette J.A. Kogelman from University of Copenhagen. Haja N. Kadarmideen thanks EU-FP7 Marie Curie Actions – Career Integration Grant (CIG-293511) for partially funding this research.

Literature Cited

Aulchenko Y.S., Ripke S., Isaacs A., et al (2007). Bioinf. 23:1294-1296.
Burnstock G. (2010) The Open Neuroscience Journal 4:24-30.
Cameron N.D. (1997). CABI, ISBN-10: 0851991696.
Civelek M. and Lusis A.J. (2013). Nat Rev Gen. 15:34-48.
Cordell H.J. (2009) Nat Rev Gen 10: 392-404.
Diez D., Wheelock A.M., Goto S., et al (2010), Mol. BioSyst. 6:289-304.
Do D.N., Ostersen T., Strathe A.B., et al. (2014). BMC Genet. 15:27.
Glas R., Sauter N.S., Schulthess F.T. et al. (2009) Diabetologica 52: 1579-1588.
Johansen, T., Hansen H.S., Richelsen, B., et al (2001). Comp. Med, 51(2): 150-155.
Kadarmideen H.N., Von Rohr P., Janss L.L.G. (2006). Mamm Gen. 17:548-564.
Kogelman L.J.A., Kadarmideen H.N., Mark T., et al (2013). Front in Genet, 4(29).
Kogelman L.J.A and Kadarmideen H.N. (2014). BMC Syst. Biol. (accepted).
Manolio T.A., Collins F.S., Cox N.J. et al. (2009). Nat. 461: 747-753.
Meuwissen T. and Goddard M. (2010). Genetics 185: 1441-1449.
Nasser J. (2008). Obes Rev 2(4):213-218.
Pers T.H., Karjalainen J., Hirschhorn J.N., et al. (2013) 63rd Annual meeting of the ASHG, 22-26 Oct 2013, Boston, MA
Segal E., Shapira M, Regev A., et al. (2003). Nat. Gen. 34:166-176.
Spurlock M.E. and Galer N.K. (2008). J Nutr 138: 397-402.
Storch J. and Xu Z. (2009) BBA-Mol Cell Biol L 1791: 671-678.
Trombetta M, Bonetti S., Boselli M., et al. (2012) PLoS ONE 7:e32755.
Wang C. and Guo F. (2013) Chin Sci Bull 58:1228-1235.
Zeng Q. and Wang X-J. (2008) Nucl. Acids res. 36: W358-W363.