Sarah Monod, Aline Beermann, Luca de Fusco
Axel Aubert and Sidonie Nicole 07.05.2020

Weighted Correlation Network Analysis (WGCNA)

Introduction

Many diseases are associated with different patterns of gene expressions. Humans have a set of about 25,000 genes, each having a different expression between them and between individuals. To find out which gene is involved in a disease, the genes would have to be inspected one by one and their expression between healthy and sick people would have to be compared. Nevertheless, this technique is complicated to implement, because it would require a significant investment of time and money. Indeed, it is complicated to analyze large sets of data and to extract interesting information by taking the expressions of the genes separately. At the same time, it is possible that interesting interactions are not considered. Researchers have created an algorithm called WGCNA for "weighted correlation network analysis" which allows to analyze large sets of data and to extract only the important information.

Based on preliminary studies on the influence of certain proteins on a given phenotype, it may be interesting to find out which genes are co-expressed or have a similar expression pattern for a specific phenotype. Using the WGCNA algorithm, it is indeed possible to find groups of genes that have a high correlation with a particular phenotype. Once these groups have been created, the groups expressing a high correlation with the chosen phenotype can be analyzed in more detail. The WGCNA algorithm is able to predict within this group of genes which genes will have a stronger influence on the others. They will thus be called "driver gene" as well as those with a very strong importance on the phenotype.

Hypothesis

The KCNF1 gene may interact with KCN genes based on evidence that they act on the same domain of the DLG1 protein. The phenotype called uric acid could be associated with the presence of hypertension and cardiovascular disease. This phenotype could be modulated by the expression of potassium channels that can be encoded by several different genes such as KCNA10, KCNF1 and KCNU1.

We hypothesize that Keratin 24 and Keratin 39 could be used as biomarkers for colorectal cancer being both co-expressed in the same module of genes in WGCNA. Additionally, TAS2R7 also found in this same gene module could cause weight gain and high hctn and therefore indirectly lead to colorectal cancer. The phenotype called weight or high hctn (homologue of the amino acid cysteine), which is a non-classical cardiovascular risk factor, could be associated with the presence of obesity or cardiovascular diseases. This phenotype could potentially be modulated by the expression of Keratin 24 and Keratin 39. Moreover, the taste receptor TAS2R7 could also play an important role for weight and cardiovascular diseases.

Method

The data coming from CoLaus were collected from the lymphocytes cell lines of 500 candidates. To proceed to the analysis of the male data, we used the WGCNA technique. To support the WGCNA results, we entered all the genes from a group into DAVID software. This will give additional information about the biological involvement of the genes. Those informations allowed us to deepen

our research by using literature and to understand the gene function on the phenotype.

Results

Based on the following matrix, we decided to focus on the group of genes “darkmagenta”, which was correlated to the phenotype “uric». Another group of gene, “yellowgreen”, was interestingly correlated with the phenotypes “wt” and “hctn”.

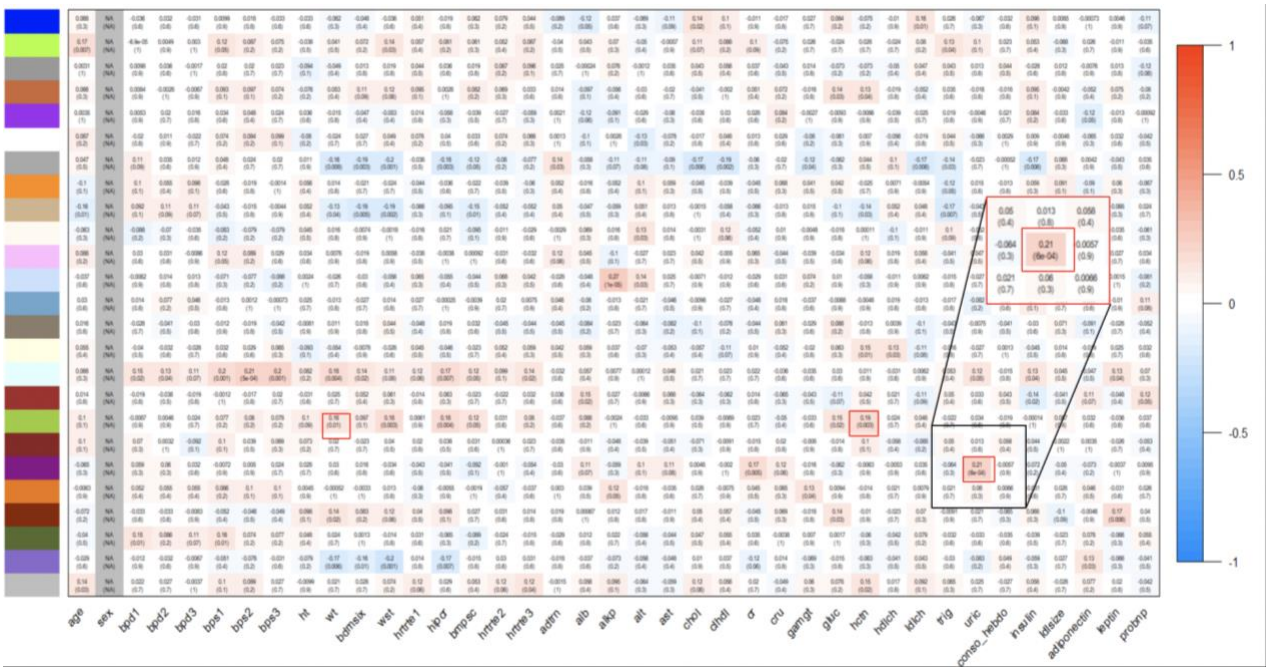

Figure 1: *The final matrix concerning the male data, each square represents the correlation between a group of genes and a phenotype. The x axis represents the phenotypes. The y axis represents the groups of genes, classified by color. On the right, the color code which indicates the value of the correlation: the blue color represents a negative correlation and the red one represents a positive correlation.*

The driver genes were searched for each group of genes.

Darkmagenta module :

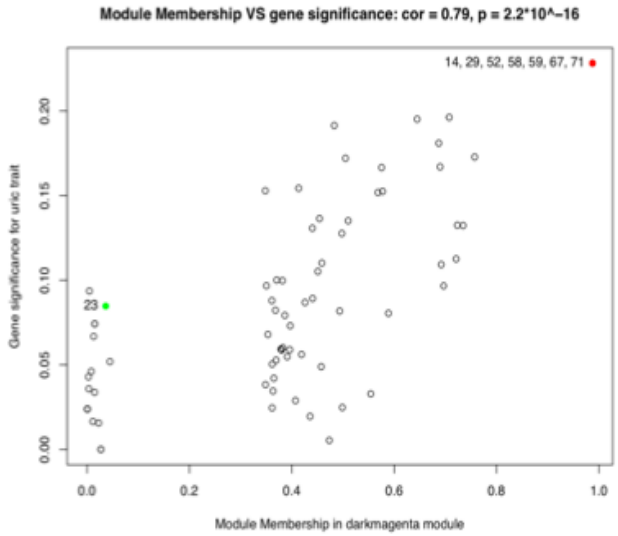

Figure 2: *Plot of each gene in the group with a high correlation with the phenotype URIC (0.21). The x axis represents the importance of the gene inside our group of genes. The y axis represents the gene significance for the uric trait.*

We observe, in this first plot, the different genes in the group. The red dot is the most relevant for this study. They are the “driver genes”. The green one is relevant as it is functionally related to one of our driver genes.

Yellowgreen module :

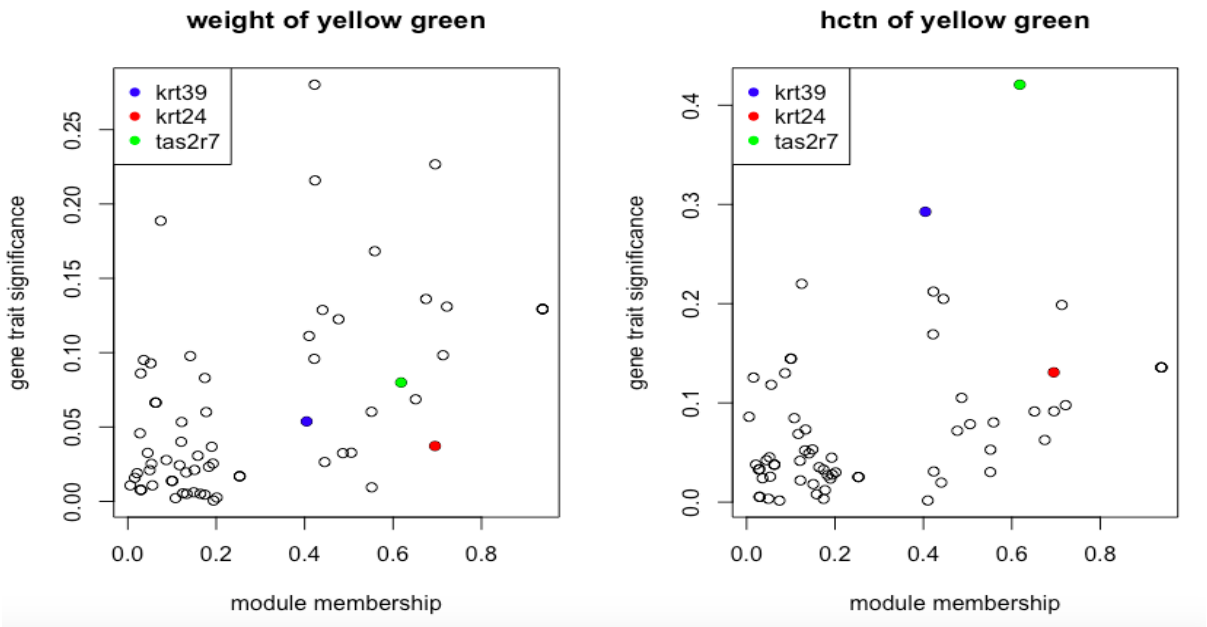

Figure 3: *Plot of each gene in the group with a high correlation with the phenotype weight (0.16) and hctn (0.18). The x axis represents the importance of the gene inside our group of genes. The y axis represents the gene significance for the weight/hctn trait.*

Discussion

KCNF1:

The group “darkmagenta”, in which the gene KCNF1 is found, is correlated with the expression of the uric phenotype. It has been shown that uric acid may be associated with hypertension and cardiovascular disease.

The KCNF1 gene is a gene of particular interest because its protein can form heteromultimers with other types of KCN proteins. Another interesting point is that it interacts with the DLG1 protein which is a multidomain protein essential for normal development. It has been discovered that DLG1 is involved in many functions of the body's organization such as the recruitment of potassium ion channels and the regulation of cell proliferation.

Studies on potassium ion channels have shown that some KCN proteins also interact with DLG1. In the following 2 studies [Ademuyiwa S. Aromolaran and al.] and [Luiz H. C. Vasconcelos and al.], it was shown that potassium ion channels are highly implicated in obesity. It shows that these channels are involved not only in the development of obesity but also in its continuity. Their involvement occurs in many ways, such as the participation in cell proliferation, in food intake control, in body energetic expenditure and many others.

Among these results one of them exposed the fact that the potassium ion channels are required to pass from the G phase to the S phase during the cell cycle which allows the cells to proliferate. In the case of obesity, proliferation of adipocytes is a crucial point for its development.

We suggest that changes in the expression of potassium ion channel genes may be involved in the pathophysiology of obesity. Indeed, KCNF1 gene may interact with KCN genes based on evidence that they act on the same domain of the DLG1 protein. Thus, change in the recruitment rate of potassium ion channels could allow the development and the maintenance of obesity.

Keratin24:

Keratin 24 is a gene usually highly expressed in keratinocytes, placenta, colon, and spleen. This gene has a major role in maintaining the epidermis, causing keratinocytes differentiation, senescence and apoptosis (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5376294/).

This specific type of keratin is not one of the usual cancer markers but has recently been identified in a study to be linked to colorectal cancer (Y. Hong et al. *Human cancer biology*, 2007). Keratin 24 is part of a group of seven genes, CYR61, UCHL1, FOS, FOS B, EGR1, VIP, and KRT24, that were found to be upregulated in cancer patients during a study. They cause, linked to other tumorigenesis factors (Wnt, PI3K, MAP kinase, hypoxia, G protein-coupled receptor) a disruption of cellular homeostasis in the intestinal mucosa.

This study means that Keratin 24 could potentially be used as a marker in order to identify colorectal cancer in subjects at an early stage. Finding Keratin 24 in this module of genes by WGCNA is really interesting, because obesity and weight gain are known risk factors for colon cancer.

Keratin39:

BY WGCNA, we also found Keratin 39 expressed in the same gene module as KRT24. It could mean that k39 is also a signal of colon cancer. Keratin 39 is nowadays known for its cytoskeletal function in human hair (https://www.nextprot.org/entry/NX_Q6A163/). It is also a type I cytoskeletal keratin, as keratin 24.

Knowing Keratin 39 is expressed in the same WGCNA gene module than Keratin 24, this gene could somehow also be involved in colon cancer. Even though there was no direct interaction to be found between Keratin 24 and Keratin 39, KRT39 was demonstrated to interact with high affinity with PTEN (https://thebiogrid.org/219771/publication/lncrnas-directed-pten-enzymatic-switch-governs-epithelial-mesenchymal-transition.html). This interaction is quite interesting as PTEN being a known tumor suppressor was found to be altered in colorectal cancer (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3893597/). Restoring the wild-type function of PTEN could be a therapeutic strategy for patients with colon cancer (https://www.ncbi.nlm.nih.gov/pubmed/30623366). Moreover, PTEN was also found to have an important role in regulating obesity and colon cancer (https://www.ncbi.nlm.nih.gov/pubmed/22970944).

In fact, PTEN has a role in inhibiting the constitutive activation of the AKT pathway. AKT is a kinase that will phosphorylate its targets once being active leading to cell growth, decreased apoptosis and cell proliferation. Cancerous cells can use this kind of pathway to spread more rapidly. Thus, PTEN allows us to turn off this pathway to stop proliferation of cells. A constitutive activation of the AKT pathway was in fact found to lead to colorectal cancer with the involvement of another keratin gene, Keratin 80 ((https://www.nature.com/articles/s41419-018-1030-y). The AKT pathway could therefore be linked to colorectal cancer with perhaps a possible involvement of Keratin 39 strongly interacting with PTEN.

TAS2R7:

We also found by WGCNA that the gene TAS2R7 is highly expressed in our database when the people had high weight and especially that this gene was one of the most important genes for hctn in our module, having a very high gene trait significance with hctn. TAS2R7 is in fact a bitter taste receptor, that has a role in regulating the release of appetite regulating hormones.

Obese people usually depict an altered taste, and it was already found that this family of taste receptor TAS2R being usually highly expressed in cases of obesity could be the target for anti-obesity drugs (https://www.sciencedirect.com/science/article/abs/pii/S0196978120300334). As obesity and cardiovascular disease potentially leads to colon cancer, TAS2R7 could be seen as an indirect cause of colon cancer.

Perspectives and further researches

To confirm the interaction between KCNF1 and other KCN proteins (such as KCNA1, KCNA3, KCNA5 and KCND2), we could first investigate the possibility of a direct interaction between the proteins. To do so, we could use the Yeast-two hybrid technique.

Nevertheless, for the colorectal cancer hypothesis, it is important to take into consideration that we only

analyzed the data from healthy people having in our case a high weight and a high hctn cardiovascular factor. Thus, we would need to observe the expression of these genes in patients with colorectal cancer and look if the expression of these genes is still high or even higher in cases of cancer. Therefore, with colorectal cancer subjects, we could look if these three genes are highly expressed in cases of colon cancer using quantitative RT-PCR by looking at their mRNA expression pattern. As the two major phenotypes analyzed by WGCNA were weight and hctn, and as we are trying to link obesity and cardiovascular diseases with colorectal cancer, it could be interesting to analyze obese patients or cardiovascular disease patients having colorectal cancer and look at the expression pattern of these three genes.

It could also be of interest in which stages of colorectal cancer these genes could have a role, whether it is in early, middle or late stages of colorectal cancer. As analysis of cancer patients in humans encounters many ethical problems, it would be easier to start with a first analysis using cancer cell lines. By immunohistochemistry, we could also look at the specific structure of the tissue by using an antibody against the protein we want to test, for example with a coupled GFP fluorescent antibody. Then, with a fluorescent microscope, we would be able to observe in which cells our protein is specifically expressed. Furthermore, this could allow us to compare between a colon cancer condition and a normal healthy control condition.

Finally, as both hypotheses were tested by WGCNA on male subjects, we could also further on look if our results for the potassium ion channel and the genes involved colorectal cancer are also valid for females, by running the same WGCNA analysis on our female database. This will allow us to see if our assumptions and conclusions apply to both sexes or if there is a sex-related effect.