Microarray Integrated Analysis of a Gene Network for the CD36 Myocardial Phenotype

Imane Sabaouni1*, Brigitte Vannier2, Ahmed Moussa3 & Azeddine Ibrahimi1

1Medical Biotechnology Lab (MedBiotech), Rabat Medical & Pharmacy School, Mohammed Vth University in Rabat, Morocco; 2Receptors, Regulation and Tumor Cells (2RTC) Laboratory, University of Poitiers, France; 3LabTIC Laboratory, ENSA, Abdelmalek Essaadi University, Tangier, Morocco; Imane SABAOUNI – E-mail: imanesabaouni@yahoo.com *Corresponding author

Received July 18, 2016; Revised July 29, 2016; Accepted July 30, 2016; Published October 11, 2016

Abstract: CD36 is a multifunctional membrane-type receptor glycoprotein that reacts with oxidized low-density lipoprotein and long-chain fatty acid (LCFA). However, much remains to be understood about the molecular mechanism of the cardio-myopathy observed in CD36-KO mice. In this study, we identify different genes pathways involved in response to CD36 cardio-myopathy phenotype by identifying the differences among biological processes, molecular pathways and networks of interactions that emerge from knocking CD3 and using different bioinformatics tools such as STRING, GeneMANIA and Cytoscape. We were able list all the CD36-regulated genes, their related function and their specific networks. Data analysis showed that CD36-regulated genes differentially expressed are involved in biological processes such as FA metabolism, angiogenesis/apoptosis and cell structure. These results provide the first look at mechanisms involved in CD36 deficiency and development of cardio-myopathy and the opportunity to identify new therapeutic targets.

Keywords: CD36, cardio-myopathy, genes networks, genes pathways, metabolism, angiogenesis/apoptosis.

Background: Hypertrophic cardio-myopathy (“HCM”) is characterized by a myocardium hypertrophy. To date the molecular mechanisms underlying this pathology remain elusive [1-6]. It is most well known as a leading cause of sudden cardiac death in young athletes [7]. The occurrence of hypertrophic cardio-myopathy is a significant cause of sudden unexpected cardiac death in any age group and is a cause of disabling cardiac symptoms. Younger people are likely to have a more severe form of hypertrophic cardio-myopathy. HCM is frequently asymptomatic until sudden cardiac death. As a consequence, re-current screening has been suggested for certain populations [8].

The CD36 gene encodes a membrane glycoprotein (also known as platelet glycoprotein IV or IIIb), that is expressed in a wide variety cell types, including platelets, monocytes, erythroblasts, capillary endothelial cells and mammary epithelial cells [9-10]. In the heart CD36 is also expressed both in epithelial and muscle cells [11]. In these latter, CD36 has been shown to be a long chain fatty acids (FA) receptor/transporter [12]. In endothelial cells, it has been proposed as one of thrombospondin receptor [13]. CD36 is a multifunctional membrane-type receptor glycoprotein that reacts with thrombospondin, collagen, oxidized low-density lipoprotein and long-chain fatty acid (LCFA). 1,2 LCFA is one of the major cardiac energy substrates; hence, LCFA metabolism may have an important role in cardiac diseases. We present here a patient with hereditary hypertrophic cardiomyopathy (HCM) and type I CD36 deficiency that showed no myocardial LCFA accumulation.

CD36-KO mice exhibit between 60 and 80% reduction in beta-methyl-p-[123I]-iodophenyl-pentadecanoic acid (BMIPP) uptake by heart tissue [14] which is paralleled by an increase in the heart/body index and the left ventricular size [13]. We propose in this study to define the molecular defects that lead to cardiac hypertrophy and consequent of the fatty acid transporter CD36 deficiency. To define this molecular defect, we used two microarrays technologies (Affymetrix, Agilent) in order to identify genetic alterations related to the myocardial phenotype of CD36-Ko mice.

In a previous study [15], heterogeneity of the arrays data was analyzed by splitting them on three levels: genes, genes set, and network/pathway. We were able to identify the CD36-regulated candidate genes linked to Hypertrophic cardio-myopathy. At the second level we were able to cluster group of genes (gene sets) that may have similar functions [15]. In this study, we aim to ascertain whether each gene set from each subtype is significantly
enriched in a list of selected phenotypes. The third level of analysis was to construct gene networks of the proposed CD36 regulated genes from three selected web-based tools: Ingenuity Pathway Analysis (IPA), Search Tool for The Retrieval of Interacting Genes (STRING), and Gene Multiple Association Network Integration Algorithm (GeneMANIA).

Materials and Methods:
Dataset:
To identify a comprehensive set of genes that are differentially regulated by CD36 expression in the heart, we used two microarray technologies (Affymetrix and Agilent) to compare gene expression in heart tissues from CD36 Knock-Out (KO-CD36) versus wild type (WT-CD36) mice. The obtained results using the two technologies were similar with around 35 genes differentially expressed using both technologies [15]. Absence of CD36 led to down-regulation of the expression of three groups of genes involved in pathways of FA metabolism, angiogenesis/apoptosis and structure. These data are consistent with the fact that the CD36 protein binds FA and thrombospondin 1 involved respectively in lipid metabolism and anti-angiogenic activities [15]. Summary of dataset analysis methodology used in this study is shown in Figure 1.

![Figure 1](image1.png)

**Figure 1:** Summary of dataset analysis methodology used in this study

Gene Set Enrichment Analysis and Enrichment Map
Gene set enrichment analyses (GSEA) [16, 17] are commonly used to determine the biological characterization, statistical significance, and concordant differences between an experimental gene set and a selected gene list from annotated gene sets knowledge bases to red on Molecular Signatures Database (MSigDB). GSEA can be downloaded from http://www.broadinstitute.org/gsea/downloads.jsp. The Jaccard coefficient is used to compare the similarity between two sample gene sets A and B and defined as the intersection between group A and B divided by their union. The results from GSEA are then visualized through the enrichment map [18], a Cytoscape plugin for network visualization. The ranked experimental gene list along with the enriched gene sets from GSEA is used to build the network of gene sets (nodes) where edges represent their similarity. The size of a node varies by gene set size and the thickness of the edge represents the degree of correlation between two gene sets.

Networks and Pathway Analysis
Analysis of Network Invoked by CD36-Regulated Genes
The biological knowledge of gene and protein interactions is growing rapidly and there are many tools and curated databases available on a large scale. Insightful knowledge gained from studying gene sets rather than individual genes using network-based approaches can reveal network patterns and relevant molecular pathways from the experiment gene sets. In this study, we utilized two different freely accessible and user-friendly web tools as follows. Gene Multiple Association Network Integration Algorithm (GeneMANIA) [19, 20] (http://www.genemania.org/) is a web-based tool for prediction of gene function or implemented as a Cytoscape plugin tool. Based on single gene or gene set query from 7 organisms, it shows results for interactive functional associative network according to their co-expression data from Gene Expression Omnibus (GEO), physical and genetic interaction data derived from BioGRID, predicted protein interaction data based on orthology from I2D, co-localization, shared protein domain, and GO function. Search Tool for the Retrieval of Interacting Genes (STRING) version 9.1 [21, 22] (http://string-db.org/) is an online protein-protein interaction database curated from literature and predicted associations from systemic genome comparisons. The user can query using single or multiple name(s) and protein sequence(s). The protein interactions can be displayed according to their confidence, evidence, actions, or interactions.

Results and Discussion:
Identification of CD36-Regulated Genes through a Refined Analysis of Data
Our initial analysis of this time series gene expression data for differentially expressed gene identification followed the same method used in a previous paper [15]. All files were processed and normalized by Robust Multiarray Average (RMA) in R as in the original study. The selected normalization method may have an effect on downstream analysis, for example, reverse engineering analyses [15]; however, investigating this effect is beyond the scope of this study. We found that combining CD36+/+ and CD36-/- data compromises the accuracy of selection of differentially expressed. Table 1 summarizes the 30 differentially expressed genes from the combined dataset.

Networks and Pathways Analysis
Network analysis can help understand the molecular and cellular interactions [23]. It can be visualized to represent entities (nodes) and their relationships (arcs). The advent of high-throughput technology has led to a large increase in publicly available information. Each data type can capture different aspect of functional roles of interested genes. In this section, we investigated functional interaction among genes and proteins in the cell using available data and knowledge bases.
and perfused hearts, the existence of competition between source such as glucose. In 1963, Ran production as ATP, which causes the use of an alternative energy disturbances result in a significant decrease in energy carnitine long chain and an important membrane alteration. CoA causes inhibition of B oxidation. This leads to a decrease in free carnitine and accumulation of intra-mitochondrial acylcarnitine long chain and an important membrane alteration. These disturbances result in a significant decrease in energy production as ATP, which causes the use of an alternative energy source such as glucose. In 1963, Randle and his colleagues [24], have brought to light, using in vitro experiments on rat isolated and perfused hearts, the existence of competition between glucose and fatty acids with long chain fatty acids being preferentially metabolized. Moreover, according to Dyck et al. [25], CD36 plays an important role in the choice of substrate in the heart [25]. WT-CD36 which normally draws more of its energy from fatty acids is forced to uses more glucose in the absence of CD36 (The heart of the CD36-KO). Our work confirms these results as we see the over-expression of genes that stimulate glucose metabolism such as IRS1, IRS2, IRS3, IDE etc.

The results of genes and network that are functionally associated with the gene set from early response of CD36 are shown in Figures 3 and 4. We compared the two networks from STRING and GeneMANIA based on the interactions they revealed; here we used CD36 as the centre gene in the comparison. All interactions found in STRING were found in GeneMANIA as described in more detail in Table 2 (cf. Annex). Comparison between the results from both tools can be used to confirm the functional associations of the interested gene sets. There is evidence of overlap and uniqueness in the interactions revealed by the two web-based tools.

**Discussion**

In this report, we were able to show that our data is consistent with the role CD36 protein in the FA metabolism, angiogenesis/apoptosis and structure. In cardiac muscle cells, CD36 is known for its lead role as a receptor / transporter of long-chain fatty acids in heart cells and involvement in metabolism in general. Indeed, the ability of cardiomyocytes to adapt its energy demand is a determinant for myocardial function and the Fatty acids are the main energy source of the heart with oxygen. Thus, the CD36 absence results in a loss of absorption of long chain fatty acids and cause deleterious effects on the heart muscle, while a decrease in the availability of acyl-CoA causes inhibition of B-oxidation. This leads to a decrease in free carnitine and accumulation of intra-mitochondrial acylcarnitine long chain and an important membrane alteration. These disturbances result in a significant decrease in energy production as ATP, which causes the use of an alternative energy source such as glucose. In 1963, Randle and his colleagues [24], have brought to light, using in vitro experiments on rat isolated and perfused hearts, the existence of competition between glucose and fatty acids with long chain fatty acids being preferentially metabolized. Moreover, according to Dyck et al. [25], CD36 plays an important role in the choice of substrate in the heart [25]. WT-CD36 which normally draws more of its energy from fatty acids is forced to uses more glucose in the absence of CD36 (The heart of the CD36-KO). Our work confirms these results as we see the over-expression of genes that stimulate glucose metabolism such as IRS1, IRS2, IRS3, IDE etc.

Our results confirmed that showing that CD9 that encodes the CD9 protein and CD36 share the same expression profiles. Studies on platelets (for solubilization of the platelet membrane) identify CD9; a11b3 and a6b1 integrins are CD36 partners [27]. Immunofluorescence studies show that CD9 and CD36 αδβ1 integrin are co-located in endothelial cells. Thus it suggests that CD9, CD36, and alpha 6 beta 1 might form a complex in endothelial cells upon the binding of TSP1. Our work also also showed a differential expression of MAPK2 (also known as ESK2), a member of the serine /threonine kinase family, and

**Figure 2: Interaction network Visualization using Cytoscape².** Cytoscape and enrichment map were used for visualization of the results; only gene sets from Gene Ontology were used. Nodes represent enriched GO gene sets, whose size reflects the total number of genes in that gene set. Edge thickness represents the number of overlapping genes between gene sets calculated using Jaccard coefficient.

**Figure 3**
an activator of P38 which in turns activates a cascade that results in increased cell size [28]. In monocytic cells, the same results were observed since it was determined the existence of a physical association between the intra-cytoplasmic carboxyl terminus domain of CD36 and a signaling complex containing Lyn and MEKK2, upstream of the P38 MAP kinase cascade was de [29].

Figure 3: Gene network of CD36 derived from GeneMANIA. A gene network from GeneMANIA shows the relationships for genes from the list (nodes) connected (with edges) according to the functional association networks from the databases.

Data analysis showed that many Structural genes were altered in CD36-KO. Indeed, failing heart differs from the normal heart in function as well as in structure as failing heart is most often remodeled with hypertrophy. Cardiac imaging with the increases of the ventricular wall thickness and smaller ventricular chambers can clinically recognize hypertrophy. Our work confirms cardiac such remodeling as we see the expression of genes that stimulate cardiac cell structure such as: MYL4, TNNT2, HAND1, PDLIM3 AND PDGFRA.

Figure 4: Results from STRING search of Protein interaction for CD36. The figure illustrates the protein interaction upon querying STRING protein network (evidence view) in Mus Musculus with 25 proteins. Additional information from other resources can be retrieved for each protein and interaction. Nodes represent proteins and different line colours denote the type of evidence for the interaction.

Conclusion
In summary, we utilized the strengths of existing network/pathway tools and databases to gain insight into processes related to cardio-myopathy have distinct molecular interaction patterns visible from various systems levels, including gene (microarray analysis), gene set, molecular pathway, and gene networks have shown that cardio-myopathy could involve three pathways; FA metabolism, angiogenesis/apoptosis and structure. Discriminating between the three pathways can help to improve the understanding of a drug’s mechanism and further improve targeting in therapeutics drug research.

References
[1] Richardson P, et al. 1996 Circulation 93(5): 841 [PMID 8598070]
[2] Maron BJ, 2002 JAMA 287(10): 1308 [PMID 11886323]
[3] Sherrid MV et al. 2003 Ann Thorac Surg. 75(2): 620 [PMID 12607696]
[4] Wigle ED et al. (1985). 28(1): 1 [PMID 3160067]
[5] Wigle ED et al. 1995 92(7): 1680 [PMID 7671349]
### Table 1: Gene description with corresponding reference

| Gene Name | Accession Numbers | Synonyms / Identifier | Description | R |
|-----------|-------------------|-----------------------|-------------|---|
| CD36      | ENSG00000175445, ENSP00000309757, ENSP00000428237, ENSP00000428496, ENSP00000428557 | [11] | CD36 molecule (thrombomodulin receptor) (472 aa) | [a] |
|           |                   |                       | updated on 21-Jun-2015 |     |
|           |                   |                       | diacylglycerol-Oacyltransferase 1; Catalyzes the terminal and only committed step in triacylglycerol synthesis by using diacylglycerol and fatty acyl CoA as substrates. In contrast to DGAT2 it is not essential for survival. May be involved in VLDL (very low density lipoprotein) assembly. In liver, plays a role in esterifying excess fatty acids to glycerol. Functions as the major acyl CoA renot acyltransferase (ARAT) in the skin, where it acts to maintain retinoid homeostasis and prevent retinoid toxicity leading to skin and hair disorders (488 aa) |     |
|           |                   |                       | alcohol dehydrogenase 4 (class II), pi polypeptide (380 aa) |     |
|           |                   |                       | insulin receptor substrate 3 |     |
|           |                   |                       | insulin receptor substrate 1, May mediate the control of various cellular processes by insulin. When phosphorylated by the insulin receptor binds specifically to various cellular proteins containing SH2 domains such as phosphatidylinositol 3-kinase p85 subunit or GRB2. Activates phosphatidylinositol 3-kinase when bound to the regulatory p85 subunit (By similarity) |     |
|           |                   |                       | interleukin 2, Produced by T-cells in response to antigenic or mitogenic stimulation, this protein is required for T-cell proliferation and other activities crucial to regulation of the immune response. Can stimulate B-cells, monocytes, lymphokine-activated killer cells, natural killer cells, and gloma cells |     |
|           |                   |                       | FAT tumor suppressor homolog 4 (Drosophila), May function in the regulation of planar cell polarity. Cadherins are cell-cell interaction molecules (By similarity) |     |
|           |                   |                       | pyruvate dehydrogenase (lipoamide) beta, The pyruvate dehydrogenase complex catalyzes the overall conversion of pyruvate to acetyl-CoA and CO(2), and thereby links the glycolytic pathway to the tricarboxylic cycle |     |
|           |                   |                       | lipoprotein lipase, The primary function of this lipase is the hydrolysis of triglycerides of circulating chylomicrons and very low density lipoproteins (VLDL). Binding to hepatic surfate proteoglycans at the cell surface is vital to the function. The apolipoprotein, APOC2, acts as a coactivator of LPL activity in the presence of lipids on the luminal surface of vascular endothelium (By similarity) |     |
|           |                   |                       | lipoprotein lipase, The primary function of this lipase is the hydrolysis of triglycerides of circulating chylomicrons and very low density lipoproteins (VLDL). Binding to hepatic surfate proteoglycans at the cell surface is vital to the function. The apolipoprotein, APOC2, acts as a coactivator of LPL activity in the presence of lipids on the luminal surface of vascular endothelium (By similarity) |     |
| Uqcrh     | ENSG00000135218, ENSP00000309757, ENSP00000428237, ENSP00000428496, ENSP00000428557 | [17] | ubiquinol-cytochrome c reductase hinge protein, This is a component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex), which is |     |

©2016
Table 3: R = Reference

| Gene          | Description                                                                 |
|---------------|------------------------------------------------------------------------------|
| Ugtc1         | Enzyme that catalyzes the conversion of uridine to thymidine.              |
| Cyk           | Protein that regulates the cell cycle and proliferation.                   |
| Thbs1         | Transmembrane protein involved in cell adhesion and migration.            |
| Sdc4          | Cell surface glycan that regulates cell adhesion.                         |
| Pdgfra        | Platelet-derived growth factor receptor, involved in cell proliferation.   |
| Hand1         | Homeodomain protein involved in muscle development.                       |
| Arn2          | Protein involved in the respiratory chain.                                 |
| Rag1          | Guanine nucleotide diphosphate exchange protein.                           |
| Api5          | Protein involved in the cell cycle and apoptosis.                         |
| Map3k2        | Protein involved in the cell cycle and apoptosis.                         |
| Merk          | Protein involved in cell motility and chemotaxis.                         |
| Myl4          | Protein involved in muscle contraction and migration.                     |
| Tron2         | Protein involved in cell migration and differentiation.                   |
| Pfllim3       | Protein involved in muscle contraction and migration.                     |

Open access
Table 2: Interactions for CD36 (early response) network using GeneMania and STRING

| Interactions | GENEMANIA          | STRING                      |
|--------------|--------------------|-----------------------------|
| CD36 > Osgal1 | Co-localization    | Co-mentioned in PubMed Abstracts |
| CD36 > lpl    | Co-expression      | Co-mentioned in PubMed Abstracts |
| CD36 > Uqcr2  | Co-expression      | Co-mentioned in PubMed Abstracts |
| CD36 > Scab1  | Predicted, shared protein domain | - |
| CD36 > Merk   | Predicted          | Experimental/Biochemical Data: putative homologs were found interacting in other species; Co-mentioned in PubMed Abstracts |
| CD36 > CD9    | Predicted          | - |
| CD36 > Hand1  | Predicted          | - |
| CD36 > Ralg1  | Predicted          | - |
| CD36 > Map7k2 | Predicted          | - |
| CD36 > I2     | Predicted          | - |
| CD36 > Igf    | Predicted          | - |
| CD36 > Thbs1  | Predicted          | Experimental/Biochemical Data: putative homologs were found interacting in other species; Co-mentioned in PubMed Abstracts |
| CD36 > Scd4   | Predicted          | Co-mentioned in PubMed Abstracts Association in Curated Databases |
| CD36 > B3Glu1 | Predicted          | Co-mentioned in PubMed Abstracts Association in Curated Databases |
| CD36 > Pdghfa | Predicted and Co-expression | - |
| CD36 > TNNT2  | Co-expression      | - |
| CD36 > Sdc1   | Predicted          | Experimental/Biochemical Data: putative homologs were found interacting in other species; Co-mentioned in PubMed Abstracts Association in Curated Databases |
| CD36 > Sdc3   | Predicted          | - |
| TNNT2 > MYL4  | Co-expression      | Experimental/Biochemical Data: putative homologs were found interacting in other species; Co-mentioned in PubMed Abstracts Association in Curated Databases |
| UQCR1C > UQCR1H | Co-expression | - |
| CD9 > UQCR1H | Co-expression | - |
| PDHB > UQCR1H | Co-expression | - |
| UQCR1C > PDHB | Co-expression | - |
| HAND1 > TNNT2 | Co-expression | - |
| PDGFR > THBS1 | Co-expression | - |
| LPL > THBS1 | Co-expression | - |
| B3Glu1 > PDGFR | Co-expression | - |
| LPL > MYL4 | Co-expression | - |
| SDC4 > MERTK | Co-expression | - |
| SDC4 > PDGFR | Co-expression | - |
| SDC4 > IFNG | Co-expression | - |
| HAND1 > MYL4 | Co-expression | - |
| HAND1 > PDGFR | Co-expression | - |
| HAND1 > THBS1 | Co-expression | - |
| HAND1 > HS1 | Co-expression | - |
| HAND1 > LPL | Genetic interactions | - |
| UQCR1C > DGAT1 | Genetic interactions | - |
| HAND1 > ADH4 | Genetic interactions | - |
| HAND1 > IL2 | Genetic interactions | - |
| LPL > FA14 | Genetic interactions | - |
| PDGFR > FA14 | Genetic interactions | - |
| LPL > TNNT2 | Genetic interactions | - |
| MAPK2 > HRE | Genetic interactions | - |
| LPL > MYL4 | Genetic interactions | - |
| TNNT2 > MYL4 | Genetic interactions | - |
| UQCR1C > rAGL1 | Genetic interactions | - |
| IFNG > THBS1 | Genetic interactions | - |
| SDC4 > API5 | Genetic interactions | - |
| UQCR1C > API5 | Genetic interactions | - |
| Protein 1 | Protein 2 | Interaction Type |
|----------|----------|-----------------|
| LPL       | API5     | Genetic interactions |
| MERTK    | RG1      | Genetic interactions |
| IFNG     | UQCR1    | Genetic interactions |
| MYL4     | FAT4     | Genetic interactions |
| UQCRH    | API5     | Genetic interactions |
| IFNG     | UQCR1    | Genetic interactions |
| IRS1     | API5     | Genetic interactions |
| PDGFRA   | MAP3K2   | Genetic interactions |
| MAP3K2   | FAT4     | Genetic interactions |
| MERTK    | PDHLM3   | Genetic interactions |
| CD9      | MERTK    | Genetic interactions |
| PDGFRA   | IDE      | Genetic interactions |
| FAT4     | API5     | Genetic interactions |
| IRS1     | FAT4     | Genetic interactions |
| PDGFRA   | PDHB     | Genetic interactions |
| PDHB     | FAT4     | Genetic interactions |
| SDC4     | THBS1    | Pathway          |
| TNNT2    | MYL4     | Pathway          |
| IFNG     | PDGFRA   | Pathway          |
| IFNG     | IL2      | Pathway          |
| IRS1     | IL2      | Pathway          |
| UQCR1    | IDE      | Shared protein domains |
| PDGFRA   | MERTK    | Shared protein domains |
| IFNG     | IL2      | Shared protein domains |
| THBS1    | FAT4     | Shared protein domains |

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License.