Identification of the Expression Patterns of Fatty Acid Oxidation Genes in a Heterogenic Cardiomyopathy Patient Cohort

The potential of PPARA-modulating compounds as a therapeutic intervention for cardiomyopathies

Bioinformatics profile internship report

Alyssa van den Brink

MSc student Regenerative Medicine & Technology

January 2023

regenerative medicine utrecht

UMC Utrecht

Universiteit Utrecht

Alyssa van den Brink, BSc
6148034, a.vandenbrink4@students.uu.nl
Regenerative Medicine & Technology, Utrecht University

Harakalova/van Steenbeek
Department of Cardiology, division Heart & Lungs, UMC Utrecht
Supervisor: Magdalena Harakalova, MD, PhD
Examinor: Frank van Steenbeek, PhD

Abstract

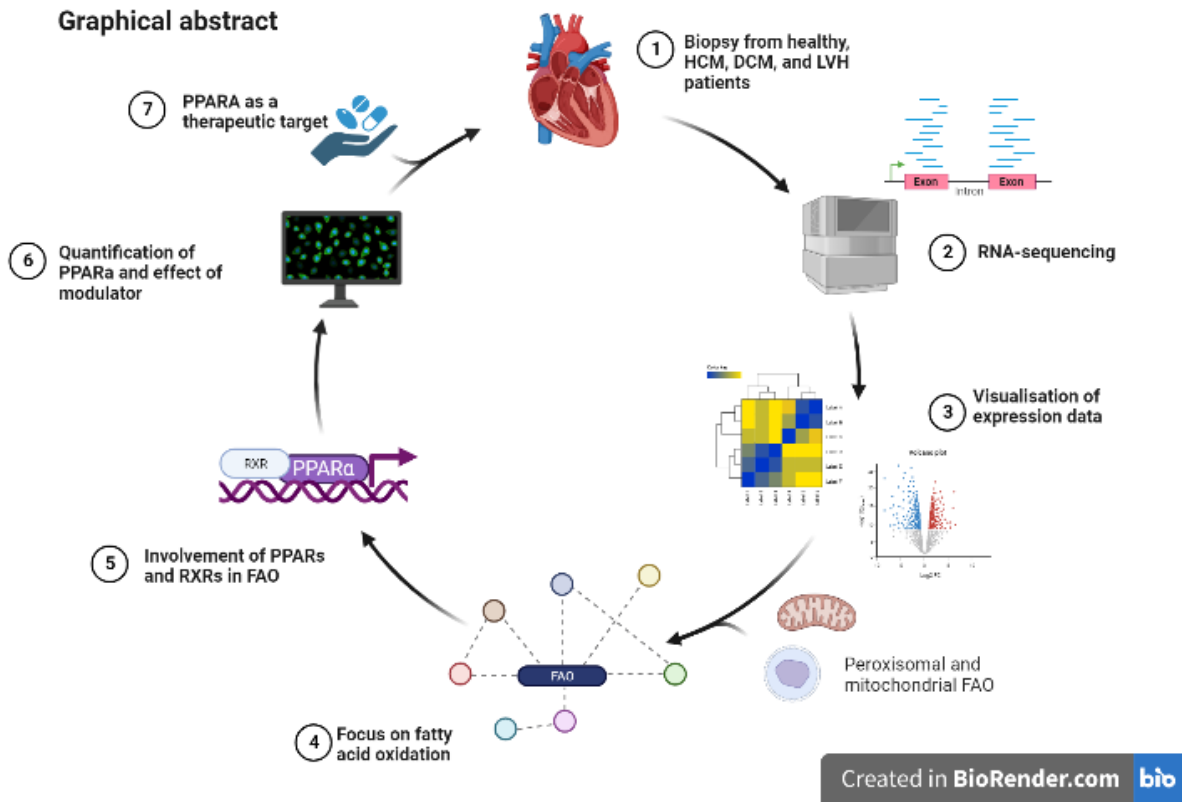
Cardiomyopathy is one of the major causes of heart failure that affects approximately 26 million patients worldwide. In a healthy situation, the heart depends on fatty acid oxidation (FAO) to maintain its energy demand. In failing hearts, the heart reverts to a fetal-like metabolic state in which FAO is downregulated and the heart switches to glucose metabolism. Peroxisome proliferator-activated receptor alpha (PPARA), a regulator of FAO, has recently been found to be hypoacetylated and its downstream effectors downregulated in dilated cardiomyopathy. However, the exact mechanisms remain largely unknown. The main purpose of this internship was to identify the specific FAO expression patterns in a large heterogenic patient cohort and investigate the involvement of PPARA regulation. The secondary goal was to set up a transcriptomic pipeline within Galaxy, a user-friendly bioinformatics platform, and test whether this platform can be used in diagnostics in the future. Differential expression analysis and gene enrichment of RNA-sequencing data show a type-specific FAO expression pattern of DCM and LVH versus HCM and confirm a downregulation of FAO related processes. Public RNA-sequencing data after the use of a bezafibrate, a PPARA agonist, which has been integrated, shows the upregulation of genes that are downregulated in our data. Combined, the results confirm the downregulation of FAO, but indicate a disease-type-specific FAO expression pattern. The results from the Galaxy pipeline show that this pipeline is suitable for transcriptomic analyses. Furthermore, we highlight the potential therapeutic role of bezafibrate in cardiomyopathy.

Keywords: cardiomyopathy, transcriptomics, fatty acid beta-oxidation, metabolism, peroxisome-proliferator activator-alpha

Layman's summary

Heart failure is a condition that affects approximately 26 million patients worldwide. Patients experience extreme tiredness or shortness of breath, because the heart's function to pump blood is weakened. Within five years, approximately 50% of the patients die of heart failure. Cardiomyopathy is a disease that can cause heart failure, which is why it is important to prevent and treat cardiomyopathy. Cardiomyopathy can arise from a mutation (variant) or it is a consequence of another disease. There are various forms of cardiomyopathy; the heart muscle can be thickened or thinned affecting the heart's pump function, the heart can beat arrhythmic, or the heart is not properly filled with blood. It is seen in cardiomyopathy that the heart makes less use of fat metabolism to generate energy. Recently, a major regulator of fat metabolism, peroxisome proliferator-activator α (*PPARA*) was seen to be less expressed and downstream players were affected. In this report, all analyses were performed on a user-friendly platform called Galaxy, with the intention of finding out whether this platform can be used for similar projects in the hospital. Furthermore, it is highlighted how genes involved in fat metabolism behave across different forms of cardiomyopathy and whether PPARA can be used as a target for intervention. This was done by comparing patient data with healthy control data and filtering out the genes of interest. The results show that fat metabolism genes behave differently depending on the form of cardiomyopathy, and confirm that fat metabolism is downregulated. In combining public data that used bezafibrate, a drug that alters PPARA, we see that it has potential as a drug for cardiomyopathy.

Graphical abstract



Transcriptomics workflow of this project: *1) Biopsies were taken from controls, HCM-, DCM-, and LVH patients. 2) RNA was sequenced by Illumina. The RNA-seq data was processed and visualized, with differential expression analysis, heatmaps and volcano plots. 4) Then a more targeted analysis was done with a focus on mitochondrial and peroxisomal fatty acid oxidation, for which a STRING gene interaction figure was used. 5) Afterwards, the involvement of PPARA and RXRs was investigated by literature research, data integration, motif enrichment and imaging. 6) Immunofluorescent stainings were made for PPARA, which were quantified using CellProfiler. 7) The end goal of this analysis was to further research whether PPARA is a potential therapeutic target, that then can be used in various forms of cardiomyopathy. (BioRender)*

Table of content

List of abbreviations

Abbreviation	Explanation
ACM	arrhythmogenic cardiomyopathy
DBD	DNA-binding domain
DCM	dilated cardiomyopathy
DEG	differentially expressed gene
DSP	desmoplakin
EFA	essential fatty acids
EU	European Union
FAO	fatty acid β-oxidation
FGSEA	Fast Gene Set Enrichment Analysis
GDPR	General Data Protection Regulation
GO	Gene Ontology
GOseq	Gene Ontology analysis for RNA-seq
HADHA	hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit alpha
HADHB	hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit beta
HCM	hypertrophic cardiomyopathy
LBD	ligand-binding domain
LV	left ventricle
LVH	left ventricular hypertrophy
MD	multiple dimension
MYBPC3	cardiac myosin binding protein C
MYH7	myosin heavy chain 7
MYL2	myosin Light Chain 2
NES	normalized enrichment score
PCA	principle component analysis
PLN	phospholamban
PPAR	peroxisome proliferator-activated receptor
PPRE	peroxisome proliferator response element
QC	quality control
RNA-seq	RNA-sequencing
ROS	reactive oxygen species
RV	right ventricle
RXR	retinoid X receptor
RZ	remote zone to the left ventricle
SMN	sarcomere mutation-negative
TazKD	tafazzin knock-down
TF	transcription factor
TFBM	transcription factor binding motif
TNNI3	troponin I3
TNNT2	troponin T2
TSS	transcription start site
TTN	titin
WT	wild-type

Introduction

Heart failure is a condition that affects approximately 26 million patients worldwide. With a mortality rate of 50% after a 5-year follow-up, it is essential to research ways to prevent and treat heart failure. One of the underlying causes of heart failure is cardiomyopathy [1]. Cardiomyopathies can be classified into primary or secondary categories. The primary category includes genetic, acquired, and mixed cardiomyopathies, and the secondary category holds cardiomyopathy due to a systemic condition. The most common primary genetic and mixed cardiomyopathies are arrhythmogenic right ventricular cardiomyopathy (ARVC), hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), and restrictive cardiomyopathy (RCM), respectively [2]. HCM is the most common cardiomyopathy, with an approximated prevalence of 1:500 adults. In most cases, HCM is inherited, though there is a broad clinical heterogeneity with nongenetic phenotypes. HCM is characterized by left ventricular hypertrophy (LVH), diastolic dysfunction, and abnormal vascular response [3–5]. In a third of all cases of heart failure, the underlying cause is DCM [6]. DCM is defined by systolic dysfunction with an enlargement or dilation of the left ventricle and can have ischemic or non-ischemic (often genetic) aetiology [7]. The causative variant is familial in about 25-30% of non-ischemic DCM patients [6,8]. The patient cohort in this report includes 9 genetic forms for HCM and DCM and two nongenetic forms. The genetic variants (mutations) are located in known cardiomyopathy genes, such as myosin binding protein C3 (*MYBPC3*), myosin heavy chain 7 (*MYH7*), myosin light chain 2 (*MYL2*), troponin I3 (*TNNI3*), troponin T2 (*TNNT2*), titin (*TTN*), desmoplakin (*DSP*), Ischemic, and phospholamban (*PLN*). Nongenetic forms include sarcomere mutation-negative (SMN) and LVH. This form of LVH is distinct from HCM in that the underlying cause for LVH is hypertension, and LVH can be a characteristic of HCM [4,9].

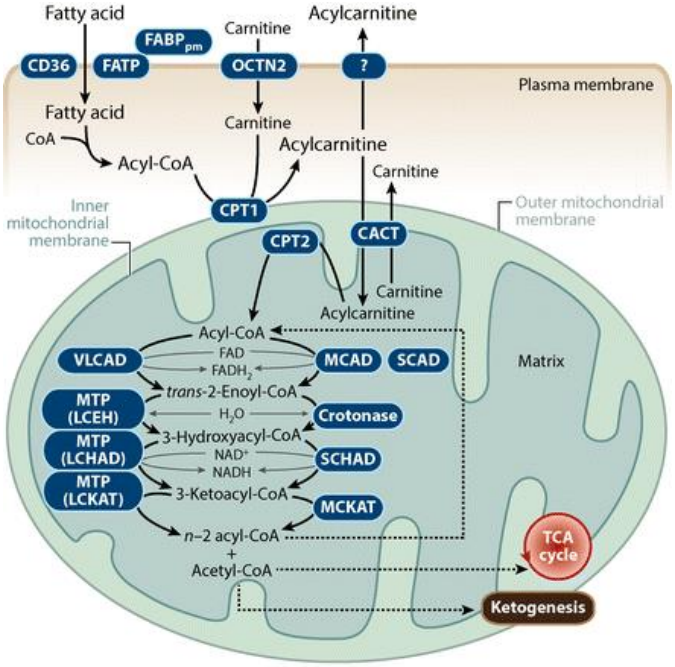


Figure 1 Mitochondrial FAO*: The cell membrane and mitochondrion are visualised. Fatty acids are transported into the cell via FAT/CD36 and FABP. FACS ads a CoA group (not in the figure) and CPT1 adds a carnitine to the fatty acids in order to facilitate transport into the mitochondrium. The carnitine is removed by CPT2. What follows isβ-oxidation in the mitochondrial matrix, catalysed by multiple proteins (ACADM, ACADS, ACADVL, MTP, and more. The end products are further metabolised in other cycles and processes [13]. ACADM = MCAD, ACADS = SCAD, VLCAD = ACADVL.*

Fatty acid β-oxidation (FAO) is the preferred metabolic pathway of the adult heart, supplying the majority of produced adenosine triphosphate (ATP). Yet, the heart contains great substrate flexibility and can switch to metabolising glucose, lactate, and ketone bodies [10,11]. The process of FAO consists of three steps: the uptake into the cytosol, the transport across the mitochondrial membrane, and the oxidation inside the mitochondria [12]. There are approximately 20 enzymes/proteins involved in this process (Figure 1) [13]. Fatty acids are transported into the cell via fatty acid transporters (*FAT/CD36*) and fatty acid binding proteins (*FABPs*). A coenzyme A (CoA) group and carnitine are added to further transport into the mitochondria by fatty acyl CoA synthetase (*FACS*) and carnitine palmitoyl transferase 1 (*CPT1*), respectively. The acylcarnitine is transported over the inner membrane of the mitochondria via carnitine translocase.

As acylcarnitine is converted back into long-chain acyl-CoA by carnitine palmitoyl transferase 2 (*CPT2*), it enters β-oxidation in the mitochondrial matrix [12,14]. Enzymes involved in this step are acyl-CoA dehydrogenases (*ACADM, ACADS, ACADVL*) and the mitochondrial trifunctional proteins (*MTPs*), namely hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit alpha (*HADHA*) and hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit beta (*HADHB*) [15]. FAO results in NADH and FADH2 that can be further used to produce ATP and supply the cell with energy [12,14]. FAO genes are under tight transcriptional control; up- or downregulation of FAO genes often translates to an up- or downregulated FAO. Peroxisome proliferator-activated receptors (PPARs) are key players in metabolic homeostasis and function [11].

PPARs are nuclear, ligand-activated transcription factors (TFs) that belong to a nuclear hormone receptor superfamily, consisting of three isoforms: PPARα, PPARγ, and PPARβ/δ, encoded by their respective genes PPARA, PPARG, and PPARD, respectively. PPARs are ubiquitously expressed, and these tissues include the liver, intestine, skeletal muscle, adipose tissue, vascular wall, and heart. It is considered that each PPAR has its own distinct function in metabolism; PPARα is a regulator of energy homeostasis through fatty acid transport, FAO, and ketogenesis. PPARγ regulates adipogenesis, such as lipid storage, insulin sensitivity, and glucose metabolism, and PPARδ increases lipid and glucose metabolism and switches muscle fibres from glycolytic to oxidative [10,16].

Once PPARs bind their ligand and become activated, they heterodimerise with another nuclear receptor, retinoid X receptor (RXR/NR2B) [17]. As a complex, they alter the transcription of target genes by binding to peroxisome proliferator response elements (PPREs), which consist of a repetition of AGG(A/T)CA interspaced by one or two nucleotides [10,17]. PPAR consists of six domains; The A/B domain contains the AF-1 region and has (in)dependent ligand binding activity, which is influenced by phosphorylation from MAPK [18] (Figure 2A). The A/B domain plays a role in determining the target gene specificity between the PPAR isoforms. However, a complete understanding of how the distinction in gene expression between the three isoforms is made remains unclear [19].

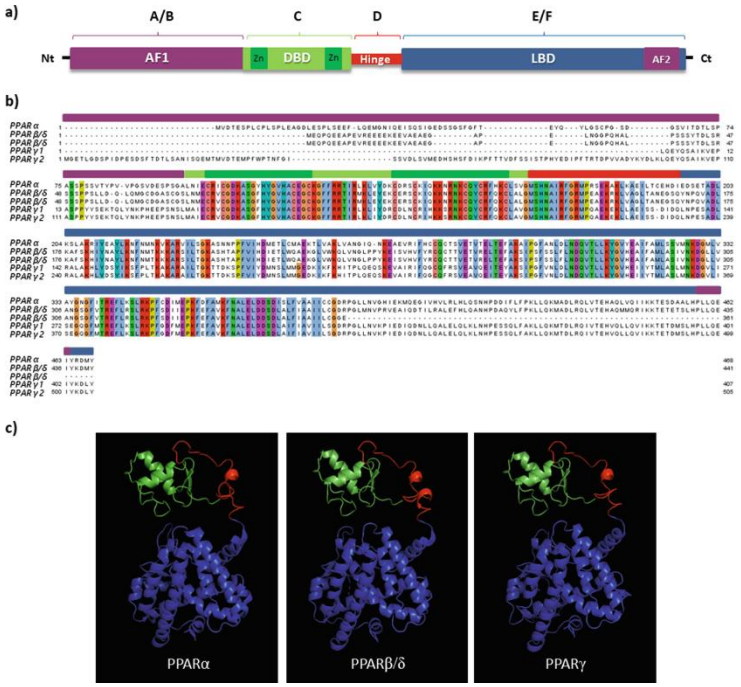


***Figure 2: PPARs sequence and structure homology**. A) The six domains of PPAR. Purple displays the AF-1 and AF-2 regions, green the DBD region, red is the hinge region, and blue is the LBD region. B) The sequences of the PPAR isoforms with their sequence homology coloured. Bars above the sequences correlate to the colours of the different regions. C) Structure of the three isoforms, PPARA, PPARD, and PPARG. Green correlates to the DBD region, blue to the LBD region and red to the hinge region [20].*

The DNA binding domain (DBD)/domain C has two zinc fingers that bind to the PPREs, which is linked to the ligand binding domain by the hinge region. The hinge region/domain D is the binding spot for corepressors and can be phosphorylated. AF-2/domain E/F determines the ligand specificity and undergoes conformational changes upon binding of the ligand, regulating interactions with co-

activators. The E/F domain is also called the ligand-binding domain (LBD) [18]. The domains and structure reflect the homology of the isoforms, as the sequence for the DBD is most conserved between the three PPARs. The LBD has significant sequence variation, which explains why the isoforms have ligand selectivity (Figure 2B and 2C)[20].

Ligands for PPARs can be natural or synthetic. Natural ligands include a wide range of fatty acids, like essential fatty acids (EFAs), mono- or poly(un)saturated fatty acids, oxylipins and prostaglandins [21]. Synthetic ligands are mainly fibrates for PPARα, and thiazolidinediones for PPARγ. Thiazolidinediones have derivatives, glitazones (PPARγ) and glitazars (dual PPARα/PPARγ). Besides these three classes of synthetic ligands, many other compounds are being tested [21–23]. These compounds include specific PPAR modulators but also dual agonists and pan-agonists (agonists for all three isotypes). Depending on what ligands bind determines what cofactor interactions are possible and regulatory response follows [10]. As for the coregulators, these are a broad class of proteins that can be divided into subgroups based on whether they are essential, where they bind on PPAR and function. Repressors prevent PPAR from binding to the PPRE, and activators can prevent PPARs from degradation in the cytosol before they can be translocated to the nucleus. In addition, they assist in binding to PPRE and associating with other proteins like histone methylases, histone acetyltransferases, and DNA-helicases. Activators also function in transcriptional regulation [24].

As stated before, PPARs are regulators of FAO in cardiomyocytes, which is the main mechanism for the heart's energy demand. During rest or exercise, PPARs switch the metabolic state of the heart to either predominantly use fatty acids or lactate. In a failing heart, as seen in cardiomyopathy, the heart returns to a fetal-like metabolic state. This state is characterised by a high expression of glycolytic genes and low expression of FAO and mitochondrial genes due to an increase in hypoxia-inducible factor (*HIF*) and a decrease in PPARα-PPARγ-coactivator 1α (*PGC1α*) activity, respectively. Though the direct role of PPARα in this FAO-glycolysis switch is not yet proven, it is seen that PPARα-deficient mice have a lower expression of genes involved in mitochondrial and peroxisomal FAO [10,25]. Peroxisomal FAO differs from mitochondrial FAO as peroxisomal FAO only processes very long-chain FAs and does not contribute to ATP synthesis; instead, it regulates cellular thermogenesis [24]. Of all the hundreds of genes that PPARα regulates, PPARα controls eight proteins typically localized in peroxisomes (*ACOX1, EHHADH, ACAA1B, SCP2, CAT, MLYCD, PEX11A*) [24]. Mitochondrial genes that are less expressed in patients with cardiomyopathy are, for example, very long chain acyl-CoA dehydrogenase (*VLCAD/ACADVL*), carnitine acylcarnitine translocase (*CACT/SLC25A20*), organic cation transporter (*OCTN2/SLC22A5*), *HADHA/HADHB*, and *CPT2* [13].

In (yet) unpublished research conducted by members of the group I am doing my internship with, chromatin immunoprecipitation and sequencing (ChIP-seq) and RNA-sequencing (RNA-seq) data of *PLN*-R14del and control hearts was analysed. In the detected hypoacetylated regions, transcription factor binding motifs (TFBMs) were found that related to TFs involved in metabolism, adipogenesis, and mitochondrial structure. One of these TFBMs belonged to PPARA, which, after immunofluorescent staining, was found to be less localized within intercalated disks and the nucleus compared to healthy hearts. Downstream targets of PPARA, such as HADHA/HADHB/MLYCD/PNPLA2, showed hypoacetylation and decreased mRNA levels [26].

It has been established that FAO is downregulated in cardiomyopathy and heart failure [27]. However, a direct link with PPARα, or how exactly FAO is downregulated in cardiomyopathy, has never been made. The research and this gap in knowledge about how exactly FAO is downregulated in

cardiomyopathy resulted in the main research question for this project: What are the expression patterns of FAO genes during various forms of heart failure? And: What proteins that cause the downregulation of FAO are potentially suitable targets for PPARA-targeted treatment?

To answer these questions, the project was set up in two objectives relating to the two research questions. For the first objective, which was the original research question of the project, patient-specific RNA-seq data from a large heterogenic cohort was analysed and visualised for differentially expressed genes (DEGs), focusing on FAO genes and PPARA. Gene enrichment was done to zoom in on the regulation of different FAO-related metabolic processes. For the second objective I decided to take a more exploratory approach. I collected a list of PPARs-modulating compounds and a list of articles that used PPARA modulators in different settings. This data and other information I gathered regarding PPARA as a therapeutic target was used to publish a mini-review [28]. Subsequently, a publicly available dataset from one of these articles was used to compare it to the UNRAVEL transcriptomic study. Afterward, motif enrichment with the MoLoTool was tested to check for PPARA (specific) regulation. At last, I helped with staining and imaging on the confocal microscope. A small pipeline was made on CellProfiler to quantify the PPARA signal within the nucleus in relation to the size of the nucleus.

Materials and methods

Sample information

RNA was isolated from biobanked frozen human cardiac tissue from different centres, using two different protocols. Consent was given by all individuals. The libraries were prepared in collaboriation with the Epigenomics facility at UMCU. Sequencing was done by USEQ. The library consists of single-end RNA-seq data from Illumina. The total sequencing library consists of 118 samples, of which 28 samples are technical replicates. The samples originate from 94 patients (59 males and 35 females), of which 102 biopsies were taken either from the septum (*n* = 69), left ventricle (*n* = 23), remote zone to the left ventricle (*n* = 4), or right ventricle (*n* = 6). Patients came from various locations, London (*n* = 4), Rotterdam (*n* = 50), St. Louis (*n* = 9), Sydney (*n* = 6), and Utrecht (*n* = 27). From the patient group, 24 patients are considered controls, 18 patients are diagnosed with DCM, 50 patients are diagnosed with HCM, and 4 patients are diagnosed with LVH (Figure 3, Supplementary Table Library overview). Included forms are ischemic DCM, MYBPC3, MYH7, PLN, TTN, DSP, MYL2, SMN HCM, TNNI3, TNNT2, TTN, TMEM43, and LVH. Two samples are labelled Uncertain, due to uncertainty about the causal variant; TTN or TMEM43, and MYH7 or MYBPC3. Some of the libraries have already been analysed and used in (un)published papers [26,29–38].

One technical replicate was removed before analysis due to a faulty file containing only 21,733 genes instead of 63,678 genes. The final library then consisted of 117 samples, 27 technical replicates, and 94 patients (16 biological replicates) (Supplementary Table Library overview).

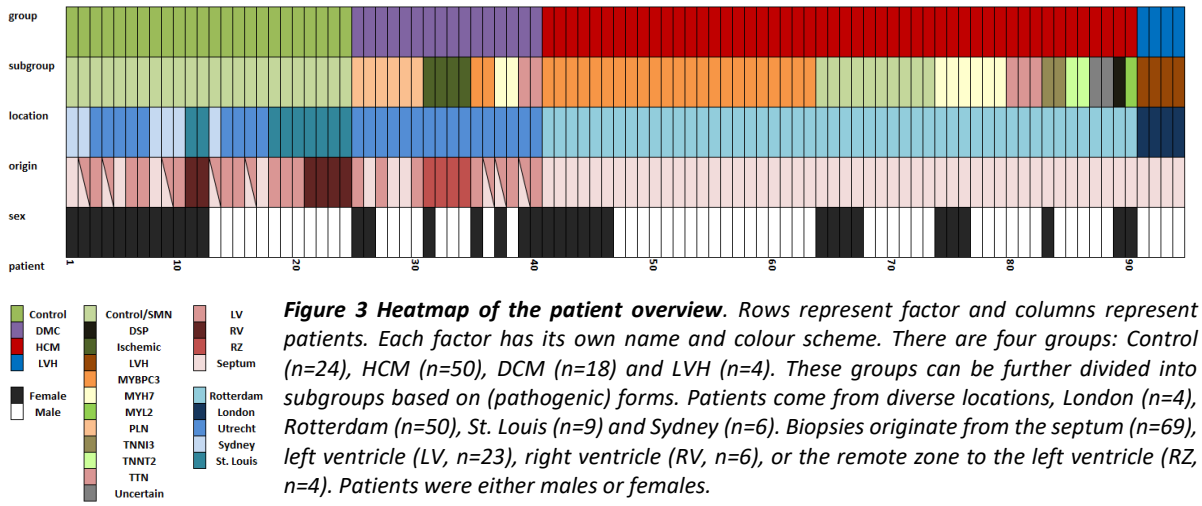


Figure 3 Heatmap of the patient overview. *Rows represent factor and columns represent patients. Each factor has its own name and colour scheme. There are four groups: Control (n=24), HCM (n=50), DCM (n=18) and LVH (n=4). These groups can be further divided into subgroups based on (pathogenic) forms. Patients come from diverse locations, London (n=4), Rotterdam (n=50), St. Louis (n=9) and Sydney (n=6). Biopsies originate from the septum (n=69), left ventricle (LV, n=23), right ventricle (RV, n=6), or the remote zone to the left ventricle (RZ, n=4). Patients were either males or females.*

Galaxy

For the main analyses of this project, the usegalaxy.eu portal has been used [39]. All data is stored under the username avdbrink. The histories are private and only accessible upon request. An overview of the histories can be found in Supplementary File FAOgeneslist: Table 1. For the descriptions of the analyses done in Galaxy, default settings were used unless stated otherwise.

Data formatting

Galaxy requires a certain format for differential expression analysis. All count files need to contain a header, and all summary lines at the end should be removed. All files need to have equal rows. It is recommended to use *Select last* (Operation = Keep last lines, Number of lines = 5) and *Select first* (Select first = 3, a dataset has a header = true) to check whether the files have the correct formatting.

A script in Python was written to automatically add a header line to all 117 files (Supplementary File Rnotebook: Script 1). Summary lines can be removed in Galaxy by using *Select* (Select lines from = all data files, that = NOT matching, the pattern = ^_, Keep header line = True). The workflow can be found in the Galaxy history: DESeq data. After formatting the datasets in Python and Galaxy, the correct Galaxy input files were copied into new histories for analysis.

Replicates

Three groups were created to perform analyses on (all samples, all biopsies, and all patients). In the biopsies group, all technical replicates (same sample, multiple sequencing runs) were merged together by averaging the counts. In the patients groups, additionally, all the biological replicates (same patient, different biopsy) were merged by averaging the counts. This was performed in Galaxy using *Column Join* (input = technical replicates/patient duplicates, identifier column = 1, number of header lines in each input file = 1, add column name to header = no) and *Compute* (input = output of Column Join, Input has a header line with column names = yes, Expressions – Add expression = int((c2+c3)/2), Mode of operation = append). In case of a class error, i.e., an error in converting strings to integers, the output of *Column Join* was downloaded and the averaged counts were calculated in excel (=INTEGER(SOM(A1:A2)). Workflow can be found in the Galaxy history: Sum/average replicates.

Outlier removal

Outlier removal and verification were done in four steps. First, principal component analysis (PCA) plots were generated using DESeq2 on all samples, all biopsies, and all patients. The PCA plots show clustering based on the top 500 genes, which is the unchangeable default of DESeq2 in Galaxy. The three analyses were pre-filtered on the sum of all samples in that group (n = 117, n = 102, n = 94). Second, another method for visual outlier representation was used; dendrograms were made using hierarchical clustering with average linkage in R (*WGCNA).* This clustering was done based on 5,000 genes, selected from DESeq2 analysis on the comparison with the most samples, HCM-control, and selecting the top 5,000 genes. Third, the outliers selected by the PCA plots in DESeq2 were removed stepwise from limma analysis on all biopsies. The results were compared against the box plots and MDS plots without outlier removal. Fourth, a limma analysis was done on all biopsies, and the outliers based on the MDS plots were then removed from another limma analysis. All limma analyses were performed with the voom method without sample quality weights. Counts were TMM-normalised. Workflows can be found in the Galaxy histories: Analysis 1: DESeq on all samples, Analysis 2: DESeq on all biopsies, Analysis 3: DESeq on patients, and Analysis 4: Limma (biopsies, with/out outlier removal).

Dendrograms were created in R in combination with Galaxy. The DESeq2 result file of HCM-Control (of all samples, all biopsies, all patients) was used to select the first 5,000 genes with *Select first* (select first: 5,000, dataset has header: yes). This list was combined with the count matrix of all samples/all biopsies/all patients to create a count matrix containing the first 5,000 genes of DESeq2. The output made with the following tools was uploaded into R: *Compare two datasets* (compare: count matrix of all samples, using column: 1, against: Output of *Select first,* and column: 1, to find: matching rows of 1st dataset). Dendrograms were made with the *hclust* function of the WGCNA package (Supplementary File R notebook: Script 2) [40].

Quality control

All samples were tagged according to their factors in Galaxy. The different factors (i.e., origin, location, sex, and runs) were visually analysed by PCA plots and sample-to-sample distances plots generated by DESeq2. All factors were analysed on all samples, with a pre-filtering value of 117. Additionally, the factor origin was run on all biopsies, with a pre-filtering value of 102. The factor sex and location were run on all patients, with a pre-filtering value of 94. Workflows can be found in the Galaxy histories: Analysis 1: DESeq on all samples, Analysis 2: DESeq on all biopsies, and Analysis 3: DESeq on patients.

Sex difference

Limma analysis was done on the comparison Female-Male with a pre-filtering value of 1 CPM in a minimum of 33 samples. DEGs were run through the STRING database and further analysed with Glimma volcano plots. Limma was set to voom without sample quality weights. The adjusted P-value threshold was set at 0.05 and corrected with the Benjamini and Hochberg method. Counts were TMM-normalised. All additional output options were selected: Glimma interactive plots, density plots, CpmsVsCounts plots, box plots, MDS extra, MD plots for individual samples, heatmaps and strip charts. An annotation file was supplied to limma for the Glimma interactive plots. The annotation file was created by using *Cut* (input = a counts file, cut columns = c1, delimited by = tab), followed by *AnnotateMyIDs* (input = output of *Cut*, file has header = true, organism = human, ID Type = Ensembl Gene, output columns = SYMBOL, GENENAME). Workflow can be found in the Galaxy history: Analysis 4b: Limma patients factor: sex.

FAO genes list

The FAO genes list comprised the 76 genes associated with the GO term GO:0006635 and our own additions based on work from the group that highlighted the involvement of these genes in cardiomyopathy (KLF15, PPARs, and RXRs). The Ensembl IDs were annotated using BioMart with the filters for Gene type and transcript type set to protein_coding, and the attributes: transcript type, gene name, gene synonym, gene stable ID, and chromosome (Supplementary File FAOgeneslist: Table 2).

Differential gene expression

Differential expression analysis was performed with limma on all biopsies without the outliers (n = 100). Pre-filtering value was set on 1 CPM in a minimum of 4 samples. Limma was applied with the voom method without sample quality weights. Counts were TMM-normalised with robust settings. The adjusted P-value was corrected with Benjamini and Hochberg and was set on a threshold of 0.05. All additional output options were selected: Glimma interactive plots, density plots, CpmsVsCounts plots, box plots, MDS extra, MD plots for individual samples, heatmaps, and stripcharts. The limma annotation file was provided. Comparisons were set on HCM-Control, DCM-Control, and LVH-Control. Workflow can be found in Analysis 4: Limma (biopsies, with/out outlier removal).

Heatmaps

Heatmaps of the top 50 most differentially expressed genes in HCM, DCM, and LVH were made based on the limma DE tables that were filtered on an adjusted p-value (Benjamini and Hochberg) of 0.05 and sorted based on Log2(FC). For this *Filter* (Input = output of *Limma*, DE tables, with the following condition = c7 < 0.05, number of header lines to skip = 1) and *Sort in descending or ascending order* (Input = output of *Filter*, number of header lines = 1, on = column 4, in = Descending order, Flavor = Fast numeric sort (-n)). Ensembl IDs with a NA for Gene Symbol and Gene Name were filtered out by

Filter (Input = output of *Sort,* with the following condition = c2!= 'NA', number of header lines to skip = 1. The 25 most upregulated and 25 most downregulated genes were selected with *Select first (*Select first = 25, from = output from *Filter*, dataset has a header = True*)* and *Select last (*Text file = output from *Filter*, Operation = Keep last lines, Number of lines = 25*)*. These two lists were combined with *Concatenate datasets tail-to-head (*Concatenate dataset = output from *Select first,* Select = output from *Select last).* These lists were then matched with the normalized counts table using *Compare two datasets to find common or distinct rows* (Compare = output from *Limma*: Normalised counts, using column = 1, against = output from *Concatenate dataset*, and column = 1, to find = matching rows of 1st dataset), which yielded the normalized counts for the top 50 most differentially expressed genes. The normalized counts for the FAO genes list heatmap was made with *Compare* (Input = Limma on biopsies: Normalized counts, using column = 1, against = ensemblIDs of the FAO genes list, and column = 1, to find = Matching rows of the 1st dataset). These lists were downloaded and loaded into R. There, heatmaps were made with the *Coolmap* function of the *Limma* package (Supplementary File R notebook: Script 3 and Script 4). Workflows can be found in Analysis 4: Limma (biopsies, with/out outlier removal) and Analysis X: heatmap of the FAOgeneslist.

Gene sets

Gene sets were made for Gene Ontology analysis through RNA-seq (GOseq) and fast pre-ranked gene set enrichment analysis (FGSEA). Gene Ontology (GO) terms of interest were selected through the Gene Ontology website by looking at the sub-classification of the fatty acid metabolic process (GO:0006631), fatty acid oxidation (GO:0019395), and fatty acid beta-oxidation (GO:0006635). Additional GO terms related to glucose metabolism, amino acid metabolism, ketone metabolism, and fatty acid related were selected. All 23 selected GO terms were run through Ensembl BioMart GRCh38.p13, as the hg19 version of BioMart does not allow filtering on GO terms. Selected attributes were Gene stable ID and Gene name (Supplementary File FAOgeneslist: Table 3). In the GO terms fatty acid metabolic process and long-chain fatty acid metabolic process, there are four genes that do not have a gene name and, upon investigation, have no expression data in our count files. Therefore ENSG00000276490, ENSG00000281938, and ENSG00000284341 are excluded from the corresponding gene sets. ENSG00000258653 (a novel protein [41]) does appear in the count matrix, so it will be kept in. The GO terms glycolytic process and cellular amino acid metabolic process contain 10 genes without a gene name. ENSG00000282835, ENSG00000283189, ENSG00000286112, and ENSG00000284512 are excluded and ENSG00000266953, ENSG00000255835, ENSG00000111780, ENSG00000260643, ENSG00000255730, ENSG00000249319, and ENSG00000269547 are included based on the previous criteria.

The gene set of fatty acid beta-oxidation was adapted to match the FAO genes list, therefore, the PPARs, RXRs, and KLF15 were added to the gene set.

Additionally, five Hallmark gene sets and three PPARA-related datasets were acquired from the MSigDatabase (Supplementary File FAOgeneslist: Table 3). This resulted in a total list of four sets (fatty acid metabolic related gene sets, additional gene sets, hallmark gene sets, and PPARA related gene sets), which include 31 gene sets.

Fast pre-ranked gene set enrichment analysis (FGSEA)

Gene sets were Gene Matrix Transposed (GMT) formatted for FGSEA (Supplementary Table Gene enrichment input). The first column has the gene set name, the second column has the GO term or

link to MSigDB, and the following columns have Ensembl IDs. Each row represents a gene set. The second FGSEA input was created using *Cut* (input = limma-voom DE table for HCM-Control, DCM-Control, LVH-control; cut columns = c1, c6; delimited by = tab). This output holds the Ensembl IDs and the t-statistic of the original limma DE output table. Next *Sort data in ascending or descending order* was used (input = output of *Cut*, number of header lines = 1, on column = 1, in = descending order, flavor = fast numeric sort (-n)). FGSEA is performed with the three outputs of *Sort* for all three comparisons and the GMT formatted gene sets, with the following settings: *FGSEA (*Ranked genes = output from *Sort*, file has header = True, Gene sets = gene set files, minimum size of gene set = 1, maximal size of gene set = 500, number of permutations = 1000, outputs plots = true, plot top most significant pathways = 16/10, Output RData file = False). Results from the FGSEA analysis read in R and with the *ggplot2* package bubble plots were made (Supplementary File R notebook: Script 5).

GOseq

Limma outputs three DE tables for the three comparisons. *Compute* (Input = limma-voom DE table, Input has a header line with column names = yes, add expression = c8 < 0.05), mode of operation = append, the new column name = status) is used to add a Boolean column stating which genes are significantly differentially expressed. Then *Cut* (input = output of Compute, cut columns = c1, c10, delimited by = tab) is used to select the Ensembl ID and Status columns. The assembled gene sets are formatted according to the input requirements; Ensembl IDs in the first column and associated GO term in the second column (Supplementary Table Gene enrichment input).

Gene set enrichment with GOseq needs a gene lengths file to correct for length bias. The reference genome in Gene Transfer Format (GTF) of hg19 was retrieved from: http://ftp.ensembl.org/pub/grch37/current/gtf/homo_sapiens/?C=S;O=D. The gene length file was generated with the tool *Gene length & GC content* (Select a built-in GTF file or one from your history = history, Select a GTF file = Homo_sapiens.GRCh37.87.gtf, analysis to perform = length). Analysis was performed with *GOseq* (Differentially expressed file = output from *Cut*, Gene lengths file = output from *Gene length & GC content,* Gene categories = history, Gene category file = gene sets file in tabular format, Use Wallenius method = True, Use hypergeometric method = False, Sampling number = 0, Select a method for multiple hypothesis testing correction = Benjamini-Hochberg [FDR] (1995), Count genes without any category = False, Output Top GO terms plot = True, Produce diagnostic plots = True, Extract the DE genes for the categories (GO/KEGG terms?) = True, Output RData file = False). The Wallenius method is a method of approximation that states that all genes within the same category have the chance of being chosen but that this chance is different from choosing genes outside this category.

This tool only outputs the top 10 categories in the Top GO terms plots. Therefore, the fatty acid metabolism-related gene sets are split in two so that all terms will be visualised in a plot.

PPAR-modulating compounds

A semi-systematic search was conducted on PubMed to look for PPAR-modulating compounds. This search included natural or synthetic ligands, single/dual/pan-agonists or antagonists of PPARα, PPARδ, and PPARγ. Information was noted about binding affinity for other proteins, clinical trials, publishing date, and the PMID.

PPARα-modulating compounds in pre-existing datasets

A semi-systematic search on PubMed was executed to collect research papers that have publicly available transcriptomics dataset of PPARA-modulating compounds. The search included PPARA-specific agonists, dual-agonists and pan-agonists (Wy-14,643, pemafibrate, ciprofibrate, fenofibrate, saroglitazar, clofibrate, chiglitazar/Bilessglu, lobeglitazone, GW409544, lanifibranor, gemfibrozil, bezafibrate/GW7647, BMS631707, KRP101, AVE8134, biphenyl derivative, elafibranor/GFT505/DY121, palmitoylethanolamide, LY518674, ZYH7, K111, macuneos, N-oleoylethanolamine, astaxanthin). Three antagonists (IS001, GW6471, AA452) yielded no results. With the compounds the following search terms were used: RNA-seq OR microarray OR transcriptomics OR transcriptome AND PPAR. Papers that used the compound in combination with another non-PPAR-modulating compound were excluded.

Integration of public dataset

From the PPARA-modulating compounds in pre-existing datasets one article was selected that had RNA-seq data on heart tissue. Supplementary files (Supplementary Table 2, 3, and 4) from Schafer, C. *et al.* containing the log(FC) and P-value of the RNA-seq data of tafazzin knockdown (TazKD) mice vs wild-type (WT) and bezafibrate treated vs untreated TazKD mice were downloaded [42]. From all three datasets, and the DE tables from limma for HCM, DCM, and LVH the FAO genes were selected with *Compare* (input = DE tables, using column = 1, against = FAO genes list, and column = 1, to find = matching rows of 1st dataset) in Galaxy. For the public data the gene symbols of the FAO genes list were used, and for the DE tables from limma the Ensemble IDs.

Motif enrichment

First, the transcription start site (TSS) was retrieved from the Ensemble website (assembly GRCh37/hg19). The first transcript ID per gene was chosen, and the first or last coordinate of the transcript region was chosen as TSS, depending on whether the genes were on the forward or reverse strand. For HADHA/HADHB a TSS was selected that was exactly in between the end coordinate of HADHA and the beginning coordinate of HADHB. To these TSS site 2,000 bp was added up- and downstream for the promotor region (Supplementary Table Motif Enrichment: Table 1A). The following promotor regions were made into BED format and uploaded onto Galaxy (Supplementary Table Motif Enrichment: Table 1B). With the tool *bedtools getfasta* (input = TSS sites in BED format, choose the source for the FASTA file = server indexed files, fasta_id = human (homo sapiens): hg19) the corresponding sequences were retrieved. Motif enrichment was executed on the HOCOMOCO website, with the sequence motif location (MoLo) tool. The motifs for PPARA, PPARG, RXRA, RXRB, RXRG, and KLF15 were used (Supplementary Table Motif Enrichment: Table 2).

CellProfiler

Nuclear size and PPARA quantification in the nucleus were analysed and measured using CellProfiler (4.2.5). All images were uploaded into CellProfiler and the metadata was automatically set-up to retrieve the case and channel data. First the nuclei were identified with *IdentifyPrimaryObjects* from the DAPI channel. Then, the nuclei were masked out in the PPARA channel with *MaskImage*. PPARA was selected using *IdentifyPrimaryObjects* in the DAPImask image, creating an object of PPARA signal only in the region of the nucleus. To quantity the PPARA signal in the nuclei, the nuclei and PPARA signal were related to each other with *RelateObjects* using the nuclei and PPARA objects. With *MeasureObjectSizeShape* the nuclei object and the related nuclei-PPARA object were measured. All

data was exported with *ExportToSpreadsheet*. The specific settings and pipeline can be found in Supplementary File CellProfiler.

All figures were created in Galaxy (version …), Excel (Microsoft 365) and RStudio (R version 4.1.2), and potentially adjusted with Inkscape (version 1.2).

Results

Pre-filtering increases the amount of differentially expressed genes

Both DESeq2 and limma offer the option to pre-filter the data. When pre-filtering all the samples in DESeq2 on a row sum of 117, and limma on 1 CPM in a minimum of 4 samples. The total number of genes goes down after pre-filtering in both DESeq2 and limma. The number of DEGs increases as well in DCM-Control and LVH-Control in limma by around 1,000-2,000 DEGs. Even though there is a relative increase in DEGs in DESeq2, the number of DEGs stays almost the same in DESeq2 after pre-filtering. HCM-Control in limma decreases in both the total number of genes and DEGs after pre-filtering. Thus, pre-filtering shows that the number of DEGs increases compared to no pre-filtering, with the exception of the HCM-Control comparison in limma (Supplementary Figure 1).

PCA plots and dendrograms of samples do not match

In the library of 117 samples, 27 are technical replicates and 16 are biological replicates. DESeq2 was used to generate PCA plots of all samples, all biopsies, and all patients to look at the clusters without technical and/or biological replicates. The groups were created by averaging the technical and biological replicates together.

Outlier identification was performed through visual identification with the output of DESeq2 (Galaxy) and dendrograms based on hierarchical clustering and average linkage (R) (Supplementary Figure 2). DESeq2 uses the top 500 most variable genes to generate the PCA plot. For the hierarchical clustering, the top 5,000 most differentially expressed genes were taken based on the HCM-Control comparison in DESeq2. The sample do5126IVS does not cluster according to its groups in “all samples” and “all biopsies” (Supplementary Figure 2A), and another sample from the same patient, do5126LV, does not cluster in “all biopsies” (Supplementary Figure 2B). Sample PLN4S lies apart in “all patients” (Supplementary Figure 2C). However, these samples do not show as outliers in the dendrograms.

Removal of DESeq2 outliers does not affect DEGs in limma

Limma is the analysis tool that will be used for eventual differential expression analysis, therefore the stepwise removal of the outliers from DESeq2 (do5126IVS, do5126LV, and PLN4S) is checked. All samples, except for two HCM samples, are normalized and the removal of the do5126 samples and PLN4S does not show a change (Supplementary Figure 3A). Limma outputs MD plots in four dimensions, which show 62HCM and 166HCM as outliers based on all samples (Supplementary Figure 3B). After the do5126 samples are removed (Supplementary Figure 3C) or additionally, PLN4S is removed (Supplementary Figure 3D), the clustering does not change, and 62HCM and 166HCM remain outliers. With the stepwise removal of the DESeq2 outliers, the amount of DEGs on the different disease comparisons decreases by an average of 8.8%, and 11.7% (Supplementary Figure 3E).

Removal of limma outliers affects DEGs

The removal of the limma outliers, 62HCM and 166HCM, caused all the samples to be normalized (Supplementary Figure 4A). The MD plots in the four dimensions show clear clusters when 62HCM and 166HCM are removed. Only in the second and third dimensions is there one sample that does not cluster well, 6056IVS (Supplementary Figure 4B and 4C). The amount of DEGs increases by an average of 27.1% once 62HCM and 166HCM are removed from the analysis (Supplementary Figure 4D). This led to the limma outliers being excluded.

Quality control

Based on all samples, PCA plots of all different factors, location, origin, sex, and runs, were made with DESeq2 (Supplementary Figure 5). The PCA plot for location shows two separate clusters for Rotterdam and Utrecht. St. Louis, Sydney, and London clusters show in between each other. Samples based on origin seem to largely overlap, there is a large diagonal cluster for septum with most of the samples of left ventricle and right ventricle clustering slight above it or to the right. For run only run_11 seems to be slightly below the rest of the clusters. In the sex PCA plot it seems like the female samples cluster more closely together, and the male samples are slightly above it.

Additionally, PCA plots were made for location and sex based on all patients, and a PCA plot for origin was made based on biopsies (Supplementary Figure 6). This was done because origin is a biopsy-related factor, and location and sex patient-related factors. The PCA plot for origin looks almost identical to the one based on all samples. In location the separate clusters for Utrecht and Rotterdam are now more clearly visible. For sex the PCA plot shows the same result, the clusters largely overlap, though male samples seem to cluster a bit lower than the female samples.

The sample-to-sample-distances plots, generated by DESeq2, were manually adjusted to only view the samples that DESeq2 and limma previously showed as outliers (Figure 5B and 6B). If not manually adjusted, Galaxy outputs the plots in such a way the sample names cannot be distinguished from each other. The outliers from limma also show as outliers in these sample-to-sample-distances, which was eventually used as an extra confirmation to remove them from the limma analysis.

No disease-related DEGs in sex comparison

There seems to be a slight clustering of females and males in the DESeq2-generated PCA plot. After setting the comparison Female-Male (n = 33, n = 59) on a pre-filtering value of 33, the limma analysis generated MD plots show two overlapping, but distinct clusters (Supplementary Figure 7B). The quality of this analysis looks normal (Supplementary Figure 7A, 7C). The report further reveals that there are 28 DEGs (9 upregulated and 19 downregulated) (Figure 3A and 3B).

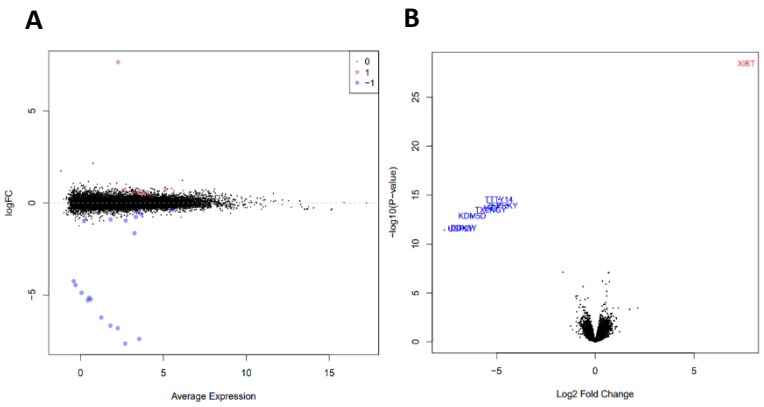


Figure 3 DEGs of the sex comparison: *The significantly differentially expressed genes in the Female-Male comparison. A. MD plot of the DEGs. B. Volcano plot of the DEGs.. Blue: downregulation. Red: upregulation. (Galaxy)*

Running the 28 DEGs through the STRING database, all but 5 genes lie on the sex chromosomes and are involved in sex-related processes. The five autosomal genes (*IFDR1, TMEM140, WDR31, RGS6*, and *DIPK1C*) are not known to be implicated in any cardiac diseases. Figure 4A reflects that the top 10 DEGs are sex-related and separate the males from the females. However, 3 male samples and 4 female samples showed opposite results compared to their group. XIST and EIF1AY expression, a female and male marker, respectively, were analysed (Figure 4B and 4C). The strip charts show that 3 female samples and 3 male samples do not cluster with their group in EIF1AY expression, and 1 female and 4

male samples do not cluster with their group in XIST expression. This shows the importance of including an analysis like this in the pipeline, to check for labelling mistakes.

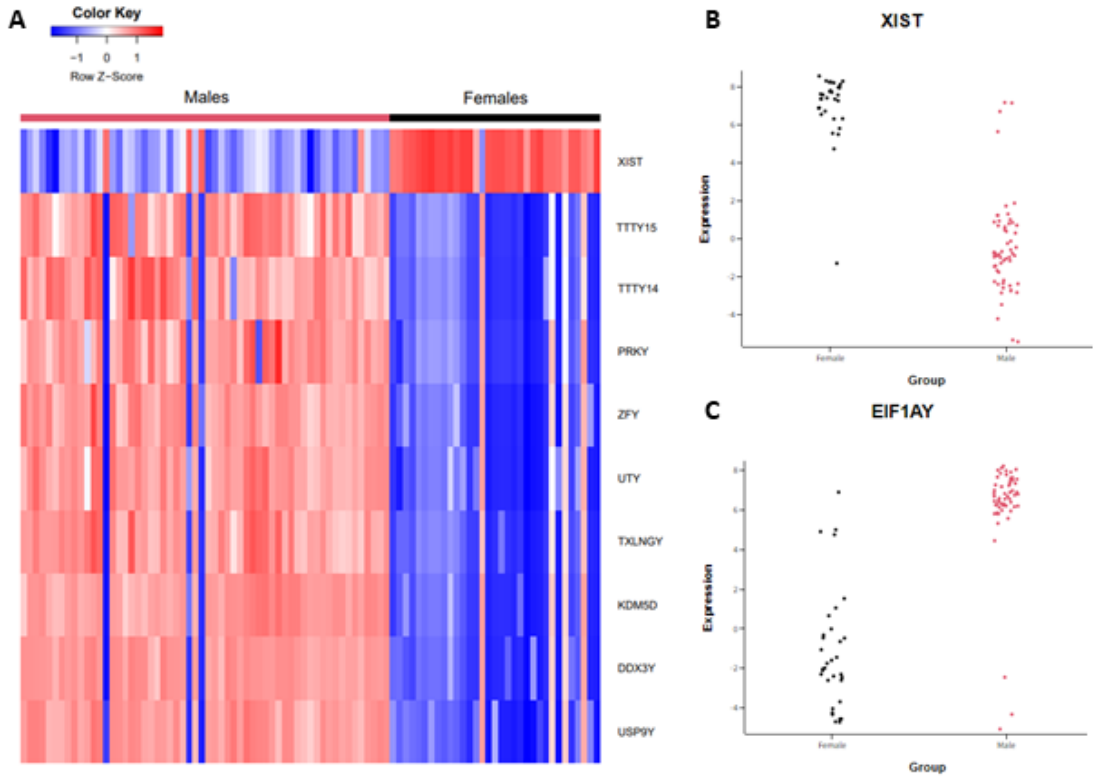


Figure 4 Expression of DEGs: *Differentially expressed genes in the Female-Male comparison. A. Heatmap of the top-10 DEGs based on adjusted P-value. B. Stripchart of XIST. C. Stripchart of EIF1AY. (Galaxy)*

Differential expression analysis

The top 50 up- and downregulated genes in HCM, DCM, and LVH were selected from the DE table from limma, and expression patterns were visualised in heatmaps (Figure 5). Clustering was done on rows, and columns are shown in a pre-selected order. The group, subgroup, and location of the samples are shown in the three bars below the columns. The figures show a clear distinction between controls and disease in all three comparisons (Figure 5). In the HCM-Control comparison, the expression of the bottom four genes in the Control group seems to relate to the location of the samples. Low expression in the St. Louis samples, higher expression in the Sydney samples, and lower expression in the Utrecht samples (Figure 5A). In the DCM-Control comparison, the control samples from Utrecht and three samples from Sydney show fewer striking colours relating to expression, when compared to the other samples (Figure 5B). The bottom cluster of genes resembles the DCM samples, while the top cluster resembles the control samples. This pattern of less expression and resembling the other control samples is also visible in the HCM-Control and LVH-control comparisons (Figure 5A and 5C). No FAO genes are found among the top 50 DEGs in all three comparisons. Only myosin heavy chain 6 (*MYH6*), a known cardiomyopathy gene, is one of the top 25 most downregulated genes in HCM-Control (Figure 5A). In the top upregulated genes in LVH-Control, shown in the bottom cluster, there are five genes that show heterogeneous expression across the control samples (Figure 5C). Four of the five genes (*DDX3Y*, *KDM5D*, *TXLNGY*, *USP9Y*, and *UTY*) are Y-linked, and all five show relation to male-specific diseases. Although these findings are not metabolism related, this overview shows a clear distinction between control and disease samples and shows the nonconforming expression of the control samples from Utrecht.

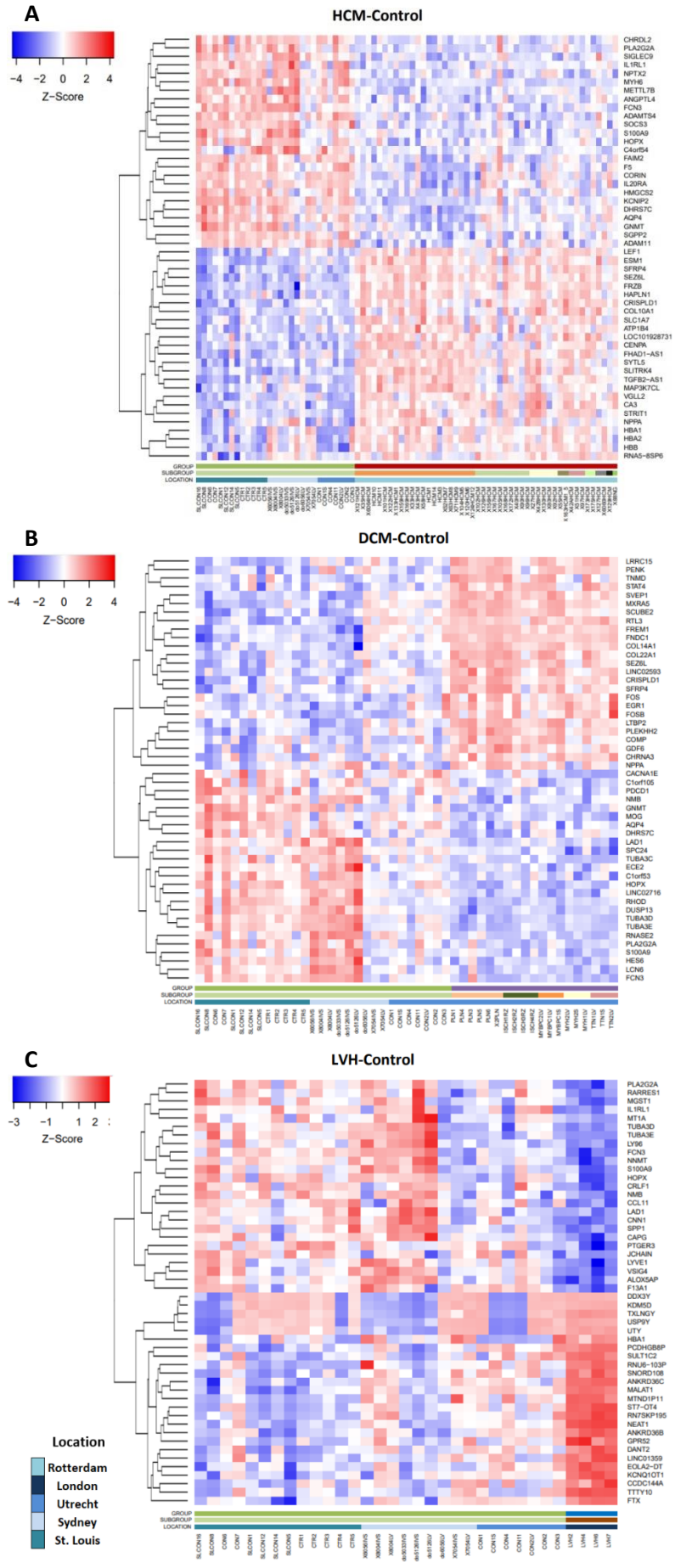


***Figure 5 Heatmaps of the top 50 differentially expressed genes**: The top 50 up- and downregulated genes are clustered on rows, and shown on the right side. Samples are shown in a pre-selected order based on group, subgroup, and location, shown in the three bars below the columns. The different colours represent different groups. A. Visualisation of the DEGs in HCM-Control, B. DCM-Control, C. LVH-Control. (R)*

Differential expression of FAO genes list

A targeted analysis of the differential expression was achieved through selection of the FAO genes list in the normalized counts. In the heatmap, clustering was executed on the genes and samples were sorted based on the group, subgroup and location as shown in the bars below the columns (Figure 6). Four genes (*ABCB11, SLC27A2, ACOXL, FABP1*) are not present in the heatmap, therefore 77 genes from the FAO genes list are visualised. The clustering of genes shows a separation between the top 37 genes and the bottom 40 genes, in green and purple bars on the left, respectively. The control group shows heterogeneity based on location. The control samples from St. Louis show high expression across the two clusters, the samples from Sydney show low expression in the purple cluster and high expression in the green cluster. The control samples from Utrecht, show a similar pattern as DCM and LVH, a high expression in the purple cluster and a low expression in the green cluster. Overall, the control samples show high expression in both clusters. The green cluster is downregulated in DCM and LVH and upregulated in HCM. The purple cluster is upregulated in DCM and LVH and downregulated in HCM. The control group shows a high expression for both the clusters. DCM and LVH show similar expression patterns compared to HCM. The HCM group also shows heterogeneity, although it is not clear whether it relates to the subgroups. Most samples in the HCM group show a relatively low expression in the purple cluster, and a high expression in the green cluster. HCM samples with the *MYBPC3* variant show either low or high expression in the purple cluster of genes, and high expression in the green cluster. The HCM-SMN samples do not clearly show low or high expression across the two gene clusters. One sample in the DCM-PLN group shows a lower expression across all genes compared to the PLN group.

Taking a further look into the two gene cluster, STRING analysis shows that the distinction between the two gene clusters is not based on peroxisomal and mitochondrial genes, or positive and negative regulation of FAO. However, there are specific clusters of genes in the green or purple gene cluster (Supplementary Figure 8). All the peroxisomal PEX genes (*PEX2, PEX5, and PEX7)*, three of the five acyl-CoA dehydrogenases (*ACADS, ACADM, ACADVL*), the peroxisomal and mitochondrial acetyl-CoA acyltransferases (*ACAA1, ACAA2, ACAT1, ACAT2*), are in the green cluster. The peroxisomal acyl-CoA oxidases (*ACOX1, ACOX2, ACOX3),* two of the PPARs (*PPARA, PPARD),* two of the RXRs (*RXRA, RXRG),* and *HADHA* and *HADHB* are in the purple cluster with *AKT1*. *AKT2, PPARG, RXRB,* and *KLF15* are in the green cluster. The ABCD transports are also split between the two clusters, *ABCD1* and *ABCD2* in the green cluster and *ABCD3* in the purple cluster. Thus, there is a specific split between the FAO genes in terms of up- or downregulation in cardiomyopathy and the expression changes based on the group, HCM, DCM, or LVH.

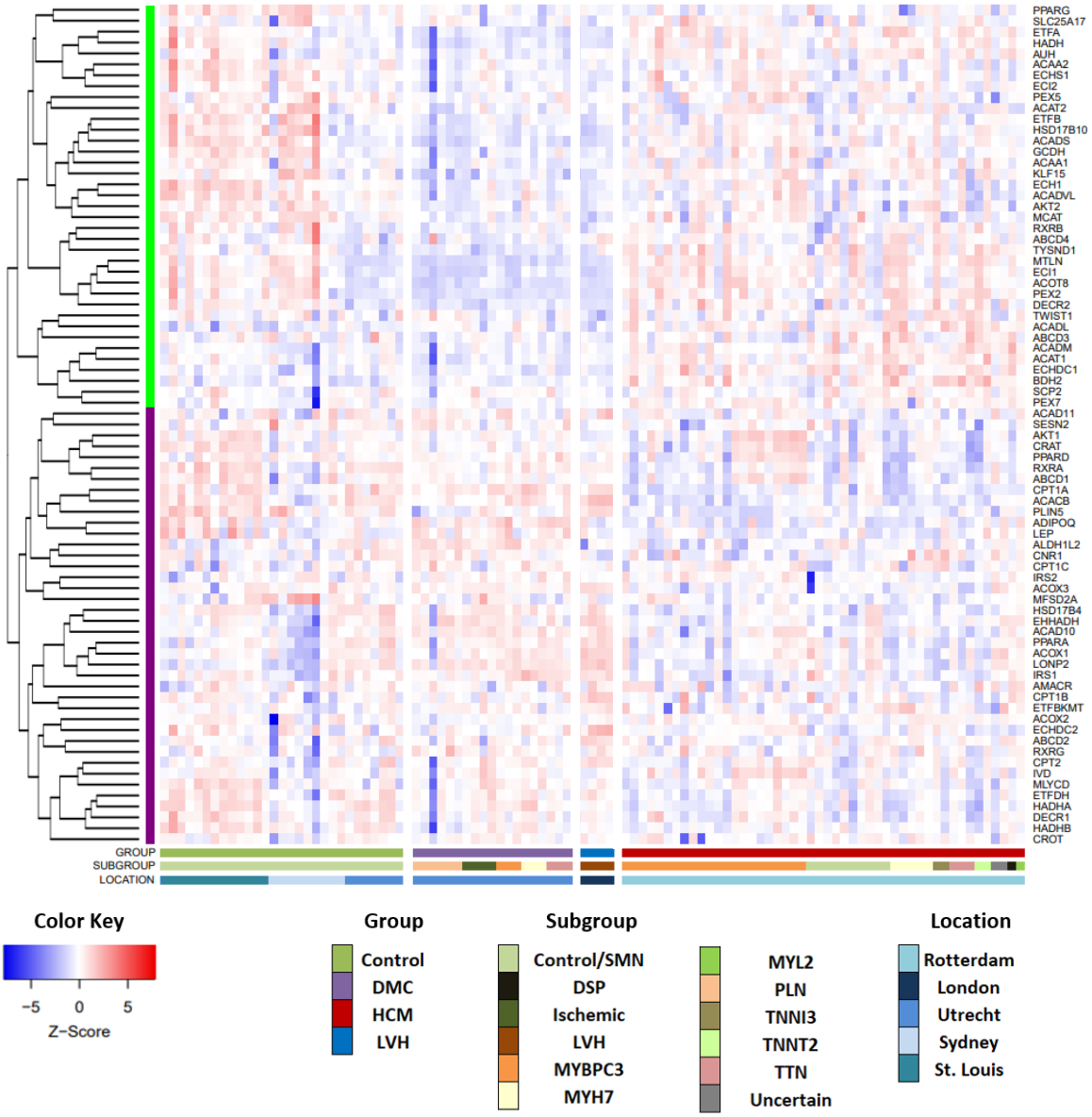


Figure 6 Expression of the FAO genes list across the four groups*: the expression of the FAO genes list is visualised according to Z-score. Dendrograms on the left represent the clusters of genes, and divides the genes in two clusters and shown by the green and purple bars. Below the heatmap are coloured bars representing the group, supgroup or location of the samples. (R)*

Specific DEGs in HCM, DCM, and LVH

A deeper look into the expression of the FAO genes list in the different comparisons shows that there are 5 DEGs in HCM-Control, 6 in DCM-Control, and 6 in LVH-Control with a significant P-value (P-value < 0.05) and a fold change > 1, these are the genes lined in red (Figure 7). Decreasing the fold change limit to 0.5 shows a big increase in the number of significant genes (HCM = 18, DCM = 25, LVH = 13). A majority of the DEGs in HCM and DCM are downregulated, while in LVH there are 7 upregulated genes and 6 downregulated genes. *ADIPOQ* is downregulated in HCM, but upregulated in DCM. *RXRA* is one of the genes downregulated in HCM, but *RXRG* is upregulated in DCM. *MTL* is upregulated in HCM, and downregulated in DCM. *ADIPOQ* and *MTLN* are two of the few genes with a fold change > 1. There are genes that are downregulated in all three groups (*ACADVL, ECH1),* or in HCM and DCM (*PLIN5, PPARG, ACADVL, ECH1).* All the downregulated genes in LVH are also downregulated in DCM. None of the genes that are upregulated overlap between the three groups. These results further

confirm the pattern that is seen in the FAO genes heatmap; HCM has a different gene expression profile than DCM and LVH.

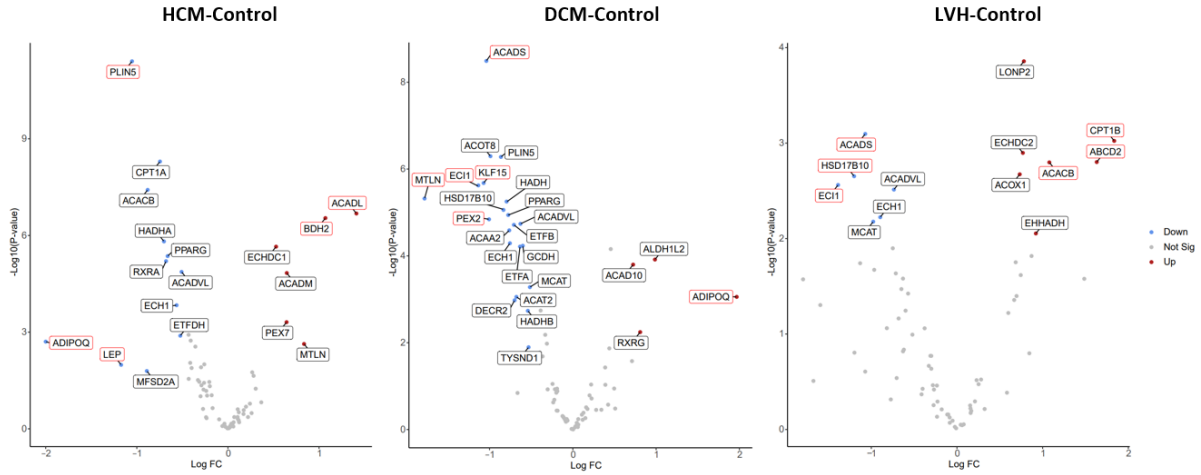


Figure 7 Volcano plots of the significant DEGs in the three groups: *Volcano plots of FAO genes in the three groups (HCM, DCM, and LVH). Labelled genes have a P-value < 0.05 and a FC > 0.5. Genes that are lined with red have a FC > 1. The X-axis shows the Log(FC) and the Y-axis shows the -Log10(P-value). Blue dots represent downregulated genes and red dots represent upregulated genes. (Galaxy)*

Gene enrichment analyses

For gene enrichment with FGSEA, the genes are ranked according to the t-statistic, which means that the genes are ranked based on the ratio of the Log2(FC) to its standard error. An enrichment score (ES) reflects whether a gene is represented at the top or bottom of the ranked list; a low ES translates to downregulation and a high ES means upregulation. FGSEA for the 31 gene sets shows a significant enrichment score for 15 sets in HCM, 11 sets in DCM, and 6 sets in LVH (Figure 8, Supplementary Table Gene enrichment result: Table 1). All these sets have a normalized enrichment score (NES) of lower than -1. Three gene sets (adipogenesis, fatty acid metabolism, glycolysis) have a significant negative NES across all three diseases (Figure 8C). Five gene sets have shared enrichment in HCM and DCM (fatty acid metabolic process, lipid oxidation, fatty acid oxidation, long-chain fatty acid metabolic process, FAO genes list, cellular amino acid metabolic process) (Figure 8A and 8B). Enrichment in one gene set (glycolytic process) is shared in HCM and LVH, and for two gene sets (oxidative phosphorylation and ROS pathway) it is shared in DCM and LVH (Figure 8B and 8C). Regardless of the adjusted P-value, most gene sets show a negative NES. The only gene sets that show a positive NES in one or two groups are negative regulation of fatty acid metabolic process, fatty acid omega-oxidation, regulation of glucose metabolic process, and PPARA 1.

GOseq is a method for gene enrichment that corrects for length bias. GOseq shows no significant enrichment for any of the 23 gene sets (fatty acid metabolic related gene sets and additional gene sets) (Supplementary Figure 9, Supplementary Table Gene enrichment results: Table 2). The lowest adjusted P-value is 0,068 for glycolytic process in LVH. Almost all gene sets in LVH have an adjusted P-value of 1, and all gene sets in FA metabolic related gene set 2 have an adjusted P-value of 1 in DCM. HCM has four gene sets with an adjusted p-value below 0,5: ketone body metabolic process, regulation of glucose metabolic process, positive regulation of fatty acid oxidation, and response to fatty acid.

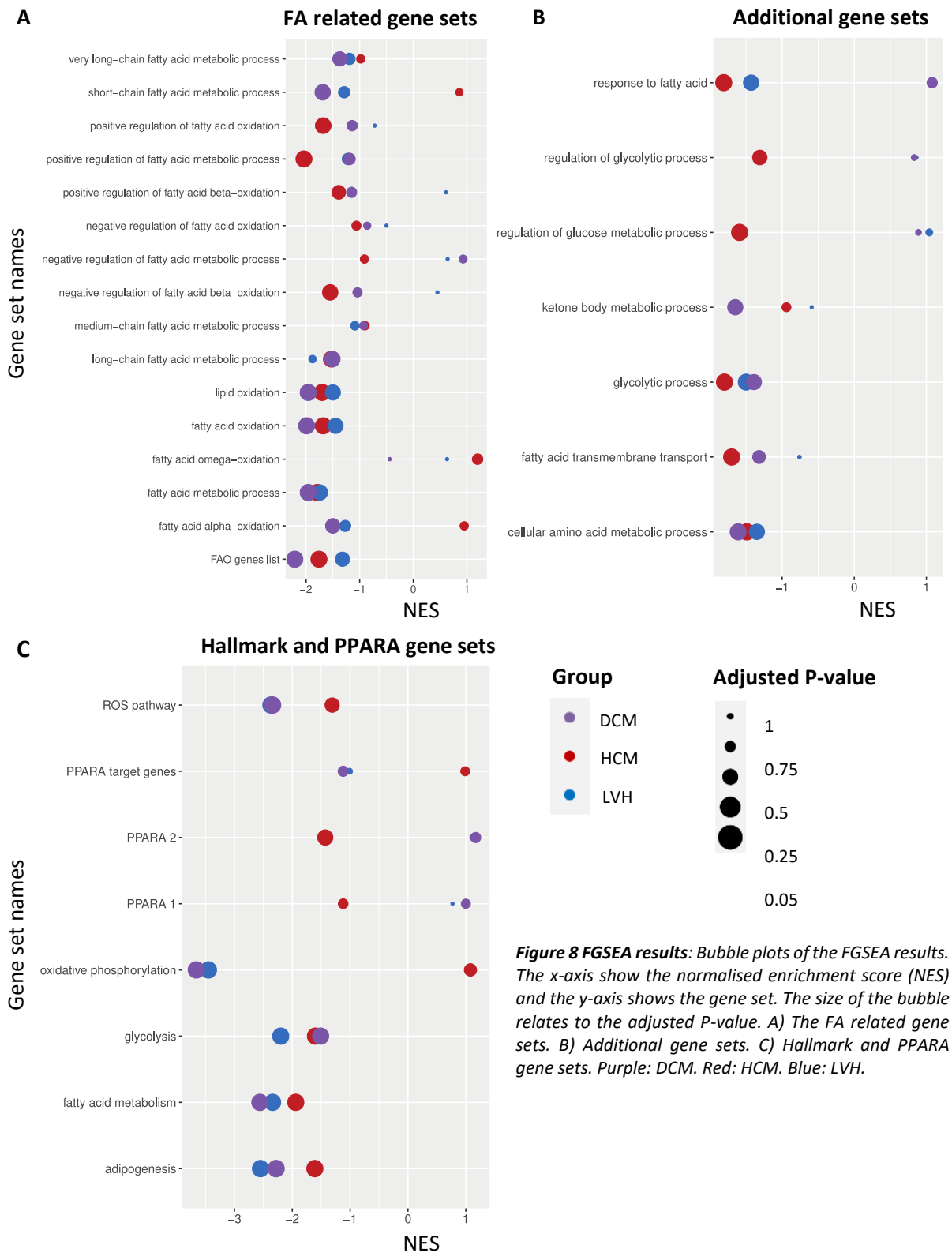


***Figure 8 FGSEA results**: Bubble plots of the FGSEA results. The x-axis show the normalised enrichment score (NES) and the y-axis shows the gene set. The size of the bubble relates to the adjusted P-value. A) The FA related gene sets. B) Additional gene sets. C) Hallmark and PPARA gene sets. Purple: DCM. Red: HCM. Blue: LVH.*

PPAR-modulating compounds

To properly understand genes involved in PPARα transcriptional regulation, the second objective of this project was to identify all natural and synthetic modulators of PPARα, and its isotypes PPARγ and PPARδ. Once a list was compiled, the second objective was to find all available datasets on the use of PPARα-modulating compounds so that these could be used to compare our results with.

From the semi-systematic PubMed search, a total of 88 PPAR agonists and 4 antagonists were found (Supplementary Table PPAR: Table 1). For this project, all specific PPARδ and PPARγ agonists are excluded as the focus is mainly on PPARα-mediated substrate flexibility. All agonists that are only mentioned in one paper or agonists of which it is not clear whether it is a direct ligand are excluded. This results in a table with 48 PPARα agonists, and 3 PPARα antagonists. Of these 48 compounds, four compounds are pan-agonists (bezafibrate, lanifibranor, chiglitazar, and indeglitazar), and 24 dual agonists (PPARα/PPARγ and PPARα/PPARδ). Eight of the synthetic PPARα agonists have been approved for clinical use in a variety of diseases, and twelve others have undergone clinical trials. Even though PPARα agonists are being used/tested for clinical use, none of the compounds is currently being developed or tested for cardiomyopathies.

Integration of public dataset

The search for transcriptome-wide datasets involving PPARα-modulating compounds yielded a list of 48 articles, including various organisms and various tissues (Supplementary Table PPAR: Table 2). From the 48, two articles were found where the influence of a PPARA-modulating compound was tested in mice cardiac tissue [42,43]. No articles were found in human cardiac tissue. The two articles used either bezafibrate or fenofibrate against cardiac hypertrophy or Barth syndrome. The selected article, from Schafer, C. *et al.*[42], uses TazKD mice as a model for Barth syndrome. This Barth syndrome model in mice exhibits DCM, similarly human patients often present DCM as well [42,44].

The expression data of the bezafibrate-treated mice, and our own data was compared (Figure 9). A majority of the FAO genes list was not found to be significantly differentially expressed in the Schafer, C. *et al.*[42] dataset by bezafibrate, thus there is no data on fold change from these genes. The fold change of all FAO genes in our dataset are visualised in the figure, significance (P-value ≤ 0.05) is marked by a black bar on top. The heatmap is ordered by the log FC of the DCM – Control group. From the 25 FAO genes that bezafibrate upregulates, 12 of 17 downregulated genes are significant in DCM, 7 of 19 downregulated genes are significant in LVH, and 10 of 16 downregulated genes are significant in HCM. There are genes that bezafibrates upregulate that are significantly upregulated in our data as well, like *IRS1*, *RXRG, ALDH1L2* in DCM (*n = 8*), *IRS1* in LVH (*n = 4*), and *ACAT1* in HCM (*n = 7*). Thus, of the genes that bezafibrate affects, more genes are (significantly) downregulated on our data. Nevertheless, bezafibrate also upregulates genes that are upregulated in our data.

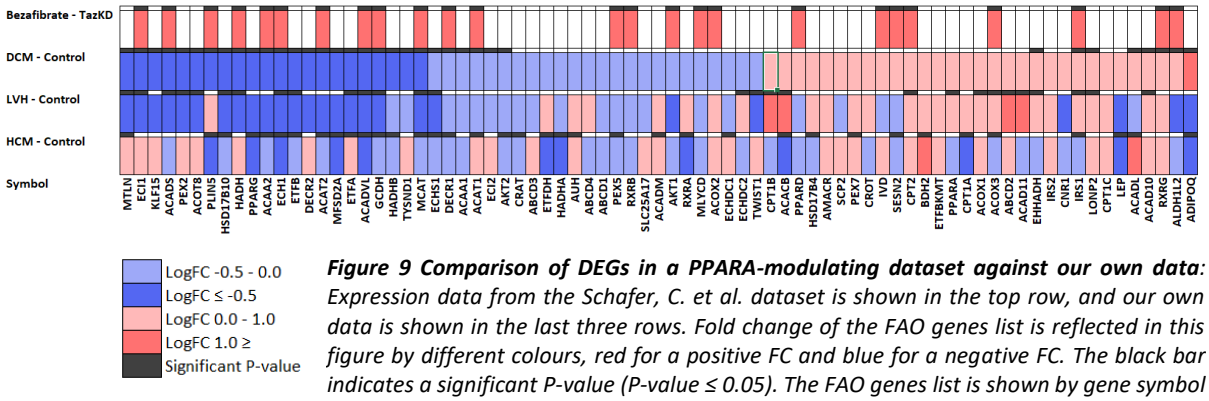


***Figure 9 Comparison of DEGs in a PPARA-modulating dataset against our own data**: Expression data from the Schafer, C. et al. dataset is shown in the top row, and our own data is shown in the last three rows. Fold change of the FAO genes list is reflected in this figure by different colours, red for a positive FC and blue for a negative FC. The black bar indicates a significant P-value (P-value ≤ 0.05). The FAO genes list is shown by gene symbol below the heatmap in columns.*

Motif enrichment

PPARA is known to regulate peroxisomal and mitochondrial FAO genes. To check the transcriptional regulation of PPARA, four known PPARA targets were chosen from peroxisomal and mitochondrial FAO, and each present in a different cluster (green or purple) in Figure 6 (*ACOX1*, *SCP2*, *CPT2*, *HADH*).

As a control, a FAO gene was selected to was known to not be directly regulated by PPARA; HADHA/HADHB. HADHA/HAHDB have a bidirectional promotor, thus one promotor was taken for these genes.

By retrieving the TSS from the region of the first transcript, an overlapping PPARA+RXR motif was found in one of the five promotor regions, *HADH* (Supplementary Table Motif enrichment: Table 3). PPARG+RXRA overlapped occurred in *ACOX1* and *SCP2.* KLF15 overlapped with RXRA on five occurrences, in three genes (*ACOX1, CPT2, HADHA/HADHB).* Furthermore, KLF15 overlapped with PPARG and RXRA in *ACOX1,* and with PPARA and PPARG in *HADHA/HADHB.* As for orientation of the motifs, all motifs were found in the opposite orientation as the gene, except for one motif in *ACOX1* and one motif in *HADHA/HADHB.* The results from HOCOMO, including found motif, P-value and coordinates, can be found in Supplementary Table Motif enrichment: Table 4.

CellProfiler

PPARA in the nucleus was successfully quantified with the CellProfiler pipeline (Supplementary File CellProfiler). However, other staining's and additional pipelines are needed to fully analyse PPARA in cardiomyocytes.

Discussion

For this project, a novel pipeline was set up and recommendations were made to analyse RNA-seq data from a retrospective patient cohort of 94 patients, with a total of 102 biological samples. The analyses have shown the possibilities of a transcriptomics project within a less coding-intensive environment named Galaxy. With the use of this pipeline, the expression patterns of FAO genes in HCM, DCM, and LVH are elucidated. Additionally, the pipeline has outlined the importance of comprehensive quality control when dealing with a large heterogenic cohort; how to deal with outliers and a check for labelling. The results show a split in the regulation of FAO genes that has not been shown to this extent before. First, it was shown that the significant FAO DEGs in HCM and DCM are downregulated, which was further confirmed by gene set enrichment analysis. Secondly, FA related gene sets primarily show negative normalized enrichment scores, indicating enrichment of downregulated genes in that particular gene set. Next, a more exploratory approach was taken to study the relationship between FAO genes and PPARA, which is the critical transcription factor regulating FAO signalling, by collecting a list of PPARA modulators and public transcriptomics datasets that used PPARA modulators. Combined with the integration of a public dataset that used bezafibrate, a PPARA agonist, the potential of PPARA as a therapeutic target was investigated. Lastly, a motif enrichment analysis was tried out in order to analyse PPARA regulation in FAO genes. This type of analysis and the interpretation of the results needs to be more extensively examined in the future.

Our novel meta-analysis pipeline has successfully shown the different regulation patterns between the three types of heart failure. The results indicate that there are similar FAO expression patterns between DCM and LVH, whereas HCM showed distinct FAO expression patterns. This shows in both the heatmap, where there is a contrasting expression pattern between HCM and DCM with LVH, and the volcano plots. Of the significant downregulated DEGs in HCM (n = 12) and DCM (n = 21), only four are shared between them. This distinct profile of HCM and DCM is in line with recent literature [28,45,46]. Research showed that KLF15 is downregulated in DCM, but upregulated in HCM [26,30]. The regulation of the peroxisomal acyl-CoA oxidases (*ACOX1, ACOX2, ACOX3*) is shown to be upregulated in DCM when compared to non-failing hearts, but remains unchanged in HCM [28,47]. These findings align with our data. We also observe the upregulation of the ACOX genes in DCM hearts. Besides, *ACOX1* is significantly upregulated in LVH. Additionally, *KLF15* is significantly downregulated in DCM, although our results do not show a significant change in HCM. *KLF15* is a transcription factor involved in controlling cardiac metabolism and cooperates with *PPARA* to regulate lipid metabolism. Studies have shown that KLF15 expression protects against hypertrophy and fibrosis [48]. Furthermore, we see more opposite expression profiles between HCM and DCM in our data. *ADIPOQ* is significantly downregulated in HCM and significantly upregulated in DCM. The opposite is true for *MTLN* and two of the *PEX* genes. *PEX7* is significantly upregulated in HCM, while *PEX2* is significantly downregulated in DCM. The PEX genes are involved in peroxisomal biogenesis and import of substrates into the peroxisome [49]. Mitoregulin (*MTLN)* interacts with the mitochondrial trifunctional proteins in the regulation of FAO, of which a previous study has shown downregulation in DCM [26,50]. In contrast with these findings, gene enrichment does not confirm the distinct profiles of HCM and DCM. Despite a few nonsignificant opposite enrichment scores for HCM and DCM in the FA-, additional-, and hallmark gene sets, the general consensus is a negative enrichment score for all three diseases.

Beside a downregulation of FAO and an upregulation of glucose metabolism, studies have revealed alterations in other metabolic pathways [11,51–53]. Ketone metabolism is downregulated in both HCM

and DCM, but amino acid metabolism, and oxidative metabolism is downregulated in DCM and upregulated in HCM [28,54–57]. Consistently, we show a positive enrichment score for HCM in oxidative phosphorylation and for DCM in regulation of glycolytic process and regulation of glucose metabolism. Interestingly, we did not observe positive enrichment scores for ketone metabolism in HCM and DCM, or positive enrichment scores for glucose metabolism and amino acid metabolism in HCM. Instead, all gene sets showed negative enrichment scores for both DCM and HCM. While these results might suggest that these metabolic pathways are similarly downregulated in HCM, DCM, and LVH, it is plausible that the two chosen gene sets per metabolic pathway are not representative of the complex metabolic pathways. It is recommended to repeat the gene enrichment analysis with a more complete gene set collection better presenting glucose, ketone, and amino acid metabolism. Furthermore, to investigate whether the first results of a different FAO regulation in HCM and DCM align with gene enrichment, the enriched genes within each gene set can be compared between the two diseases.

Our results show a split in the FAO genes list in terms of regulation according to disease. There is interplay between mitochondrial and peroxisomal FAO, and each pathway uses different fatty acids as substrates [58–60]. For the first time, we specifically looked at the expression of mitochondrial and peroxisomal FAO genes, and the implication of it. Contrary to the hypothesized link that the split is due to a separation in peroxisomal and mitochondrial genes, our data shows an equal distribution of peroxisomal and mitochondrial genes in each cluster. There seems to be a yet undefined reason for this split. In our data we see a split in the peroxisomal ABCD genes, which transport lipids into the peroxisomal matrix [61]. *ABCD1* and *ABCD2* are downregulated in DCM and LVH, and *ABCD3* is downregulated in HCM. Downstream genes are the ACOX genes and *SCP2*, which are upregulated in DCM and downregulated in HCM, and vice versa. This shows that proteins that directly interact with each other show opposite expression profiles within this observed split. As for the PPARs and RXRs, the members are also present in the split. *PPARG* is significantly downregulated in DCM, and *PPARA* is slightly upregulated in DCM and LVH, although this upregulation is not significant. This relates with previously mentioned unpublished research by our group, where the expression of *PPARA* seemed unchanged, but acetylation levels were different and downstream effectors were downregulated [26]. These downstream effectors, *HADHA* and *HADHB*, also show significant downregulation in our data. For which *HADHA* is downregulated in HCM, and *HADHB* in DCM. *HADHA* was also significantly downregulated in a multi-omics paper comparing 27 HCM patients with 13 healthy controls [62]. The dysfunction or deficiencies in mitochondrial trifunctional proteins, that are formed by *HADHA* and *HADHB*, show severe clinical symptoms and can present itself as HCM or DCM, amongst other types of cardiomyopathy [63]. MTPs catalyse the last steps in mitochondrial FAO and are therefore crucial [64]. Further research is needed to investigate this split in FAO genes and the implications thereof.

The explorative investigation of PPARA regulation and modulation has shown that there are already a lot of PPARA modulators being researched, developed, and approved for clinical use. The integration of a public dataset shows that a PPARA agonist can specifically elevate expression of FAO genes that are downregulated in cardiomyopathy, suggesting a positive effect of bezafibrate in cardiomyopathy. However, of all the datasets found, none of the datasets were on human cardiac tissue. This poses as a limitation in the direct translation of this result to the use in humans. As mentioned before, PPARA modulators like bezafibrate have already been approved for clinical use e.g. dyslipidemia [65]. The use in humans has thus already been deemed safe. Additionally, there are ongoing clinical trials for the use of bezafibrate in diseases that reflect the impaired mitochondrial function and lipid accumulation as seen in cardiomyopathy, like mitochondrial disease and neutral lipid storage disease with myopathy

[66–68]. As the effect of fibrates has not been researched in human cardiomyopathy, these findings show promising results for the use in cardiomyopathy and emphasise the need for more research into PPARA as a therapeutic target. The last arising question involving PPARA modulation, is the specificity of the PPARs and RXRs. As shown the PPARs and RXRs show structure and sequence homology and their bindings motifs are extremely similar [18,20,69]. Subsequently, motif enrichment was done to find out more about PPAR regulation. Our results did not return the expected results yet, as we did not find any PPARA motifs in genes that are known to be regulated by PPARA. There were motifs found on the reverse orientation as the gene, although this is not of importance for gene regulation as the orientation of the motifs does not matter [70]. Future projects should focus on perfectioning the motif enrichment analysis to further define the regulation of PPARs and their specificity.

This project is the first time that a meta-analysis was performed on a large retrospective cohort, that included this many genetic variants. There are studies done in large cohorts, but these are either single-centre, or do not contain all three disease types and as many variants as this cohort [71–73]. The heterogenicity of the cohort is both a strength and a limitation. The heterogeneity poses as an opportunity to elucidate disease-wide disturbances and changes, instead of finding variant-specific variation in small cohorts. However, by using a heterogenic cohort, some form of bias has to be accepted. For instance, biopsies from four different locations in the heart were used. Therefore, this could also be a cause of variance in our data. This highlights the importance of an extensive quality control and looking for possible confounders in the data. In the part of outlier removal, we show clustering based on 500 and 5,000 genes, where we do not see any outliers based on 5000 genes. An explanation for this might be that a higher selection of genes, causes less clear outlier. Therefore, it is better to cluster on a large group of genes. Another result from the quality control analyses is that origin and run do not cause variance in the data. Although it looks like location might be a confounder due to the separate clusters for Rotterdam and Utrecht, it can be expected that this is actually variance caused by disease, since all samples from Rotterdam are HCM samples and Utrecht only supplied control samples. This shows the hyper-specialisation of hospitals and how relevant it is to assemble and promote more inclusive single-centre biobanks. Another way this reflects in the results, is in the untargeted analysis on the top 50 DEGs. The control samples from Utrecht do not show the same results as the other control samples from other locations. This might be due to the tissue extraction protocol from Utrecht, in comparison to the direct tissue sample handling happening in other locations.

Due to lack of samples in the LVH group, the reliability of the results for this group are questionable. The lack of samples in LVH, compared to the number of samples in HCM and DCM, might have led to robustness into detecting DEGs between LVH and control. Moreover, the generalizability of the results from the FAO genes heatmap is limited by the overall limited expression and thus dull colours. It is hard to draw concise conclusions based on this. At last, future studies should take into account that the genes in GO:terms are based on different organs, and therefore not heart specific. This led us to include genes in our FAO genes list that are not expressed in the heart (*ABCD11, ACOXL, FABP1, SLC27A2)*. Due to no expression of these genes in the heart, the genes were filtered out during limma analysis as is seen in the heatmap and gene interaction figure.

In conclusion, my project presents a novel pipeline for the meta-analysis of a large heterogenic patient cohort. The data showed distinct expression profiles within three types of heart failure, and with gene enrichment the downregulation of FAO was further confirmed. Herein I emphasize the importance of

an extensive quality control with a more detailed description on how to deal with outliers and how to check the sex-labelling of your samples. The more explorative and integrative research highlighted the importance of further investigation into PPARA regulation. The use of PPARA-modulating compounds or metabolic-altering drugs might benefit cardiomyopathy patients and open new avenues for drug repurposing.

Future perspectives for Galaxy

For this project, the aim was to create a workflow on a local Galaxy docker on an Azure DRE environment. Since 2018 the General Data Protection Regulation (GDPR) has come into effect in the European Union (EU). GDPR governs the regulation and protection of the personal data of natural persons. This entails that people hold the right to decide what happens with their personal information and that all institutions need to uphold and respect that right. GDPR states certain rules and regulations about the storage and processing of personal data. Personal data may not be uploaded or analysed on open and accessible servers [74]. UMC Utrecht, in collaboration with Erasmus MC Rotterdam and Radboudumc Nijmegen, has created anDREa B.V., the company behind Azure DRE. Azure DRE offers researchers a GDPR-compliant digital environment where personal data can be securely stored, shared between collaborators and accessed from anywhere [75]. The combination of Galaxy and Azure DRE allows research to be safely conducted and shared between collaborations, without violation of data protection laws. The generation of bioinformatics pipelines within this Galaxy Docker offers the future possibility of bedside analyses.

The integration of scripted analyses in user-friendly tools has made Galaxy an interesting software for future diagnostics. Researchers in the field can make workflows that can supply clinicians who have little to no bioinformatics background with a fully operational pipeline. It is a great step towards personalized medicine, where patient material can be analysed and visualised bedside within hours. The opportunities are limitless. My project has shown that it is possible to analyse a large cohort within this platform, and that multiple different analysis and quality checks can be properly done. However, working on this report, several things have stood out that would be great additions to the Galaxy platform for better usability and improved analytics.

Limitations and recommendations for Galaxy

Since 2005 Galaxy is an open-source platform that strives for accessibility, reproducibility, transparency and scalability. The platform has integrated more than 8,000 packages in a user-friendly interface called tools, that allow users to easily integrate and reproduce data analyses in their research. Whole pipelines of analyses can be saved and edited into workflows that can be run by anyone and anywhere if access is granted. Galaxy currently hosts three servers in Australia, Europe, and the United States, though the server can also be installed on clusters or local clouds. Besides data analysis and visualization tools, Galaxy offers workshops and has set up large, comprehensive training materials and modules that enable users to start and learn new analyses and tools [39]. In this section I provide a list of limitations and recommendations for the Galaxy platform.

Copying of datasets

The current version of Galaxy works with the beta history panel. This panel shows the current history datasets and offers options to, for instance, copy the history, export the tool citations or set the permissions of the history. The previous version of Galaxy used the legacy history panel, which had

some features that were beneficial compared to the beta history panel. One such function was the ability to select history datasets and copy them to new or existing datasets. In the beta history panel it is only possible to copy whole histories, not single datasets. This feature allows for the easy structing of analyses in Galaxy and continuing analyses in different histories without having to deal with all the data of another history.

Analysis and visualisation with DESeq2 versus limma

Usually, for RNA-seq analysis DESeq2 is more often used than limma. In Galaxy the usability and visualisation option of limma offer an advantage so that the use of limma is preferred over the usage of DESeq2. DESeq2 provides the PCA plots and sample-to-sample distances that are good to use for quality control and outlier removal. Limma offers the possibility to set multiple comparisons in one analysis run, multiple QC plots and the MDS plots, and with an annotation file limma offers Glimma interactive plots. For improvement, it would be best if DESeq2 allowed for multiple factor analysis and returned the output of all the comparisons. Also, the visualisation option for DESeq2 leaves something to be desired.

Plots and heatmap2

Galaxy has tools to visualise data and with various tools it offers to output plots immediately. However, with a lot of these tools it is not possible to change colour schemes, change font sizes or select data within the tool (e.g. the heatmap2 tool). When making a heatmap with heatmap2, and the supplied dataset exceeds approximately 20 samples and 20 genes, the samples become unreadable and it selectively chooses which genes to output on the axis. So, it is not possible to relate rows back to genes. This limitation could be easily solved by providing the option to change font sizes.

Rscript and RData

A possible solution to the visualisation problem, is to download the RData and Rscript from Galaxy and manually change the plots in R studio. Despite this, the option for RData and Rscript is not available for every tool. It is therefore recommended that this will become available in more tools.

GOseq returns NA with supplied gene categories

The GOseq tool on Galaxy performs gene enrichment. This tool needs a datafile that states which genes are differentially expressed in Booleans and a gene length file for length bias correction. Further, there is an option to let the tool get categories to file the genes in, or to supply your own categories file. When a categories file is supplied, the tool errors in one of the outputs. It is an option to output the DE categories list, which is a table that states which genes were found in what GO terms. If categories were manually supplied, the lists only returns NA values. It is thus not possible to trace back which of the DEGs were categorized/found in the GO terms.

Acknowledgements

Foremost, I would like to thank my supervisor Magdalena Harakalova for her guidance, positivity and understanding during these past six months. I am grateful for the opportunity to work on a paper and to get published! I would like to thank Jiayi, Talitha, Karen, and Frank as well for the helpful discussions and feedback, and for always being willing to make time for me. The support from the group has been greatly appreciated. Also, I am grateful to have found friends in my fellow students Danielle, Maaike, Luka, Quinn and Babette.

References

1. Snipelisky, D., Chaudhry, S.-P. & Stewart, G. C. The Many Faces of Heart Failure. *Card. Electrophysiol. Clin.* **11**, 11–20 (2019).
2. Brieler, J., Breeden, M. A. & Tucker, J. Cardiomyopathy: An Overview. *Am. Fam. Physician* **96**, 640–646 (2017).
3. Yamada, T. & Nomura, S. Recent Findings Related to Cardiomyopathy and Genetics. *Int. J. Mol. Sci.* **22**, 12522 (2021).
4. Teekakirikul, P., Zhu, W., Huang, H. C. & Fung, E. Hypertrophic Cardiomyopathy: An Overview of Genetics and Management. *Biomolecules* **9**, 878 (2019).
5. Santos Mateo, J. J., Sabater Molina, M. & Gimeno Blanes, J. R. Hypertrophic cardiomyopathy. *Med. Clin. (Barc.)* **150**, 434–442 (2018).
6. Rosenbaum, A. N., Agre, K. E. & Pereira, N. L. Genetics of dilated cardiomyopathy: practical implications for heart failure management. *Nat. Rev. Cardiol.* **17**, 286–297 (2020).

7. Hershberger, R. E., Hedges, D. J. & Morales, A. Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nat. Rev. Cardiol.* **10**, 531–547 (2013).
8. Peters, S. *et al.* Familial Dilated Cardiomyopathy. *Heart Lung Circ.* **29**, 566–574 (2020).
9. Yildiz, M. *et al.* Left ventricular hypertrophy and hypertension. *Prog. Cardiovasc. Dis.* **63**, 10–21 (2020).
10. Montaigne, D., Butruille, L. & Staels, B. PPAR control of metabolism and cardiovascular functions. *Nat. Rev. Cardiol.* **18**, 809–823 (2021).
11. Lopaschuk, G. D., Ussher, J. R., Folmes, C. D. L., Jaswal, J. S. & Stanley, W. C. Myocardial Fatty Acid Metabolism in Health and Disease. *Physiol. Rev.* **90**, 207–258 (2010).
12. Doenst, T., Nguyen, T. D. & Abel, E. D. Cardiac Metabolism in Heart Failure - Implications beyond ATP production. *Circ. Res.* **113**, 709–724 (2013).
13. Houten, S. M., Violante, S., Ventura, F. V. & Wanders, R. J. A. The Biochemistry and Physiology of Mitochondrial Fatty Acid β-Oxidation and Its Genetic Disorders. *Annu. Rev. Physiol.* **78**, 23–44 (2016).
14. Fillmore, N., Mori, J. & Lopaschuk, G. D. Mitochondrial fatty acid oxidation alterations in heart failure, ischaemic heart disease and diabetic cardiomyopathy. *Br. J. Pharmacol.* **171**, 2080–2090 (2014).
15. Rakhshandehroo, M., Knoch, B., Müller, M. & Kersten, S. Peroxisome Proliferator-Activated Receptor Alpha Target Genes. *PPAR Res.* **2010**, 612089 (2010).
16. Wagner, N. & Wagner, K.-D. The Role of PPARs in Disease. *Cells* **9**, 2367 (2020).
17. Schoonjans, K., Martin, G., Staels, B. & Auwerx, J. Peroxisome proliferator-activated receptors, orphans with ligands and functions. *Curr. Opin. Lipidol.* **8**, 159–166 (1997).
18. Pawlak, M., Lefebvre, P. & Staels, B. Molecular mechanism of PPARα action and its impact on lipid metabolism, inflammation and fibrosis in non-alcoholic fatty liver disease. *J. Hepatol.* **62**, 720–733 (2015).

19. Hummasti, S. & Tontonoz, P. The Peroxisome Proliferator-Activated Receptor N-Terminal Domain Controls Isotype-Selective Gene Expression and Adipogenesis. *Mol. Endocrinol.* **20**, 1261–1275 (2006).

20. Lamas Bervejillo, M. & Ferreira, A. M. Understanding Peroxisome Proliferator-Activated Receptors: From the Structure to the Regulatory Actions on Metabolism. in *Bioactive Lipids in Health and Disease* (eds. Trostchansky, A. & Rubbo, H.) 39–57 (Springer International Publishing, 2019). doi:10.1007/978-3-030-11488-6_3.

21. Grygiel-Górniak, B. Peroxisome proliferator-activated receptors and their ligands: nutritional and clinical implications – a review. *Nutr. J.* **13**, 17 (2014).

22. Henke, B. R. Peroxisome Proliferator-Activated Receptor α/γ Dual Agonists for the Treatment of Type 2 Diabetes. *J. Med. Chem.* **47**, 4118–4127 (2004).

23. Garcia-Vallvé, S. *et al.* Peroxisome Proliferator-Activated Receptor γ (PPARγ) and Ligand Choreography: Newcomers Take the Stage. *J. Med. Chem.* **58**, 5381–5394 (2015).

24. Tahri-Joutey, M. *et al.* Mechanisms Mediating the Regulation of Peroxisomal Fatty Acid Beta-Oxidation by PPARα. *Int. J. Mol. Sci.* **22**, 8969 (2021).

25. Finck, B. N. & Kelly, D. P. Peroxisome proliferator-activated receptor alpha (PPARalpha) signaling in the gene regulatory control of energy metabolism in the normal and diseased heart. *J. Mol. Cell. Cardiol.* **34**, 1249–1257 (2002).

26. Pei, J. *et al.* Transcriptional regulation profiling reveals disrupted lipid metabolism in failing hearts with a pathogenic phospholamban mutation. 2020.11.30.402792 Preprint at https://doi.org/10.1101/2020.11.30.402792 (2020).

27. Sack, M. N. *et al.* Fatty Acid Oxidation Enzyme Gene Expression Is Downregulated in the Failing Heart. *Circulation* **94**, 2837–2842 (1996).

28. Gaar-Humphreys, K. R. *et al.* Targeting lipid metabolism as a new therapeutic strategy for inherited cardiomyopathies. *Front. Cardiovasc. Med.* **10**, (2023).

29. Feyen, D. A. M. *et al.* Unfolded Protein Response as a Compensatory Mechanism and Potential Therapeutic Target in PLN R14del Cardiomyopathy. *Circulation* **144**, 382–392 (2021).

30. Pei, J. *et al.* Multi-omics integration identifies key upstream regulators of pathomechanisms in hypertrophic cardiomyopathy due to truncating MYBPC3 mutations. *Clin. Epigenetics* **13**, 61 (2021).

31. van Driel, B. O. *et al.* Metabolomics in Severe Aortic Stenosis Reveals Intermediates of Nitric Oxide Synthesis as Most Distinctive Markers. *Int. J. Mol. Sci.* **22**, 3569 (2021).

32. Schuldt, M. *et al.* Proteomic and Functional Studies Reveal Detyrosinated Tubulin as Treatment Target in Sarcomere Mutation-Induced Hypertrophic Cardiomyopathy. *Circ. Heart Fail.* **14**, e007022 (2021).

33. Schuldt, M. *et al.* Mutation location of HCM-causing troponin T mutations defines the degree of myofilament dysfunction in human cardiomyocytes. *J. Mol. Cell. Cardiol.* **150**, 77–90 (2021).

34. Vigil-Garcia, M. *et al.* Gene expression profiling of hypertrophic cardiomyocytes identifies new players in pathological remodelling. *Cardiovasc. Res.* **117**, 1532–1545 (2021).

35. Pei, J. *et al.* H3K27ac acetylome signatures reveal the epigenomic reorganization in remodeled non-failing human hearts. *Clin. Epigenetics* **12**, 106 (2020).

36. Manduchi, E. *et al.* A comparison of two workflows for regulome and transcriptome-based prioritization of genetic variants associated with myocardial mass. *Genet. Epidemiol.* **43**, 717–726 (2019).

37. Hemerich, D. *et al.* Integrative Functional Annotation of 52 Genetic Loci Influencing Myocardial Mass Identifies Candidate Regulatory Variants and Target Genes. *Circ. Genomic Precis. Med.* **12**, e002328 (2019).

38. Parbhudayal, R. Y. *et al.* Variable cardiac myosin binding protein-C expression in the myofilaments due to MYBPC3 mutations in hypertrophic cardiomyopathy. *J. Mol. Cell. Cardiol.* **123**, 59–63 (2018).

39. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* **50**, W345–W351 (2022).

40. Langfelder, P. & Horvath, S. Tutorial for the WGCNA package for R: I. Network analysis of liver expression data in female mice.

41. ENSG00000258653 Gene - GeneCards | G3V3Y1 Protein | G3V3Y1 Antibody. https://www.genecards.org/cgi-bin/carddisp.pl?gene=ENSG00000258653.

42. Schafer, C. *et al.* The Effects of PPAR Stimulation on Cardiac Metabolic Pathways in Barth Syndrome Mice. *Front. Pharmacol.* **9**, 318 (2018).

43. Parry, T. L. *et al.* Fenofibrate unexpectedly induces cardiac hypertrophy in mice lacking MuRF1. *Cardiovasc. Pathol. Off. J. Soc. Cardiovasc. Pathol.* **25**, 127–140 (2016).

44. Pang, J., Bao, Y., Mitchell-Silbaugh, K., Veevers, J. & Fang, X. Barth Syndrome Cardiomyopathy: An Update. *Genes* **13**, 656 (2022).

45. Chaffin, M. *et al.* Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy. *Nature* **608**, 174–180 (2022).

46. Quttainah, M. *et al.* Transcriptomal Insights of Heart Failure from Normality to Recovery. *Biomolecules* **12**, 731 (2022).

47. Veldhoven, P. P. V. Biochemistry and genetics of inherited disorders of peroxisomal fatty acid metabolism [S]. *J. Lipid Res.* **51**, 2863–2895 (2010).

48. Zhao, Y. *et al.* Multiple roles of KLF15 in the heart: Underlying mechanisms and therapeutic implications. *J. Mol. Cell. Cardiol.* **129**, 193–196 (2019).

49. Waterham, H. R., Ferdinandusse, S. & Wanders, R. J. A. Human disorders of peroxisome metabolism and biogenesis. *Biochim. Biophys. Acta* **1863**, 922–933 (2016).

50. Friesen, M. *et al.* Mitoregulin Controls β-Oxidation in Human and Mouse Adipocytes. *Stem Cell Rep.* **14**, 590–602 (2020).

51. J, van der V. *et al.* Metabolic changes in hypertrophic cardiomyopathies: scientific update from the Working Group of Myocardial Function of the European Society of Cardiology. *Cardiovasc. Res.* **114**, (2018).

52. Tran, D. H. & Wang, Z. V. Glucose Metabolism in Cardiac Hypertrophy and Heart Failure. *J. Am. Heart Assoc.* **8**, e012673 (2019).

53. Lopaschuk, G. D., Karwi, Q. G., Tian, R., Wende, A. R. & Abel, E. D. Cardiac Energy Metabolism in Heart Failure. *Circ. Res.* **128**, 1487–1513 (2021).

54. Schugar, R. C. *et al.* Cardiomyocyte-specific deficiency of ketone body metabolism promotes accelerated pathological remodeling. *Mol. Metab.* **3**, 754–769 (2014).

55. Selvaraj, S., Kelly, D. P. & Margulies, K. B. Implications of Altered Ketone Metabolism and Therapeutic Ketosis in Heart Failure. *Circulation* **141**, 1800–1812 (2020).

56. Ampong, I. Metabolic and Metabolomics Insights into Dilated Cardiomyopathy. *Ann. Nutr. Metab.* **78**, 147–155 (2022).

57. Previs, M. J. *et al.* Defects in the Proteome and Metabolome in Human Hypertrophic Cardiomyopathy. *Circ. Heart Fail.* **15**, e009521 (2022).

58. Wanders, R. J. A., Vaz, F. M., Waterham, H. R. & Ferdinandusse, S. Fatty Acid Oxidation in Peroxisomes: Enzymology, Metabolic Crosstalk with Other Organelles and Peroxisomal Disorders. *Adv. Exp. Med. Biol.* **1299**, 55–70 (2020).

59. Okumoto, K., Tamura, S., Honsho, M. & Fujiki, Y. Peroxisome: Metabolic Functions and Biogenesis. *Adv. Exp. Med. Biol.* **1299**, 3–17 (2020).

60. Schrader, M., Costello, J., Godinho, L. F. & Islinger, M. Peroxisome-mitochondria interplay and disease. *J. Inherit. Metab. Dis.* **38**, 681–702 (2015).

61. Tawbeh, A., Gondcaille, C., Trompier, D. & Savary, S. Peroxisomal ABC Transporters: An Update. *Int. J. Mol. Sci.* **22**, 6093 (2021).

62. Ranjbarvaziri, S. *et al.* Altered Cardiac Energetics and Mitochondrial Dysfunction in Hypertrophic Cardiomyopathy. *Circulation* **144**, 1714–1731 (2021).

63. Bursle, C. *et al.* Mitochondrial Trifunctional Protein Deficiency: Severe Cardiomyopathy and Cardiac Transplantation. *JIMD Rep.* **40**, 91–95 (2017).

64. Adeva-Andany, M. M., Carneiro-Freire, N., Seco-Filgueira, M., Fernández-Fernández, C. & Mouriño-Bayolo, D. Mitochondrial β-oxidation of saturated fatty acids in humans. *Mitochondrion* **46**, 73–90 (2019).

65. Hong, F., Xu, P. & Zhai, Y. The Opportunities and Challenges of Peroxisome Proliferator-Activated Receptors Ligands in Clinical Drug Discovery and Development. *Int. J. Mol. Sci.* **19**, 2189 (2018).

66. Steele, H. *et al.* Metabolic effects of bezafibrate in mitochondrial disease. *EMBO Mol. Med.* **12**, (2020).

67. Ma, G., Ja, R., Sp, M. & Aa, F. Neutral lipid storage disease with myopathy and dropped head syndrome. Report of a new variant susceptible of treatment with late diagnosis. *J. Clin. Neurosci. Off. J. Neurosurg. Soc. Australas.* **58**, (2018).

68. T, van de W. *et al.* Effects of bezafibrate treatment in a patient and a carrier with mutations in the PNPLA2 gene, causing neutral lipid storage disease with myopathy. *Circ. Res.* **112**, (2013).

69. Kulakovskiy, I. V. *et al.* HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).

70. Lis, M. & Walther, D. The orientation of transcription factor binding site motifs in gene promoter regions: does it matter? *BMC Genomics* **17**, 185 (2016).

71. Ho, J. E. *et al.* Biomarkers of cardiovascular stress and fibrosis in preclinical hypertrophic cardiomyopathy. *Open Heart* **4**, e000615 (2017).

72. Leoni, C. *et al.* Genotype-cardiac phenotype correlations in a large single-center cohort of patients affected by RASopathies: Clinical implications and literature review. *Am. J. Med. Genet. A.* **188**, 431–445 (2022).

73. Ho, C. Y. *et al.* Myocardial Fibrosis as an Early Manifestation of Hypertrophic Cardiomyopathy. *N. Engl. J. Med.* **363**, 552–563 (2010).

74. EU data protection rules. https://commission.europa.eu/law/law-topic/data-protection/eu-data-protection-rules_en.

75. Azure DRE. http://www.andrea-cloud.eu/azure-dre/.

Supplementary information

Supplementary Figures

Supplementary Table 1

Supplementary Table 2