



PROJECT FINAL REPORT

Grant Agreement number: 277936

Project acronym: META-PREDICT

Project title: Developing predictors of the health benefits of exercise for individuals

Funding Scheme: [HEALTH.2011.2.4.3-3 HEALTH.2011.2.4.2-2 HEALTH] Molecular and physiological effects of lifestyle factors on diabetes/obesity. Evaluation and validation studies of clinically useful biomarkers in prevention and management of cardiovascular disease Health.

Period covered: from 01/12/2011 to 01/07/2016 (incl. extension 6 months)

Name of the scientific representative of the project's co-ordinator¹,

Olav Rooyackers

Title and Organisation: Professor Karolinska Institutet, Stockholm, Sweden

Tel: +46 8 58586182

Fax: +46 8 7795424

E-mail: olav.rooyackers@ki.se

Project website address: www.metapredict.eu

¹ Usually the contact person of the coordinator as specified in Art. 8.1. of the Grant Agreement.

4.1 Final publishable summary report

4.1a - Executive Summary

Structured supervised exercise-training results in modulation of a variety of biomarkers for long term health. The response of each biomarker (e.g. aerobic capacity, insulin sensitivity etc.) is highly variable such that each subject demonstrates an extremely wide range of outcomes. The FP7 project '**Meta-predict**' has produced a new resource which will enable the discovery of the underlying molecular reasons for the variability in biomarkers for health and produce predictive tools for monitoring and tailoring exercise prescription to yield better outcomes with a particular focus on time-efficient exercise. The primary global molecular tool utilised was transcriptomics and this was applied to independent cohorts studying the impact of resistance (RT), endurance (ET) and high-intensity time-efficient (HIT) exercise training in sedentary over-weight adult men and women. The primary health biomarkers were changes in aerobic fitness and/or glucose tolerance (blood insulin area-under-the-curve) and secondary endpoints include changes in fasting insulin status (HOMA), blood pressure and body-composition in response to supervised training. 412 muscle/blood legacy samples were profiled on the U133+2 Gene-chip, while additional muscle/blood samples were profiled on 968 HTA 2.0 and 470 Affymetrix miRNA 2.0 gene-chips. Each gene-chip profile is accompanied by 10-to-200 biochemical and physiological phenotype measures, including targeted quantitative metabolomics (n=43) and ~50,000 insulin and c-peptide assays. We developed and implemented several novel informatics solutions for determining exon-usage (~310,000 protein coding exons per sample) and 'maps' of ~120,000 non-coding RNA features in each sample (lncRNA, 3'UTR and 5'UTR). Thus we have successfully created the largest and most comprehensive biological database of adult humans responding to supervised exercise training, including validating that the first practical and genuinely time-efficient (5-by-1min cycle sprints) exercise model for improving fitness and metabolism is comparable or superior to traditional time-consuming ET.

A novel linear feature (i.e. RNA or Metabolite) selection model was developed in R, and this allows for the identification of an ensemble of molecular markers, assayed in the pre-training sample, to predict changes in aerobic fitness and insulin 'action'. We are currently establishing if a non-responder for aerobic capacity with HIT training has the same molecular profile as a subject unable to improve aerobic capacity with traditional high-volume ET. We established that neither glucose tolerance nor HOMA could be predicted from simple physiological characteristics (e.g. BMI or VO₂max) and RNA and metabolomics signatures are being combined with BMI to provide a novel diagnostic of insulin resistance e.g. 'pre-diabetes'. A model of HOMA responses to HIT, ET and RT will be compared to establish if the ~40% of adults that show no improvement in insulin resistance with ET have the possibility to respond to either HIT or RT. This will lead to follow-up randomised clinical trials for this response-predictor. We made seminal observations for the field of human physiology and human genomics, including establishing that insulin secretion is directly and disproportionately down-regulated with HIT compared with insulin levels, while we have noted that >7,000 transcripts are regulated in people that improve VO₂max with HIT, including specific changes in the region of the transcript responsible for interactions with microRNAs and the translational machinery. 315 lncRNA (long non-coding RNA) were regulated in VO₂max responders, and a subset appeared to interact in a sense-antisense manner with a *cis* protein coding gene. Despite unforeseen delays totalling 1yr, additional high-profile publications, patent applications

and educational activities will emerge in the coming 1-2 years and this will be updated on the www.metapredict.eu website.

4.1b - A summary description of project context and objectives

Introduction

The general importance of physical activity has prompted guidelines from the US Surgeon General, American Heart Association, Centre for Disease Control, the UK Chief Medical Officer and Department of Health. These guidelines are **generic with only small variations provided for different age groups and some chronic-disease states**. The evidence basis includes several decade-long lifestyle interventions that have demonstrated an impressive reduction in Type 2 diabetes, e.g. DPS (38) DPP (7) and LookAhead (18) trials. In contrast, these studies found no reduction in cardiovascular disease burden (e.g. stroke and heart disease) and this limits the projected health-related cost savings that can be expected from intensive lifestyle programmes.

Indeed, numerous exercise-intervention studies have demonstrated that even under ideal conditions **one-size-fits-all** training programmes yield a dramatic range of physiological responses (3, 25, 26, 35). For example, ~25% people do not lower blood pressure (3) and ~25% do not gain muscle mass (25) during **supervised** exercise training. Indeed many people demonstrate one or more **adverse** (3) physiological outcomes while many more exhibit **negligible** gains (5, 15, 25, 33, 35). The extent of this range of responses is so dramatic that numerically speaking adverse responders are observed, where blood pressure **increases** (3), insulin sensitivity **declines** (29, 39) and muscle function can be found to be **impaired** (10). We believe that the variable and/or disappointing efficacy is largely related to one-size-fits all exercise/life-style programs producing both positive and adverse responses such that the net gain in “health” for each **individual** is unpredictable or suboptimal. **This was the raison d’être for meta-predict.**

The specific diagnosable mechanisms which explain the highly **variable** set of physiological responses to **standardised** lifestyle modification programmes and/or drug intervention is a developing field (13, 35). Selecting candidate genes from our earlier analysis (11), the Look-Ahead investigators found that genetic factors could be shown to modify the cardiovascular benefits of their **one-size-fits-all** lifestyle programme (24), while we have shown that gains in fitness are also influenced by genetics (35). RNA, metabolomic and epigenetic OMIC diagnostics are potentially more powerful (1, 12, 13, 35) than DNA based analysis, as they integrate both genomic and personal environmental influences on physiological health.

In meta-predict we focus predominately on global profiling of **coding** and **non-coding** RNA. While protein coding genes are highly conserved from worm to man, non-coding DNA distinguishes us even from our closest relative – and thus is considered to be critical for influencing phenotype. Practically, global RNA profiling is technically superior to proteomics. Further, it should be appreciated that analysis on RNA at the exon level (as we have done) allows us to examine alternative exon usage (with its impact on protein function) something that classic protein detection technology can not inform on. **As well as creating an “Exercise OMICS” bio-bank that contains > 0.5 Billion data-points (WP2,4,6,8,10) from almost 1000 human subjects, we have developed novel analytic methods for producing diagnostic/prognostic predictors (WP7/9), linked aspects of responder biology to/from human and out-bred rodent models and thus completed a research strategy that will yield the tools that will help personalise exercise prescription.**

We brought together leading EU and North American investigators (Figure 1) to study multi-modal exercise determinants of diabetes, obesity and cardiovascular disease risk factors. **The project deliverables will include new diagnostics that can define higher risk populations.** We will achieve this through the identification and validation of predictors that predict the nature of the health benefits of increased physical activity. This can help, for example, with optimisation of the prescription of the type and frequency of physical activity, for the treatment of e.g. insulin resistance.

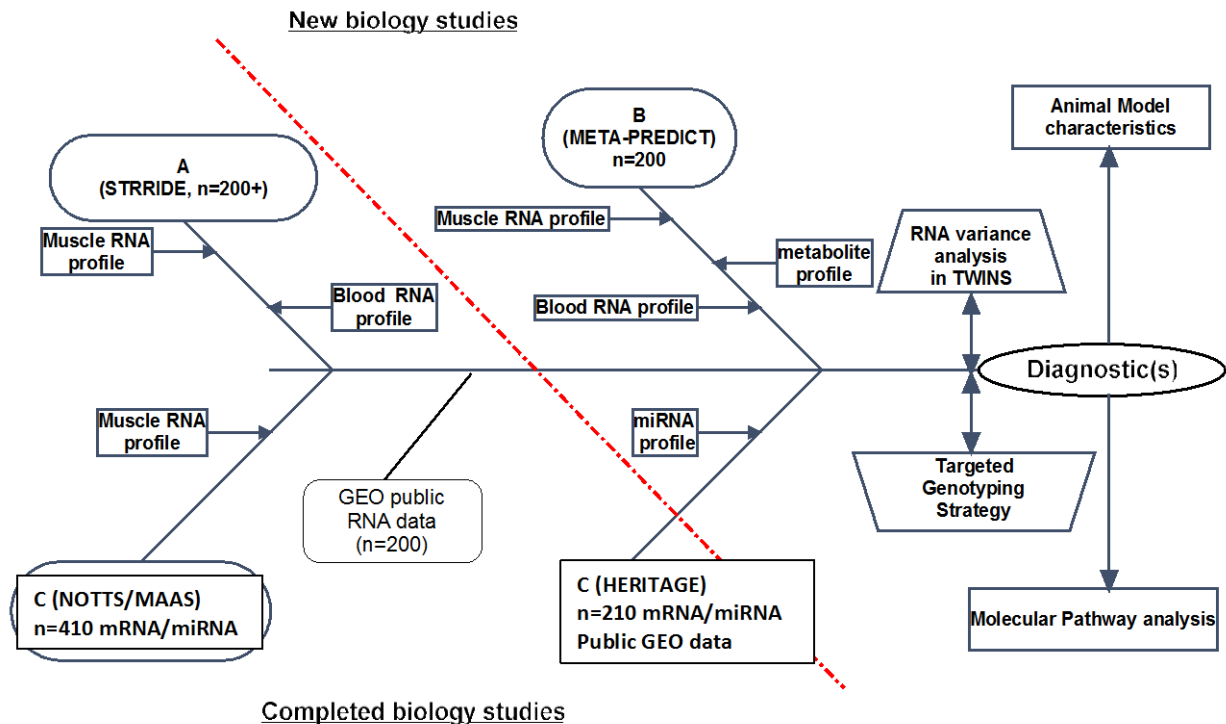


Figure 1. A high-level overview of the META-PREDICT. We have independent exercise life-style modification programs (A-D) with common clinical end-points that feed into a 'predictor' pipeline where we generate diagnostic screens from RNA, mDNA and metabolites. This will lead to personalised diagnostics, greater understanding of the underlying biology, and biomarkers for validating a novel drug screening in vivo model. Studies A and part of D have existing RNA gene-chip profiles; Study C has miRNA and mRNA deposited in the public domain (GEO). Study B will be the largest study of time-efficient cycle-sprint training in at-risk (metabolic disease) subjects studied using HTA 2.0 RNA, miRNA and mDNA chips.

Objectives

Our primary objective was to provide robust models that explain or diagnose the divergent effects of physical activity on metabolism using state-of-the-art strategies i.e. to carry out research that was genuinely integrated and was focused on being translational.

To achieve our objectives we had to carry out detailed OMICS profiling of materials obtained from several life-style intervention studies. Each of our clinical cohorts was utilised to produce predictors molecules for a diverse set of situations relevant to offsetting or improving insulin resistance or enhancing aerobic fitness in subjects at risk of Type 2 diabetes or insulin resistance. Our key objects were:

1. To produce robust nucleic acid and metabolomic data that can predict of the magnitude of the physiological responses to exercise related life-style modification using tissue and blood and to enable future selection of optimal exercise programmes.
2. To generate a biobank that combined diverse of studies exercise-training and measurement technologies so that the discovery, validation and implementation of each predictor could be as accurate and cost-effective as possible
3. To generate the data that allows for the study of biological networks and pathways linked to why >20% of the population are non-responders for the impact of exercise training on **insulin action** and propose potential drug-targets and offer alternative therapy.
4. To generate the data that allows for the study of biological networks and pathways linked to why >20% of the population are non-responders for the impact of exercise training on **aerobic fitness** and propose potential drug-targets and offer alternative therapy.
5. To utilise molecular data from **high-responders** to better understand the molecular determinants of successful life-style related gains in fitness (aerobic capacity) with the aim of being able to produce positive reinforcement tools (diagnostics or prognostics) for helping manage cardiovascular disease patient rehabilitation or disease prevention
6. To compare the **biological networks** and pathways from human studies and a new rodent model of 'exercise resistance' with the hope of validating the use of the model for future drug-screening and to generate basic understanding around the biology
7. Provide a large clinical data set characterising the impact of **time-efficient exercise** on metabolic and cardiovascular health parameters in adults at risk for Type II diabetes
8. Produce the largest **physiologically characterised biobank** (electronic and biological) of 'exercise biology' to facilitate the development of **future** technologies that aid the investigation and treatment of cardiovascular and metabolic disease.

4.1c - A description of the main S&T results/foreground

This part will summarise the main achievements and finding for the 10 individual work packages.

4.1c – WP 1 Management

WP1 conclusion: Despite several conflicts in the consortium, the strategy to focus on the scientific output has allowed us to produce the largest data set for omics changes to different types of exercise allowing us to develop predictors for responders and non-responders to training. A major effort has been made to obtain high quality data to allow for development of predictors with high specificity and sensitivity. The project has had major media interest including a BBC documentary (<http://www.bbc.com/news/health-17177251>).

Objective 1: To provide leadership, management and governance to the META-PREDICT consortium.

Coordination. During the whole project the coordination has been changed several times. The initial coordinator (Prof. James Timmons) was managing the project from the University of Birmingham first (months 1-7) and from Loughborough University after that (months 8-28). After this Prof. Timmons left his academic position and contributed to the project via the SME XRGenomics Ltd and the coordinator became Prof. Olav Rooyackers and Karolinska Institutet (KI) took over as coordinating university. In addition, XRGenomics LTD was voted in by the general assembly as the new partner representing Prof Timmons.

Consortium. During the project several partners left. Birmingham University (month 7) and Loughborough University (month 28) left due to a change in coordinator. McMasters University left the project in month 22 since they were not able to deliver the subjects for the HIT training that they initially promised to supply. Louisiana State University was voted out by the general assembly due to failing to declare a conflict of interest about the PI advising and collaborating with a commercial company in the USA aiming for commercializing the same IP as Meta-Predict. This latter process had to be redone due to procedural misunderstandings leading to the EU commission rejecting the initial exit request (month 28). A new general assembly vote and request for an amendment of the grant agreement was finally submitted in month 42 due to several administrative delays. In the same amendment request an extension of the project without extra budget for 6 months was applied for. Due to the severe delays in the process, the extra 6 months was essential to finish the scientific work. In addition, delays in recruitment of the subject for WP3, since many subjects screened were too active, and technical problems in the equipment for the metabolomics analyses (WP2) made the extension crucial. This final request was approved by the commission in month 52 leaving only 3 month to finish the work.

The exit of the Louisiana State University as a partner was decided by a majority vote in the general assembly, but some partners did not agree with the exit. This led to several conflicts within the consortium and throughout the whole process. Among others, this led to the change of the PI at one of the partners (MPI) with prof. Steen Knudsen leaving the project and thereby compromising the scientific work in WP7. The initial promise to share code for making predictors and applying this to the Meta-Predict cohorts fell away and the work had to be redone by another partner (XR Genomics). Despite these problems we have been able to finish up all the sample collection, sample analyses and most of the bio-informatics by changing the work logistics. For example smaller workshop meeting focusing on specific questions and specific work tasks such as data handling and bio-informatics were organized with dedicated persons working on these tasks.

Objective 2: To ensure that work-packages are complete and communication is optimal across the consortium

Communication within the project has been compromised by several conflicts as mentioned above. This is combination with the nature of the project has made the progress of the work packages a challenge. The project is build up with first the collection of clinical material, both from a new high intensity training study and several older cohorts. Only after collection of the clinical material, transcriptomic, genomic and metabolomics analyses could be initiated. Not until all analyses were performed, we could initiate the last part of the project integrating all the data with data organization and bio-informatic analyses. This logic build-up of the process, makes any delays in the early parts, delay the final work packages. Several delays were present during the whole project due to partners leaving, more difficult recruitment of subjects and technical difficulties. In addition, we underestimated the amount of work for the final parts of the project (data organization and bio-informatics) and thereby also under-budgeted this parts. Despite this we were able to perform the majority of the work in the different work packages. All the clinical data from the different cohorts are collected and all the samples are analyzed. A major part of the design of the predictors is based on transcriptomics. During the progress of the project we decided to use a newly developed chip (Affymetrix Human Transcriptome Array 2.0) for the transcriptomics analyses allowing us to include also coding and non-coding genes including RNA expression signal for individual exons. This yields ~250,000-400,000 exon measures per tissue sample, depending on the tissue type, i.e. 10x more data per sample for the same lab-cost. For the genomics we decided to analyze DNA methylation rather than SNP analyses due to time restrains and methylation being more under the influence of environmental changes and therefor closer related to mRNA changes. All the clinical data from all the cohorts has been collected in a database and is available for the predictor work. The bio-informatics part of the project has been challenged with the need to develop new code and the increase of the data due to the new chips. All codes are developed including code for making the transcriptomic data of much higher quality than ever before due to elimination of low and non-expressed genes, correction for GC rich genes and normalizing to a tissue specific back ground. This cleaned data will give much better quality of data and better a chance to detect predictors that will have high sensitivity and specificity. Preliminary results on predictors are presented in the relevant work packages.

Objective 3: To implement a quality assurance process so that the data are compatible with novel diagnostic development

To obtain predictors that will have high specificity and sensitivity both the omics data but also the clinical data has to be of the highest possible quality. Within the project lots of effort has been made to guarantee this and we have succeeded well in this. This means that we have a very large database from several large training cohorts with high quality omics and clinical/physiological data. To guarantee the quality several actions were undertaken in the different work packages.

Clinical data. For the HIT training study (WP3) 3 workshops were initiated to discuss and establish all the SOPs (standard operating procedure) for the training and the measurements. Two of these work shops were to ensure all research assistants (RA) at the different centers were involved in the process and to make sure all were fully aware of the SOPs and its content. One person travelled to all the centers to ensure that the procedures were followed. We implemented an interim analysis of the HIT training to ensure we would get physiological responses. At this interim meeting we decided to adjust the training protocol and were successful to obtain the responses needed for the predictors in a sufficient number of subjects. During the training study all data was collected and checked continuously using an on-line data system to ensure that all data was

actually collected and available. Within this we had both an electronic CRF (case report folder) and uploads of collected data files.

For several of the other cohorts (STRIDDE II and STRIDDE III) the insulin analyses were performed with old and difficult to validate assays. For these two cohorts we decided to rerun all baseline insulin analyses with the same method used in the HIT cohort to make this physiological outcome parameter of the same quality and comparable.

Analyses. For the metabolomics analyses new methods were established. We decided to include both internal and external standards to get real quantification of all the metabolites. The importance of using real concentrations is that if future metabolites are used for predicting, also other analytical methods measuring the quantity of the metabolite can be utilized. This might be an advantage especially when singular metabolites are used for the predictors and better assessable methods are preferred. An extended literature search was performed to choose the best and most applicable insulin assay (see WP2 for details).

Bio-informatics. As mentioned above new code and algorithms were developed and checked to improve the quality of the transcriptomics data. This complicated process took months of extra work. Many transcripts on the chips are not or minimally expressed in the tissue at hand and produce lots of noise with high variation. Since many of the statistical tools are based on variations analyses (e.g. principle component analyses, PCA) these genes interfere significantly with the results and are actually not relevant due to low or no expression. We therefore took the effort to clean up the data for these transcripts and also for genes not expressed in the tissue of interest. In addition, transcripts with a high G and C content have a better affinity and therefore higher values. We adapted an algorithm from Affymetrics to adjust for this. This has not been done before and this left us with very high quality transcriptomic data.

Objective 4: To ensure that communication with the EU is optimal and the project finances are managed appropriately

Finance during the latter half of the project was managed by Karolinska Institutet with a broad experience with financial management of EU grants. Some smaller adjustments in the budget were made to redirect budget to the more labor intense tasks that we underestimated in WP7 and WP9.

Objective 5: To ensure maximum scientific impact of the consortium research programme

Due to the nature of the project, the main scientific production is at the end of the project. Some papers on smaller subjects have been published during the project, but the majority of the papers containing the predictors and the integrated biology will be coming out in the next 2 years. Since all the data is collected and quality controlled at the highest possible level, this process will now be continuous.

Objective 6: To ensure the knowledge generated from the project is disseminated widely, to decision makers, health advisers, biotechnology investors and the scientific and public communities

Several aspects of the project have had major media interest including news articles in most European countries but also the US. The High Intensity Training has a potentially high impact on society since it requires a minimal investment of people's time (in our protocol about 15 minutes

per week) and significantly improves physiological parameters thought to prevent metabolic disease. Also the idea that some people do not respond to training has attained wide media interest. The BBC has dedicated a documentary specifically on these subjects and partners (prof. James Timmons) from Meta-Predict have contributed to this. Also a successful book written by Michael Mosley on High Intensity Training is introduced by Prof. James Timmons (<https://www.fast-exercises.com/>). Within the project we now show in the largest High Intensity Training study ever that the physiological benefits are several (improved physiological capacity, improved insulin sensitivity and even beneficial body composition changes) that the future dissemination will be large. The future plan is to further promote the beneficial effect of High Intensity Training further. In addition, once the predictors are established the plan is to commercialize these and broad collaboration with health care professionals and biotechnology investors will be sought.

In addition, new software (coding in the program R) for dealing with non-coding RNA informatics using HTA chips was made, published and made available to the scientific community. Several international researchers have shown interest in the iGEMS software (<http://nar.oxfordjournals.org/content/early/2016/04/19/nar.gkw263.full>). R code for microRNA chip data analyses has been developed and made available as well. (<https://github.com/iaingallagher/wCCS/blob/master/README.Rmd>).

We will use our website (www.metapredict.eu) to share not only the results of the project but also the produces SOPs and the iGEMS code.

4.1c – WP 2 Metabolomics

WP2 conclusion: In WP2 Metabolomics analyses was established to quantitative measure 43 metabolites and applied to the HIT study (WP3) and the rat studies (WP5). In addition this work package contained all centralized insulin and c-peptide analyses to guarantee comparable measures of insulin sensitivity among the different cohorts.

The main objective of WP2 was to establish the metabolomics analyses and apply this to the HIT study (WP3) mainly. In addition, metabolomics for the TWINN study (WP4) and the rat studies (WP5) have been added. Also the central analyses of the insulin and c-peptide levels for the HIT study and the re-analyses of insulin levels for the STRIDDE studies and the insulin analyses for the REHIT study have been part of WP2. All analyses, except the metabolomics for the TWINN study were performed at the Karolinska Institutet. Objective 1 was to establish the analyses and apply to the studies and objective 2 was to supply metabolomics input into WP7 (predictors) and 9 (biology).

Objective 1 was to establish the metabolomics analyses and apply to the studies

Metabolomics: For this a new HPLC and mass spectrometer (Ultimate 3000 UHPLC and TSQ Vantage, Thermo-Fisher) were purchased for the analyses of fatty acids and acylcarnitines (see also deliverable D2.2). Analysis for amino acids was already established on a HPLC (Alliance 2695 and fluorescence detector 474, Waters). The new analyses were established after a SOP from the Metabolomics Core of the Mayo Clinic. All these analyzes include internal standards and external standard curves in the appropriate range allowing for quantification of the metabolites analyzed. In addition, several metabolites (non-esterified free fatty acids (FFA), triacylglycerides, LDL- and HDL-cholesterol) were analyzed using standard analyses on an automated analyzer (Konelab

20XTi, Thermo-Fisher). For all metabolomics analyses quality control samples were included. For SOPs and methods see deliverables D2.3, D2.4 and D2.5.

Insulin/c-peptide: For this an extended literature search was performed to identify the most reliable method since variation of insulin concentration between the different assays is significant and unfortunately often ignored. Following this we choose to use two methods: an automated ELISA based analyses for all the samples taken during the OGTT test in the HIT study due to the large number of samples (about 20000). We were able to borrow a large capacity analyzer (Immulite 2000 Xpi, Siemens, Stockholm) from the manufacturer specific for these analyzes, which allowed us to analyze insulin and c-peptide in all samples semi-automated. All kits used for this were from the same batch. This analysis gives very good correlation with the gold standard method using mass spectrometry (19). However, the lower detection limit is often too high for baseline insulin levels. We therefore decided to use a high sensitivity manual ELISA kit (Dako, Stockholm, Sweden) for all baseline samples. The same kit was used for analyzing the insulin samples from the OGTT in the REHIT study and to re-analyze the baseline insulin levels of the STRIDDE studies. The latter was done to make the physiological outcome parameter for insulin sensitivity (HOMA) that will be used for developing the predictors comparable for the main studies. The other studies used the same kit for their baseline insulin analyses. Both methods were validated by comparison with the absolute gold standard for insulin analyses by measuring concentrations of insulin in WHO standard. Acceptable correlation with the WHO standard were obtained; $r^2 = 0.9995$ and $r^2 = 0.9923$ for Immulite 2000Xpi and Dako ELISA, respectively. Both methods correlated well for the WHO standards ($r^2 = 0.989$) and reasonable for the baseline samples from the HIT study ($r^2 = 0.654$).

Analyses. All samples were marked with barcodes and analyzed blinded for physiological outcomes for the person analyzing. All samples were randomized for analyses.

Objective 2 was to supply metabolomics input into WP7 (predictors) and 9 (biology).

Quality of analyses. For the metabolomics analyses plasma samples from healthy volunteers and critically ill patients from a previous study were pooled and run with every batch of samples. Most metabolites showed a coefficient of variation below 10% with a few exceptions for metabolites with low concentrations. The day to day variation for the QC was within 2xSD and showed no consistent change over time. For the insulin/c-peptide QC samples from the kit were run with the samples every day and showed a coefficient of variation over the whole analyses period of were between 5.5-6.6% for insulin and between 5.4.7 and 8.3% for c-peptide.

TWIN study (see also deliverable D2.6). Large scale metabolomics analyses has been performed on the TWIN cohort discordant for physical activity by the original group in Finland and published (16) (Figure 2). From all metabolites were the branched chain amino acids (leucine, isoleucine and valine) strongly correlated to the physical activity.

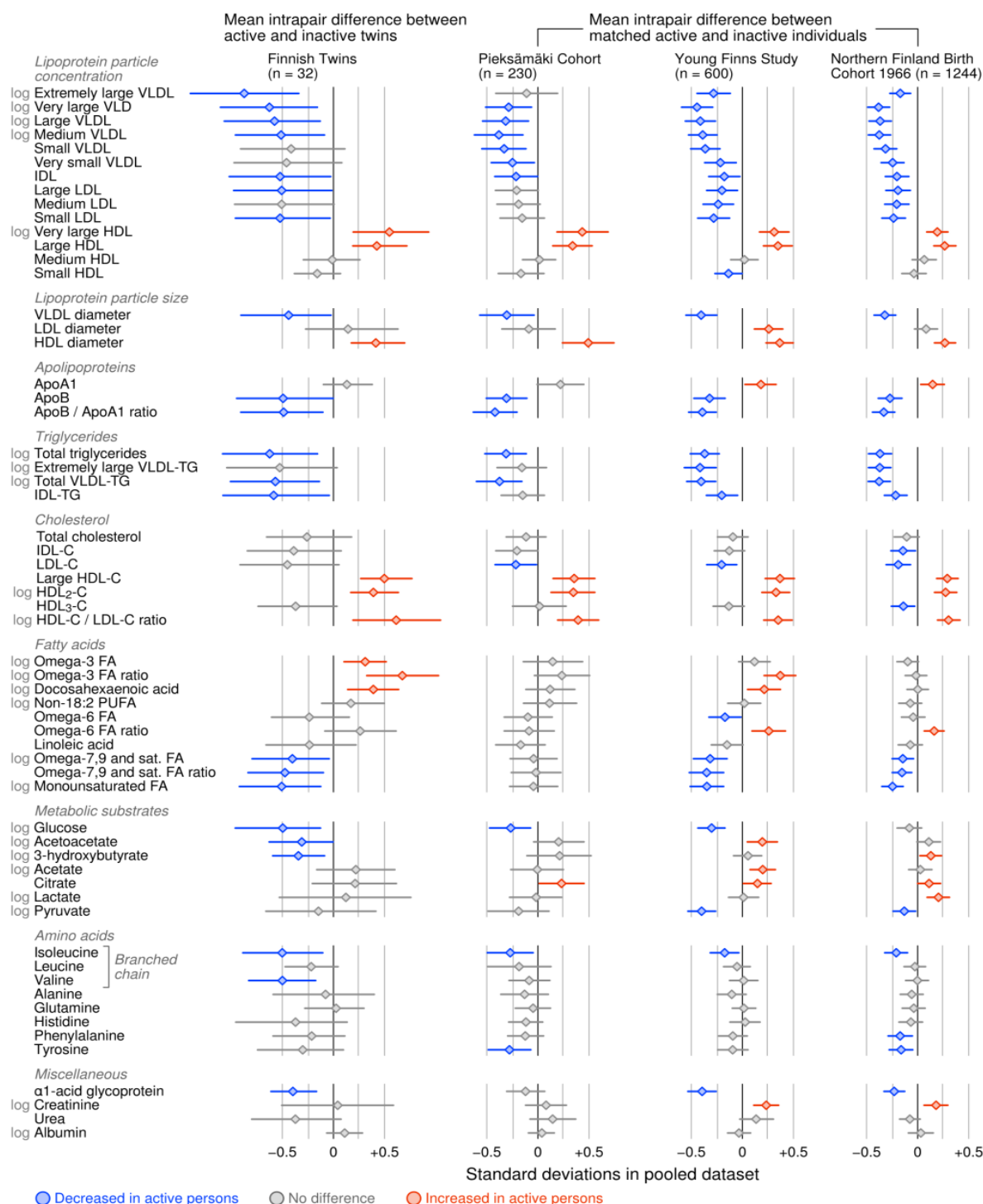


Figure 2. Metabolomics analyses in a smaller cohort of Twins with significant difference in level of physical activity (left) which was confirmed in larger cohorts of matched individuals with high and low levels of physical activity. From original publication (16).

High responder (HRT) and low responder (LRT) rat metabolomics (see also deliverable D2.9). For these analyses, plasma samples from 20 best responding HRTs (10 males, 10 females) and 20 least responding LRTs (10 males, 10 females) were analyzed. Many metabolomics difference were observed between male and female rats and therefore all the following analyses were done for the genders separately.

For the pre-training samples, two metabolite concentrations (asparagine and threonine) were significantly different between male HRTs and LRTs before HIT-training. In females, no significant differences were found.

The HIT training induced the most significant changes in branched-chain amino acid concentrations in the male rats. Between HRT and LRT males, the HIT-training induced changes were largely similar. In the females, differential changes were obvious, with the largest differences seen in acylcarnitines (Table 1).

Table 1. Interval-training induced concentration differences between HRT and LRT female rats. The concentrations for the metabolites are given as mean±SD in µM. TRP:tryptophane.

Metabolite(µM)	FemaleHRT		FemaleLRT		p-value	FDR
	PRE	POST	PRE	POST		
Lauroylcarnitine	0.04±0.02	0.02±0.01	0.03±0.02	0.02±0.01	3.97E-09	1.71E-07
Stearoylcarnitine	0.11±0.10	0.05±0.04	0.07±0.04	0.04±0.02	5.30E-08	1.14E-06
Octanoylcarnitine	0.02±0.01	0.02±0.01	0.02±0.01	0.03±0.02	1.23E-07	1.77E-06
Myristoylcarnitine	0.12±0.09	0.04±0.03	0.10±0.07	0.04±0.02	7.69E-07	8.26E-06
Hexanoylcarnitine	0.05±0.02	0.05±0.02	0.05±0.03	0.09±0.06	1.27E-05	0.0001094
TRP	107.4±9.7	98.8±9.5	102.5±15.1	98.4±13.1	0.0016521	0.01184

HIT study (see also deliverable D2.8). Insulin sensitivity before and after the HIT training was assessed using an oral glucose tolerance (OGTT). The area under the curve (AUC) for insulin and c-peptide were significantly lower following the 6 weeks HIT training. A striking relationship between the relative reduction in insulin AUC and C-peptide AUC was observed (Figure 3). However, the molar reduction in C-peptide observed during the OGTT was ~3 times greater than the observed reduction in insulin AUC. This indicates that insulin secretion was reduced following training more so than insulin concentration, and therefore insulin clearance was likely reduced.

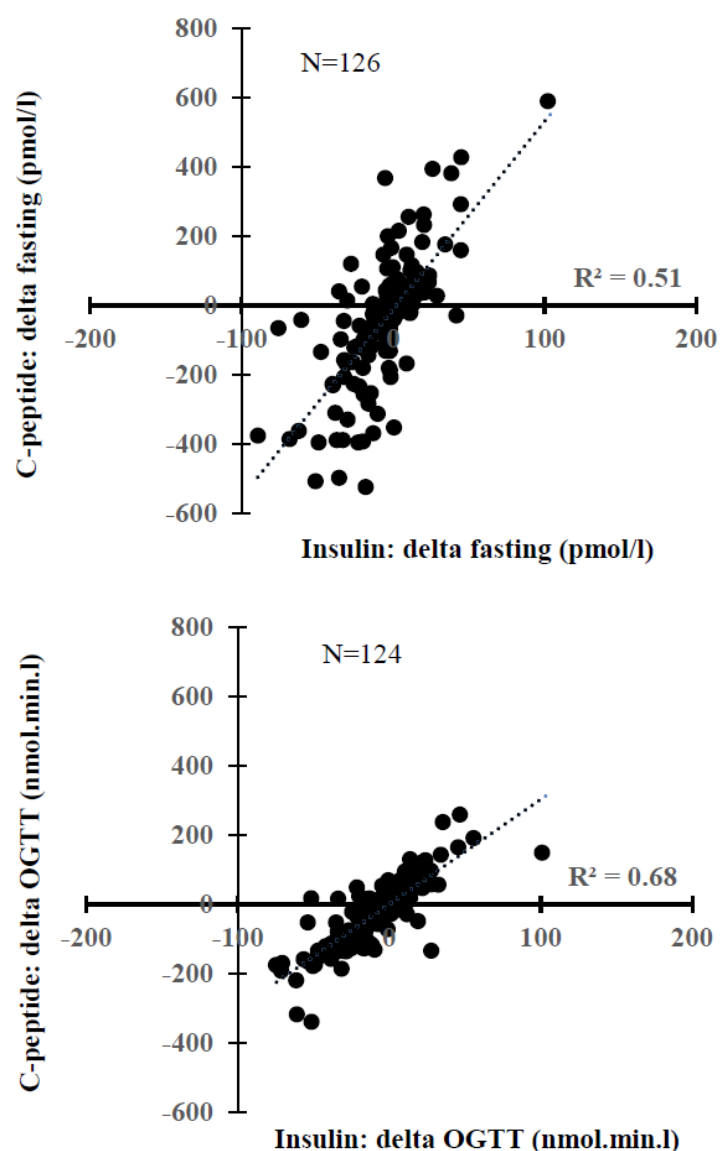


Figure 3: Human Studies - Presentation of the linear relationship between changes in plasma insulin and plasma C-peptide concentrations following 6 weeks of '5-by-1' HIT protocol. Oral glucose tolerance test (OGTT). Value is based on Pearson correlation coefficients (coefficient of determination).

For the effect of HIT training on the metabolomics, only minor significant changes were observed. HDL increased due to the 6 weeks HIT training, 2 of the acylcarnitines increased and glutamate decreased (see deliverable D2.8 for details and numbers).

Principle component analysis (PCA) of the metabolomics data shows that the metabolites do not cluster according to center, but do cluster into different metabolic clusters. These data show that there is biological variation in the metabolomics data, making it more likely to find predictors for physiological responses to training on its own or in combination with transcriptomics in WP7.

4.1c – WP 3 HIT study

WP3 conclusion: In work package 3 we have recruited 223 subjects among 5 centers and trained with 2 high intensity protocols. The shorter and more intense protocol (5 times 1 minute; 3 times a week) significantly improved exercise capacity, insulin sensitivity and clearance, lean body mass and mean arterial pressure. We also showed a novel finding in that insulin clearance is affected by the training.

Objective 1 To employ appropriate staff to conduct the clinical experiments and carry out the medical procedures.

Each clinical test center recruited and employed appropriate staff (RA) in accordance to the national/local rules (see also deliverable D3.2).

Objective 2 To set-up identical exercise and glucose equipment, experimental procedures and analytical methods across the 4 (+ MAC) individual clinical test centers.

The same testing/training equipment (Corival, Lode, Groningen, the Netherlands) were purchased to all test centers. Standard operation processes (SOP) were composed for all the procedures, tests and analyses by all the test centers PIs (see also deliverable D3.3 and D2.3). The final accepted SOPs were documented on the locked Meta-predict file server (Zoho Creator). To ensure that all RA (from the different test centers) understood and followed the SOP in exactly the same way, a one day obligatory training meeting were organized (16/9 2012 Stockholm). During this meeting all the protocols were discussed to identify any site-specific problems. To follow up any aspect of the training program, analytic methods or experimental procedures two meeting were conducted during the clinical trial period (RA work-shop on MP research components in Loughborough 4-5/1 2012 and 3-4/9 2013 Copenhagen Denmark).

To further optimize “identical” experimental procedures between the centers, a test person travelled to all test centers and performed key testing for the training program to ensure an acceptable inter-reliability between the centers. Throughout the study an electronic CRF (case report folder) was filled out on-line and all experimental and data files were continuously uploaded to a file server (Zoho Creator) to ensure correct documentation/archiving of the raw data. The server was locked for any changes except for the coordinator (deliverable D3.4).

Objective 3 To successfully recruit and screen 240 sedentary and overweight male and female volunteers up to the age of 50 years for inclusion into a time-efficient exercise training study (known as the META-PREDICT cohort).

Seven-hundred and seventeen people were screened to recruit 223 participants across 5 geographical regions in Europe. Recruitment commenced in 2012 and the last participant completed in 2014 (see also deliverables D3.1, D3.5 and D3.6). Participants were classified as sedentary (<600 METs (metabolic equivalents) min⁻¹·wk⁻¹) using a modified International Physical Activity Questionnaire, and had a fasting blood glucose value consistent with World Health Organization (WHO) criteria for impaired glucose tolerance (IGT; >5.5<7.0 mmol·l⁻¹), and/or a BMI >27 kg·m⁻². Exclusion criteria included evidence of treatment of cardiovascular, respiratory, gastrointestinal, renal disease, history of malignancy, coagulation dysfunction, musculoskeletal or neurological disorders, recent steroid or hormone replacement therapy, or any condition requiring long-term drug prescriptions. All participants gave their written, informed consent to participate (see also deliverables D3.1). Many of the screened subjects had an activity that was too high (>600 METs) and this delayed the recruitment and the finalization of the WP by several months.

Objective 4 To complete a 6 week, high intensity aerobic training programme in 200 of these volunteers.

The study complied with the 2008 Declaration of Helsinki and was approved by the relevant ethics committees for the University of Nottingham (D8122011 BMS), Karolinska Institutet (2012/753-31/2), the University of Copenhagen (H- 3-2012-024), the University of Las Palmas de Gran Canaria

(CEIH-2012-02) and Loughborough University (12/EM/0223). The study design including time points for blood and skeletal muscle sampling and experimental procedures, OGTT, DEXA, VO₂ and physical activity monitoring (see also deliverables D3.1, D3.3 and D3.4). Forty-six participants were recruited and 40 completed the baseline visits and HIT Protocol 1 (Table 2).

Table 2. Subject characteristics for the HIT study

	Non-exercise (n=13)	'7-by-1' HIT (n=40)	'5-by-1' HIT (n=137)
Gender (men/women)	4 / 9	20 / 20	63 / 74
Age (y)	25 (20-51)	38 (20-53)	38 (18-51)
Height (m)	1.66 (1.52-1.81)	1.71 (1.53-1.94)	1.72 (1.51-2.01)
BMI (kg·m⁻²)	32.7 (27.3-41.4)	29.8 (27.0-45.5)	31.2 (26.5-48.1)
Fasting glucose (mmol·L⁻¹)	4.6 (4.0-5.2)	4.5 (3.8-5.4)	4.6 (3.3-5.6)
2hr OGTT glucose (mmol·L⁻¹)	6.3 (5.2-9.4)	6.5 (4.2-10.5)	6.8 (3.7-10.1)
IPAQ Score	248 (160-1188)	382 (73-594)	330 (0-597)

This HIT protocol was developed from the literature and consisted of three supervised cycling sessions per week for six weeks at ~100% of the work-load required to elicit $\dot{V}O_2$ max (based on a pre-training 25-W stepwise protocol), with 7 bouts and 1 min recovery between each bout. Interim analysis found limited change in $\dot{V}O_2$ max (see Results below); therefore, this protocol was discontinued. A further 164 participants were screened for HIT Protocol 2 with 137 completing the entire study (Table 2). Compared to Protocol 1, the volume of exercise training was reduced while the intensity was increased. All sessions began with a 2-min warm-up at 50 W followed by five 1 min high-intensity cycling bouts with 90 s recovery between each bout (3 bouts performed in week 1). The intensity, determined in session 1 of week 1, started at 85% of the work required to elicit $\dot{V}O_2$ max (W_{max}), and increased by 10% until the participant was unable to complete a 1-min bout (training load was 121±11% of pre-training $\dot{V}O_2$ max).

The initial training load did not correlate with changes in any primary or secondary outcome. Training load for each subsequent session was at the highest completed 1 min bout from the previous session and a further 10% increase was implemented after 2 weeks. Thirteen participants were allocated at random to serve as a non-training comparison group for the study (Table 2). These participants underwent all screening and assessment procedures but did not participate in any exercise sessions.

Objective 5 To quantify (implement algorithms) the impact of the 6 week training programme on maximal oxygen consumption and insulin sensitivity (using duplicate aerobic capacity and baseline oral glucose tolerance and intra-venous glucose tolerance tests).

There were no significant differences in baseline characteristics between the participants in the three groups. Protocol 1 ('7-by-1') yielded a modest improvement in $\dot{V}O_{2\max}$ (+5.8%, $P<0.0001$) (Figure 2) and BMI, but not in MAP, HOMA-IR nor plasma insulin AUC. Protocol 2 ('5-by-1') increased $\dot{V}O_{2\max}$ to a significantly greater extent (+10.0%, $P<1E-23$) (Figure 4) and upper-leg lean mass (+3.0%; $P<1E-19$) and reduced mean arterial pressure (-3.2%; $P<0.0001$) and HOMA-IR (-6.8%, $P=0.03$). Changes in OGTT-derived insulin AUC (-9.1%, $P=0.0004$) and C-peptide (-4.6%, $P=0.006$) strongly co-varied ($R^2=0.68$; $P<0.0001$) yet C-peptide was reduced ~3 times more than insulin. Responses were similar in women and men; they were substantially maintained for 3 weeks after training. There were examples of individuals that demonstrated gains in $\dot{V}O_{2\max}$, reduced MAP, HOMA-IR and OGTT insulin AUC while there were also several examples of poor/adverse responses for each health biomarker, including rare examples of individuals that demonstrated no gain in $\dot{V}O_{2\max}$, increased MAP and increased HOMA-IR and OGTT insulin AUC values (Figure 5). (see also deliverable D3.7)

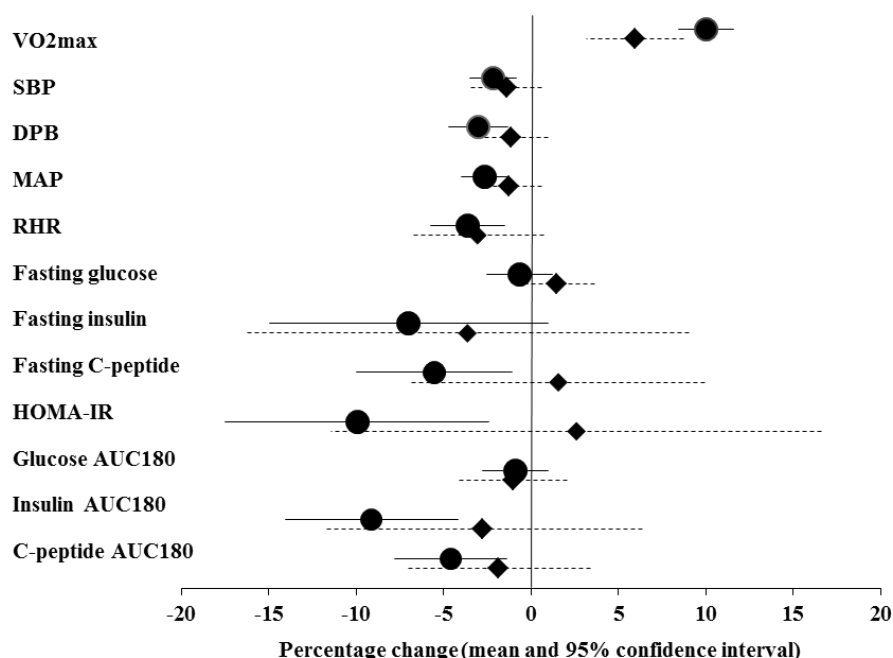


Figure 4: Comparison of the efficacy of 6 weeks of the '5-by-1' HIT high intensity cycle based exercise training protocol (black circles) with the less effective '7-by-1' HIT (black diamonds) protocol in obese or over-weight adults. The values are presented as mean $\pm$ 95% CI for the main training responses. AUC: area under the curve during oral glucose tolerance testing (180 min); SBP: Supine systolic blood pressure; DBP: Supine diastolic blood pressure; MAP: Supine mean arterial pressure; RHR: resting heart rate.

Objective 6 obtain muscle biopsy samples and body composition measurements (DEXA) pre and post training for RNA profiling (WP8), miRNA profiling (WP6) DNA gene sequencing (WP10) (and produce a future metabolomics and DNA epigenetic analysis biobank).

Muscle biopsy samples and body composition measurements (DEXA) were done at time points indicated in Figure 6 for the WP7-10). After administering lidocaine, a muscle biopsy was taken from vastus lateralis using the Bergstrom needle technique. Body composition data was obtained using a dual energy x-ray absorptiometry (DXA) scan and analysed using enCORE software (Lunar, GE Healthcare, Bucks, UK).

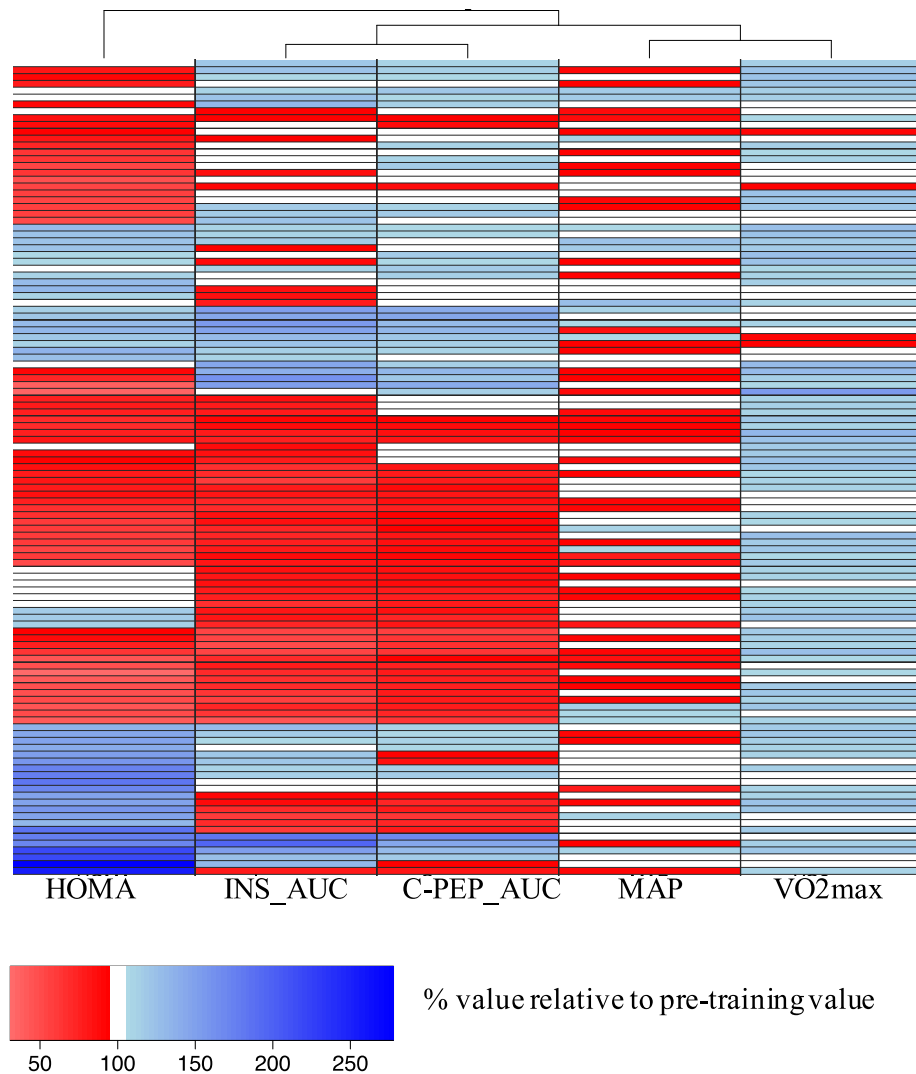


Figure 5: Heat-map representation of the individual responses to exercise training for 5 variables following ‘5-by-1’ HIT ($\dot{V}O_{2max}$, MAP, Oral glucose tolerance test (OGTT) Insulin AUC₁₈₀ and HOMA-IR). It can be observed that each subject demonstrates a differential response to the 5 clinical parameters. Some subjects improve all five, most improve 2 or 3 while some do not improve any of these health biomarkers.

Objective G To produce blood and plasma samples, pre and post-training for metabolomics analysis (WP2).

Venous blood samples were then taken from a retrograde venous cannula (dorsal hand vein) with the hand situated in a heated unit for metabolomics analysis at time points indicated and used in accordance to WP2.

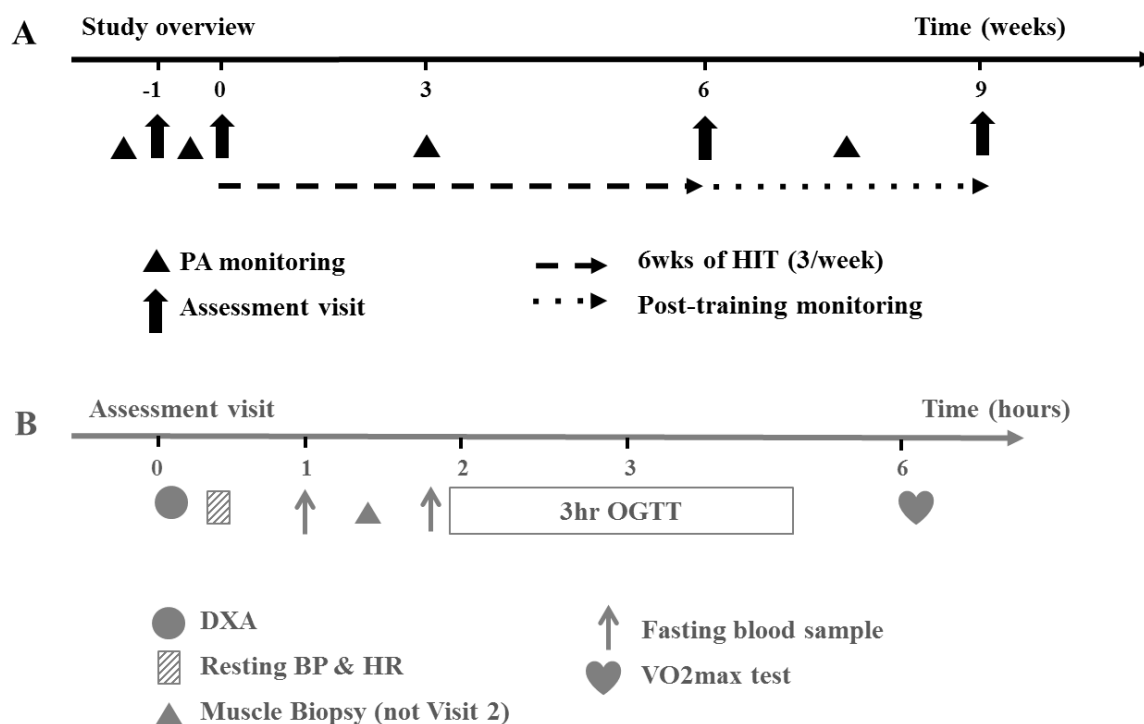


Figure 6. META-PREDICT HIT study design. Six weeks of supervised HIT was preceded by two study visits (baseline measurements, one week apart) and followed by another two study visits, post-training (~3 days after final session) and post-training monitoring (3 weeks after the final session). After an overnight fast, a DXA scan was followed by two fasted blood samples and a muscle biopsy, a standard 75-g OGTT, a standardised meal and a $\dot{V}O_2$ max-test. The DXA scan and muscle biopsies were not performed during Study visit 2.

4.1c – WP 4 TWINN studies

WP4 conclusion: *In this work package monozygotic TWIN discordant or concordant for physical activity were identified from larger databases and phenotyped in more detail and sampled for transcriptomic and metabolomics profiling to determine the genetic component of the predictors. The active twins had higher fitness level, lower percentage body fat and better insulin sensitivity compared to their inactive co-twins.*

Objective 1 (see also deliverable 4.5): Identify 60 monozygotic twin pairs that demonstrate either similar or divergent levels of obesity or physical activity.

Young adult twin pairs discordant or concordant for physical activity were identified by a follow-up survey for the Finnish *FinnTwin16* twin cohort (for details see deliverable D4.2). We did a follow-up survey for the population-based cohort study among 3383 Finnish twin individuals (1578 men) from five birth cohorts (1975-1979), who earlier had answered questionnaires at mean age of 24.4 y (SD 0.9), and the current survey was at mean age of 33.9 y (SD 1.2). Waist circumference was self-measured at each time-point. Secondly, from the *FinnTwin16* cohort we identified pairs concordant or discordant for their physical activity habits on the basis of the questionnaires and a telephone interview, who then participated in detailed two day clinical examinations plus biopsy visit in Jyväskylä (*FITFATTWIN* study). In addition, obesity discordant and concordant pairs have been studied in Helsinki using partly identical procedures.

From the *FinnTwin16* cohort data we have reported data on the associations between changes in physical activity and waist circumference during the 10-year period (Figure 7). Decreased activity was linked to greater waist gain compared to increased activity (3.6 cm, $P < 0.001$ for men; 3.1 cm, $P < 0.001$ for women). Among same-sex activity discordant twin pairs, twins who decreased activity gained an average 2.8 cm (95% CI 0.4 to 5.1, $P = 0.009$) more waist than their co-twins who increased activity ($n = 85$ pairs); among MZ twin pairs ($n = 43$), the difference was 4.2 cm (95% CI 1.2 to 7.2, $P = 0.008$). So, among young adults, an increase in leisure-time physical activity or staying active during a decade of follow-up was associated with less waist gain, but any decrease in activity level, regardless baseline activity, led to waist gain that was similar to that associated with being persistently inactive. Also more detailed report on which type of sports participation predicts low waist gain, is in manuscript phase.

From the *FITFATTWIN* study we reported the main data on the monozygotic male twin pairs who were most discordant for leisure-time physical activity. The main aim was to investigate how physical activity level is associated with body composition, glucose homeostasis, and brain morphology. According to pairwise analysis, the active twins had higher fitness level, lower body fat% ($P = 0.029$) and HOMA index ($P = 0.031$), and higher Matsuda index ($P = 0.021$) compared to their inactive co-twins. Striatal and prefrontal cortex (sub-gyral and inferior frontal gyrus) brain gray matter volumes were larger in the non-dominant hemisphere in active twins compared to inactive co-twins with a statistical threshold of $P < 0.001$. So, among healthy adult male twins in their mid-thirties, a greater level of physical activity is associated with improved glucose homeostasis and modulation of striatum and prefrontal cortex gray matter volume independent of genetic background. The findings may contribute to later reduced risk of type 2 diabetes and mobility limitations.

Another finding was that in particular intraperitoneal fat was reduced as a consequence of physical activity (in depth analysis of intra-abdominal fat accumulation measured from MR images). Maybe more importantly the precognitive sensory brain functions seem to be associated with the level of physical activity. Physical activity history (3-yr-LTMET), physiological measures and somatosensory mismatch response (sMMR) in EEG were recorded in 32 young healthy twins. In all participants, 3-yr-LTMET correlated negatively with body fat%, $r = -0.77$ and positively with VO_{2max} , $r = 0.82$. Trend toward larger sMMR was seen in inactive compared to active participants. This finding was significant in a pairwise comparison of 9 monozygotic twin pairs discordant for physical activity. Larger sMMR reflecting stronger synchronous neural activity may reveal diminished gating of precognitive somatosensory information in physically inactive healthy young men compared to the active ones possibly rendering them more vulnerable to somatosensory distractions from their surroundings (34). Overall, the brain has an important role related to physical activity motivation and ability to exercise.

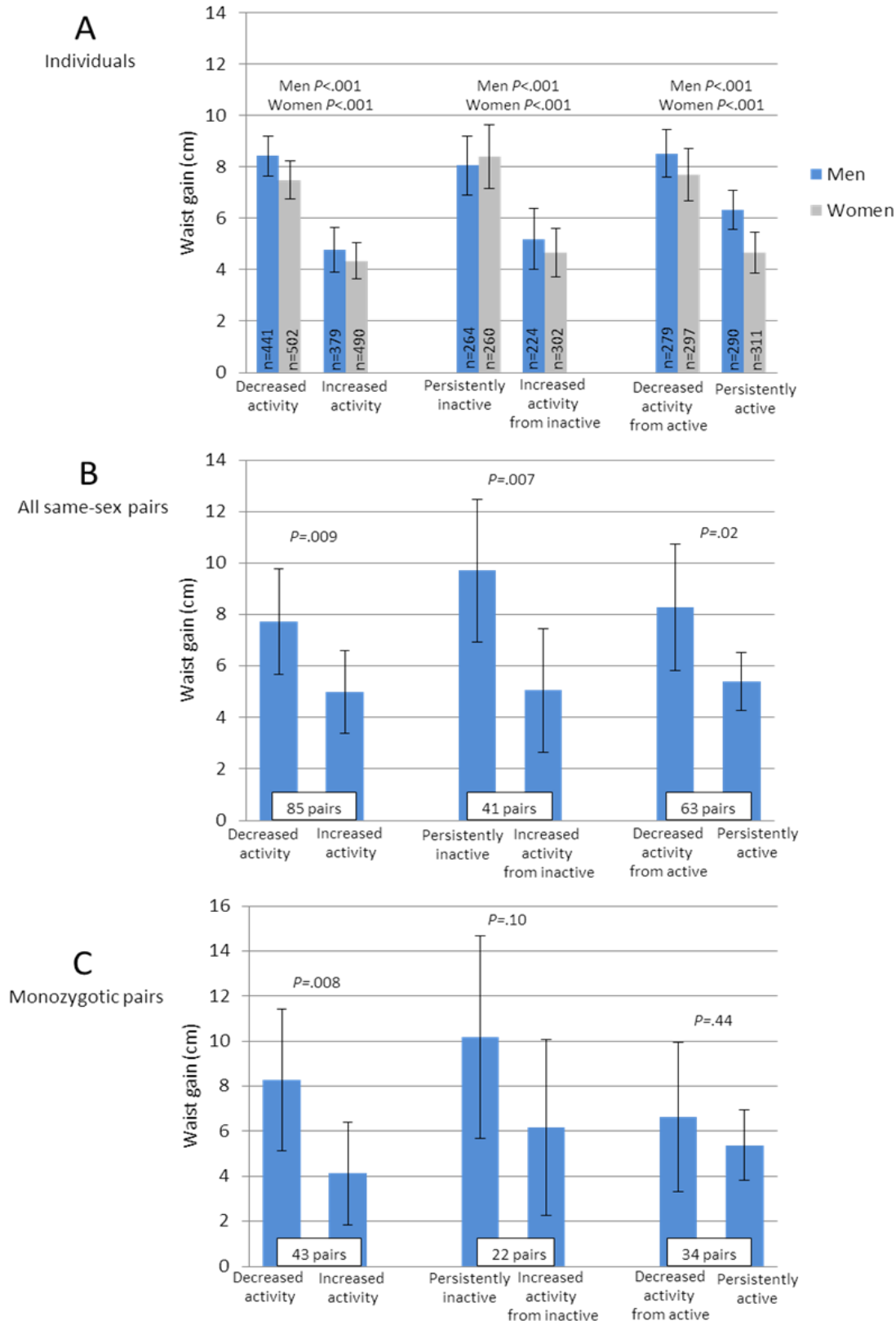


Figure 7. Differences in waist gain (cm, mean and 95%CI) during follow-up (decreased activity: changed from upper third to a lower one; increased activity: changed from a lower third to upper one). (A) Sex-specific differences among individuals taking into account clustered observation of twin pairs. (B) Pairwise difference among leisure-time physical activity discordant same-sex twin pairs. (C) Pairwise differences among leisure-time physical activity discordant monozygotic twin pairs (from (28)).

Objective 2: A) Blood and skeletal muscle sampling B) Integrating of data.

A (see deliverable D4.3, D4.4 and D8.2): The twin blood, muscle and fat samples have been used for the development of an integrated model for identification of alternative exon usage events (31) and the exon data will be used for developing predictors of training-related improvements in fitness and metabolism (see WP 7 and 9).

B: Multi-dimensional investigation of twin material was also used to study the effects of physical activity to support the Meta-predict aims as we do not yet know exactly which mechanisms are most strongly related to the prediction of responses to exercise. In addition to young adult twins we also used the older Finnish twin cohort with a clinical sub-study to long-term physical activity discordant twins to produce data to support the aims of the Meta-predict consortium. This includes reports on the association between dyspnea score (as fitness indicator) and as well as metabolomics study among physical activity discordant twin pairs and individuals identified from other population cohorts.

The combination of metabolite differences (including 134 metabolic measures and ratios) between active members of twin pairs and their inactive co-twins was investigated by permutation analysis with two test statistics: the internal setting constructed within the twin data showed a P-value of 0.003 and the external validation using a population-based (global) metabolic profile as an independent reference yielded a P-value of 0.006. The pairwise analysis among twin pair was repeated using matched pairs with long-term discordance in physical activity from other population cohorts. The pairwise comparisons showed many differences in the metabolomics measures consistent in different datasets (see also deliverable D2.6).

4.1c – WP 5 Rat studies

WP5 conclusion: Low and high responder rats were successfully phenotyped and samples were analyzed for metabolomics and microRNA. Metabolomics showed changes in acylcarnitines due to HIT training which is similar in the humans (WP2). MicroRNA pathway analyses showed TGF β related pathways including SMAD to be affected also similar to our preliminary human results (WP9).

The general objective in WP5 was to complete the characterization of our novel “high and low responder-to-exercise” aerobic model to provide a screening tool for drug development against exercise resistance. It included 4 tasks/objectives: 1-2: High and low responder rat phenotype description pre and post exercise training and metabolic phenotype of HRT and LRT rats respect to high fat feeding. Task 3-4: Provide samples for checking consistency with the human high versus low responder profiles and discovery of conceivable drug/exercise therapy targets

Objective 1: The effects of training on the physiology in the HRT and LRT rats .

Four individual experiments were performed to complete basic physiological characterization of High Responder and Low Responder (HRT/LRT) rodent model(14)(14), and their responses to HIT and other training regimens. Rats were 6-8 months old at the start of the experiments. Normally these rats were phenotyped for their inherited training response at the University of Michigan. Phenotyping training on a treadmill starts at the age of 11 weeks, and training protocol starts at speed 10 m/min (~30-40% of the maximal running capacity), lasting 20 min. Phenotyping includes three training sessions a week, for 8 weeks, with gradually increasing speed and duration (being finally 20 m/min for 30 min). Response to training is characterized by a speed ramped treadmill

test before and after the training. However, to characterize the animal model in full, we used both unphenotyped and phenotyped rats, and we also phenotyped one group of adult rats in Finland.

Results (also see deliverables D5.2 and D5.5): Body composition: Fat% was higher in HRT males than in HRT females before the HIT-intervention ($p < 0.001$). HIT decreased fat% in HRT males, thus the gender difference in body fat did not exist after the intervention. In LRTs there was no gender difference in fat% either before or after the intervention. HIT-training increased lean mass more in LRT females than in LRT males ($p = 0.03$) contrary to HRTs ($p = 0.065$). In general, HIT was more effective decreasing fat% in HRT than in LRT ($p = 0.019$). HIT as such had no effect on weight gain neither in male nor female HRTs or LRTs.

Gain in aerobic running capacity: In HIT-training the training load was individually adjusted, and both rat lines (HRT and LRT) responded to it. In unphenotyped male rats - in both rat lines - HIT-training resulted in an increase of ~5m/min in speed, in relative terms the increase was 29% in HRT and 26% in LRT. In phenotyped HRT rats, the response was also ~5 m/min, corresponding ~18% increase in both HRT males and females. In LRT rats, females gained less than males: the increase was ~2 m/min, 5%, in females and ~2.5 m/min, 10%, in males. In addition, HIT increased running capacity more in female HRTs than in female LRTs ($p < 0.05$).

Spontaneous activity: In general, there seemed to be no ratline differences in the activity of control animals. HIT decreased active time in comparison with controls (main effect of HIT: $p = 0.007$). However, the intensity of activity (activity index during active time) appeared to be highest in HIT-trained HRTs and lowest in HIT-trained LRTs (HIT-HRT vs. HIT-LRT, $p = 0.046$).

Fasting blood insulin and glucose, glucose tolerance and insulin sensitivity: HIT-intervention had no effect on fasting blood glucose and insulin concentrations. In the 120 min glucose tolerance test, HIT-training diminished blood glucose concentrations at time points +30 min and +120 min in HRT female rats. However, no significant effect of HIT was seen in the calculated glucose tolerance. HIT had no significant effect on insulin sensitivity.

Blood lipids and metabolites: HIT did not affect total blood cholesterol concentrations. In HIT-trained LRT females HDL level was lower than in HIT-trained HRT females (2.0 vs. 2.5, $p = 0.029$). In LDL no effects of training were detected. Triglyceride level differed between males and females in sedentary LRT rats ($p = 0.01$), and in LRT males HIT decreased triglyceride level ($p = 0.016$). In HIT-trained LRT males glycerol levels were lower than in HIT-trained LRT females ($p = 0.021$), and the similar trend was found in HRTs ($p = 0.064$).

Angiogenesis and fibre typing: the LRT rats had an impaired angiogenesis response to 6 weeks of treadmill training. In addition the LRT rats had a 7% Type 1 fibre whereas the HRT had >20% (17).

Objective 2: The metabolic phenotype of HRT and LRT rats - the effects of the high fat diet in the HRT and LRT rats (see also deliverable 5.6).

The aim of this experiment was to find out if female HRT and LRT rats respond differently to long-term (8-10 weeks) high-fat diet. Before the high-fat diet intervention, LRT rats were heavier having more lean mass and fat than HRT animals ($p < 0.001$ in all parameters). High-fat diet induced similar increases in body weight both in LRT and HRT groups. No great differences in the changes of body composition were found. After the high-fat diet intervention VO₂ (oxygen consumption) was

measured for two days. During the second day of measurement, HRTs had significantly higher VO₂ than LRTs when calculated either per total body mass ($p < 0.017$) or per lean body mass ($p < 0.050$). No differences were found in respiratory exchange ratio (RER) or total activity between the groups. Prior to the high-fat diet intervention, HRT rats had higher fasting blood glucose concentration than LRTs (101.0 ± 1.2 and 92.6 ± 1.2 mg/dL (mean $\pm$ SE), respectively, $p < 0.001$). After the high-fat diet intervention, no difference in the blood glucose was seen between the groups. However, diet induced a significant increase in resting blood glucose concentration in both groups to 121.6 ± 2.2 mg/dL in HRTs and 119.4 ± 3.7 mg/dL in LRTs. No difference in blood glucose concentrations were seen during the insulin tolerance test except at 15 min when HRTs had significantly lower blood glucose ($p < 0.01$).

Objective 3-4 Provide samples for second species verification.

Metabolomics: In order to clarify metabolic differences between HRT and LRT strains, a predefined set of metabolites were analyzed in serum of male and female before and after HIT-training (see also deliverable D5.4 and D2.9). Initial analysis showed that there was a difference between males and females, thus the genders were separately assessed. Two metabolite concentrations (asparagine and threonine) were significantly different between male HRTs and LRTs before HIT-training. In females no significant differences were found. Before training, concentrations of several metabolites were different between male and female LRTs, α -aminobutyrate (AABA) showing the greatest difference.

Also in HRT males and females some metabolites were showing different concentrations. Interestingly, HIT-training induced different profile changes in the metabolome of males and females. In males, the most significant changes were seen in branched-chain amino acid concentrations. Between HRT and LRT males, the HIT-training induced changes were largely similar; between females differential changes were obvious with the largest differences in acylcarnitines (see WP2).

Molecular measures: microRNA chips have been run for the high fat diet fed rats (for details see deliverable D9.2). SLC35E4 is down-regulated which is interesting as it is a transporter of carbohydrate and amino-acid of unknown specificity and thus may play a role in substrate metabolism that has yet to be fully understood.

Myofibrillar, sarcoplasmic, mitochondrial, and collagen protein synthesis were measured from HRT and LRT gastrocnemius muscle. In general, female rats tended to have significantly higher collagen and mitochondrial protein turnover rates than males ($p = 0.072$ and 0.098 , respectively), while HIT-training as such did not affect protein turnover rates. In order to screen acute high intensity exercise-induced responses and activation of the signaling pathways regulating glucose and lipid metabolism, angiogenesis, and cell survival and growth, western blotting was used to semi-quantify protein phosphorylation at key residues of signaling intermediates (such as PI3K-Akt, PGC1 α , AMPK, mTOR and MAPK pathways) in aim later to bench mark the model with human data (see also deliverable D5.7). HRT and LRT rats performed one exercise bout with HIT protocol. Immediately after HIT exercise, rats were necropsied and gastrocnemius muscles were collected for analyses. Non-exercising HRT and LRT rats served as controls. Protein analyses focused on molecular mediators underpinning presumable differential insulin sensitivity between the rat lines. Statistically significant increases in HRT compared to LRT were found in p-AKT, p-p70S6K1, p-p38MAPK, p-JNK and p-CaMKII, which are activated forms of these signaling proteins (Figure 8). Significant increases in HRT were found also in HIF-1 α and PPAR γ . These results provide new

information of the skeletal muscle pathways regulating physical activity-induced adaptive processes.

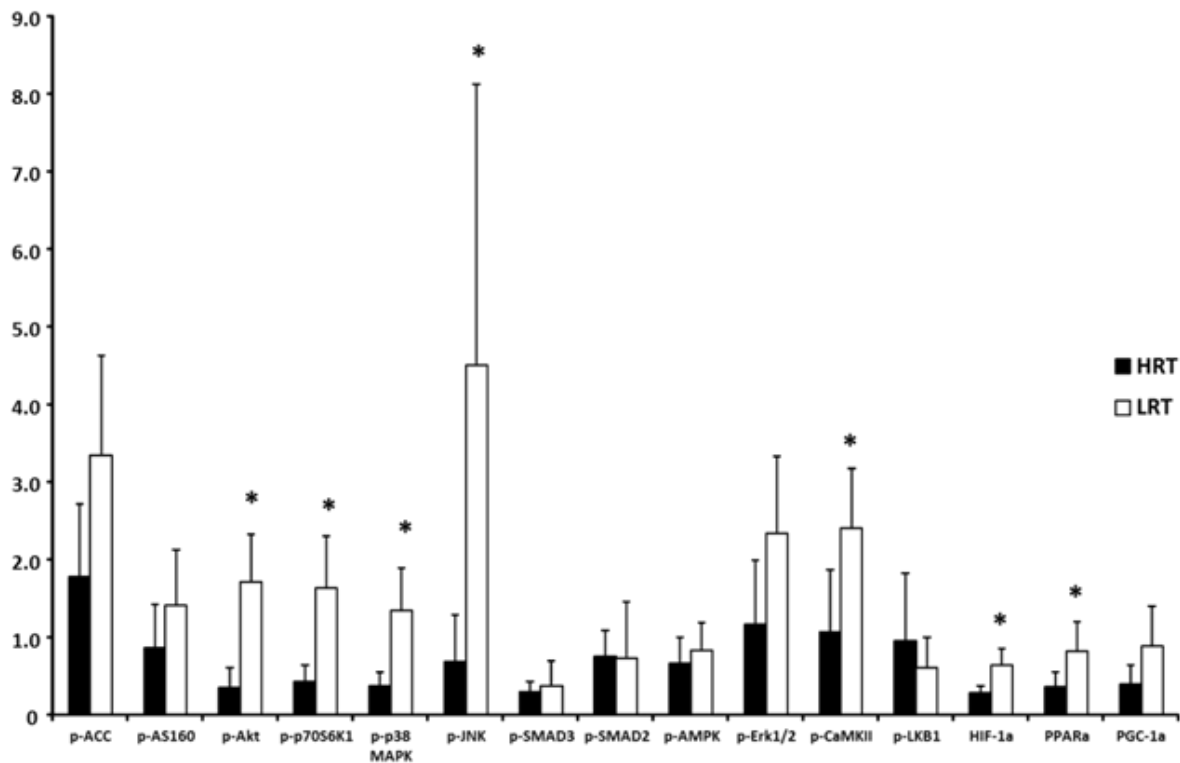


Figure 8. Content of selected signaling proteins [mean (SD)] in skeletal muscle (m.gastrocnemius) of high and low responder rats following acute high intensity exercise. * = statistically significant difference between the rat lines.

4.1c – WP 6 MicroRNA

WP6 conclusion: The clinical samples and OMIC data were successfully processed. Due to time constraints (WP7/9) only the basic informatics analysis has been complete to confirm validity of raw data. Data will be integrated into future analyses as appropriate.

MicroRNAs are post-transcriptional regulators of protein synthesis. In the absence of global proteomics (clinical biopsy samples are typically too small) they provide a link between mRNA and protein production changes. We have developed an approach known as the weighted cumulative context score (wCCS, (8)) method to predict which protein pathways are targeted by microRNA changes in vivo. At the start of the project we evaluated both Exiqon and Affymetrix miRNA chip technology in both rat (2012/13) and human muscle (2011-2015) samples. While there is no gold-standard platform and both these platform yield distinct results (Figures 9 and 10) we progressed with the Affymetrix miRNA 2.0 chip as they were higher throughput, more reproducible and more cost effective.

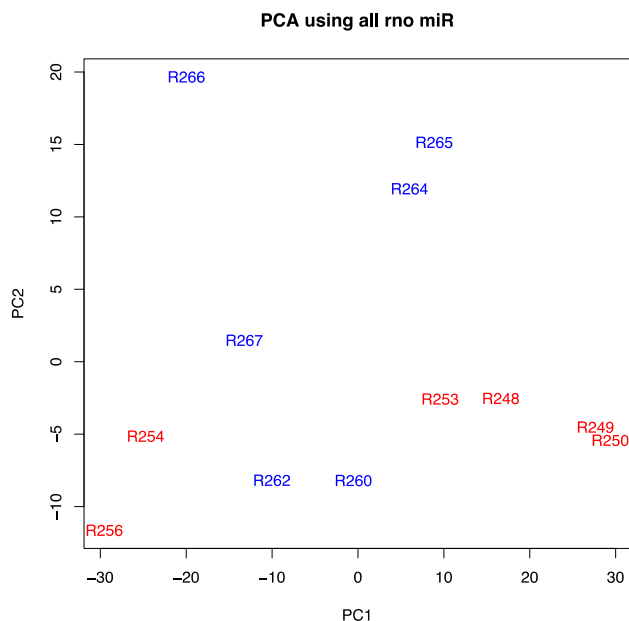


Figure 9. The impact WP5 D5.3 experiment on Rat miRNA PCA (Affymetrix)

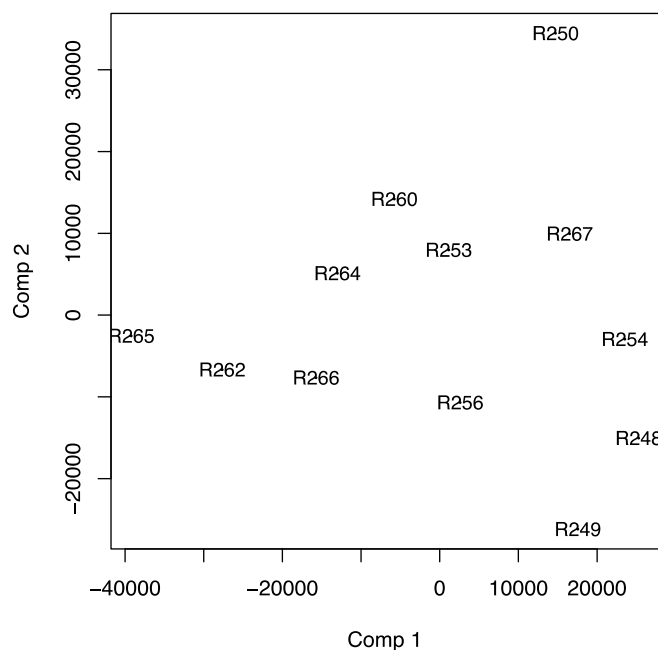


Figure 10. The impact WP5 D5.3 experiment on Rat miRNA PCA (Exiqon)

A subset of subjects undergoing supervised classic aerobic training that alters insulin resistance and aerobic fitness (n=740) to a highly variable extent were profiled using the Affymetrix miRNA 2.0 chip. A total of 110 chips were run from HERITAGE (Figure 11) using 400ng RNA input. Initial QC analysis has been carried out indicating that 2 subjects RNA failed QC and hence 4 chips were to be removed from down-stream analysis (WP7/9).

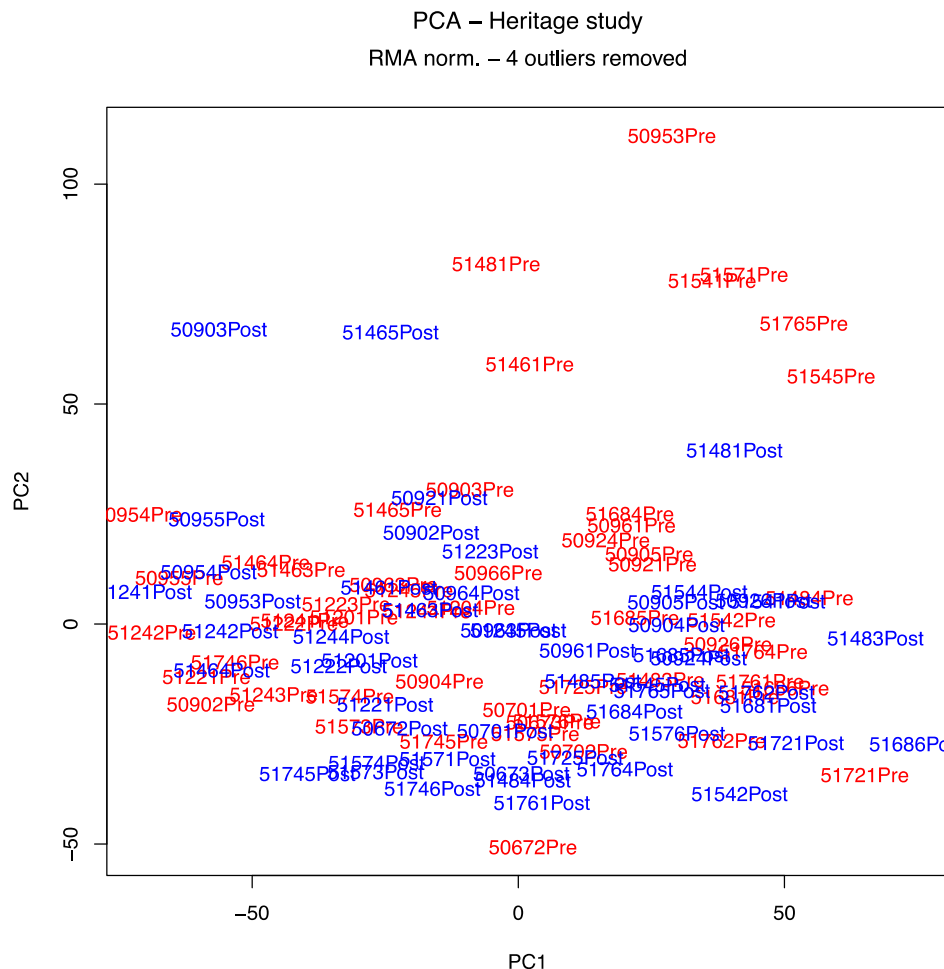


Figure 11 - Affymetrix miRNA array 2.0 PCA plot – Heritage Samples

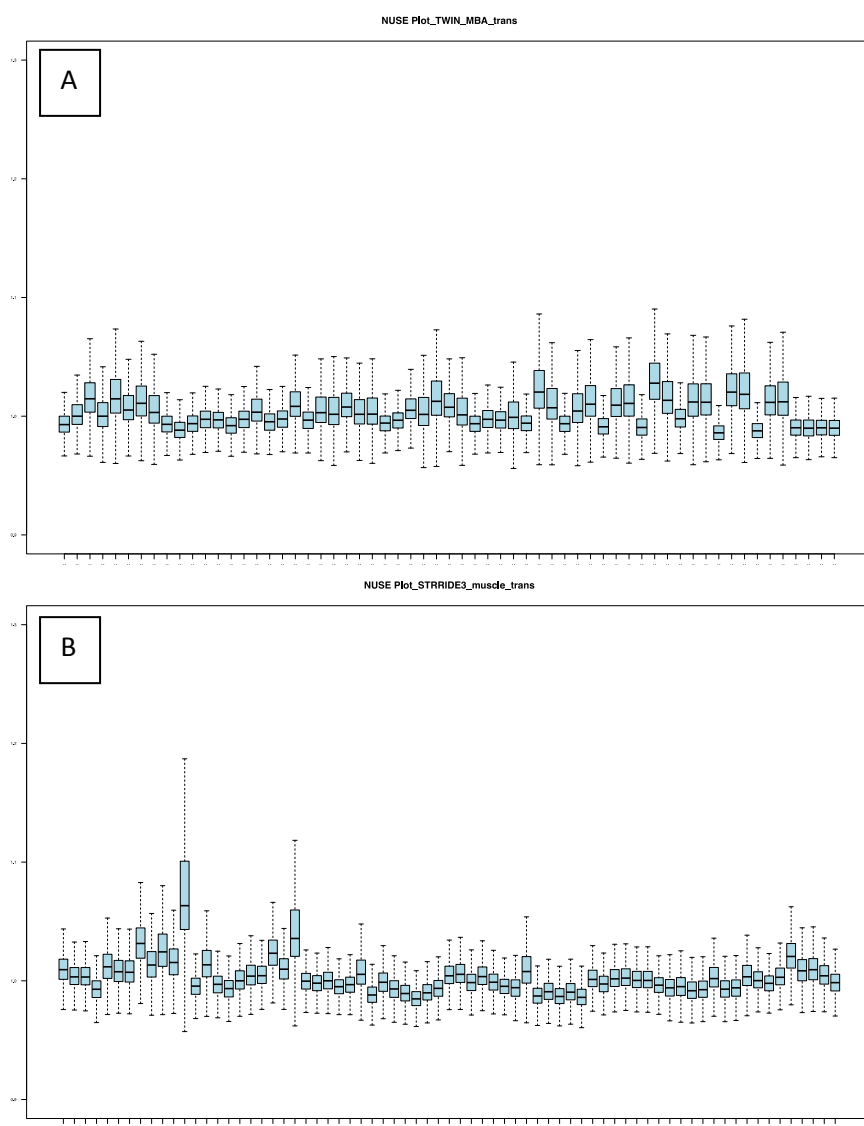
As well as the n=110 miRNA chips profiled from HERITAGE muscle (Deliverable D6.2), we produced n=85 muscle RNA samples from Derby resistance training study (as per Figure 1 of Part-B of the original grant application) and n=86 muscle from resistance training (van Loon MAAS samples - See Figure 1 in Part B of the application) and n=189 from the MP HIT study (including a subset of n=118 that were pre/post samples)(Deliverable D6.3). Because we also used the more extensive Affymetrix Human Transcriptome Array 2.0 with blood and muscle RNA, which provides extensive data on **coding** and **non-coding** genes, we did fewer miRNA chips than original planned.

A total of n=470 miRNA chips were run. For miRNA network biology (WP9) we have developed a method known as wCCS (weighted cumulative context score) in 2010 (8). The R-code for this analysis method has been further developed into a beta-R package to enable faster down-stream analysis in WP9. The code is awaiting final QC prior to being released as R-package publication but is available via GitHub (<https://github.com/iaingallagher/wCCS/blob/master/README.Rmd>). Pathway analyses using this software on rat microRNA can be found in Deliverable D9.2.

4.1c – WP 8 Transcriptomics (WP7 activity came sequentially next – see below)

WP8 conclusion: Due to the additional inputs from Professors Kraus (Duke University, USA) and van Loon (Maastricht University, NL), as well as a successful WP3 we more than met our target for providing tissue and blood samples for RNA processing. From this point, ~90% of the samples were successfully turned into QC passed gene-chip profiles.

For WP8 RNA was processed from blood and muscle samples and HTA 2.0 GeneChip data using a TRIzol (Invitrogen) reagent based protocol modified to limit interphase contamination, followed by an alcohol re-precipitation step (for numbers see deliverables (D8.1, D8.2, D8.3, D8.4 and D8.5). The RNA pellet is re-suspended in non-DEPC treated water and RNA concentration is determined using a Nanodrop spectrophotometer (LabTech International, UK). QC was determined in WP7/9 (**NUSE** plots, see examples Muscle NUSE data below, Figure 12) using the chip-data (most robust approach: better than RIN). Sample ID was cross-checked with gender-specific RNA markers. From ~1030 HTA 2.0 Chips run, 6 were disregarded for this reason while the rest were rejected based on QC (reflecting RNA quality).



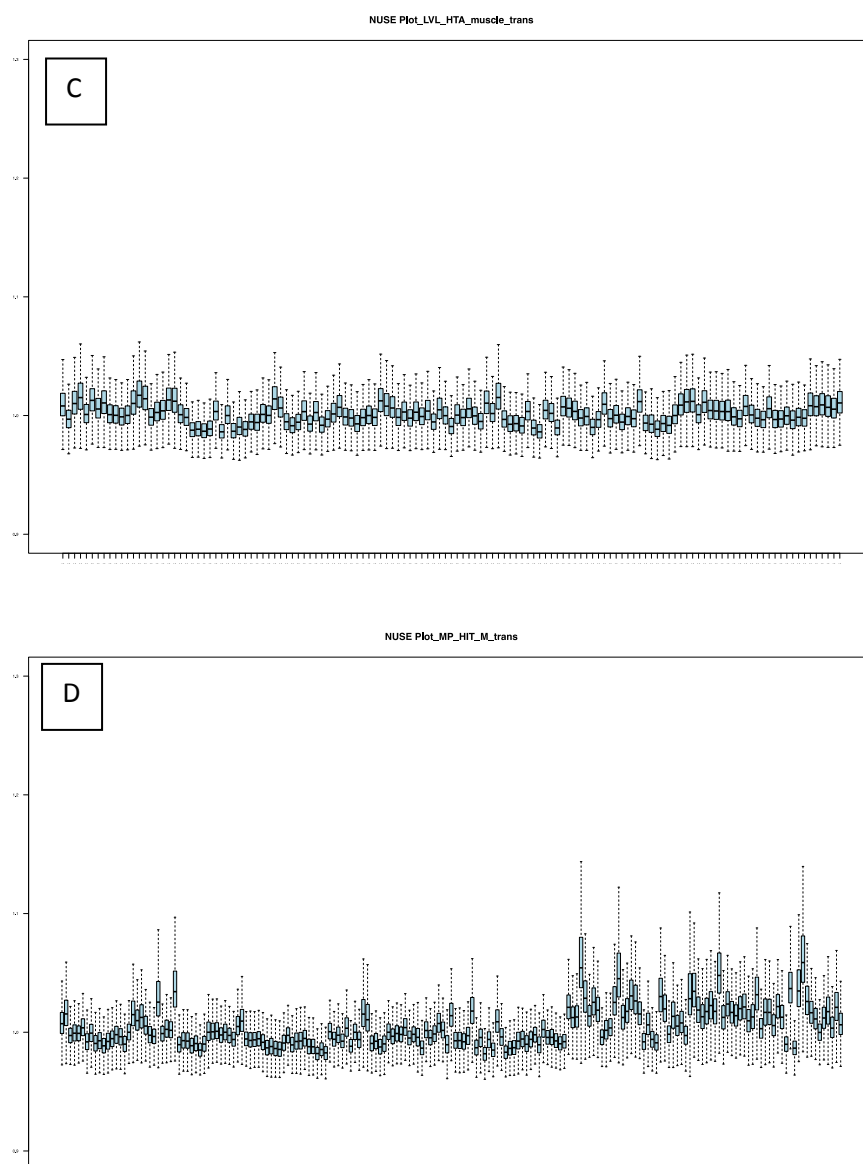


Figure 12. NUSE plots for muscle samples from A) the TWINN study, B) STRIDDE III, C) LVL Maastricht study and D) HIT study.

4.1c – WP 7 Predictors (utilised WP8 raw data)

WP7 conclusion: *The MPI Bioinformatician left the project prior to the initiation of the RNA informatics work and so the work-package was carried out by XRGenomics LTD resulting in a ~18 month delay until computer code could be developed and validated. XRGenomics LTD successfully delivered novel R code for building linear classification models, novel data-processing methods for the HTA 2.0 Chip (See WP9) and some prototype models. The majority of outputs for WP7 were pre-deemed CO (Confidential) and confidential data has been deposited with the EU commission in the Deliverable reports. What follows in the non-confidential data only.*

MPI had developed computational code over the previous decade, suitable for processing the OMIC and clinical data into predictor linear models of clinical response. When their bio-informatician left the project in November 2013, this task was taken over by another partner. While XRGenomics LTD was experienced at writing and implementing **non-linear** predictor

informatics (e.g. (30)) they now had to develop and write computational code for producing **linear** classification models. Thus, the linear code for the project was therefore produced, tested and adapted in several stages.

The first stage following curation of the clinical data into an Access database and quality controlling the values by two independent scientists, was examination of the relationship between insulin and clinical variables (to established potential co-variates for the OMIC models). For fasting insulin homeostasis there were some modestly correlating clinical variables (e.g. Vo2max, Age and BMI). This U133+2 Microarray data has been QC and deposited online (<https://www.ebi.ac.uk/arrayexpress/files/E-GEOD-47969/E-GEOD-47969.idf.txt>).

Using the pre-existing DUKE muscle U133+2 chip data (n=134) and our previously published aerobic data-sets (n=41) we developed a novel muscle based predictor of aerobic adaptation (December 2014). This new set of 144 ps will be used to evaluate the new Meta-predict related samples. For example, using new (WP8) DUKE blood chip data (n=124 baseline from 289) we will take this aerobic predictor signature genes and attempt to reproduce this classifier from the blood RNA profile (figure 13).

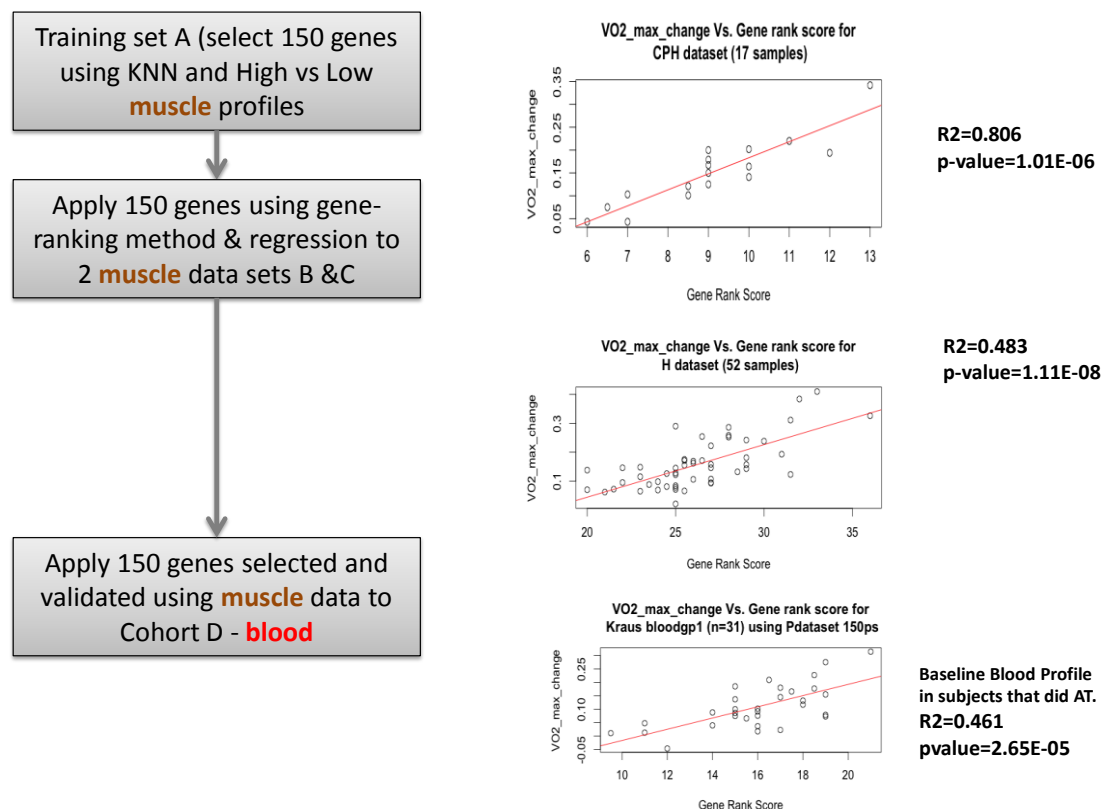


Figure 13. Using new (WP8) DUKE blood chip data (n=124 baseline from 289) we took the muscle based aerobic predictor signature genes and attempt to reproduce this classifier from the blood RNA profile.

The benefit of moving a muscle based RNA classifier over to blood, is that we have demonstrated a greater degree of independent validation in first showing a good performance using muscle based RNA profiles. RNA expression variance is related to genetic variance and thus it is also expected that shared patterns will be identified. Thus WP7 relies on high quality transcriptomic data and metabolomics data (WP2) to refine existing predictors (e.g. aerobic fitness) or produce

novel (insulin sensitivity) predictors of human responses to life-style modification (supervised exercise programmes).

RNA has been extracted from the samples and analyses using the Affymetrix Human Transcriptome Array 2.0 (see WP8). This has delivered RNA expression of 389,798 Exons across the three tissues using the XRGenomics custom CDF software. The first task is to extract reliable RNA expression data from the chip. Most labs will use the standard pipelines to do this but that data is highly flawed. The probes are wrongly annotated and many millions of probes do not provide a signal in a given tissue.

The Affymetrix Human Transcriptome Array 2.0 - making muscle specific data

There are many components of RNA that define the transcript. There are exons, splice-variants, 3' untranslated region and 5' untranslated region. Our R code is ~800 lines that update the annotation of each of the 6.9 Million probes on the chip, determine if the probe signal is above background and assign the probe to a probe-set (designed to reflect Exons, Transcripts, Noncode-Genes or UTR's)(Figures 14-17).

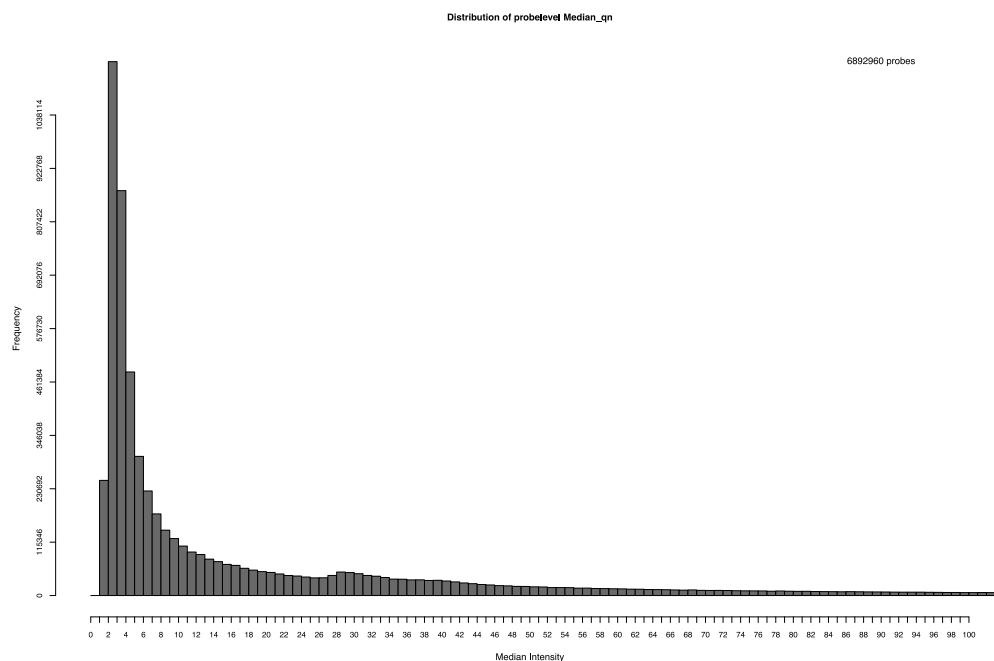


Figure 14. Demonstrating the large peak of background probe signal seen in the 'raw' HTA 2.0 chip data in MP muscle samples.

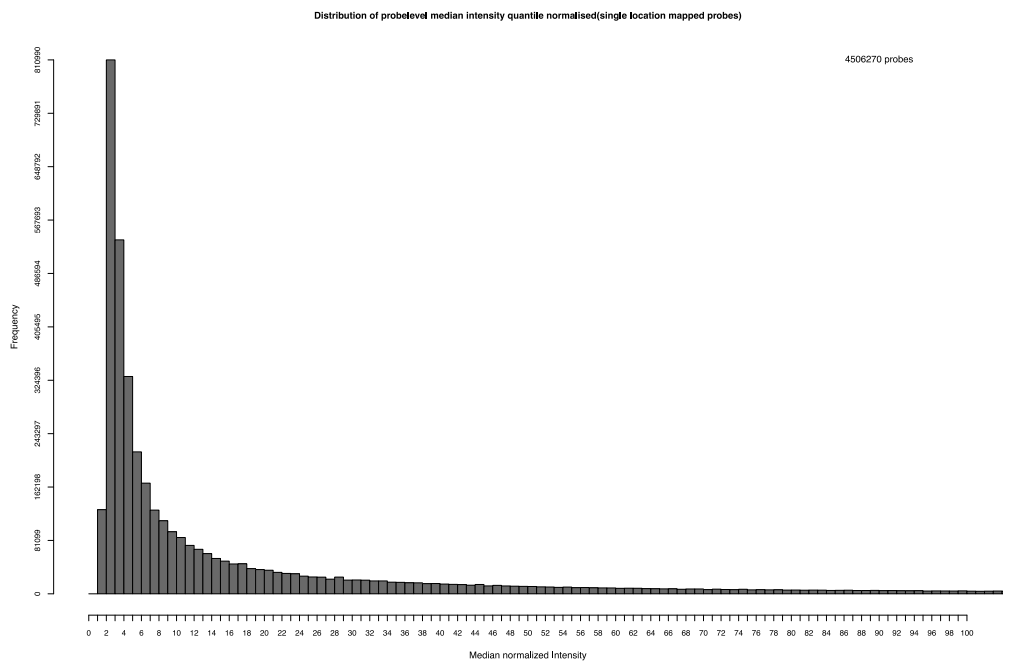


Figure 15. Demonstrating the impact of re-alignment of the original HTA 2.0 probes to the latest genome sequence - ~2.4Million probes are removed.

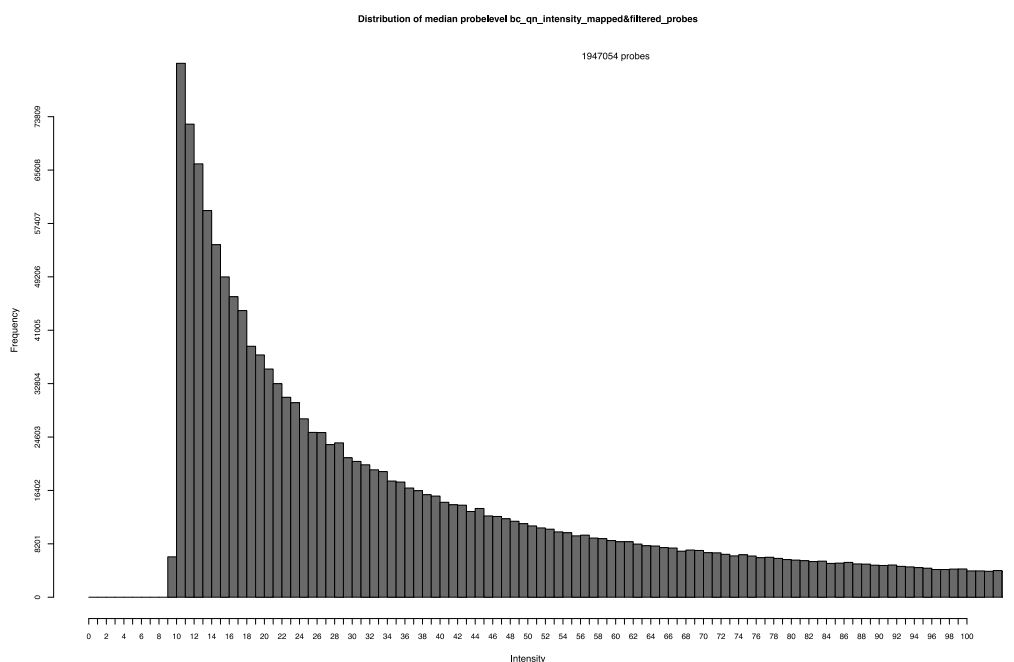


Figure 16. Demonstrating the impact of processing the filtered HTA 2.0 probes to reflect the probes expressed in a tissue specific manner - ~2.3Million probes are removed.

GC_ALL_MUSCLE_HTA_cv0.25i500_cv0.25i10_ENSE_Grch38

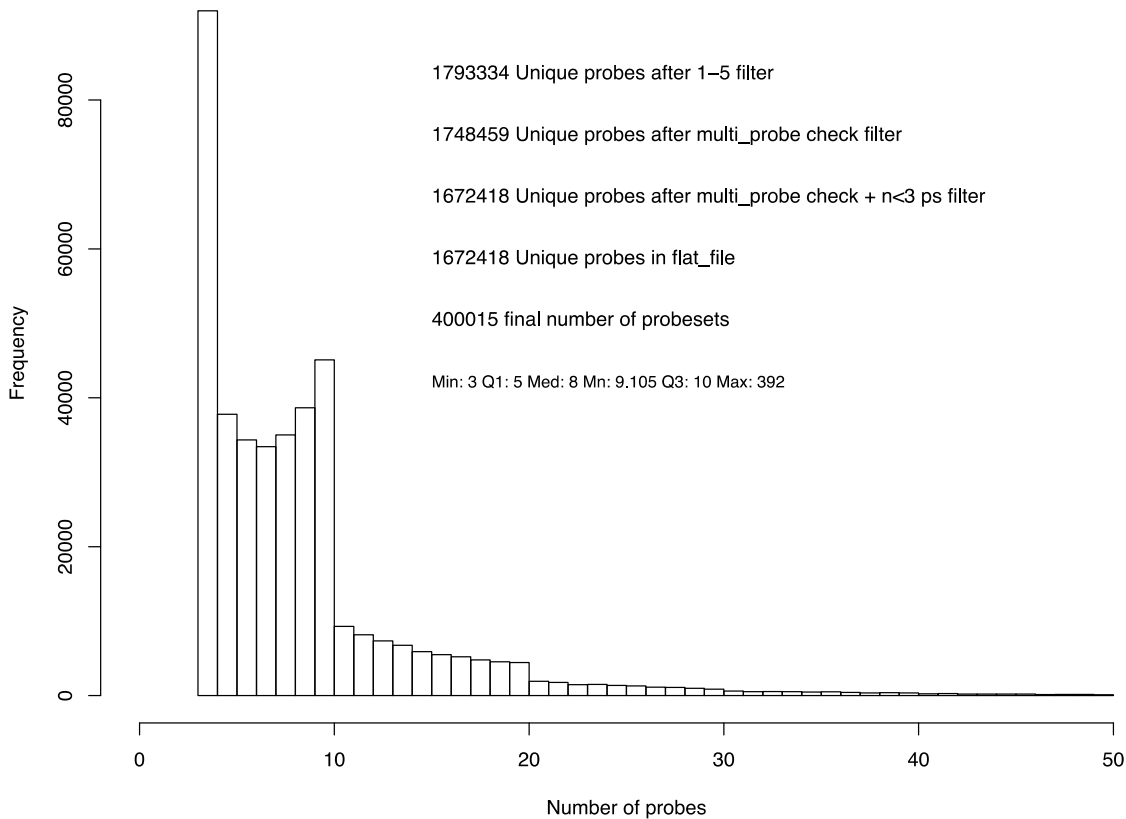


Figure 17. A muscle specific HTA 2.0 ‘probe’ map that is then assembled into probe-sets to detect individual exons. Only exons (probesets) with $n=3$ probes or more, aligned to 1 place in genome and containing probes above background signal are included.

To extract this information from the Affymetrix Human Transcriptome Array 2.0 we developed a number of tools that allowed us to quantify each component of the RNA molecule separately feed them into the classification models. Thus linear models using the following “chip maps” (CDFs) is now possible.

- GC_ALL_MUSCLE_HTA_cv0.25i500_cv0.25i10_ENSE_Grch38
- GC_ALL_MUSCLE_HTA_cv0_25i500_cv0_25i10_ENSG_Grch38
- GC_ALL_MUSCLE_HTA_cv0.25i500_cv0.25i10_ENST_Grch38
- GC_ALL_MUSCLE_HTA_cv0_25i500_cv0_25i10_3utr_Grch38
- GC_ALL_MUSCLE_HTA_cv0_25i500_cv0_25i10_5utr_Grch38
- GC_ALL_MUSCLE_HTA_cv0.25i500_cv0.25i10_Noncode2016_EnsGRCh38v83proteinencoding_Grch38

Production of these new Maps for the Affymetrix HTA 2.0 product reflects the new challenge of integrating exon level signals into ‘whole’ transcriptional units (Figure 18). Failure to carry out this procedure results in combining non-expressed units with expressed units, adding technical variation and reducing specificity of the detectable signal.

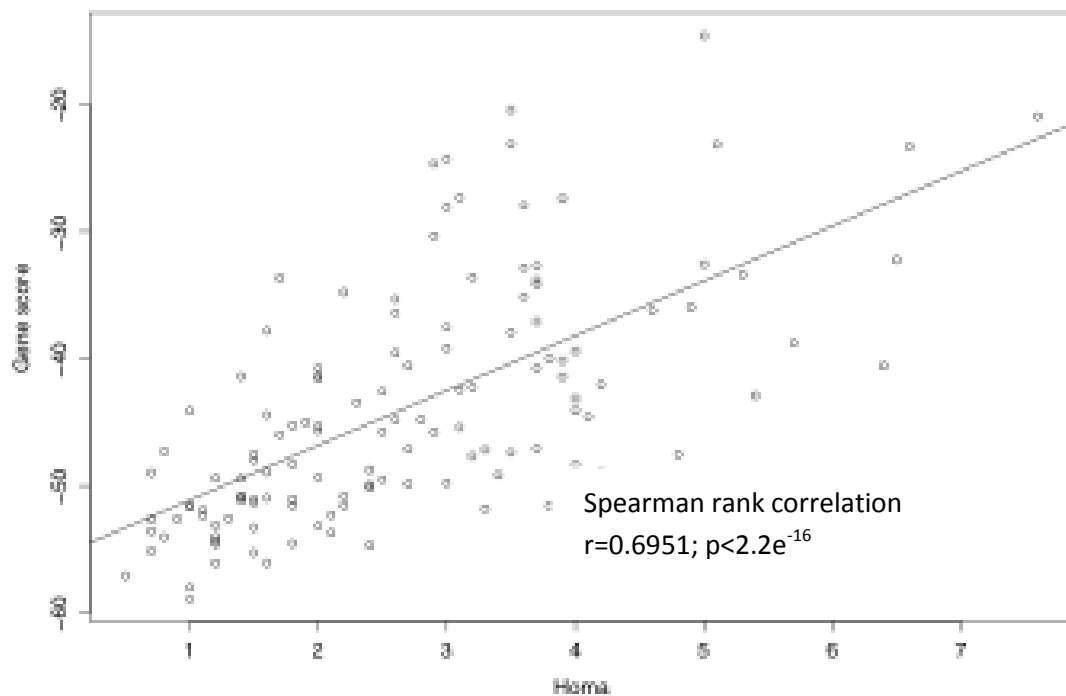


Figure 18. An example of a proto-type blood RNA versus HOMA linear model (Using STRRIDE II blood data). This is a simple Knn + Ranking model was generated prior to custom CDF methodology and will be further developed to replicate across cohorts using the new linear modeling code and CDFS.

The new linear code for the project was then produced, tested and adapted in several stages and is presented in detail in Deliverable D7.4 and figure 19. This task was completed in early spring 2016 and thus there is limited time available to apply it to the available and fully QC'd OMIC data sets (See WP2, WP8 and WP9). That work is ongoing during 2016 and 2017.

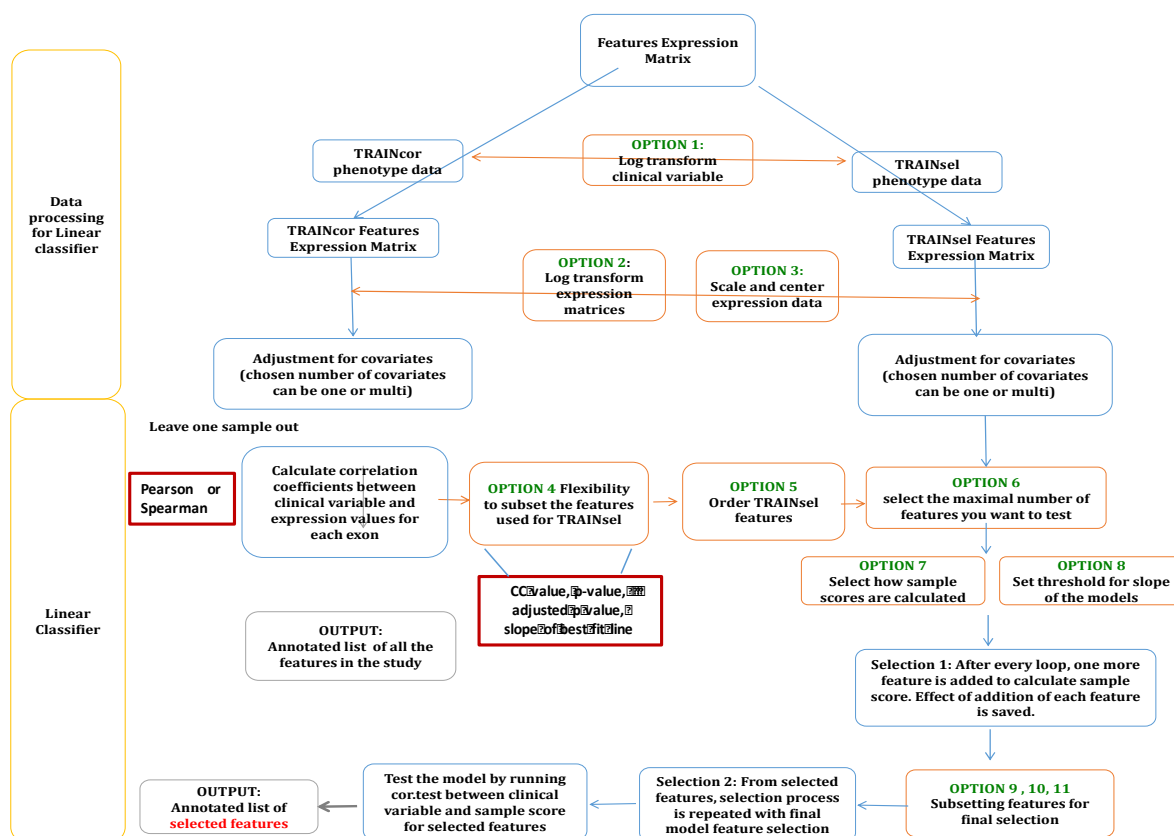


Figure 19. XRGenomics computational strategy for linear code

The metabolomics for the HIT study (WP3) included the analyses of both the metabolites and the insulin and c-peptide analyses for measurement of the insulin sensitivity. During the process of establishing the assay for the insulin it became clear that different assays give very different results.

Due to this we decided to rerun all the baseline insulin samples from the STRRIDE II and III studies to make the measure of insulin sensitivity between the different cohort that will be used to make predictors for responders and non-responders in insulin sensitivity comparable. This meant that more samples had to be analysed and the final data was not available for statistical modelling June 2015 (insulin) and May 2016 (all metabolites).

By this time point the informatics resources were fully consumed on the RNA data processes so only preliminary data analysis is presented below. Predictor analysis will be completed during 2016 for patenting and/or publishing. For obvious reasons the integration of predictors for changes in insulin sensitivity, glucose-tolerance, aerobic fitness (and blood pressure) will be carried out once the other individual predictor signatures are established. This will give us, theoretically, the 'maximum' performance assay for predicting these clinical outcomes (albeit they are less practical to carry-out). The integration of the data will also aid in understanding the underlying biology of these responses. This goal is further developed in WP9 and significant progress has been made at XRGenomics LTD to analyse the transcriptomic responses to aerobic adaptation and insulin resistance, with a specific focus on coding vs non-coding genes and gene splicing (See Deliverables D9.3 onwards).

4.1c – WP 9 Integrated biology

WP9 conclusion: Common to WP7, key was extracting reliable raw data and then carrying out novel analyses that would inform about the biology of ‘responders’ to exercise w.r.t. aerobic capacity and insulin biology. The largest available ‘biobank’ of human exercise data has been created, including a data-based with >0.5 billion data points associated with key clinical variables. Successful new methods were also published (31) -training 3 PhD students along the way - along with other manuscripts drafted. Publication will be timed dependent on the progress in WP7 and considerations of patenting of predictor signatures. What follows is a non-confidential summary of the activities in WP9 (more details can be found in Deliverables D9.3, D9.4 and D9.5).

Producing a new in-depth analysis of the human muscle transcriptome – and its responses to exercise training and insulin biology

As described in WP7 we produced a new methodology to study RNA and its regulation in response to metabolic status and exercise. Production of this high-quality and detailed view of the transcriptome could be done using either microarrays or RNA sequencing. However, RNA sequencing is still an immature science, and there are major challenges working with human tissues that express large biases in RNA expression. Using materials and expression from WP4 (Twins) and WP8 we established that our choice of the Affymetrix Human Transcriptome Array 2.0 could detect more exon usage events than the data produced by the NIH Tissue Sequencing Quality Control Consortium and that sequencing was not able to produce a linear signal for lower expressed exons/transcripts (Figure 20).

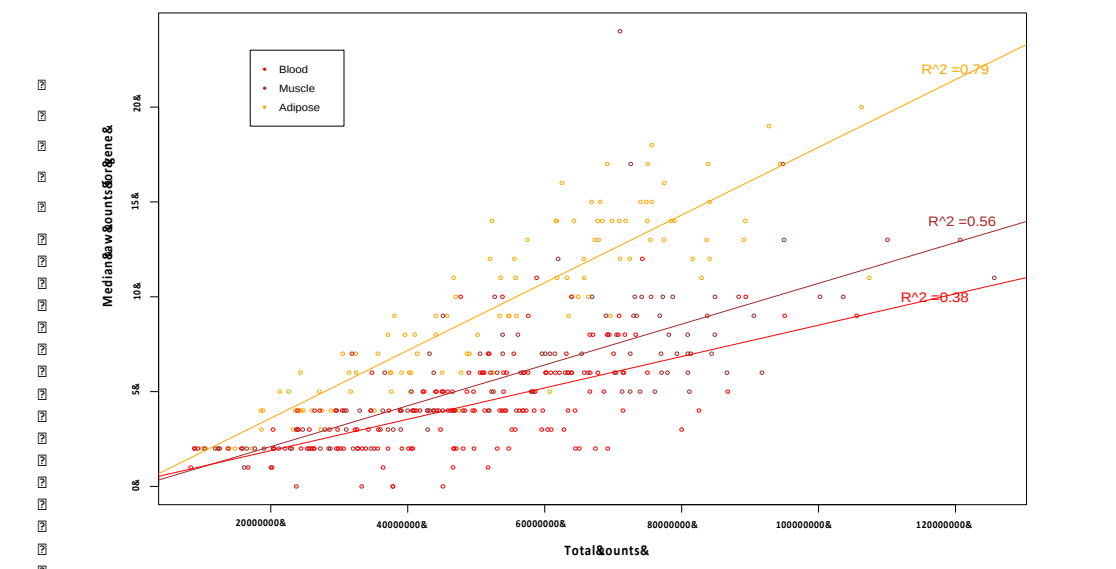


Figure 20. RNA sequencing not able to produce a strongly linear signal for lower expressed human tissue genes (especially muscle and blood).

By selecting and optimizing the most suitable technology for capturing the tissue transcriptome we have maximized our data for building classification models (WP7) as well as exploring the biology of insulin resistance and exercise responses to an unparalleled level (See below).

Global Transcript responses to HIT training

For the first time, due to the customised mapping of the Affymetrix Human Transcriptome Array 2.0 we were able to explore the regulation of not only full coding and noncoding transcript units in human skeletal muscle (e.g. Ensembl transcript ID and noncode ID) but we were able to examine if the untranslated ends (UTR) of each transcript (the sequences responsible for interacting with miRNA and the translational machinery) were differentially regulated even when the abundance of the main transcript was not (i.e. the signal changes because the UTR is shorter or longer (or a higher fraction of the transcript has a robust UTR)).

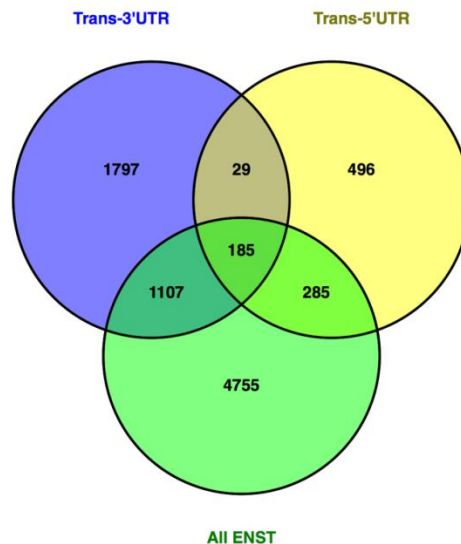


Figure 21. Full-Transcript, 3'UTR-transcript and 5'UTR-transcript regulation in response to 6 weeks of 5by1 HIT (WP3) in the skeletal muscle of subjects that demonstrate a gain in aerobic capacity (FDR 10%, FC>1.2)

As can be seen in Figure 21, there were **1797** “transcripts” which uniquely demonstrate differential regulation when only the 3' UTR is measured, while 496 when only the 5'UTR is measured. When the entire transcript is measured as a single unit, >5,000 Ensembl transcript IDs are regulated. As we will describe below, the biological profile of the 5'UTR regulated ‘genes’ was distinct from the 3'UTR. It is noteworthy that standard protein detection technology would not be capable of such a detailed and quantitative analysis.

Using a robust cut-off of 5%FDR (false discovery rate) and 1.2FC (fold change) we examined the genes that had regulated 5'UTR and the Ingenuity database. We have independently shown that the Ingenuity database up-stream analysis tool produces robust and valid information on the molecular regulators of a set of transcripts (21, 25). Up-stream analysis examines the transcript responses in your experiment, including directional change, and assesses the similarity to the impact of individual genes on gene expression networks using bibliometric data. This is more reliable than the original network-analysis tools and co-expression analysis because the p-value generated takes into account the directional change as well as the Z-score for the degree of overlap (adjusted for gene set size).

For example, HTT – is the huntingtin protein, and it has many molecular functions (including beta-tubulin binding, caspase-6 cleavage site, Cdk5 phosphorylation site, p53 binding and PAR domain binding). Mir-223 was identified as an up-stream regulator of the 5'UTR response to exercise training but not the 3'UTR response (below). Activation of mir-223 production is thought to impact

on physiological hypertrophy of cardiac muscle and indeed we noted up to 10% increase in thigh lean mass after 6 weeks of HIT training (WP3).

About 50% of the upstream regulators noted for the 5'UTR analysis were common to the 3'UTR analysis, however the 3'UTR signature was more strongly enriched in factors previously associated with exercise including overlap with our earlier map of endurance exercise training (11). This is not surprising as the earlier arrays were based on 3' cDNA libraries and 3' biased sequencing and design. This initial data provides a wealth of confirmatory and novel data for competition of the pathway analysis for publication and confirmation of the validation of the array data, from this new generation gene-chip and novel transcript mapping process, for use in the predictor modelling (confirmation given the over-lap with our 2005 and 2011 data sets on the endurance transcriptome).

Long non-coding RNA – sense/anti-sense interactions

The Affymetrix Human Transcriptome Array 2.0 also detects long noncoding RNAs and some miRNA precursor molecules. We carried out an analysis to establish which long noncoding RNAs (lncRNA) were robustly detected in human skeletal muscle and regulated by exercise. Many lncRNA molecules are transcribed Cis to protein coding genes, and these genes can appear to be co-regulated (Figure 22).

Of those lncRNA regulated we established that a subset were correlated with their cis-located antisense orientated antisense protein coding gene. This information allows us to evaluate what protein coding or functional pathways might be influenced by lncRNA.

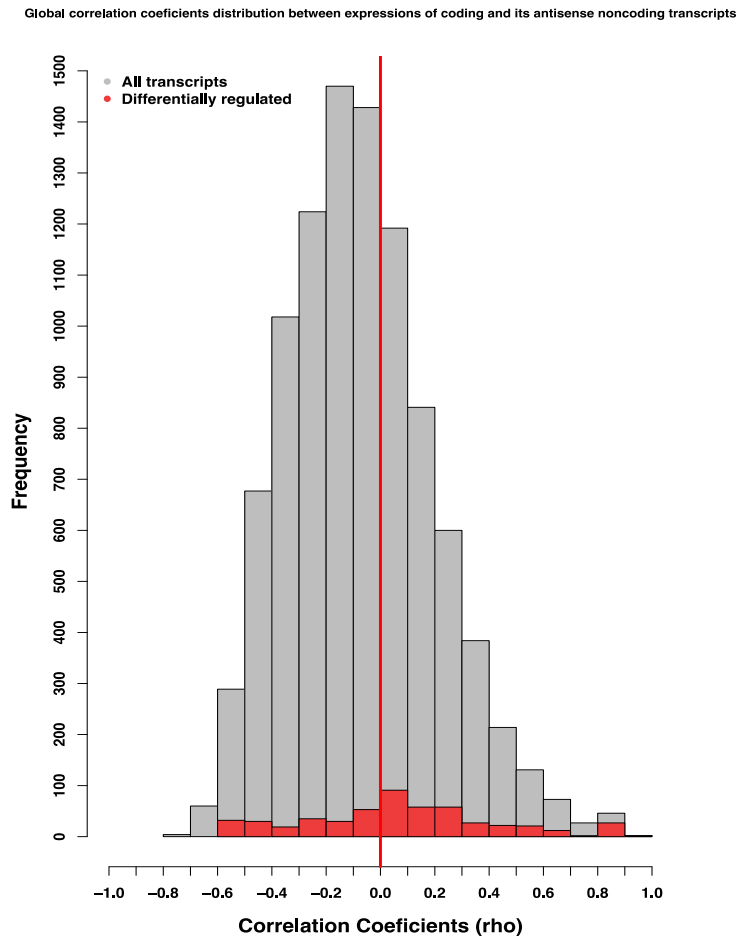


Figure 22. Global correlation between *lncRNA* and *CIS* protein coding genes in human skeletal muscle (and HIT exercise regulated genes in red).

A more global analysis of *lncRNA* *CIS* associated coding genes was carried out using DAVID and the detectable protein coding *CIS* transcriptome (to control for Ontology bias, See Timmons et al (37)). We noted that the main category was related to proteins with coiled-coil structures, including structural components of skeletal muscle. Ingenuity analysis (Figure 23) of this list of genes identified that they were regulated by Integrin linked kinase (ILK) and thus the noncoding RNA appears to be involved in regulating protein coding genes that connect integrins to the cytoskeleton and biological processes that we have previously identified as components of the responses to endurance exercise in human muscle (36).

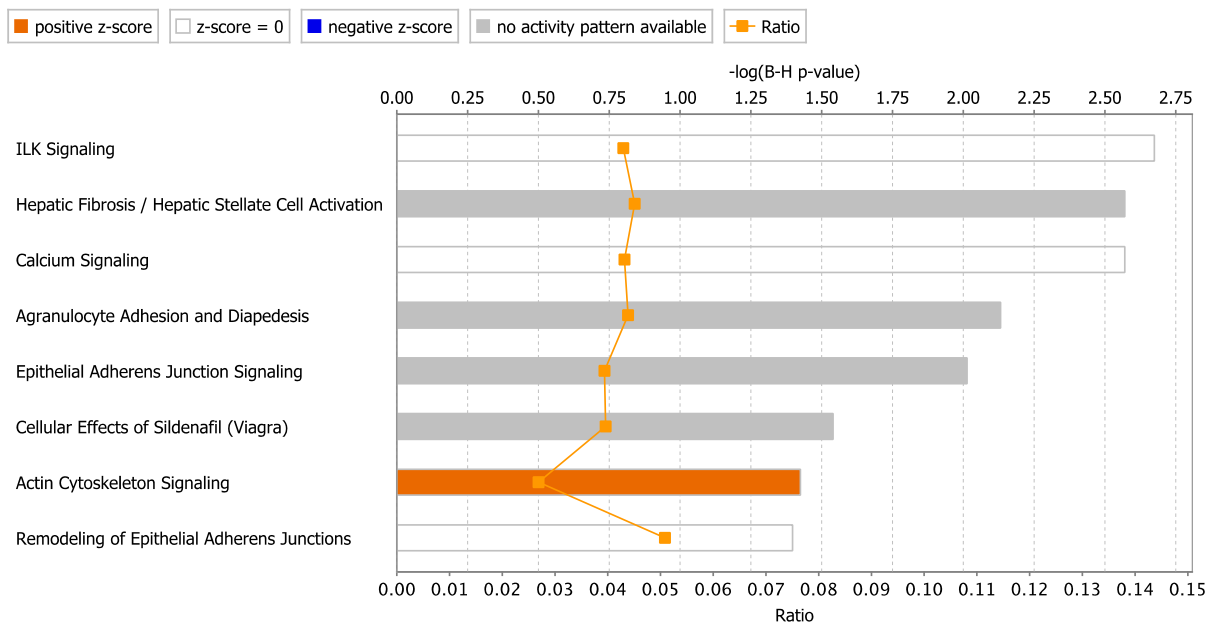


Figure 23. Up-stream regulators associated with the protein-coding genes that are co-regulated with the lncRNA exercise responsive genes in subjects that demonstrate a robust improvement in aerobic capacity

The aim of meta-predict was not only to produce a database and OMIC data set that would enable biotechnology activities (for predictors) but also to produce the most comprehensive biological ‘road-map’ of the impact of exercise on human muscle and insulin sensitivity. The main technology was detailed RNA profiling combined with targeted metabolomics. DNA analysis is not informative enough to be useful for many environmental influenced physiological traits while proteomics is not quantitative and specific enough for routine use in clinical studies. Thus to yield novel publishable units of data analysis from Meta-predict the first aim was to build a comprehensive and novel strategy for measuring a complete picture of RNA molecules in human tissue samples. As mentioned in WP9 D9.3 RNA sequencing is currently not robust enough for use in blood and muscle. As described in WP7, WP8 and used in WP9 D9.3 we produced novel analysis methods to enable profiling of each component of the transcript. To illustrate why this process was important consider Figure 24 below. It demonstrates the exon by exon profile of a typical protein coding gene. In human muscle tissue you can see that not all exons are expressed and this is why adding together all exons to calculate a ‘gene’ expression score for calculating differential expression is invalid.

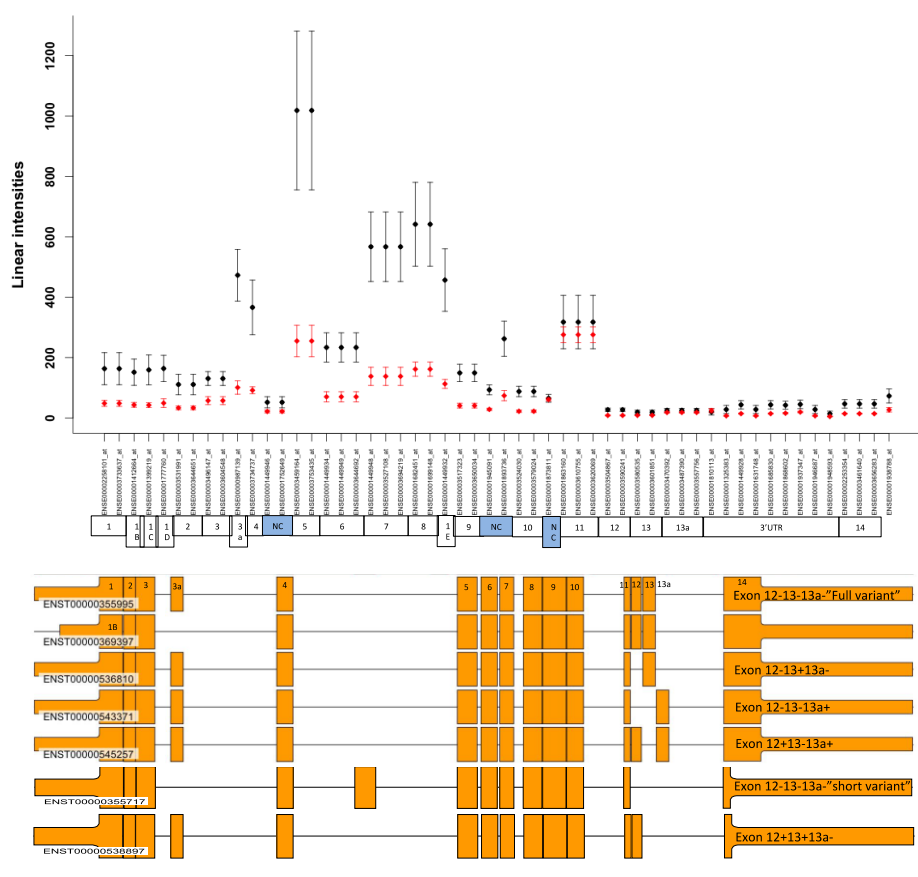


Figure 24. Probe-sets for TCF7L2 in muscle and adipose (WP5).

As you can see from Figure 24, many exons are not expressed in either tissue and thus adding these signals together is invalid. From a technical perspective these sorts of signal also add in additional background or batch related noise and thus their removal is critical to generating high quality array data in Exon or Tiling arrays.

Development and implementation of the custom mapping of probes, probe-sets and hence transcriptional units was described in detail in WP7. The first publishable unit from this work was focused on validating the use of these transcript maps for studying muscle gene exon usage. This would be critical for exploring insulin resistance and glucose tolerance responses to exercise (See below).

iGEMS (Integrated Gene and Exon Model of Splicing) represents the new method we developed to use the Exon/HTA data for studying gene splicing and exon usage (Figure 25). The method outperforms any published method.

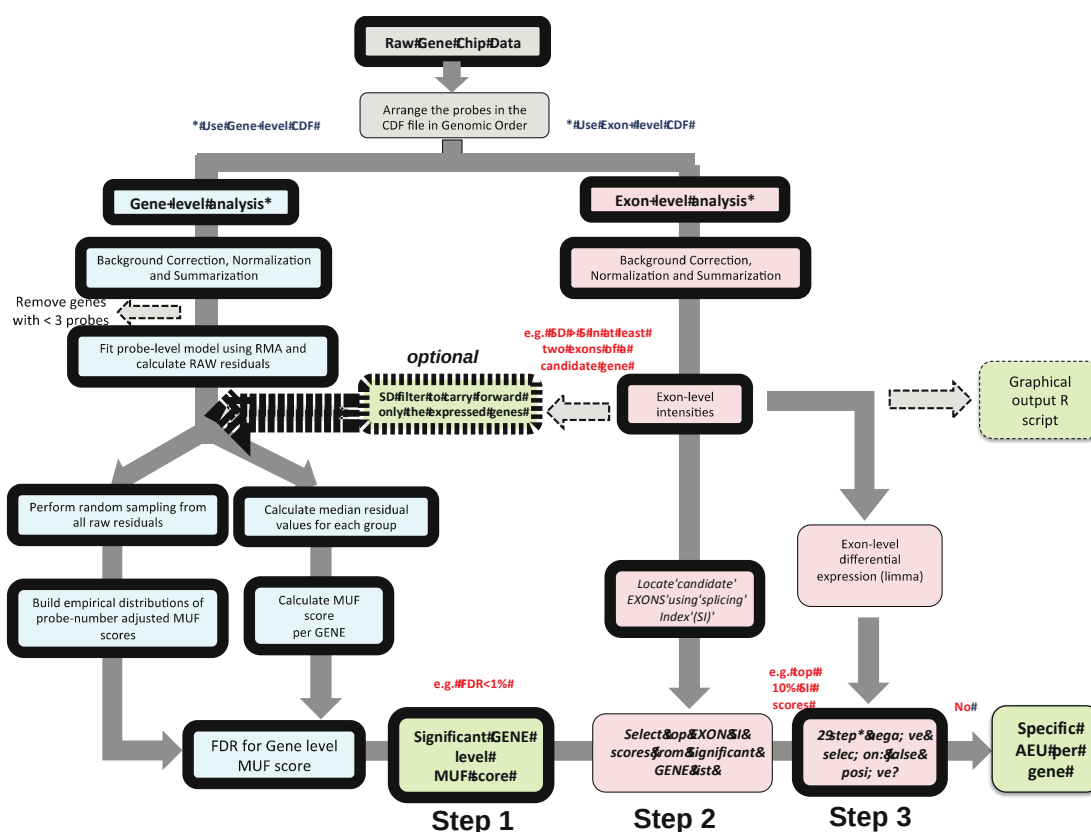


Figure 25. iGEMS analysis pipeline

The software we wrote can be located by following links at the article website: <http://nar.oxfordjournals.org/content/early/2016/04/19/nar.gkw263.full>

Application of iGEMS to human muscle in people with low vs high insulin sensitivity.

Referring to Figure 1 of Part B of the original application, we had several clinical cohorts that were part of the database/analysis plans for Meta-predict. One subcohort is the LVL MAAS cohort, where subjects (young and old) are profile at baseline and following resistance training. NUSE plots indicate the data is of high quality.

The RNA signal was extracted from the HTA 2.0 data using the custom CDFs for 'gene level' and 'exon level' using the iGEMS pipeline (31)(Figure 25). In this first analysis, we examined the relationship between RNA expression and insulin sensitivity (as defined by HOMA and insulin levels during an OGGT, AUC).

Given a central role for skeletal muscle in glucose and insulin homeostasis in healthy humans, decades of analyses using the muscle biopsy technique with biochemical assays (2), invasive physiology (9) or genomic methods applied to tissue or DNA (8, 27, 32, 41) have attempted to uncover the molecular pathways contributing to the pathophysiology of Type II diabetes (T2D) and reduced insulin sensitivity in particular ("insulin resistance").

The presentation in 2003 that the global down-regulation (20) of the expression of all oxidative phosphorylation (OXPHOS) genes was the only distinguishing feature of T2D muscle, raised the possibility that specific molecular defects in OXPHOS represent the key molecular pathway leading to skeletal muscle insulin resistance. Not only has the OXPHOS RNA observations of Mootha et al proven irreproducible in larger T2D cohorts (8, 22) there is also limited evidence that any specific biochemical defect exists in mitochondrial isolated from human skeletal muscle (4, 6, 40).

While the Mootha et al (20) observation can be explained as being an artifact of the hyper-insulinaemic clamp (23, 27) the lack of a global molecular signature for skeletal muscle insulin resistance is a strange paradox. One potential explanation was the design of the gene-chips used to profile the tissue samples and the lack of exon specific data.

It was therefore clear that in Meta-predict we would need to take a novel approach to the transcriptome. Application of the custom CDF and iGEMS pipelines yielded the first evidence that insulin resistance is associated with alterations in RNA – and that this occurs largely at the level of the individual exon. One example of a gene related to in vivo insulin resistance is presented below (Figure 26). **DECR1** is a mitochondrial 2,4-dienoyl-CoA reductase involved in fatty acid oxidation. While part of the annotated 5' UTR region is largely not expressed in skeletal muscle, the average mRNA expression is not different between insulin sensitive and insulin resistance subjects. However it is very clear that alternative use of several Exons are occurring (e.g. blue box (two exon sequences, measured with 3 Ensembl IDs (a fault with Ensembl naming strategy)) and this confirms that the ratio of splice variants being used is altered by insulin resistance.

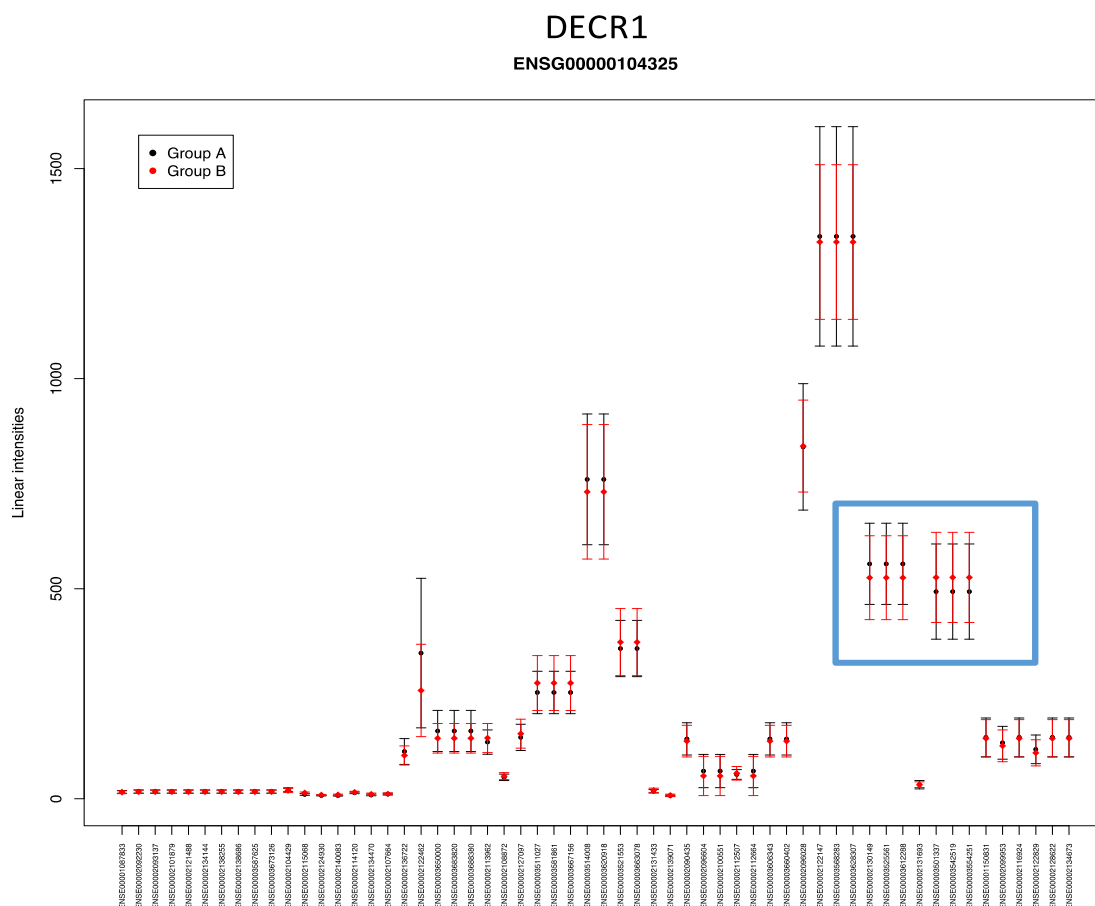


Figure 26. Plot of DECR1 in people with high versus low insulin sensitivity. The blue box indicates alternative splicing for several Exons.

Application of iGEMS to muscle tissue related to high responders in VO2max.

Another example is application of iGEMS to muscle tissue related to high responders in VO2max (Figure 27). VDAC3 (Voltage-dependent anion-selective channel protein 3) forms a channel through the mitochondrial outer membrane that transports ATP and other small metabolites.

VDAC3 is not, when the entire transcript is considered, differentially regulated by exercise. However, the ratio of individual exons are altered by exercise training and these link the potential ratio of known transcript variants. In particular there is movement from a noncoding variety of the transcript towards producing more of a coding variant.

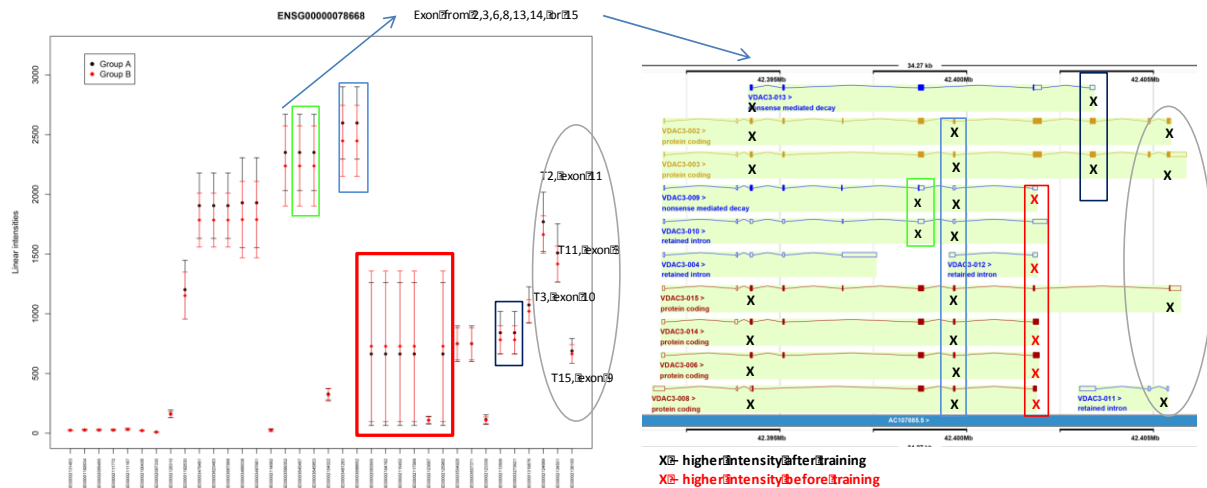


Figure 27. Application of iGEMS to muscle tissue related to high responders in VO2max

MicroRNA (miRNA) network biology.

For miRNA network biology (WP9) we have developed a method known as wCCS (weighted cumulative context score) in 2010 (8). The R-code for this analysis method has been further developed into a beta-R package to enable faster down-stream analysis in WP9. The code is awaiting final QC prior to being released as R-package publication but is available via GitHub (Gallagher). We utilized the code to process the rodent WP5 D5.3 experiment for wCCS (Figure 28).

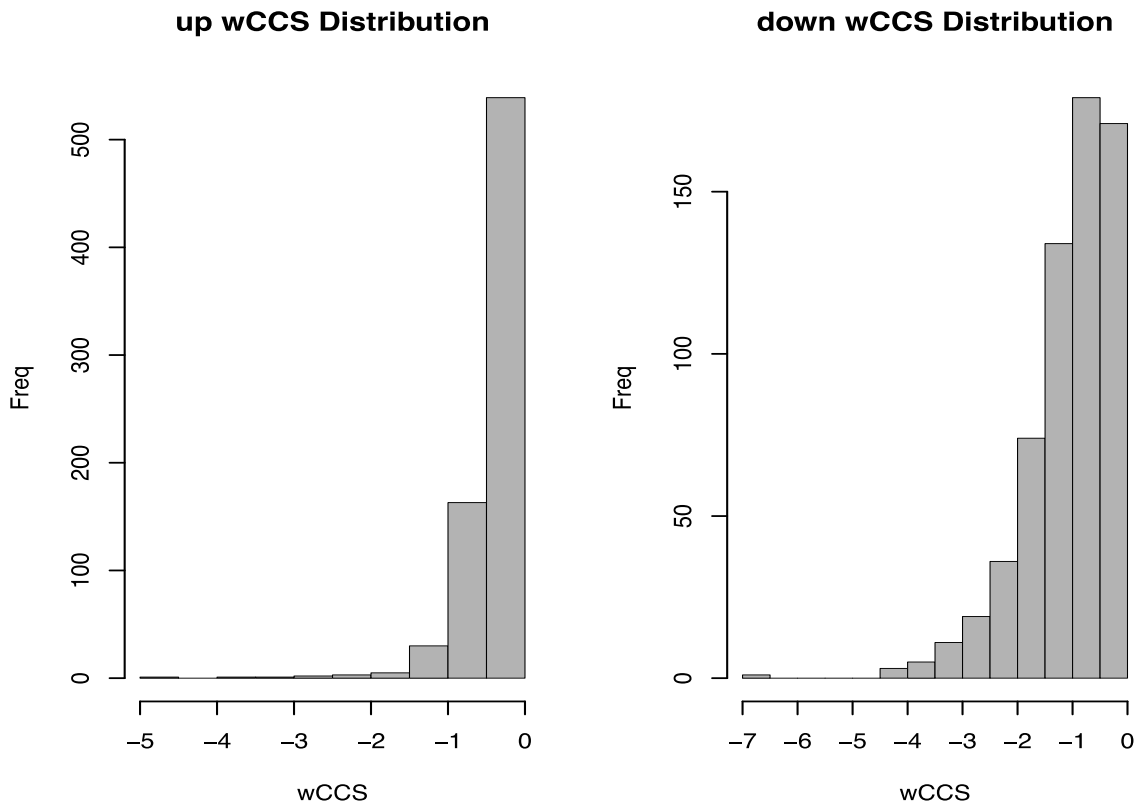


Figure 28. The wCCS for up and down regulated miRNAs WP5 D5.3 experiment.

Analysis is ongoing, cross-referencing the pathways with the 2013 Diabetes J publication on the same model, looking at gene expression changes w.r.t. insulin resistance and our clinical data (published and current). Note that in vivo miRNA changes are associated with shifts in protein expression and not typically changes in mRNA abundance (unlike plants). Those that have reported a relationship between miRNA and mRNA are typically reporting a technical artifact of excessive transfection of a single miRNA displacing endogenous miRs in the RISC complex. More results on this are shown in Deliverable D9.2.

4.1c – WP 1 Genomics

WP10 conclusion: *WP10 originally had a budget that went to partners at Pennington BRC in USA. When Pennington left the project in 2015, due to the long process no more specific budget was available for WP10 activity. However DNA has been extracted from muscle samples from pre and post-exercise training for a large subset of the HIT trial (N=83) and we have processed the data on Illumina 450k chips for DNA methylation analysis.*

The original objective of the Genomics work package is to apply targeted analysis given the sample sizes involved in Meta-predict and molecular classification models. The partner responsible for this subsequently proposed Whole Gene Sequencing instead. Following their exit, we switched the focus to DNA methylation as this would better complement the RNA data, especially given the cohort samples sizes. Due to lack of resources we have only produced the DNA and 450K chip raw data for this work-package. This still creates the largest available DNAm data set responding to 'exercise'. The data will be progressed towards publication in the next 2 years.

4.1d – Potential impact

Impact of Scientific Society. Within the project we have developed several new tools for increasing quality of transcriptomics data that has already been made available to the scientific society. New software (coding in the program R) for dealing with non-coding RNA informatics using HTA chips was made, published and made available to the scientific community. (<http://nar.oxfordjournals.org/content/early/2016/04/19/nar.gkw263.full>). In addition, R code for microRNA chip data analyses has been developed and made available as well. (<https://github.com/iaingallagher/wCCS/blob/master/README.Rmd>).

During the project we have developed a series of SOPs that we will make available for the scientific society via our website ones the major physiology paper is published.

Several scientific papers (6) have been published in open access, one more is submitted. Many more papers will be published in the next 2-3 years including the metabolomics (WP2), the predictors (WP7) and the integrated biology (WP9).

Depending on the findings and the success of finding funds or collaboration, we are planning to patent and commercialize several of the predictors for responses to exercise for insulin sensitivity, exercise capacity and cardiovascular health markers. Is successful in this this will have a socio-economic impact in new companies and jobs.

Impact on General Society. The project already had major media interest due to the usage of the time-efficient high intensity training and the largest study done showing clear benefits for health biomarkers. Also the potential to predict non-responders to certain types of exercise has obtained lots of media attention (see section 4.2B for details). One of the PIs (James Timmons) has close contact with Michael Mosley that has been active in promoting the idea of the time-effective HIT training in the form of a BBC documentary and a book (<https://www.fast-exercises.com/>) and they are planning to do more on this to improve the public awareness of this form of training.

We are planning to add a more general public summary and interpretation of our finding on to our website in the near future (www.metapredict.eu).

References

1. **Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, Shin SY, Dempster EL, Murray RM, Grundberg E, Hedman AK, Nica A, Small KS, Mu TC, Dermitzakis ET, McCarthy MI, Mill J, Spector TD, and Deloukas P.** Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS genetics* 8: e1002629, 2012.
2. **Bjorntorp P, Schersten T, and Fagerberg SE.** Respiration and phosphorylation of mitochondria isolated from the skeletal muscle of diabetic and normal subjects. *Diabetologia* 3: 346-352, 1967.
3. **Bouchard C, Blair SN, Church TS, Earnest CP, Hagberg JM, Hakkinen K, Jenkins NT, Karavirta L, Kraus WE, Leon AS, Rao DC, Sarzynski MA, Skinner JS, Slentz CA, and Rankinen T.** Adverse metabolic response to regular exercise: is it a rare or common occurrence? *PLoS One* 7: e37887, 2012.
4. **Boushel R, Gnaiger E, Schjerling P, Skovbro M, Kraunsoe R, and Dela F.** Patients with type 2 diabetes have normal mitochondrial function in skeletal muscle. *Diabetologia* 50: 790-796, 2007.
5. **Davidson PK, Gallagher IJ, Hartman JW, Tarnopolsky MA, Dela F, Helge JW, Timmons JA, and Phillips SM.** High responders to resistance exercise training demonstrate differential regulation of skeletal muscle microRNA expression. *Journal of applied physiology* 110: 309-317, 2011.

6. **De Feyter HM, Lenaers E, Houten SM, Schrauwen P, Hesselink MK, Wanders RJ, Nicolay K, and Prompers JJ.** Increased intramyocellular lipid content but normal skeletal muscle mitochondrial oxidative capacity throughout the pathogenesis of type 2 diabetes. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 22: 3947-3955, 2008.
7. **Diabetes Prevention Program Research G, Knowler WC, Fowler SE, Hamman RF, Christophi CA, Hoffman HJ, Brenneman AT, Brown-Friday JO, Goldberg R, Venditti E, and Nathan DM.** 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet* 374: 1677-1686, 2009.
8. **Gallagher IJ, Scheele C, Keller P, Nielsen AR, Remenyi J, Fischer CP, Roder K, Babraj J, Wahlestedt C, Hutvagner G, Pedersen BK, and Timmons JA.** Integration of microRNA changes in vivo identifies novel molecular features of muscle insulin resistance in type 2 diabetes. *Genome medicine* 2: 9, 2010.
9. **Hughes VA, Fiatarone MA, Fielding RA, Kahn BB, Ferrara CM, Shepherd P, Fisher EC, Wolfe RR, Elahi D, and Evans WJ.** Exercise increases muscle GLUT-4 levels and insulin action in subjects with impaired glucose tolerance. *The American journal of physiology* 264: E855-862, 1993.
10. **Karavirta L, Hakkinen K, Kauhanen A, Arija-Blazquez A, Sillanpaa E, Rinkinen N, and Hakkinen A.** Individual responses to combined endurance and strength training in older adults. *Medicine and science in sports and exercise* 43: 484-490, 2011.
11. **Keller P, Vollaard NB, Gustafsson T, Gallagher IJ, Sundberg CJ, Rankinen T, Britton SL, Bouchard C, Koch LG, and Timmons JA.** A transcriptional map of the impact of endurance exercise training on skeletal muscle phenotype. *Journal of applied physiology* 110: 46-59, 2011.
12. **Kim J, Ghasemzadeh N, Eapen DJ, Chung NC, Storey JD, Quyyumi AA, and Gibson G.** Gene expression profiles associated with acute myocardial infarction and risk of cardiovascular death. *Genome medicine* 6: 40, 2014.
13. **Knudsen S, Jensen T, Hansen A, Mazin W, Lindemann J, Kuter I, Laing N, and Anderson E.** Development and validation of a gene expression score that predicts response to fulvestrant in breast cancer patients. *PLoS One* 9: e87415, 2014.
14. **Koch LG, Pollott GE, and Britton SL.** Selectively bred rat model system for low and high response to exercise training. *Physiological Genomics* 45: 606-614, 2013.
15. **Kraus WE, Houmard JA, Duscha BD, Knetzger KJ, Wharton MB, McCartney JS, Bales CW, Henes S, Samsa GP, Otvos JD, Kulkarni KR, and Slentz CA.** Effects of the amount and intensity of exercise on plasma lipoproteins. *The New England journal of medicine* 347: 1483-1492, 2002.
16. **Kujala UM, Makinen VP, Heinonen I, Soininen P, Kangas AJ, Leskinen TH, Rahkila P, Wurtz P, Kovanen V, Cheng S, Sipila S, Hirvensalo M, Telama R, Tammelin T, Savolainen MJ, Pouta A, O'Reilly PF, Mantyselka P, Viikari J, Kahonen M, Lehtimäki T, Elliott P, Vanhala MJ, Raitakari OT, Jarvelin MR, Kaprio J, Kainulainen H, and Ala-Korpela M.** Long-term leisure-time physical activity and serum metabolome. *Circulation* 127: 340-348, 2013.
17. **Lessard SJ, Rivas DA, Alves-Wagner AB, Hirshman MF, Gallagher IJ, Constantin-Teodosiu D, Atkins R, Greenhaff PL, Qi NR, Gustafsson T, Fielding RA, Timmons JA, Britton SL, Koch LG, and Goodyear LJ.** Resistance to aerobic exercise training causes metabolic dysfunction and reveals novel exercise-regulated signaling networks. *Diabetes* 62: 2717-2727, 2013.
18. **Look ARG, Wing RR, Bolin P, Brancati FL, Bray GA, Clark JM, Coday M, Crow RS, Curtis JM, Egan CM, Espeland MA, Evans M, Foreyt JP, Ghazarian S, Gregg EW, Harrison B, Hazuda HP, Hill JO, Horton ES, Hubbard VS, Jakicic JM, Jeffery RW, Johnson KC, Kahn SE, Kitabchi AE, Knowler WC, Lewis CE, Maschak-Carey BJ, Montez MG, Murillo A, Nathan DM, Patricio J, Peters A, Pi-Sunyer X, Pownall H, Reboussin D, Regensteiner JG, Rickman AD, Ryan DH, Safford M, Wadden TA, Wagenknecht LE, West DS, Williamson DF, and Yanovski SZ.** Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *The New England journal of medicine* 369: 145-154, 2013.
19. **Miller WG, Thienpont LM, Van Uytendaele K, Clark PM, Lindstedt P, Nilsson G, Steffes MW, and Insulin Standardization Work G.** Toward standardization of insulin immunoassays. *Clinical chemistry* 55: 1011-1018, 2009.
20. **Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P,**

- Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, and Groop LC.** PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics* 34: 267-273, 2003.
21. **Nakhuda A, Josse AR, Gburcik V, Crossland H, Raymond F, Metairon S, Good L, Atherton PJ, Phillips SM, and Timmons JA.** Biomarkers of browning of white adipose tissue and their regulation during exercise- and diet-induced weight loss. *The American journal of clinical nutrition*, 2016.
 22. **Nguyen LL, Kriketos AD, Hancock DP, Caterson ID, and Denyer GS.** Insulin resistance does not influence gene expression in skeletal muscle. *Journal of biochemistry and molecular biology* 39: 457-463, 2006.
 23. **Parikh SV, Abdullah SM, Keeley EC, Cigarroa JE, Addo TA, Warner JJ, Khera A, de Lemos JA, and McGuire DK.** Effect of glucose-insulin-potassium (GIK) infusion on biomarkers of cardiovascular risk in ST elevation myocardial infarction (STEMI): insight into the failure of GIK. *Diabetes & vascular disease research* 4: 222-225, 2007.
 24. **Peter I, Papandonatos GD, Belalcazar LM, Yang Y, Erar B, Jakicic JM, Unick JL, Balasubramanyam A, Lipkin EW, Delahanty LM, Wagenknecht LE, Wing RR, McCaffery JM, Huggins GS, and Look ARG.** Genetic modifiers of cardiorespiratory fitness response to lifestyle intervention. *Medicine and science in sports and exercise* 46: 302-311, 2014.
 25. **Phillips BE, Williams JP, Gustafsson T, Bouchard C, Rankinen T, Knudsen S, Smith K, Timmons JA, and Atherton PJ.** Molecular networks of human muscle adaptation to exercise and age. *PLoS genetics* 9: e1003389, 2013.
 26. **Potteiger JA, Jacobsen DJ, Donnelly JE, Hill JO, and Midwest Exercise T.** Glucose and insulin responses following 16 months of exercise training in overweight adults: the Midwest Exercise Trial. *Metabolism: clinical and experimental* 52: 1175-1181, 2003.
 27. **Rome S, Clement K, Rabasa-Lhoret R, Loizon E, Poitou C, Barsh GS, Riou JP, Laville M, and Vidal H.** Microarray profiling of human skeletal muscle reveals that insulin regulates approximately 800 genes during a hyperinsulinemic clamp. *The Journal of biological chemistry* 278: 18063-18068, 2003.
 28. **Rottensteiner M, Pietilainen KH, Kaprio J, and Kujala UM.** Persistence or change in leisure-time physical activity habits and waist gain during early adulthood: a twin-study. *Obesity* 22: 2061-2070, 2014.
 29. **Sigal RJ, Kenny GP, Boule NG, Wells GA, Prud'homme D, Fortier M, Reid RD, Tulloch H, Coyle D, Phillips P, Jennings A, and Jaffey J.** Effects of aerobic training, resistance training, or both on glycemic control in type 2 diabetes: a randomized trial. *Annals of internal medicine* 147: 357-369, 2007.
 30. **Sood S, Gallagher IJ, Lunnon K, Rullman E, Keohane A, Crossland H, Phillips BE, Cederholm T, Jensen T, van Loon LJ, Lannfelt L, Kraus WE, Atherton PJ, Howard R, Gustafsson T, Hodges A, and Timmons JA.** A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. *Genome biology* 16: 185, 2015.
 31. **Sood S, Szkop KJ, Nakhuda A, Gallagher IJ, Murie C, Brogan RJ, Kaprio J, Kainulainen H, Atherton PJ, Kujala UM, Gustafsson T, Larsson O, and Timmons JA.** iGEMS: an integrated model for identification of alternative exon usage events. *Nucleic acids research* 44: e109, 2016.
 32. **Sparks LM, Xie H, Koza RA, Mynatt R, Hulver MW, Bray GA, and Smith SR.** A high-fat diet coordinately downregulates genes required for mitochondrial oxidative phosphorylation in skeletal muscle. *Diabetes* 54: 1926-1933, 2005.
 33. **Swank AM, Horton J, Fleg JL, Fonarow GC, Keteyian S, Goldberg L, Wolfel G, Handberg EM, Bensimhon D, Illiou MC, Vest M, Ewald G, Blackburn G, Leifer E, Cooper L, Kraus WE, and Investigators H-A.** Modest increase in peak VO₂ is related to better clinical outcomes in chronic heart failure patients: results from heart failure and a controlled trial to investigate outcomes of exercise training. *Circulation Heart failure* 5: 579-585, 2012.
 34. **Tarkka IM, Savic A, Pekkola E, Rottensteiner M, Leskinen T, Kaprio J, and Kujala UM.** Long-term physical activity modulates brain processing of somatosensory stimuli: Evidence from young male twins. *Biological psychology* 117: 1-7, 2016.

35. Timmons JA, Knudsen S, Rankinen T, Koch LG, Sarzynski M, Jensen T, Keller P, Scheele C, Vollaard NB, Nielsen S, Akerstrom T, MacDougald OA, Jansson E, Greenhaff PL, Tarnopolsky MA, van Loon LJ, Pedersen BK, Sundberg CJ, Wahlestedt C, Britton SL, and Bouchard C. Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans. *Journal of applied physiology* 108: 1487-1496, 2010.
36. Timmons JA, Larsson O, Jansson E, Fischer H, Gustafsson T, Greenhaff PL, Ridden J, Rachman J, Peyrard-Janvid M, Wahlestedt C, and Sundberg CJ. Human muscle gene expression responses to endurance training provide a novel perspective on Duchenne muscular dystrophy. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 19: 750-760, 2005.
37. Timmons JA, Szkop KJ, and Gallagher IJ. Multiple sources of bias confound functional enrichment analysis of global -omics data. *Genome biology* 16: 186, 2015.
38. Uusitupa M, Peltonen M, Lindstrom J, Aunola S, Ilanne-Parikka P, Keinanen-Kiukaanniemi S, Valle TT, Eriksson JG, Tuomilehto J, and Finnish Diabetes Prevention Study G. Ten-year mortality and cardiovascular morbidity in the Finnish Diabetes Prevention Study--secondary analysis of the randomized trial. *PLoS One* 4: e5656, 2009.
39. Van Dijk JW, Manders RJ, Canfora EE, Mechelen WV, Hartgens F, Stehouwer CD, and Van Loon LJ. Exercise and 24-h glycemic control: equal effects for all type 2 diabetes patients? *Medicine and science in sports and exercise* 45: 628-635, 2013.
40. van Tienen FH, Praet SF, de Feyter HM, van den Broek NM, Lindsey PJ, Schoonderwoerd KG, de Coo IF, Nicolay K, Prompers JJ, Smeets HJ, and van Loon LJ. Physical activity is the key determinant of skeletal muscle mitochondrial function in type 2 diabetes. *The Journal of clinical endocrinology and metabolism* 97: 3261-3269, 2012.
41. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson Bostrom K, Bravenboer B, Bumpstead S, Burt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganter M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieveer A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shriver P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature genetics* 42: 579-589, 2010.