EurHEALTHAging final report
Main Science and Technology Results

Table of contents:

WP1. Data Mining and Sample Logistics. ---------- 2
WP2. Technology coordination and Genetics ---------- 4
WP3. Epigenetic Assays and Validation: ---------- 7
WP4. Post Translational Modifications & Validation ---------- 9
WP5. Metabolomic Assays and Validation ---------- 11
WP6. Pathway analysis and statistical modelling---------- 13
WP7. Data Integration and Health Benefit Evaluation---------- 25

WP1 Data Mining and Sample Logistics.

Task 1.1. A GWAS on birthweight (bw) was carried out in TwinsUK and replicated in HCS. The results from this work were published in Twin Research in 2014. Common genetic variants identified by the EGG consortium to be associated with birthweight have been selected for WP2. Table 1 includes the list of markers identified to be validated by WP2. In addition to SNPs associated with birthweight other SNPs selected from GWAS on height, blood pressure and obesity were also chosen to be tested.

Table 1: list of markers selected for further de novo replication with birthweight in the EurHEALTH cohorts

gene	chr	pos	SNP	alleles	EAF	Total N	N studies	Change in BW Z-score effect allele (g)	SE	P-value	Heterog P value
CCNL1	3	158281469	rs900400	C/T	0.39	61142	41	-0.072 (-35)	0.006	3.6E-38	0.49
CENPM	22	40666118	rs5758511	A/G	0.27	69308	43	-0.027 (-13)	0.006	2.9E-06	0.01
ADCY5	3	124579510	rs9883204	C/T	0.76	61509	42	-0.059 (-29)	0.006	5.5E-20	0.69
HMGA2	12	64644614	rs1042725	T/C	0.50	68655	43	-0.047 (-23)	0.005	1.4E-19	0.82
CDKAL1	6	20811931	rs6931514	G/A	0.29	68822	43	-0.050 (-24)	0.006	1.5E-18	0.57
CALCR	7	93079576	rs7780752	T/C	0.64	67937	40	-0.029 (-14)	0.006	2.2E-07	0.51
5q11.2	5	57207835	rs4432842	C/T	0.30	53619	35	-0.034 (-16)	0.006	4.6E-08	0.40
LDLRAD2	1	22020351	rs6684152	G/A	0.28	67507	40	-0.024 (-12)	0.006	3.1E-05	0.42
LCORL	4	17602508	rs724577	C/A	0.73	55877	39	-0.042 (-20)	0.006	4.6E-11	0.16
CPEB3	10	93920259	rs7087501	C/T	0.57	67610	40	-0.016 (-8)	0.005	1.7E-03	0.01
RALA	7	39680346	rs13311977	G/T	0.89	58110	34	-0.032 (-15)	0.009	5.3E-04	0.05
NKAIN2	6	124829088	rs519959	A/G	0.78	48899	34	-0.025 (-12)	0.007	5.0E-04	0.01
ADRB1	10	115795046	rs1801253	G/C	0.26	49660	30	-0.041 (-20)	0.007	3.6E-09	0.97
PARG	10	50710396	rs11599257	A/G	0.68	56518	36	-0.023 (-11)	0.006	2.0E-04	0.01
HLA	6	32848420	rs2857211	G/A	0.68	54580	36	-0.019 (-9)	0.006	2.1E-03	0.15
*SLC2A4**	17	7125205	rs5415	T/C	0.31	39252	25	-0.034 (-17)	0.007	4.9E-06	0.17
KY	3	135888563	rs1562489	T/C	0.33	54688	36	-0.015 (-7)	0.006	1.2E-02	0.02
*CPB1**	3	150056878	rs2137585	G/A	0.63	40543	26	-0.029 (-14)	0.007	3.4E-05	0.04
*ZNF659**	3	21474710	rs2076829	C/T	0.19	40434	26	-0.036 (-17)	0.009	2.4E-05	0.06
GALNT13	2	154221196	rs16834011	T/G	0.24	53861	36	-0.018 (-9)	0.007	6.2E-03	0.07
HHIP	4	145821313	rs6537307	A/G	0.50	54606	36	-0.020 (-10)	0.006	5.1E-04	0.05

Task 1.2. An epigenomewide scan on birthweight was carried out in discovery and replication samples from identical twins discordant for birth weight. DNA from white blood cells was obtained in 24 (discovery) and 21 (replication) BW-discordant female monozygotic (MZ) twin-pairs, and 86 unrelated female subjects from the TwinsUK cohort. Methylation profiles were characterized using the Infinium Human Methylation 450 BeadChip Kit. Following quality-control checks, within-twin-pair methylation differences were correlated to BW differences. Meta-analysis across the discovery and replication twin datasets was performed to identify birth-weight differentially methylated regions (BW-DMRs). Altogether, 12 CpGs (within/near genes *HSF1, THADA, BRSK2, MAFF, FNINP1, RAPGEF6, MPP1, ENDOD1, TMEM126A, SLC9A3,* and *RAB3A*) were differentially methylated in the twin-pair meta-analysis ($P < 1\times10^{-8}$). Of these, 54.5% share the same direction of association between methylation and BW in the unrelated dataset of 86 female subjects. We further validated the direction of BW-methylation associations at several of these regions using methylated DNA

immunoprecipitation followed by deep sequencing (MeDIP-seq) in the 21 BW-discordant MZ twin-pairs.. The top 350 probes associated have been selected for assay development by WP2 and validation by WP3. Details on the probes that have passed QC for further development are presented under WP2. A manuscript is being drafted with these results. (see WP6)

Task 1.3. Metabolomic screens using the Metabolon and Biocrates panels on birthweight were performed in twins and the top 25 metabolites forwarded to WP5. The results of the Metabolon analyses have been published and the open access article is available at:
http://ije.oxfordjournals.org/content/early/2013/06/29/ije.dyt094.long

Task 1.4 Shipments of DNA and biospecimen samples from the clinical collections to the SMEs for genetic, methylation, metabolomic and post-translational modification testing have been carried out.

Task 1.5 Work on the NGS pipeline was carried out. A report with QC aligment and call report was submitted to the EU portal.

Task 1.6 NextGen Seq Association Analysis. Single point has one significant hit but no information on biological relevance. Borderline hit of interest: GLI2: zinc finger (7.41E-06) Mutations in this gene may contribute to sterol accumulation and atherosclerosis, and have been observed in patients with sitosterolemia.SKAT (collapsing of rare variants)- a number of hits but most significant orfs, FAM proteins (no biological information). Borderline hit of interest: NBPF1 Implicated in a number of developmental and neurogenetic disorders

WP2. Technology coordination and Genetics

Task 2.1

T2.1 comprises the validation of candidate marker panels selected from data of previous GWAS and MEDIP-Seq. The selection of the marker panels was headed by KCL and was done in accordance with the other partners of the consortium. Electronic lists of epigenetic as well as SNP candidate markers were defined and forwarded to AIT for assay design. The lists were extended by markers identified by thorough literature research. Thus the lists are based on robust statistical analyses and already identified possible marker candidates. AIT compiled a SOP for the preliminary preparation of the samples to ensure that proper amounts of DNA were shipped to the corresponding project partners.

Results: Based on the previous genome wide GWAS and MEDIP-Seq data statistical robust candidate marker panels have been selected which indicates a significant difference in birth weight (low vs. high), birth height (small vs. large), blood pressure and BMI. List with candidate markers for both epigenetic and SNP validation studies have been forwarded to AIT. The assay design is accomplished and a SOP for the shipment of the samples to the corresponding partners of the consortium has been drafted. Shipment of the samples to TATAA and AIT for the epigenetic and SNP analysis is already completed.

Task 2.2

The objective of T2.2 is the development of workflow based tools for facilitated assay design and data analysis. All tools (TBS Primer Design; TBS data analysis; assay design for MSRE and MSP qPCR; SNP data analysis tool; Epigenetic qPCR data analysis) developed within that project are made available on XworX, a workflow based, cloud enabled analysis platform which can be downloaded from http://www.xworx.org/#!download/ch3t. In order to guarantee scalability, flexibility & reusability of pipelines all tools are submitted with default inputs for easy and user friendly handling.

Each tool consists of a short description at the beginning, as well as a more detailed description with additional information at the end. Tools can be started by clicking the green triangle at the XguI window and the results will be shown at the XtablE or XvieW window, respectively. An overview of all bioinformatics tools can be found at http://www.xworx.org/#!tools/ceap.

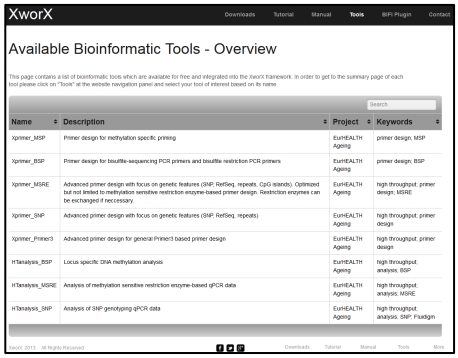


Figure 1.1: Overview of the developed bioinformatics tools in EurHEALTHAging project.

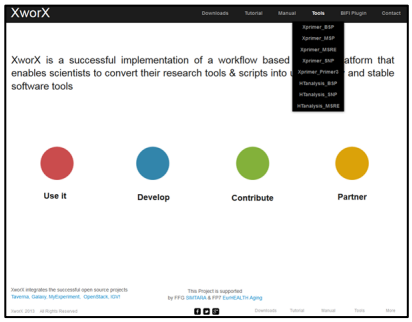


Figure 2.2: Access to the tool specific webpages for each bioinformatics tool.

In addition, by using the dropdown options under the header entry "Tools" (Figure) tool specific webpages for each bioinformatics tool providing information about: Input, Output, Tool version, Programming language, License, Screencast link, Download link for the basic tool code, Runtime and Reference information, Recommendations for tool usage and contact information can be found.

Summary: We have finished the

- TBS primer design tool
- MSRE-and MSP based qPCR design tool
- SNP qPCR design tool
- TBS data analysis pipeline
- SNP qPCR analyses pipeline
- Epigenetic qPCR analysis pipeline

All workflows are now available as tools on the XworX platform, which can be downloaded from "http://www.xworx.org/#!download/ch3t". Furthermore, we added a tool specific webpage for each bioinformatics tool providing additional information and screencasts.

T2.3 Technical validation of candidate markers using Targeted Bisulfite Sequencing (TBS)
– preliminary report in 1st period
– accomplished in 2nd reporting period

Summary of progress towards objectives

Aims

1. Assay design for candidate markers of relevant group comparisons
2. Technical confirmation of candidate markers using Targeted Bisulfite Sequencing (TBS).
3. Selection of technically validated epigenetic candidate sequences for the design of high-throughput MSRE based (or MSP-) qPCR.

To support the 450k methylation data with additional epigenome wide information, a selected panel of twins pairs (n=24) were analysed by Illumina's HumanMethylation450 BeadArrays (450k BeadChip). This was a major improvement to the initially stated project proposal, as the use of genome wide methylation BeadChips massively improves the available data and allows robust statistics. The 450k Bead Chip interrogates 485577

single CpG's distributed over the whole genome and returns the methylation state of the individual cytosine in terms of percentage. DNA of the 48 samples underwent quality assessment and was quantified using measurements based on DNA intercalating fluorescence dyes (PicoGreen). All samples passed the quality check and were deaminated according to the recommendations of Illumina. The deamination procedure with sodium bisulphite converts unmethylated cytosines into uracil by removing the amino group from the nucleotide, whereby methylated cytosines remain unaffected. To ensure sufficient amounts of DNA whole genome amplification was performed with subsequent enzymatically fragmentation of the DNA, followed by several washing procedures to ensure highly purified DNA. In the next step the processed DNA was hybridized onto the 450k BeadChips. Finally the BeadChips were scanned on an Illumina iScan. Data preprocessing and analysis was done according to Illumina's recommendations and BRB Array Tools, this included transformation of signal intensities into beta-values, quantile normalization, inference statistics, hierarchical clustering and ROC analysis. The genome wide data was analysed with respect to the different birth weight and body height of the twins. The 450k data was also analysed for any correlations with the Holloway – Isle of Wight study.

Results

Statistical evaluation of the data revealed the highly discriminative methylation pattern between twins with low and high body weight/height. However the absolute differences of approximately 10% in methylation intensities were smaller than expected.. Based on the available data derived from methylation arrays a selection of markers was nominated in accordance with the participating partners of WP2 for further confirmation on a large independent sample cohort. A technical validation of the MEDIP-Seq and BeadChip results by TBS was planned. But the small methylation differences (~10%) derived from the 450k BeadChips are hardly detectable by qPCR based methods. Therefore, we started a technical validation run with the qPCR based method, to evaluate which methylation differences are capable by those methods. The technical validation by TBS was executed after that confirmation experiments done by the qPCR based methods. Consequently, we conducted the confirmation of the 450k BeadChip data by two independent methods. This was an improvement compared to the initial stated project plan.

450 k vs MSRE qPCR

The methylome of 24 twin pairs has been analysed by Illumina's 450k BeadChip technology, which is an improvement compared to the initial project proposal. The use of state-of-the-art technology enabled us to perform a cost efficient whole genome analysis of the methylome. This yielded high statistical power and a robust selection of features for subsequent independent testing using independent samples and an independent technology. A technical validation of the 450k BeadChip by targeted bisulfite sequencing (TBS) was conducted on the same 24 twin pairs. Assays for TBS were designed using the XworX platform. The basic for the assay design were the 90 MSRE-qPCR assays included in the final marker panel (see T2.4). 63 out of the 66 TBS assays have been successfully tested and optimized in the lab. Methylation values received by TBS and by 450k BeadChips showed a high concordance (R=0.77-0.98) and confirmed the marker selection, which derived from the 450k BeadChip.

T2.4 Design, setup and validation of methods for nano-liter MSRE qPCR for epigenetic high throughput testing
- accomplished in 2nd reporting period

Aims:

1. Selection of technically validated epigenetic candidate sequences for the design of high-throughput MSRE based (or MSP-) qPCR.
2. Assay design for candidate markers of relevant group comparisons

For the methylation assays 305 epigenetic candidate markers with differentially methylated loci (DML) were nominated based on the data of the previous analysis (T2.1 and additionally conducted 450k experiments). The MSRE-primer-design tool ("Xprimer_MSRE") – developed in T2.2 – was used to design and retrieve primer sequences for all 305 candidate markers. qPCR assays were successfully designed for all 305 epigenetic candidate markers. Independently of the number of sequences for which assays should be designed, the "Xprimer_MSRE" tool takes approximately 37s for the primer design of each input sequence. Consequently, "Xprimer_MSRE" supports the user to design the best possible primer pair for the given DNA

sequence. Subsequently it was possible to evaluate the suitability of the designed assays *in silico* by evaluating the presence of SNPs in the primer sequence, the presence of repeats and the number of given cut sites for the methylation sensitive restriction enzymes (HpaII, Hin6I, AciI, HpyCH4IV).

Subsequently, each MSRE qPCR assay which passed the *in silico* evaluation was subjected to MIQE conform primer testing in single as well as in multiplexed approaches using DNA isolated from peripheral blood. All primers were tested on three different days with three different DNA samples (1 male, 1 female and DNA from 1 commercial available cell line [MCF7]) in serial dilution series (start concentration 10 ng/reaction, dilution factor: 4; calibration points: 4).

Minimum selection criteria for MSRE qPCR assays were as follows:

- Discriminative power (measured on the Illumina 450k BeadChip): gene regions with highly significant p-values ($p<0.01$) and at least a 10% difference in methylation signal intensities between the groups.
- Regions including SNPs which could interfere with restriction digestions and/or primer sequences were avoided
- At least two cut sites for the MSREs HpaII, Hin6I, AciI, HpyCH4IV.
- Long stretches of repeats and assays completely on repeats were avoided, due to the danger of reduced qPCR performance.
- Similar properties across the different assays (including primer length [18-28 bp], amplicon length [65-130 bp] and melting temperature [63-67°C]) to enable multiplexed high throughput qPCR assays.

The successfully tested qPCR assays were applied to the same twin DNA, which was also used for the 450k BeadChip experiments to confirm the findings from the BeadChips by qPCR based methods. That is a deviation to the initially stated project proposal, which describes that the confirmation will be done by TBS only. The reason for that were the small differences in methylation state between the envisioned groups.

For technical confirmation by the MSRE based qPCR the results from the 450k BeadChip have been reproduced in the same sample set. A member of TATAA visited AIT by the end of September 2013 for training purposes. The training at AIT covered the whole workflow. The large scale sample analysis was performed after that training at the labs of TATAA in Göteborg.

T2.5 GWAS/SNP replication

In previous studies GWAS experiments on a large panel of twins were conducted. Those GWAS experiments served as basis for the SNP marker selection within the EurHEALTHAgeing project. The marker selection was done in accordance with KCL, and additional SNPs with potential impact on birth weight and birth height were selected from literature. The final panel of selected SNPs contained 94 SNPs with impact on birth height and birth weight. SNPtype assays were provided by Fluidigm (San Francisco). Fluidigms high throughput SNPtype genotyping assays were based on a specific target amplification to warrant the enrichment of the targets of interest, even in the case that only small amounts of DNA are available. Enrichment was realized by a multiplexed preamplification with primers specific for the targets of interest (STA and REV primers). The preamplification was done in a 5 µl scale in a 384 well format. 1 µl of the 5 µl preamplification was diluted 1:100 and was sufficient for several hundred analyses. The readout was based on the amplification with allele specific primers (Fwd primer), amplifying in dependence of the presence of the corresponding allele either the wildtype (homozygote), the SNP (homozygote) or both (heterozygote). The detection consists of a simple endpoint measurement. The readout was done on micro-fluidic chips covering 24 assays x 192 samples, making the platform the ideal choice for high throughput testing.

Summary: 94 SNPs were selected and subjected to SNPtype assay design (**Error! Reference source not found.**). Seventy-eight SNPtype assays were successfully designed of which 72 were passed the quality criteria in several test runs and showed high concordance with SNP data derived from microarrays. Those 72 assays were applied to 3168 samples, which were provided by the University of Southampton. Sample testing has been finished in calendar week 37. Data processing is still in progress and will be made available to KCL and UOULU by the end of September at the latest.

Three different control DNAs isolated from peripheral blood were included in each Fluidigm Chip. Concordant call rates for the control DNAs were >98%. D2.5 due in M30 was delivered 11 month earlier than stated in the project proposal. After finalizing the measurements the received data was forwarded to the bioinformatics section of Southampton as stated in the project proposal.

Different significant assocations between the investigated SNPs and the parameters body weight, BMI, sysBP and diaBP have been identified.

WP3- Epigenetic Assays and Validation:

The main objective for work package 3 was to analyse samples provided from work package 1 for the methylation markers identified in work package 2 using the optimised procedures developed and to generate genomewide methylation data for both singletons and twins.
Task 3.1: Reagents for pre-amplification and high-throughput qPCR have been optimised and compared to current leading commercial reagents. Especially the pre-amplification mastermix require special conditions to be compatible with the reagent carryover from the restriction enzyme step prior to pre-amplification.
The results from the validation were reported in deliverable D3.2. Most assays worked well and all of them, except ABL1, will be used for the analysis of the samples from cohort studies. ABL1 which did not work at all was replaced by a negative control, which was lacking in the original design. Due to the difference in Cq values between samples for the same assay, for the continuation of this project both cleaved and uncleaved samples will be analysed together to be able to correctly determine the degree of methylation.

In addition to bias the most important property of the pre-amplification step is the reproducibility.. Both number of assays with a standard deviation above 0.5 and the average standard deviation for all assays are similar or better for the developed mastermix than for the leading commercial mastermixes.

The Fluidigm Biomark instruments used for qPCR analysis of pre-amplified material is not compatible with regular SYBR Green I based qPCR mastermixes. The most commonly used dye for the Biomark is EvaGreen. A mastermix based on EvaGreen was developed and compared to one of the leading EvaGreen based commercial mastermixes. Figure 3.2 shows the correlation between the developed mastermix and a commercial mastermix commonly used for the Biomark. In general the Cq-values for a qPCR reaction starts to loose reproducibility above Cq 30-34, but for the Biomark Cq-values are normally at least 5 cycles lower which means that reproducibility starts to increase at cycle 25-29. This can be seen in Figure 3.3 where the reproducibility of the qPCR replicates shows a good correlation until about Cq 25, while at higher Cq-values the qPCR replicates start to differ.
The correlation between the developed and commercial mastermixes shows a good correlation, especially at lower Cq-values. At higher Cq-values the uncertainty of the Cq determination increases the variability of the data. Further optimisation was performed to increase signal level without losing reproducibility. The optimised reagents were sent to AIT for evaluation with the methylation assays developed in WP2.

Task 3.2: The standard operating protocol (SOP) describing all steps for the DNA methylation analysis of samples from the project using quantitative polymerase chain reaction (qPCR) was finalised as planned. The SOP is described deliverable D3.1. The steps of the analysis consists of restriction enzyme cleavage of the DNA using methyl sensitive restriction enzymes, pre-amplification of the digested DNA samples and qPCR analysis of the pre-amplified samples. The main instrument for this method is the high-throughput platform BioMark from Fluidigm. The assays and procedure used in the SOP are validated in Task 3.1 and deliverable D3.1.

Task 3.3: During the second period 8832 sample reactions (from 4416 cleaved and uncleaved samples, respectively) obtained from WP1 were analysed by methylation sensitive restriction enzyme (MSRE) quantitative real-time PCR (qPCR) according to the standard operating procedure (SOP) described in deliverable D3.1. In order to achieve the best result for each sample regardless of any possible concentration differences of the samples each sample was divided into two parts where one part was digested using the MSREs and the other was treated with the same buffers and temperatures without the MSREs. The level of methylation was then based on the difference in Cq between the two parts and calculated assuming 90% efficiency, based on the assay validation in deliverable D3.2.
Among the 96 assays were 6 controls. Two of them, JUB and IRF4, are positive controls, i.e. they have no site for the restriction enzymes and should have the same Cq values for both digested or undigested samples. One, TBP, targets an un-methylated region with a site of the enzymes which should be completely cleaved in the digested samples. Another control assay, SNRPN, targets an imprinted region which means the ideal methylation level is 50% and consequently the difference in Cq between the digested and undigested samples

should be 1 (or 1.1 for 90% efficiency of the assay). One assay, XIST, targets the X-chromosome and therefore differentiates between men and women, where women's values should look like SNRPN and the men's should look like JUB or IRF4. As mentioned in deliverable D3.2 one of the control assays (ABL1) did work at all and that was therefore not analysed in the samples, but instead replaced with water as a negative control, blank.
The results of the analyses are reported in deliverable D3.3 and deliverable D3.3 Appendix 1. The raw data was transferred to partner 6, AIT, for calculation of methylation degrees. These were further transferred to the cohort partners for statistical analysis.

Task 3.4 Genomewide DNA methylation quantification using the Infinium 450k
DNA was extracted from whole peripheral blood (stored in EDTA tubes) by standardized salting out methods. Genome-wide DNA methylation levels were measured using Illumina Human Methylation 450K array (Illumina). In short, in two batches, samples (500ng of DNA per sample) were first bisulfite treated using the Zymo EZ-96 DNA-methylation kit (Zymo Research, Irvine, CA, USA). Next, they were hybridized to the arrays according to the manufacturer's protocol. The methylation percentage of a CpG site was reported as a beta-value ranging between 0 (no methylation) and 1 (full methylation).
Quality control of the samples was done with Genome Studio and all samples had a call rate >99%, and complete bisulphite conversion.
Quality control of the probes was done based on the detection p-value calculated with Genome Studio. Probes with a detection p-value of more than 0.01 in more than 1% of the samples were excluded.
Methylation beta values were generated using the 450K DataProcessing pipeline, which is based on the pipeline of Tost and Touleimat (*2012, Epigenomics*), with background correction. We excluded individuals with leukemia or received chemotherapy. Lastly, we excluded samples whose correlation with our reference population was $r<0.80$

Quality Control for Ilumina 450k data
Prior to analysis, we cleaned the data for potential outliers by first clustering the subjects from each dataset by hierarchical clustering based on the pairwise Pearson's correlation to produce the correlation heatmap and boxplot for the methylation distribution in each subject. Secondly, we correlated the first two principal components (PC) to the covariates to identify systematic batch effects. The plate, position on plate, and bisulfite conversion levels are associated with methylation levels, so these three batch effects and age are included as covariates in all the analysis. In all datasets, methylation probes that mapped to multiple locations within 2 mismatches, with missing values, mapped to Y chromosome, and with detection p-value>0.05 are removed. Finally, a total of 475,529 probes including probes on the X chromosomes are carried through to the birthweight-DMR analysis. 396 samples were analysed using the 450k and were forwarded to participant #1 for epigenome wide analysis of birthweight and other traits Results are summarized under WP6.

WP4. Post Translational Modifications & Validation

Deliverable D4.1. The main objective of the first delivery was to optimize and technically validate candidate biomarkers measuring specific PTMs.The main tasks of the delivery D4.1 were do optimized and validate the citrullinated vimentin assay (VICM) and the nitrosylated type III collagen assay (CO3-nys), as well as doing some preliminary test of the assays in relevant samples.

It was decided that VICM was an interesting assay to measure in samples this showed some predictive value for structural progression in ankylosing spondylitis. However the CO3-nys assay showed little biological relevance and it was decided that this part of the project should be discontinued. Instead other PTM assays were proposed as being more interesting for the project and therefore focus will change to those markers in the coming period.

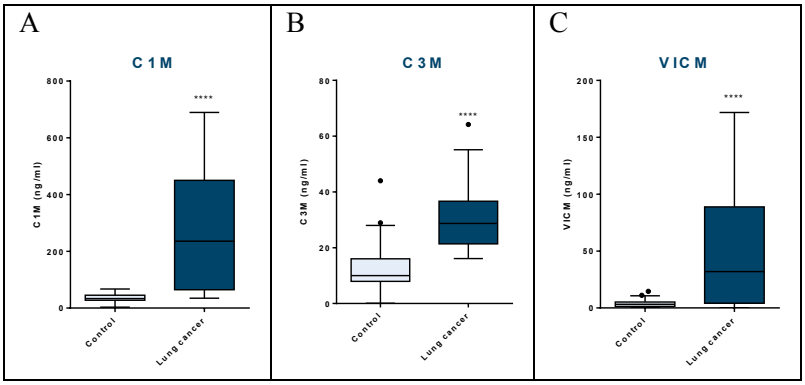
Deliverable D4.2. The main tasks of the deliverable D4.2 were to test CO3-nys, C3M and VICM in a sample cohort collected from an elderly population. The purpose of the sample testing was to pre-validate the assays before measurement in the cohorts allocated to the project. Furthermore an additional task was added to the WP. As CO3-nys failed initial evaluation, new biomarker candidate should be proposed. First we tested the two markers in a cohort of 65 elderly and investigated whether the markers were associated with age, sex or BMI. There was no significant correlation between the markers and those parameters (table 4.2).

Next we measured whether levels of the markers would be higher in individuals with above normal degree of inflammation (hsCRP>5) as compared to individuals with normal degree of inflammation (hsCRP<5).

Deliverable D4.3. The main tasks of the delivery D4.3 were to develop and test a range of novel collagen turnover marker of type I, IV, V and VI collagen. Furthermore, we wanted to continue the evaluation of the VICM assay.

The levels of MMP generated fragments of type I (C1M) and III (C3M) collagen were significantly elevated in serum from lung cancer patients as compared to controls (**Figure 4.1A-B**) with an average increase of 8-fold and 2-fold, respectively. These findings indicate that altered collagen turnover is ongoing in lung cancer. The levels of citrullinated and MMP generated fragments of vimentin (VICM) were significantly elevated (13-fold on average) as compared to controls (**Figure 4.1C**). These data together with previously presented data indicate that these markers may be markers of an unhealthy phenotype and may contribute in the prediction of disease events.

Figure 4.1.



We have prepared and tested the several proposed markers. We found that C1M, VICM and C3M were robust and showed good technical performance, and felt comfortable to measure those in the cohort samples of the project.

Deliverable D4.4. The main tasks of the delivery D4.4 were to test a selective panel of biomarkers in several studies; 1) The UK-Twin study and 2) The HCS study. Following biomarkers were measured:

- C1M (Type I collagen degradation, measures a MMP-generated fragment of type I collagen)
- C3M (Type III collagen degradation, measures a MMP-generated fragment of type III collagen)
- VICM (Citrullinated vimentin degradation, measures a MMP generated citrullinated vimentin fragment)

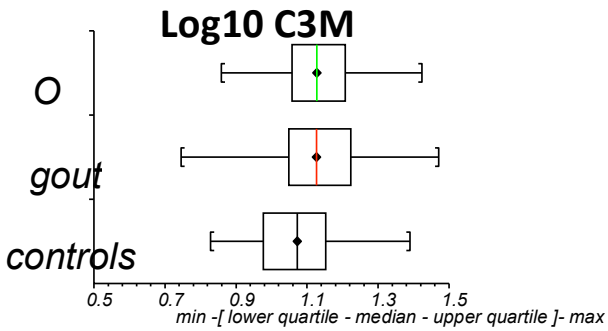
All biomarker kits were produced under GMP and validated before run in the cohort samples following internal SOPs at Nordic Bioscience. This includes setting limits for reruns and failed sample detection.

The biomarkers tested showed only a weak correlation with birthweight, both C1M and VICM are appear associated with birthweight in the twin cohort. Also VICM was associated with CVD.

Deliverable D4.5. The main tasks of the delivery D4.5 was to test two of the best biomarkers from D4.1 to D4.4, C1M and C3M, in the UK-Twin study (N=2500). Tested for relationship with birth weight and various age related co-morbidities (GOAL cohort, mean age 69 years)

Gout and OA are the two most prevalent rheumatological conditions, both of them show a strong association with age and both are usually accompanied by cardiometabolic comorbidities. In the case of gout a 30-70% increased risk in CVD mortality has been reported. We investigated the role of the two collagen degradation markers in these two conditions. Association between C3M with OA after adjustment for age, and BMI gave

an OR =66.5 [6.9-643], p <0.0003 and with gout OR =82.8 [9.9-695], p <2.4 x 10[-5].



However C1M was strongly associated with prevalence of COPD and C3M appears to be associated with prevalence of epilpetic seizures. The increased risk seen was also consistent with the negative association seen with birthweight according to DOHaD. There was no association to other co-morbidities such as Heart disease, hypertension, osteoporosis and type II diabetes (table 4.3).

Table 4.3

Trait	C1M			C3M		
	OR	95%-CI	P-value	OR	95%-CI	P-value
COPD	3.46	1.49-8.03	0.0004	1.48	0.44-4.98	Ns
Epilepsy	1.32	0.79-2.22	Ns	2.01	1.01-4.01	0.048

In summary: We developed and tested several biomarkers measuring PTMs. We found a strong relationship between the biomarkers C1M and C3M, an unhealthy outcome. This may indicate that these biomarkers may act as diagnostic and prognostic tools.A technician, as well as part of a scientist, were recruited as part of this project. These will continue their employment of Nordic Bioscience to further explore the diagnostic potential of the investigated biomarkers.

WP5 Metabolomic Assays and Validation

Task 5.1. Definitiono f 25 metabolites. These were derived from the analysis of nominally significant metabolites associated with birthweight as described in WP1. The list of these metabolites was forwarded to SL to develop the necessary assays.

Task 5.2: Development and validation of mass spectrometric MRM-assays for 25 metabolites (Duration: months 6-24)

For each metabolite, collision-induced dissociation (CID) was optimized and the most characteristic fragment ions were selected for building up the MRM-assays. Ultimate goal was the inclusion of as many metabolites as possible in one chromatographic run. Some metabolites could be sensitively detected without derivatization (**method 1**). However, for increased sensitivity and chromatographic separation, most metabolites had to be derivatized (butylated; **method 2**). Serum was used as sample matrix. Assay-validation and sample measurement comprised following steps:

Selection of materials (standards, instruments)

Name	Company	Ordering number
Standards Method 1		
Urea	PlusOne	17-1319-01
Uric acid	Sigma	U2625
DL-2-Hydroxybutyric acid sodium salt	Alfa Aesar	A18636
C-Glycosyl tryptophan	RIKEN (personal agreement with RIKEN and Shino MANABE, scientist)	-
DL-p-Hydroxyphenyllactic acid	Aldrich	H3253-100MG
Hexadecanedioate	Aldrich	177504-1G
Standards Method 2		
Amino acid standard acidic and neutral	Sigma	A6407
DL-Kynurenine	Sigma	61250
N-Acetylglycine	Aldrich	A16300

H-Glu (Leu-OH)-OH	Bachem	G-1950
H-Glu (Val-OH)-OH	Bachem	G-2015
Phenylacetyl-L-Glutamine	Santa Cruz	sc-212551
N-Acetylthreonine	Chem-Impex	#03262
2-Methylbutyroyl carnitine	Medical Isotopes	5242
Acetyl-L-carnitine	Dr. Herman J. ten Brink	-
Dodecanoyl-L-carnitine	Dr. Herman J. ten Brink	-
Isovaleryl-L-carnitine	Dr. Herman J. ten Brink	-
Oleoyl-L-carnitine	Aldrich	597562
Propionyl-L-carnitine	Dr. Herman J. ten Brink	-
Glutarylcarnitine	TRC	G597600
Succinylcarnitine	TRC	S688830
Pyridoxic acid	Sigma	P9630
Deuterated internal standards		
DL-Aspartic acid-d3	Cambridge Isotope Laboratories	DLM-832
Glycine-d2	Cambridge Isotope Laboratories	DLM-1674
L-Glutamic acid-d3	Cambridge Isotope Laboratories	DLM-3725
L-Kynurenin-d4	Buchem	-
L-Tyrosine (Ring-d4)	Cambridge Isotope Laboratories	DLM-451
N-Acetyl-d3-threonine-2,3-d2	Medical Isotopes	D5843
Phenylacetyl-d5 L-glutamine	Medical Isotopes	D30083
Hexadecanoyl-L-carnitine-d3	Dr. Herman J. ten Brink	-
Acetyl-L-carnitine-d3	Dr. Herman J. ten Brink	-
Isovaleryl-L-carnitine-d9	Dr. Herman J. ten Brink	-
Propionyl-L-carnitine-d3	Dr. Herman J. ten Brink	-
2-Methylbutyroyl N-(methyl-d3)-carnitine	Medical Isotopes	D32092
NSK-B-G mix standard	Cambridge Isotope Laboratories	NSK-B-G
Pyridoxic acid-d2	buchem	4-PYRA-D2-001
Urea 15N2	Cambridge Isotope Laboratories	NLM-233
Uric acid 15N2	Cambridge Isotope Laboratories	NLM-1697
g-Hydroxybutyric acid-d6 sodium salt	Lipomed	GHB-752-NA-100
Testosteroneglucuronide-d3	NARL	97-000056
Hexadecanedioate-d28	CDN Isotopes	D-5186
Other		
Methanol HPLC-grade	Merck	1.06035.1000
Water HPLC-grade	Biosolve	23214102
Acetonitrile HPLC-grade	Biosolve	01204101
3N HCl/1-butanol	Fluka	87472-50ml-F
V-vials	Phenomenex	AR0-3740-13 1000/PK
Bördelk./Crimp Caps N11	Macherey-Nagel	70231
LC-MS/MS unit	Thermo Scientific, TSQ Vanatge	TQU02485
LC-MS/MS unit	Thermo Scientific, TSQ Quantum	-
Column	Dionex	063189
Column	Zorbax	966967-906

Sample preparation

100 µl of human serum were mixed with 10 µl of internal standard. For protein precipitation, 850 µl of methanol were added, the mixture was vortexed two times for 5 seconds and centrifuged for 3 minutes at 10°C and 14.000 rcf. 200 µl of the supernatant were transferred into a new V-vial and dried under vacuum for 1 hour at 45°C. The dried material was re-dissolved in 100 µl of 20% methanol and 6 metabolites were measured directly using LC-MS/MS (**method 1**). 700 µl of the supernatant were transferred into a second V-vial and dried under vacuum for 2 hours at 45°C. For derivatisation, 100 µl of 3N HCl/1-butanol were added, the mixture was vortexed, incubated at 60°C for 7.5 minutes and dried under vacuum for 45 min. at 45°C. The dried material was re-dissolved in 100 µl of 20% acetonitrile/0.1% formic acid and 19 metabolites were measured using LC-MS/MS (**method 2**).

Calibration: To generate calibration curves for method 1, six different concentrations (standards 1, 2, 3, 4, 5 and 6) of the requested substances were mixed with 10 µl of internal standard in new V-vials. The mixture was dried under vacuum for 1 hour at 45°C. The dried material was re-dissolved in 100 µl of 20% methanol and the 6 metabolites were directly measured using liquid chromatography mass spectrometry (LC-MS/MS). As an example, the calibration curve of C-glycosyl tryptophan is shown in figure 1. To generate calibration curves for method 2 eight different concentrations (standards 1, 2, 3, 4, 5, 6, 7 and 8) of the requested substances were mixed with 10 µl of internal standard in new V-vials. The mixture was dried under vacuum for 2 hours at 45°C. For derivatisation, 100 µl of 3N HCl/1-butanol were added, the mixture was vortexed, incubated at 60°C for 7.5 minutes and dried under vacuum for 45 minutes at 45°C. The dried material was re-dissolved in 100 µl of 20% acetonitrile/0.1% formic acid and 19 metabolites were measured using LC-MS/MS. As example, the calibration curve of propionylcarnitine is shown in figure 2.

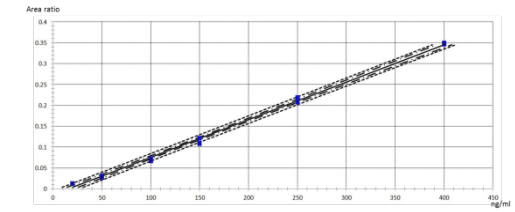


Figure 1: Calibration curve of C-glycosyl tryptophan. On the x-axis C-glycosyl tryptophan concentration [ng/ml] and on the y-axis the area ratio (ratio between signal of standard and internal standard) are shown.

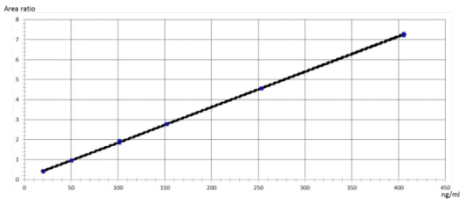


Figure 2: Calibration curve of propionylcarnitine. On the x-axis propionylcarnitine concentration [ng/ml] and on the y-axis the area ratio (ratio between signal of standard and internal standard) are shown.

LC-MS/MS

*Method 1*For direct analyses, an LC-MS/MS instrument of the TSQ Vantage series (Thermo Scientific), a reversed-phase C-18 column (Dionex Acclaim PolarAdvantage II, 4.6 x 50mm, 3µm) and following chromatographic conditions were used: 100% water/0.2% formic acid (1 min, isocratic), linear gradient to 100% methanol/0.1% formic acid (3 min), 100% methanol/0.1% formic acid (2.5 min, isocratic) and linear gradient back to 100% water/0.2% formic acid (1.5 min).

*Method 2*For analyses of derivatised samples, an LC-MS/MS instrument of the TSQ Quantum Access Max series (Thermo Scientific), a reversed phase C-8 column (Zorbax XDB, 4.6 x 75mm, 3.5µm) and following chromatographic conditions were used: 15% acetonitrile and 85% water (0.1 min), linear gradient to 100% acetonitrile (6.5 min), 100% acetonitrile (1.5 min, isocratic) and linear gradient back to 15% acetonitrile and 85% water (1 min).

Validation: Each standard was measured three times and the *limits of quantitation and detection* (LOQs and LODs, see table below) were determined using ValiData3.00. To evaluate *matrix effects*, these standards were mixed with 100 µl of human control serum from a healthy male volunteer and also analysed. No matrix interferences were observed. To guarantee *repeatability*, human control serum was analysed for 5 times on three different days. Repeatability was confirmed. To test the *precision* of the method, human control serum was measured ten times in a row. *Robustness* was tested by varying three steps during sample preparation, namely vortexing of samples (one time/30 seconds vs. two times/5 seconds), centrifugation (6 minutes instead of 3 minutes), and derivatisation time (10 minutes instead of 7.5 minutes). Each modification was analysed three times. Results showed that the methods were precise and robust.

Table 1: LOQs and LODs for the 25 metabolites

Metabolite	LOQ	LOD
Method 1		
Urea	10 µg/ml	3 µg/ml
Uric acid	7 µg/ml	2 µg/ml
2-Hydroxybutyric acid	879 ng/ml	240 ng/ml
3,4-Hydroxyphenyllactic acid	9 ng/ml	2 ng/ml
C-glycosyltryptophane	32 ng/ml	9 ng/ml

Hexadecanedioate	10 ng/ml	3 ng/ml
Method 2		
Glycine	1 µg/ml	260 ng/ml
Tyrosine	1579 ng/ml	438 ng/ml
N-Acetylthreonine	13 ng/ml	4 ng/ml
Kynurenine	9 ng/ml	3 ng/ml
N-Acetylglycine	270 ng/ml	75 ng/ml
Aspartic acid	1690 ng/ml	470 ng/ml
Glutamic acid	2 µg/ml	445 ng/ml
Phenylacetylglutamine	87 ng/ml	24 ng/ml
g-gluamylleucine	81 ng/ml	23 ng/ml
g-glutamylvaline	97 ng/ml	27 ng/ml
Pyridoxic acid	3 ng/ml	1 ng/ml
Acetylcarnitine	64 ng/ml	18 ng/ml
Propionylcarnitine	7 ng/ml	2 ng/ml
2-Methylbutyroylcarnitine	5 ng/ml	1 ng/ml
Isovalerylcarnitine	2 ng/ml	1 ng/ml
Succinylcarnitine	3 g/ml	1 ng/ml
Glutarylcarnitine	4 ng/ml	1 ng/ml
Dodecanoylcarnitine	5 ng/ml	1 ng/ml
Oleoylcarnitine	9 ng/ml	2 ng/ml

Task 5.3: High throughput quantification of 25 pre-selected candidate metabolites in 5000 serum samples by mass spectrometry (Duration: months 26-32)
After method development and validation, all ca. 5000 samples were analysed and results delivered to project partners It was agreed that all samples are returned after analysis and this was accomplished by month 35.

The analysis of these results in the three cohorts combined is presented in WP2 (task 6.2)

WP6 Pathway analysis and statistical modelling

Task 6.1 Pathway Analysis

Our pathway analysis approach is based on collective effects of the groups of genes interlinked by functional relationships (see [1]). To maximise the advantage of pathway analysis, we concentrated on the "group behaviour" of genes, their ability to interact and pre-existing annotation placing the genes into the same biological pathway, linking to the same cellular function. The benefits of the use of pathway and ontological analyses of genomics data have been extensively presented in the past [2-3].

Single genes that do not map into any statistically significant pathway will still be considered significant if reproducible and independently validated in more than one of the approaches used (methylation, NGS, GWAS, metabolomics) as they may happen to be in large pathways where the rest of the genes have no effect, and so the pathway is not significant, yet being highly significant in themselves. However, for our analysis pipeline we will leave such genes out since our approach is based on collective effects of the groups of genes interlinked by functional relationships, which is inapplicable to some genes lacking information on function, regulation and interaction with other genes.

Analysis of biological pathways will be performed using The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (**DAVID**) v6.7 (4)

Genes through GWAS

List of genes associated with birthweight

ANGPT4	*CCNL1*	*HMGA2*	*CDKAL1*
CPOE	*CENPM*	*CALCR*	*5q11.2*
CDK2	*ADCY5*	*LDLRAD2*	*LCORL*
GRB10	*CPEB3*	*RALA*	*NKAIN2*

OSBPL5	*PARG*	*HLA*	*KY*
REG1B	*GALNT13*	*HHIP*	
RUNX2			

GO TERM	Pathway	P_Value
OMIM_DISEASE	Many sequence variants affecting diversity of adult human height	2.80E-03

Genes Through Epigenetics

A disease enrichment analysis was performed using web-based gene set analysis toolkit by comparing our top 51 hits to the entrez gene database. Several disease-associated genes were found and the table below lists the top associated diseases (at least 3 genes are associated with the disease, adjust P-value< 0.01) and genes. Additionally, we found two genes, GJA4 and ITGA4, previously reported to be associated with infertility (adj. P-value= 0.0085), and NEUROG3 and DIO2 to associate with Type 2 diabetes (adj. P-value= 0.0363).

List of diseases associated top birthweight-differentially methylated regions in Twins.

Disease	Gene	Adj. P-value
Stroke;	PRKCH, GJA4, LTBP2, SORCS2	0.0011
Stroke NOS;		0.0011
Cerebral Infarction		0.0011
Subarachnoid Hemorrhage	PTPRN2, RGS12, SORCS2	0.0014
Diabetes Mellitus;	PTPRN2, BACH2, NEUROG3, DIO2	0.0022
Endocrine disturbance NOS;		0.0032
Endocrine system Diseases;		0.0032
Endocrine disorder NOS		0.0032
Autoimmune Disease	PTPRN2, ZMIZ1, BACH2, ITGA4	0.0032
Type I Diabetes Mellitus	PTPRN2, BACH2, NEUROG3	0.0032
Genetic predisposition to disease	LSP1, GJA4, PTBP2, ZMIZ1, BACH2	0.0039
Infarction	PRKCH, GJA4, SORCS2	0.005
Metabolic diseases	PTPRN2, BACH2, NEUROG3, DIO2	0.0063
Skin Diseases (genetic)	LSP1, ZMIZ1, MTA1	0.0076

*Adj. P-value: P-value adjusted by multiple tests

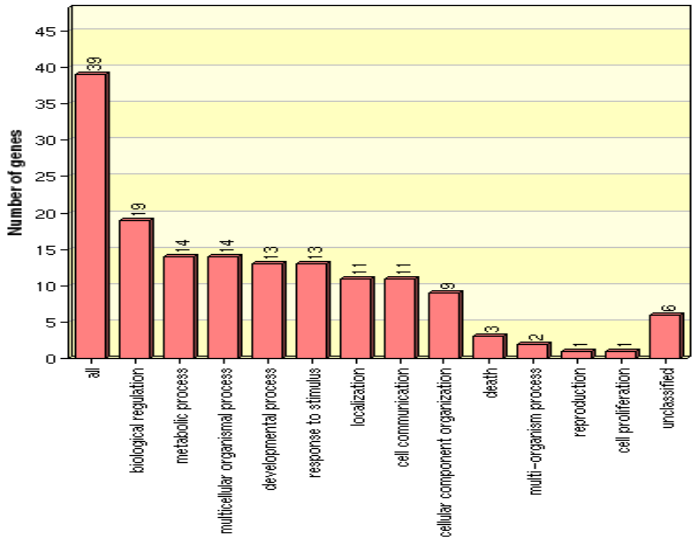


Figure showing a summary of the biological process categories of the 39 genes involved. Among these genes, 14 of them are involved in metabolic processes, and 13 are involved in developmental processes.**. Top 39 genes involved biological process**

Combined Genetic and epigenetic molecular pathways

		P_Value
GOTERM_BP_FAT	activation of adenylate cyclase activity	6.10E-03
GOTERM_BP_FAT	positive regulation of adenylate cyclase activity	6.40E-03
GOTERM_BP_FAT	positive regulation of cyclase activity	6.60E-03
GOTERM_BP_FAT	positive regulation of lyase activity	7.00E-03
OMIM_DISEASE	Many sequence variants affecting diversity of adult human height	9.60E-03
GOTERM_BP_FAT	regulation of adenylate cyclase activity	1.80E-02
GOTERM_BP_FAT	regulation of cyclase activity	1.90E-02
GOTERM_BP_FAT	regulation of cAMP biosynthetic process	2.00E-02
GOTERM_BP_FAT	regulation of lyase activity	2.00E-02
GOTERM_BP_FAT	regulation of cAMP metabolic process	2.00E-02
GOTERM_BP_FAT	regulation of cyclic nucleotide biosynthetic process	2.30E-02
GOTERM_BP_FAT	regulation of nucleotide biosynthetic process	2.30E-02
GOTERM_BP_FAT	regulation of cyclic nucleotide metabolic process	2.40E-02
GOTERM_BP_FAT	regulation of nucleotide metabolic process	2.50E-02

Discussion of Significant Pathways

- **activation and positive regulation of adenylate cyclase activity**
  Any process that increases the frequency, rate or extent of adenylate cyclase (AC) activity that is an integral part of a G-protein coupled receptor signaling pathway. This term can be used to annotate ligands, receptors and G-proteins that lead to activation of adenylate cyclase activity within a signaling pathway.
- **positive regulation of lyase activity**
  Any process that modulates the frequency, rate or extent of lyase activity, the catalysis of the cleavage of C-C, C-O, C-N and other bonds by other means than by hydrolysis or oxidation, or conversely adding a group to a double bond. They differ from other enzymes in that two substrates are involved in one reaction direction, but only one in the other direction. When acting on the single substrate, a molecule is eliminated and this generates either a new double bond or a new ring.
- **Many sequence variants affecting diversity of adult human height**
- **regulation of cAMP biosynthetic process**
  Any process that modulates the frequency, rate or extent of the chemical reactions and pathways involving the nucleotide cAMP (cyclic AMP, adenosine 3',5'-cyclophosphate).
- **regulation of cyclic nucleotide biosynthetic and metabolic process**
  Any process that modulates the frequency, rate or extent of the chemical reactions and pathways resulting in the formation of cyclic nucleotides.

Discussion

The significant pathways are involved in blood pressure, obesity and height which are in agreement with published findings linking early growth factors with adulthood disease.

Adult height has been found to be inversely associated with mortality. Data from the Danish Medical Birth Register have shown that there is a strong positive association between birth weight and adult height; for subjects with low birth weight and that genetic and/or environmental factors operating both during the pre- and postnatal period may be responsible for the association between birth length and adult height [5]

There is a very clear link between adult blood pressure and birthweight with low birthweight being a significant risk factor for adult hypertension [6,7] and obesity [8].

The results presented here shed light into the molecular pathways in common between these two phenomena.

Furthermore, we find that there is concordance in the pathways if not in the individual genes identified by genetic and epigenetic approaches.

References

1. Ptitsyn AA, Weil MM, Thamm DH. Systems biology approach to identification of biomarkers for metastatic progression in cancer. BMC Bioinformatics. 2008 Aug 12;9 Suppl 9:S8.
2. Draghici S, Khatri P, Martins RP, Ostermeier GC, Krawetz SA: Global functional profiling of gene expression. Genomics 2003, 81(2):98-104
3. Manoli T, Gretz N, Grone HJ, Kenzelmann M, Eils R, Brors B: Group testing for pathway analysis improves comparability of different microarray datasets. Bioinformatics 2006, 22(20):2500-2506
4. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. (2009) *Nat Protoc.* 4(1):44 -57

5. Sørensen HT, Sabroe S, Rothman KJ, Gillman M, Steffensen FH, Fischer P, Sørensen TI. Birth weight and length as predictors for adult height. Am J Epidemiol. 1999 Apr 15;149(8):726-9.
6. Sato R, Maekawa M, Genma R, Shirai K, Ohki S, Morita H' Suda T, Watanabe H. Final Height and Cardiometabolic Outcomes in Young Adults with Very Low Birth Weight (<1500 g). PLoS One. 2014 Nov 14;9(11):e112286. doi: 10.1371/journal.pone.0112286
7. McNamara BJ, Gubhaju L, Chamberlain C, Stanley F, Eades SJ. Early life influences on cardio-metabolic disease risk in aboriginal populations--what is the evidence? A systematic review of longitudinal and case-control studies. Int J Epidemiol. 2012 Dec;41(6):1661-82. doi: 10.1093/ije/dys190. Epub 2012 Dec 3.
8. Botton J, Scherdel P, Regnault N, Heude B, Charles, MA Postnatal weight and height growth modeling and prediction of body mass index as a function of time for the study of growth determinants. Nutr Metab. 2014;65(2-3):156-66. doi: 10.1159/000362203. Epub 2014 Nov 18.

Task 6.2. statistical modelling

Genetic and life-course analyses leading to healthy ageing – approach from early life

The intrauterine period is a vulnerable period of development. Any adverse environment can permanently change the body's organ structure and function, expressed as an increased disease risk later in life. Studies show that variability in growth patterns in early life is associated with obesity, and other cardiovascular diseases in adulthood, but the genetic and environmental determinants of these processes are largely unknown. In EurHealthAgeing we aimed to identify genetic, metabolic and environmental factors associated with early growth in infancy and childhood and later metabolic outcomes in adulthood. The analyses using early childhood growth data were based on Northern Finland Birth Cohort 1966 (NFBC1966) and NFBC1986 data while the rest of the analyses e.g. on birth measures and ageing related outcomes used also TwinsUK and Hertfordshire Cohort Study (HCS). Genome-wide analyses (GWAS) on childhood growth included also other birth cohorts studies externally to EurHealthAgeing form Early Growth Genetics (EGG) consortium (as planned due to statistical power issues).

Our results multivariable modelling show that several maternal and paternal factors, such as socioeconomic status, height, smoking, parity and pre-eclampsia, have direct and independent associations with postnatal height growth patterns, some of which had their association mediated by size-at-birth variables. It was observed that an obesogenic environment *in utero* and during a child's growth exerts a 'programming' effect on the glucose-insulin axis as well as cardiovascular risk factors in adolescence. We identified genetic variants (e.g. in *FTO*-gene) that showed an age-dependent association with adiposity in early childhood, while others clearly had their effect on adult adiposity mediated by early growth phenotypes. These analyses emphasise the clinical importance of early growth markers as it may inform public health policy aimed to improve the pre-pregnancy environment and also monitor infant growth during the first years of life (Das, thesis 2014, Das et al 2015). Our discoveries on *FTO*-gene led us to investigate further its' function. These results demonstrate that this gene has a role in adipose tissue which modifies the response of white adipose tissue to high-fat feeding. *Fto*-deficiency in mouse model increases the expression of genes related to adipogenesis preventing adipocytes to become hypertrophic after high-fat diet (Ronkainen et al 2015, in press). Our results emphasize the importance of further work on molecular mechanisms.

For further statistical analyses, a birth weight lowering genetic risk score was calculated for Northern Finland Birth Cohort 1966 (NFBC1966) and Hertfordshire Cohort Study (HCS). The risk score was based on 17 of 22 SNPs formerly identified as being associated with birth weight (Horikoshi et al 2012, where NFBCs and HCS were included), reduction in the number of SNPs being due to availability in HCS. The association of the risk score with birth weight itself was validated in both cohorts. As birth weight itself is associated with various adult phenotypes, it was hypothesized that these same genetic factors affecting birth weight could be associated with health-related outcomes in later life; i.e. may explain partially the observed associations between foetal growth and adult phenotypes.

The first set of outcomes was adult adiposity measures: BMI, weight, hip circumference and waist circumference and in NFBC1966 also blood pressure. The results for adiposity measures showed significant negative association with the birth weight lowering risk score and all adiposity measures in the NFBC66 (n=4912). However, there was no evidence for any association in the HCS (n=2397). This could be due to the age difference in the two cohorts; the NFBC1966 measurements were done at the age of 31 years, while in the HCS the phenotypes in question were measured at much higher age (~ age of 60 years). The above analyses are not conducted before and show novel, slightly unexpected but plausible results. The detailed analyses are presented in the Appendix.

In NFBC1966, the effect of the birth weight lowering genetic risk score on adult blood pressure was also examined. Evidence for negative association between the risk score and systolic blood pressure was found,

even after various adjustments. We did not find any statistically significant association for diastolic blood pressure. All above work is being prepared for publication .As the next step, the same risk score was calculated for Northern Finnish Birth Cohort 1986 (NFBC1986) and the association with the same set of phenotypes will be examined. As the data for NFBC1986 is measured at the age of 16, this will provide more information on the association at yet another time point compared to NFBC1966 and HCS.

Both NFBC1966 and NFBC1986 have data measured on multiple time points, ranging from maternal variables to the most recent 46-year study for NFBC1966 collected between 2012 and 2014. For these longitudinal datasets, structural equation models (SEM) have been constructed until the age of 31 years, available for the present study, in order to gain more depth for understanding the life-course association with the genetics and early-life variables and the later life phenotypes related to "healthy ageing". This work will continue by using the latest data which should be available in 2015 for the analyses. Later on, a subset of these wider life-course models could be used in cohorts with no longitudinal data available. The latest completed analyses during the reporting period using SEM on 5198 participants from the NFBC1966 with data on birth weight, height and weight measurements until adolescence, systolic and diastolic BP at 31 years and several other covariates including genetic risk score for blood pressure (Figure 1). Construction of these final models includes multiple sets of analyses, very careful insight into the individual associations between the variables, and testing of the model performance (the baseline developmental work is described in the thesis by Marika Kaakinen in 2013). Negative direct effects of birth weight on adult systolic BP were observed (standardised regression coefficients: −0.08 (−0.14 to −0.03) in males and −0.04 (−0.09 to 0.01) in females, equalling −1.99 (−3.32 to −0.65) and −1.01 (−2.33 to 0.32) mm Hg/kg, respectively). Immediate postnatal growth was associated with adult BP only indirectly via growth later in life. In contrast, growth from adiposity rebound onwards had large direct, indirect and total effects on adult BP. Current body mass index was the strongest growth-related predictor of adult BP (0.36 (0.30 to 0.41) in males and 0.31 (0.24, 0.37) in females, equalling 1.29 (1.09 to 1.48) and 0.81 (0.63 to 0.99) mm Hg/(kg/m2), respectively). As a summary, our path analytical approach provides evidence for the importance of both foetal growth and postnatal growth, especially from adiposity rebound onwards, in determining adult BP, together with genetic predisposition and behavioural factors. These analyses also give much more in-depth understanding about the interplay of different contributing factors and show how important it is to follow-up groups at risk from very early beginning of life. This kind of work is the starting point for effective primary prevention (Kaakinen et al 2014).
In addition to the above longitudinal analyses we have been working on environmentally and lifestyle related factors in adult life which may be causally related to ageing related conditions. One of these potential contributing factors is vitamin D status. Our results suggest that higher 25(OH)D leads to reduced risk of hypertension, providing support for important non-skeletal effects of vitamin D (Vimaleswaran et al. 2014).

Metabolomic data analyses (see Tables below)

Based on Menni et al paper (2013), 25 metabolites associated with ageing were identified and the data for these metabolic variables in TwinsUK (n NA), HCS (n NA) and NFBC86 (n=2394) was provided by Seibersdorf Laboratories. The association with these metabolites and birth measures (birth weight, birth length, ponderal index, head circumference) were examined. The analyses were stratified by sex. For birth weight, the female results were meta-analysed and 12 metabolites were identified being associated with birth weight.

After the meta-analyses, in order to examine the risk for a low birth weight, a birth weight metabolomic risk score for females was constructed in each cohort separately. This was done by treating the lowest tertile of birth weight vs. the rest as a binary dependent variable (value 1 if individual belong in the lowest birth weight tertile). Stepwise regression was used in order to find out if any of the 12 metabolites could be excluded from the risk score, nevertheless no support for excluding any variable was found by these stepwise methods. However, due to severe collinearity, one metabolite was removed from the final model, leaving 11 metabolites in the final risk score model. The AUC in a ROC curve was then calculated for each three cohorts. The next step in the metabolites analysis is to examine the relationship with the metabolites and the genetic risk score in NFBC1986. These works are being written up for publication. We have also participated in important methodological work related to assessing of multivariate gene-metabolome associations with rare variants. This work proposes a new statistical approach based on Bayesian reduced rank regression to assess the impact of multiple SNPs on a high-dimensional phenotype. Because of the method's ability to combine information over multiple SNPs and phenotypes, it is particularly suitable for detecting associations involving

rare variants. The work demonstrates the potential of our method and compare it with alternatives using the Northern Finland Birth Cohort with 4702 individuals (Marttinen et al 2014).

Methylation data analyses

For DNA methylation, the data for 90 assays and 6 control assays was provided by TATAA for TwinsUK (n=1034), NFBC1966 (n=1521), NFBC1986 (n=1527) and HCS (n=269). The associations with the markers and birth weight were examined and meta-analysed for TwinsUK, NFBC1966 and NFBC1986 data, using a brief QC of scaling and quantile-normalising the assay values. Evidence for association was found for a CpG site near *MIA3* gene where variants have been associated with cancer development and coronary artery disease.

Methylation data using Illumina 450K array (HM450K) data has recently become available for both NFBC1966 (n=816) and NFBC1986 (n=552) (in late 2014) as well as in HCS and TwinsUK. A thorough QC pipeline process is set up based on our experiences using this array. The first results show that almost all previously published top hits associated with maternal smoking during pregnancy (using cord blood DNA and HM450K; Richmond et al 2014) show also association in NFBC1966 in samples taken at age 31 years. These results are extremely valuable from public health point of view. The QC process for our HM450K data is further ongoing in order to provide usable data for epigenome-wide association analyses. The original plan was not to use HM450K arrays at larger scale because their usability was not yet established when the current study was designed. However, during the project's time more evidence was gathered and it because evident that it will be feasible to carry on with larger scale analyses. Our ongoing work also approves HM450K's value also in population based studies. In the future, the longitudinal life-course models described earlier could be appended with both metabolomic and epigenetic data. The analyses done until this point give excellent starting points in creating these models, and the wide variety of data enables creation of complex models which, given such a complex set of phenotypes as "healthy ageing", are severely needed in order to understand the very nature of healthy ageing.

Overall comparison of results between datasets

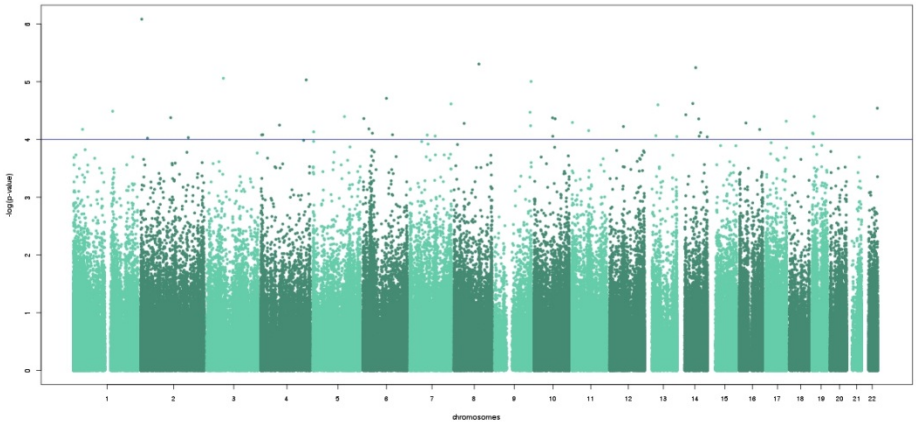
Overall the analysis of genetic, epigenetic and metabolomics data showed consistent effects for singletons and twins for metabolites. A metabolomics profile in adult age reaches 60% of the area of under the curve for birthweignt. On the other hand, genetic factors are clearly different between singletons and twins, but are reproducible within singletons and are associated with blood pressure. Epigenetic analysis identified only two probes consistently associated with birthweight in both singletons and twins. The genetic risk score analyses with adiposity related adult outomes showed different patters by age of the cohort.

EPIGENETIC ANALYSES

EWAS for birthweight in Twins and singletons using 450k data

The genomewide methylation generated in Task 3.4 was used to perform epigenomewide scans in both twins and singletons

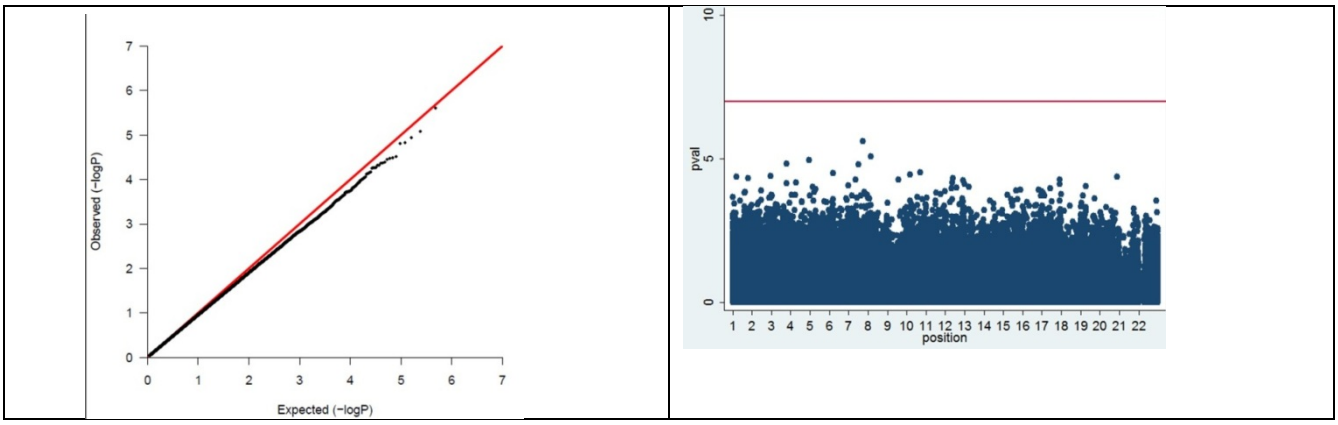
EWAS for birthweight in twins: Manhattan plot for birthweight EWAS in TwinsUK. Manhattan plot of bw-EWAS. Each green dot represents is a transformed –log10(p-value) value of one CpG site. Blue line indicates p-value of 10^{-4} and red line indicates Bonfferoni correction criteria P = 1.05×10^{-7}. None of the hits reached genomewide significance.



List of top probes in the twins

IlmnID	CHR	Gene NAME	β_Observatory	β_Replication	P-value
cg26174880	2	-	-0.708	-0.646	8.27E-07
cg06699564	8	-	-0.193	-0.852	4.95E-06
cg12165758	14	PRKCH	-0.574	-0.677	5.71E-06
cg22145181	3	-	-0.184	-0.842	8.73E-06
cg01324261	4	SCRG1	0.638	0.610	9.30E-06
cg14410072	9	C8G;FBXW5	-0.786	-0.450	9.90E-06
cg23366832	6	BACH2	0.459	0.702	1.94E-05
cg01510588	14	C14orf183	0.756	0.443	2.38E-05
cg12415687	7	PTPRN2	-0.588	-0.608	2.42E-05
cg21258821	13	KBTBD6	0.722	0.481	2.52E-05
cg26621897	X	TMSB15A	0.363	0.739	2.74E-05
cg12961733	22	-	-0.515	-0.653	2.88E-05
cg05630111	1	LASS2	-0.555	-0.620	3.24E-05
cg06866628	9	GTF3C5	-0.350	-0.738	3.38E-05
cg16280098	14	PABPN1	-0.667	-0.519	3.73E-05
cg15222563	19	TRAPPC5	0.534	0.625	4.00E-05
cg06062821	5	-	0.561	0.606	4.01E-05
cg02120071	2	-	0.486	0.655	4.20E-05
cg06973667	10	NEUROG3	0.473	0.663	4.22E-05
cg15323253	6	-	0.756	0.410	4.35E-05
cg17045635	10	ZMIZ1	0.640	0.536	4.39E-05
cg07171024	14	DPF3	-0.595	-0.575	4.40E-05
cg12846139	17	BAHCC1	-0.747	-0.415	4.83E-05

EWAS for birthweight in singletons: The same techniques were used to perform an EWAS on birthweight in the singleton cohorts. Again, no genomewide significant hits were found and very little overlap in the top 10000 probes with those from twins.



LIST OF TOP HITS IN SINGLETONS:

Gene	CPG_ISLAND	TargetID	chr	Cjr [psotopm_36	bw_pvalue	bw_beta	bw_se
		cg15287987	7	117791566	2.47E-06	4.0791	0.7919
		cg25967146	8	23773851	8.28E-06	-3.1752	0.6574
ODZ3		cg16793274	4	183951366	1.15E-05	3.1993	0.6746
DHX36		cg27297043	3	155517786	1.49E-05	-3.2903	0.7044
SEMA3A		cg16346212	7	83662191	1.55E-05	2.9506	0.6332
CYP2C19		cg04189838	10	96513337	3.05E-05	-3.4565	0.7726
TAPBP	N_Shore	cg26315802	6	33389950	3.24E-05	3.2391	0.7268
IGBP1	Island	cg19629755		69269900	3.34E-05	1.8673	0.4197

LOC100128811	Island	cg06967120	10	25504265	3.54E-05	3.3740	0.7612
MLPH	N_Shore	cg17187163	2	238059746	4.06E-05	-3.3565	0.7639
BMP7	Island	cg20340302	20	55274556	4.24E-05	-2.8304	0.6459
		cg12804647	1	48243188	4.28E-05	-3.3038	0.7544
GPR25	N_Shore	cg24550456	1	199108567	4.79E-05	-2.9578	0.6803
KRT83	Island	cg16160105	12	51001317	4.79E-05	3.4751	0.7994

The lack of agreement in the genes found between twins and singletons suggest that epigenetic markers of birthweight, as are also genetic influences, are different in twins and singletons and should be analysed separately. This was confirmed by the meta anlaysis of the hits analysed by MSRE based (or MSP-) qPCR (see below)

A note of caution is that the EWAS results derive in total from less than 200 singleton samples and 296 twins. One of the starting assumptions of the project was that epigenetic effect sizes on birthweight would be much larger than genetic effect sizes and therefore even a few hundred individuals would be sufficient to identify consistent epigenetic effects. This has not proven to be the case and one important conclusion from this study is that larger discovery sample sizes are needed.

Epigenetic replication of initial discovery hits in data generated by MSRE based qPCR

Using the MSRE qPCER methylation data we combined data from all cohorts (total n= 3400) for the probes that passed QC that were generated by D3.3. The meta-analysis results for the association of candidate markers with birth weight (TwinsUK, NFBC1966, HSC and NFBC1986) are shown in the table below.

Marker	P-value	Beta	SE
ABHD1	0.737085744	0.006759832	0.020135613
AGAP1	0.996476861	8.79047E-05	0.019907835
ASXL1	0.471239524	0.014506611	0.02013511
ATP9B	0.468891731	0.014522085	0.020050285
BAI3	0.509812045	0.013157687	0.019962198
BCAN	0.962405699	-0.0009521	0.020199531
BLCAP	0.888311968	-0.002792969	0.019887213
C11orf92	0.481901528	0.014070445	0.020007821
C1orf53	0.437383539	0.015521126	0.019985513
C21orf67	0.559641627	0.011721698	0.020092946
CD32	0.241915817	0.023543095	0.02011866
CRB2	0.664595767	0.008713903	0.02009771
DDB2	0.262007354	0.022535881	0.020091554
DIP2C	0.291600224	0.021045675	0.019955737
EIF2AK1	0.520164932	-0.012917211	0.020086133
EP4	0.273352979	-0.022011635	0.020095089
ESYT1	0.356925003	0.018509664	0.020092115
F5	0.43575647	0.015544673	0.019944952
FAM172A	0.112387259	0.031385014	0.019769416
FBXL21	0.514669563	0.01309028	0.020089941
FOXG1	0.433742262	-0.015721538	0.020083627
FOXQ1	0.314666838	0.020166048	0.020056183
HES7	0.383123622	0.017526147	0.020095216

HIP1	0.062598512	0.037507479	0.02014331
HLA.DPB1a	0.342394548	-0.019117511	0.02013547
HLA.DPB1b	0.122101699	-0.031139947	0.020142112
HLA.DPB1c	0.064544658	-0.037203842	0.020127614
HSD17B12	0.269026542	0.022243256	0.020124047
HSPA1A	0.433256989	-0.015733446	0.020077641
IRF4	0.980961258	-0.000475232	0.019914446
KCNN2	0.562593088	0.011575289	0.019992068
LDB1	0.462430681	0.014750708	0.020073086
LOC644145	0.850606075	0.003781261	0.020076229
MBP	0.456062468	0.014958583	0.020069316
MIA3	**0.002200304**	**0.061630631**	**0.020128944**
MRPL11	0.992977744	-0.00017814	0.020240506
MTHFD1	0.565546446	0.011500589	0.020014159
MYOM2	0.343226982	-0.018943889	0.019987045
NAV2	0.992285217	0.00019293	0.019953194
NETO2	0.138130776	0.029501295	0.019895826
NHSL1	0.913595924	0.002173091	0.020027781
NUDT12	0.400931634	-0.016787799	0.019986466
NXPH1	0.840757915	0.004028197	0.020048344
OXTR	0.229058416	0.024023348	0.019973049
PCDHGA4	0.190823705	0.026114059	0.019962639
PDE4C	0.351437649	0.018751882	0.020124272
PFDN1	0.096931444	0.03320262	0.020002598
PRKCQ	0.277179108	0.021676679	0.019947651
PTCHD3	0.361575109	-0.018196431	0.019944186
RAB3Ca	0.712338829	0.007406633	0.020087584
RAB3Cb	0.621374143	0.009929851	0.020104829
RABIF	0.159838577	0.028282623	0.020121173
RGS2	0.547686274	0.011935226	0.019851326
RNF216L	0.836113333	0.004177228	0.020192766
SARM1	0.124975135	0.030868802	0.020120169
SCYL1	0.813336194	0.004739042	0.020070105
SGEF	0.687859475	0.008097946	0.020156092
SLAIN1	0.528171702	0.012658068	0.020066715
SNRPN	0.887211913	0.00283658	0.019999444
SOCS3	0.940459415	-0.001503304	0.020126633
SORCS2	0.208087526	0.02523878	0.020049203
T	0.06099941	0.037496029	0.020013906
TBCCD1	0.74568301	-0.006500714	0.020043081
TBP	0.053654375	0.038586832	0.019997162
TMEM17	0.688208767	0.008048574	0.020056893
TRIM72	0.88554187	0.004250734	0.029529748
VENTX	0.13673696	0.029921666	0.020107929
XIST	0.737410414	-0.005439433	0.016223324
ZNF345	0.839358665	0.00404293	0.019944016

ZNF549	0.684729808	0.00814746	0.02006677
ZNF577	0.954753423	0.00113814	0.020059418
ZNF66	0.243316068	0.023459221	0.020106664

Overall none of these probes achieves Bonferoni significance. The only probe achieving nominal significance ($p<0.002$) after meta analysis of all 3 cohorts is MIA3, encoding melanoma inhibitory activity family, member 3. SNPs in this gene are associated with cardiovascular artery disease and myocardial infarction (Li X, et al Meta-analysis identifies robust association between SNP rs17465637 in MIA3 on chromosome 1q41 and coronary artery disease. Atherosclerosis. 2013 Nov;231(1):136-40.)

Birth weight genetic risk score analysis

THE ANALYSIS OF THE ASSOCIATION OF LOWER BIRTH WEIGHT RISK SCORE WITH LATER LIFE ADIPOSITY MEASURES:

Analysis done for cohorts: NFBC1966 (n=4912) and Hertfordshire Cohort Study (n=2397). A genetic risk score for lower birth weight was calculated using 17 birth weight SNPs which were available genotyped in HCS. The association of these individual SNPs with later life adiposity measures were also examined. The selected responses were BMI, weight, waist circumference and hip circumference.

Grs= genetic risk score for birthweight

RESPONSE: BMI	NFBC1966 (n=4912)			HCS (n=2397)		
	B	SE	p	B	SE	p
grs	-0.00225	0.00080	**0.00476**	0.00088	0.00112	0.43195

Betas, standard errors and p-values for the association with adult bmi. Adjusted for sex and age. P-value bold, if <0.05 after Bonferroni correction.

RESPONSE: WEIGHT	NFBC1966 (n=4912)			HCS (n=2397)		
	B	SE	p	B	SE	p
grs	-0.17879	0.06160	**0.00372**	0.05650	0.08752	0.51867

Betas, standard errors and p-values for the association with adult weight. Adjusted for sex, age and height. P-value bold, if <0.05 after Bonferroni correction.

RESPONSE: WAIST CIRCUMFERENCE	NFBC1966 (n=4912)			HCS (n=2397)		
	B	SE	p	B	SE	p
grs	-0.12727	0.05629	**0.02381**	0.04650	0.08235	0.57238

Betas, standard errors and p-values for the association with adult waist circumference. Adjusted for sex, age and height. P-value bold, if <0.05 after Bonferroni correction.

RESPONSE: HIP CIRCUMFERENCE	NFBC1966 (n=4912)			HCS (n=2397)		
	B	SE	p	B	SE	p
grs	-0.00131	0.00039	**0.00085**	0.00020	0.00059	0.73720

Betas, standard errors and p-values for the association with adult hip circumference. Adjusted for sex, age and height. P-value bold, if <0.05 after Bonferroni correction.

The analysis of the association of lower birth weight risk score with adult blood pressure:

Analysis done for: NFBC1966 (n=4259), using the same risk score as earlier.

	Systolic			Diastolic		
Adjustments	**Beta**	**95 % CI**		**Beta**	**95 % CI**	
sex	-0.17	-0.3	-0.04	-0.08	-0.2	0.03
sex, birth vars	-0.19	-0.32	-0.06	-0.09	-0.2	0.03
sex, birth vars, adult vars	-0.14	-0.26	-0.02	-0.05	-0.17	0.06
sex, birth vars, adult vars, mother vars	-0.15	-0.28	-0.03	-0.05	-0.16	0.06

Betas and their 95 % confidence intervals for the effect of birth weight risk score on systolic and diastolic blood pressures at 31 years with different adjustments. Birth vars: birth weight, gestational age, birth length. Adult variables: bmi at 31 years, smoking status, alcohol consumption. Mother variables: parity, family SES, mother's BMI.

References:

Das S.: Genetic and environmental correlates of growth patterns leading to obesity. A thesis based on Northern Finland Birth Cohort Data, completed in 2014 (Imperial College London/University of Oulu), under primary supervision of Professor Marjo-Riitta Jarvelin.

Das S *et al*: Common variants at *LEPR and LEPROT, FTO, TFAP2B* and *GNPDA2* are associated with childhood growth trajectories, linking early growth with adulthood obesity and metabolism. Submitted, 2015.

Horikoshi M, ...Järvelin M-R, Timpson NJ, Prokopenko I, Freathy RM.: New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. Nat Genet. 45, 76-82, 2013. doi:10.1038/ng.2477. Epub ahead of print. (IF 35.532)

Kaakinen M, Sovio U, Hartikainen A-L, Pouta A, Savolainen MJ, Herzig K- H, Elliott P, De Stavola B, Läärä E, Järvelin M-R.: Life course structural equation model of the effects of prenatal and postnatal growth on adult blood pressure. J Epidemiol Community Health 2014;68:1161–1167. doi:10.1136/jech-2013-203661. (IF 3.294).

Marttinen P, Pirinen M, Sarin AP, Gillberg J, Kettunen J, Surakka I, Kangas AJ, Soininen P, O'Reilly PF, Kaakinen M, Kähönen M, Lehtimäki T, Ala-Korpela M, Raitakari OT, Salomaa V, Järvelin M-R, Ripatti S, Kaski S.: Assessing multivariate gene-metabolome associations with rare variants using Bayesian reduced rank regression. Bioinformatics. 2014 Jul 15;30(14):2026-34. doi: 10.1093/bioinformatics/btu140. Epub 2014 Mar 24. (IF 5.323)

Menni, C. *et al.* Metabolomic markers reveal novel pathways of ageing and early development in human populations. International journal of epidemiology 42, 1111-1119, doi:10.1093/ije/dyt094 (2013).

Richmond RC. *et al*. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). Hum Mol Genet. 2014 Dec 30. pii: ddu739. [Epub ahead of print]

Ronkainen J, Huusko T, Soininen R, Mondini E, Cinti F, Mäkelä KA, Kovalainen M, Herzig K-H, Järvelin M-R, Sebert S, Savolainen M, Salonurmi T.: Fat mass- and obesity-associated gene *Fto* affects the dietary response in mouse white adipose tissue. Nature Communications, in press 2015.

Vimaleswaran KS.....Järvelin M-R, Hyppönen E.: Association of vitamin D status with arterial blood pressure and hypertension risk: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2014; 2: 719–29.

The association of birth measures and later life metabolomics:

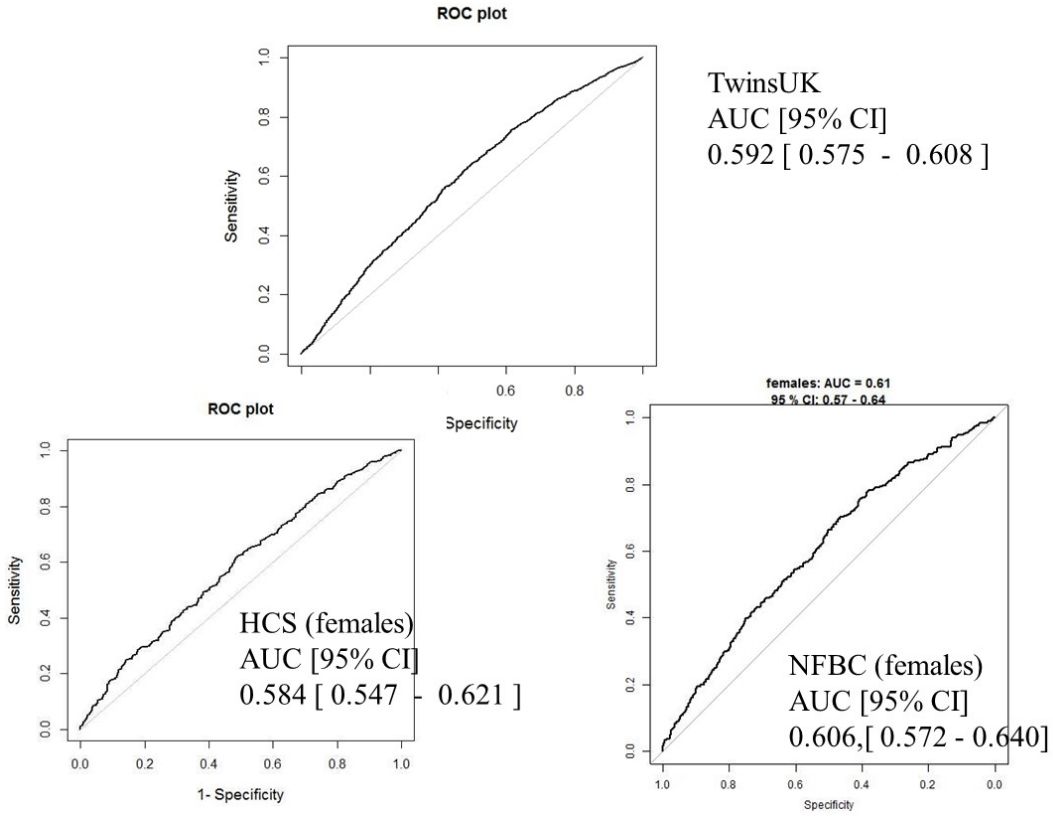
The effect of birth measures on 25 selected metabolites were examined in NFBC1986 (n=2391), HCS (n not available) and TwinsUK (n not available). The metabolites were selected by King's College London.

Meta-analysis: 12 metabolites significant after the meta-analysis with NFBC1986 (n=1170), HCS (n=1011) and TwinsUK (n=5036):

METABOLITE	**PValue_**	**Beta**	**SE**
Aspartate	0.0350	-0.0286	0.0136
Glycine	0.0001	0.0468	0.0122

g-Glutamylleucine	0.0003	-0.0476	0.0130
g-Glutamylvaline	0.0064	-0.0305	0.0112
Isovalerylcarnitine	0.0002	-0.0535	0.0142
Methylbutyroylcarnitine	7.08E-05	-0.0520	0.0131
Propionylcarnitine	0.0006	-0.0486	0.0141
Urea	4.63E-05	-0.0529	0.0130
Uric-acid	0.0002	-0.0547	0.0146
2-hydroxy-butyric-acid	2.05 E-06	-0.0625	0.0113
Glycosyltryptophan	0.0068	-0.0307	0.0113
3-4-hydroxyphenyllactate	0.0024	-0.0401	0.0132

A receiving operator characteristic curve was then performed to investigate how much were metabolites in adult life explained by low birth weight. ROC curves for birth weight metabolomic risk score in females from TwinsUK, NFBC86 and HCS. The response: low birth weight tertile. The explanatory variables: 11 metabolites listed in the meta-analysis table, excluding g-Glutamylvaline for severe collinearity, adjusted for bmi (see also table below).



These metabolites are also individually associated with blood pressure or insulin resistance in TwinsUK and HSC as the following table shows. A manuscript reporting these results is currently being drafted.

TwinsUK	**SBP**		**DBP**		**HOMA-IR**	
	Beta[95%CI]	**P**	**Beta[95%CI]**	**P**	**Beta[95%CI]**	**P**
2-hydroxy-butyric acid	0.59[0.08,1.09]	**2.32E-02**	0.22[-0.11,0.55]	**2.32E-02**	-0.04[-0.08,-0.01]	**2.35E-02**
3,4 - hydroxyphenyllactate	1.04[0.54,1.54]	**5.19E-05**	0.78[0.45,1.12]	**5.19E-05**	0.08[0.05,0.11]	**6.51E-08**
Aspartate	0.63[0.19,1.07]	**5.55E-03**	0.58[0.28,0.88]	**5.55E-03**	0.11[0.08,0.14]	**2.58E-13**
g-Glutamylleucine	0.36[-0.14,0.86]	1.59E-01	0.29[-0.05,0.63]	1.59E-01	0.1[0.07,0.14]	**3.57E-08**
g-Glutamylvaline	0.45[-0.04,0.94]	7.51E-02	0.49[0.16,0.82]	**7.51E-02**	0.15[0.11,0.18]	**6.56E-15**

glycine	-0.24[-0.75,0.28]	3.65E-01	-0.32[-0.64,0]	3.65E-01	-0.08[-0.11,-0.05]	**9.16E-08**
glycosyltryptophan	0.25[-0.28,0.79]	3.54E-01	0.2[-0.16,0.56]	3.54E-01	0.01[-0.02,0.05]	4.63E-01
Isovalerylcarnitine	-0.09[-0.57,0.39]	7.21E-01	0.06[-0.26,0.39]	7.21E-01	0.08[0.05,0.1]	**1.86E-07**
Methylbutyroylcarnitine	0.42[-0.07,0.92]	9.32E-02	0.39[0.06,0.72]	9.32E-02	0.11[0.07,0.14]	**2.48E-11**
Propionylcarnitine	0.09[-0.41,0.59]	7.30E-01	0.11[-0.22,0.44]	7.30E-01	0.09[0.06,0.12]	**4.13E-09**
urea	0.61[0.07,1.16]	**2.79E-02**	0.26[-0.08,0.6]	**2.79E-02**	0.04[0,0.07]	**3.32E-02**
uric Acid	1.09[0.57,1.6]	**3.36E-05**	0.68[0.34,1.02]	**3.36E-05**	0.1[0.07,0.13]	**6.40E-11**

HCS	**SBP**		**DBP**	
metabolite	**Beta**	**P**	**Beta**	**P**
3,4-hydroxyphenyllactate	1.58[0.41,2.74]	**8.32E-03**	0.65[0.04,1.27]	**3.75E-02**
aspartate	0.42[-0.74,1.58]	4.78E-01	-0.62[-1.23,-0.01]	**4.75E-02**
g-glutamylleucine	0.18[-1.29,1.65]	8.09E-01	0.65[-0.14,1.43]	1.07E-01
g-glutamylvaline	-0.77[-1.99,0.44]	2.14E-01	-0.13[-0.78,0.52]	6.98E-01
glycine	0.68[-0.54,1.9]	2.72E-01	0.25[-0.38,0.89]	4.33E-01
glycosyltryptophan	1.76[0.53,3]	**5.24E-03**	-0.01[-0.67,0.64]	9.65E-01
hydroxybutyric acid	0.81[-0.3,1.92]	1.52E-01	0.12[-0.49,0.73]	6.98E-01
isovalerylcarnitine	-0.06[-1.25,1.13]	9.21E-01	0.06[-0.59,0.72]	8.51E-01
methylbutyroylcarnitine	-0.16[-1.39,1.06]	7.95E-01	-0.04[-0.67,0.6]	9.07E-01
propionylcarnitine	-0.72[-1.92,0.48]	2.41E-01	-1.1[-1.75,-0.46]	**7.94E-04**
urea	0.68[-0.49,1.84]	2.56E-01	0.08[-0.54,0.71]	7.94E-01
uric acid	1.27[0.17,2.36]	**2.33E-02**	0.99[0.43,1.55]	**5.89E-04**

Summary of WP6 :

1-The common pathways related to epigenetics and genetic associations with birthweight in twins highlight molecular pathways related adult height. The pathways related only to epigenetics are linked to various adult diseases including cardiovascular disease

2- Joint analyses of singletons and twins found consistency in associations with birthweight within singletons for all three technologies (genetics, epigenetics, metabolomics) but only for metabolomics between twins and singletons. Genetic results in singletons are consistent an a genetic risk score for birthweight correlates with measures of metabolic disease and blood pressure in adults A manuscript summarising these results is currently in preparation.

3- Epigenetic analyses resulted in only one gene being nominally (but not Bonferroni) significant between twins and singletons. This is the MIA3 gene which is known to be associated with cardiovascular disease

4- Metabolomics results in all three cohorts are consistent, and a metabolomic signature in adults is seen related to birthweight (AUC= 0.6). The metabolites associated with birthweight are also associated with blood pressure and insulin resistance in adults. A manuscript summarising these results is currently in preparation.

WP7- Data Integration and Health Benefit Evaluation

The main objective of Tasks 7.1-7.3 is to implement an integrated platform that will make available to its users an extensive database of the correlations between early life events and ageing outcomes linking datapoints to methylation, metabolomic, genetic and post-translational modification (PTM) results, allowing future comparison with data derived from other studies. In order to achieve this the following steps/objectives have to be completed:

1. Study the available data from the cohorts and identify what kind of information and datasets will be integrated into the database **[DONE]**
2. Define the User and Technical Requirements and Applications needed **[DONE]**
3. Design the System Architecture and Mock-ups of the applications' GUI **[DONE]**
4. Develop, Integrate and test the applications of the system and database **[DONE]**
5. Define traits and then start using the platform and evaluate the results **[DONE]**

Summary: The main objective of this WP was to evaluate the impact of the identified biomarkers on health benefit and the possibility of their use within the context of the ageing European population and given the prevalence of the age-related diseases under study by EurHEALTHAgeing. In WP7, an integrated platform was implemented in order to help the researchers to compare the data derived from EurHealth studies with those that come from other studies based on the correlations between early life events and ageing outcomes linking datapoints to methylation, metabolomic, genetic and post-translational modification (PTM) results. All tasks have been completed as planned in the original workplan.
Task 7.1 (Requirements Analysis) has been completed and the corresponding deliverable has been submitted on time. The state of the art of existing protocols and data formats related to each partner's dataset has been studied and used in order to define the correct data structures for the SW to be developed.

This input was used in Task 7.2 (Static and Dynamic MockUps) in order to design the Graphical User Interface (GUI) of the software and use this to collect the "users' feedback".

Task 7.3. (Technical Implementation, Testing and Deployment) is the actual EurHealthAgeing platform running on a web server. The users can access the database through a web browser and the credentials that have been provided to them. The corresponding deliverable D7.3 serves as the user manual of EurHealthAgeing Portal that contains all essential information for the user to make full use of the system.

In Task 7.4. (System Operation and Evaluation) the users completed the population of the database using the system provided services and applications.

The operation of the platform services was evaluated based on the evaluation methodology and criteria that were defined in the internal deliverable of T7.2.

Task 7.1. Requirements Analysis:

The main focus of this task was to study the available data from the cohorts and identify what kind of information and datasets will be integrated into the database thus providing an encompassing set of requirements for the structures and data that was used. Moreover all the valuable sources for the different domains involved in the project in accordance with the user requirements were identified and incorporated in our system.
Task 7.2 .Design the System Architecture and Design Mock-ups of the applications' GUI
n this Task, the state of the art of existing protocols and data formats related to each partner's dataset was studied. Furthermore we examined the existing technical state of the art and tried to manipulate them according to our platform's needs. In this Task we defined the key building blocks of the architecture which were the basis for the technical development within T7.3. During T7.2 the system's architecture and Database Schema were defined while the GUI was formed using static MockUps of the EurHealthAgeing Portal.
Task 7.3. Technical Implementation, Testing and Deployment
By using the available information and results driven from the previous WPs we developed our platform and the meta-database that stores the results of EurHealthAgeing studies. During this task, the implementation of the database schema, the web page development and the web server configuration took place in order to deliver the final platform to the users. The corresponding deliverable D7.3 includes a description of the system functions and capabilities, contingencies and alternate modes of operation, and step-by-step procedures for system access and use.
Task 7.4. System Operation and Evaluation
During this task users populated the database with their data and corresponding findings using the system provided services and applications. This was an opportunity for the users to actually perform a hands-on evaluation of the provided functionalities, ease-of-use and the other criteria as described in the evaluation criteria internal document. The outcome of this evaluation is summarized in D7.4 Evaluation Report

WP10- **Ethics**
During the length of the project no ethical issues have arisen. The external ethical advisor reviewed all the relevant documentation and found it satisfactory. She has participated at the project miterm and final reviews and assessed that for each workpackage and each participant institution proper ethical and data management protocols and procedures are in place and being performed. She has drafted a final report.