 doi:10.20944/preprints201809.0082.v1

Article

Transcriptomics as precision medicine to classify *in vivo* models of dietary-induced atherosclerosis at cellular and molecular levels

Alexei Evsikov [1,2], Caralina Marín de Evsikova [1,2]*

[1] Epigenetics & Functional Genomics Laboratory, Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, Florida, 33612, USA;
[2] Department of Research and Development, Bay Pines Veteran Administration Healthcare System, Bay Pines, FL 33744, USA

* Correspondence: cmarinde@health.usf.edu; Tel.: +1-813-974-2248

Abstract: The central promise of personalized medicine is individualized treatments that target molecular mechanisms underlying the physiological changes and symptoms arising from disease. We demonstrate a bioinformatics analysis pipeline as a proof-of-principle to test the feasibility and practicality of comparative transcriptomics to classify two of the most popular *in vivo* diet-induced models of coronary atherosclerosis, apolipoprotein E null mice and New Zealand White rabbits. Transcriptomics analyses indicate the two models extensively share dysregulated genes albeit with some unique pathways. For instance, while both models have alterations in the mitochondrion, the biochemical pathway analysis revealed, Complex IV in the electron transfer chain is higher in mice, whereas the rest of the electron transfer chain components are higher in the rabbits. Several fatty acids anabolic pathways are expressed higher in mice, whereas fatty acids and lipids degradation pathways are higher in rabbits. This reflects the differences between two translational models of atherosclerosis. This study validates transcriptome analysis as a potential method to precisely identify altered cellular and molecular pathways in atherosclerotic disease, which can be used to individualize treatment even in the absence of genetic data.

Keywords: atherosclerosis, coronary aortic disease, gene set enrichment analysis, heart disease, Apoe mouse, transcriptomics, RNA-seq analysis, pathway enrichment analysis, mouse, precision medicine, New Zealand White rabbit

1. Introduction

Precision medicine is the ability to classify individuals according to their underlying susceptibility, prognosis, or targeting potential treatment response. Unlike DNA sequencing technology that focuses on the genome, RNA sequencing produces the snapshot of the full transcriptome, and has the capability to fulfill precision medicine to classify patients at both molecular and cellular levels. Development of RNA sequencing pipelines is important for implementation of transcriptomics as precision medicine [1], which can be used successfully to classify patient attributes and predict therapeutic response and ultimate outcomes. Classifying patients based on symptoms is limited because symptoms often arise from numerous origins or multimodal pathways, as the case with atherosclerosis.

Atherosclerosis is a costly disease in the United States, at $9 billion per year in hospitals stays [2], and its related morbidities, such as heart attack and stroke, totaling for $43.5 billion of total

hospital costs per year [3,4]. Atherosclerosis is a silent disease as initially there are no symptoms as the artery narrows from the gradual accumulation of plaques, which consist mainly of fat, cholesterol and calcium, and often harbor bacteria [5]. While its etiology is complex, inflammation, arising either from lifestyle factors like stress, obesity, illness, or allergens, is currently proposed as one of the initial triggers for atherosclerosis [6]. The current working model suggests plaques may build up in the arterial epithelial wall after damage; these plaques harden, narrowing the arteries and restricting blood flow. As oxygenated blood flow decreases over time, by middle age symptoms begin to emerge, depending upon the location of atherosclerotic plaques, which provoke stroke, peripheral artery disease, kidney problems, heart disease and coronary artery disease [5]. While the exact cause underlying atherosclerosis is unknown, there are many associated risk factors that increase its likelihood because of damage inflicted to arterial epithelial lining. These risk factors are smoking tobacco products, diet, age, family history and genotype [7-9]. Notably, many of factors are related to metabolism and energy regulation, such as excessive body weight, obesity, elevated circulating glucose from either insulin resistance, pre-diabetes, and diabetes, suggesting that energy balance and regulation is a necessary, but not critical, component in triggering atherosclerosis [10]. In fact, as childhood obesity rates have risen during past few decades, likewise the incidence of atherosclerosis in youth increased [11].

Current research focuses on the molecular, cellular, and physiological origins of atherosclerosis and its pathology. Fundamental questions focus on environmental and genetic triggers proximal and ultimate causes inducing artery damage, the development of plaques and its dynamic remodeling that may lead to rupture and formation of blood clots. These vascular events cause two of the major morbidities and mortalities consequences of atherosclerosis, ischemic stroke and heart attack. Given the complex, multimodal disease, one needs reliable model systems to replicate and experimentally test concepts and new therapeutics based on emerging knowledge of the integrated systems underlying it ultimate cause and proximal mechanisms inducing it pathology and symptomology.

Biomedical researchers in both clinical and basic settings need to choose models that recapitulate the specific characteristics of disease, and its pathology, under scientific scrutiny. Transcriptomics is a robust method to measure the common and unique pathways among different translational models. Depending upon the hypothesis and biomedical question, researchers need to choose a model system to detect changes in the target molecular and cellular pathways. Thus, transcriptomics can classify individual and simultaneously facilitate discovery, testing, and validation of new therapeutics for patients with specific characteristics at cellular and molecular levels one needs to choose a system to detect changes in the target molecular, cellular and physiological pathways.

We developed a resource guide for transcriptomics and bioinformatics to use gene expression levels to substantiate biochemical, biological, cellular, molecular, and physiological changes across clinical, experimental, and model systems [1]. In this current study, we employed our guide as a proof-of-principle example for suitability, feasibility, and practicality of comparative transcriptomics to detect and evaluate gene expression overlap to reveal both common and unique biological, cellular, and molecular pathways and gene networks in a translational model between two species, apolipoprotein E (*Apoe*) null mice and New Zealand White rabbits, of coronary atherosclerosis induced by high fat and high cholesterol diets.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

2. Materials and Methods

2.1. Experimental models

2.1.1. Mice

RNA-seq data used in this study were published (BioProject ID: PRJNA371776; [12]), and we used control mouse samples only (experiment IDs: SRX2544726, SRX2544727, SRX2544728). The following brief description of samples is from the original report [12]. The male mice have mixed genetic background of C57BL/6J, C57BL/6N, 129S4 and FVB/N due to breeding in of multiple transgenes and floxed alleles, and their genotype is *Col15a1*$^{wt/wt}$, *Myh11*-CreER^{T2}, ROSA26-STOP^{flox}-eYFP, *Apoe*$^{-/-}$. Six-week-old male mice were treated with tamoxifen to induce CreER^{T2} translocation to the nucleus in *Myh11*-CreER^{T2}-expressing tissues, where it excises the stop in ROSA26-STOP^{flox}-eYFP [13] locus allowing for expression of YFP. Absence of *Apoe* (from B6.129P2-*Apoe*^{tm1Unc}/J, Jax® Mice Stock No: 2052) leads to marked increase in total plasma cholesterol levels that are unaffected by age or gender [14], which makes B6.129P2-*Apoe*^{tm1Unc}/J mice a popular model in atherosclerosis research. Mice were placed on a Western diet consisting of 21% milk fat and 0.15% cholesterol for 18 weeks. Mice were euthanized by CO_2 inhalation, and their brachiocephalic arteries, aortic arch, and carotid arteries dissected and flash frozen in liquid nitrogen. Total RNA was extracted using TRIzol, and sequencing library prepared using Illumina kit with ribosomal reduction and strand specificity.

2.1.2 Rabbits

RNAseq data were published (BioProject ID: PRJNA274427; [15]). In our study, we used high fat, high cholesterol-fed New Zealand White (NZW) rabbits aorta RNAseq data only (experiment IDs: SRX864779, SRX864780, SRX864782, SRX864783) [15]. Briefly, four NZW rabbits were fed with a cholesterol-rich diet containing 0.3% cholesterol and 3% soybean oil for 16 weeks. Aortic arches were collected for RNA extraction. For the library preparation, 3 μg total RNA was used. Library preparation was performed with the TruSeq RNA LT V2 Kit.

There was no information about the sex of the animals for these RNAseq data. To identify sex, we used the genomic sequence from OryCun2.0 genome assembly harboring *Xist* gene, and an EST (CU464548) from rabbit pre-implantation embryo SSH library [16] corresponding to the transcript of Y-chromosome-linked *Ddx3y* gene, to query rabbit RNAseq datasets using SRA BLAST [17]. The gene, *Ddx3y*, being Y-linked, is expressed exclusively in males, whereas *Xist* is expressed from inactive X-chromosome in females [18]. All four datasets contained *Ddx3y* reads (average 350 reads per dataset), and did not contain reads corresponding to *Xist*. From these data, we concluded all four RNAseq datasets were for aortic arches from male rabbits.

2.2 RNA-seq analysis

2.2.1 Overall transcriptomics strategy

Our analysis (Figure 1) is based on a collection of robust, publicly available tools for deep transcriptome analysis; most were created to be used by researchers with only moderate bioinformatics experience. We provide detailed description of each step below.

2.2.2 Genome alignments

FASTQ sequence data were downloaded from the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena). Sequence alignments were performed using RNA STAR [19] tool from

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

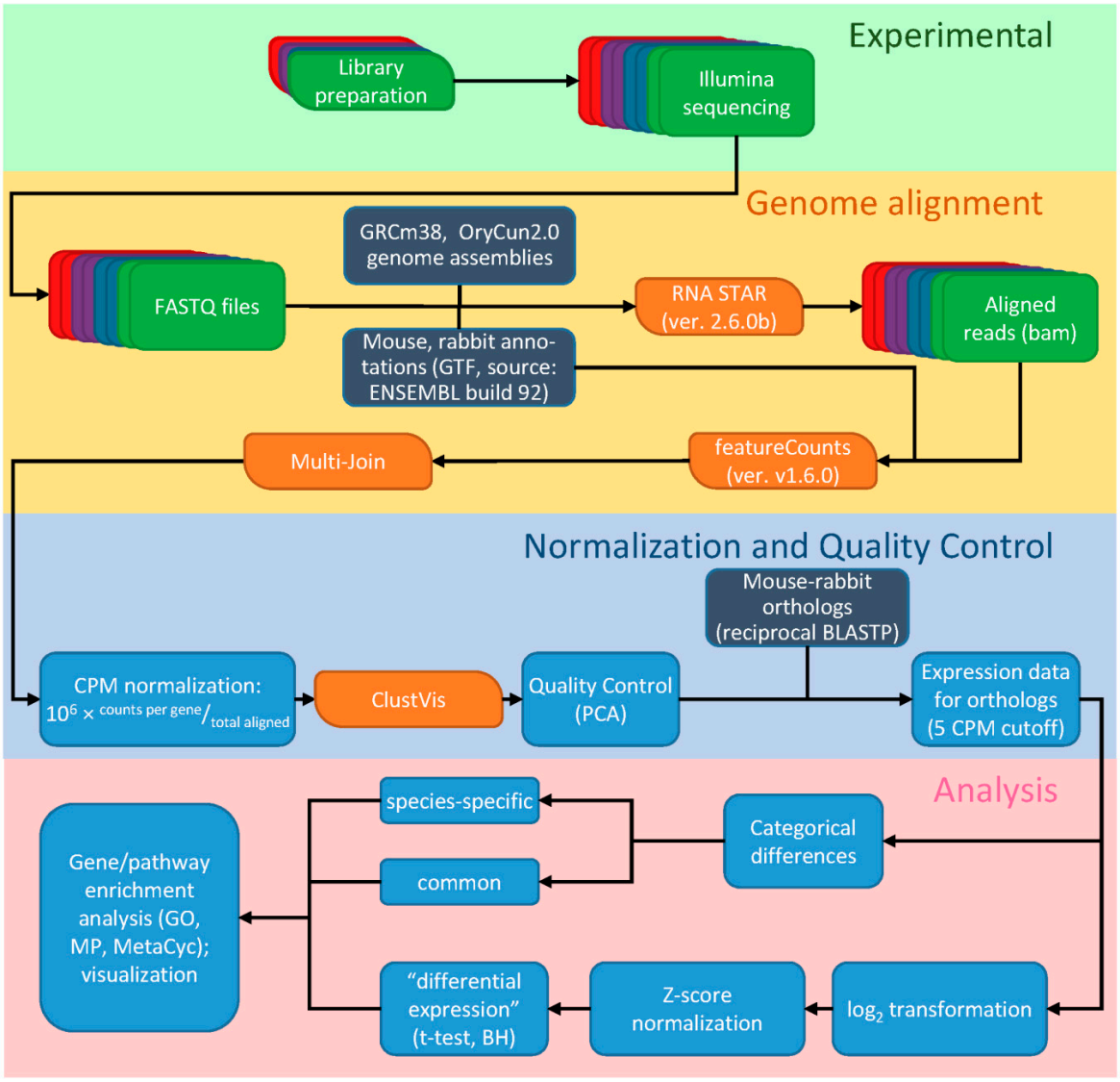

Figure 1. Gene Expression Analysis Pipeline. RNAseq data for each biological replicate of mouse and rabbit aortas were aligned to the GRCm38 (mouse) and OryCun2.0 (rabbit) genome assemblies using STAR [19]. Datasets of splice sites for alignment of spliced reads were obtained from ENSEMBL. Gene expression data were extracted from the files of aligned reads using featureCounts [21], normalized to CPM values, and quality control of replicates was performed with ClustVis [23]. Expression data for 15,179 established orthologous genes were extracted, and analyzed in two ways. In the first approach, we categorized all orthologs by expression status into three groups, expressed in mouse, expressed in rabbit, or expressed in both species. In the second approach, transformed expression data were used to test differences in gene expression levels between mice and rabbits.

within Galaxy platform [20]. Reads were aligned to the respective reference genomes (GRCm38, a.k.a. mm10, for mouse data; OryCun2.0 for rabbit data) using the following parameters: RNA STAR version: 2.6.0b; single-end or paired-end reads: paired; gene model (gff3, gtf) file for splice junctions: yes (see below); length of the genomic sequence around annotated junctions: 100; count number of reads per gene: false; additional output parameters (formatting and filtering): no; other parameters: default. Gene model files for splice junctions, i.e. coordinates of known mouse and rabbit transcripts (in GTF format), were downloaded from ENSEMBL ftp site (release 92). To calculate gene expression, we counted the numbers of reads aligned to regions flagged as exons in GTF files using the program featureCounts [21] version 1.6.0.6 from within Galaxy using the following parameters: gene annotation file: yes; output format: Gene-ID "\t" read-count; create gene-length file: False; count fragments instead of reads: Disabled; only allow fragments with both reads aligned: False; exclude chimeric fragments: True; GFF feature type filter: exon; GFF gene identifier: gene_id; report on feature level: False; allow read to contribute to multiple features: False; count multi-mapping

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

reads/fragments: Enabled; assign fractions to multi-mapping reads: True; minimum mapping quality per read: 12; minimum bases of overlap: 30.

2.2.3 Data normalization and quality control

To account for the impacts depth of sequencing, which affects read numbers of individual transcripts, we use normalized expression data, specifically counts per million (CPM) [22], to perform quality comparison of datasets. To calculate CPM values, we used the following formula:

$$E_{g,s} = 1{,}000{,}000 \times C_{g,s} / T_s$$

where $E_{g,s}$ is a CPM value of a gene in a biological replicate; $C_{g,s}$ is the number of reads mapping to all exons of this gene in this biological replicate; T_s is the total number of reads aligned (anywhere in the genome) from this biological replicate (i.e., the number of aligned reads in RNASTAR output "binary alignment map" bam files). This procedure also transforms data from counts to continuous scale.

For quality control, we used ClustVis [23], a tool for clustering of complex data such as RNAseq, based on principal component analysis, and visualization of results. Any samples that fall outside of 95% confidence interval on two-dimensional PCA plot are flagged as outliers and removed from further analysis.

2.2.4 Data filtering and transformation

RNAseq data requires a cutoff threshold to avoid numerous false positives caused by over-dispersed values at the low CPM values and we used a cutoff of 5 CPM, corresponding to a 2 FPKM threshold typically used in RNAseq analysis pipelines [22]. To filter out the genes with extreme low expression, specifically less than 5 CPM expression, we use average CPM values for a gene across all mice, and again for rabbit, samples. The union of two gene lists was used for further analysis. For data transformation, we use logarithm to the base of 2 for the CPM values. To avoid the logarithm of zero exception, all zero values are replaced with a minimal non-zero value in a given RNAseq dataset [24] (i.e., CPM value corresponding to the read count of 1). For calculating Z-scores, we used the formula

$$z_{g,s} = (x_{g,s} - \mu)/\sigma$$

where $x_{s,g}$ is a logarithm to the base of 2 of the CPM value of a gene (g) in a sample (s), and μ and σ are the average and standard deviation, respectively, of the logarithm to the base of 2 for geometric means of CPM values for each gene across all mouse or rabbit samples [25,26].

2.3 *Establishment of mouse-to-rabbit orthology.*

To establish phylogenetic relationship among mouse and rabbit genes, we have downloaded all mouse protein sequences, and all rabbit protein sequences, from ENSEMBL ftp site (release 92) [27]. We used ENSEMBL as a source because each protein sequence in ENSEMBL is cross-annotated to a corresponding gene, which ensures precise ID mapping. Using reciprocal BLAST approach, we performed 1) BLASTP comparison of each rabbit protein to all mouse proteins, 2) extracted the top result in each search (i.e., mouse protein which is a potential homolog) and compared it to all rabbit proteins, 3) extracted the top result in this search, 4) translated protein IDs to ENSEMBL gene IDs. BLAST analysis was performed using a stand-alone NCBI BLAST+ package (version 2.7.1) for Windows [28]. One-to-one homology indicates orthology, and mouse and rabbit genes were flagged as orthologs, if and only if, the genes in steps 1) and 3) above were the same, and no mouse gene in step 2) paired with more than one rabbit gene. Finally, each ENSEMBL gene ID of a mouse ortholog was converted to Mouse Genome Informatics (MGI; [29]) gene ID using MGI Batch Search tool [30]. We established 15,179 orthologous pairs among mouse and rabbit genes (Supplementary Table 1). Expression data for these 15,179 orthologs were used in all comparative studies.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

2.4 *Statistical and Gene/Pathway Enrichment Analyses.*

2.4.1 Categorization by expression status

To split genes in three groups by their expression status (present in mouse only, present in rabbit only, or present in both species), we generated the list of genes that are expressed at a level of at least 5 CPM in at least one species (averaged across samples). Within this list, genes which were expressed at the level of 5 CPM or higher in mice, but less than 5 CPM in rabbits, were categorized as "mouse-only"; genes expressed at the level of 5 CPM and higher in both species were categorized as "common"; and the rest were categorized as "rabbit-only".

2.4.2 Statistical analysis of gene expression differences

To compare two different animal models of atherosclerosis using data from different laboratories, RNAseq data were normalized and transformed as described above. For gene expression analysis, we used t-test module of scipy.stats [31] in Python on z-score data to find significant expression differences. To control for multiple testing, we used Benjamini-Hochberg correction [32] implemented in StatsModels package for Python [33]. For downstream analysis, genes with FDR $q < 0.1$ and t-test $p < 0.01$ were considered significantly different.

2.4.3 Visual Annotation Display (VLAD) analysis

VLAD, accessible via MGI web portal, is a powerful tool to find common functional themes in the lists of genes by analyzing statistical over- or underrepresentation of ontological annotations [34]. Currently, users can choose among Gene Ontology (GO) [35] and Mammalian Phenotype Ontology (MP) [36] annotations for mouse genes, Gene Ontology annotations for human genes, or upload a file of own annotations (in open biomedical ontology [37] 'obo' format). Unlike other packages for ontological enrichment, VLAD allows analysis of more than one query (i.e., several lists of genes may be analyzed and visualized simultaneously), as well as permits user to provide own "universe set", i.e. gene list to test queries. For our studies, the "universe set" was the list of all orthologous gene pairs (Supplementary Table 1). For GO analysis, we searched for overrepresentation among terms with experimental evidence (i.e., codes EXP, "Inferred from experiment"; IDA, "Inferred from direct assay"; IMP, "Inferred from mutant phenotype"; TAS, "Traceable author statement"). For MP categories, we searched only among terms with the following evidence codes: IMP, "inferred from mutant phenotype"; TAS, "stated by author"; and EE, "shown by experimental evidence".

2.6 BioCyc analysis

BioCyc is a collection of Pathway/Genome Databases (PGDBs), which link biochemical pathways, reactions, and compounds with genes and proteins on the species level, as well as software tools to analyze these connections [38]. We used MouseCyc database [39] available on MGI portal to analyze gene lists via Metabolism → Cellular Overview → Omics Viewer tool of MouseCyc.

3. Results

3.1. Overall statistics of datasets.

On average, 80% of rabbit reads, and 90% of mouse reads, aligned to OryCun2.0 and GRCm38 assemblies, respectively (Supplementary Table 2). Lower alignment rate for rabbit samples is likely due to incomplete coverage of the genome in OryCun2.0 assembly. Of 53,801 mouse genes represented in the mouse GTF file, 29,338 genes had coverage of at least one read in at least two of the three RNASeq datasets. For rabbit data, of 23,669 genes, 16,307 had coverage of at least one read in at least two of the four RNASeq datasets. Count data for each sample and species (Supplementary

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

Tables 3 and 4) were filtered to include only orthologous genes between mouse and rabbit species, because this step allows direct comparison between the models of atherosclerosis. Count data for each gene were normalized to the total number of aligned reads per sample, and count per million (CPM) values were used to quality control these RNAseq samples using ClustVis [23]. As expected, this analysis (Figure 2) revealed largest variability among data (Principal Component 1) arising from species differences, while Principal Component 2 mainly reflects variability among biological replicates. We detected no outliers among the samples and proceeded with further analysis.

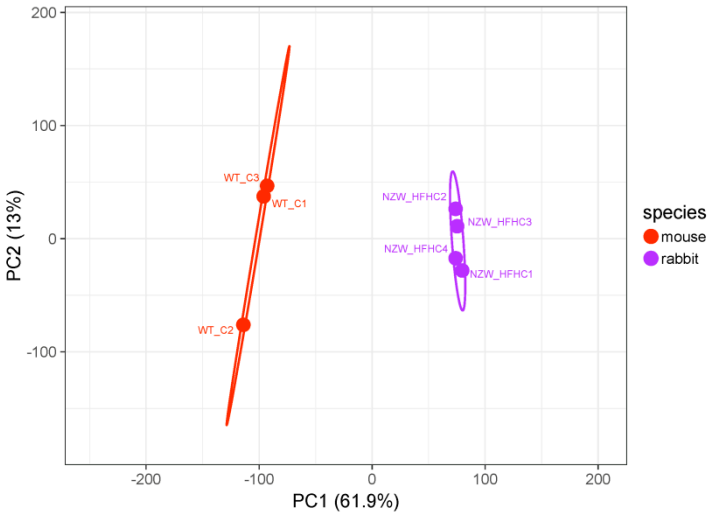

Figure 2. Principal Component Analysis of gene expression data. Principal components were calculated by singular-value decomposition. X axis (Principal Component 1) and Y axis (Principal Component 2) account for 61.9% and 13% of variation, respectively. Prediction ellipses denote probability at 0.95, a new observation from the same group will fall inside the ellipse. N = 7 gene expression datasets.

3.2. Categorization of gene expression.

RNAseq data tend to be over-dispersed at very low CPM values, and requires certain cutoff threshold to avoid numerous false positives [22]. We chose a cutoff of 5 CPM, which approximately corresponds to 2 FPKM threshold typically used in RNAseq analysis pipelines. When applied to data, we were able to categorize genes in three distinct groups (Figure 3), those whose expression is present in mice and absent in rabbits (1,337 genes); those with common expression (7,172 genes); and those present in rabbits but absent in mice (1,218 genes). The fact that 75% of genes are common reflects similarity of gene expression programs in the aortas between these two models of atherosclerosis. Complete list of categorized genes is in Supplementary Table 5.

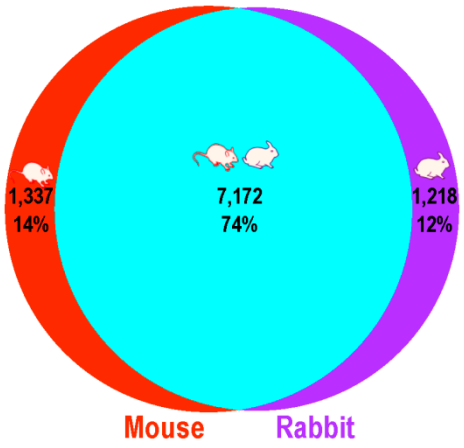

Figure 3. Categorization of genes by expression status. Among orthologs expressed at an average ≥5 CPM threshold in at least one species, 7,172 genes were designated as "common"; 1,337 genes whose expression was above threshold in mice, but below threshold in rabbits were designated as "mouse-only"; and 1,218 genes whose expression was above threshold in rabbits, but below threshold in mice were designated as "rabbit-only".

To identify meaningful pathways among common and species-specific lists, we analyzed the lists for enrichment of specific Gene Ontology annotations using VLAD application [34] (Figure 4A-C). We identified a total of 472 significantly overrepresented Biological Process, 167 Cellular Component, and 21 Molecular Function categories ($p < 0.01$, $q < 0.1$; Supplementary Table 6). Most of the overrepresented Biological Process and Cellular Component categories (428 and 99, respectively), and all Molecular Function categories, were from the common expression group. All top 25 categories in Biological Process category are related to "GO:0008152 Metabolic process" category, being either more specific, descendant metabolic related-categories or regulators of these metabolic processes (Figure 4A). Interestingly, among common genes in this category, one-third are associated with the "MP:0002127 abnormal cardiovascular system morphology" phenotype (see below), which is significantly higher than expected. Among top Cellular Component categories, several were for species-specific groups. Surprisingly, among mouse-only genes, categories "GO:0097458 neuron part" and "GO:0044456 synapse part", were significantly overrepresented (Figure 4B). Seventeen of these genes (*Add2*, *Cacna1b*, *Dagla*, *Kcna1*, *Ldlr*, *Mapk10*, *Mapt*, *Ngf*, *Ngfr*, *Nrcam*, *Nrxn1*, *Prom1*, *Scg2*,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

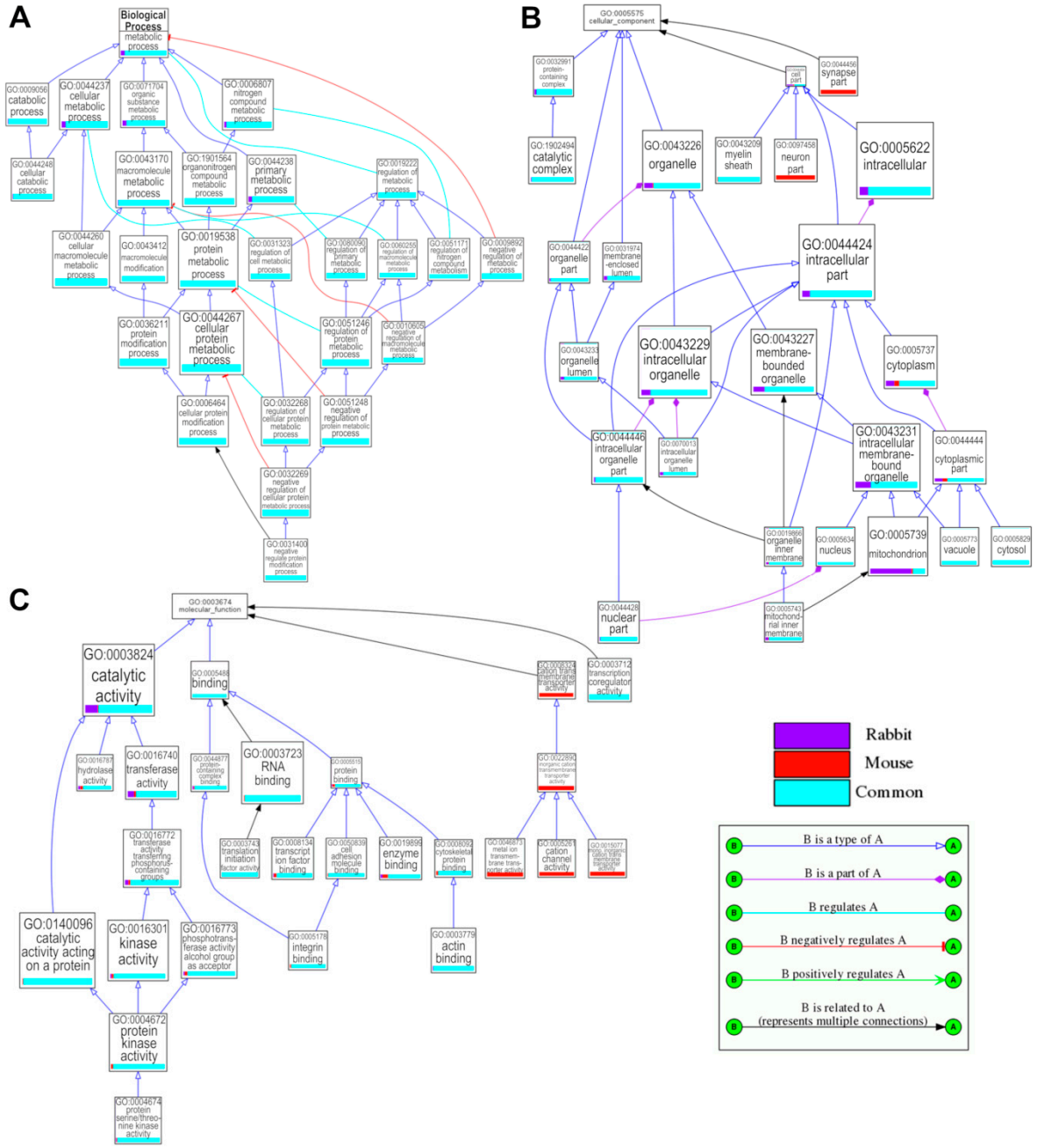

Figure 4. Gene Ontology (GO) enrichment analysis for mouse, rabbit, and common using categorization of expression levels. For each GO module, Biological Process (A), Cellular Component (B) and Molecular Function (C), only the top 25 significant terms with lowest p-values are shown. The box size reflects its relative statistical significance with the largest box with the lowest p value and the colored bar within the box indicates the proportion of contribution to a specific gene set (purple: Rabbit, red: Mouse, blue: Common). Arrows connecting boxes represent different types of relationship among GO terms. For more detail and interactive module, see Supplemental Table 6 and Supplemental HTML1.

Snap25, *Syt1*, *Uchl1*, *Uhmk1*) are also associated with "MP:0002127 abnormal cardiovascular system morphology" category. Moreover, corresponding mouse-only genes were also enriched in "GO:0022008 neurogenesis" category, its descendants, and related processes, such as "GO:0007411 axon guidance" and "GO:0048812 neuron projection morphogenesis" (Supplementary Table 6). Among rabbit-only genes, the overrepresented category is "GO:0005739 mitochondrion" and its descendants (Figure 4B, Supplementary Table 6).

We have also explored Mammalian Phenotype (MP) Ontology annotations using the same strategy (Figure 5). A total of 1,049 MP categories were significantly overrepresented, again most of them (1,000) were in the common genes group (Supplementary Table 7). As expected, "MP:0005385 cardiovascular system phenotype" and it's descendant category, "MP:0002127 abnormal cardiovascular system morphology" discussed above, were among top 25 overrepresented

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

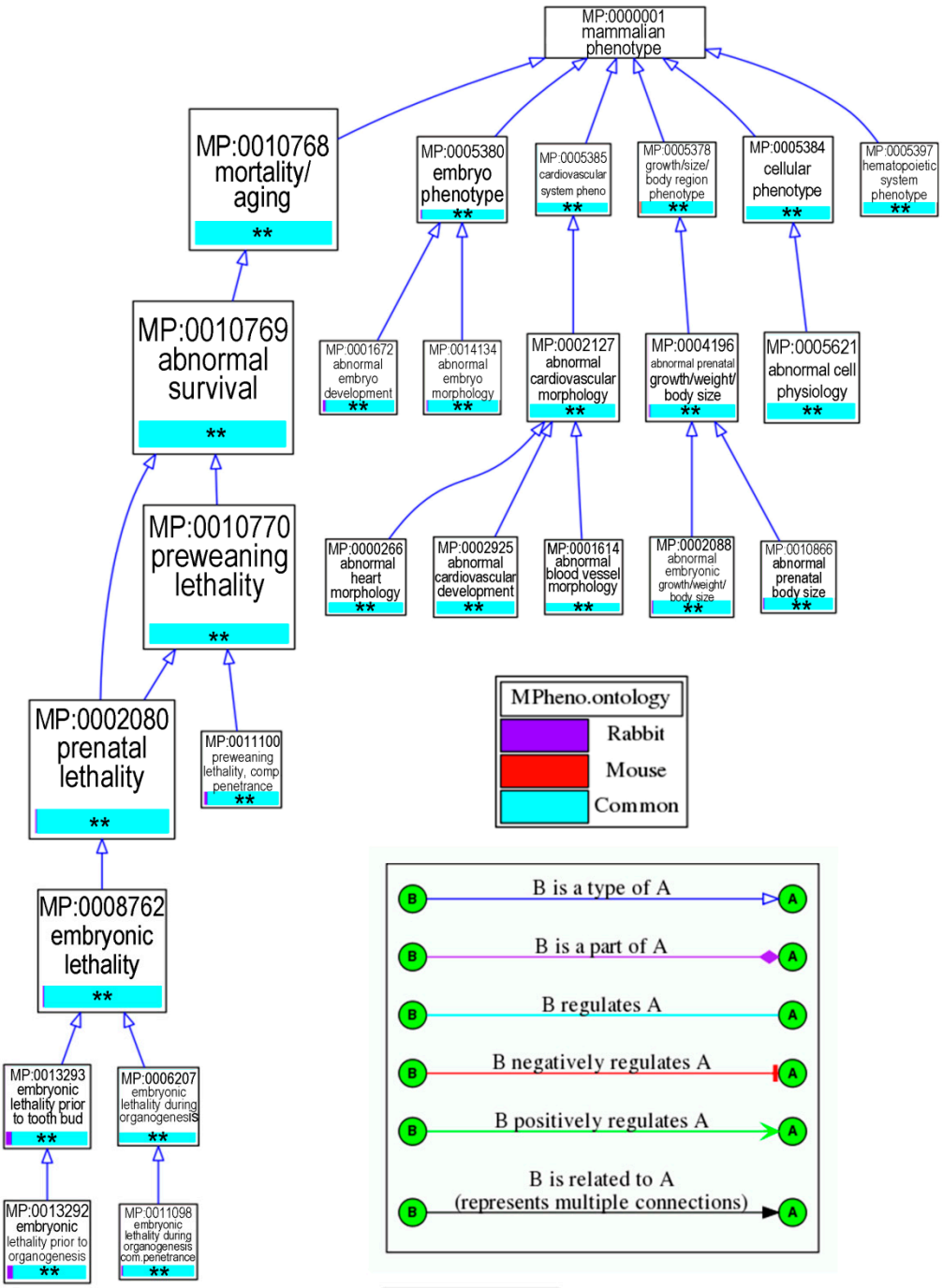

Figure 5. Mammalian Phenotype (MP) ontology enrichment analysis for mouse, rabbit, and common using categorization expression levels. For each MP category only the top 25 significant terms with lowest p-values are shown. The box size reflects its relative statistical significance with the largest box with the lowest p value and the colored bar within the box indicates the proportion of contribution to a specific gene set (purple: Rabbit, red: Mouse, blue: Common). Arrows connecting boxes represent different types of relationship among MP terms. For more detail and interactive module, see Supplemental Table 7 and Supplemental HTML2.

categories (Figure 5) found no significantly overrepresented MP categories among rabbit-only genes. Among mouse-only genes, we again found MP categories related to neuronal function, such as "MP:0005386 behavior/neurological phenotype", "MP:0003633 abnormal nervous system physiology". Due to the nature of hypergeometric statistical test employed by VLAD, in both GO and MP analyses, broad gene categories tend to dominate the top tiers of low *p*, low *q* values and more narrow, descendent gene categories tend to be located lower on the list. For example, 7 genes are currently annotated to the category "MP:0011572 abnormal aorta bulb morphology" (*Fbn1*, *Lox*, *Lrp1*, *Smarca4*, *Tgfb2*, *Tgfbr1*, *Tgfbr2*), and all 7 genes are present in the common group of genes; however, because of relatively high *p* and *q*, this category may be easily overlooked in its 995th

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

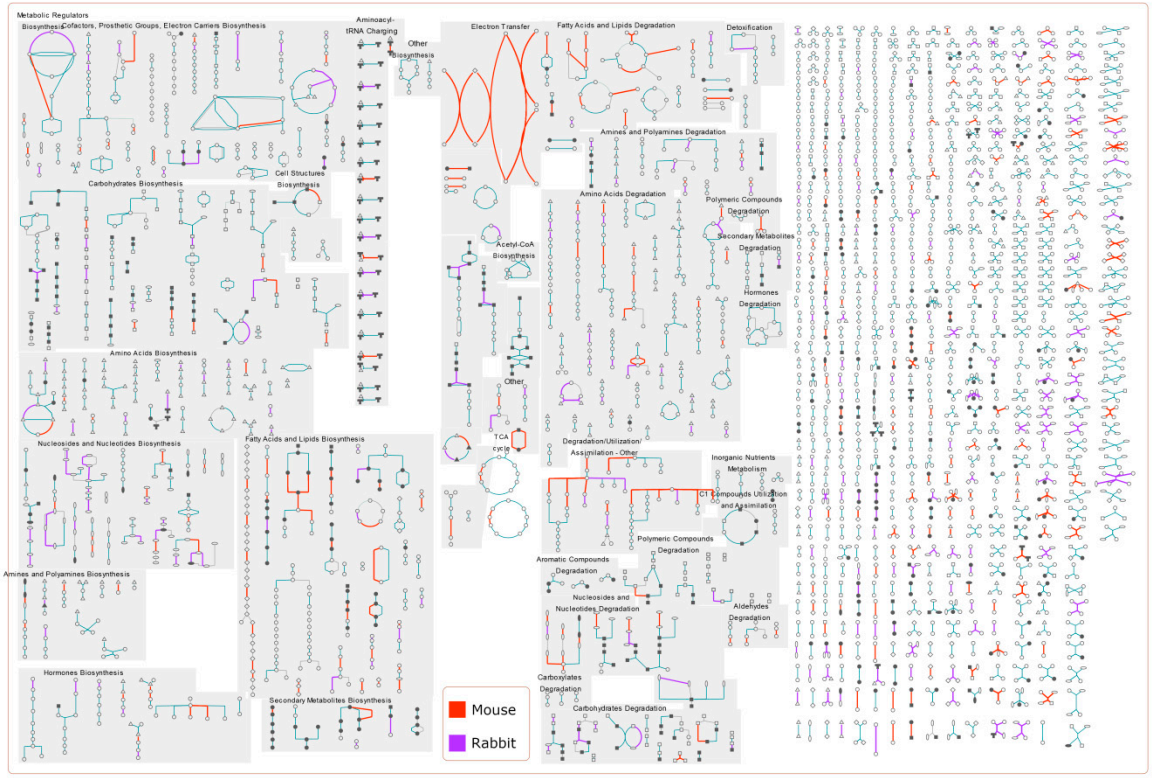

Figure 6. MouseCyc, enrichment tool for biochemical pathway analysis and visualization, using gene expression levels for mouse vs. rabbit samples. The tool depicts all the reactions and pathways with mouse only genes in red, and rabbit-only genes in purple.

position when sorted by *p*-value despite being statistically significant (Supplementary Table 7). The whole continuum of overrepresented categories can be further explored in the provided interactive html files (GO: Supplementary HTML1; MP: Supplementary HTML2).

To identify potential differences in biochemical pathways affected, we analyzed the mouse-only and rabbit-only genes using MouseCyc, a database and tool for biochemical pathway analysis and visualization [39] (Figure 6). This analysis reveals mouse-only genes are involved in mitochondrial electron transfer chain (specifically cytochrome b), fatty acids and lipids degradation, and fatty acids and lipids biosynthesis. Rabbit-only genes are more prominent in glycolysis, γ-glutamyl cycle, and several nucleosides and nucleotides biosynthesis pathways.

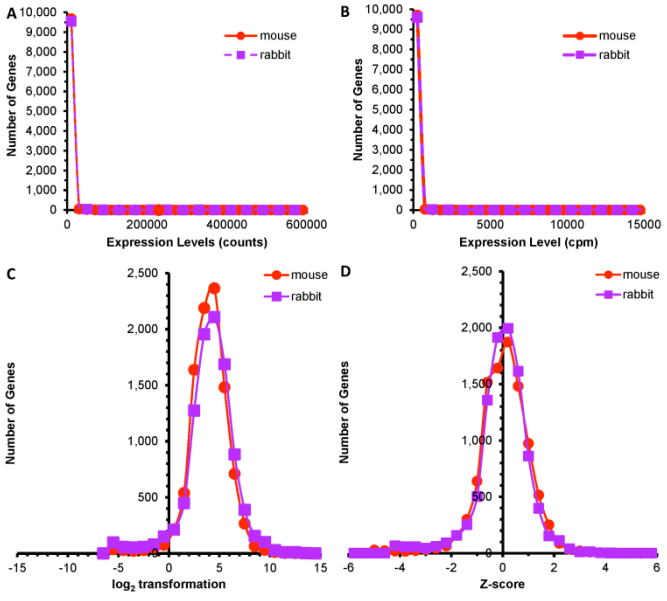

Figure 7. Data transformation from raw counts to Z-scores. Mouse and rabbit raw count data (A) for genes with average ≥5 CPM expression in at least one species were normalized to CPM value (B), CPM values were log2-transformed (C), with a CPM value corresponding to 1 read was added to all zero values (a.k.a. "pseudo count" [24]); z-scores (D) were calculated using geometric means from log2-transformed data for individual gene expression levels in mouse and rabbit datasets.

3.3. *Differential gene expression*

To discover quantitative changes in gene expression between mouse and rabbit models of atherosclerosis, mouse and rabbit count data for genes with average ≥5 CPM expression in at least one species were normalized to CPM value, except for a count of 1 read was added to all zero count values (a.k.a. a "pseudo count" [24]); CPM values were log2-transformed, and z-scores were calculated against mouse and rabbit

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

references for gene expression, which were log_2-transformed geometric means of individual gene expression levels (CPM) in mouse and rabbit datasets, respectively (Figure 7). This procedure accounts for potential differences in sequencing depth between samples (Figure 7A) and changes the distribution of gene expression values from approximately negative binomial (Figure 7B) to normal (Figure 7C), and then harmonizing resulting distributions (Figure 7D) for further statistical testing. Gene expression levels were compared using t-test, and corrected for multiple testing. Subtraction products of z-score means between mouse and rabbit samples serve as the quantitative measures of the difference in expression of individual genes. This procedure revealed 1,441 genes were expressed relatively higher in rabbits, while 1,587 genes were expressed higher in mice ($p < 0.01$, FDR $q < 0.1$; Figure 8, Supplementary Table 5). The 6,699 genes with no significant difference in expression between mice and rabbits were designated as common.

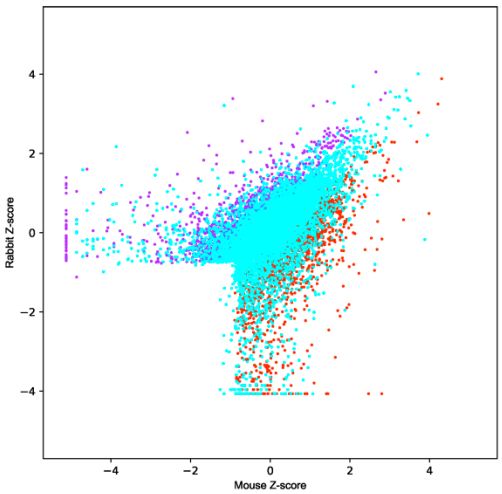

Figure 8. Subtraction plots of z-score means between mouse and rabbit samples for expression of individual genes. Similar number of genes were expressed in species-specific manner, rabbits (purple, 1,441 genes) vs. mouse (red, 1,587 genes, $p < 0.01$, FDR $q < 0.1$). Genes with no significant difference in species expression between were designated as common (blue, 6,699 genes).

GO enrichment analysis (Figure 9 A-C) revealed 225 significantly overrepresented Biological Process, 101 Cellular Component, and 12 Molecular Function categories ($p < 0.01$, $q < 0.1$). Of these, 89 Biological Process categories were overrepresented among genes expressed at higher levels in mice; no overrepresented Biological Process categories were found for rabbit; 136 categories were overrepresented among common genes. Among overrepresented Cellular Component categories, 58 were for common genes, and 20 and 23 categories were overrepresented for genes with higher expression in mice and rabbits, respectively (Supplementary Table 8). Among overrepresented Molecular Function categories, 151 were for common genes, and one each was for genes with higher expression in mice and rabbits, respectively. "GO:0008152 Metabolic process" was again significantly overrepresented among common genes. However, differential expression analysis proves to be more fine-tuned to experimental system because, e.g., "GO:0019222 regulation of metabolic process" is overrepresented among genes upregulated in the mouse (Figure 9A), rather than common genes in the previous analysis (Figure 4A). The topology and nodes overlapped for Cellular Component categories, and "GO:0005739 mitochondrion" was again overrepresented among genes with higher expression in rabbit (Figure 9B). This analysis also revealed genes with "GO:0003712 transcription coregulator activity" (Figure 9C; this group includes *Aebp2*, *Arnt*, *Atf7ip*, *C1d*, *Cbfa2t2*, *Crebbp*, *Ctnnb1*, *Ddx5*, *Gon4l*, *Hcfc1*, *Hipk2*, *Hr*, *Jmy*, *Kat2b*, *Kdm5a*, *Limd1*, *Mkl2*, *Myocd*, *Naca*, *Ncoa2*, *Ncoa3*, *Ncor2*, *Nrip1*, *Nsd1*, *Rad54l2*, *Raly*, *Rbm39*, *Scai*, *Sin3a*, *Tbl1xr1*, *Tcf4*, *Trim28*, *Trrap*, *Ube3a*, *Zfp281*) have higher expression in mice. Of these genes, six are annotated with "MP:0005385 cardiovascular system phenotype" (*Arnt*, *Crebbp*, *Ctnnb1*, *Mkl2*, *Myocd*, *Naca*, *Ncor2*, and *Trim28*). To navigate through all overrepresented GO categories, please see Supplementary HTML3.

Analysis of Mammalian Phenotype ontologies revealed an interesting bias: of 537 overrepresented MP categories ($p < 0.01$, $q < 0.1$), 386 are associated with the genes higher expressed in mice, and 151 with common genes; no single overrepresented MP category was associated with genes expressed higher in rabbit. "MP:0010768 mortality/aging" and its descendants were the prevalent categories among top 25 (Figure 10), recapitulating previous result (Figure 5). Similarly, "MP:0002127 abnormal cardiovascular system morphology" was overrepresented in this analysis as well (Figure 10, Supplementary Table 9). However, this analysis revealed that both of these categories are overrepresented in *both* common genes and genes with higher expression in the mouse (Figure 10). "MP:0001785 edema" was an overrepresented category among genes with high expression in mice (Figure 10). Among these, 17 genes with higher expression in mice (*Ago2*, *C2cd3*, *Cflar*, *Ctnnb1*, *Flrt2*, *Itgav*, *Kmt2d*, *Kras*, *Map3k7*, *Mib1*, *Mkl2*, *Naca*, *Notch2*, *Pdgfra*, *Pkn2*, *Por*, *Wnk1*)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

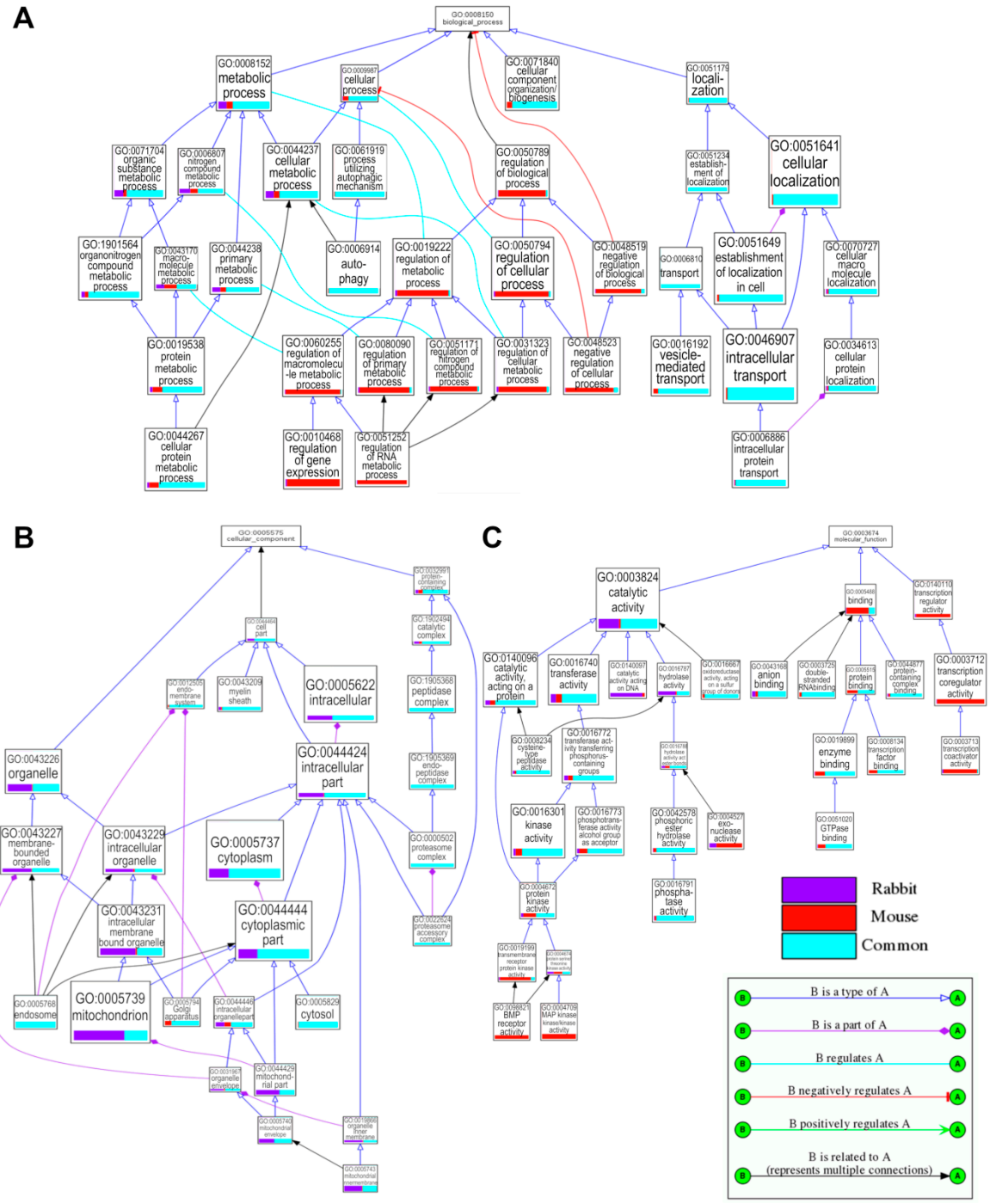

Figure 9. Gene Ontology (GO) enrichment analysis for differentially expressed genes among mouse, rabbit, and common gene sets. For each GO module, Biological Process (A), Cellular Component (B) and Molecular Function (C), only the top 25 significant terms with lowest p-values are shown. The box size reflects its relative statistical significance with the largest box with the lowest p value and the colored bar within the box indicates the proportion of contribution to a specific gene set (purple: Rabbit, red: Mouse, blue: Common). Arrows connecting boxes represent different types of relationship among GO terms. For more detail and interactive module, see Supplemental Table 8 and Supplemental HTML3.

belong to a particularly interesting "MP:0001787 pericardial edema" (Supplementary HTML4); indeed, atherosclerosis and edema conditions are closely linked [40].

MouseCyc analysis of differentially expressed genes allowed completing and clarifying the biochemical pathway differences between mice and rabbits (Figure 11). For example, it made clear that in the mitochondrion, in the electron transfer chain, it is the last step (Complex IV), which is higher in mice, whereas the rest of the electron transfer chain pathway is higher in the rabbits. BioCyc analysis confirms, complements, and importantly extends the VLAD the results in the VLAD analysis of Gene Ontology data in Figure 4B and 9B to specific biochemical pathways and molecules.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

Interestingly, several fatty acids biosynthesis pathways are expressed higher in mice, while fatty acids and lipids degradation pathways are higher in rabbits (Figure 10). This reflects the differences between two translational models of atherosclerosis.

4. Discussion

As a proof of principle of precision medicine at the molecular and gene network level, we used transcriptomics to classify two of the most popular *in vivo* diet-induced models of coronary atherosclerosis, apolipoprotein E (*Apoe*) null mice [14] and NZW rabbits [41], fed with high fat and high cholesterol diets [12,15]. This comparison is a suitable model to evaluate strengths and weaknesses of transcriptomics usage for precision medicine in a clinical setting because data were generated by samples from heterogeneous genomic population, different laboratories, using different molecular biology kits for RNAseq library preparation, and sequencing instruments. Consequently, our analysis mimics the challenges of meaningful bioinformatics evaluation across different RNAseq datasets of the same human disease. An additional challenge exemplified in this study is comparison of two different species, with diverse alleles, used to model the same condition.

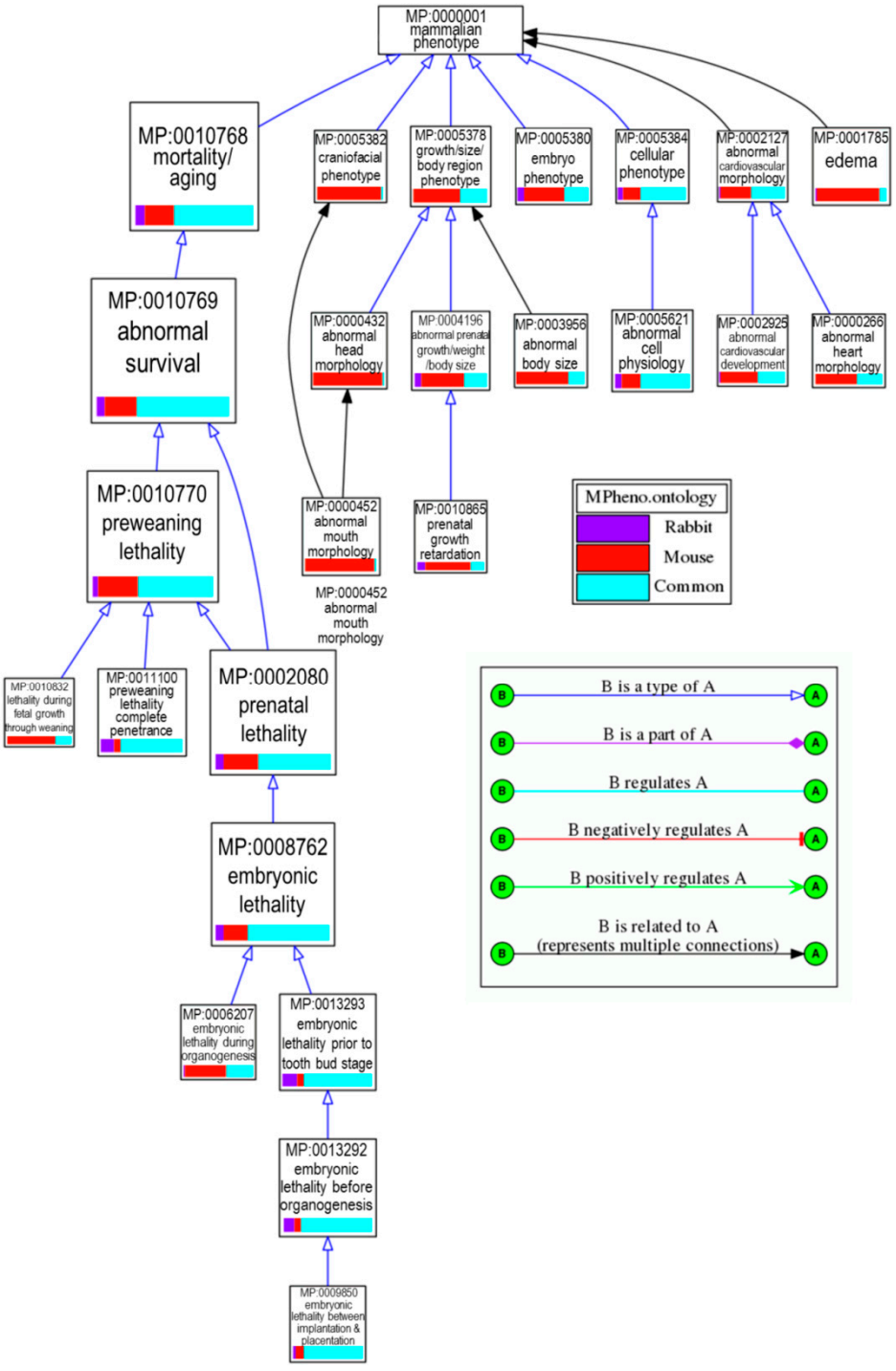

Figure 10. Mammalian Phenotype (MP) ontology enrichment analysis for differentially expressed genes among mouse, rabbit, and common. For each MP category only the top 25 significant terms with lowest p-values are shown. The box size reflects its relative statistical significance with the largest box with the lowest p value and the colored bar within the box indicates the proportion of contribution to a specific gene set (purple: Rabbit, red: Mouse, blue: Common). Arrows connecting boxes represent different types of relationship among MP terms. For more detail and interactive module, see Supplemental Table 9 and Supplemental HTML4.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2018 doi:10.20944/preprints201809.0082.v1

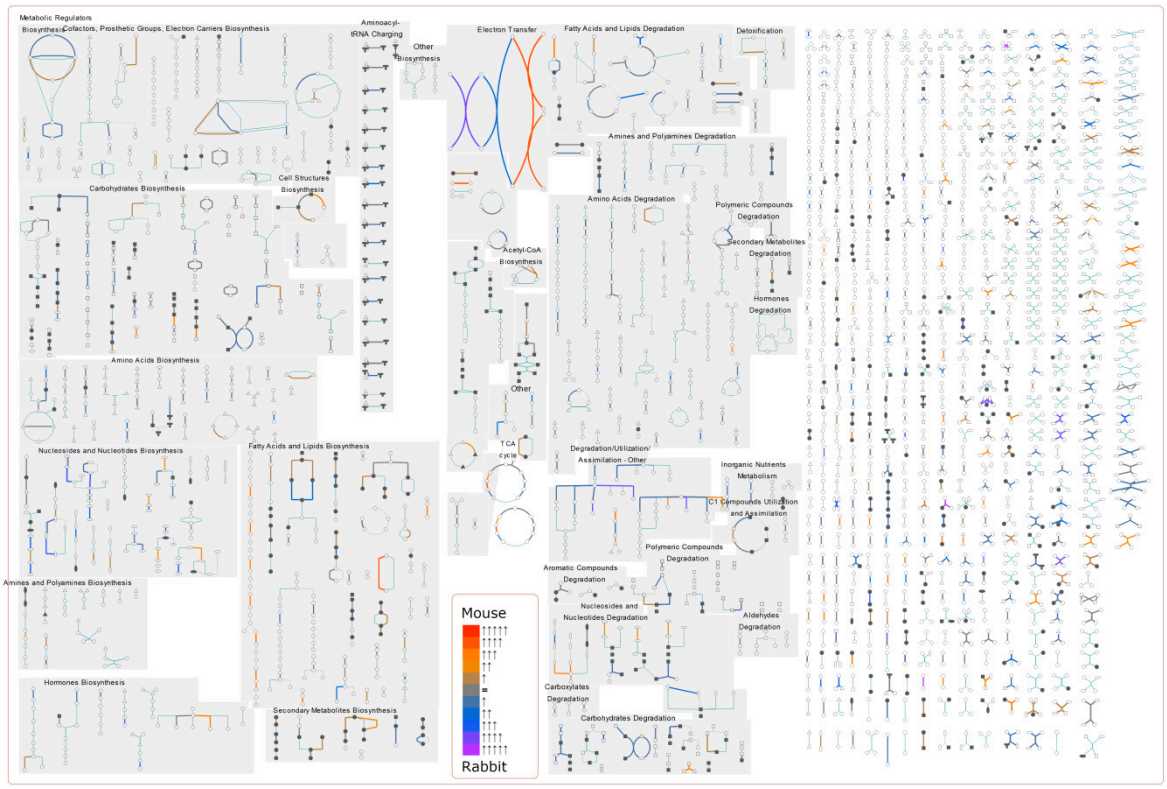

Figure 11. MouseCyc enrichment tool for biochemical pathway analysis and visualization using differentially expressed genes for mouse vs. rabbit samples. Color scheme corresponds to relative expression in the mouse vs rabbit data based on z-score difference.

In summary, from our comparative transcriptome analysis, we discovered that both *in vivo* diet-induced models of coronary atherosclerosis, apolipoprotein E (*Apoe*) null mice [14] and NZW rabbits [41], share a substantial overlap in dysregulated biological processes, pathways, and molecules. Furthermore, our study demonstrates transcriptome analysis can discover specific cellular and molecular pathways and genes with unrecognized roles in atherosclerosis. For example, using the results from Gene Ontology and Mammalian Phenotype, genes associated with axonal guidance (*Chl1*, *Dpysl5*, *Efna5*, *Epha4*, *Epha5*, *Epha7*, *Ephb2*, *Fzd3*, *Gap43*, *Gli2*, *Isl1*, *Klf7*, *L1cam*, *Nfasc*, *Ngfr*, *Nrcam*, *Plxna4*, *Scn1b*, *Sema5a*, *Sema6a*, *Tubb3*, *Unc5c*) may have unknown roles in the pathology of coronary atherosclerosis in the *Apoe* null mouse model (Figures 4B, Supplemental Table 6). Lipid metabolism is a focal area of therapeutic target testing in atherosclerosis. Both *in vivo* models exhibit lipid metabolism derangements, albeit our comparative analysis using MouseCyc pinpoint anabolic pathways are relatively higher in mice, whereas catabolic pathways are relatively higher in rabbits. Overall rabbit model has relatively more common pathways than unique species-specific affected pathways compared to the mouse model, as revealed by Gene Ontology, Mammalian Phenotype, and BioCyc annotations.

One of the main points to clarify is that we applied Gene Ontology and Mammalian Phenotype annotations of *mouse* genes to *both* mouse *and* rabbit gene sets, inferring close similarities for gene functions between mouse and rabbit models of atherosclerosis. Annotation of rabbit genes is underrepresented in all curated databases, therefore we used annotations of mouse genes to infer upon rabbit orthologs, and thus we relied on "inferred by sequence similarity" principle for ontological annotation of rabbit genes. If more experimental, rabbit gene-specific annotations existed, our analysis may be more precise and resolve some of the obvious biases, such as genes with higher expression in rabbits not having corresponding overrepresented Mammalian Phenotype categories. If such annotations for rabbit genes existed at the same level of detail as for the mouse genes, the result may have more exact and detailed results. However, since laboratory mice are a

predominating mammalian model system, mouse genes will have inherently better ontological annotations based on experimental evidence comparing to other model species.

Another source of differences in gene expression between these two models of atherosclerosis is their genetic difference due to the *Apoe* mutation in mice. For example, microarray analysis of aortic endothelial cells from wild-type and *Apoe*$^{-/-}$ mice revealed ~800 differentially expressed genes [42]. Likewise, many differences in our comparison may be due to the effects of *Apoe* mutation upon direct and secondary changes in gene expression from disruption of *Apoe* regulation. Nevertheless, global similarity in gene expression between these two models of diet-induced atherosclerosis, and common pathways identified by functional genomics analysis, provide a compelling example of the power of transcriptomics in comparative atherosclerosis research.

Treatments that target molecular mechanisms underlying the physiological changes, in addition to treating symptoms arising from pathophysiology, are a central promise of personalized medicine – indeed, in theory genomic data can reveal specific disease-associated genotypes to optimize the treatment plan [43]. Albeit, in human population only a few genotypes associated with severe atherosclerosis have been identified, such as multiple alleles of apolipoprotein E (*APOE*, Slooter 1998, Elosua 2004), angiotensin-converting enzyme I (*ACE*) [44] and aryl hydrocarbon receptor (*AHR*) polymorphisms (Huang, Shui et al 2015). Importantly, many non-genetic factors, such as environmental exposures [7], diet [8] and lifestyle [9], strongly affect onset, symptomology and severity of atherosclerosis. This study validates transcriptome analysis as a robust alternative method to identify specific cellular and molecular facets of atherosclerotic disease, which can be used individualize treatment and develop novel avenues of therapeutic intervention even in the absence of genetic data. Current diagnostic and medical laboratory technologies used in clinical setting provide only a small snapshot into the state of disease. The untapped potential of transcriptomics to personalized medicine resides within its revelations of all the local and global changes to cellular, molecular, and biochemical pathways occurring from disease.

Supplementary Materials: The following are available online, Suppl Table 1. List of Orthologs, Suppl Table 2. Aligned RNAseq Datasets, Suppl Table 3. Read Counts for Mouse Genes, Suppl Table 4. Read Counts for Rabbit Genes, Suppl Table 5. Gene Expression Categorization for Statistics for Orthologs
Suppl Table 6. Overrepresented Gene Ontology Terms by Expression Levels, Suppl Table 7. Overrepresented Mammalian Phenotype Terms by Expression Levels, Suppl Table 8. Overrepresented Gene Ontology Terms by Differential Gene Expression Analysis, Suppl Table 9. Overrepresented Mammalian Phenotype Terms by Differential Gene Expression Analysis and interactive modules HTML 1. VLAD GO GXSD, HTML 2. VLAD MP GXSD, HTML 3. VLAD GO Diff GXD, HTML 4. VLAD MP Diff GXD.

Author Contributions: Conceptualization, C.M.E.; Methodology, A.V.E. and C.M.E.; Software, A.V.E. and C.M.E.; Formal Analysis, A.V.E. and C.M.E.; Data Curation, A.V.E., C.M.E. Writing – Original Draft Preparation, A.V.E. and C.M.E.; Writing – Review & Editing, A.V.E. and C.M.E.; Visualization, A.V.E. and C.M.E.; Supervision, C.M.E; Project Administration, C.M.E.; Funding Acquisition, A.V.E. and C.M.E.

Funding: This research was funded, in part, by Impact Assets.

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Marín de Evsikova, C.; Lockhart, J.; Jaimes, G.; Raplee, I.D.; Evsikov, A.V. The transcriptomic toolbox: Resources for interpreting large gene expression data within a cellular, molecular and disease context for cardio-metabolic disease atherosclerosis. *Journal of Personalized Medicine* **2018**, *submitted*.
2. Ohsfeldt, R.L.; Gandhi, S.K.; Fox, K.M.; Bullano, M.F.; Davidson, M. Medical and cost burden of atherosclerosis among patients treated in routine clinical practice. *Journal of Medical Economics* **2010**, *13*, 500-507.
3. Kochanek, K.D.; Murphy, S.L.; Xu, J.Q.; Arias, E. Mortality in the united states, 2013. National Center for Health Statistics: Hyattsville, MD, 2014.
4. Torio, C.M.; Moore, B.J. National inpatient hospital costs: The most expensive conditions by payer, 2013. Agency for Healthcare Research and Quality: Rockville, MD, 2016.
5. *Tresch and aronow's cardiovascular disease in the elderly, fifth edition*. 5 ed.; CRC Press: Boca Raton, 2014; p 800.
6. Pant, S.; Deshmukh, A.; GuruMurthy, G.S.; Pothineni, N.V.; Watts, T.E.; Romeo, F.; Mehta, J.L. Inflammation and atherosclerosis—revisited. *Journal of Cardiovascular Pharmacology and Therapeutics* **2014**, *19*, 170-178.
7. Brook, R.D.; Rajagopalan, S. Particulate matter air pollution and atherosclerosis. *Current Atherosclerosis Reports* **2010**, *12*, 291-300.
8. Kritchevsky, D. Diet and atherosclerosis. *Am J Pathol* **1976**, *84*, 615-632.
9. Ornish, D.; Brown, S.E.; Billings, J.H.; Scherwitz, L.W.; Armstrong, W.T.; Ports, T.A.; McLanahan, S.M.; Kirkeeide, R.L.; Gould, K.L.; Brand, R.J. Can lifestyle changes reverse coronary heart disease?: The lifestyle heart trial. *The Lancet* **1990**, *336*, 129-133.
10. Wild, S.H.; Byrne, C.D. Risk factors for diabetes and coronary heart disease. *BMJ* **2006**, *333*, 1009-1011.
11. Friedemann, C.; Heneghan, C.; Mahtani, K.; Thompson, M.; Perera, R.; Ward, A.M. Cardiovascular disease risk in healthy children and its association with body mass index: Systematic review and meta-analysis. *BMJ : British Medical Journal* **2012**, *345*.
12. Durgin, B.G.; Cherepanova, O.A.; Gomez, D.; Karaoli, T.; Alencar, G.F.; Butcher, J.T.; Zhou, Y.-Q.; Bendeck, M.P.; Isakson, B.E.; Owens, G.K., *et al.* Smooth muscle cell-specific deletion of col15a1 unexpectedly leads to impaired development of advanced atherosclerotic lesions. *American Journal of Physiology-Heart and Circulatory Physiology* **2017**, *312*, H943-H958.
13. Srinivas, S.; Watanabe, T.; Lin, C.-S.; William, C.M.; Tanabe, Y.; Jessell, T.M.; Costantini, F. Cre reporter strains produced by targeted insertion of eyfp and ecfp into the rosa26 locus. *BMC Developmental Biology* **2001**, *1*, 4.
14. Piedrahita, J.A.; Zhang, S.H.; Hagaman, J.R.; Oliver, P.M.; Maeda, N. Generation of mice carrying a mutant apolipoprotein e gene inactivated by gene targeting in embryonic stem cells. *Proceedings of the National Academy of Sciences* **1992**, *89*, 4471-4475.
15. Wang, Z.; Zhang, J.; Li, H.; Li, J.; Niimi, M.; Ding, G.; Chen, H.; Xu, J.; Zhang, H.; Xu, Z., *et al.* Hyperlipidemia-associated gene variations and expression patterns revealed by whole-genome and transcriptome sequencing of rabbit models. *Scientific Reports* **2016**, *6*, 26942.
16. Bui, L.C.; Leandri, R.D.; Renard, J.P.; Duranthon, V. Ssh adequacy to preimplantation mammalian development: Scarce specific transcripts cloning despite irregular normalisation. *BMC Genomics* **2005**, *6*, 155.

17. NCBI. Sra blast. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&BLAST_SPEC=SRA&LINK_LOC=blasttab
18. Ng, K.; Pullirsch, D.; Leeb, M.; Wutz, A. Xist and the order of silencing. *EMBO reports* **2007**, *8*, 34-39.
19. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. Star: Ultrafast universal rna-seq aligner. *Bioinformatics* **2013**, *29*, 15-21.
20. Goecks, J.; Nekrutenko, A.; Taylor, J.; Galaxy, T. Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **2010**, *11*, R86.
21. Liao, Y.; Smyth, G.K.; Shi, W. Featurecounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923-930.
22. Mortazavi, A.; Williams, B.A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nature Methods* **2008**, *5*, 621.
23. Metsalu, T.; Vilo, J. Clustvis: A web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic acids research* **2015**, *43*, W566-570.
24. Rapaport, F.; Khanin, R.; Liang, Y.; Pirun, M.; Krek, A.; Zumbo, P.; Mason, C.E.; Socci, N.D.; Betel, D. Comprehensive evaluation of differential gene expression analysis methods for rna-seq data. *Genome Biology* **2013**, *14*, 3158.
25. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biology* **2010**, *11*, R106.
26. Robinson, M.D.; Oshlack, A. A scaling normalization method for differential expression analysis of rna-seq data. *Genome Biology* **2010**, *11*, R25.
27. Zerbino, D.R.; Achuthan, P.; Akanni, W.; Amode, M R.; Barrell, D.; Bhai, J.; Billis, K.; Cummins, C.; Gall, A.; Girón, C.G., *et al.* Ensembl 2018. *Nucleic acids research* **2018**, *46*, D754-D761.
28. NCBI. Blast® command line applications user manual. National Center for Biotechnology Information (US): Bethesda (MD), 2008.
29. Smith, C.L.; Blake, J.A.; Kadin, J.A.; Richardson, J.E.; Bult, C.J.; the Mouse Genome Database, G. Mouse genome database (mgd)-2018: Knowledgebase for the laboratory mouse. *Nucleic acids research* **2018**, *46*, D836-D842.
30. MGI. Mouse genome informatics. http://www.informatics.jax.org
31. Millman, K.J.; Aivazis, M. Python for scientists and engineers. *Computing in Science & Engineering* **2011**, *13*, 9-12.
32. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **1995**, *57*, 289-300.
33. Seabold, S.; Perktold, J. In *Statsmodels: Econometric and statistical modeling with python*, 9th Python in Science Conference, 2010.
34. Richardson, J.E.; Bult, C.J. Visual annotation display (vlad): A tool for finding functional themes in lists of genes. *Mammalian Genome* **2015**, *26*, 567-573.
35. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T., *et al.* Gene ontology: Tool for the unification of biology. *Nature Genetics* **2000**, *25*, 25.
36. Smith, C.L.; Goldsmith, C.-A.W.; Eppig, J.T. The mammalian phenotype ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biology* **2004**, *6*, R7.

37. Smith, B.; Ashburner, M.; Rosse, C.; Bard, J.; Bug, W.; Ceusters, W.; Goldberg, L.J.; Eilbeck, K.; Ireland, A.; Mungall, C.J., *et al.* The obo foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology* **2007**, *25*, 1251.
38. Caspi, R.; Billington, R.; Ferrer, L.; Foerster, H.; Fulcher, C.A.; Keseler, I.M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Mueller, L.A., *et al.* The metacyc database of metabolic pathways and enzymes and the biocyc collection of pathway/genome databases. *Nucleic acids research* **2016**, *44*, D471-D480.
39. Evsikov, A.V.; Dolan, M.E.; Genrich, M.P.; Patek, E.; Bult, C.J. Mousecyc: A curated biochemical pathways database for the laboratory mouse. *Genome Biol* **2009**, *10*, R84.
40. Falk, E.; Shah, P.K.; Fuster, V. Coronary plaque disruption. *Circulation* **1995**, *92*, 657-671.
41. Yanni, A.E. The laboratory rabbit: An animal model of atherosclerosis research. *Laboratory Animals* **2004**, *38*, 246-256.
42. Erbilgin, A.; Siemers, N.; Kayne, P.; Yang, W.P.; Berliner, J.; Lusis, A.J. Gene expression analyses of mouse aortic endothelium in response to atherogenic stimuli. *Arterioscler Thromb Vasc Biol* **2013**, *33*, 2509-2517.
43. Libby, P.; Ridker, P.M.; Hansson, G.K. Progress and challenges in translating the biology of atherosclerosis. *Nature* **2011**, *473*, 317.
44. O'Malley, J.P.; Maslen, C.L.; Illingworth, D.R. Angiotensin-converting enzyme dd genotype and cardiovascular disease in heterozygous familial hypercholesterolemia. *Circulation* **1998**, *97*, 1780-1783.