

2016

Integration of flow studies for robust selection of mechanoresponsive genes

Nataly Maimari
Imperial College London


Ryan M. Pedrigi
University of Nebraska-Lincoln, rpedrigi@unl.edu

Alessandra Russo
Imperial College London

Krysia Broda
Imperial College London

Rob Krams
Imperial College London, r.krams@imperial.ac.uk

Follow this and additional works at: <http://digitalcommons.unl.edu/mechengfacpub>

 Part of the [Mechanics of Materials Commons](#), [Nanoscience and Nanotechnology Commons](#), [Other Engineering Science and Materials Commons](#), and the [Other Mechanical Engineering Commons](#)

Maimari, Nataly; Pedrigi, Ryan M.; Russo, Alessandra; Broda, Krysia; and Krams, Rob, "Integration of flow studies for robust selection of mechanoresponsive genes" (2016). *Mechanical & Materials Engineering Faculty Publications*. 293.
<http://digitalcommons.unl.edu/mechengfacpub/293>

This Article is brought to you for free and open access by the Mechanical & Materials Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Mechanical & Materials Engineering Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Integration of flow studies for robust selection of mechanoresponsive genes

Nataly Maimari^{1, 2}, Ryan M. Pedrigi¹, Alessandra Russo¹, Krysia Broda¹ and Rob Krams²

Dept. of Computing¹ and Bioengineering², Imperial College London

Correspondence:

Professor Rob Krams

Chair in Molecular Bioengineering

Dept. Bioengineering

Imperial College London

Room 3.15

Royal School of Mines

Exhibition Road

SW7 2AZ London

e-mail: r.krams@imperial.ac.uk

<tel:+442075941473>

Summary: Blood flow is an essential contributor to plaque growth, composition and initiation. It is sensed by endothelial cells, which react to blood flow by expressing >1000 genes. The sheer number of genes implies that one needs genomic techniques to unravel their response in disease. Individual genomic studies have been performed but lack sufficient power to identify subtle changes in gene expression. In this study, we investigated whether a systematic meta-analysis of available microarray studies can improve their consistency.

We identified 17 studies using microarrays, of which 6 were performed in vivo and 11 in vitro. The in vivo studies were disregarded due to the lack of the shear profile. Of the in vitro studies, a cross-platform integration of human studies (HUVECs in flow cells) showed high concordance (>90%). The human data set identified >1600 genes to be shear responsive, more than any other study and in this gene set all known mechanosensitive genes and pathways were present. A detailed network analysis indicated a power distribution (e.g. the presence of hubs), without a hierarchical organization. The avg. cluster coefficient was high and further analysis indicated an aggregation of 3 and 4 element motifs, indicating a high prevalence of feedback and feed forward loops, similar to prokaryotic cells.

In conclusion, this initial study presented a novel method to integrate human-based mechanosensitive studies to increase its power. The robust network was large, contained all known mechanosensitive pathways and its structure revealed hubs, and a large aggregate of feedback and feed forward loops.

Keywords: bioinformatics, network biology, network structure, mechanobiology, microarrays, human.

Introduction

Blood flow plays a key role in atherosclerosis, both in the initial and advanced stages of the disease, and during subsequent interventions to alleviate the disease (1-3). The information generated by the blood flow is transmitted by endothelial cells, which are exquisitely sensitive to the magnitude and direction of the blood flow, through a process called mechanotransduction. Mechanotransduction is the propensity of cells to react to mechanical cues, and this mechanism consists of a variety of mechanosensors that drive signaling pathways enabling the cell to react to the mechanical cues by changing its protein signature. Endothelial cells, the mechanosensitive nature of which has been extensively studied, reacts to blood flow by switching on between 1000-2000 genes, activating >8 transcription factors and >10 signaling pathways. The sheer number of signaling molecules involved in a cellular response implies that we need to use a genomic approach to study mechanotransduction and use bioinformatics techniques to infer signaling networks (1-3).

In a previous paper (2), we summarized all findings on genome-wide studies and concluded that the large variability in results reported between researchers was due to a large variability of bioinformatics analysis, necessitating the need to standardize this approach. Furthermore, integrative analysis is a relatively inexpensive way to gain new biological insight by making use of data accumulated through the years. Increasing sample size and statistical power allows for the discovery of subtle differences in gene expression and results in more reliable and generalized findings that are less likely to be due to study-specific artifacts (4-9). In addition, each individual study addresses only a small part of the global solution, therefore the merging of diverse conditions on the same biological problem allows for the interrogation of a greater proportion of the “solution space” that is relevant to the biological problem under investigation (8).

This paper describes an initial approach to increase the power of individual microarray studies through a pipeline that integrates Affymetrix, Illumina and Agilent platforms and applies them to previously reported microarray studies performed to identify mechanosensitive signaling pathways.

How to aggregate single microarray studies.

Initial selection of studies

Studies have been collected from three databases: PubMed, Gene Expression Omnibus (GEO) (10) and ArrayExpress (11). A systematic search has been conducted to find all studies whose abstract and/or title contain at least one of the following keywords: endothelial cell, fluid mechanical forces and gene expression. Supplemental Figure 1 illustrates the flow chart for study collection and characterization. In total, PubMed, GEO and ArrayExpress returned 209, 142 and 60 studies respectively. After removing duplicates and screening for relevance, 82 studies were identified and further evaluated. Amongst these, 9 papers were reviews and 16 original research papers had incorrect datatype (e.g., methylation or miRNA expression datasets) and therefore were excluded. The remaining 57 studies were formally reviewed and summarized resulting in a final set of 17 studies considered for building the integrated mechanoresponsive gene expression profile. Exclusion criteria of the studies were: no raw data available, sample size less than three, incompatible factors (e.g., knockdown genes) and studies that included first generation microarrays with low coverage.

Figure 1 illustrates the different shear stress conditions across the studies considered, grouped by species and microarray platform. Studies have conducted analysis on Affymetrix, Agilent, Illumina and spotted cDNA platforms using human umbilical venous endothelial cells (HUVEC), porcine arterial endothelial cells (PAEC) and bovine arterial endothelial cells (BAEC). Shear stress conditions have been grouped in a descriptive phenotype variable, which consists of five levels: static, low, moderate, high and complex shear stress. Low, moderate and high refer to the magnitude of shear stress used for steady uniform shear. Low shear includes samples from 1 to 10 dyne/cm², moderate shear captures samples from 11 to 35 dyne/cm² and high includes samples with shear between 75 and 284 dyne/cm². The complex category includes all non-uniform shear stress waveforms (e.g., oscillatory, pulsatile). As can be seen in Figure 1, there is no consistency in the shear stress conditions used across studies.

For the purpose of this study, we have focused on the analysis of single color arrays (Affymetrix, Agilent and Illumina). This allowed us to test the feasibility of the integrative approach without the additional complexity of combining data from two-color, which are obtained as log ratios with expression values obtained from single color arrays. Of the remaining single color arrays, only a small number of samples had been subjected to complex shear (e.g., oscillatory shear or shear with gradient). Hence, only samples with uniform shear have been retained for integration.

The primary factor considered in the analysis was shear stress pattern that had been characterized by three parameters: waveform, directionality and magnitude. Secondary factors that were considered include: species, flow system and platform.

Platform specific preprocessing

Affymetrix datasets were normalized using the Robust Multiarray Analysis (RMA) algorithm (12) implemented in Affy R package (13), as it has consistently been shown to perform amongst the top ranked methods in terms of sensitivity to biological variation (14-17). In addition, it has been shown to improve cross-platform comparability (18) and it has been used by studies that successfully performed cross-platform integration of expression values (19-21).

Illumina datasets were normalized using methods implemented in the Lumi R package (22). Raw data were log₂ transformed and quantile normalized. No background correction was used. This approach has been shown to perform well before (23) and (22).

Agilent datasets were normalized using methods implemented in the Limma R package (24). Raw data were background corrected using the *'normexp'*, log₂ transformed and quantile normalized. Agilent is frequently used for its two-color arrays, which allow the hybridization of two samples on the same chip. Therefore the recommendations for two-color arrays have been used, since the error model for single channel Agilent is expected to be very similar (25, 26).

Annotation of Microarray Platforms

All probes have been mapped at the gene level using human Entrez IDs as the common identifier between platforms. Entrez ID has been selected since it is stable over time, well curated and has previously been shown to achieve high mapping rates between platforms (27). Probes mapping to the same Entrez ID were combined into a single value by taking the mean. A summary of the final mapping statistics is shown in supplemental Table 1, depicting the total number of Entrez IDs covered by each array, alongside the number of mapped and excluded probes.

Following the successful annotation of each individual platform, we compared the coverage of the four human arrays (Affy U133, Lumi HT12, Lumi Ref8 and Ag44k). Collectively, the probes on the four arrays target a total of 22,478 unique Entrez IDs, out of which 14,030 are common amongst all four human platforms (Supplemental Figure 2).

There is an overlap of 7,743 genes between human and bovine platforms and an overlap of 12,641 genes between human and mouse. The 7,743 genes common between bovine and human arrays capture 85% of the bovine array, but only 55% of the human arrays. This is because of the low coverage of the bovine array, which includes a total of 9,902 genes. The mouse array has a more extensive coverage of human genes (a total of 15,963) and therefore 90% of the common human is captured by the mouse array.

Systematic biases and cross-platform correction

Systematic biases and cross platform correction was performed using ComBat (28-30). ComBat was chosen based on its demonstrated ability to perform well in small sample studies (28, 29, 31) and the fact that it uniquely allows a design with multiple covariates (e.g., species, shear regime) is very important for our study, which has an imbalanced design. Its ability to remove batch effects due to different experimental runs has been demonstrated in Affymetrix arrays (28, 29, 32, 33), Illumina arrays (34) and two-color arrays (35). In all of the above papers, Combat improved the consistency of gene lists suggesting that the biological signal was still intact. Combat also demonstrated the potential for cross-platform normalization based on merging technical replicates (19, 21).

Finally, individual studies were assessed for technical and biological outliers by arrayQualityMetrics (36). ArrayQualityMetrics implements automatic detection of outliers based on empirically calculated thresholds. MDS plots were used to assess the uniformity of replicates and ensure that samples were clustered by experimental condition. The relative contribution of each factor to the overall variability was assessed using the Principal Variance Component Analysis (PVCA) method (37). Inter-study concordance was assessed through mean-mean expression plots. The goodness of the line of best fit (r^2) was calculated as a measure of inter-study concordance.

Finally, differentially expressed shear responsive genes are detected using the moderated t-test of the LIMMA package. The moderated t-statistic has the same interpretation as an ordinary t-statistic except that the standard errors for every gene are adjusted using an empirical Bayes method which takes into account information from all the genes. Shrinking the standard errors towards a pooled estimated makes the analysis more stable (38). P-values were corrected for multiple testing using the Benjamini–Hochberg procedure. Functional enrichment analysis for Gene ontology terms, canonical pathways and transcription factor binding sites was performed with the online tool WebGestalt (39). Annotations were selected for enrichment based on FDR p-value less than 0.05.

Results of the different steps in microarray normalization

The different steps described above will have different effects in the total variance between microarray studies and within microarray studies. These effects will be discussed in more detail below.

Normalization of individual studies

Prior to merging, the studies were checked for technical and biological outliers using exploratory clustering of samples with multidimensional plots to assess the uniformity of within-study replicates. In addition, variance component analysis was performed using principal variance component analysis (PVCA) (37). The Mun09 dataset includes samples subjected to 12 dyne/cm²

shear stress with static cultures as controls. Sample clustering of normalized data revealed the existence of batch effects in the dataset, which contributes to 27.8% of total variation. The MDS plot in Supplemental Figure 4 demonstrates sample separation related to experimental run, with Run 1 (circles) forming a particularly distinct group. Following batch effect correction, the variation due to experimental run has been reduced to 0%, with clear separation of samples by the four experimental conditions. Following batch correction, the shear stress factor contributed to 86% of total variation.

White11 investigates the shear stress response at 15 dyne/cm² and 75 dyne/cm². Sample clustering of normalized data demonstrates consistent grouping of samples by shear stress magnitude (x-axis direction in Supplemental Figure 5). There is a minor batch effect captured by the second principle component (y-axis direction in Supplemental Figure 5). Following batch effect correction, the proportion of variation in expression data relating to phenotype increased from 62.6% to 81.5%, with a corresponding decrease in the variation linked to batch effects (from 7.9% to 1.3%).

The study of Conway10 includes three shear stress environments: static control, 1 dyne/cm² and 75 dyne/cm². Initial clustering of samples indicated a batch effect that was quantified by variance analysis at 34.6% of the total variation (Supplemental Figure 6). Following batch effect correction, the samples of all shear stress levels form three spatially distinct groupings, whose percentage of variation related to shear stress, increased from 48.5% to 86.1% (Supplemental Figure 6). There was still some interaction between shear stress and batch effect quantified at 15%.

Lorenz15 studies the gender-specific effects of shear stress response. HUVECs isolated from three female and three male donors, were subjected to 6 dyne/cm² shear stress or kept in static conditions, resulting in a paired experimental design. Exploratory MDS plots on Lorenz15 showed unexpected shifts in expression values in response to experimental conditions for which we were unable to correct for by batch adjustment. Initial analysis of all samples in the dataset (Supplemental Figure 3) indicated that two of the male samples (Run 4 and Run 5) demonstrated no response to shear stress. Further, one of the female samples (Run1, circles) demonstrated an enhanced response with a shifted distribution that was identified as a technical outlier by array

quality metrics. Given the evidence of unreliability in the expression data, the entire Lorenz15 dataset was eliminated from the pipeline.

The final dataset considered for the integration of flow studies conducted in human endothelial cells was the Ni11 study, which includes samples at 15 dyne/cm² and oscillatory shear stress. For the purpose of this study, only the 15 dyne/cm² samples were retained. As we were unable to retrieve information on experimental runs from the publication associated with this study (40), no variance component analysis was conducted. Sample clustering did not indicate the existence of batch effects; uniform sheared samples clustered together and away from the oscillatory shear stress samples (data not shown).

Integration of studies to generate a merged expression dataset

Following cross-study normalization, sample distributions showed a great degree of similarity, with a clear separation of samples into three respective shear stress phenotype's. In variance component analysis, 66.2% of variation in normalized data has been linked to shear stress phenotype (Figure 2). There is, however, some study-effect remaining in the data as indicated by the 24% variation attributed to an interaction between phenotype and platform.

Four samples out of the 26 samples have been identified as outliers and removed from further analysis. One of the White11 15 dyne/cm² samples has clustered away from the rest of the moderate shear stress samples (Figure 3A). In addition, the three low shear stress samples from Conway10 study clustered with the moderate shear stress samples and therefore were removed as outliers (data not shown). As a consequence, >90% of variation was attributed to the shear stress phenotype (Figure 3).

In order to explore further the concordance between the different studies, we have analyzed the agreement between the expression values of all the 15 dyne/cm² samples across studies. In total, there are 9 samples, three samples from each of Conway10, Ni11 and White11 studies (Figure 4). Prior to cross-study normalization, the average r^2 value was 0.60 (Figure 4^a), which increased to an average of 0.94 after applying cross-study normalization (Figure 4^b), indicating the high agreement between studies from different platforms. Similar results were

obtained for the concordance between the static samples in Conway10 and Mun09. Following cross-study normalization, r^2 increased from 0.70 to 0.96.

In the second phase of the project, we have attempted to expand the set of integrated studies to cross-species datasets. Two additional studies have been considered, Dolan12 and Dolan13, which subject bovine aortic endothelial cells to a wide range of shear conditions (static controls, 20 dyne/cm², 35 dyne/cm², 100 dyne/cm² and 284 dyne/cm²). After an extensive analysis employing all tools described above, only 147 differentially expressed genes were detected. In that aggregated dataset, the main known mechanosensitive genes were not differentially expressed and we decided not to pursue this analysis due to this outcome.

Identification of shear responsive genes

In order to verify that the biological variation has not been affected by the normalization steps, shear responsive genes have been identified through differential expression analysis of the static and moderate shear stress samples in the integrated dataset. In total, 1,628 shear responsive genes have been identified with a $p < 0.05$. A query for the most related diseases in the set of differentially expressed genes returned cardiovascular diseases with an enrichment score of 32.51 ($p < 0.001$). Enrichment analysis for transcription factor binding sites identified key transcription factors known to be involved in the endothelial cell shear response including AP1, NF- κ B, p53, MEF2, NRF2 and PPAR γ (41-44). Amongst the top unregulated genes were KLF2 (fold change=5.2), KLF4 (fold change=16) and NRF2 (fold change=1.9), which have been attributed to regulate up to 70% of the atheroprotective endothelial cell transcriptional program (42). The list includes upstream regulators of these transcription factors, including ERK5, MEF2, KEAP1, HDAC5, as well as downstream genes controlled by these transcription factors, which have been associated with shear stress and include eNOS (endothelial nitric oxide synthase 3), THBD (thrombomodulin), HMOX1 (heme oxygenase (decycling) 1), NQO1 (NAD(P)H dehydrogenase, quinone 1), CYP1B1 (cytochrome P450, family 1, subfamily B, polypeptide 1), SELE (E-selectin), ICAM1 (intercellular adhesion molecule 1) and plasminogen activator inhibitor type 1 (SERPINE1) (45-48).

Functional enrichment analysis for Gene Ontology biological processes and KEGG/Wikipathway canonical pathways revealed a strong involvement of phosphoproteins ($p < 0.001$) with a widespread response of all five MAPK pathways (enrichment score (ES)=2.88, $p < 0.001$), and processes regulating cell communication ($p < 0.001$) including adherent junction (ES = 4.12), focal adhesions (ES= 2.75), regulation of actin cytoskeleton (ES=2.39), tight junctions (ES=2.48), and integrin mediated cell adhesion (ES=3.49). Top enriched signaling pathways include the Keap1-Nrf2 Pathway (ES= 6.15), oxidative stress (ES= 4.36), Wnt Signaling (ES=2.44), and TGF-beta signaling pathways (ES=3.27). These findings are in line with current literature (49-51).

In summary, the pipeline presented in this study provides an implementation of methods to integrate Affymetrix, Illumina and Agilent at the gene expression level, which has not been previously reported within the literature. It identified ~1650 differentially expressed genes, one of the most ever reported (52, 53). Furthermore, the identified, differentially regulated genes contained all known mechanosensitive pathways, demonstrating the effectiveness of our approach.

Network-analysis

We subjected the aggregated human genes to a network analysis applying weighted gene co-expression analysis (54) and visualized a network consisting of the 10,000 highest expressers with Cytoscape(55). As can be seen in Figure 5, the network is highly interconnected, displays a power law distribution (panel B), while the cluster coefficient was independent of node size (panel C). The power law distribution means that a few nodes have a high interaction with other nodes, while the majority of nodes have only minimal interaction with their neighbors. The highly interactive nodes or hubs have been used successfully to identify new leads for drug development. This is based on the concept of robustness of networks, where low interacting nodes were not influential by deleting them but hubs were (56-58). In addition, the avg. cluster coefficient was independent of degree of freedom of the nodes, which specify the absence of a hierarchical structure (59-61). As the avg. cluster coefficient, and thus the presence of higher order structures

was higher than expected (e.g. on basis of a random network distribution) we counted the presence of three node (negative and positive feedback loops) and four node motifs (feed-forward loops). We detected >50 of these motifs, which is concordant to the presence of feedback loops in our detected mechanosensitive pathways (e.g. MAPK, Wnt, and NF-kB).

In conclusion, the present approach enables to study robust gene networks. The new findings suggest this approach could be used in the future to study gene networks in disease where rewiring and loss of function of several pathways and hence small communities have been reported in several other fields than cardiology.

Legends

Figure 1. Experimental factors in the flow studies considered for building the integrated dataset

Each study is represented by a different box, which is colored according to the microarray platform used. The species and shear stress regime used is also indicated. Shear stress conditions have been grouped in a descriptive phenotype variable, which consists of five levels: static, low, moderate, high and complex shear stress. Low, moderate and high refer to the magnitude of shear stress used for steady uniform shear. The number in each box shows the specific shear stress value (in dyne/cm²) used. The complex category includes all non-uniform shear stress waveforms (e.g. oscillatory, pulsatile).

Figure 2. Merging of human expression datasets.

(A) Sample clustering of the merged human endothelial cell dataset. For each study the normalized batch corrected data were used. Each point on the plot is a sample, where colour corresponds to the shear stress environment. B) Visualizes the same MDS plot as in (A), with individual samples colored according to the different studies instead of shear stress environment. C) The bar chart plot the proportion of variance (y-axis) captured by each of the known factors (x-axis). D) Expression values, density distribution plots for each sample, coloured according to individual studies.

Figure 3. Cross-study normalization of the integrated human endothelial cell dataset

(A) Sample clustering of the integrated human endothelial cell dataset following cross-study normalisation. Each point on the plot is a sample, where colour corresponds to the shear stress environment. B) Visualises the same MDS plot as in (A), with individual samples coloured according to the different studies instead of shear stress environment. C) The bar chart plot the proportion of variance (y-axis) captured by each of the known factors (x-axis). D) Expression values, density distribution plots for each sample, coloured according to individual studies.

Figure 4. Mean-mean scatter plots of 15 dyne/cm² samples across studies.

A) Mean-mean plots before cross-platform normalisation. B) The same pairwise comparisons after cross-platform normalisation. Each plot shows the mean expression levels of 15 dyne/cm² samples from one study, plotted against the mean expression level of 15 dyne/cm² samples from another study. The axis labels indicate the studies compared. Under perfect conditions, the points on the mean-mean scatter should fall on the theoretical $y=x$ line, plotted in blue. The red line corresponds to the line of best fit to the data. The goodness of fit (r^2), is showed under each plot and it quantifies the fraction of variance explained by the line of best fit. r^2 is a measure of inter-study concordance.

Figure 5: Network analysis of robust mechanosensitive genes.

A) Displayed is the network structure of the 10,000 highest gene expressors obtained from the fusion of 4 human microarray studies (see text for details). On the top the network is present and below the number and composition of motifs are displayed. Note the high number of 3, 4, and 5-element motifs in this network, B) The inverse relationship between (\log) number of degrees (number of connections per node) and (\log) number of nodes is representative of a power law. This implies that hubs, which may be used for drug development, C) The relationship between (\log) number of degrees and average cluster coefficient indicates that a hierarchical clustering in this network is present.

Supplemental Figure 1. Flow Chart for study collection and characterisation. Studies were collected from three sources: PubMed, ArrayExpress and GEO database. Search commands used are: ^amicroarray*, expression profil*, transcriptome, whole-genome, transcript*, transcriptional analysis, Affymetrix, Illumina, transcriptomics, ^b: endothelial, endothelium, HUVECs, MAECs, BAECs, PAECs, ^c: shear stress, flow NOT flow cytometry, laminar, disturbed, hemodynamics,

mechanosensitive, mechanical force, atheroprotected, athero-protected, atherosusceptible, atherosusceptible, atherosusceptibility, atheroprone.

Supplemental Figure 2. Entrez ID overlap between the array versions included in the integration study.

A) Overlap between the four human array versions. 14030 are common amongst all four arrays referred to as “common-human” list. The individual mapping rates are shown in the table. The rate is calculated by obtaining the set of common Entrez IDs between the two platforms and dividing it by the total number of Entrez IDs in the source platform. B) Overlap between common human Entrez IDs and bovine (top) (or mouse (bottom)) Entrez IDs which have been mapped to human Entrez IDs, using NCBI homolog information. Cross-species mapping rates to individual human arrays are shown in the table

Supplemental Figure 3. Lorenz15 sample clustering

The study consists of three female and three male samples subjected to 6 dyne/cm² shear stress or kept in static conditions, resulting to a paired experimental design. Sample clustering of all samples (A), female samples only (B) and male samples only (C) is plotted using multidimensional scaling plots. Each point on the plot is a sample, where colour corresponds to sheared stress environment and the symbol corresponds to the experimental run of the sample. The inconsistent trend in the shear stress responses have been highlighted by pairing static and low shear samples from each donor to emphasize the directionality of the response. Male samples (C) show a diverse range of responses to shear stress, indicating underlying problems in experimental data. One of the female samples ((B), Run1, circles) demonstrates an enhanced response with a shifted distribution which has been identified as a technical outlier by array quality metrics.

Supplemental Figure 4. Mun09 study: Sample clustering and variance component analysis.

The bar charts show the proportion of variance (y-axis) captured by each of the known study factors (x-axis), before (A) and after (C) batch effect correction. Batch effects define the experimental run of a given sample, Passage defines the age of the cells used, which were either young or senescent, and shear stress defines the shear environment the cells were subjected to. Sample clustering before (B), and after (D), batch effect correction is plotted using multidimensional scaling plots. Each point on the plot is a sample, where colour corresponds to the shear stress environment and the symbol corresponds to the experimental run of the sample.

Supplemental Figure 5. White11 study: Sample clustering and variance component analysis

The bar charts plot the proportion of variance (y-axis) captured by each of the known study factors (x-axis), before (A) and after (C) batch effect correction. Sample clustering before (B), and after (D), batch effect correction is plotted using multidimensional scaling plots. Each point on the plot is a sample, where colour corresponds to the shear stress environment and the symbol corresponds to the experimental run of the sample.

Supplemental Figure 6. Conway10 study: Sample clustering and variance component analysis

The bar charts show the proportion of variance (y-axis) captured by each of the known study factors (x-axis), before (A) and after (C) batch effect correction. Sample clustering before (B), and after (D), batch effect correction is plotted using multidimensional scaling plots. Each point on the plot is a sample, where colour corresponds to sheared stress environment and the symbol corresponds to the experimental run of the sample.

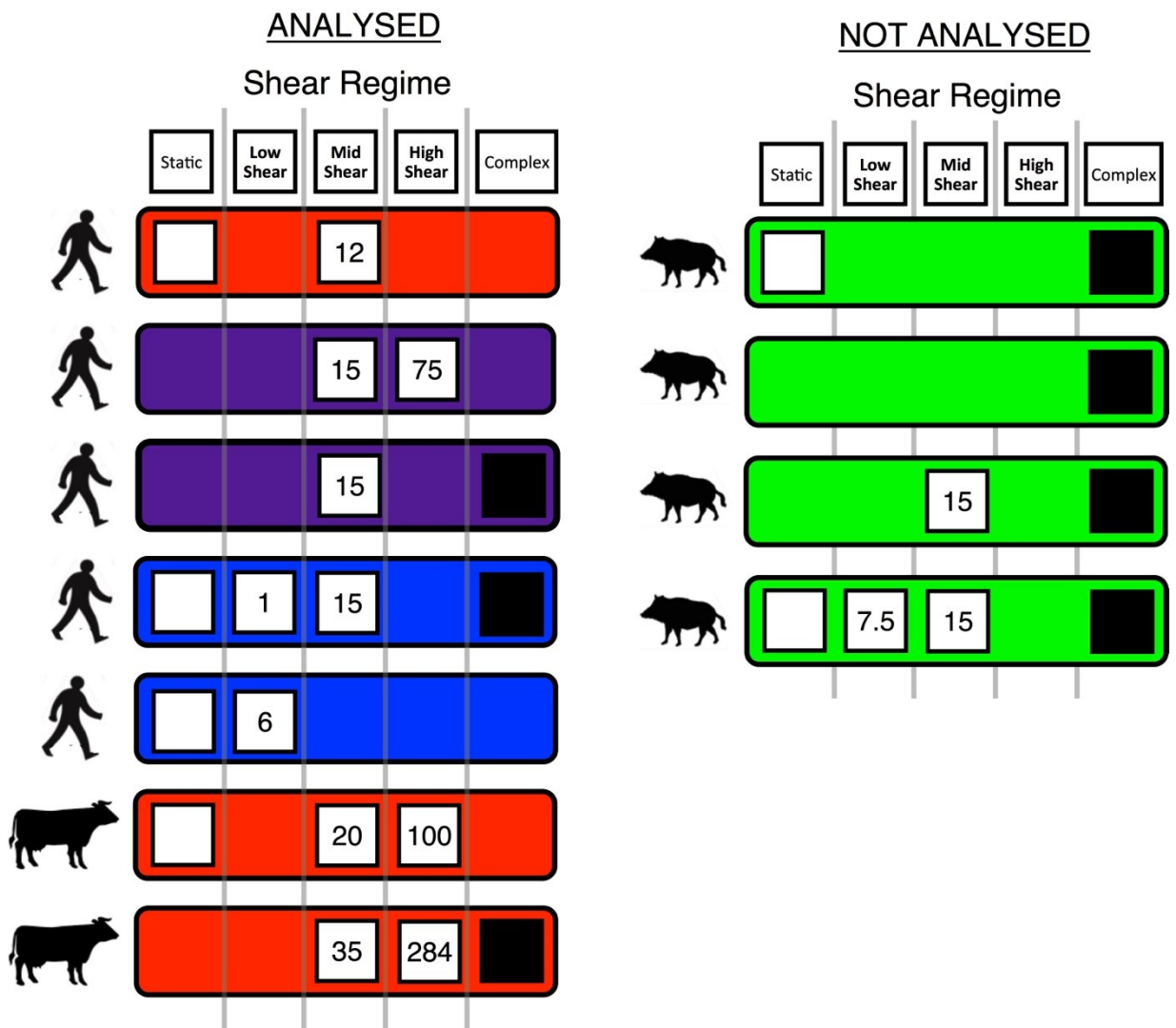
References

1. Pedrigi RM, Poulsen CB, Mehta VV, et al. Inducing Persistent Flow Disturbances Accelerates Atherogenesis and Promotes Thin Cap Fibroatheroma Development in D374Y-PCSK9 Hypercholesterolemic Minipigs. *Circulation* 2015.
2. Frueh J, Maimari N, Homma T, et al. Systems biology of the functional and dysfunctional endothelium. *Cardiovasc Res* 2013; 99(2): 334-41.
3. Segers D, Lipton JA, Leenen PJ, et al. Atherosclerotic Plaque Stability Is Affected by the Chemokine CXCL10 in Both Mice and Humans. *Int J Inflam* 2011; 2011: 936109.
4. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proceedings of the National Academy of Sciences* 2006; 103(15): 5923-8.
5. van Vliet MH, Reyal F, Horlings HM, et al. Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. *BMC Genomics* 2008; 9(1): 375.
6. Xu L, Tan A, Winslow RL, et al. Merging microarray data from separate breast cancer studies provides a robust prognostic test. *BMC Bioinformatics* 2008; 9(1): 125.
7. Belcastro V, Siciliano V, Gregoretti F, et al. Transcriptional gene network inference from a massive dataset elucidates transcriptome organization and gene function. *Nucleic Acids Research* 2011; 39(20): 8677-88.
8. Lukk M, Kapushesky M, Nikkilä J, et al. A global map of human gene expression. *Nature Publishing Group* 2010; 28(4): 322-4.
9. Schmid PR, Palmer NP, Kohane IS, et al. Making sense out of massive data by going beyond differential expression. *Proceedings of the National Academy of Sciences* 2012; 109(15): 5594-9.
10. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* 2002; 30(1): 207-10.
11. Brazma A. ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Research* 2003; 31(1): 68-71.
12. Irizarry RA, Wu Z, Jaffee HA. Comparison of Affymetrix GeneChip expression measures. *Bioinformatics* 2006; 22(7): 789-94.
13. Gautier L, Cope L, Bolstad BM, et al. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004; 20(3): 307-15.
14. Bolstad BM, Irizarry RA, Åstrand M, et al. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003; 19(2): 185-93.
15. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; 4(2): 249-64.
16. McCall MN, Almudevar A. Affymetrix GeneChip microarray preprocessing for multivariate analyses. *Briefings in Bioinformatics* 2012; 13(5): 536-46.
17. Lim WK, Wang K, Lefebvre C, et al. Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics* 2007; 23(13): i282-i8.
18. Shippy R, Fulmer-Smentek S, Jensen RV, et al. Using RNA sample titrations to assess microarray platform performance and normalization techniques. *Nature Biotechnology* 2006; 24(9): 1123-31.
19. Rudy J, Valafar F. Empirical comparison of cross-platform normalization methods for gene expression data. *BMC Bioinformatics* 2011; 12(1): 467.
20. Sîrbu A, Ruskin HJ, Crane M. Cross-Platform Microarray Data Normalisation for Regulatory Network Inference. *PLoS ONE* 2010; 5(11): e13822.
21. Turnbull AK, Kitchen RR, Larionov AA, et al. Direct integration of intensity-level data from Affymetrix and Illumina microarrays improves statistical power for robust reanalysis. *BMC Medical Genomics* 2012; 5(1): 1-.
22. Dunning MJ, Barbosa-Morais NL, Lynch AG, et al. Statistical issues in the analysis of Illumina data. *BMC Bioinformatics* 2008; 9(1): 85.

23. Schmid R, Baum P, Ittrich C, et al. Comparison of normalization methods for Illumina BeadChip HumanHT-12 v3. *BMC Genomics* 2010; 11(1): 349.
24. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 2015; 43(7): e47-e.
25. Ritchie ME, Silver J, Oshlack A, et al. A comparison of background correction methods for two-colour microarrays. *Bioinformatics* 2007; 23(20): 2700-7.
26. Zhu Q, Miecznikowski JC, Halfon MS. A wholly defined Agilent microarray spike-in dataset. *Bioinformatics* 2011; 27(9): 1284-9.
27. Allen JD, Wang S, Chen M, et al. Probe mapping across multiple microarray platforms. *Briefings in Bioinformatics* 2012; 13(5): 547-54.
28. Johnson WE, Li C. Adjusting batch effects in microarray experiments with small sample size using Empirical Bayes methods. *Batch Effects and Noise in Microarray Experiments: Sources and Solutions*: 113-29.
29. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2006; 8(1): 118-27.
30. Leek JT, Johnson WE, Parker HS, et al. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012; 28(6): 882-3.
31. Lazar C, Meganck S, Taminau J, et al. Batch effect removal methods for microarray gene expression data integration: a survey. *Briefings in Bioinformatics* 2013; 14(4): 469-90.
32. Chen C, Grennan K, Badner J, et al. Removing Batch Effects in Analysis of Expression Microarray Data: An Evaluation of Six Batch Adjustment Methods. *PLoS ONE* 2011; 6(2): e17238.
33. Taminau J, Lazar C, Meganck S, et al. Comparison of Merging and Meta-Analysis as Alternative Approaches for Integrative Gene Expression Analysis. *ISRN Bioinformatics* 2014; 2014(1, article 3): 1-7.
34. Kitchen RR, Sabine VS, Sims AH, et al. Correcting for intra-experiment variation in Illumina BeadChip data is necessary to generate robust gene-expression profiles. *BMC Genomics* 2010; 11(1): 134.
35. Larsen MJ, Thomassen M, Tan Q, et al. Microarray-Based RNA Profiling of Breast Cancer: Batch Effect Removal Improves Cross-Platform Consistency. *BioMed Research International* 2014; 2014: 1-11.
36. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. *Bioinformatics* 2009; 25(3): 415-6.
37. Li J, Bushel PR, Chu T-M, et al. Principal variance components analysis: Estimating batch effects in microarray gene expression data. *Batch Effects and Noise in Microarray Experiments: Sources and Solutions* 2009: 141-54.
38. Smyth GK. *Limma: linear models for microarray data*. Springer 2005; pp. 397-420.
39. Zhang J, BurrIDGE KA, Friedman MH. In vivo differences between endothelial transcriptional profiles of coronary and iliac arteries revealed by microarray analysis. *AJP: Heart and Circulatory Physiology* 2008; 295(4): H1556-H61.
40. Ni CW, Qiu H, Jo H. MicroRNA-663 upregulated by oscillatory shear stress plays a role in inflammatory response of endothelial cells. *AJP: Heart and Circulatory Physiology* 2011; 300(5): H1762-H9.
41. Boon RA, Fledderus JO, Volger OL, et al. KLF2 Suppresses TGF- Signaling in Endothelium Through Induction of Smad7 and Inhibition of AP-1. *Arteriosclerosis, Thrombosis, and Vascular Biology* 2007; 27(3): 532-9.
42. Boon RA, Horrevoets A, others. Key transcriptional regulators of the vasoprotective effects of shear stress. *Hämostaseologie* 2009; 29(1): 39-43.
43. Fledderus JO, Boon RA, Volger OL, et al. KLF2 Primes the Antioxidant Transcription Factor Nrf2 for Activation in Endothelial Cells. *Arteriosclerosis, Thrombosis, and Vascular Biology* 2008; 28(7): 1339-46.
44. Heo K-S, Fujiwara K, Abe J-i. Disturbed-Flow-Mediated Vascular Reactive Oxygen Species Induce Endothelial Dysfunction. *Circulation Journal* 2011; 75(12): 2722-30.
45. CLARK PR, JENSEN TJ, KLUGER MS, et al. MEK5 is Activated by Shear Stress, Activates ERK5 and Induces KLF4 to Modulate TNF Responses in Human Dermal Microvascular Endothelial Cells. *Microcirculation* 2011; 18(2): 102-17.
46. García-Cardena G, Comander J, Anderson KR, et al. Biomechanical activation of vascular endothelium as a determinant of its functional phenotype. *Proceedings of the National Academy of Sciences* 2001; 98(8): 4478-85.

47. McCormick SM, Eskin SG, McIntire LV, et al. DNA microarray reveals changes in gene expression of shear stressed human umbilical vein endothelial cells. *Proceedings of the National Academy of Sciences* 2001; 98(16): 8955-60.
48. Wang W, Ha CH, Jhun BS, et al. Fluid shear stress stimulates phosphorylation-dependent nuclear export of HDAC5 and mediates expression of KLF2 and eNOS. *Blood* 2010; 115(14): 2971-9.
49. Chatzizisis YS, Coskun AU, Jonas M, et al. Role of Endothelial Shear Stress in the Natural History of Coronary Atherosclerosis and Vascular Remodeling. *Journal of the American College of Cardiology* 2007; 49(25): 2379-93.
50. Chiu JJ, Chien S. Effects of Disturbed Flow on Vascular Endothelium: Pathophysiological Basis and Clinical Perspectives. *Physiological Reviews* 2011; 91(1): 327-87.
51. Pan S. Molecular mechanisms responsible for the atheroprotective effects of laminar shear stress. *Antioxidants & redox signaling* 2009; 11(7): 1669-82.
52. Frueh J, Maimari N, Homma T, et al. Systems biology of the functional and dysfunctional endothelium. *Cardiovascular Research* 2013; 99(2): 334-41.
53. Frueh J, Maimari N, Lui Y, et al. Systems and synthetic biology of the vessel wall. *FEBS letters* 2012; 586(15): 2164-70.
54. Zhao W, Langfelder P, Fuller T, et al. Weighted gene coexpression network analysis: state of the art. *J Biopharm Stat* 2010; 20(2): 281-300.
55. Shannon PT, Grimes M, Kutlu B, et al. RCytoscape: tools for exploratory network analysis. *BMC Bioinformatics* 2013; 14: 217.
56. Hammann F, Drewe J. Data mining for potential adverse drug-drug interactions. *Expert Opin Drug Metab Toxicol* 2014; 10(5): 665-71.
57. Uhart M, Bustos DM. Protein intrinsic disorder and network connectivity. The case of 14-3-3 proteins. *Front Genet* 2014; 5: 10.
58. Sporns O. The human connectome: origins and challenges. *Neuroimage* 2013; 80: 53-61.
59. Honardoost MA, Naghavian R, Ahmadinejad F, et al. Integrative computational mRNA-miRNA interaction analyses of the autoimmune-deregulated miRNAs and well-known Th17 differentiation regulators: An attempt to discover new potential miRNAs involved in Th17 differentiation. *Gene* 2015.
60. Villaveces JM, Koti P, Habermann BH. Tools for visualization and analysis of molecular networks, pathways, and -omics data. *Adv Appl Bioinform Chem* 2015; 8: 11-22.
61. Bracken CP, Khew-Goodall Y, Goodall GJ. Network-Based Approaches to Understand the Roles of miR-200 and Other microRNAs in Cancer. *Cancer Res* 2015; 75(13): 2594-9.

Figure 1



Key to Microarray Platform

- Affymetrix
- Illumina
- Agilent
- Two colour

Species

- Human
- Bovine
- Porcine

Figure 2

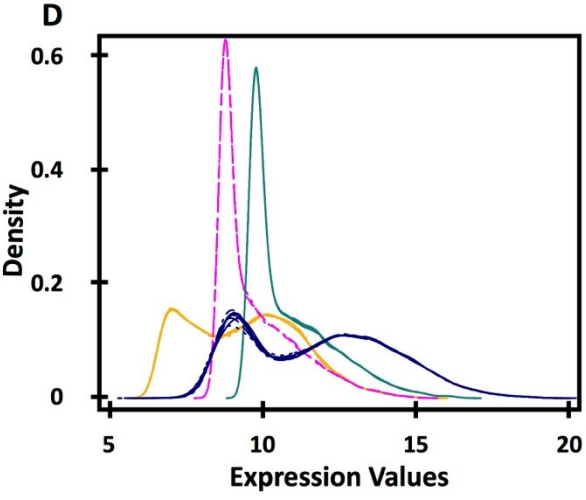
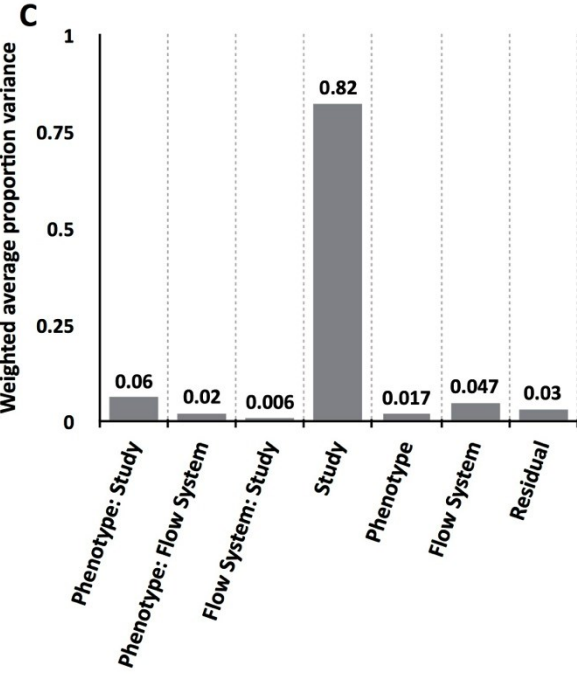
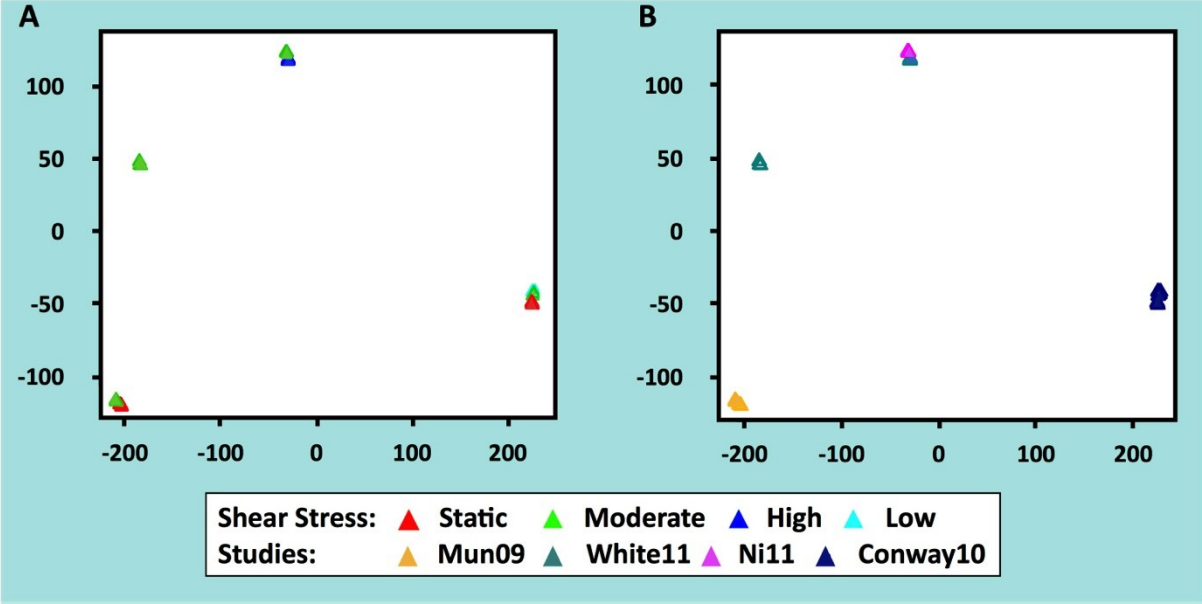


Figure 3

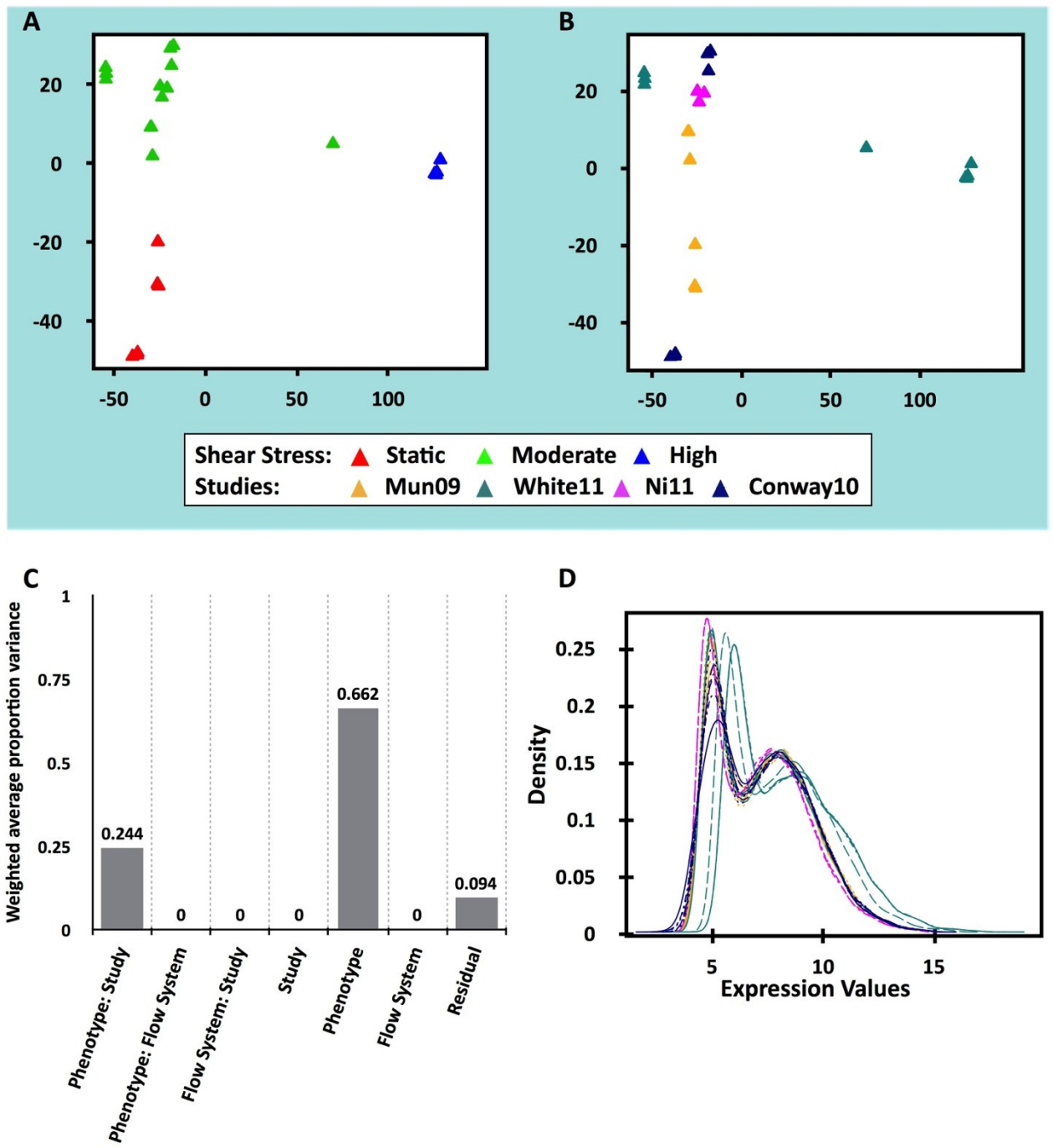


Figure 4

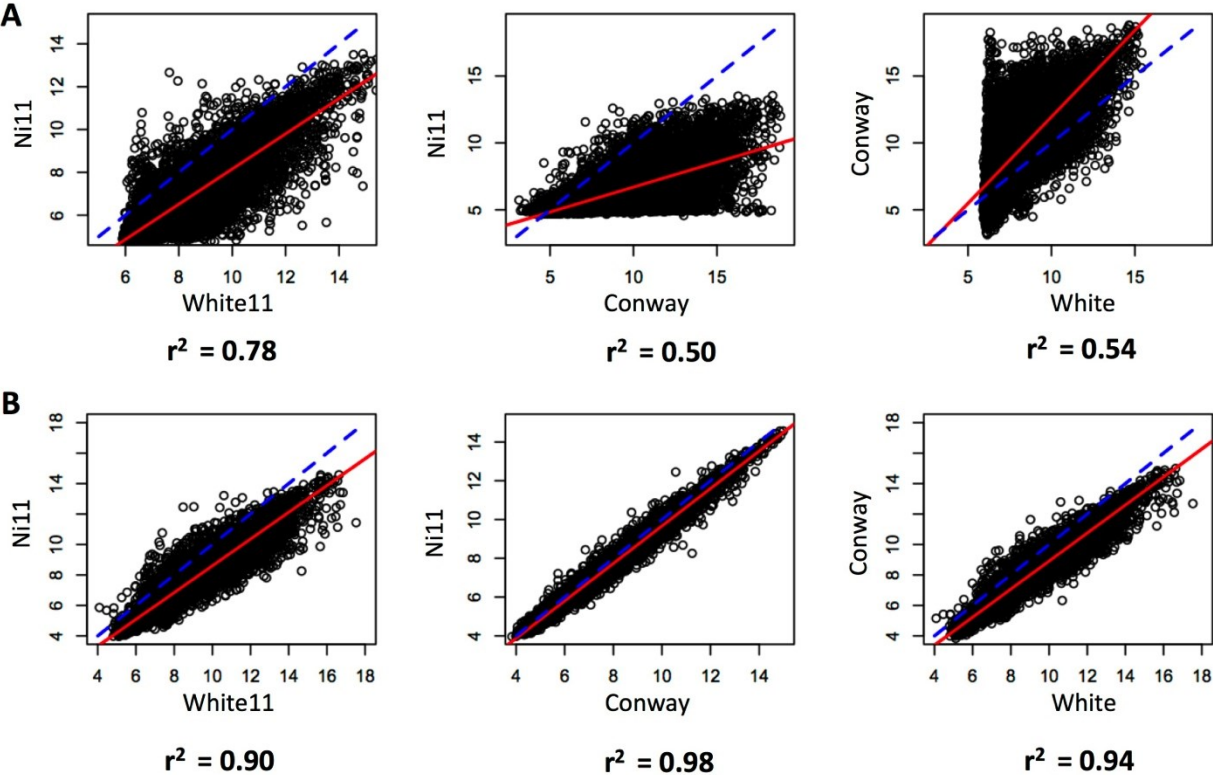
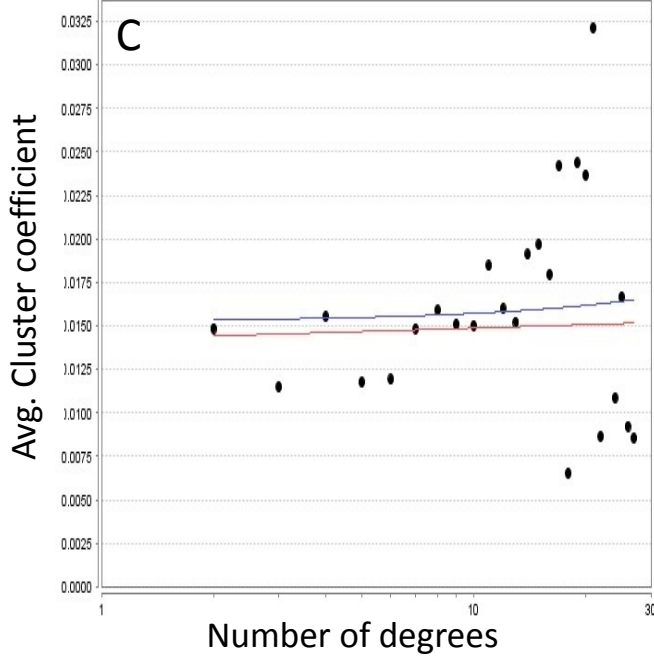
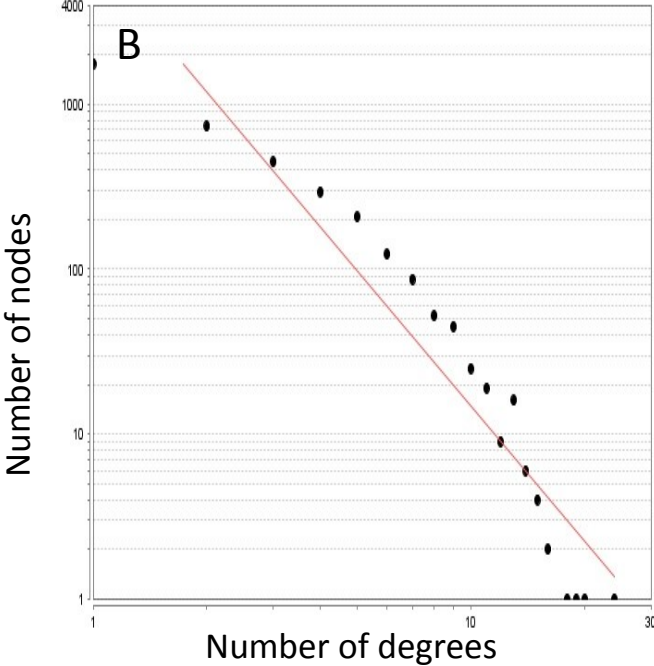
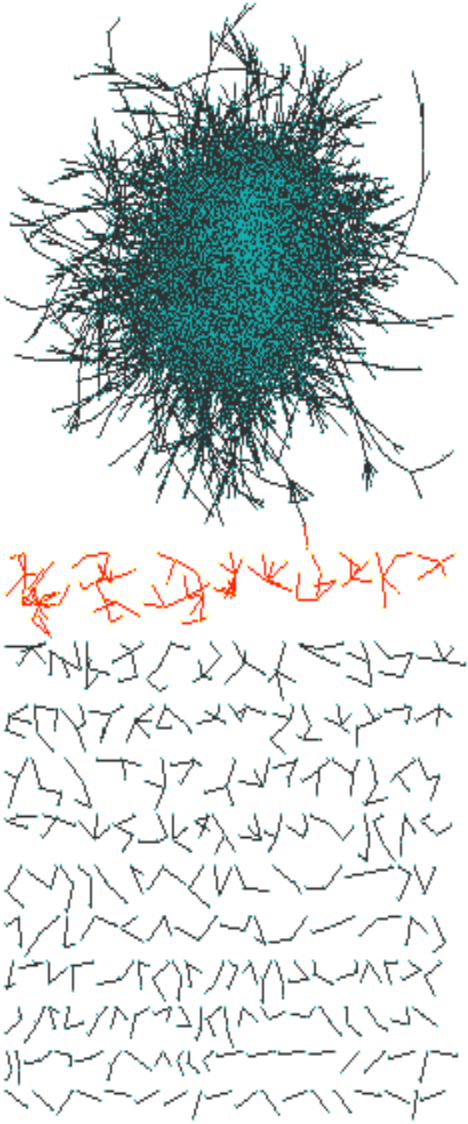


Figure 5

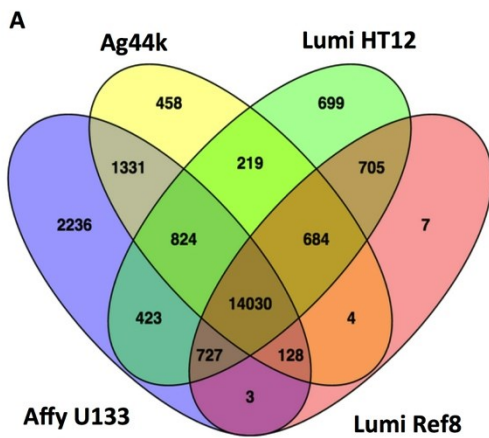
A



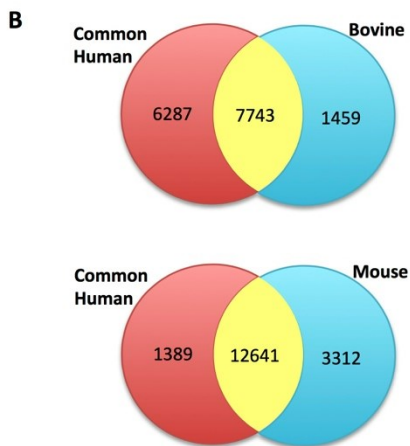
Supplemental Figure 1



Supplemental Figure 2

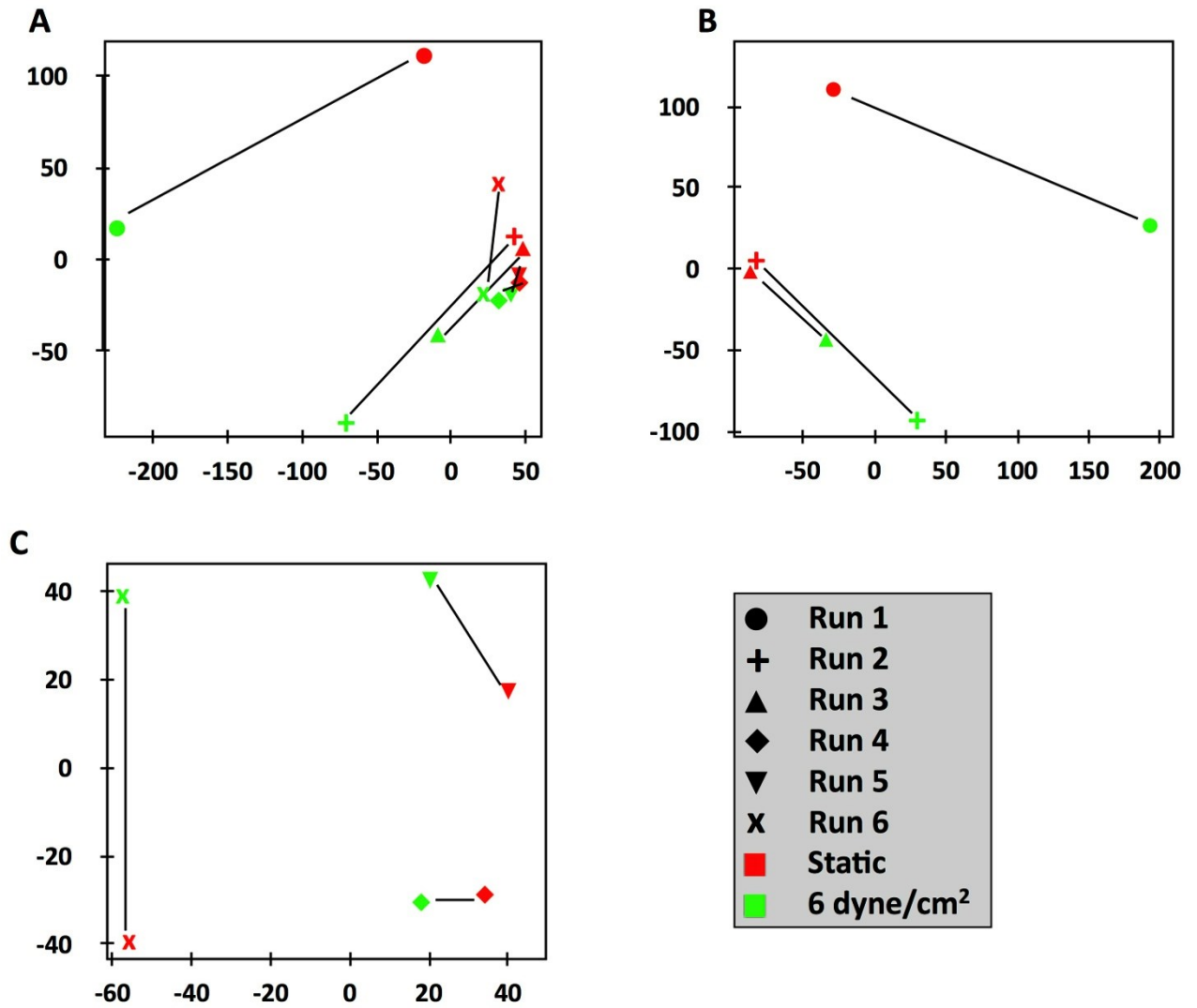


Source	Target			
	Affy U133	Lumi HT12	Lumi Ref8	Ag44k
Affy U133	---	0.81	0.76	0.83
Lumi HT12	0.88	---	0.88	0.86
Lumi Ref8	0.91	0.99	---	0.91
Ag44k	0.92	0.89	0.84	---

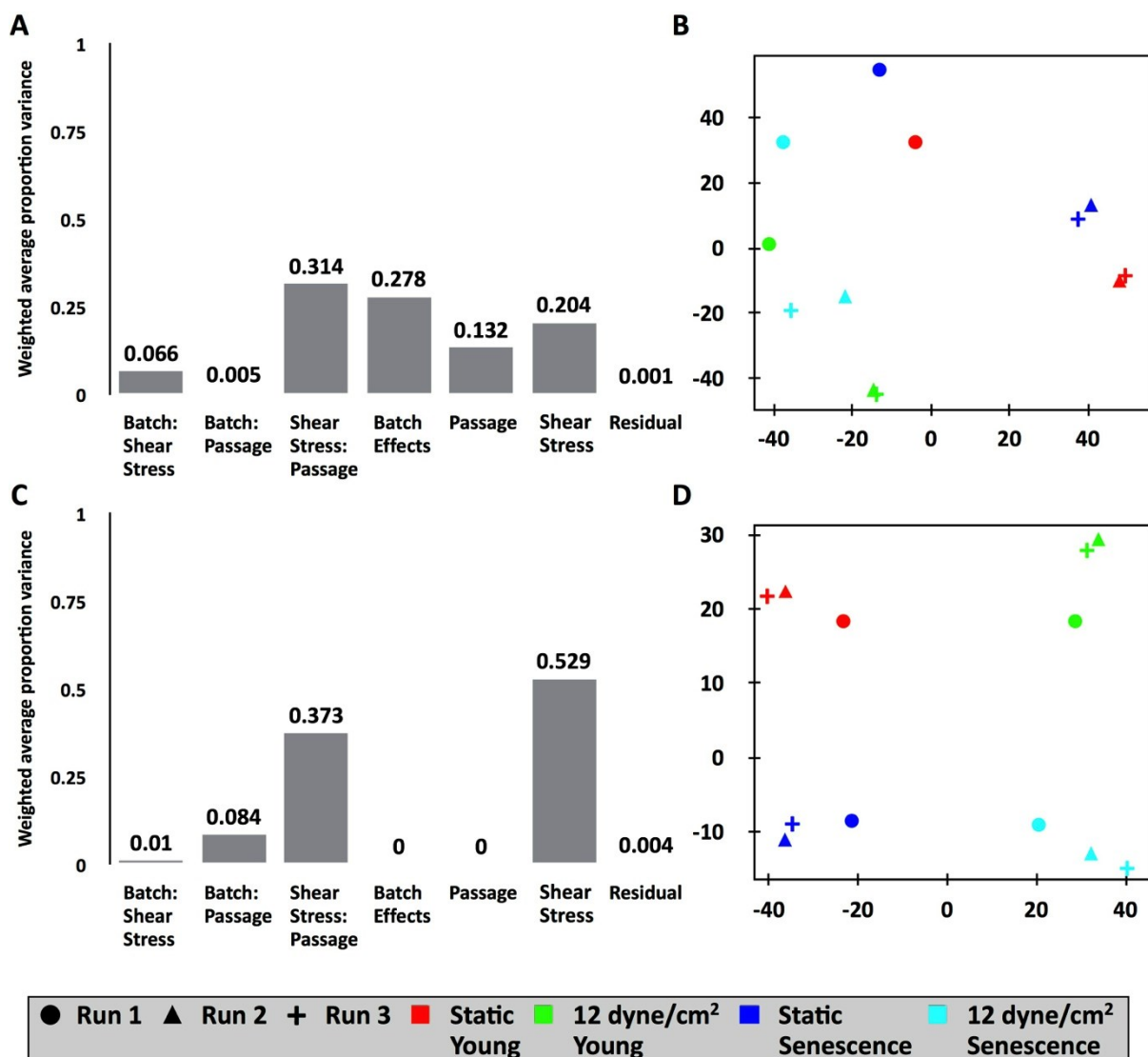


Source	Target			
	Affy U133	Lumi HT12	Lumi Ref8	Ag44k
Bovine	0.97	0.95	0.93	0.9
Mouse	0.93	0.92	0.92	0.86

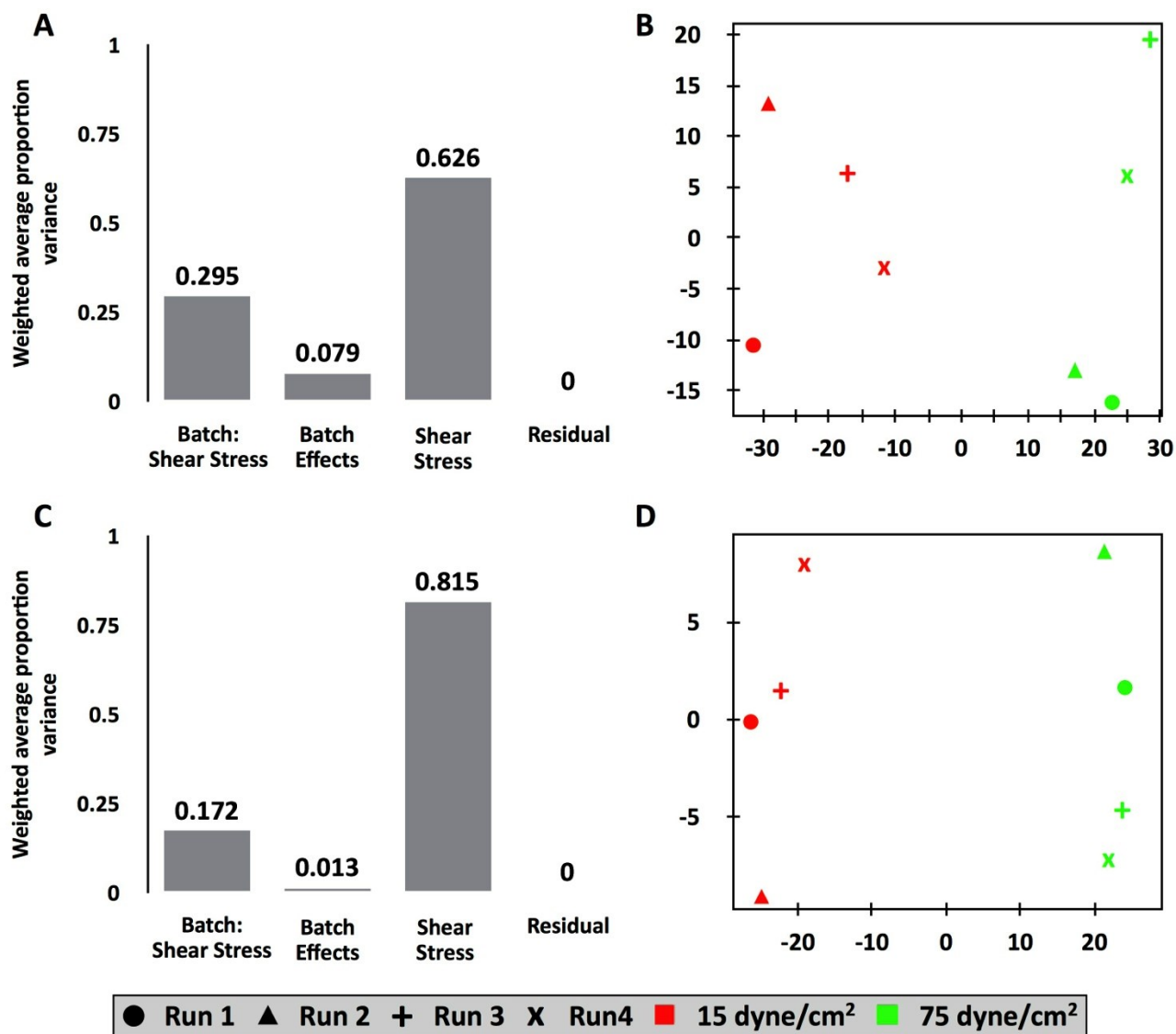
Supplemental Figure 3



Supplemental Figure 4



Supplemental Figure 5



Supplemental Figure 6

