Published in final edited form as:

Methods Enzymol. 2011 ; 489: . doi:10.1016/B978-0-12-385116-1.00007-8.

Discovery Approaches to UPR in Athero-Susceptible Endothelium *In Vivo*

Mete Civelek^{*,†,||}, Elisabetta Manduchi^{‡,§}, Gregory R. Grant^{‡,§}, Christian J. Stoeckert Jr.^{‡,§}, and Peter F. Davies^{*,†,¶}

^{*}Institute for Medicine and Engineering, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[†]Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[‡]Center for Bioinformatics, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[§]Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[¶]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Abstract

The endothelium is a monolayer of cells that lines the entire inner surface of the cardiovascular and lymphatic circulations where it controls normal physiological functions through both systemic and local regulation. Endothelial phenotypes are heterogeneous, dynamic and malleable, properties that in large- and medium-sized arteries lead to a central role in the development of focal and regional atherosclerosis. The endothelial phenotype in athero-susceptible sites is different from that in nearby athero-resistant regions. Understanding the *in vivo* gene, protein, and metabolic expression profiles of susceptible endothelium is, therefore, an important spatiotemporal challenge in atherosclerosis research. Recent studies have demonstrated that endoplasmic reticulum (ER) stress and the UPR are characteristics of susceptible endothelium. Here, we outline global genomic profiling, pathway analyses, and gene connectivity approaches to the identification of UPR and associated pathways as discrete markers of athero-susceptibility in arterial endothelium.

1. Introduction

Atherosclerosis is not a diffuse disease; it has been noted for over a century that lesion development is associated with arterial curvatures, asymmetries, and branches where the nonuniform arterial geometry generates patterns of blood flow that are considerably more complex than elsewhere. Since it is well established that endothelial cells are highly sensitive to flow/shear stress, a hemodynamics contribution to localized susceptibility is likely. Athero-susceptible endothelium *in vivo* expresses a different repertoire of cell phenotypes than that in nearby protected locations (Davies, 2009). Identification of important differences in gene and protein expression and the mechanisms responsible requires both global profiling and classic cell and molecular approaches. Recently, systems biology and discovery science methodologies identified the unfolded protein response (UPR) as a prominent differential component of endothelial phenotype in regions *in vivo* that are susceptible to atherosclerosis. Here, we first summarize the findings that led to UPR

identification. This is followed by a description of the transcriptomics approaches and bioinformatics analyses employed. Mention is then made of the conventional cell and molecular biochemistry used to validate and extend the results of the genomics predictions.

1.1. Site-specific adaptive ER stress and UPR phenotype in athero-susceptible endothelium *in vivo*

ER stress is an adaptive protective mechanism that arises because of excessive protein biosynthesis or interference with normal protein-folding mechanisms in the ER lumen in response to multiple kinds of cellular stress. In athero-susceptible regions of the arterial circulation, such stresses are likely generated by locally complex hemodynamics that create adverse biomechanical forces and, through the formation of flow separation zones, promote the retention of pro-pathological biochemicals such as free radicals. These and other stresses result in excessive newly synthesized and/or misfolded polypeptides in the ER lumen that exceed its protein-folding capacity. The resulting activation of UPR is an ubiquitous adaptive cell response that activates a set of compensatory intracellular signaling pathways. The UPR elicits a coordinated transcriptional upregulation of ER chaperones and folding enzymes to promote the correct assembly of unfolded polypeptides and prevent incompletely folded proteins from aggregating, thereby assisting cell survival in an adverse environment.

In the unstressed state, the ER chaperone Binding Protein (BiP; also known as heat shock protein A5, HSPA5, and glucose-related protein 78, GRP78) binds to each of three ER stress transducers. These are ER transmembrane proteins each having an ER-luminal domain for the sensing of unfolded proteins and a cytosolic domain for signaling. Bound BiP maintains the inactive state of the transducers. In ER stress, BiP dissociates from the chaperones to bind unfolded/misfolded polypeptides in the ER lumen, causing chaperone phosphorylation. Activation of ATF6 (activating transcription factor 6), IRE1 (inositol requiring kinase 1), and PERK (protein kinase-like ER kinase) together with the downstream consequences of their activation constitute the UPR. The products of the activated UPR transducers converge as transcriptional regulators in the nucleus to upregulate ER chaperones and UPR transducer synthesis and to ubiquitinate unfolded proteins for degradation through the proteasome; both processes relieve ER stress accumulation of unfolded proteins and restore ER protein equilibrium to a normal range. Failure leads to apoptosis through transcriptional induction of the transcription factor CHOP (C/ERB homologous protein), inflammation through activation of NFkB, and generation of reactive oxygen species (ROS) through excessive protein oxidation in the ER (Malhotra and Kaufman, 2007).

1.2. Genomics approach to arterial endothelial phenotype

In a multisite study (Civelek *et al.*, 2009) on 45 normal adult swine, endothelium in susceptible regions of the aortic arch (AA), proximal brachiocephalic artery, aorto-renal branch region, and abdominal aorta were analyzed relative to protected sites of the common carotid artery, descending thoracic aorta (DT), and the distal renal artery. All athero-susceptible regions are associated with complex disturbed blood flow. From this multisite study, the most abundant common feature of the endothelium of all athero-susceptible regions was the upregulation of genes associated with ER processing of proteins, ER stress, and the UPR. Differential gene expression analysis identified a highly connected and coordinated network of genes upregulated in the susceptible regions. Three independent pathway mining approaches—Gene Ontology (GO) terms overrepresentation (using DAVID), gene set enrichment analysis (GSEA), and Ingenuity Pathway Analysis (IPA)—identified ER stress and the UPR to be overrepresented functional categories in athero-susceptible endothelium including genes that function in protein folding, synthesis, and posttranslational protein modification.

To validate the transcriptome analyses, endothelial cell proteins were isolated from AA and DT and also from the athero-susceptible aorto-renal branch and the protected distal renal artery. At each athero-susceptible disturbed flow site, BiP transcript and/or protein expression was significantly upregulated. Western blot demonstrated significantly elevated phospho ATF6, phospho IRE1 and its target, spliced XBP-1. However, the third transducer pathway PERK was not activated. Overall, this study, approached without preconceived expectations of differential expression of genes and proteins associated with ER stress/UPR, suggests that stresses associated with flow disturbance *in vivo* elicit activation of the UPR, an ER response common to other forms of stress, and that chronic UPR is a signature for athero-susceptible endothelial phenotype *in vivo*. A schematic of the overall experimental approach is outlined in Fig. 7.1. The reader is also referred to Civelek *et al.* (2009) for specific UPR output data associated with the procedures outlined in this chapter in addition to the public domain bioinformatics data sets referred to in the text.

2. Procedures for the Isolation of Endothelial Cells and Preparation of RNA for Microarray Hybridization

In this section, the endothelial isolation procedure from swine arterial tissue is outlined. Contamination by smooth muscle cells and leukocytes should be avoided. In order to preserve the integrity of RNA for down-stream applications, RNase-free conditions must be strictly followed. We recommend preparing buffers in DEPC-treated water, cleaning the surgical instruments with RNAaseZap (Ambion, Inc.), and general care for avoiding contamination by changing gloves frequently.

2.1. Arterial tissue preparation

Ascending, descending, and abdominal aortas with their branches and carotid artery are harvested within 30–45 min after animal death. Immediately following excision from the animal, they should be cooled on ice to inhibit metabolism. The vessel lumen is rinsed with ice-cold RNAase-free PBS. Surrounding tissue is dissected away and the vessels are cut open longitudinally with blunt artery scissors (Fine Scientific Tools) to minimize damage to endothelial cells then pinned onto waxed trays and rinsed once again with cold RNase-free PBS.

2.2. Endothelial cell harvest

Endothelial cells are gently scraped (fine scalpel blade angled in the direction of the “stroke”) from discrete arterial regions identified to correspond to sites susceptible to atherosclerosis (or any other variable of interest to the investigator). For our work, this information is known from the geometry of the arteries, various *in vivo* imaging modalities of blood flow (where flow disturbances predict sites of athero-susceptibility), and by direct observations of the nuclear shape of the endothelial cells. The last is conducted by immersing the artery for 1 min in a solution of 10 µg/ml DAPI (4,6-diamidino-2-phenylindole) dye (Sigma) in PBS, a fluorescent stain that binds strongly to DNA; flow disturbance is indicated by the absence of alignment between endothelial cells in contrast to most arterial locations where the nuclear alignment reflects directional undisturbed laminar flow (Passerini *et al.*, 2004). For most of our studies, sample size ranges between 0.1 and 1 cm² (several hundred to ~10,000 cells). Since small regions yield insufficient endothelial RNA for direct labeling and hybridization to micro-arrays, mRNA amplification is necessary (below). We have retained good fidelity of linear amplification from as little as 1 ng total RNA which corresponds to about 100 cells. To avoid sample collection and processing bias as well as other confounding factors, it is advisable if possible to harvest randomly from comparative groups of animals. Although this approach introduces increased variance in gene expression, the emerging results are likely to represent that of a general

population. In our studies, no more than three samples came from the same animal; in most cases, each sample was from a single animal (45 animals in total). Scrapes from various regions from multiple animals were pooled to obtain a reference sample for microarray hybridizations.

2.3. Assessing endothelial cell purity

Periodically throughout the harvest procedure, samples of scraped endothelial cells are spread on glass microscope slides and fixed in ice-cold acetone. In order to assess the purity, cells are double stained using an anti-porcine CD31 and von Willebrand factor antibodies for the detection of endothelium. This double staining is important since certain leukocyte subtypes also express CD31. Anti-porcine alpha-actin antibody for the detection of smooth muscle cells and anti-porcine CD45 antibody for the detection of leukocytes are used. DAPI staining indicates the total number of isolated cells in each optical field on the slide. Counting of the cells with different staining patterns indicates the presence/absence of contaminating tissue and blood cells. The gentle scraping isolation technique yields on average 96.5% endothelium (Fig. 7.2).

2.4. Endothelial RNA extraction and quality control

Freshly isolated cells are transferred directly to a lysis buffer containing the RNase inhibitors guanidine isothiocyanate and β -mercaptoethanol (0.143 M; Absolutely RNA Nanoprep Kit, Stratagene, La Jolla, CA) and stored on dry ice. Total RNA is isolated using the Absolutely RNA Nanoprep or Microprep kit depending on the size of the samples (cell numbers) according to manufacturer's instructions. Briefly, an equal volume of 70% RNAase-free ethanol is added to thawed cell lysates. They are loaded onto a silica-based fiber matrix, which binds RNA during centrifugation. Contaminating DNA is digested by a 15-min DNase treatment at 37 °C. Proteins and DNA are removed by high- and low-salt buffer washes. Total RNA is purified by two successive elutions of 10 μ l for Nanoprep or 50 μ l for Microprep kits in 65 °C elution buffer and RNAase-free water. Our experience suggests that warming the elution buffer to 65 °C greatly enhances the amount of recovered RNA from cells.

2.5. Integrity of isolated endothelial RNA

Total RNA integrity is evaluated using an Agilent Bioanalyzer 2100 and RNA 6000 Nano Labchips (Agilent Technologies, Palo Alto, CA) according to manufacturer's instructions. Total RNA is judged to be intact if two ribosomal bands (28S and 18S) are present in approximately a 2:1 ratio and if the RNA integrity number is above 9.5. RNA quantity is measured using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Inc., Rockland, DE). Total RNA with 260/280 and 260/230 wavelength (nm) ratios, which denote the purity of isolated nucleic acid, between 1.8 and 2.1 and concentrations higher than 10 ng/ μ l is used in subsequent procedures.

2.6. Messenger RNA amplification and evaluation

It is not possible to obtain enough RNA from small arterial regions used in this study. Therefore, linear RNA amplification is used to increase the RNA amount for subsequent microarray hybridizations while preserving the ratio of transcripts intact. Total RNA (range 1–100 ng) is amplified using the MessageAmp aRNA Kit (Ambion, Austin, TX). This procedure is based upon the antisense RNA (aRNA) linear amplification procedure described by Van Gelder *et al.* (1990). Poly(A)RNA is reverse-transcribed with an oligo (dT) primer containing a T7RNA polymerase promoter sequence. RNase H treatment cleaves the mRNA into small fragments that serve as primers during second-strand synthesis, resulting in a double-stranded cDNA template for T7-mediated linear

amplification by *in vitro* transcription. Aminoallyl UTP nucleotides are incorporated for subsequent dye conjugation steps. Typically 2–5 µg aRNA is produced from one round of amplification of 100 ng RNA; a second round of amplification for small samples yields substantial amounts (>20 µg) of aRNA with little loss of fidelity (Polacek *et al.*, 2003). aRNA is quantified using the Nanodrop ND-1000 and evaluated for size distribution using Agilent RNA 6000 Nano Labchips. In addition to the sample preps, *pooled reference* RNA is amplified once, collected and frozen in 10 µg aliquots for reference use.

2.7. Amplified RNA fluorescent dye conjugation

Amplified RNA is dried using a vacuum dryer at low heat setting. Higher temperatures degrade RNA and, therefore, should be avoided. RNA is reconstituted in 9 µl coupling buffer (from the MessageAmp kit). Mono-functional NHS ester Cy3 or Cy5 dye, reconstituted in 11 µl DMSO, is added (Amersham CyTM Dye Post-labeling Reactive Pack, GE Healthcare, UK). Reference aRNA and sample aRNA are labeled with Cy3 and Cy5 fluorescent dyes, respectively. Samples are incubated in the dark for 30 min at room temperature. Hydroxylamine (4.5 µl; 4 molar) is added for 15 min to quench the dye coupling reaction. Dye-coupled RNA is purified into nuclease-free water using the aRNA filter cartridges (MessageAmp kit) in order to remove excess dye. Final volume of dye-conjugated RNA is typically 150 µl.

3. Microarray Hybridization and Feature Extraction

A reference design is used in which samples are labeled with Cy5 and the amplified RNA from pig common reference RNA is labeled with Cy3. Pig common reference RNA consists of aRNA amplified from endothelial total RNA which is pooled from all arterial sites in the study. Samples are vacuum dried to 27 µl at a low heat setting. In order to facilitate hybridization efficiency, samples are fragmented using fragmentation reagents (Ambion, Cat# 8740, Austin, TX). Three microliter of 10× fragmentation reagent is added and the samples are incubated at 70 °C for 15 min. The fragmentation reaction is inhibited by the addition of 3 µl stop reagent. Each Cy5-conjugated sample is combined with an equal volume of Cy3-conjugated reference RNA. Nuclease-free water is added to each sample to a total volume of 70 µl followed by the addition of 1 µl of 10 mg/ml herring sperm DNA. Seventy-one microliter of 2× hybridization buffer (Proteomics Research Services, PRS-16003050) is added to the samples which are incubated at 95 °C for 5 min followed by centrifugation at 10,000 *g* for 1 min. Samples are loaded onto custom-printed or commercial swine microarrays. Microarrays are hybridized in a Genomics Solution Hyb Station (Ann Arbor, MI) using a step-down protocol (42, 35, 30 °C each for 5 h) which ensures uniform hybridization across all probes. Microarrays are later washed with medium stringency buffer (PRS-16004001, Proteomics Research Services, Ann Arbor, MI) at 30 °C for 2 min, followed by a high stringency buffer wash (PRS-16004501, Proteomics Research Services, Ann Arbor, MI) at 25 °C for 2 min. Finally, microarrays are washed with post wash buffer (PRS-16003501, Proteomics Research Services, Ann Arbor, MI) at 25 °C for 2 min. They are dipped into deionized water for 30 s and dried by centrifugation at 500 µg for 1 min. In our study, a total of 98 samples were hybridized. Sample hybridizations were performed in batches of 10–12 in random order over several days to avoid experimental bias (day-to-day variations in hybridization conditions occur remarkably arbitrarily despite best practices).

Arrays are scanned with an Agilent DNA Microarray Scanner at 5 µm resolution (single pass) with 100% laser power and 100% photo multiplier tube sensitivity. Images are analyzed with Agilent Feature Extraction Software (version 9.1) with raw fluorescence intensity values determined using the “CookieCutter” method of spot analysis. Each TIF image file is examined for the quality of hybridization. If artifacts (e.g., substantial uneven hybridization) are present, those microarrays are discarded. GAL file grid, which contains

probe annotations, is fitted by hand for each microarray image to ensure correct alignment for each spot.

4. Bioinformatics Analysis

4.1. Annotation of the porcine microarray

Since the pig genome is fully sequenced, commercial swine microarrays are now readily available and often cost effective; manufacturer's procedures should be followed. We custom-print porcine oligonucleotide microarrays (University of Pennsylvania Microarray Core Facility) using Qiagen's Pig Array-Ready Oligo Set on 16.94 × 52.94 mm Codelink slides. This set includes 70-mer probes for 10,665 genes from The Institute for Genomic Research (TIGR) porcine database. The average melting temperature for the primers is 78 °C. They are designed to have minimal hairpin structure and cross-hybridization and are 3' biased (within 1000 base pairs of the 3' end) to ensure adequate signal tolerance for some RNA degradation. Several control Cy3 spots, Stratagene alien controls, and 133 custom 70-mer probes including genes known to play key roles in endothelial function and in atherosclerosis are also printed onto the microarrays. Oligos are suspended in 50 mM sodium phosphate buffer at a final concentration of 8.33 μM for printing. The resulting array has 12,288 spots arranged in 32 subgrids, each with 17 rows and 22 columns.

Several sequencing projects have contributed to the sequencing of the porcine transcriptome in recent years (Tuggle *et al.*, 2007). Frequent updating of the porcine expressed sequence tags (ESTs) necessitated the most recent annotation of the 70mers used in printing the microarrays for subsequent bioinformatics analyses. Each of the printed 70mers is derived from 64,746 (currently 110,744) Tentative Clusters (TCs) built from 575,730 (currently >1 million) ESTs and 6854 expressed transcripts from a total of 257 cDNA libraries (*Current Porcine Gene Index: SsGI Release 14.0*; March 11, 2010: Dana Farber Cancer Institute; <http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=pig>). Using the 70mer sequences, a text file in FASTA format is created to store the printed oligomer sequence information. In order to obtain the most recent annotation, these sequences are first compared to the available 1185 porcine Reference Sequence Collection (RefSeq) from the National Center for Biotechnology Information (NCBI) using the Basic Alignment Search Tool (BLAST) with a required match of 64 bases with 94% identity. As a result, 830 printed oligomers were annotated with porcine RefSeq IDs. Second, the same sequences were compared to human RefSeqs with a stringency of blast *p*-value less than 0.001. This resulted in the annotation of 4180 printed oligomers with human RefSeq IDs. Third, 70mers were translated into peptide sequences using "blastx" in all six possible reading frames and the resulting peptide sequences were compared to UniProt100 database, which contains the translation of coding sequences of multiple genomic databases. Matches with blast *p*-value of less than 0.001 were retained. Since Uni-Prot100 contains information about multiple species, a word comparison script was used to collapse the matches preferentially to porcine, human, mouse, and rat species for a total of 4111 UniProt IDs. In rare instances, matches for other species are allowed if porcine, human, mouse, or rat were not available. Finally, similar to step three, the TC sequences (instead of the 70mers) are translated into peptide sequences using "blastx" in all six possible reading frames and the resulting peptide sequences were compared to UniProt100. Using these approaches, 8962 of the 10,798 printed oligomers are fully annotated.

Functional annotation of the microarray is achieved by mapping the Uniprot IDs to GO IDs. The GO describes gene products based on their associated biological processes, cellular components, and molecular functions in a species-independent manner. Microarray spots mapped to 3153 unique GO "biological process," 1885 GO "molecular function," and 599 GO "cellular component" IDs.

Information about the microarray and its full annotation can be found with accession number A-CBIL-16 at ArrayExpress <http://www.ebi.ac.uk/microarray-as/ae/>.

4.2. Microarray data preprocessing

One thousand four hundred and ninety spots which correspond to control spots (blanks, Cy3, and Strategene Alien) are first filtered out. For each assay, saturated spots in at least one channel are set to NA. For each channel, the mean signal measure from the Agilent Feature Extraction software is used as input signal intensity. No background subtraction is performed. M and A values are calculated from the raw data using Eqs. (7.1) and (7.2) where R and G are signals of the Cy5 (red) and Cy3 (green) channels, respectively.

$$M = \log_2 R - \log_2 G \quad (7.1)$$

$$A = \frac{\log_2 R + \log_2 G}{2} \quad (7.2)$$

The M values are normalized with print-tip loess normalization using the Bioconductor *marray* package (version 1.12.0) for R (version 2.4.0).

4.3. Analysis of differential gene expression

Differential expression analysis is performed using Patterns of Gene Expression (PaGE version 5.1.6; <http://www.cbil.upenn.edu/PaGE>; Grant *et al.*, 2005). PaGE is a false discovery rate (FDR)-based method of controlling false positives. It uses a permutation-based algorithm to estimate the FDR. In PaGE, for any specified constant, permutations of the data matrix are used to estimate the rate of false positives in any set of genes having a T-statistic greater than the constant. An appropriate constant is chosen to guarantee the desired FDR, typically ranging between 0.25 and 0.05 (75–95% confidence). Confidences are then assigned to all genes in the set. For example, if 100 genes are discovered as differentially expressed with a confidence of 0.75 (i.e., FDR of 0.25), the expected number of false positives is 25. PaGE also produces “levels” of differential expression, based on the confidence parameter. Detailed information about the PaGE algorithm can be found at http://www.cbil.upenn.edu/PaGE/doc/PaGE_documentation_technical_manual.pdf. Differentially expressed genes in various comparisons are obtained using processed M values as input to PaGE.

The above procedures usually result in a long list of genes differentially expressed above a significant threshold set by the investigator through PaGE. Further data processing procedures allow the physiological and pathological implications of subsets of these genes to be evaluated through their commonalities, associations, and network connectivities, and the identification of dominant differential pathways. The use of several of these procedures is next outlined in the context of UPR as a prominent differentially expressed characteristic of athero-susceptible endothelium.

4.4. Identification of enriched biological themes

Differentially expressed genes are interrogated for overrepresented biological themes using database for annotation visualization and integrated discovery (DAVID) and based on GO terms (Huang da *et al.*, 2009). The DAVID functional annotation clustering tool highlights the most relevant GO terms associated with a differentially expressed gene list. Details of the DAVID algorithm can be found at <http://david.abcc.ncifcrf.gov/>.

4.5. Gene set enrichment analysis

GSEA is an algorithm that performs differential expression analysis at the level of gene sets (Subramanian *et al.*, 2005). The input to GSEA consists of a collection of gene sets and microarray expression data with replicates for two conditions to be compared. GSEA employs a permutation-based test which uses Kolmogorov–Smirnov running sum statistic to determine which of the gene sets from the collection are differentially expressed between the two conditions. GSEA differs from differential gene expression analysis in the sense that it might identify genes which are part of a differentially expressed set but which might not be identified as significantly differentially expressed alone. The details of the GSEA algorithm can be found at <http://www.broad.mit.edu/gsea/>.

For GSEA of various comparisons, for example, athero-susceptible versus athero-protected arteries, gene sets made up of 15–500 genes were created using GO mappings of the microarray. A total of 548 GO biological process (BP), 256 GO molecular function (MF), and 142 GO cellular component (CC) gene sets were used for this analysis. For each comparison, GSEA is performed separately for BP, MF, and CC gene sets. Gene sets are analyzed at an FDR set by the investigator.

4.6. Ingenuity pathway analysis

Differentially expressed genes are analyzed for interactions (from scientific literature) with each other and other molecules using IPA (Ingenuity Systems, www.ingenuity.com; Calvano *et al.*, 2005). A list of differentially expressed genes with multispecies UniProt identifiers (to cover homologies) with their corresponding confidence and expression values is uploaded into the IPA application. A confidence cutoff is set by the investigator to identify genes which are used to interrogate the relationships curated to include literature findings. Details of the IPA algorithm can be found at <http://www.ingenuity.com/library/index.html>. In our particular endothelial experiments, IPA identified 73% of the upregulated genes in athero-susceptible endothelium to form a tightly connected network of interactions based on known gene–protein and protein–protein direct relations (Fig. 7.3) many of which are linked to ER stress and UPR (Civelek *et al.*, 2009).

5. Another Approach to Gene Connectivity: Weighted Gene Coexpression Network Analysis

In recent studies (Civelek *et al.*, submitted for publication), we are using an additional approach to global endothelial gene expression across the arterial system. Weighted gene coexpression network analysis (WGCNA) is a systems biology approach to construct a weighted gene coexpression network to identify groups of genes (modules) whose expression is highly correlated using the WGCNA R package (Langfelder and Horvath, 2008). A comprehensive tutorial with several examples for the methodology is available at <http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/Rpackages/WGCNA/>.

Eighty seven samples were used for a recent network analysis. For each probe, a connection strength measure is determined by the pairwise correlations between expression profiles. Overall connectivity (k) for each probe is calculated by taking the sum of its connection strength with all other probes in the network (see WGCNA R tutorial). About 5579 probes with $k > 5$ are used for further analysis (Ghazalpour *et al.*, 2006). A similarity matrix is constructed by calculating the biweight midcorrelation, which is robust to outliers, for all pairwise comparison of probe expression across all microarray samples (Wilcox, 2005). This correlation matrix is then transformed into a matrix of connection strengths (adjacency matrix) using a power function (connection strength = (correlation)^b), resulting in a weighted network. The parameter b was chosen by using the scale-free topology criterion

(Langfelder and Horvath, 2008) to be six such that the resulting network connectivity distribution approximated scale-free topology. The adjacency matrix is used to calculate the topological overlap matrix (TOM) (Yip and Horvath, 2007). Modules are groups of genes with similar patterns of connection strengths with all other genes of the network and “high topological overlap” (Zhang and Horvath, 2005). Modules are identified using the TOM in conjunction with average linkage hierarchical clustering. A dynamic tree-cutting algorithm is used to identify the modules (Langfelder *et al.*, 2008). Each module is summarized by its first eigengene (first principal component of the expression values across samples) and modules with similar gene expression identified by highly correlated eigengenes (correlation coefficient > 0.95) are merged. In our study, samples are grouped according to athero-susceptibility (susceptible vs. protected) or circulatory bed (coronary vs. noncoronary). Significant association of each module with either athero-susceptibility (susceptible = 1 and protected = 0) or circulatory bed (coronary = 1 and noncoronary = 0) is identified by calculating the Pearson correlation of module eigengene and each classification and by calculating the Student asymptotic *p*-value for the given correlations. The significant modules are then assessed for enrichment in GO terms using DAVID based on Fisher’s exact test utilizing the 5579 probe annotations as reference gene list (Huang da *et al.*, 2009). This is then visualized.

5.1. Network visualization

Gene interactions are visualized using Cytoscape (Cline *et al.*, 2007). Intramodular connectivity for each gene is calculated taking the sum of its connection strength with all other genes in the same module. The gene connectivity is scaled by the maximum connectivity strength which results in the gene with the most connections having a connectivity value of 1. Genes with connectivity (*k*) less than 0.3 are filtered, resulting in 3192 genes (nodes) and 48,990 interactions (edges) for network visualization. The layout is obtained using an edge-weighted spring embedded algorithm using connectivity as the edge weight (Fig. 7.4). The distance between two nodes is proportional to the strength of the connection between two nodes, with highly connected genes being the closest to each other. The nodes are colored according to the module colors they are assigned by the WGCNA algorithm. For individual module network visualization, all the interactions between module genes are used without filtering. The node size is proportional to the intramodular connectivity of each gene with the others in the same module. Largest node represents the gene with the highest connectivity and is denoted as the hub gene of the module. This particular approach revealed that endothelium in coronary and noncoronary arterial beds exhibit markedly different gene expression profiles but that transcript profiles in coronary arteries are consistent with the presence of ER stress/UPR, as previously shown in non-coronary arteries.

6. Validation and Follow-up

Identification of differentially expressed pathways using global genomics directs the investigator to pathways, biomarkers, or clusters of molecules of interest for further hypothesis-driven, focused investigations. Typically, the important initial steps in this transition are to validate differential expression of representative genes by quantitative real-time PCR. The development of commercial multiplex-based qRT-PCR instrumentation has replaced gene-by-gene analyses with chip analyses. This can readily be performed on very small numbers of cells (thus retaining spatial relevance). Expression of proteins encoded by genes of interest can be quantified by Western blot if sufficient cells are available (potential loss of spatial resolution). Standard protocols for these procedures are readily available.

Our studies of athero-susceptible endothelial phenotyping of UPR were validated and expanded at both the transcript (by qRT-PCR) and protein (Westerns) levels as outlined in

the Section 1.1 (Civelek *et al.*, 2009). Besides our study, two additional reports using complementary approaches support our findings and provide further mechanistic insights about site-specific endothelial ER stress, hemodynamics, and athero-susceptibility. Feaver *et al.* (2008) used an *in vitro* model to simulate human arterial shear stress waveforms. Athero-susceptible or athero-protective flow was applied to human endothelial cells. BiP (GRP78) was found to be significantly upregulated in a sustained manner under athero-susceptible, but not athero-protective flow up to 24 h. This response was dependent on both sustained activation of p38, as well as integrin α_1 . Increased BiP expression correlated with the activation of the ER stress sensing element promoter by athero-susceptible flow as a marker of the UPR. Shear stress regulation of BiP was through increased protein stability when compared to other flow regulated proteins, such as connexin-43 and vascular cell adhesion molecule (VCAM)-1. Increased endothelial expression of BiP was also observed in athero-susceptible versus athero-protective regions of C57BL6 mice. The study supports a role for the hemodynamic environment in preferentially inducing BiP and the UPR in athero-susceptible regions before lesion development. Zeng *et al.* (2009) reported that spliced XBP-1 (sXBP-1) encodes the XBP-1 transcription factor that translocates to the nucleus to activate selective pro-apoptotic target genes as one of the three transduction arms of the UPR response. Following the observation of endothelial expression of the XBP-1 pathway of UPR in branching regions of apoE^{-/-} mice arteries and in atherosclerotic lesions that developed there, their study reported that athero-susceptible flow waveforms induced XBP-1 splicing in cultured endothelial cells. Overexpression of (activated) sXBP-1 induced apoptosis in cultured human endothelial cells. To extend the findings to an *in vivo* assay for atherogenesis, adenoviral-mediated overexpression of sXBP-1 was induced in an apoE^{-/-} murine aortic isograft model. In these animals, enhanced intimal hyperplasia and atherosclerosis developed in normally protected regions of the aorta suggesting that when the XBP-1/UPR pathway is greatly overstimulated, the adaptive protective function of UPR reverts to a pathological imbalance. While overexpression was not entirely limited to the endothelium in the isograft model, the data are supportive for a prominent role for endothelial sXBP-1. These three different but complementary approaches to endothelial ER stress provide compelling evidence for the existence of site-specific chronic adaptive UPR in endothelial cells *in vivo*; the hemodynamic environment associated with sites of athero-susceptibility likely plays a significant role.

The principles and procedures outlined for global genomics in this chapter can be applied to other levels of regulation in the same cells. For example, we have recently reported differential microRNA analyses of athero-susceptible endothelium by microarray to show posttranscriptional regulation of pro-inflammatory phenotypes by these small RNAs without preconceived expectation of their target pathways (Fang *et al.*, 2010).

Acknowledgments

We gratefully acknowledge American Heart Association Fellowship support to MC (0315286U) and National Institute of Health grants HL062250 (PFD) and HG004521 (CS, EM, GG).

REFERENCES

- Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, Miller-Graziano C, Moldawer LL, et al. A network-based analysis of systemic inflammation in humans. *Nature*. 2005; 437:1032–1037. [PubMed: 16136080]
- Civelek M, Manduchi E, Riley RJ, Stoeckert CJ Jr, Davies PF. Chronic endoplasmic reticulum stress activates unfolded protein response in arterial endothelium in regions of susceptibility to atherosclerosis. *Circ. Res.* 2009; 105:453–461. [PubMed: 19661457]

- Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, et al. Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* 2007; 2:2366–2382. [PubMed: 17947979]
- Davies PF. Hemodynamic shear stress and the endothelium in cardiovascular pathophysiology. *Nat. Clin. Pract. Cardiovasc. Med.* 2009; 6:16–26. [PubMed: 19029993]
- Fang Y, Shi C, Manduchi E, Civelek M, Davies PF. MicroRNA-10a regulation of proinflammatory phenotype in athero-susceptible endothelium in vivo and in vitro. *Proc. Natl. Acad. Sci. USA.* 2010; 107:13450–13455. [PubMed: 20624982]
- Feaver RE, Hastings NE, Pryor A, Blackman BR. GRP78 upregulation by atheroprone shear stress via p38-, α 2beta1-dependent mechanism in endothelial cells. *Arterioscler. Thromb. Vasc. Biol.* 2008; 28:1534–1541. [PubMed: 18556570]
- Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt EE, Drake TA, Lusis AJ, Horvath S. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet.* 2006; 2:e130. [PubMed: 16934000]
- Grant GR, Liu J, Stoekert CJ Jr. A practical false discovery rate approach to identifying patterns of differential expression in microarray data. *Bioinformatics.* 2005; 21:2684–2690. [PubMed: 15797908]
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 2009; 4:44–57. [PubMed: 19131956]
- Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* 2008; 9:559.
- Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics.* 2008; 24:719–720. [PubMed: 18024473]
- Malhotra JD, Kaufman RJ. The endoplasmic reticulum and the unfolded protein response. *Semin. Cell Dev. Biol.* 2007; 18:716–731. [PubMed: 18023214]
- Passerini AG, Polacek DC, Shi C, Francesco NM, Manduchi E, Grant G, Pritchard WP, Powell SJ, Chang G, Stoekert C, Davies PF. Coexisting pro-inflammatory and anti-oxidative endothelial transcription profiles in a disturbed flow region of the adult porcine aorta. *Proc. Natl. Acad. Sci. USA.* 2004; 101:2482–2487. [PubMed: 14983035]
- Polacek DC, Passerini AG, Shi C, Francesco NM, Manduchi E, Grant GR, Powell S, Bischof H, Winkler H, Stoekert CJ Jr, Davies PF. Fidelity and enhanced sensitivity of differential transcription profiles following linear amplification of nanogram amounts of endothelial mRNA. *Physiol. Genomics.* 2003; 13:147–156. [PubMed: 12700361]
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA.* 2005; 102:15545–15550. [PubMed: 16199517]
- Tuggle CK, Wang Y, Couture O. Advances in swine transcriptomics. *Int. J. Biol. Sci.* 2007; 3:132–152. [PubMed: 17384733]
- Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA.* 1990; 87:1663–1667. [PubMed: 1689846]
- Wilcox, RR. *Introduction to Robust Estimation and Hypothesis Testing.* second edn.. Elsevier/Academic Press; Amsterdam, Boston: 2005.
- Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinform.* 2007; 8:22.
- Zeng L, Zampetaki A, Margariti A, Pepe AE, Alam S, Martin D, Xiao Q, Wang W, Jin ZG, Cockerill G, Mori K, Li YS, et al. Sustained activation of XBP1 splicing leads to endothelial apoptosis and atherosclerosis development in response to disturbed flow. *Proc. Natl. Acad. Sci. USA.* 2009; 106:8326–8331. [PubMed: 19416856]
- Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 2005; 4:1–45.

Figure 7.1.

Schematic outline of discovery approach to site-specific endothelial phenotype in swine arteries. Regions of athero-susceptibility and resistance are indicated in the coronary arteries, aorta, and renal arteries. (A) Heart and great vessels, from left to right: distal right coronary (RC) artery, proximal RC, aortic arch, proximal left coronary artery, proximal left anterior descending (LAD) coronary artery, distal LAD, distal circumflex artery. (B and C) (opened) Aortic arch and descending thoracic aorta. (D and E) (opened) Proximal and distal regions of renal arteries. Transcriptome analyses of endothelium isolated from discrete regions identified protein biosynthesis, ER stress, and UPR as dominant pathway phenotypes differentially expressed in regions of athero-susceptibility. Transcript validation by quantitative real-time PCR (qRT-PCR) is followed by isolation of endothelial proteins for Western blotting, to determine regional protein and phospho-protein expression differences (indicative of UPR activation), and immunohistochemistry (IHC) of proteins *in situ*.

Figure 7.2.

Purity of isolated endothelial samples. Scraped samples were fixed onto microscope slides. Purity and contamination was assessed by antibody staining. (Left panel) CD31 (green) and von Willebrand Factor (vWF) (red) double staining for endothelial cells. (Middle panel) - smooth muscle actin (green) staining for smooth muscle. (Right panel) CD45 (green) staining for leukocytes. Average endothelial purity was 96.5% with 2.78% smooth muscle cell and 0.72% leukocyte contamination. Nuclei were observed with blue Hoechst 33258 staining. Red arrows show a smooth muscle cell and a leukocyte. Bar = 20 μm .

Figure 7.3.

IPA of endothelial gene expression. A dataset containing gene identifiers with corresponding expression values and FDRs were analyzed using IPA. An FDR value cutoff of 25% was set to identify genes whose expression was significantly differentially regulated. Networks of these focus genes were then algorithmically generated based on their connectivity. Only direct interactions of gene–protein and protein–protein were considered. Seven networks were combined to form one large network. Red color indicates EC gene upregulation in athero-susceptible regions of arteries compared to athero-protected regions. Intensity of red is proportional to expression ratio. Gray color indicates molecules present in the data set that were not significantly differentially expressed (FDR > 25%).

Figure 7.4.

Arterial endothelium gene coexpression network. Highly connected genes were portioned into modules of similar expression profiles using the WGCNA algorithm. An interaction network was constructed using gene connectivity between two genes as the distance of an edge between two genes (nodes). The colors indicate different modules of highly correlated endothelial genes.