May 2010

Identification of candidate genes involved in atherosclerotic cardovascular disease : a systems biology appproach

Pavana Anur

IDENTIFICATION OF CANDIDATE GENES INVOLVED IN ATHEROSCLEROTIC CARDIOVASCULAR DISEASE; A SYSTEMS BIOLOGY APPROACH

By

Pavana Anur

A thesis presented to the

Department of Medical Informatics and Clinical Epidemiology and the Oregon Health and Science University School of Medicine in partial fulfillment of requirements for the degree of

Masters of Science

May 2010

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master's thesis of

Pavana Anur

"IDENTIFICATION OF CANDIDATE GENES INVOLVED IN ATHEROSCLEROTIC CARDIOVASCULAR DISEASE; A SYSTEMS BIOLOGY APPROACH"

has been approved

Mentor/Advisor

Member

Member

Member

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

I would like to thank my adviser Dr.Shannon McWeeney for constant encouragement and guidance through the process of this thesis. I would also like to thank my committee members Dr.Beth Wilmot for statistical inputs, Dr. Kemal Sonmez for providing systems biology perspective for the research question and Dr. Balaji Tamarappoo for the help with biology of atherosclerosis. I also extend my acknowledgements to DMICE staff for all their support.

Abstract

Gene expression profiles and mass-spectrometry analysis of a plaque have revealed genes and proteins involved in an atherosclerotic plaque progression. Though differentially expressed genes in an atherosclerotic plaque progression have been identified previously, this is the first study that has combined gene expression data with proteins identified from different stages of plaque progression to identify pathways and candidate genes affecting state of a plaque using a network approach. We constructed an unstable plaque network with 59 genes of which 9 were transcription factors, 10 genes were over expressed and 12 genes were under expressed. Of all genes and proteins found to be associated with a plaque progression, we further identified and prioritized candidate genes which could be potential biomarker or drug targets to prevent progression of a plaque. In this study we also show that just differential expression of genes does not identify genes which are associated with the disease state but not differentially expressed.

CHAPTER 1 – INTRODUCTION

1.1 Systems biology

A system is composed of several parts. A biological system is composed of genome, transcriptome and proteome. Understanding biological system as a whole in contrast to understanding just the parts is called systems biology (Kitano, 2002). Biological systems exhibit complex traits due to modifications of its components like positive and negative regulation, phosphorylation and acetylation, transcription and translational controls. In addition biological systems are robust, that is they are not affected by slight modifications in the system, in spite of the environmental or genetic pressures, they are under (Kitano, 2004). To get a comprehensive view of molecular events driving biological processes we need to understand systems dynamics, structure and key factors which drive the system.

Using traditional bioinformatics like microarrays and mass spectrometry, numerous genes and proteins, which could be key drivers in a biological system causing disease were identified. Although the techniques revealed significant genes and proteins associated with a disease they do not reflect their relationships in a complex interacting system (T Ideker et al, 2001). With increase in quantitative high-throughput biological tools and availability of databases like KEGG (Kanehisa & Goto, 2000), BIND(Bader et al., 2001), STRING(Szklarczyk et al, 2010) and HPRD (Peri et al, 2003) a systems level integrating approach would reveal the

architecture below a complex system, thereby providing insights into molecular mechanisms and key drivers which in turn could be potential drug targets (Leroy Hood & Perlmutter, 2004).

A map in which the components of a system are represented as a node and the interactions or relations between them represented as edges is called a network. Networks provide simple visual representations of complex biological system. According to Barabasi when a new node is added to a network it preferentially attaches to a node with high connections just as a new webpage is likely to have connections to a most popular websites. Hence networks exhibit a scale free topology following a power law $P(k) \sim ck^{-}$ where P(K) is the fraction of node in the network with K connections, c is a constant and parameter whose value ranges from 2 to 3(Barabasi et al , 2000). Assembling genes and proteins into a scale free network is one of the systems approaches which allow us to identify key nodal points in a network.

Scale free network exhibit central-lethality rule. According to central-lethality rule, in nature, essential proteins tend to be more connected than non essential proteins. These proteins appear as hubs in a Protein-Protein Interaction (PPI) network. Additionally essential proteins are evolutionarily conserved and deletion of this essential protein is more lethal than a non-essential protein. Hence the essential proteins are of structural and functional importance. Also in a human interactome study it has been proven that proteins associated with human diseases are more interconnected than a non-disease protein (X. He & Jianzhi Zhang, 2006). Thus hub protein could serve as potential biomarkers for a disease.

1.2 Atherosclerosis

Coronary heart disease accounts for 51% of deaths in United States according to statistics from National Centre for Health Statistics (NCHS) [Figure 1]. Statistics from NCHS also indicate that the prevalence rate of cardiovascular disease is around 80% among people of age 60 to 80 [Figure 2]. Additionally American Heart Association (AHA) has predicted that the cost of treating cardiovascular disease will increase from $172 billion spent in 2010 to $276 billion in 2030, 61% increase in treatment costs adding financial burden on American economy (Heidenreich et al, 2011). Among the cardiovascular disease, the biggest burden of disease lies in Myocardial infarction. Atherosclerosis is the leading cause of myocardial infarction or more commonly heart attack (Lusis, 2010).

Atherosclerosis commonly referred to as hardening of the arteries is a condition affecting the arterial walls as a result of lipid deposition and chronic inflammation, which leads to formation of multiple plaques. An atherosclerotic plaque consists of a lipid core surrounded by connective tissue. Lipid core mainly composed of cholesterol, macrophage derived foam cells, cytokines and metalloproteinases. Connective tissue surrounding the lipid core is derived from smooth muscle cells. American Heart Association has classified atherosclerotic plaques into six types based on histological observations (Stary et al, 1995). Type I, II and III there is some amount of lipid deposition, endothelium remains intact. Type IV there is increase in lipid core and intimal disorganization. Type V is marked by formation of fibrous connective tissue around lipid core. Plaques with additional disruptions on the surface are

classified as type VI. Some cases calcium increases in fibrous connective tissues; such plaques are called calcific plaques. Plaques can be further classified into two distinct types of plaque, stable and unstable plaque. An unstable plaque is one in which a plaque ruptures causing blockage of coronary arteries leading to myocardial infarction and stable plaque is a plaque which does not rupture. Type IV, V and VI can be considered unstable plaques.

Atherosclerosis begins with accumulation of lipids, invasion of immune cells and conversion of smooth muscle cells into fibrous cells. Epidemiological and genetic studies have indicated several environmental and genetic factors affect atherosclerosis (Lusis, 2010). Hence atherosclerosis is a multifactorial disease; more effective strategies are required not only to prevent atherosclerosis, but also to avoid the complications arising from it.

1.3 Pathology of unstable plaque

An artery is composed of endothelial cells , the middle Smooth Muscle Cells and external adventia. Inner most layers of endothelial cells are called intima. Atherosclerosis begins when Low Density Lipoproteins (LDL) and cholesterol begin to accumulate on the artery walls (Intima). The accumulated lipoproteins get oxidized and the modified lipoproteins incites a inflammatory response in the endothelial cells resulting in release of proteins like cytokines (Gargalovic et al, 2006). This enables cells of the immune system (macrophages, T-lymphocytes, monocytes) to migrate through the endothelial cells and enter intima. The macrophage takes in the lipoprotein molecules and becomes foam cells which mark beginning

of plaque formation. Smooth muscle cells migrate from middle layers towards the intima and secrete fibrous collagen which forms a fibrous cap and forming a '**plaque**'. The foam cells and immune cells within the plaque activates an inflammatory response and release enzymes which cause rupture of fibrous cap resulting in thrombus formation, which when severe causes myocardial infarction or heart attack. In some cases it is observed that the fibrous cap is thicker and the plaque is less vulnerable to rupture. This type of unruptured plaques is called Stable Plaque and plaques that rupture are Unstable Plaque [Figure 3].

According to pathophysiological understanding instability in plaque is due to inflammation, hypoxia (decreased oxygenation) and micro vessel formation (Sluimer & Daemen, 2009) . However the exact mechanism or biology of progression of stable plaque to unstable plaques is yet not well understood. Several attempts to this at genetic level have been done. Plaque biology has been dissected with traditional bioinformatics approaches like microarray and mass spectrometry. Transcription studies have revealed significant mRNAs that might be causing plaque instability (Sluimer et al, 2007). Proteomic studies have attempted to identify proteins involved in transforming a stable plaque into an unstable plaque. (Lepedda et al, 2009;Bagnato et al, 2007;Slevin et al, 2006). For a comprehensive understanding of underlying molecular mechanism in progression of a plaque from stable to unstable state we integrate results from transcriptome and proteomic studies.

1.4 Integrative approach

In the past integrative analysis of Protein Protein Interaction (PPI) network and gene expression has given insights into functional roles underlying a pathological condition. A example of this approach is a human heart failure study where in a global PPI of human heart failure was assembled and the relations to differentially expressed genes was analyzed, leading to insights into underlying mechanisms of heart failure and identification of drug targets (Camargo & Azuaje, 2007). Another similar approach was taken combining expression profiles of pancreatic cancer and proteins associated with pancreatic cancer to find biomarkers of pancreatic cancer (Harsha et al, 2009). The goal of this study is to use a similar approach of integrating PPI and Differentially Expressed (DE) genes to identify pathways and genes involved in progression of stable plaque to unstable plaque.

We hypothesize in addition to significant differentially expressed genes there could be genes which are not differentially expressed but are functionally important, an example for this case is a transcription factor, though it controls gene expression it may not be identified in a differential expression study as they are present in low amounts within a cell. Differentially expressed gene could affect other gene expression or processes, a example for this case can be a gene differentially expressed activating other processes like inflammation but the genes involved in the process may not be differentially expressed [Figure 4]. In this study further investigation on the candidate genes in unstable areas of the plaque is done using network

approach to identify candidate gene associated with progression of a plaque but may not be differentially expressed.

1.5 Transcription factors

Biological information flow from DNA follows central dogma. Flow of information from DNA to RNA is called transcription. Flow of information from RNA to proteins is called translation. Transcription is a complex process involving RNA polymerase and other regulating factors called transcription factors. Rate of transcription is controlled by sequence preceding a gene. This region is called promoter region. Promoter region contains Upstream Activation Sequence (UAS) or Upstream Repressing Sequence, based on binding of transcription factors to these sequences, transcription is activated or repressed. Hence transcription factors affect rate of transcription by binding specifically on a promoter (Lee & Young, 2000).

In atherosclerotic plaque progression from stable to unstable state, ever stage can have characteristic gene expression profiles. Presences of transcription factor or combination of transcription factors present in plaques govern these expression profiles by stimulating multiple rounds of transcription of a gene or by repressing gene expression. Hence to understand molecular mechanism of transcription regulation in unstable plaque, identification of transcription factor binding sites or transcription factors present in unstable plaque is required.

1.6 Biomarker

Technological advances are moving trend in medicine towards predictive, preventive and personalized medicine. To achieve this goal, indicators for abnormal state or a pathological condition has to be known. In medicine a Biomarker is considered as indicator for pathological condition, proteins whose expression are high in a pathological condition and can be measured quantitatively are used as biomarkers. Disease based biomarkers have been identified based on protein expression. Matrix metalloproteins and P53 proteins are examples of disease based biomarkers (Jezierska & Motyl, 2009;Gasparini et al, 2011). Their expression is found to be significant during early stages of breast-cancer, hence can serve as a biomarker. Though there has been significant success to the endeavor of finding disease related biomarkers through gene and protein expression measurements yet several challenges remains due to underlying complexity of the disease. Measuring genes and proteins separately does not characterize entire molecular mechanisms involved during progression of a disease. Moreover it is important to identify a disease specific potential biomarker, not all proteins associated with a disease serve as biomarker. In this study we address this issue with respect to atherosclerotic plaque progression from stable to unstable state. We specifically integrate gene expression from stable versus unstable plaques and proteins from different stages of plaque progression to identify potential biomarkers indicating progression of a plaque.

CHAPTER 2 – METHODS

2.1 Literature search for gene expression data

Two databases of gene expression data, Gene Expression Omibus (GEO) and Array Express were searched for mRNA expression experiments on stable versus unstable atherscleroticplaque.

Further selection of expression study was limited to those satisfying following criteria:

1. Stable and unstable plaque used in th study should be from same patient.
2. mRNA extaracted should be from all different cells of a plaque (endothelial cells, smooth muscle cells, macrpphages etc)

These criteria avoids interpatient variability .

Two gene expression studies were found satisfying above criteria. First gene expression data was got from supplementary tables of differential expression study of stable and unstable plaques, raw data for this study was unavailable (Papaspyridonos et al, 2006) . Plaques in this study were classified by macroscopic observations by vascular surgeons. Plaque with intact fibrous cap was classified stable and those with ulcerated surface were classified unstable. This classification was further confirmed by histological analysis. Total of 11 plaques were used for RNA extraction, 4 of them were stable and 7 unstable. The plaques were obtained from 3 patients. RNA extracted was hybridized to Affymetrix array U133. Though

interplaque and intraplaque approach was used in this study, we used expression results from intraplaque analysis where in plaques from same patient was used.

Second gene expression data was got from array express , accession number E-MEXP-268 (Ijäs et al, 2007). Plaques in this study were identified based on prior clinical symptoms. Patient who suffered ipsilateral stroke or ischemia were considered positive for presence of unstable plaque. From 4 positive patients both stable and unstable plaques were obtained for the study. These plaques were further subjected to histological examinations. RNA from these plaques was hybridized to Affymetrix U95Av2 arrays.

CEL files obtained from array express was normalized using Affy Bioconductor package in R (Gentleman et al., 2004) . Normalization was done to adjust individual probe hybridizations so as to make biologically meaningful interpretations of the signal intensities. Multest package was then used to perform paired t-test on normalized intensities of stable and unstable plaques; this was followed by False Discovery Rate (FDR) correction. Fold change of genes was calculated from ratio of mean of intensities from stable plaque to mean of intensities from unstable plaques and converted to log 2.

2.2 Protein Protein Interaction (PPI) Network

Literature was searched to select studies which identified proteins present in stable and unstable plaque. Proteins identified from following studies were used as a seed for protein protein interaction network (PPI) :

1. Proteins from 35 atherosclerotic plaques were extracted and subjected to mass spectrometry . The plaques were classified into three histological categories of early , intermediate and advanced. Total of 806 unique proteins were identified to be present in atherosclerotic plaque (Bagnato et al, 2007).

2. Proteins were extracted from 5 stable and 12 unstable plaques of different patients and hybridized to glass protein mictoarrays having 512 antibodies to identify proteins which are differentially expressed . Protein expression showed 21 proteins were higly expressed in unstable plaque and 3 proteins were highly expressed in stable plaque (Slevin et al, 2006)

3. Plaques from 48 patients were obtained , histologically they were classified as 19 stable and 29 unstable plaques. Proteins extracted from plaques were subjectd to mass spectrometry. Proteins identified were further analyzed using western blot to compare expression between stable and unstable plaques. 9 proteins were found to be differentialy expressed (Lepedda et al, 2009).

4. Additionally 106 proteins associated with unstability of a plaque was identified from literature.

Total of 984 proteins , after removal of duplicates, was used as a seed for PPI. Human Protein Reference Database (HPRD) was used as a source for PPI intially . PPI in HPRD based solely on experimental evidence, 41% from in vivo experiments, 33% from in vitro experiments 24% from both in-vitro and in-vivo, 2% from yeast two hybrid experiments (Peri et al., 2003). PPI from in-vitro experiments could be false. For example yeast two hybrid assay may indicate proteins A and B interact together and proteins B and C interact together this does not imply that proteins A and C interact . Moreover rate of false positive PPI from yest two hybrid assay has been estimated to be 70% (Deane, 2002). To consider A and C as true interactions more evidence supporting the interactions has to be incorporated which can be done using STRING (Szklarczyk et al, 2010). STRING database include predicted protein interactions along with known PPI. It integrates information regarding a PPI from multiple sources such as geomic context, highthrouhput experiments, coexpression and literature .A confidence score is assigned to each PPI based on supporting evidences. This eliminates false interactions in the network resulting in high confidence network. So STIRNG was used as a source for PPI instead of HPRD .

STRING map's proteins back to gene from which they were encoded . 984 proteins used as seed mapped to 523 genes. In the network the nodes represent gene . Edge joins interacting genes or genes whose proteins form a complex. The network is further extended to include proteins interacting with seed proteins by adding two neighbors to seed proteins.

PPI network was visulized using Cytoscape (Shannon et al, 2003). Cytoscape is a open source software, serves as a tool for network visualization, manipulation and integration of data . PPI was overlaid by expression of genes whose differential expression between stable and unstable plaques was significant. However two expression datasets were not combined as they are from different array's and patients from whom plaques were obtained for gene expression analysis vary .

2.3 Transcription factor binding site prediction

Promoter region of significant Differentially Expressed (DE) genes from both gene expression datasets were examined for transcription factor binding sites. Transcription factor binding site were predicted using TRANSFAC, a database of eukaryotic transcription factors, target genes and binding sites (Matys, 2003). TRANSFAC uses MATCH algorithm, which search the promoter region of a gene for transcription factor binding site based on position weight matrix (Kel, 2003). Promoter window of 1000 base pairs was examined to identify overrepresented transcription factor binding sites. A p-value cut-off of 0.05 and human housekeeping genes as

a background were used for transcription factor binding site prediction. Predicted transcription factors were then mapped to the PPI network.

2.4 Topology of the unstable plaque network

Integrating significant differential gene expression , PPI and transcription factors gave unstable plaque network . The nodes in the network were classified based on degree of connectivity. Highest degree of connectivity in the network was around 80; very few genes had such high connectivity [Figure 5]. To capture more unstable plaque associated hub genes any gene with greater than 50 neighbors was chosen as superhubs. Degree connectivity less than 20 was satisfied by 50 % of the network nodes, to keep criteria for hub gene stringent node with connectivity degree greater than 20 and less than 50 were chosen as hub.

STRING assigns a combined confidence score to edges based on evidence from literature, genomic context, high throughput experiments and lab experiments. Edges with a score of less than 0.7 were removed to obtain a network with moderate confidence. This was followed by two filtering steps:

1. Nodes which are not differentially expressed or those which are not interacting with any of the differentially expressed genes were removed. Interactions with such nodes are present in the network because of their presence in STRING database and may not be associated with unstable plaques.

2. Sample size of first and second data set for gene expression was 3 and 4 respectively. Due to small sample sizes statically power would be low if we consider superhubs and hubs having differentially expressed neighbors from single dataset. Hence such superhubs and hubs were eliminated.

Gene expression and proteins used as a seed for PPI are associated with unstable plaque, but PPI data from STRING are not plaque associated. Therefore the network was filtered to remove any node which is not differentially expressed or not connected to any of differentiallly expressed gene , interaction of such node exist in the network because they were present in STRING and not associated to plaque progression.

2.5 Comparison of unstable plaque network to random network

Radom networks of unstable plaque network were generated by shuffling of edges between the nodes using random network plug-in of Cytoscape. This method of generating random networks is called a stub-rewiring approach where in maintaining degree of connectivity of nodes the edges are shuffled; in this process edges of nodes may vary and may not have degree as in original network which would indicate the nodes are noise and not true signals (Banks et al, 2008).

2.6 Gene ontology

Superhub and hub genes of unstable plaque network were accessed for over representing gene ontology categories. Overrepresented biological processes involved in a plaque progression were identified by hypergeometric test between annotations of superhubs and hub genes in network versus whole human genome annotation followed by FDR correction.

2.7 Mapping to pathways

Genes in unstable plaque PPI network were mapped on to cellular signaling pathways to identify pathways involved in conversion of a stable plaque to unstable plaque. DAVID, was used for mapping genes to pathways. DAVID is a bioinformatics tool released in 2003; it is aimed at extracting biological annotations from a list of genes. Functional analysis in DAVID is influenced by quality of gene list used. Our gene list was ‘good’ gene list satisfying following characteristics as suggested by DAVID (Huang et al, 2009) :

1. Gene lists contains important genes involved in processes such as chemokine production, angiogenesis, response to wounding, inflammation, foam cell differentiation and adhesion. Hence is disease specific.

2. Genes pass significant statistical thresholds (fold change > 2, p-value < 0.05).

3. Up-regulated and down-regulated genes involved in disease associated biological processes rather than being spread across all other biological processes.

CHAPTER 3 – RESULTS

3.1 Literature search for gene expression data

Two studies satisfied our criteria of gene expression, first criteria being gene expression must be from stable vs unstable plaque, second criteria being gene expression study should include mRNA from all cells present in a plaque .

First study chosen for gene expression data was from Papaspyridonos et.al (Papaspyridonos et al, 2006). The results from their intraplaque analysis indicated 168 genes to be differentially expressed, of these 93 were down regulated and 75 up regulated. Fold change of 168 (2 duplicates removed) differentially expressed genes was greater than 2 fold. As raw data was unavailable, scale of fold change and FDR correction was not known. However differential expression of genes was further confirmed by Sybr Green assays.

Second study chosen for gene expression data was from Ijas et.al (Ijäs et al, 2007). Of 12,625 genes 28 genes had fold change greater than one on a scale of log 2. To capture more significant genes, FDR correction of 0.05 was used as criteria for choosing significant genes, which resulted in 175 significant genes, of these 99 were down regulated and 76 up regulated.

Of all significant genes from two datasets, overlap of 3 genes was found between them. MAP1B, MAGED2, PKD2 are the significant differentially expressed genes found in both

datasets [Figure 6]. Though only 3 genes overlap, gene ontology reveals 24 common biological processes between two datasets [Figure 6]. Increase in overlap of biological processes is expected as the two datasets represent same pathological condition of plaque instability.

3.2 Protein Protein Interaction network

Mass spectrometry and protein microarray analysis of protein extract from atherosclerotic plaques have revealed 824 unique proteins involved in atherosclerosis (Bagnato et al, 2007;Lepedda et al, 2009;Slevin et.al,2006). In addition to these proteins , 160 proteins associated with atherosclerotic plaques were identified from literature. A initial unstable plaque PPI network was assembled using these 984 proteins as input to STRING. The 984 proteins were mapped to 523 genes and 5384 interactions. The network was extended by adding one neighbor to include proteins interacting with seed proteins. This resulted in PPI network with 526 nodes and 5503 edges. Adding two neighbors resulted in network of 539 nodes and 5810 interactions.

Edges have a confidence scores assigned by STRING. Scores ranges between and 0.4 to 1 . The network was filtered to remove edges below the score of 0.7 to obtain a network with moderate confidence. After filtering step the final network has 483 nodes and 3373 edges.

3.3 Transcription factors

Transcription factor binding sites were predicted 1000 base pairs upstream of significant differentially expressed genes from both datasets. A total of 272 transcription factors were predicted using TRANSFAC. Of the 272 transcription factors, 11 transcription factors mapped to PPI network. After filtration steps 9 transcription factors remained in final unstable plaque network. Of them two transcription factors STAT1 and HIF1A were hubs. Other transcription factors were PPARA, REST, TAL1, HMGA1, PDX1, NR1H2 and NR1H3. EP300 and TP53 are transcription factors which were eliminated in final unstable plaque network.

3.4 Unstable plaque network

Integration of significant gene expression from both datasets, PPI network and transcription factors yielded an unstable plaque network. PPI network of proteins found in plaques served as a backbone structure, significant differential gene expression from both the datasets was overlaid on this PPI. Topological examination revealed 16 superhubs and 111 hubs. Network filtration to remove nodes which are not differentially expressed or those which are not interacting with any of the differentially expressed genes and elimination of superhubs and hubs which had differential expression genes as neighbor from one dataset only, gave final unstable plaque network. The final unstable plaque network had the following metrics:

- Superhubs : 12
    - Down regulated : 1

- Hubs : 20
    - Upregulated : 2
    - Downregulated : 2
    - Transcription factors : 2
- Downregulated genes : 9
- Upregulated genes : 8
- Transcription factors : 7

3.5 Unstable plaque network versus random network

As a check for false positives , unstable plaque network was compared to random networks by a stub-rewiring approach. A node whose degree of connectivity modified from its original connectivity would indicate noise. In our unstable plaque network all of the nodes had the degree preserved in all 1000 random network generated. Each random network was generated with 1000 shuffling of edges for each node [Table 5]. Since the degree of connectivity of nodes was preserved we can say there was no noise in network in terms of topology .This is reflected by the 0 standard deviation in topological measure of node degree across original network and random networks. However there is deviation of 0.005 in cluster coefficient indicating variation in neighbors of nodes in random networks.

3.6 Annotation of superhub and hub genes

Final 12 superhubs and 20 hubs were further validated for their roles in progression of a plaque based on their annotations. Functional annotations indicate that hub and superhub genes are involved in regulation of angiogenesis, chemokine production and foam cell differentiation [Table 6]. All the processes which would be expected to take place during progression of a plaque.

3.7Mapping to pathways

To obtain comprehensive view of biological processes involved in conversion of a stable plaque to unstable plaque, hub and superhub genes were mapped to cellular signaling pathways using DAVID. Pathways mapped were:

1. Focal adhesion
2. Cytokine-cytokine receptor interaction
3. Sustained angiogenesis
4. Apoptosis
5. VEGF signaling pathway

Cellular signaling pathways are not isolated; they are interconnected forming a network of interaction. An external stimulus from oxidized lipids and hypoxia can trigger the activation of the interconnected pathways.

CHAPTER 4 – DISCUSSION

4.1 Gene expression profiles and protein concentrations

Despite limitations in our knowledge of dynamic biological functions affecting state of a plaque, integrating genes with PPI reveals underlying interconnectivity of the perturbed molecular network during a plaque progression.

Gene expression profiles of atherosclerotic plaques have revealed significant DE genes in an atherosclerotic plaque (Martinet, 2002; McCaffrey et al, 2000; Peri et al, 2003). However gene expression and protein concentrations are not correlated because of variation in gene copy number; post translational modifications and variation in splicing events. This is further reflected in poor mapping of gene expression profiles on to PPI network. Of 178 significant DE genes from Ijas et.al dataset only 7 genes mapped to PPI network. Of 168 significant DE genes from Papaspyridonos et.al 22 genes mapped to PPI network. Hence instead of correlating gene expression and protein concentrations we scaffold gene expression profiles with PPI network to identify genes involved in a plaque progression. Network is further extended by adding genes interacting with seed genes. This approach has revealed EGF and HSPD1 novel genes which are not part of our seed list.

EGF was novel superhub not included in our seed list. As a response to inflammation monocytes begin to infiltrate into plaques and transforms into macrophage. EGF receptors

are found to be present on surface of the monocytes and macrophages which can be used as indicator of plaque progression. Presence of EGF in atherosclerotic plaques has been confirmed by macrophage co-localization staining (Lamb et al, 2004).

Pub med search indicates 110 citations indicating role of HSPD1, the heat shock proteins in atherosclerosis. Recent review on heat shock proteins indicated their role in immunogenic and protective response to stress induced in atherosclerosis. Hence it could be a potential drug target (LU & KAKKAR, 2010).

4.2 Topology

Topology of the network indicates not all hub and superhub genes are significantly differentially expressed. Therefore this confirms our hypothesis; solely relying only on DE genes data would mean loss of information.

Unstable plaque network has hub and superhub genes having DE genes as neighbors from both dataset. Superhub and hub gene with DE genes from single gene expression dataset were eliminated since sample size of gene expression datasets were small, affecting statistical power.

HMOX1 and SOD2 are highly expressed in plaque to overcome oxidative stress. While CCL5, CASP3 and EGFR are superhub and hub genes downregulated in unstable plaque PPI network.

4.3 Odds of data

Examination of overlap between significant DE genes showed only 3 genes overlap between two gene expression datasets. Such low overlap could be because of :

1. Plaques used for gene expression analysis were obtained from different sets of patients.

2. Different methods were used for plaque classification as stable or unstable.

3. Two gene expression datasets could be reflecting different mechanisms involved in a plaque progression.

4. In addition to proteins identified to be present different stages of plaque progression from mass spectrometry and protein microarrays, 160 proteins included in our seed list was from literature. These 160 were based on clinical perspective, hence may not be comprehensive.

4.4 Gene ontology of superhub and hub genes

Biological activity of superhub and hub genes in affecting state of a plaque was confirmed by their functional roles. Angiogenesis, secretion of chemokines and foam cell differentiation, specific functions which promote plaque vulnerability were considered as most important. These are in line with histomorphological features identified by Shah et.al. (Shah, 2003) According to Shah et.al increase neovascularity (angiogenesis), inflammatory cell infiltration (mediated by cytokines) and high lipid content dispositions a stable plaque to unstable plaque by increasing vulnerability and leading to rupture. The study by Sluimer et.al and Mause et.al (Sluimer et.al, 2010;Mause & Weber, 2009) has shown that angiogenesis alone can increase vulnerability of a plaque by providing way for entry of leukocytes and erythrocytes. We can say this study confirmed pathophysiological mechanism of unstable plaque as predicted by Shah et.al .

4.5 Mapping of superhub and hub genes to Pathways

Superhub and hub genes mapped to the following pathways:

i. **Focal adhesion**

Focal adhesion refers to recruitment of proteins serving as adhesion molecules. Lipid deposition and oxidation on the endothelium causes inflammatory response triggering adhesion cascade. Adhesion molecules like ICAM1 and VCAM present on the surface of endothelial cells interact with monocytes and T-lymphocytes enabling them to enter the

plaque through junctions of endothelium cells (Watanabe & Fan, 1998). Increased amount of soluble ICAM1 were detected in plasma of subjects suffering from atherosclerosis indicating ICAM1 could serve as a biomarker for atherosclerosis (S.-J. Hwang et al, 1997). IL8, BCL2, IL1B, COL1A1, TGFB1, THBS1, TGFB1, IFNG, SERPINE1 are other focal adhesion molecules identified in this study which are associated with plaque progression.

ii. Cytokines

Activated macrophages and lymphocytes within a plaque secrete cytokines. Role of cytokines in an atherosclerotic plaque has been studied since 1980. Cytokines play a regulatory role. They regulate apoptosis, angiogenesis, permeability of endothelial cells, activation of adhesion molecules and MMP expression (Tedgui et.al, 2006). IL6, HIF1A, HMOX1, IL1B cytokines associated with plaque instability appeared as hub and superhub genes in unstable plaque PPI network .

iii. Angiogenesis

Angiogenesis means formation of new vessels. Pathway assessment of the unstable plaque PPI network and gene ontology indicate angiogenesis to be a critical process associated with progression of a plaque. IL6, HIF1A, HMOX1, VEGFA, IL1B, FGF2 are hub and superhub genes associated with angiogenesis.

iv. Apoptosis

As atherosclerotic plaque progresses, there is increase in lipoproteins in the necrotic core which undergo oxidation forming oxysterols which have apoptotic effects (Reviews & Biology, 1999). In addition to oxysterols, cytokines, reactive oxygen species and growth factors leads to deregulation of apoptosis and activation of fas/fas pathway in plaque affecting plaque stability (Geng, 2001) . CASP3, CASP1, FAS are apoptosis related superhub and hub genes present in unstable plaque PPI network.

v. VEGF signaling pathway

Smooth muscle cells derived fibroblast connective tissue forms cap of a plaque. With progression of a plaque the thickness of the cap increases decreasing oxygen within core of the plaque resulting in condition called hypoxia. Hypoxia induces transcription factor HIF1A binds to promoter region of VEGFA causing increase in expression of VEGFA (Olsson et.al, 2006). Increase in VEGFA triggers angiogenesis increasing plaque instability.

All the pathways are interconnected forming a cascade. Superhub and hub genes have prominent biological roles in deciding a state of a plaque. Though their presence in unstable plaque has been confirmed and their roles in atherosclerosis been studied, no study has validated the genes to be a marker for deciding a state of a plaque. Undoubtedly through further validation these genes could be used as potential markers whose presence would indicate conversion of a stable plaque to an unstable plaque.

CHAPTER 5- CONCLUSION

5.1 Limitations of the study

Data limitation

Although search in published and publicly available data indicated several mRNA expression studies which have been done identifying significant differentially expressed genes in a atherosclerotic plaque , we limited our search for publicly available mRNA expression data. The search was limited to differentially expressed genes between stable and unstable human atherosclerotic plaques. This was done to identify key genes and pathways which indicate plaque progression from stable to unstable state. Of the few studies looking at differential expression of mRNA in stable versus unstable plaques Papaspyridonos et.al was one of the most appropriate for our study but due to improper deposition of data and lack of clear explanation of deposited data, we used additional data set from Ijas et.al to probe into our question.

Proteins used as seed for PPI are limited to the ones which are physiologically validated in a plaque based on mass spectrometry and glass bead microarray techniques. It may not be a comprehensive list of proteins involved in plaque progression.

Analysis limitations

Sources for PPI present in STRING database are from experiments like co-immunoprecipitation, yeast two hybrid, high-throughput experiments and predicted interactions. From the experiments and predictions interacting proteins are identified but not their directionality. Hence in unstable plaque PPI network directionality of the network is not known. This also limits the PPI to the ones present in the STRING database, which may not be comprehensive.

Sample size in both the studies was small affecting statistical power. Plaques from 3 patients were included in Papspyridonos et.al study and plaques from 4 patients were included in Ijas et.al study. This limited us from co-expression analysis of the significant differentially expressed genes.

5.2 Future directions

An atherosclerotic plaque progresses through several stages from its initial state of lipid deposition to formation of an unstable plaque. Analysis of samples from a single stage would provide a snapshot of association of genes to plaque progression. This would be insufficient for obtaining comprehensive view of biological processes contributing to a state of a plaque. To overcome this limitation samples from series of patients and series of plaques at every stage would be required.

Several common biological processes appear to occur in complex diseases like cancer and atherosclerosis. For example angiogenesis, process where new vessels are formed is seen in cancer and atherosclerosis. To satisfy nutritional requirement of new cells formed in cancer, new vessels are formed. To overcome hypoxia or reduced amount of oxygen in plaque due to fibrous cap formation, plaques develop new vessels. Hence to differentiate a plaque specific processes patient selection has to be constrained. Additionally careful phenotyping of plaque is required. Ideal samples would be plaques from patients suffering from atherosclerosis only and no other disease conditions.

Superhub and hub genes can be further validated by laboratory experiments using mouse models or immunohistochemical techniques to confirm their role in plaque progression.

The integrative approach can be extended to include metabolomic data and high throughput sequence data to gain more insights into molecular functions involved in plaque progression.

PPI databases are ever growing with a continuous cycle of computational predictions and experimental validations. This necessitates reproducing this study in future including new validated and predicted PPI which may reveal new biomarkers.

REFERENCES

Bader, G. D., Donaldson, I., Wolting, C., Ouellette, B. F., Pawson, T., & Hogue, C. W. (2001). BIND--The Biomolecular Interaction Network Database. *Nucleic acids research*, *29*(1), 242-5. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=29820&tool=pmcentrez&rendertype=abstract.

Bagnato, C., Thumar, J., Mayya, V., Hwang, S.-I., Zebroski, H., Claffey, K. P., et al. (2007). Proteomics analysis of human coronary atherosclerotic plaque: a feasibility study of direct tissue proteomics by liquid chromatography and tandem mass spectrometry. *Molecular & cellular proteomics : MCP*, *6*(6), 1088-102. doi: 10.1074/mcp.M600259-MCP200.

Banks, E., Nabieva, E., Chazelle, B., & Singh, M. (2008). Organization of physical interactomes as uncovered by network schemas. *PLoS computational biology*, *4*(10), e1000203. doi: 10.1371/journal.pcbi.1000203.

Barabasi, a, Albert, R., & Jeong, H. (2000). Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications*, *281*(1-4), 69-77. doi: 10.1016/S0378-4371(00)00018-2.

Camargo, A., & Azuaje, F. (2007). Linking gene expression and functional network data in human heart failure. *PloS one*, *2*(12), e1347. doi: 10.1371/journal.pone.0001347.

Deane, C. M. (2002). Protein Interactions: Two Methods for Assessment of the Reliability of High Throughput Observations. *Molecular & Cellular Proteomics*, *1*(5), 349-356. doi: 10.1074/mcp.M100037-MCP200.

Gargalovic, P. S., Imura, M., Zhang, B., Gharavi, N. M., Clark, M. J., Pagnon, J., et al. (2006). Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(34), 12741-6. doi: 10.1073/pnas.0605457103.

Gasparini, B. G., Weidner, N., Bevilacqua, P., Maluta, S., Palma, P. D., Caffo, O., et al. (2011). Size , and Peritumoral Lymphatic Vessel Invasion Are Relevant Prognostic Markers in Node-Negative Breast Carcinoma. *Society*, *12*(3), 454-466.

Geng, Y. J. (2001). Biologic effect and molecular regulation of vascular apoptosis in atherosclerosis. *Current atherosclerosis reports*, *3*(3), 234-42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11286645.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, *5*(10), R80. doi: 10.1186/gb-2004-5-10-r80.

Harsha, H. C., Kandasamy, K., Ranganathan, P., Rani, S., Ramabadran, S., Gollapudi, S., et al. (2009). A compendium of potential biomarkers of pancreatic cancer. *PLoS medicine*, *6*(4), e1000046. doi: 10.1371/journal.pmed.1000046.

He, X., & Zhang, Jianzhi. (2006). Why do hubs tend to be essential in protein networks? *PLoS genetics*, *2*(6), e88. doi: 10.1371/journal.pgen.0020088.

Heidenreich, P. a, Trogdon, J. G., Khavjou, O. a, Butler, J., Dracup, K., Ezekowitz, M. D., et al. (2011). Forecasting the Future of Cardiovascular Disease in the United States: A Policy Statement From the American Heart Association. *Circulation*. doi: 10.1161/CIR.0b013e31820a55f5.

Hood, Leroy, & Perlmutter, R. M. (2004). The impact of systems approaches on biological problems in drug discovery. *Nature biotechnology*, *22*(10), 1215-7. doi: 10.1038/nbt1004-1215.

Huang, D. W., Sherman, B. T., & Lempicki, R. a. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, *4*(1), 44-57. doi: 10.1038/nprot.2008.211.

Hwang, S.-J., Ballantyne, C. M., Sharrett, A. R., Smith, L. C., Davis, C. E., Gotto, A. M., et al. (1997). Circulating Adhesion Molecules VCAM-1, ICAM-1, and E-selectin in Carotid Atherosclerosis and Incident Coronary Heart Disease Cases : The Atherosclerosis Risk In Communities (ARIC) Study. *Circulation*, *96*(12), 4219-4225. Retrieved April 19, 2011, from http://circ.ahajournals.org/cgi/content/abstract/96/12/4219.

Ideker, T, Galitski, T., & Hood, L. (2001). A new approach to decoding life: systems biology. *Annual review of genomics and human genetics*, *2*, 343-72. doi: 10.1146/annurev.genom.2.1.343.

Ijäs, P., Nuotio, K., Saksi, J., Soinne, L., Saimanen, E., Karjalainen-Lindsberg, M.-L., et al. (2007). Microarray analysis reveals overexpression of CD163 and HO-1 in symptomatic carotid plaques. *Arteriosclerosis, thrombosis, and vascular biology*, *27*(1), 154-60. doi: 10.1161/01.ATV.0000251991.64617.e7.

Jezierska, A., & Motyl, T. (2009). SO LY L SO LY L, *15*(2).

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27-30. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=102409&tool=pmcentrez&rendertype=abstract.

Kel, a E. (2003). MATCHTM: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Research*, *31*(13), 3576-3579. doi: 10.1093/nar/gkg585.

Kitano, H. (2002). Systems biology: a brief overview. *Science (New York, N.Y.)*, *295*(5560), 1662-4. American Association for the Advancement of Science. doi: 10.1126/science.1069492.

Kitano, H. (2004). Biological robustness. *Nature reviews. Genetics*, *5*(11), 826-37. doi: 10.1038/nrg1471.

Lamb, D. J., Modjtahedi, H., Plant, N. J., & Ferns, G. a a. (2004). EGF mediates monocyte chemotaxis and macrophage proliferation and EGF receptor is expressed in atherosclerotic plaques. *Atherosclerosis*, *176*(1), 21-6. doi: 10.1016/j.atherosclerosis.2004.04.012.

Lee, T. I., & Young, R. A. (2000). T RANSCRIPTION OF E UKARYOTIC P ROTEIN -C ODING G ENES. *Young*, 77-137.

Lepedda, A. J., Cigliano, A., Cherchi, G. M., Spirito, R., Maggioni, M., Carta, F., et al. (2009). A proteomic approach to differentiate histologically classified stable and unstable plaques from human carotid arteries. *Atherosclerosis*, *203*(1), 112-8. doi: 10.1016/j.atherosclerosis.2008.07.001.

LU, X., & KAKKAR, V. (n.d.). The Role of Heat Shock Protein (HSP) in Atherosclerosis: Pathophysiology and Clinical Opportunities. *Current medicinal chemistry*, *17*(10), 957-973. Bentham. Retrieved February 16, 2011, from http://cat.inist.fr/?aModele=afficheN&cpsidt=22629707.

Manuscript, A. (2010). NIH Public Access. *Molecular Biology*, *407*(6801), 233-241. doi: 10.1038/35025203.Atherosclerosis.

Martinet, W. (2002). Gene Expression Profiling of Apoptosis-Related Genes in Human Atherosclerosis: Upregulation of Death-Associated Protein Kinase. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *22*(12), 2023-2029. doi: 10.1161/01.ATV.0000041843.44312.12.

Matys, V. (2003). TRANSFAC(R): transcriptional regulation, from patterns to profiles. *Nucleic Acids Research*, *31*(1), 374-378. doi: 10.1093/nar/gkg108.

Mause, S. F., & Weber, C. (2009). Intrusion through the fragile back door: immature plaque microvessels as entry portals for leukocytes and erythrocytes in atherosclerosis. *Journal of the American College of Cardiology*, *53*(17), 1528-31. American College of Cardiology Foundation. doi: 10.1016/j.jacc.2009.01.047.

McCaffrey, T. a, Fu, C., Du, B., Eksinar, S., Kent, K. C., Bush, H., et al. (2000). High-level expression of Egr-1 and Egr-1-inducible genes in mouse and human atherosclerosis. *The Journal of clinical investigation*, *105*(5), 653-62. doi: 10.1172/JCI8592.

Olsson, A.-K., Dimberg, A., Kreuger, J., & Claesson-Welsh, L. (2006). VEGF receptor signalling - in control of vascular function. *Nature reviews. Molecular cell biology*, *7*(5), 359-71. doi: 10.1038/nrm1911.

Pacheco, E., Kutys, B., Duimel, H., & M, P. (2010). leakage, *53*(17), 1517-1527. doi: 10.1016/j.jacc.2008.12.056.Thin-walled.

Papaspyridonos, M., Smith, A., Burnand, K. G., Taylor, P., Padayachee, S., Suckling, K. E., et al. (2006). Novel candidate genes in unstable areas of human atherosclerotic plaques. *Arteriosclerosis, thrombosis, and vascular biology*, *26*(8), 1837-44. doi: 10.1161/01.ATV.0000229695.68416.76.

Peri, S., Navarro, J. D., Amanchy, R., Kristiansen, T. Z., Jonnalagadda, C. K., Surendranath, V., et al. (2003). Development of Human Protein Reference Database as an Initial Platform for Approaching Systems Biology in Humans. *Genome Research*, 2363-2371. doi: 10.1101/gr.1680803.17.

Review, I. (2009). Novel concepts in atherogenesis: angiogenesis and hypoxia in atherosclerosis. *Journal of Pathology, The*, (January), 7-29. doi: 10.1002/path.

Reviews, C., & Biology, M. (1999). No Title. *Critical Reviews in Biochemistry*, *34*(1), 25-32.

Shah, P. (2003). Mechanisms of plaque vulnerability and rupture*1. *Journal of the American College of Cardiology*, *41*(4), S15-S22. doi: 10.1016/S0735-1097(02)02834-6.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, *13*(11), 2498-504. doi: 10.1101/gr.1239303.

Slevin, M., Elasbali, a B., Miguel Turu, M., Krupinski, J., Badimon, L., & Gaffney, J. (2006). Identification of Differential Protein Expression Associated with Development of Unstable Human Carotid Plaques. *American Journal Of Pathology*, *168*(3), 1004-1021. doi: 10.2353/ajpath.2006.050471.

Sluimer, J. C., Kisters, N., Cleutjens, K. B., Volger, O. L., Horrevoets, A. J., Akker, L. H. van den, et al. (2007). Dead or alive: gene expression profiles of advanced atherosclerotic plaques from autopsy and surgery. *Physiological genomics*, *30*(3), 335-41. doi: 10.1152/physiolgenomics.00076.2007.

Stary, H. C., Chandler, A. B., Dinsmore, R. E., Fuster, V., Glagov, S., Insull, W., et al. (1995). A Definition of Advanced Types of Atherosclerotic Lesions and a Histological Classification of Atherosclerosis : A Report From the Committee on Vascular Lesions of

the Council on Arteriosclerosis, American Heart Association. *Circulation*, *92*(5), 1355-1374. Retrieved April 17, 2011, from http://circ.ahajournals.org/cgi/content/abstract/92/5/1355.

Szklarczyk, D., Franceschini, a, Kuhn, M., Simonovic, M., Roth, a, Minguez, P., et al. (2010). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research*, *39*(November 2010), 561-568. doi: 10.1093/nar/gkq973.

Tedgui, A., Mallat, Z., National, I., & Sante, D. (2006). Cytokines in Atherosclerosis : Pathogenic and Regulatory Pathways. *Physiological Reviews*, 515-581. doi: 10.1152/physrev.00024.2005.

Watanabe, T., & Fan, J. (1998). Atherosclerosis and inflammation mononuclear cell recruitment and adhesion molecules with reference to the implication of ICAM-1/LFA-1 pathway in atherogenesis. *International journal of cardiology*, *66 Suppl 1*, S45-53; discussion S55. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9951802.

TABLES

Superhubs	**Nodes satisfying the criteria**	**Superhubs from seed Genes**	**Superhubs from added neighbors**
Superhubs 80 having neighbors	1	1	0
Hubs having 70 neighbors	5	5	0
Hubs having 60 neighbors	7	7	0
Hubs having 50 neighbors	16	15	1

Table 1 : Count of superhubs satisfying criteria of having n (80,70,60,50) neighbors.

Hubs	**Nodes satisfying the criteria**	**Hubs from seed proteins**	**Hubs from added neighbors**	**Percentage of hubs from added neighbors not in seed list**
Hubs having 5 neighbors	333	318	15	4%
Hubs having 10 neighbors	249	234	15	6%
Hubs having 15 neighbors	174	166	8	4%
Hubs having 20 neighbors	127	122	4	3%
Hubs having 25 neighbors	79	76	3	2%

Hubs having 30 neighbors	56	55	1	1%
Hubs having 35 neighbors	42	41	1	2%

Table 2 : Count of hubs satisfying criteria of having n (5,10,15,20,25,30,35) neighbors and percentage of hubs which are not part of seed list.

Topological Measure	**APPI Network**	**Random APPI Networks Average**	**Random Networks Standard Deviation**
Clustering Coefficient	0.421452	0.103042	0.005267
Average Degree	13.86037	13.86037	0
Degree Distribution	-1.09379	-1.09379	0
Mean Shortest Path	3.150165	2.71451	0.010678

Table 3 : Average and standard deviation of topological measures of PPI network and 1000 random networks of PPI network .

Hub gene	Downregulated neighbors from Papaspyridonos et.al	Upregulated neighbors from Papaspyridonos et.al	Downregulated neighbors from Ijas et.al	Upregulated neighbors from Ijas et.al
ICAM1	JAM3	HMOX1	CCL5	
FGF2	EGFR	F11R	CCL5	
IFNG	SOD3	SOD2	CCL5	
VIM	EGFR,TPM2,TPM1,TAGLN	ENO1	CASP3	
IL2	DSTN		CCL5	
CYCS	HSPB1	CTSB,SOD2	CASP3,UQCRQ	
VCAM1	JAM3,EGFR	SOD2	CCL5	
STAT1	EGFR		CASP3,CCL5	
TLR4	RGFR	HMOX1	CCL5	MSR1
PPARG	EGFR	HMOX2	CASP3	
SOD2	SOD3	HMOX1	CASP3	
HIF1A	EGFR	ENO1,HMOX1	CASP3,CCL5	
ACTA1	TPM2,TPM1,EGFR,TAGLN,HSPB1,MYLK		CASP3	
TNF	EGFR	HMOX1	CASP3,CCL5	
FAS	EGFR,HSPB1		CASP3	
HMOX1		SOD2	CASP3	
EIF3S10	EGFR		CASP3	
CASP1		CTSB	CASP3	
SOD1	SOD3,HSPB1	IDH1,SOD2	CASP4	
HSPD1	HSPB1	SOD2	CASP3	PPIA

Table 4 : Hubs with DE neighbors in both gene expression data sets.

Superhub	Total number of nodes connected to superhub	Down regulated nodes connected to superhub from Papaspyridonos et.al	Up regulated nodes connected to superhub from Papaspyridono et. al	Down regulated nodes connected to superhub from Ijas et.al	Up regulated nodes connected to superhub from Ijas et.al
IL6	73	EGFR	HMOX1	CCL5	
AKT1	61	EGFR,SOD3,HSPB1,PFMK,GAS6	HMOX1,SOD2	CASP3	MSR1
TP53	58	EGFR	SOD2	CASP3,CCL5	
VEGFA	58	CTSB,EGFR	HMOX1	CCL5	
TGFB1	58	HMOX1	FMOD	CCL5	PPIA
MMP9	58	EGFR	CTSS,CTSB	CASP3,CCL5	
IL1B	58	EGFR	SOD2,HMOX1		GRAP2
CD44	58	EGFR		CASP3	
ITGB1	56	JAM3,MYLK	F11R	CASP3	
IL8	56	JAM3,MYLK	F11R	CCL5	
EGF	54	EGFR	HMOX1	CASP3	
F2	53	EGFR		CASP3,CCL5	

Table 5 : Superhubs with DE neighbors in both gene expression data sets.

Biological Process	Genes	Corrected p-value
Regulation of Angiogenesis	IL6,HIF1A,HMOX1,VEGFA, IL1B,FGF2	3.98E-11
Regulation of Chemokine production	IL6,HIF1A,HMOX1,IL1B	1.94E-08
Foam cell differentiation	TGFB1	3.35E-03

Table 6 : Functional annotation of superhub and hub genes

FIGURES

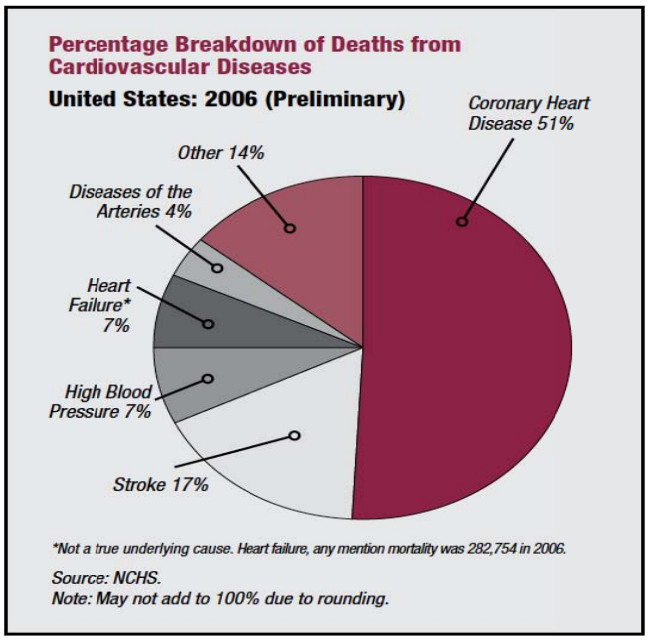

Figure 1 : Statistics for deaths due to cardiovascular disease types

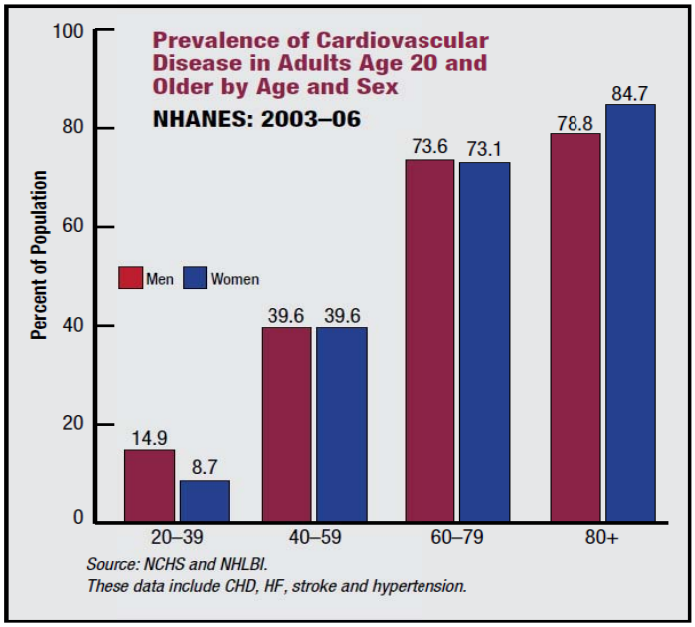

Figure 2 : Prevalence of cardiovascular diseases

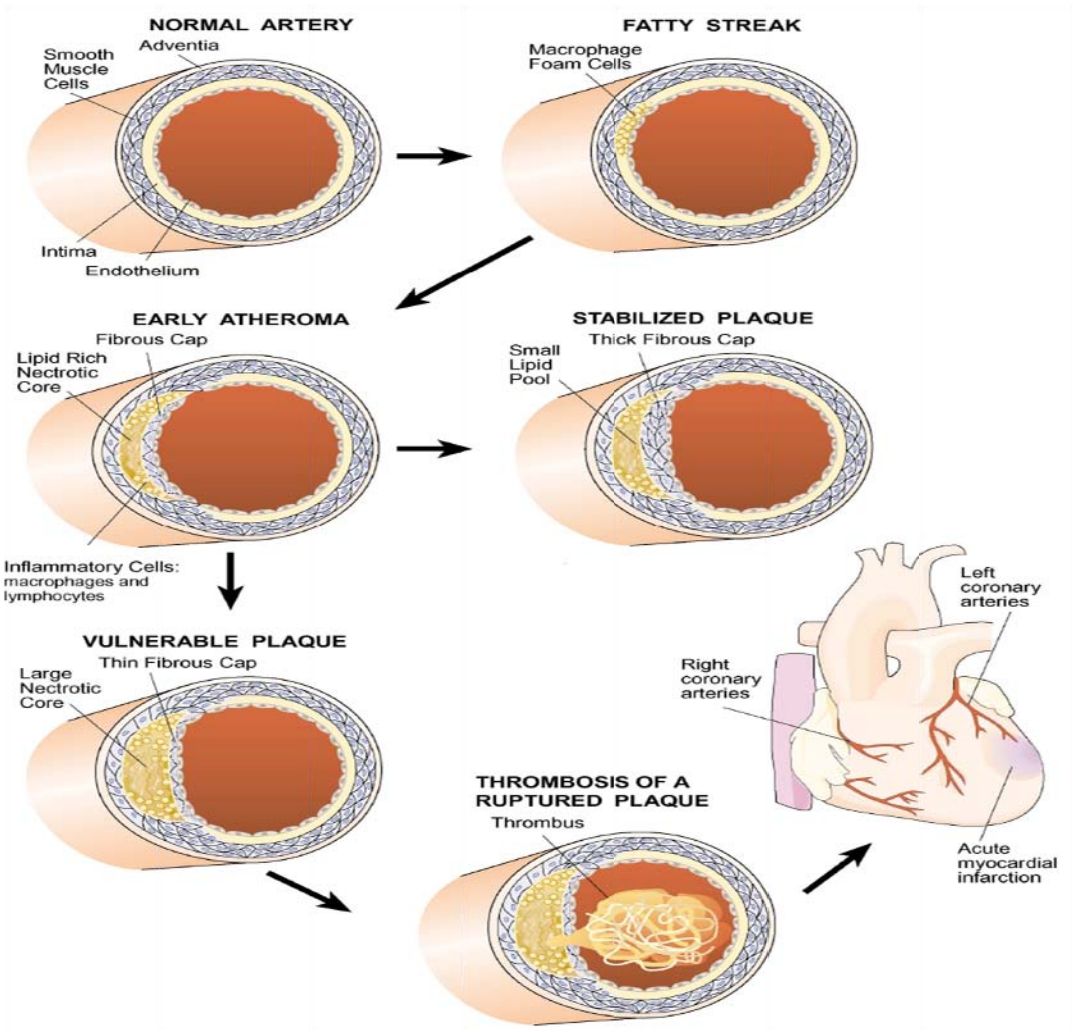

Figure 3: Atherosclerotic plaque formation and progression

Adopted from: Sluimer J & Daemen M . Novel concepts in atherogenesis: angiogenesis and hypoxia in atherosclerosis. *Journal of Pathology, The*. 2009;(January):7-29.

Figure 4 : Illustration of our hypothesis A) gene x not differentially expressed , affecting differential expression of other genes B) Differentially expressed gene y , affecting other processes.

Adopted from : Ramsey S A, Gold ES, Aderem A. A systems biology approach to understanding atherosclerosis. *EMBO molecular medicine*. 2010;2(3):79-89. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20201031.

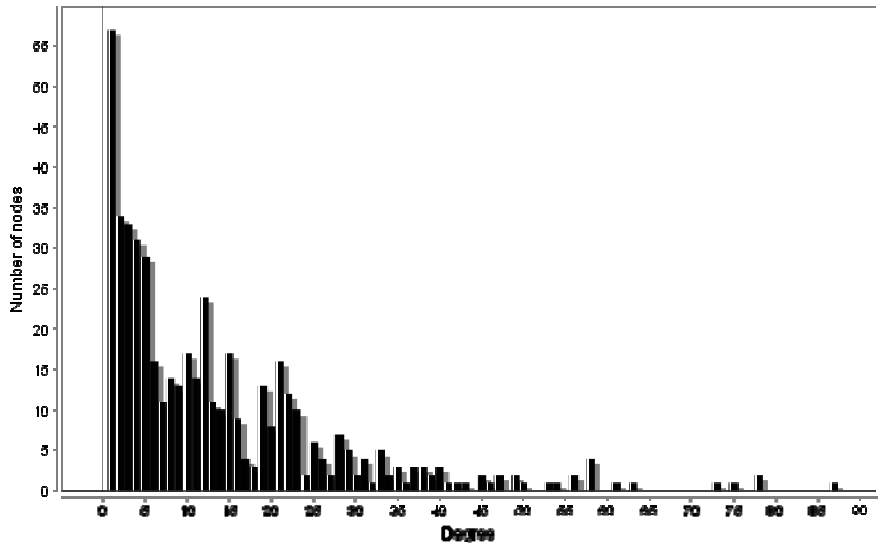

Figure 5 : Histogram representing distribution of degree (connected neighbors) of genes in the PPI network.

A

B

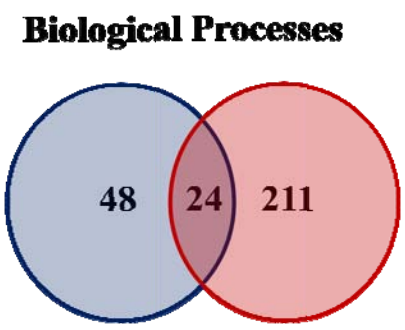

Figure 6: A) Count of DE genes from two gene expression datasets used in this study. Red indicates count of up-regulated genes and green indicates count of down regulated genes B) Overlap of biological processes between two gene expression data set

Figure 7: Count of proteins used to build PPI network. Blue indicates count of proteins, identified to be present in different stages of plaque progression using mass spectrometry and microarrays. Pink indicates count of proteins identified from literature.

Figure 8: Flow chart of method used in this study

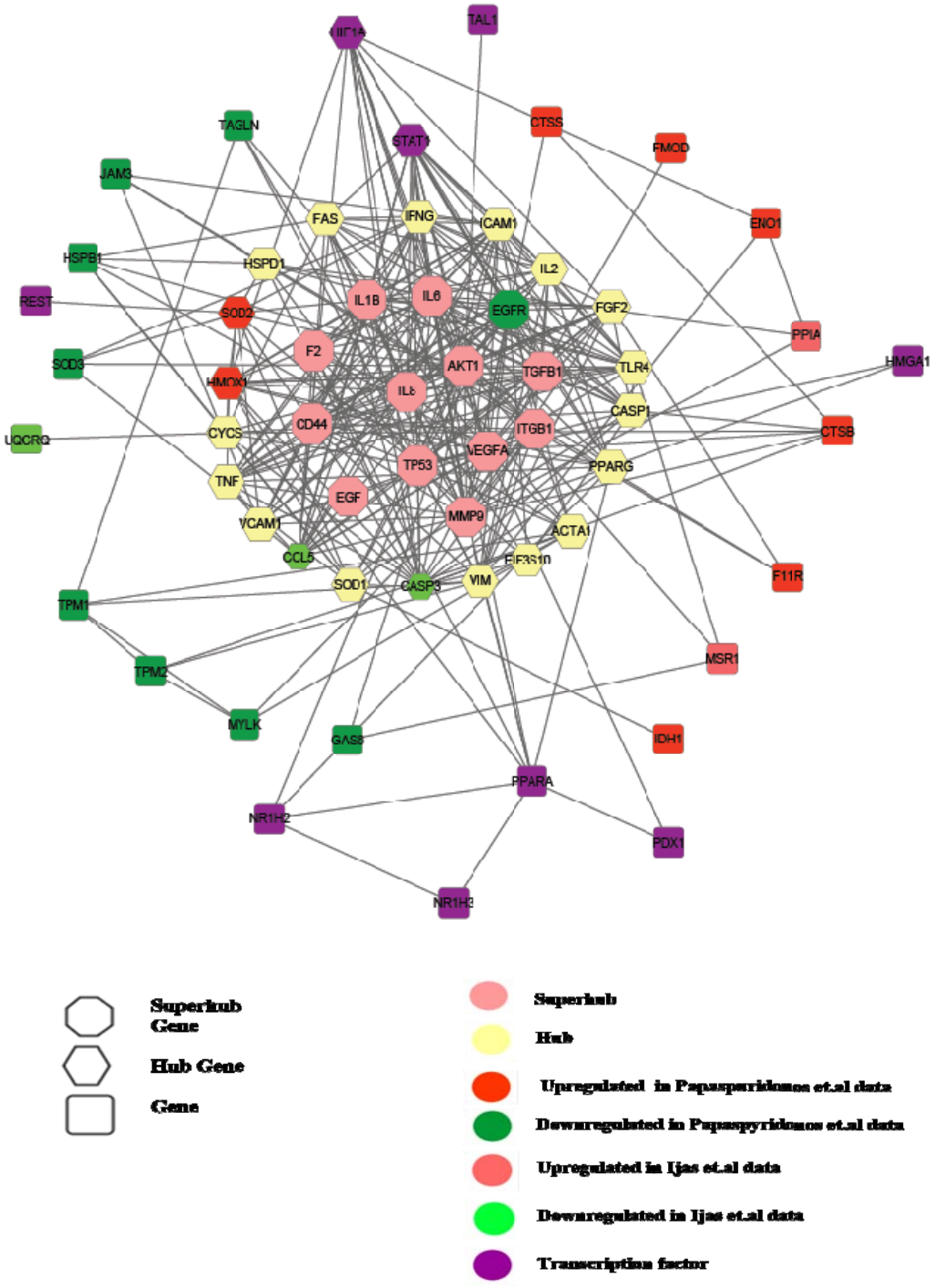

Figure 9: Unstable plaque PPI network