

Spring 2008

# Differentially expressed genes in aortic cells from atherosclerosis-resistant and atherosclerosis-susceptible pigeons

Janet Lynn Anderson  
*University of New Hampshire, Durham*

Follow this and additional works at: <https://scholars.unh.edu/dissertation>

---

## Recommended Citation

Anderson, Janet Lynn, "Differentially expressed genes in aortic cells from atherosclerosis-resistant and atherosclerosis-susceptible pigeons" (2008). *Doctoral Dissertations*. 416.  
<https://scholars.unh.edu/dissertation/416>

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact [nicole.hentz@unh.edu](mailto:nicole.hentz@unh.edu).

**DIFFERENTIALLY EXPRESSED GENES IN AORTIC CELLS FROM  
ATHEROSCLEROSIS-RESISTANT AND  
ATHEROSCLEROSIS-SUSCEPTIBLE PIGEONS**

**BY**

**JANET LYNN ANDERSON**

**B.A., University of New Hampshire, 1991  
M.S., University of New Hampshire, 2003**

**DISSERTATION**

**Submitted to the University of New Hampshire  
in Partial Fulfillment of  
the Requirements for the Degree of**

**Doctor of Philosophy  
in  
Animal and Nutritional Sciences  
May, 2008**

UMI Number: 3308366

### INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

**UMI**<sup>®</sup>

---


UMI Microform 3308366

Copyright 2008 by ProQuest LLC.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

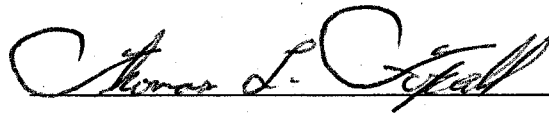
ProQuest LLC  
789 E. Eisenhower Parkway  
PO Box 1346  
Ann Arbor, MI 48106-1346

This dissertation has been examined and approved.



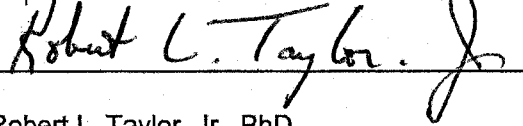
---

Dissertation Director, Samuel C. Smith, PhD.  
Professor, Department of Animal and Nutritional Sciences  
Professor, Department of Biochemistry and Molecular Biology



---

Thomas L. Foxall, PhD.  
Chair, Professor, Department of Animal and Nutritional Sciences



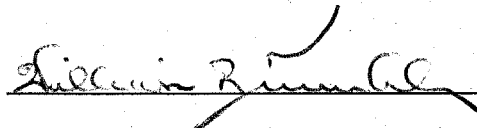
---

Robert L. Taylor, Jr., PhD.  
Professor, Department of Animal and Nutritional Sciences  
Professor, Genetics Program



---

William K. Thomas, PhD.  
Director, Hubbard Center for Genome Studies  
Chair, Professor, Department of Biochemistry and Molecular Biology



---

William R. Trumble, PhD.  
Professor, Department of Biochemistry and Molecular Biology

## DEDICATION

This Dissertation is fondly dedicated to the life, pursuits, and optimism of  
Elizabeth Cooke Smith\*  
Research Scientist, Mentor, Friend, 1929-2005

### The Weaver

Colored threads are treadled,  
Like so many times before, she  
Weaves.  
This time is different.  
Colored threads warp to fluorescent dyes  
Red, Green, Yellow, Blue  
Nucleotides: spinning strands of DNA,  
Embedding truth, deep within the tapestry.  
The cipher - invisible yet ever present,  
Drives her onward, into the loom, and deeper.  
She leans over her work, into her life.  
She is confident that the  
Pattern will present itself over time.  
Emergent. Eternal.  
Gracefully, and without question,  
She continues to  
Weave.

\*B.S. Pre-med Chemistry, Saint Lawrence University, 1951

\*M.S. Biochemistry, Pennsylvania State University, 1954

\*PhD. Biochemistry, Pennsylvania State University, 1959

## ACKNOWLEDGEMENTS

### Funding Sources:

1. NIH #1R15HL072786-01  
Candidate Gene(s) for Pigeon Atherosclerosis  
Duration: May 1, 2003 – April 30, 2007
2. UNH Graduate School Dissertation Fellowship, AY 2005-2006
3. Animal & Nutritional Science Department Teaching Assistantship (2004)
4. UNH Graduate School TA Summer Fellowship (2004)

### Resources:

1. Detailed RDA protocol provided by Craig V. Byus, PhD.  
Professor of Biomedical Sciences and Biochemistry  
University of California, Riverside
2. Sequencing performed by the DNACore Service Center  
Hubbard Center for Genome Studies, University of New Hampshire

## TABLE OF CONTENTS

DEDICATION.....	iii
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES.....	ix
LIST OF FIGURES.....	xi
ABSTRACT.....	xii

CHAPTER	PAGE
INTRODUCTION.....	1
I. LITERATURE REVIEW.....	4
Human Atherosclerosis.....	4
Introduction to Human Atherosclerosis.....	4
Human Atherogenesis.....	5
The Observed Beginnings: Foam Cells and Lesion Development.....	5
Risk Factors and Atherogenesis.....	7
Theories of Human Atherogenesis.....	10
Genetic Defects in Human Atherogenesis.....	17
Models of Human Atherogenesis.....	24
Themes, Gaps, & Inconsistencies in Current Knowledge of Human Atherogenesis.....	43
Investigating the Genetic Differences between Two Similar Populations.....	46

Research Objectives.....	56
Hypotheses.....	56
Experimental Design.....	57
<b>II. METHODS AND MATERIALS.....</b>	<b>58</b>
Pigeon Colonies.....	58
Cell Culture System.....	58
Preparation of Total RNA.....	59
Preparation of Complementary DNA (cDNA).....	62
Representational Difference Analysis (RDA).....	65
Amplicon Preparation.....	65
Primer Information.....	66
Driver Preparation.....	67
Restriction Digest Parameters.....	69
Adapter Ligation Protocol.....	69
Phase Extraction Parameters.....	70
Precipitation Parameters.....	71
Tester Preparation.....	71
Round One Hybridization.....	72
Amplification of Round One Hybridization Products.....	72
Purification & Assessment of Round One Difference Products.....	74
Subsequent Rounds of RDA.....	75
Cloning the RDA Difference Products.....	76
Sequencing of Cloned Difference Products.....	79



Sequence Analysis and Characterization of RDA Difference Products.....	82
Raw Sequence.....	83
Vector Trimming.....	83
BLASTn.....	84
Sequence Quality.....	84
TBLASTx.....	86
Unidentified Difference Products.....	87
Annotation of Difference Products.....	88
Assignment of Pigeon Transcripts to the Chicken Physical Map.....	91
Statistical Analysis of Differential Expression.....	93
Determination of False Positives.....	94
III. RESULTS.....	95
IV. DISCUSSION.....	141
Biological Relevance of Results to Atherogenesis.....	141
Determination of Candidate Genes Contributing to Pigeon Atherosclerosis.....	141
The Five Major Themes Associated with Atherogenesis in the Pigeon.....	150
Theme One: Energy Metabolism.....	150
Theme Two: Phenotypic Modification.....	153
Theme Three: Immune Response.....	158
Theme Four: Signaling, Receptors, and Transcription Factors.....	160
Theme Five: Protein Modification & Metabolism.....	165

Theories of Human Atherogenesis Supported by Pigeon RDA Results.....	170
The Chicken as a Comparative Genomics Resource for the Pigeon.....	172
Limits to Data Interpretation.....	175
Inherent Limitations of RDA.....	175
Experimentally Introduced Limitations.....	177
Reproducibility & Efficacy of RDA for Identifying Differentially Expressed Genes in the Pigeon.....	179
CONCLUSIONS.....	180
LIST OF REFERENCES.....	182
APPENDICES.....	202
APPENDIX A:    ANIMAL RESEARCH DOCUMENTATION.....	203
APPENDIX B:    SELECTED PROTOCOLS.....	204
APPENDIX C:    RAW DATA TABLES.....	205
APPENDIX D:    ANNOTATION TABLES.....	219

## LIST OF TABLES

<u>Table</u>	<u>Title</u>	<u>Page</u>
Table 1	Comparison of Selected Characteristics of Atherosclerosis.....	27
Table 2	Comparison of Strategies to Investigate the Transcriptome.....	54
Table 3	RDA Primer Sets.....	66
Table 4	Raw Sequence Data and BLAST Results.....	96
Table 5a	Differentially Expressed Genes Exclusive to White Carneau.....	97
Table 5b	Differentially Expressed Genes Exclusive to Show Racer.....	99
Table 5c	Differentially Expressed Genes Found in WC and SR.....	101
Table 6	General Biological Functions (KEGG Level 1).....	102
Table 7	Types of Metabolism (KEGG Level 2).....	103
Table 8	Types of Metabolism and Specific Pathways (KEGG Level 3)..	105
Table 9	Functional Cell Compartment of WC and SR Differentially Expressed Genes (Pathway Studio).....	111
Table 10	WC and SR Genes Expressed by Functional Cell Compartment (Pathway Studio).....	112
Table 11	Types of Metabolism (Pathway Studio).....	116
Table 12	WC and SR Genes Expressed by Types of Metabolism (Pathway Studio).....	117
Table 13	Types of Biological Processes (Pathway Studio).....	120
Table 14	WC and SR Genes Expressed by Biological Process (Pathway Studio).....	124
Table 15	Comparison of WC, SR, and Chicken (Gg) Gene Count per Chicken Chromosome.....	131

Table 16	Comparison of WC and SR Transcript Count per Chicken (Gg) Chromosome.....	134
Table 17	Predicted Amino Acid Sequences of WAG-65N20 Clones.....	136
Table 18	Predicted Amino Acid Sequences of Un-Annotated Clones.....	137
Table 19a	Predicted Amino Acid Sequences of Unidentified WC Transcripts.....	138
Table 19b	Predicted Amino Acid Sequences of Unidentified SR Transcripts.....	138
Table 20	Technical Replication Analysis.....	140
Table 21	Candidate Genes Determined by Copy Number and Reproducibility.....	142
Table 22	Candidate Genes Determined by Chromosomal Position.....	145
Table 23	Five Major Themes as Determined by KEGG and Pathway Studio.....	147
Table 24	Top Candidates Determined from Multiple Perspectives.....	149
Table 25	Coding Region of Avian ND4 Gene within Mitochondrion.....	173
Table 26	Cell Culture and RDA Experimental Parameters.....	177
Table B.27	SLOWCOOL.....	204
Table C.28	Differential Gene Expression in Cultured Pigeon Aortic Cells (Combined Totals).....	205
Table C.29	Differential Gene Expression in Individual Experiments.....	211
Table C.30	Chi Square Analysis of Unidentified Difference Products.....	217
Table D.31	Distribution of Differentially Expressed Pigeon Transcripts by Chicken Chromosome.....	219

## LIST OF FIGURES

<u>Figure</u>	<u>Title</u>	<u>Page</u>
Figure 1	Maturation and Degeneration of Pigeon Aortic Cells.....	42
Figure 2	RDA Experiment.....	57
Figure 3	Comparison of WC, SR, and Chicken (Gg) Gene Count per Chicken Chromosome.....	129
Figure 4	Comparison of WC and SR Gene Count per Chicken Chromosome.....	130
Figure 5	Comparison of WC and SR Transcript Count per Chicken Chromosome.....	133
Figure 6	Avian ND4 Phylogram (CLUSTALW).....	174

## ABSTRACT

### DIFFERENTIALLY EXPRESSED GENES IN AORTIC CELLS FROM ATHEROSCLEROSIS-RESISTANT AND ATHEROSCLEROSIS-SUSCEPTIBLE PIGEONS

By

Janet Lynn Anderson

University of New Hampshire, May, 2008

Representational Difference Analysis (RDA) was used to identify genes that were differentially expressed between White Carneau (WC) and Show Racer (SR) pigeon aortic smooth muscle cells. The gene(s) responsible for atherosclerotic resistance in cultured SR smooth muscle cells (SMC) were hypothesized to be silent or down regulated in the WC. In the reciprocal experiments, it was hypothesized that the gene(s) contributing to the spontaneous atherosclerotic phenotype in cultured WC SMC would not be expressed in the SR.

Total RNA was extracted from primary cultured cells of each breed, converted to cDNA, and compared in four reciprocal RDA experiments. Seventy-four transcripts were identified exclusively in the WC cells, and 63 were unique to the SR. Genes representing several biochemical pathways were

distinctly different between aortic cells from susceptible (WC) and resistant (SR) pigeons.

The most striking genetic differences were observed in energy metabolism and smooth muscle contractility. The WC cells derived their energy from glycolysis, while the SR cells utilized oxidative phosphorylation to produce energy. Myosin light chain kinase and alpha actin were exclusively expressed by the SR SMC, whereas beta actin and collagen were dominant in the WC. Because of the compressed *in vitro* time frame compared with *in vivo* development, it was not obvious whether insufficient ATP synthesis is preventing the WC aortic cells from performing their contractile function or if the lack of functional contractile elements in the WC causes the mitochondrial ATP synthesis to down regulate. Either way, energy production was successfully coupled to muscle contraction in the SR, but not in the WC. This difference was observed prior to lipid accumulation in the WC cells, and appears to be a major contributing factor in pigeon atherogenesis.

One hundred forty five pigeon transcripts were homologous to the chicken. However, the mitochondrial genes expressed in the pigeon were more closely related to non-domestic birds such as the turkey vulture and oriental stork. Despite this categorical exception, the recently published chicken genome was an ideal resource for identifying differentially expressed genes in the pigeon. The results were interpreted in the context of current hypotheses of human atherogenesis. The pigeon transcripts can also be used in comparative studies of avian genomics.

## INTRODUCTION

Atherosclerotic cardiovascular disease is the leading cause of death in the United States and other economically developed countries. One of the earliest manifestations of this disease is the formation of lesions, known as plaques, in the arterial wall. These plaques accumulate lipid, proteoglycans, and fibrous material which can lead to a restriction in blood flow, increasing the likelihood of a heart attack or stroke. Many environmental factors such as diet, lack of exercise, stress, and smoking habits are known to contribute to the severity of the lesion once it is formed. However, the mechanism(s) that initiate lesion development is unknown, and quantifiable risk factors such as elevated blood cholesterol, triglycerides, and blood pressure fail to predict over fifty percent of heart disease mortality.

Despite the prominent role of heritability in the “pathophysiology of atherosclerosis, clinical risk assessment and therapeutic decision making are still based on classical risk factors.”<sup>1</sup> Until the genetic basis for susceptibility to atherosclerosis is understood, correlation of various risk factors with specific metabolic or pathological features will be difficult to assess, and efforts for prevention will remain equivocal. The individual and synergistic effects of multiple genes, various linkages, and gene-environment interactions that contribute to the atherosclerotic phenotype are difficult to separate. Work reported to date has looked at gene expression in human atherosclerotic



plaques compared to non-involved arterial segments, or at the identification of genes and loci correlating with lesion severity in knockout mice. In both cases, the research focus has been on end point gene expression, rather than initiation of the disease itself. This is problematic because it is only when the genetic contributions to lesion initiation *and* progression are understood that the true impact of dietary and lifestyle choices can be determined. Clinical trials have demonstrated that the “efficacy of dietary factors to prevent cancer is probably stage dependent”<sup>2</sup>, and it is reasonable to expect that this will also be the case for heart disease.

An animal model that could facilitate characterization of the initiating mechanisms of atherosclerosis without the complications of environmental or genetic modifications would allow the systematic analysis of lesion pathogenesis under a wide variety of conditions. The domestic pigeon (*Columba livia*) is an appropriate model of human atherosclerosis for several reasons. The foremost reason is susceptible and resistant breeds exist within the same species, eliminating the need to create an artificial atherosclerotic phenotype in the laboratory. In addition, susceptibility to spontaneous (non-induced) atherosclerosis in the pigeon is inherited as an autosomal recessive trait. Susceptibility resides at the level of the arterial wall and is manifested as lesions at the celiac bifurcation of the descending aorta. These lesions are pathologically and biochemically similar to the human lesion, yet occur in the absence of elevated blood cholesterol and other traditional risk factors. Comparisons of genetic and metabolic differences between the susceptible and

resistant pigeon breeds are further facilitated because preliminary cellular features of atherogenesis are also seen in pigeon aortic cell cultures.

Representational Difference Analysis (RDA) is a positive selection technique to determine differential gene expression between two closely related species. RDA will be used to characterize the genetic contribution to spontaneous atherosclerosis in the pigeon by comparing the gene expression between susceptible and resistant pigeon aortic cells. Once a panel of differentially expressed genes has been identified, it will become feasible to monitor atherogenesis and disease progression in response to diet, stress, and other environmental factors. Since gene structure, function, and pattern of expression is extensively conserved across living organisms, candidate genes isolated from the pigeon can be screened against the human genome and analyzed in terms of current hypotheses of human atherogenesis.

## CHAPTER I

### LITERATURE REVIEW

#### **Human Atherosclerosis**

##### Introduction to Human Atherosclerosis

Atherosclerosis is the most common form of heart disease, a general term encompassing a variety of pathologies affecting the heart and circulatory system. More specifically, atherosclerosis is a disease of the arteries, and is most likely to develop at branch points and other regions of low shear stress along the arterial tree, such as the celiac bifurcation of the aorta, and in coronary and carotid arteries<sup>3,4</sup>.

Atherosclerotic lesions begin to develop during childhood as lipid-filled foam cells making up "fatty streaks"<sup>5,6</sup>, and slowly progress through fibrous lesions into complex plaques consisting of multiple cell types, intra- and extracellular cholesterol esters, calcium deposits, proteoglycans, and extensive connective tissue. The final and terminating event of atherosclerosis is blood vessel occlusion, often caused by plaque rupture, which can lead to a heart attack, stroke, or embolism, depending on the location of the affected artery. However, not all fatty streaks progress to the stage of an advanced lesion<sup>7</sup>, and the rate of progression/regression, although well correlated with classical risk factors, is unique to each individual.

Clinical symptoms of atherosclerosis do not usually appear until later in life<sup>5,8</sup>. Therefore, research and intervention strategies have focused on delaying the progression of plaque formation rather than preventing the appearance of foam cells or fatty streaks in the first place. There is a strong familial component to all forms of heart disease, and many genetic disorders have been identified that contribute to lesion progression and the probability of plaque rupture in the general population. However, little is known about the specific genes that determine disease predisposition or how these genes interact with each other and the environment to initiate atherosclerotic foam cell formation in any one individual.

### Human Atherogenesis

The Observed Beginning: Foam Cells and Lesion Development. In human lesions, early foam cells originate primarily from vascular smooth muscle cells (VSMC)<sup>9</sup>. They are the first cell type to appear in susceptible regions of the aorta<sup>10,11</sup>, and the most abundant cell type in the developing fatty streak<sup>9,12-15</sup>. As early as 1964, using electron microscopy, Balis et al. noted that VSMC were often filled with lipid when there was no lipid in either existing macrophage cells or in the extracellular space, but the reverse was never observed. Therefore, the following series of events was proposed<sup>10</sup>:

1. VSMC accumulate lipid
2. Sloughing of endothelial cells, compromising VSMC membranes
3. Lipid released from VSMC and enters extracellular space
4. Macrophage cells arrive and phagocytose the lipid
5. Lipid accumulates in the macrophage cells

There has been little reason to reconsider this sequence of events except for two additional observations. First, multiple investigators have since reported that abnormal accumulation of VSMC in susceptible regions of the aorta precedes the actual lipid accumulation<sup>11,12</sup>. Second, atherosclerotic foam cells can be derived from both VSMC and macrophage cells<sup>9,16</sup>, depending on the physical location<sup>17</sup> and the cause of initiation.

Plaques that develop along the descending thoracic aorta tend to have more macrophage cells than VSMC, whereas plaques along the abdominal aorta and coronary arteries are comprised mostly of VSMC, with very few macrophage cells present in early stages<sup>9</sup>. Thoracic plaques are very rare in humans, and those that do progress are usually secondary to other chronic conditions such as hypertension and hyperlipidemia<sup>9</sup>.

It is interesting that although VSMC are the first cell type to accumulate lipid and initiate the fatty streak, a new researcher to the field of atherogenesis would not be aware of this key event. Most of the current literature describes only the mechanisms of macrophage foam cell development. For example, using the key words "macrophage foam cells" and "myogenic foam cells" to search the PubMed Database (*NCBI: September 2007*) results in 2048 hits, and 7 hits, respectively. In December 2006 the same search resulted in 1957 hits and 7 hits, further revealing the research trend. This emphasis on macrophage foam cells rather than myogenic foam cells could be for many reasons. Macrophage cells are the predominant cell type in common animal models of human atherosclerosis, especially transgenic mice. In these animals,

macrophage-derived foam cells are quick to develop into lesions and are easy to induce with a high-fat and/or high-cholesterol diet<sup>18-20</sup>. Also, unlike VSMC, which can alternate between two phenotypes, contractile and synthetic, macrophage cells do not change during the progression of the disease, and so are easier to identify in the laboratory under controlled conditions.

This research bias towards macrophage-derived foam cells is problematic because the pathogenic mechanism of lipid accumulation appears to be different for the two cell types<sup>5,10</sup>. Unfortunately, much of the clinical and genetic work on atherogenesis has focused exclusively on the macrophage cell and the immune response as the primary event. However, preliminary lesions observed in the aortas of humans aged 15-34 consist almost entirely of VSMC, with macrophages rarely present at this stage of the disease<sup>9</sup>. Therefore, rather than being an initiative mechanism in humans, the arrival of macrophage cells appears to be a secondary response, as they are far more common in advanced plaques than in early lesions<sup>5,9,21,22</sup>. In order to develop a more complete picture of atherogenesis relevant to humans, further investigation into the mechanisms of VSMC proliferation and foam cell development independent from the influence of macrophages is warranted.

Risk Factors and Atherogenesis. There are several physiological conditions and lifestyle patterns, which, if present, can increase an individual's risk of developing atherosclerosis. Major factors such as high blood cholesterol (high LDL/low HDL), high blood pressure, and diabetes are not just "used to

diagnose or monitor the disease but are used to estimate risk of developing heart disease and to estimate the risk of death (*American Heart Association Website, January 2007*). Collectively, these risk factors, along with physical inactivity, maleness, smoking, obesity, stress, heredity, and advanced age have been extensively researched and statistically correlated to specific stages of lesion development, plaque stability, and overall disease outcome in the general population.

In spite of genetic influences on traits such as LDL/HDL levels, blood pressure, and adiposity<sup>23</sup>, progress has been made on minimizing the effects of the controllable risk factors in order to disrupt, delay, reverse, or otherwise interrupt plaque rupture and aortic occlusion in “high risk” individuals. This has been possible largely because of the interrelationship of many of the individual risk factors, resulting in a synergistic effect. For example, controlling weight (obesity) can reduce the impact of diabetes and hypertension. Physical activity can in turn decrease weight and stress, and increase HDL levels. Despite moderate success, especially in the realm of cholesterol-lowering drugs, unknown genetic factors continue to influence both the age of onset and the frequency/severity of clinical symptoms<sup>1</sup>. Unfortunately, by the time most people manifest clinical symptoms, the lesions have become complex plaques, and it is too late to implement preventative measures. Therefore, the sooner susceptible individuals can be identified; the earlier treatment can begin. However, less than 50% of the mortality from coronary heart disease can be predicted by currently recognized risk factors<sup>24</sup>.

In order to understand what is happening in the at-risk population that remains unidentified under current screening methods, it is necessary to determine the specific contributions of heredity, diet, and lifestyle influences on atherogenesis and its progression. Towards this end, research emphasis has recently shifted towards identifying markers of cardiovascular disease that may be detectable prior to the manifestation of most of the classical risk factors. Markers are simply allelic variations that are known to associate with a specific disease phenotype. Markers do not necessarily cause the disease, but can be used to improve diagnosis and risk assessment<sup>23</sup>. Inflammatory markers such as C-reactive protein (CRP)<sup>24,25</sup> and markers of oxidative damage such as myeloperoxidase<sup>26</sup> and paraoxanase<sup>27</sup> have already increased the predictive power of clinicians. As more markers of atherosclerosis are identified and correlated with disease progression and outcome, the genetic variation contributing to predisposition and initial manifestation will become clear.

Until the genetic basis for susceptibility to atherosclerosis is understood, correlation of various risk factors with specific metabolic or pathological features will be difficult to assess, and prevention efforts will remain equivocal. Understanding the mechanisms of inheritance in atherosclerosis is an important step towards reducing the morbidity and mortality from the disease by customizing intervention strategies for individuals based upon unique genotypes and environmental risk exposures.



Hypotheses of Human Atherogenesis. A variety of hypotheses have been proposed in an attempt to explain how an atherosclerotic lesion is initiated in the human aorta. Although there are model-specific differences in the order of events, the pathological steps common to all theories of atherogenesis are<sup>5,8,9,12,28,29</sup>.

1. Site specific proliferation of intimal smooth muscle cells
2. Elaboration of excessive and/or abnormal extracellular components
3. Accumulation of lipids within and around cells
4. Entry of monocytes/macrophages into area of proliferation

The abnormal accumulation of lipid within smooth muscle and macrophage cells could arise from increased infiltration (influx), increased retention, decreased efflux, and/or increased lipid biosynthesis by the cells themselves.

The Lipid Infiltration Hypothesis states that cells in the arterial wall will accumulate lipid if there is a high concentration of circulating blood lipids, or a consistently elevated amount of low-density lipoproteins (LDL). In the healthy human aorta, circulating LDL particles are incorporated into SMC by receptor-mediated endocytosis. In atherosclerosis, the rate of LDL influx could overwhelm these receptors, causing the excess lipid to be taken up by scavenger receptors. Modified lipoproteins such as oxidized<sup>30</sup>, acetylated<sup>31</sup>, or particularly small (<70nm)<sup>32</sup> LDL are thought to slip through the loose junctions between endothelial cells and accumulate in the intima intact. In these cases, because of their altered conformation, it is hypothesized that the modified LDL molecules are readily incorporated into SMC by uncharacterized scavenger receptors. As foam cells develop and burst, macrophage cells are recruited to the region. The precise mechanism of how cholesterol from circulating LDL

enters the intima to be incorporated in the developing foam cell is not clear. In addition, the Lipid Infiltration Hypothesis does not, on its own, account for the observed proliferation of SMC prior to lipid accretion.

If lipid infiltration of any type is coupled with decreased HDL levels, there will be reduced cholesterol clearance (efflux) from the cells, and the sterol will remain trapped. In addition, the innermost arterial cells are in a state of chronic hypoxia<sup>33</sup>. If the fatty acids released by neutral cholesterol ester hydrolase (NCEH) are not completely oxidized, the metabolites will accumulate at an accelerated rate<sup>34</sup>, and can potentially serve as substrate for endogenous cholesterol and/or triacylglycerol (TAG) synthesis. There is compelling evidence indicating that the increase of intracellular cholesterol is at least partly the result of biosynthesis, and not uptake of circulating lipoproteins<sup>35</sup>. However, this mechanism is not explained by any of the current hypotheses of atherogenesis, which are mainly focused on the infiltration and retention of plasma lipids and their subsequent inflammatory effects.

One mechanism proposed to explain the increased retention of circulating blood lipids is the Response to Retention Hypothesis<sup>36</sup>. According to this idea, as proteoglycans (PG), especially versican<sup>32,37</sup>, accumulate in the extracellular matrix (ECM) of the proliferating smooth muscle and recruited macrophage cells, they bind to incoming LDL particles. The electrostatic interaction between the apoB component of LDL and the sulfated chains on the core PG protein binds the LDL to the cell surface<sup>38</sup> where its solubility is decreased<sup>39</sup>. PG-bound LDL is also more likely to become oxidized, and in

either case, the trapped lipoprotein is incorporated into the developing foam cell. Presumably the lipid enters the individual cells by the action of scavenger receptors, but proponents do not directly address this component. Despite this omission, the Response to Retention Hypothesis does provide a concrete mechanism for the adherence of circulating lipoproteins to the arterial intima. Therefore, advocates of this hypothesis claim that “essentially all later advances can be traced to<sup>36</sup>” the initial attraction of circulating LDL to ECM proteoglycans.

Although the aforementioned explanations of lipid infiltration and retention provide mechanistic evidence of how lipoproteins can accumulate in the arterial intima, there are key steps of atherogenesis that are not addressed. Neither hypothesis offers direct evidence for how cholesterol esters form within early foam cells, nor do they explain the initial cellular proliferation of SMC prior to lipid accumulation. Neither hypothesis explains the subsequent entry of monocytes and macrophages to the region of infiltration, nor do they explain the predictable locations of lesion development along the arterial tree.

The observation that both smooth muscle and macrophage cell types were actively recruited to the site of balloon catheterization led to the “Response to Injury” hypothesis<sup>11</sup>. According to this idea, the arterial endothelium is compromised by various perturbations such as environmental chemicals, high concentrations of blood lipids, certain types of bacteria and viruses, autoimmunity, and/or hemodynamic stress<sup>40</sup>. In response, the endothelial cells can either slough off or become porous, allowing the subsequent influx of lipoproteins and macrophages into the arterial intima.

Once initial damage has occurred, the exposed intimal cells are increasingly vulnerable to additional hemodynamic and environmental aggravation, thus perpetuating the original injury and eliciting a multi-pronged immune response. Although endothelial denudation or injury is not necessary for foam cell formation, there is an obvious and observable inflammatory response that seems to exacerbate the developing plaque.

Atherosclerosis is now considered to be a chronic inflammatory disease<sup>24</sup>, largely because in a variety of animal models, signs of inflammation occur hand in hand with hypercholesterolemia<sup>41-44</sup>. This hypothesis proposes that the immune system is activated as a direct result of lipid infiltration<sup>45,46</sup> and so readily explains the presence of monocytes and macrophages in the fatty streak. These cells express scavenger receptors that not only ingest lipid, especially oxidized LDL; they actively secrete cytokines and recruit adhesion molecules to the region. These actions are thought to be directly responsible for the increase in extracellular components observed during later stages of atherogenesis<sup>47</sup>. As the matrix between cells becomes more complex, there are more opportunities for proteoglycans to bind and transform lipid molecules, and so perpetuate the entire process of macrophage recruitment and cytokine signaling. In addition to macrophage cells, endothelial cells, SMC, and platelets in the developing lesion are all capable of synthesizing and/or releasing chemoattractants and growth factors<sup>9</sup>. These cellular interactions work together to expand the initial fatty streak to a more advanced fibrous plaque. The inflammatory response is well correlated with plaque stability, and there are

already blood tests available that will assess the risk of thrombosis based on the levels of inflammatory markers<sup>23</sup>. However, as with other hypotheses of atherogenesis, the inflammatory response does not seem to account for the initial proliferation of SMC prior to lipid accumulation and macrophage recruitment. Nor does the inflammatory response explain that, in humans, early foam cells are primarily myogenic. For these reasons, the inflammatory hypothesis explains the mechanisms of plaque progression and the likelihood of disease endpoint better than it explains atherogenesis itself.

A fourth hypothesis of atherogenesis is that the smooth muscle cells that abnormally proliferate and accumulate lipid are transformed and of monoclonal origin<sup>48,49</sup>. The SMC which develop into foam cells are phenotypically different from their counterparts in the normal media<sup>14,50</sup> in that they actively secrete ECM components such as proteoglycans and collagen<sup>51</sup>. Healthy, fully differentiated SMC are slow to proliferate and do not synthesize an extensive ECM<sup>52</sup>. Loss of cell cycle control and the ability to regulate cholesterol metabolism are early hallmarks of cancerous cells, and it is interesting that both of these pathological phenotypes are seen in atherogenesis. The idea of monoclonal SMC has not been well accepted by the field. However, technological advances in DNA sequence analyses have confirmed that although SMC in general are largely heterogenic, subsets of cells involved in atherogenic events are derived from specific clones<sup>53,54</sup> that seem to proliferate in patches<sup>55</sup>. A second observation that lends support to this controversial hypothesis is the recent observation that the DNA in atherosclerotic lesions is

hypo-methylated<sup>56</sup>. Decreased methylation is significantly correlated with increases in transcriptional activity. During the development of cancer, normally silent oncogenes can be turned on (expressed) because of under methylation<sup>57</sup>. More research is needed to determine the role of hypo-methylated DNA and clonal SMC cells in atherogenesis.

Although the monoclonal origin hypothesis explains the proliferation of an altered population of lipid-accumulating, ECM-producing SMC in the arterial intima, no specific transformation event or set of pathological conditions have been proposed that would account for the initial change in differentiation state, methylation status, and/or the increased rate of mitosis of a specific subpopulation of SMC. In addition, it is difficult to explain the reversibility of lesions with the Monoclonal Hypothesis, and there are examples in both animals and humans where atherosclerotic lesions appear to regress<sup>58-60</sup>.

Smooth muscle cell differentiation is a complex process that requires more than just a few transcription factors to be successful<sup>61</sup>. In the developing embryo, SMC are in the "synthetic" state as they are actively proliferating and synthesizing their contractile elements, myofilaments, and ECM to form the arterial intima<sup>51,62</sup>. Once the blood vessels are fully formed, the SMC reach the "contractile" state where they stop proliferating and function to facilitate muscle contraction in response to stimuli. A healthy vessel wall is able to maintain both contractility and the quiescent state<sup>50,63</sup>.

It has been observed that during atherogenesis, SMC seem to revert back to the synthetic state where they abnormally produce an extensive ECM<sup>64</sup>.

This phenotypic modulation<sup>51</sup> is believed to occur before the cells begin to replicate and migrate to the arterial intima, but the stimulus for this change is unknown. An alternative explanation is that in individuals predisposed to atherosclerosis, the SMC never fully differentiate in certain arterial regions. In either case, modified SMC are characterized by a decrease in the alpha/beta actin isoform ratio<sup>14,50,65</sup> and the loss of the intermediate filament proteins such as vinculin and desmin<sup>50,65,66</sup>. Research in this area has focused on identifying more SMC phenotypic markers<sup>52,61,67</sup> in order to both clearly define the state of differentiation, and to articulate what is driving the change in differentiation state<sup>68,69</sup>. Although this idea is narrowly focused on the role of SMC, and does not address macrophage recruitment, it is a valid attempt to describe what is happening in the earliest stages of atherogenesis at the cellular level.

Hemodynamic stress is a pervasive factor in all hypotheses of atherogenesis, because lesions typically develop at aortic regions of bi-directional flow. However, if it were simply a matter of arterial architecture, all humans would develop atherosclerosis. Indeed, any organism with a branching aorta would spontaneously develop foam cells and initiate the atherosclerotic pathology. Because this is not the case, there must be some as yet unidentified factor intrinsic to the vessel wall of resistant individuals that can withstand the effects of low shear stress and maintain the cells in a contractile phenotype.

It is clear that all of the current hypotheses have gaps. None of them provide a causative explanation that includes all features and is both necessary and sufficient to invoke the full atherosclerotic cascade. Many of them explain

lipid accumulation in the arterial wall, some of them describe the preferred site of fatty streak formation and the appearance of macrophage cells, but none of them completely describe the entire series of events that occurs in human atherogenesis. Genetic factors clearly influence cholesterol metabolism<sup>70</sup>, replication rates<sup>71</sup>, the immune response<sup>22,43</sup>, and the oxidative capacity of the mitochondria for cellular lipids<sup>72</sup>, thereby manifesting an underlying influence on all aspects of atherogenesis that warrants further discussion.

Genetic Defects in Human Atherogenesis. There is clearly a higher relative risk for atherosclerotic incidents in individuals with a familial history, than there is for those with a susceptible lipid profile<sup>1,70,73</sup>. Many studies have been conducted to explore and quantify the relationship, or concordance, between heredity and atherosclerosis. Heritability has been found to have an overall concordance rate with early-onset coronary heart disease of 0.63<sup>74</sup>. The relationship becomes even clearer after analyzing concordance in monozygotic and dizygotic twins. Twins fertilized from one egg (monozygotic) have a concordance rate of 0.83, whereas twins that arose from two separate fertilizations (dizygotic) demonstrate a concordance rate of 0.22<sup>74</sup>. These concordance values suggest a tight correlation between the genotype of an individual and the incidence of heart disease.

The fact that the concordance rate in monozygotic twins is not 1.0 (indicating 100% correlation) most likely reflects the attenuating effects of environmental risk factors on atherosclerosis initiation and progression. This



gap in causality underscores the importance of understanding the genetic profile of a client before attempting intervention, because even among those sharing the same set of alleles, the phenotype of atherosclerosis will vary depending on individual exposures.

Genetic research on human atherosclerosis has focused primarily on the role of cholesterol metabolism. It is estimated that several hundred genes<sup>70</sup> are involved in the absorption, conversion, transport, deposition, excretion, and biosynthesis, of cholesterol and other lipid substrates in the body<sup>19,75</sup>, and very few have been characterized. A defect in any of these pathways may contribute to atherosclerotic susceptibility, because the net result can be a significant increase in plasma lipoprotein concentration, especially LDL, and/or the inappropriate deposition of cholesterol in peripheral tissues such as skin, tendons, and arteries<sup>76</sup>.

Blood lipid homeostasis and cellular cholesterol metabolism are known to be highly regulated<sup>77</sup>, and genetic defects have been found to impact overall cholesterol metabolism at many locations. In humans, most plasma cholesterol is in the form of LDL, and the half-life of circulating LDL is believed to be about 2.5 days<sup>78</sup>. Some of the cholesterol component of LDL is transferred to HDL via the action of cholesterol ester transfer protein (CETP). However, as much as 70% of LDL is removed from the blood by LDL receptors (LDLR) in the liver<sup>76</sup>. A variety of single gene defects have been identified that increase the incidence of atherosclerosis by influencing the activity of the LDLR<sup>1</sup>.

Probably the most studied of these LDLR defects is familial hypercholesterolemia (FH). FH is an autosomal dominant Mendelian disorder that is caused by a mutation in the LDLR gene<sup>1,78,79</sup>. This mutation renders the hepatic receptors nonfunctional, so that they are unable to clear circulating LDL from the blood. A second type of hypercholesterolemia, Autosomal Recessive Hypercholesterolemia (ARH), results from an LDLR mutation inherited as an autosomal recessive trait<sup>76,78</sup>. ARH is similar to FH, in that both of these hereditary defects result in chronically elevated blood cholesterol. However, unlike FH, it is believed that in ARH, the LDLR are functional, but their altered location in the liver makes them inaccessible to circulating LDL<sup>76,78</sup>.

Brown and Goldstein have also identified a single gene defect known as familial ligand defective apoB-100. ApoB-100 is the primary apoprotein in human LDL<sup>80</sup>, and this inherited disorder results in a decreased ability of the LDL to be picked up by the LDLR<sup>78,81</sup>. In this situation, the hepatic receptors are in the correct location, and they are functioning properly. The defect lies in the composition and binding capacity of the apoB-100 to the LDLR. There are also apoB-100 binding sites on the ECM components versican and biglycan<sup>82</sup>. These proteoglycans have a lower affinity for the apoB-100 portion of LDL than the LDLR, but have a much higher capacity<sup>78</sup>. Therefore, genetic mutations in the apoB-100 could impact both cholesterol clearance rates (directly) and levels of lipoprotein retention in the arterial wall (indirectly).

In the healthy human aorta, LDL particles are thought to be incorporated into SMC by receptor mediator endocytosis. Chemically modified or oxidized

LDL enters via scavenger receptors. Once inside the cell, the cholesterol esters (CE) in LDL are transported to the lysosomes where they are hydrolyzed by lysosomal acid lipase (LAL), also known as acid cholesterol ester hydrolase (ACEH). This enzyme breaks each CE into its free fatty acid (usually linoleate), and free cholesterol. There are several known genetic mutations to the LAL gene that result in the abnormal accumulation of cholesterol esters in the lysosome.

Two of the more common lysosomal storage disease phenotypes of a congenital LAL defect are Wolman's Disease<sup>83,84</sup> and cholesterol ester storage disease (CESD). Both are inherited as an autosomal recessive trait, although Wolman's disease is usually fatal within the first year of life, and so not directly related to atherogenesis in the general population. However, individuals with CESD do demonstrate premature atherosclerosis, although they also accumulate CE and triglycerides (TG) in the liver, adrenal glands and intestines<sup>85</sup>. Niemann-Pick Type C is a third form of lysosomal storage disease that directly impacts cholesterol metabolism at the cellular level<sup>86</sup>. In this scenario, the CE is successfully hydrolyzed by ACEH, but the released cholesterol component is unable to leave the lysosome to travel to the endoplasmic reticulum, causing the accumulation of free cholesterol in the lysosome.

Lysosomes are also responsible for the degradation of glycosaminoglycan (GAG) chains after the core proteoglycan has been broken down by extracellular proteases such as matrix metalloproteinases (MMP) and

disintegrins (ADAM)<sup>87,88</sup>. There is an extensive repertoire of catalytic lysosomal enzymes, and their functions have been revealed mostly by observing the consequences of their absence<sup>89</sup>. Defective enzymes lead to a wide variety of diseased phenotypes known as mucopolysaccharidoses (MP) ranging from the mild Schie Disease to the severe Hurler Disease, which results in childhood mortality. In these two examples, GAGs are not properly degraded, and so will accumulate in the lysosomes and in the extracellular space. GAGs in the ECM will attract LDL that has entered the intima by binding to apoB-100 as previously described, and the cholesterol is most likely endocytosed by macrophage and SM cells within the developing plaque.

Once in the cytoplasm, cholesterol that is not needed for routine cellular functions is esterified by acyl CoA: cholesterol acyltransferase (ACAT) and stored in vacuoles. Intracellular CE remain trapped in the cytoplasm until hydrolyzed by NCEH. This enzyme releases the free cholesterol so it can be removed by HDL and taken to the liver for excretion. A pair of ATP binding cassette proteins has been identified that are believed to control this efflux of cellular cholesterol. One of these, ABCP-1 is defective in Tangier Disease<sup>90</sup>, an inherited condition where cholesterol is unable to exit the cell via reverse cholesterol transport. There is a moderate risk of atherogenesis associated with Tangier Disease, and this risk is increased in the presence of additional risk factors<sup>91</sup>.

Research is also being directed towards a range of apoproteins associated with HDL. It has been suggested that approximately 50% of the

variance of HDL composition and plasma concentration in the general population is due to genetic factors<sup>91</sup>. The primary apoprotein in HDL is apoA1, followed by apoA2, apoC, and apoE<sup>80</sup>. ApoE is an important ligand for receptor mediated clearance of HDL from arterial cells<sup>75,92</sup>. The role of apoE is of great interest to investigators of atherosclerotic resistance because most patients with familial dysbetalipoproteinemia (FD) are homozygous for the E2 isoform of apoE<sup>93</sup>. Although this defect has been shown to be relevant in some animal models, especially apoE null mice<sup>20,94</sup>, only 1-4% of humans with the E2/E2 apoE phenotype actually develop FD<sup>93</sup>. The pathological influence of apoE dysfunction is important in these genetically susceptible individuals, but may not be relevant to the more common forms of atherosclerosis in the overall population.

Any of the currently identified monogenic defects that directly or indirectly influence cholesterol metabolism and/or the inflammatory response will increase the likelihood of atherosclerotic events. However, individual genes do not always act alone, and additional genetic and/or environmental factors may be required to determine the overall susceptibility or resistance. Nuclear hormone receptors and other types of transcription factors are under investigation to determine how they exert their regulatory effects<sup>95,96</sup>. For example, although the binding capacity of apoB-100 is genetically determined<sup>78</sup>, the specific number of hepatic LDLR being expressed at any given time is dependent on dietary and hormonal factors<sup>81</sup>. In a hypothetical situation, the apoB domain of LDL may be functional (non-mutated), but without the adequate

expression of the LDLR to bind circulating LDL, the end result could still be high blood cholesterol.

Clinical studies have demonstrated that not all individuals afflicted with FH will develop early onset atherosclerosis. Of those manifesting the heterozygous form of the disease, where circulating LDL levels tend to range between 300–400 mg/dL, only 50% will actually develop cardiovascular disease<sup>75</sup>. This is because there are both hyper- and hypo- responders to the effects of dietary cholesterol on serum levels, and some individuals demonstrate relative resistance to atherogenesis, even in the face of hypercholesterolemia<sup>75</sup>. In addition to looking for genetic defects that contribute to individual predisposition and susceptibility, the goal should also be to identify protective factors that denote resistance.

The ultimate sequel of atherosclerotic events is a result of the combined effects of many genes, regulatory factors, and environmental exposures<sup>97</sup>. This synergistic influence on phenotype may give the appearance of a polygenic or multifactorial effect<sup>1,78</sup>, even when a monogenic abnormality has been clearly implicated. These interactions have made it difficult to establish a universally accepted mechanism of atherogenesis<sup>98</sup>, because the sample sizes needed to test these gene-gene and gene-environment interactions are much larger than those needed for simpler genotype-phenotype associations<sup>70</sup>. There have not been enough human population studies that allow a true “estimate of the percentage of genes where a single mutation has a contribution to the phenotype large enough to be detectable<sup>1</sup>.” It is often difficult to determine

whether a polymorphism is THE initiating factor, or if the mutation has simply associated with the susceptible or resistant phenotype under study by chance<sup>99</sup>. For these reasons, “although there has been considerable success in identifying genes for the rare disorders associated with atherosclerosis, the understanding of genes involved in the more common forms is largely incomplete<sup>100</sup>”.

Pathways that trigger atherosclerosis in the general population have yet to be elucidated<sup>27</sup>. Most genomic scale experiments have compared either full-blown plaques against non-affected aortic segments<sup>101-104</sup>, or they have analyzed differences between plaques that have ruptured and those that have not<sup>105-107</sup>. In both types of comparisons, differentially expressed genes have been identified that shed light on plaque development and mortality risk. However, genes responsible for initiating foam cell formation could not be discriminated from those involved in later events. This gap is not an oversight by the investigators, but rather reflects the limited availability of human tissue samples at early stages of atherosclerosis for relevant comparative studies. One of the major limitations of elucidating the sequence of events that occur during atherogenesis is that an investigator can “observe and study a single site in the arterial vasculature” only once<sup>11</sup>. For this and other reasons, most atherogenic research requires animal and in-vitro models of the human disease.

Models of Human Atherogenesis. No animal model of human disease can fully encompass the unique complexity of molecular machinery and the

wide range of expressed clinical phenotypes. However, many important metabolic pathways have been explained by the judicious use of animal models<sup>97</sup>. Therefore, the selection of a disease model for genetic inquiry becomes important, and the most appropriate choice will ultimately depend on the specific hypothesis or research question being investigated.

There are some general guidelines to follow when choosing an animal model of human disease. The phenotype should resemble the physiological condition of humans as closely as possible in both the normal and diseased state<sup>92</sup>. In the specific case of circulating cholesterol, genetics, and atherosclerosis, it is essential that the lipoprotein profile is comparable to humans, and the familial lineage is clearly defined<sup>99</sup>. In addition, there are practical issues to consider such as the size of the animal and housing requirements, generation times, and the specific cost of overall maintenance, including food, daily care, and experimental treatment<sup>92,108</sup>. These concerns become especially important with the development of transgenic models, in that the associated investment costs are much higher than with traditional animal studies.

There are currently several animal models being used to investigate various clinical manifestations and genetic mechanisms of human atherosclerosis. Although mice (regular laboratory and transgenic) are the most commonly used, rabbits, hamsters, miniature swine, primates, rats, dogs, and pigeons are also used. The primary contribution of these animal models has been to elucidate the role of specific molecules in atherogenesis, lesion



progression, thrombosis, and plaque rupture by direct hypothesis testing.

Selected characteristics of atherosclerosis in animal models and their relationship to the human disease are presented in Table 1 (page 27).

One of the difficulties encountered when studying lipid metabolism in animals is that most species circulate the majority of cholesterol in HDL<sup>108</sup>, whereas most human cholesterol is in the form of LDL<sup>76</sup>. For example, a decrease in plasma HDL has been associated with a reduced risk of atherosclerosis in mice<sup>109</sup>. It does make sense that relatively low levels of HDL decreased the clinical incidence of atherosclerosis because 70% of the total cholesterol in mice is in the form of HDL<sup>92</sup>.

However, in humans, decreased HDL levels are associated with an *increased* risk of atherosclerosis. Despite this marked inconsistency, the successful extrapolation of animal studies to human atherosclerosis is exemplified by the fact that Goldstein and Brown found that it was impossible to raise levels of circulating LDL, and thus increase the risk of atherosclerosis in experimental models, unless the LDLR were somehow compromised, either genetically or in response to dietary overload<sup>78,79</sup>. Subsequently, over 600 mutations in the human LDLR gene have been identified that result in varying degrees of hypercholesterolemia<sup>78</sup> similar to FH. In addition, it has been repeatedly demonstrated in hamsters, rabbits and primates, that the functional capacity of hepatic receptors is decreased in response to a diet high in fat<sup>108</sup>. It has also been demonstrated that there are individual variations in LDLR activity in response to dietary fat and cholesterol intake. For example, primates, dogs,

Table 1: Comparison of Selected Characteristics of Atherosclerosis

	Hamster		Mouse		Pig		Rabbit		Pigeon		Human	
	LDL	HDL	Regular	Transgenic	LDL	HDL	Regular	WH-HL/MI	HDL	HDL	LDL	LDL
<b>Lipoprotein Profiles</b>												
Predominant	+	+	+	+	+	+	+	+	+	+	+	+
CETP	+	+	+	+	+	+	+	+	+	+	+	+
LDLR	+	+	+	+	+	+	+	+	+	+	+	+
ApoE	+	+	+	+	+	+	+	+	+	+	+	+
ApoB-100	+	+	+	+	+	+	+	+	+	+	+	+
ApoB-48	+	+	+	+	+	+	+	+	+	+	+	+
<b>Lesions/Foam Cells</b>												
Primary Location	Arch	Root	Root	Root	Arch	Arch	Arch/Thoracic		Celiac Branch		Coronary/Celiac Branch	
Primary Cell Type	MØ	MØ	MØ	MØ	SMC	SMC	SMC/MØ		SMC		SMC	
Spontaneous	+	+	+	+	+	+	+	+	+	+	+	+
Diet-Induced	+	+	+	+	+	+	+	+	+	+	+	+
Thrombosis	+	+	+	+	+	+	+	+	+	+	+	+
Myocardial Infarction	+	+	+	+	+	+	+	+	+	+	+	+
<b>Genome Size (Gbp)</b>	3.6	3.4	3.4	3.4	3.1	3.4	3.4	3.4	3.4	1.4	3.4	3.4
<b>Wild-Type Diet</b>	Omnivore	Omnivore	Omnivore	Omnivore	Omnivore	Omnivore	Herbivore	Herbivore	Omnivore	Omnivore	Omnivore	Omnivore

and rabbits, like humans, have been shown to be hypo- or hyper- responsive to diet<sup>75,78,92,110</sup>, with some individuals demonstrating unique resistance.

In newborn humans and many animal species, hepatic LDLR have been shown to have a maximum operative capacity when circulating LDL levels are approximately 0.25 mg/dL<sup>111</sup>. Approximately 60% of plasma LDL in hamsters is removed by hepatic receptors. The hamster clearance rate for the LDLR is 3.1 mg/hr whereas the companion human LDLR only removes 0.6 mg/hour<sup>108</sup>. However, the fact that hamsters and humans share a common mechanism of LDL clearance makes the hamster a suitable model for this aspect of cholesterol metabolism. Hamsters also share the CETP with humans<sup>108</sup>. This molecule transfers the cholesterol component of LDL to HDL, a key step in reverse cholesterol transport. These homologous features are in direct contrast to the mouse, which, despite being fed a high-fat high-cholesterol diet<sup>112</sup> and its evolutionary relationship to the hamster, does not develop advanced atherosclerotic plaques resembling those found in humans unless mice with "sensitized genetic backgrounds<sup>18</sup>" are studied<sup>113</sup>.

The mouse is technically advantageous because of its small size, short generation time, large litters, and the availability of many inbred strains<sup>109</sup>. However, laboratory mice fed on a chow diet do not develop spontaneous atherosclerotic lesions. Atherosclerosis must be experimentally induced by feeding a diet containing 15% fat, 1.25% cholesterol, and 0.5% cholic acid. These non-physiological conditions create serious limitations for comparison with human studies. The most important factor may be the presence of cholate

in the diet. Cholate is enough, in and of itself, to put mice in a chronic inflammatory state<sup>109,114</sup> confounding the true role of inflammation in atherogenesis. This is further exacerbated by the fact that some mice are more sensitive to inflammatory cues<sup>115</sup> so that “although the inflammation correlates with atherosclerosis... it is possible that some of the genetic differences between susceptible and resistant mouse strains pertain to the diet used, rather than the atherogenic process as it is observed on Western diets<sup>109</sup>.”

These and other genetic differences that exist between mouse strains can cause significant problems when interpreting and comparing the results of gene expression studies<sup>116</sup>. For example, just because a specific inflammatory marker was identified in an atherosclerotic plaque and not in a healthy aorta does not mean that inflammation is causing the disease. Indeed, the molecule could be there to accelerate the cascade; but it could also be there in an attempt to reverse the pathology, or may even be responding to a cellular signal not specific to plaque progression<sup>19</sup> such as cholate. This is true even with transgenic mice because the foundation stock may be different. Also, as gene insertion is random, knock-in models do not by definition contain the gene of interest at the same locus. Therefore, simple transgenics may not be sufficient to prove the role of any given trait because of positional insertion effects on both absolute gene expression and copy number variation<sup>99</sup>. Delineating the specific function of a candidate gene is difficult, if not impossible, without being able to precisely correlate the phenotype back to the initiating mechanism of

foam cell formation; and, the heterogenic background of the mice combined with the variable responses to the atherogenic diet confound the interpretation.

Despite these often overlooked limitations of extrapolating mouse studies to the human disease, research using transgenic mice further developed the concept that atherosclerosis is not a simple lipid disorder. New hypotheses of atherogenesis must be explored in order to explain the occurrence of atherosclerotic heart disease in individuals displaying no dyslipidemia. Over twenty unique quantitative trait loci (QTL) have been identified in mice<sup>94</sup>, and most of them do not influence plasma lipid levels or blood pressure<sup>100,117</sup>. This finding has been especially interesting because these QTL were identified in hypothesis-driven experiments to look specifically at cholesterol metabolism in LDLR and/or apoE knockout mice. Many of these studies have demonstrated the strong influence of genetic factors in the arterial wall on the susceptible and resistant phenotypic differences between mouse strains<sup>118</sup>.

Knockout models theoretically mirror homozygous recessive forms of inherited disease because of a loss of gene function<sup>19</sup>. As in Familial Hypercholesterolemia (FH), LDLR null mice experience a 2X increase in plasma cholesterol levels, even on a regular diet, and the effect is exacerbated on the high-fat, high-cholesterol atherogenic diet<sup>19</sup>. The same is true for apoE null mice, although the impact of this mutation has a greater effect on plasma cholesterol compared with LDLR negative mice, with 4-5 times the normal amount of circulating lipoproteins<sup>19,20</sup>. However, preliminary studies revealed no relationship between these elevated lipid levels and lesion size in apoE null

mice<sup>119</sup>. In fact, only 2% of the homozygous apoE2 null mice developed aortic lesions at all, and the contribution of this mutation to the overall human disease burden has been questioned<sup>27,74</sup>. Subsequent studies have shown contradictory results, as the nature of the lesion appears to be dependent on the parental strain used in the experiment rather than the particular knockout gene<sup>7,94,100 116</sup>. The largest effect noticed to date in these hypercholesterolemic models is that of Macrophage Colony Stimulation Factor (MCSF) on lesion progression<sup>19</sup>. MCSF has been reported in advanced human atheromas, and this finding in mice lends experimental support to the role of the inflammatory response in atherosclerosis. However, the role of this molecule in atherogenesis per se is difficult to elucidate in the mouse, because of its chronically inflamed state.

Although not yet yielding consistent results applicable to human therapeutics<sup>120</sup>, transgenic mouse research has reinforced how important the genetic background is in determining atherosclerotic susceptibility and resistance in an individual. These studies have also suggested that the mechanism of foam cell formation is different in different individuals under different experimental and/or environmental stimuli. The importance of the specific initiating mechanisms on the developing phenotype has been further demonstrated in rabbit models of atherosclerosis.

Rabbits, like hamsters, have CETP and do develop atherosclerotic foam cells when induced by an unnatural diet<sup>108</sup>. Unlike the other animal models described in Table 1, rabbits are vegetarian, and so cholesterol is not a normal component of their wild-type diet. The Watanabee Heritable Hyperlipidemic

(WHHL) rabbit was developed through selective breeding, and does not have LDLR<sup>121</sup>. WHHL rabbits get lesions within six months along the aortic arch, but do not experience thrombosis or myocardial infarction. However, these advanced atherosclerotic phenomena are observed in a sub-strain, the WHHLM rabbit. This rabbit does get a heart attack similar to one of the endpoints of human atherosclerosis<sup>122</sup>.

One of the important contributions of the rabbit model to understanding human disease was the observation that foam cells in rabbits can be derived from smooth muscle cells (SMC) or macrophage cells (MØ), depending on the specific dietary perturbation<sup>123</sup>. This finding is in direct contrast to the mouse, where the predominant cell type in early lesions is always the macrophage cell, regardless of diet and genetic strain<sup>124</sup>. Myogenic foam cells in the rabbit are biochemically and morphologically distinct from macrophage derived foam cells, and both types of early lesions are structurally different from those produced by catheter injury<sup>123</sup>. Recognizing that different types of foam cells develop in response to different initiating mechanisms should help unravel the controversy of foam cell origin. In all probability, the predominant cell type in early atherogenesis is dependent on the pathological stimulus, and the specific model under study.

A second revelation from rabbit research has been that both MØ and SMC express receptors for the MCSF protein<sup>125</sup>. The proto oncogene *c-fms3* induces SMC migration and proliferation, as well as MØ recruitment to the atherosclerosis-prone regions of the aorta<sup>126</sup>. This is important for the

investigation of atherogenesis because both SMC and MØ are found in human lesions and the ratio of these cells change as the disease progresses. The fact that both cell types share an activation mechanism means that the presence of MCSF in an experimental sample does not by definition mean that only MØ will be recruited. This simple fact is not evident from the plethora of mouse studies, and is further evidence that multiple models are needed to grasp the complexity of human atherosclerosis, especially at the initiation stage.

Swine are unique among the other mammals depicted in Table 1 because, although they are LDL carriers like the hamster<sup>127</sup>, and most lesions develop in the aortic arch, they also develop spontaneous lesions in the abdominal aorta<sup>128</sup>. The initial foam cells are derived from intimal SMC<sup>128</sup>, and appear similar to those found in early stages of the human disease. Unfortunately, these lesions do not progress to advanced atheromas without being induced by a 4% (w/w) cholesterol diet<sup>92</sup>. Interestingly, even after 90 days on a hyperlipidemic diet, less than 5% of the cells are monocytes<sup>128</sup>. For this reason, it would appear that the pig could adequately model the gradual transition from a myogenic fatty streak to an advanced lesion with activated macrophage cells, mirroring the inflammatory response in humans over time.

Avian models of human atherosclerosis such as the chicken, turkey, quail, and pigeon are not currently in widespread use, but have a longer and richer history than most mammalian models of cardiovascular disease. The first angioplasty surgery of the aortic wall was performed in birds in 1874<sup>129</sup>. Spontaneous (non-induced) atherosclerosis in the chicken was first described in



1914<sup>130</sup>, and it has been repeatedly observed that avian lesions bear close resemblance to their human counterparts<sup>131-134</sup>. The pigeon (*Columba livia*) is especially suited for genetic studies of atherosclerosis because susceptible and resistant strains exist in the natural population<sup>132,135</sup>, eliminating the need to construct an artificial phenotype through genetic or dietary manipulation. In fact, it has been suggested that the White Carneau (WC) pigeon may be one of the most appropriate models of early human lesions<sup>92,133,136</sup>.

The WC pigeon is unique among non-primate models in that it develops naturally occurring (spontaneous) atherosclerosis at both the celiac bifurcation of the aorta and in the coronary arteries<sup>131,137</sup>. Foam cells develop into fatty streaks which progress into mature plaques in the absence of elevated plasma cholesterol and other traditional risk factors<sup>138,139</sup>. These non-induced atherosclerotic lesions are morphologically and ultrastructurally similar to those seen in humans<sup>4,140</sup> and occur at parallel anatomical sites along the arterial tree<sup>4,133,141</sup>. A variety of studies have clearly demonstrated that susceptibility in the WC resides at the level of the arterial wall<sup>138,142,143</sup>, and the lesion site specificity, severity, and disease progression as a function of age are highly predictable<sup>144,145</sup>.

Show Racer pigeons (SR) are resistant to atherosclerosis, while consuming the same cholesterol-free diet. This difference occurs despite the fact that both WC and SR have similar plasma cholesterol and lipoprotein concentrations<sup>146</sup>. WC pigeons are one of the few animal models to develop severe atherosclerosis while consuming a cholesterol free diet, and using the

SR as an experimental control enables pathological changes associated with the disease to be distinguished from changes due to the natural aging process in the pigeon. Virtually all differences between the WC and SR are at the level of arterial tissue as there are few system level differences<sup>147</sup>.

In the wild, both breeds of pigeons are hypercholesterolemic compared to humans, and, like mice and rabbits, they are primarily HDL carriers. However, pigeons are unique in that for the first three days of life, cholesterol is circulated in the form of LDL, after which time the lipoprotein profile switches to HDL<sup>148</sup> for the remainder of the pigeon's life. Neither breed has apoE<sup>149</sup> or LDLR<sup>143,150</sup>, so the effect of these variables in other models of the human disease is not a factor in the pigeon pathology. Combined data gathered on several hundred birds aged 6 months to 3 years over a twenty-year period shows that the average plasma cholesterol concentration in pigeons ranges from 201 mg/dL in the SR to 242 mg/dL in the WC (+/- 16 in both groups)<sup>151</sup>. Although these values are significantly different ( $p < 0.05$ ), they do not change during the progression of the disease, and it appears blood cholesterol does not induce foam cell development in the WC. This fact is further supported in wild mourning doves, a close relative of the pigeon. These birds have average plasma cholesterol levels of 258 mg/dL, and they do not get atherosclerosis<sup>152</sup>. Sterol balance studies have revealed that the WC excretes less neutral sterols than the SR breed<sup>153,154</sup>, but this seems to impact the differences observed in the diet-induced form of atherosclerosis to a much greater degree than it does

the susceptible phenotype of the WC to the naturally occurring form of the disease<sup>155</sup>.

The most widely studied spontaneous atherosclerotic lesion in susceptible pigeons occurs at the celiac bifurcation of the aorta, and by three years of age reaches a size to be easily visible on gross examination<sup>144,156</sup>. Early pathological and metabolic changes are apparent microscopically in this site by six months of age<sup>145</sup>. In contrast, diet-induced lesions in the WC aorta occur at various and unpredictable sites along the descending<sup>138,157,158</sup> and abdominal aortas, and are pathologically very different from non-induced lesions. Foam cells in spontaneous lesions consist primarily of modified SMC<sup>135,145</sup> while cholesterol-induced foam cells are mostly composed of macrophage cells<sup>135,158-160</sup>.

As with mice, diet-induced lesions develop more rapidly in the pigeon than their spontaneous counterparts<sup>18,160</sup>, but different mechanisms of atherogenesis appear to be involved<sup>135,144</sup>. One of the primary effects of diet induction is to shift the physiological HDL lipoprotein profile to that of primarily LDL<sup>161,162</sup>. In fact, diet supplementation with 1% cholesterol causes such a rapid onset of atherosclerotic foam cells in both breeds that it becomes virtually impossible to detect the influence of intrinsic factors<sup>163</sup> contributing to either susceptibility in the WC or resistance in the SR. Therefore, in order to identify candidate genes that predispose the WC to atherosclerosis, it is necessary to use the spontaneous pigeon model, as the introduction of an artificial diet confounds the interpretation of the earliest events occurring in atherogenesis.

Although this may seem like a novel idea in this era of cholesterol-mania, Kauniz warned thirty years ago that “the preoccupation with studying the cholesterol induced lesion may be retarding progress in the search for the true initiating factor in atherosclerosis”<sup>164</sup>.

Since 1959, many studies have been performed to systematically characterize the initiating factor in lesion development in the susceptible WC pigeon. However, the mechanism(s) leading to foam cell development in the WC is not known, and no studies have been conducted in the spontaneous model to identify the gene(s) or gene product(s) that are specific to initiation. Clarkson and Prichard observed that age and heredity were the biggest factors in atherosclerotic susceptibility<sup>131</sup>. Diet, exercise, and gender were not primary factors in the WC pathology. The effects of age and heredity were further distinguished in 1963, when Goodman and Herndon demonstrated that genetics play a larger role in lesion development than the normal aging process<sup>165</sup>. The authors hypothesized that inheritance was a polygenic trait. In 1973, Wagner et al. compared susceptibility to lesion development between the WC and SR celiac bifurcation of the aorta<sup>142</sup>. The authors found a greater number of advanced lesions in the WC than in the SR, and concluded that the genetic control conferring susceptibility or resistance in the pigeon appeared to be at the level of the artery. Further experiments by the Wagner group showed that blood cholesterol, triacylglycerol, and glucose levels were not different between the two breeds<sup>138</sup>, and that blood pressure is actually a consequence of pigeon atherosclerosis, rather than being an initiating factor<sup>139</sup>. The latter

study provided one of the first indications that although diet is not the primary factor contributing to atherosclerotic susceptibility in the pigeon, it can impact the severity of a lesion once formed; thus indicating a role in progression.

In addition to investigating the effect of risk factors on lesion development, various studies have identified metabolic differences between the arterial wall of WC and SR pigeons. *In vivo*, differences in the WC susceptible foci include increased glycosaminoglycans, especially chondroitin-6-sulfate<sup>166</sup>, greater lipid content, predominantly in the form of cholesterol esters<sup>156,167</sup>, lower oxidative metabolism<sup>168,169</sup>, relative hypoxia<sup>170</sup>, decreased acid cholesterol ester hydrolase<sup>171</sup> and neutral cholesterol ester hydrolase<sup>172</sup> activities, increased glycolysis<sup>173</sup>, decreased tricarboxylic acid cycle activity<sup>173</sup>, and the increased synthesis of prostaglandin E2, which also decreased cholesterol ester hydrolase activity<sup>174</sup>. Although these studies did not distinguish the primary or underlying problem from those that are secondary effects, increases in non-esterified fatty acids (NEFA) and in chondroitin-6-sulfate (C6S) seem to precede many of the other observed differences.

The role of excess NEFA and C6S in pigeon atherogenesis is not yet clear, although the presence of C6S in the susceptible pigeon by six weeks of age does lend support to the response to retention hypothesis. Both human and pigeon smooth muscle cells synthesize C6S as part of the ECM<sup>175,176</sup>, where it has been observed to form complexes with plasma LDL entering the vascular wall<sup>21,177-179</sup>. The question remains as to why there is more C6S in the WC aorta than in the SR aorta in the first place. The same question can be asked

regarding the presence of more NEFA. Although the excess NEFA could be a result of deficient mitochondrial oxidation, the influence of NEFA on lesion development is not understood. Clearly, more studies are needed determine what is happening at the earliest stages of pathology in the WC pigeon.

In humans, atherosclerosis is considered to be a multifactorial disease, with many genes and environmental factors contributing to the specific phenotype and ultimate endpoint. In pigeons, where individual lifestyle choices and environments are not factors, the numbers and types of genes contributing to baseline susceptibility and resistance may be easier to elucidate. Preliminary crossbreeding studies indicated a polygenic mechanism of inheritance<sup>165</sup> with resistance being the dominant trait. However, the authors noted that each breed responded differently to the effects of dietary manipulation<sup>132</sup>, so it is possible that the number of genetic differences observed may have reflected the confounding influence of diet, rather than the spontaneous expression profile.

Pigeons are not as well suited for traditional inheritance studies as hamsters, mice, and rabbits because the birds mate for life, and do not reach sexual maturity until seven months of age<sup>180</sup>. Although excess cholesterol esters can be detected biochemically at 12 weeks, three years are required in order to definitively characterize the complete atherosclerotic phenotype. However, the pigeon genome is approximately half the size of its counterpart in mammalian models, and comparative genomic studies are facilitated by the recent publication of the chicken (*Gallus gallus*) genome, which is similar in size<sup>181</sup> to that of the pigeon.

In 2001, Smith, Smith, and Taylor completed a 15-year cross breeding study at the University of New Hampshire. From the examination of grossly visible lesions (or lack thereof) at three years of age in the celiac foci of susceptible WC, resistant SR, and in F<sub>1</sub>, F<sub>2</sub>, and backcross progeny, the authors determined that susceptibility to spontaneous atherosclerosis in the pigeon was inherited as a single gene autosomal recessive trait<sup>182</sup>. This finding is in direct contrast to earlier results indicating a polygenic mechanism, but the researchers at UNH carried the experiments all the way through the backcross generation, where Herndon et al. only investigated the F<sub>1</sub> progeny. In addition, and probably of greater importance to the experimental results, all of the pigeons consumed the same cholesterol-free diet. Parallel investigation of the smooth muscle cells cultured from several tissues of the WC, SR, and F<sub>1</sub> pigeons demonstrated that lipid accumulation observed at the celiac bifurcation is a constitutive property of WC<sup>182</sup>.

The finding that spontaneous atherosclerosis in the susceptible WC appears to be the result of a single gene, and not the net result of many interacting genes, as is thought to be the case in humans, makes the pigeon model a simplest case system. Identification of the gene responsible for predisposition, and an understanding of how this gene influences the described metabolic and morphological changes could reveal an initiating mechanistic pathway that remains undetected in more confounded models of atherogenesis.

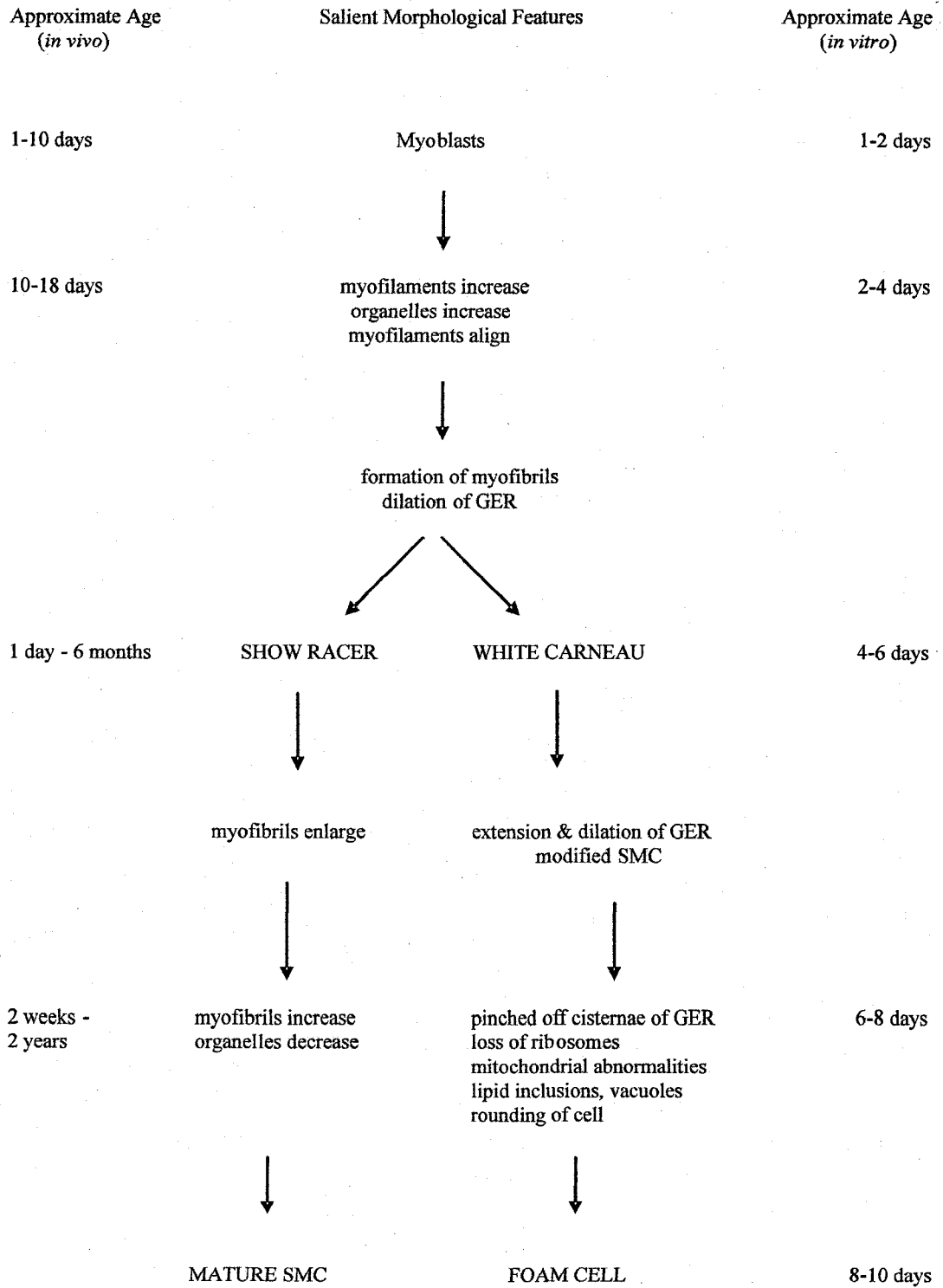
It has been demonstrated that the SMC in monolayers grown *in vitro* accumulate lipid and synthesize proteoglycans in the same manner as aortic

cells *in vivo*<sup>145,177,183,184</sup>, but at an accelerated rate. A comparison of the maturation and degeneration of pigeon aortic cells *in vivo* and *in vitro* is presented in Figure 1 (p.42). In culture, foam cell development is evident in WC SMC by 8-10 days, where several weeks are needed in order to observe the same phenomena *in-vivo*. Other differences in the WC SMC include more esterified cholesterol present in lipid vacuoles, less arachidonate, and decreased mitochondrial metabolism. Although the act of culturing aortic cells can change the SMC phenotype from contractile to synthetic<sup>65</sup>, this has not been observed in primary cultures, where the lack of sub-culture minimizes these and other types of potential genetic alterations. In addition, WC aortic cells obtained *in vitro* demonstrate a similar degenerative progression as those cells observed from the celiac bifurcation<sup>145,184</sup>, further evidence that the gene expression profile is comparable between the two model systems.

*In vitro*, there is no signal communication between SMC and endothelial cells, monocytes, hormones, neurotransmitters, other humoral factors, and whole body feedback systems<sup>51,67</sup>. The only sources of interaction are between the SMCs and SMC with the media components, resulting in cell growth and the synthesis of the ECM. This makes it possible to observe the intrinsic characteristics of WC and SR aortic cell development in a controlled, time-compressed setting, while limiting the number of genes under investigation to those specific to aortic SMC. Interestingly, although only the SMC of the WC celiac and coronary bifurcations are susceptible to atherogenesis *in vivo*, SMC taken from other WC tissue such as the gizzard or small intestine will exhibit



**Figure 1: Maturation and Degeneration of Pigeon Aortic Cells**



features similar to atherogenesis in aortic cells *in vitro*. This is not the case in the SR, where neither SMC from the celiac foci, nor SMC from any other tissue undergo phenotype modification when cultured under identical conditions.

The aforementioned experiments provide additional evidence that the genetic defect predisposing the WC to atherosclerosis is conditionally expressed in SMC. Factors that stimulate the expression of atherogenic genes at the celiac bifurcation *in vivo* appear to be present *in vitro*, as the cultured WC cells undergo degeneration parallel to their counterparts in aortic tissue<sup>145,184</sup>. In addition, the genetic factors denoting resistance in the SR remain expressed in both experimental environments. Therefore, comparing gene expression between the two breeds *in vitro* may identify the single gene responsible for atherogenesis in the WC. Future experiments to directly test the response of the causative gene to extrinsic (dietary manipulations) and intrinsic (whole body) stimuli in terms of spontaneous atherogenesis in the pigeon could then be designed.

Themes, Gaps, & Inconsistencies in Current Knowledge of Human Atherogenesis. Phenomena that occur during the progression of atherosclerosis cannot by default be used to deduce the initiating mechanism. Many studies reporting metabolic or genetic differences in atherogenesis in humans and animal models are actually reporting the occurrences of progression. The merging of progressive and initiative factors in the literature is partly because it is more convenient to study pathological changes as they

occur, but also because of the broad and varied definitions of atherogenesis. The medical field works directly with the prevention and treatment of myocardial infarction and stroke, and, therefore, tends to view atherogenesis as anything that is observed prior to a clinical event. However, multiple mechanisms can cause varied paths to a similar phenotype<sup>1,23</sup>, and each may warrant a unique treatment strategy. In order to sort out the chronology of events in different individuals, a more precise definition of atherogenesis is needed.

Foam cell development is the earliest common event across all animal models and in humans, and it would be helpful if atherogenesis referred only to the specific mechanisms leading up to and including this event. Once a fatty streak has formed, whether myogenic or macrophage based, all subsequent events, whether progression, regression, or inert, should be characterized and reported in terms that are specific to the type and location of the initial foam cell. Smooth muscle and macrophage cells appear to play different pathogenic roles in different animal models, and in different forms of the human disease. SMC are more prevalent in the abdominal aortic branches of spontaneous models, and in the abdominal and coronary atheromas found in humans. Macrophage cells predominate in animal models where foam cell development is induced by an artificial diet and/or genetic manipulation. In the human, this type of foam cell is also associated with extreme hypercholesterolemia, often in the thoracic aorta, and tends to be secondary to other pathological conditions. Macrophage cells appear in the progressive stages of atherosclerosis in all models, which

further underscores the need to clearly articulate the precise pathological step under study.

The confusion between types of cells involved in early atherosclerotic events may be explained by the fact that a non-specific antibody was often used to positively distinguish a macrophage cell from a SMC<sup>45,61,185,186</sup>. Alpha actin is the major actin isoform in a differentiated SMC, and the lopsided emphasis of monocytes in foam cell development may be because any cell that did not stain for this marker was considered to be of non-SMC origin<sup>187,188</sup>. The predominant actin isoform in the SMC switches from alpha to beta/gamma (BG) as part of the phenotypic modulation observed during the initial stages of atherosclerosis<sup>14,45,61,67,188,189</sup>. The alpha/BG actin ratio in a contractile SMC is 2:1 whereas after the cell becomes synthetic, the alpha/BG ratio is 2:7<sup>14</sup>. Because alpha actin is not expressed in a phenotypically modified SMC, it is not a reliable marker to use in distinguishing SM from macrophage cells. Therefore, many researchers may have falsely concluded that SMC were not a major cell type in the development of atherosclerosis in their model.

The literature is also confounded in terms of spontaneous and diet-induced atherosclerosis. Many authors are beginning to report spontaneous atherosclerosis in transgenic models<sup>20,40,190</sup> which calls the true meaning of spontaneous into question. Any type of experimental manipulation required to induce the desired phenotype cannot be considered spontaneous. This is because additive effects are complicating the true role of both hypercholesterolemia and inflammation in atherogenesis, and are preventing

the identification of intrinsic factors within a given individual model. Induced models (transgenic or diet) can be used to look at the interactions of genes known to be involved with atherosclerosis, and to investigate the synergistic effects of these genes with identified risk factors in a hypothesis driven fashion. However, in order to describe unknown facets of the disease, spontaneous, non-induced models are needed. Genetic factors influencing susceptibility and resistance in the absence of confounding factors remain to be identified in order to explain the prevalence of atherosclerosis in individuals that remain at risk under current screening methods.

### **Investigating the Genetic Differences between Two Similar Populations**

There are many strategies and techniques available to investigate the differences in genetic expression between susceptible and resistant individuals within a defined population. It is important to appreciate that most of these methods are not competitive, in that one is any better than another. Rather, the varied approaches provide complementary information that can be utilized in assembling the bigger picture of human atherogenesis. Experimental design must consider not only the specifics of the research question being asked, but also what resources exist for the selected organism.

If the genome has been published and well annotated, more options are available. The orthologous relationship between highly conserved functional proteins among species allows the identification of an unknown nucleotide

sequence based on its similarity to something that has already been characterized. However, if the genome of interest has not been characterized, a method must be selected that does not depend on prior knowledge of the nucleotide sequence or karyotype. An experimental technique that requires a *priori* knowledge is considered to be a closed system, and is well suited for hypothesis driven experiments when the gene of interest has been characterized<sup>191</sup>. In contrast, open systems are those that do not depend on previous knowledge, and so allow for the discovery of novel genes and expression events<sup>62</sup>.

Traditionally, molecular biologists have focused on open systems such as subtractive hybridization (SH) and differential display (DD) in order to identify restriction fragment length polymorphisms (RFLP)<sup>192</sup>. With advances in the polymerase chain reaction (PCR), sequencing and microarray technology, many eukaryotic genomes have been characterized, and there are valuable closed systems in which to test a multitude of hypotheses. When information about an organism is complete, a closed system is preferable to an open system because all of the potential interactions are traceable. However, there is still a lot that is not known, and an over emphasis on closed systems can actually prevent the discovery of new knowledge. In fact, most of the gene expression studies described to date have “merely validated differential expression of genes and pathways known to be involved in atherosclerosis, and have yet to fully exploit the power and possibility of identifying novel players (and eventually novel pathways) underlying atherosclerosis”<sup>193</sup>.

New genes are identified by either one of two ways: computer prediction based on the identification of open reading frames (ORF) or the isolation and functional analysis of expressed sequence tags (EST) in a traditional wet lab<sup>194</sup>. Despite the fact that the human genome was published in 2001<sup>195,196</sup>, and the mouse and rat genomes soon followed<sup>197,198</sup>, researchers are still finding transcripts that have not been predicted by current computer models. A number of Serial Analysis of Gene Expression (SAGE) experiments continue to isolate genuine transcripts that are not in the EST databases<sup>194</sup>. The fact that new genes are still being discovered provides strong rationale for employing discovery-based experiments in order to identify novel candidate genes involved in the pathogenesis of atherosclerosis.

The most common source of genetic variation between individuals is single nucleotide polymorphisms (SNP). Any given mutation will only be present in a small percentage of the population, and the associated risk will be clinically important only in some subjects<sup>199</sup>. However, the effect of the nucleotide sequence may be quite large in those populations that are susceptible, and it is imperative that large genetic screenings are conducted along with precise phenotype descriptions in order to determine the true impact of a polymorphic allele in the general population<sup>27,99,200</sup>. Even when the allele frequency is known for a sample population, it still may not correlate with disease. Many times a polymorphism will randomly associate with susceptible and resistant strains, and these events can be hard to distinguish from those that are actually causative<sup>99,201-203</sup>. The definitive assignment of a SNP to a phenotype is

complicated by the fact that even when carrying the deleterious allele, some individuals will only develop the associated disease in conjunction with an environmental trigger<sup>27</sup>.

In addition to using SNPs to correlate genotype and phenotype within a defined population, another source of genetic variation is that of copy number (CNV)<sup>204</sup>. CNV are structural variations at the chromosome level that can be quantified by measuring their effects. CNV affect both the allele dosage and the transcript copy number and they are widely prevalent throughout the genome. Distinct genomic regions have particular copy number variations and these could be a source of variation between species as well as between susceptible and resistant individuals within a single species<sup>204</sup>. The mechanism and purpose of CNV are not yet understood in healthy individuals, so it is too early to use them to predict disease. However, it is expected that CNV will help to explain how similar genes come to develop new or altered functions, and they will be used in conjunction with SNP to generate novel phenotypic markers throughout the genome.

Alternative splicing is an important posttranscriptional mechanism that is known to generate diverse proteins from a single transcript<sup>205</sup>. It is now believed that a substantial fraction of human genes (35-60%) produce transcripts that are differentially spliced<sup>206</sup>, to provide a third source of genetic variation between individuals. Furthermore, 70-90% of these alternative-splicing events alter the resulting protein product<sup>206</sup>. Almost all ion-channel and neurotransmitter receptor pre-mRNA undergo extensive alternative splicing to



generate several protein isoforms. Differential splicing may account for the localization of gene expression to a particular tissue and /or developmental stage<sup>191</sup>, but has also been implicated in some types of genetic diseases<sup>98,207</sup>.

Sometimes mutations in regulatory regions of the genome can cause the altered expression of many genes, thus appearing to be a multifactorial trait, yet still inherited in a classic Mendelian fashion<sup>208,209</sup>. Alternative splicing is one example of this type of regulation, as is chromatin remodeling, transcription initiation, RNA elongation and editing, and instructions for protein degradation<sup>191,209</sup>. Each level is composed of a “dynamic system of self-organizing proteins, the output of which is governed by laws that are still poorly understood”<sup>210</sup>. Careful analysis of many nucleotide sequences in many organisms has revealed that the distribution of exons, promoters, gene start sites, and other genomic features is highly variable, and functions to regulate overall genetic expression<sup>211</sup>.

Many genes may be common between organisms, especially those of the same breed. Sequence and functional similarity is what gives comparative genomics its strength<sup>212</sup>. However, when trying to isolate differentially expressed genes from closely related species, it may be more efficient to look at those genes expressed only in the tissue of interest rather than the overall genomic content. Genomic DNA is more stable than messenger RNA (mRNA); however, it is not expressed at all times in all cells. In contrast, mRNA provides a snapshot of transcriptional activities at the moment of cell lysis<sup>213</sup>. Therefore, RNA expression profiling has the potential to “identify genes involved in disease

pathogenesis resulting from both environmental and genetic factors, along with associated interactions."<sup>214</sup>

It is important to keep in mind that the amount of mRNA present does not necessarily correlate with the amount of protein made<sup>215</sup> for many reasons. There are varying degrees of translation efficiency, and different messages have different half-lives. For example, the mRNA for c-fos has a half-life of just a few minutes, whereas beta actin mRNA is stable for several hours<sup>191</sup>. The specifics of mRNA degradation are considered to be an important aspect of genetic regulation, but many of the details of this pathway have not yet been elucidated. Once the primary protein sequence is translated, there are a multitude of post-translational modifications (PTM). Sulfation, phosphorylation, glycosylation and proteolytic cleavage are just a few directed changes that provide the cell with flexibility, and guide the peptide to its functional form. Most proteins that will be secreted from the cell or embedded in the plasma membrane will be glycosylated<sup>191</sup>. The genetic signals for specific PTM events are not known and it is not always possible to predict how each protein will be modified after it is translated based on the nucleic acid or amino acid sequence. Therefore, although general transcription rates do correspond to general rates of protein synthesis, it is important not to make quantitative comparisons between the two types of molecules unless they are being analyzed in parallel.

Although targeting the transcriptome will not necessarily identify the regulatory regions in non-coding DNA<sup>208</sup>, this strategy significantly decreases the complexity of the starting material, as only around 2% of the genome

actively codes for protein<sup>216</sup>. The specific profile of these mRNA transcripts constitutes a fingerprint, or signature of the phenotype. Theoretically, after the genes composing that fingerprint are characterized, the upstream regulatory events can be deduced from the unique signature.

As the genes involved in atherogenesis are presumed to be different than those genes responsible for progression, changes in gene expression will occur as the pathological events progress and as a consequence of normal development and cellular differentiation of the vasculature<sup>217</sup>. The interpretive challenge is not to confuse the genes that caused the foam cell to appear at discrete loci in the first place, with those being expressed as a secondary or downstream effect. It is therefore necessary to look at entire pathways instead of just the individual players<sup>218</sup> as investigating the effects of only one gene at a time can provide fragmented knowledge<sup>219</sup> that is potentially misleading<sup>200</sup>. The goal of cardiovascular genomics is to establish the genetic baseline of susceptible and resistant phenotypes. By interrogating the system as a whole, as is possible with a discovery based approach, gene-gene interactions and overall profile patterns that are consistent with either a susceptible or resistant phenotype can be detected<sup>99,220</sup>. Subsequent perturbations can then be imposed in order to observe the influence of genotypic variation on initiation, progression, and regression of human atherosclerosis in a hypothesis-directed approach.

There are basically three types of biochemical approaches to investigate the expression levels of differential transcripts<sup>221</sup>. (1) Hybridization experiments take advantage of the chemistry of Watson-Crick base pairing in order to isolate transcripts of both known and unknown (random primers) nucleotide sequences. (2) PCR methods depend on the controlled and exponential amplification of the target transcript to generate enough product to be analyzed, and (3) sequence based methods use capillary gel electrophoresis to determine the identity and order of the nucleotides in a sample based on their densitometric difference, although novel sequencing methods are being developed as the race for the \$1000 genome continues<sup>222</sup>. Newer techniques are able to employ a combination of these three primary approaches to maximize the output from the varying amounts provided by biological samples.

The primary methods used to investigate the transcriptome are presented in Table 2 (p.54). Eight methods are presented, and categorized in terms of the following criteria: open or closed system; hybridization (H), PCR (P), or sequence (S) based technique; the ability to detect alternative splicing; the possibility of absolute quantification; and throughput capacity<sup>191,221,223</sup>. Sample throughput refers to the number of different biological samples that can be analyzed in one experiment, whereas transcript throughput represents the number of differentially expressed transcripts that can be analyzed after completing one experiment<sup>221</sup>. The classification of high, medium, and low throughput is not quantitative, as they are simply generalizations of the relative amounts of information when compared to the other methods.

**Table 2: Comparison of Strategies to Investigate the Transcriptome**

	SAGE	CHIP	QPCR	RNAi	DD	SH	RDA	NOR
System Type	Open	Closed	Closed	Closed	Open	Open	Open	Closed
Sample Throughput	Low	Med	High	Low	Low	Low	Low	Low
Transcript Throughput	High	High	Low	Low	High	High	High	Low
Detect Alternate Splicing?	No	No*	Yes	No	Yes	Yes	Yes	Yes
Technical Basis	S	H	P	H	P	H	H, P, S	H
Absolute Quantification?	Yes	No	Yes	No	No	No	No	No

**Legend:** SAGE = Serial Analysis of Gene Expression; CHIP=Microarray Analysis; qPCR=Quantitative Polymerase Chain Reaction; RNAi=RNA Interference; DD=Differential Display; SH=Subtractive Hybridization; RDA=Representational Difference Analysis; NOR=Northern Blot Analysis

\*Microarray technology is only as specific as its composition. If individual exons are plated on the slide, alternative splicing can be detected<sup>224</sup>.

Representational Difference Analysis (RDA) is a positive selection technique that couples the specificity of subtractive hybridization with the sensitivity and amplification power of PCR to isolate differentially expressed transcripts. RDA draws from the strengths of all three primary methods of investigating the transcriptome, and as an open system, provides the opportunity for discovering genes not previously implicated in atherosclerosis. It is considered to be a positive enrichment strategy because it selects for what is different between two samples, rather than degrades what is the same, as is the case with negative enrichment designs.

RDA was originally developed to find subtle differences between whole genomes<sup>225</sup>, and has successfully identified disease markers in pedigree dogs<sup>226</sup> and strain differences between palm trees<sup>227</sup>. The technique has been modified to use cDNA as a starting template<sup>228</sup>, and as such, is capable of picking up variations in gene expression between tissues or individuals. RDA

provides an advantage over SH and SAGE by reducing the bias of abundant transcripts<sup>217</sup>. It allows more opportunities to capture rarely expressed transcripts<sup>229,230</sup>, and variations in copy number, as well as the presence/absence of a message from a candidate gene. RDA can detect differences in less than one copy per cell<sup>223,231</sup>, which is important considering that anywhere from 80-95% of transcripts are present in less than 5 copies per cell<sup>191,223,232</sup>.

There have been many modifications to the original cDNA RDA protocol in order to minimize the isolation of false positives and accommodate smaller amounts of clinical tissue<sup>233-235</sup>. In general, there are three steps common to all RDA protocols: representation; PCR coupled subtractive hybridization; and screening of the difference products. Methods used to verify and characterize transcripts vary by researcher, including northern blots, virtual southern blots, microarrays and qPCR<sup>62,234,236,237</sup>. Northern blots and microarray analysis are not always sensitive enough to detect rarely expressed messages that are captured during the enrichment process of RDA<sup>62,223</sup> so preliminary sequence analysis can be performed in order to identify the top candidate among the difference products<sup>238,239</sup>. Although the copy number variation detected in RDA cannot be considered absolute because of the multiple PCR steps and the variable number of clones actually picked and analyzed, general trends in regulation have been consistently confirmed using qPCR. Upregulated transcripts identified in an open system such RDA can then be plated on a microarray chip<sup>240,241</sup> to monitor changes in gene expression under varying

environmental conditions, and/or used to design fluorescent probes and primers in order to provide absolute quantitative results with qPCR.

### **Research Objectives**

1. To isolate genes that are differentially expressed between cultured aortic smooth muscle cells (SMC) from White Carneau (WC) and Show Racer (SR) pigeons by Representational Difference Analysis (RDA).
2. To identify the differentially expressed genes by comparative genomics (Bioinformatics).
3. To correlate the function of differentially expressed genes with mechanisms for initiation of spontaneous atherosclerosis in the susceptible WC pigeon.

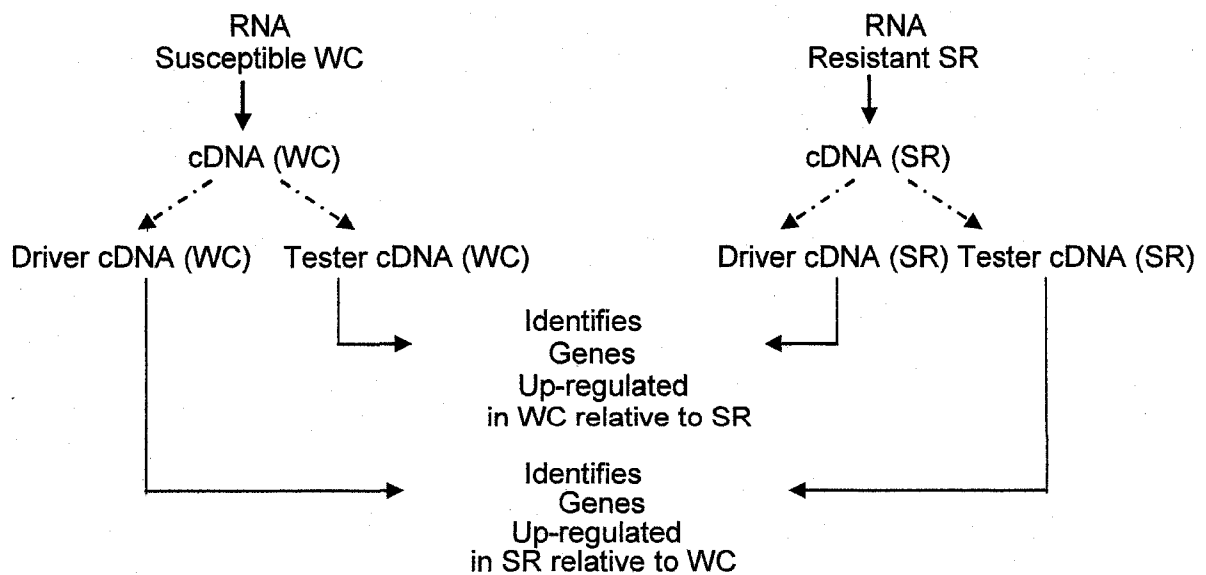
### **Hypotheses**

1. The gene(s) responsible for atherosclerotic resistance in cultured SR SMC are silent or down regulated in the WC SMC.
2. The gene(s) contributing to the spontaneous atherosclerotic phenotype in cultured WC SMC are not expressed in SR SMC.

## Experimental Design

1. Representational Difference Analyses (RDA) was performed to compare gene expression in SMC cultured from susceptible WC and resistant SR aortas.
2. The reciprocal RDA experiment (diagrammed in Figure 2) was done in quadruplicate on several pools of WC and SR cells. A list of candidate genes for susceptibility/resistance to pigeon atherogenesis was compiled.
3. The functional role of genes that are differentially expressed between the two breeds was deduced.

Figure 2: RDA Experiment:





## CHAPTER II

### METHODS AND MATERIALS

#### Pigeon Colonies

White Carneau (WC) and Show Racer (SR) pigeons were obtained from the UNH colonies, which are housed in fly coops at ambient temperature and allowed free access to water, Purina Pigeon Chow Checkers, and Kaytee Red Grit. The colonies were established in 1962 with birds obtained from the Palmetto Pigeon Plant in Sumter, South Carolina and have remained closed flocks. The colonies are maintained under the supervision of the UNH Animal Care and Use Committee (Approval #050601).

#### Cell Culture System

One to three day-old WC and SR pigeons were used to prepare primary cultures of aortic smooth muscle cells according to the method developed in this laboratory<sup>183</sup>. One culture preparation of 18-25 flasks requires 3-5 pigeon aortas. Thirty-two explants were planted in each 24cm<sup>2</sup> flask and grown for 7 days at which time flasks that exhibited confluent monolayers were washed three times with Hanks Balanced Salt Solution (HBSS) to remove the growth media. Cells were serially harvested in 1ml Trizol™ (*Invitrogen*) for RNA extraction. Briefly, cells from the first flask were scraped into Trizol™ and

poured into the second flask. This process was continued until all of the cells from 8-10 flasks were concentrated into one 1.7mL microcentrifuge tube. A second 1mL Trizol™ was used to serially harvest the remaining flasks, and the extractions were not pooled until total RNA was assessed.

### Preparation of Total RNA

An RNA extraction procedure must allow for complete cellular disruption while effectively inhibiting the ribonucleases (RNAses) that are released upon membrane lysis<sup>242</sup>. Trizol™ is an effective acid-phenol-guanidium-thiocyanate (APGT) RNA extraction reagent<sup>237</sup>. Guanidium thiocyanate is a chaotropic salt that denatures protein:nucleic acid complexes while simultaneously distorting the tertiary structure of RNAses, rendering them inactive<sup>242,243</sup>. Once the cellular RNA is exposed and protected from endonuclease degradation, the acid-phenol component of Trizol™ reacts with chloroform (0.2mL per 1.0 mL Trizol™) and partitions the DNA, protein, and RNA according to density gradient and pH. The RNA partitions in the aqueous phase with the guanidium thiocyanate and is recovered by precipitating overnight with isopropanol. There are many versions of APGT lysis buffers commercially available, and Trizol™ is routinely used in this laboratory to extract total RNA from both aortic tissue (1mL: 70 mg) and cultured aortic cells (1 mL: 8-10 flasks/50-65mg).

Following the overnight incubation in isopropanol, total RNA was sedimented by centrifugation at 16,000xg for 18 minutes (*Eppendorf Model #5415D*), and the resultant "pellet" was washed in 70% ethanol to remove the

guanidium thiocyanate. Each pellet was dissolved in 125uL of molecular grade water (18.2 megOhm), and mixed with 12.5uL 3M sodium acetate (pH 5.3) and 313uL of ice-cold 100% ethanol to initiate a second purification step. The sample was vortexed for 10 seconds, incubated at 4°C for 30 minutes, centrifuged for 18 minutes as before, and the pellet was washed in 70% ethanol. The supernatant was removed using a vacuum with a 20uL-filtered pipet tip, and the rim of each tube was wicked dry against a clean paper towel. The purified pellets of total RNA were dissolved in 20uL water.

The concentration of total RNA was estimated using the optical density (OD) of the sample at the wavelength of 260nm (A260) in a 3mM sodium phosphate buffer, pH 7.5<sup>244,245</sup>. Absorbance values must be within the range of 0.1-1.0 in order to provide accurate information about the amount of RNA in solution. The OD was also measured in 10nm intervals from 220-320nm to determine the presence of contaminants that interfere with the absorbance of light by the sample at 260nm. Sources of these contaminants include components of the biological starting material that were not completely removed from the RNA extract such as protein (280nm) and polysaccharide (230nm) in addition to residuals from the extraction process itself including phenol (270nm) and salts (240nm)<sup>213,245,246</sup>.

Although spectrophotometric quantitation is not absolute, checking the relative absorbance of these common contaminants can increase the reliability and, thus, usability of the A260 reading for calculating the yield of RNA. In addition, impurities in the RNA preparation can inhibit many of the downstream

manipulations required for RDA such as reverse transcription and the polymerase chain reaction. Polysaccharides are a significant component of pigeon aortic tissue, and are routinely detected in RNA extracts prior to the second purification step<sup>247</sup>. The 230/260 ratio should be  $< 0.5$ <sup>248</sup>, and, in general, the ratio of "good" absorbance values at 230:260:280 approximate 1:2:1<sup>246</sup>. Although a 260/280 ratio of  $2.0 \pm 0.15$ <sup>213</sup> is often used to assign quality to an RNA sample, it is at best a gross estimate of protein contamination<sup>249</sup> and was reported, but not used as the primary indicator of RNA purity.

Spectrophotometry provides an estimate of the amount of nucleic acid extracted, but because all nucleic acids absorb light at 260nm, further analysis is required to determine whether the sample is in fact RNA, and not genomic DNA, and to determine whether the product is intact or degraded. Therefore, both spectrophotometric and electrophoretic analyses of the sample were performed before the integrity of the RNA preparation was accepted. Since nucleic acids have a net negative charge, electrophoresis can separate the DNA and various forms of RNA on the bases of the density, mass, and shape of each molecule. These separations correspond to specific sedimentation rates in ultracentrifugation, which are expressed as Svedberg coefficients (S)<sup>213</sup>.

An aliquot (0.5ug) of RNA was run on a 1.2% agarose gel at 85 volts for 25 minutes and stained in an ethidium bromide solution (0.01mg/ml). The predominant 18S and 28S ribosomal RNA (rRNA) bands were visualized under ultraviolet light (302nm) for evidence of degradation and DNA contamination

using the MultiDoc-It Digital Imaging System (*UVP Model M-20*). For the RNA to be considered pure, the 28S band should be about twice the intensity of the 18S band<sup>213</sup>, and both bands should be discrete with no sign of smearing. If the RNA is degraded (smeared), it must be discarded. Residual DNA in the sample can be observed at the origin of the well. Additional treatment with DNase can be performed to remove the genomic DNA, but this was not necessary for any of the RNA preparations used for the RDA experiments. The amount of total RNA available for RDA decreases with each purification step, but potential yield must be sacrificed for a quality template, especially when working with the pigeon, whose genome has not been sequenced. RNA samples determined to be pure by both spectrophotometric and electrophoretic assessments were stored at  $-80^{\circ}\text{C}$  until the next step in the RDA procedure.

### **Preparation of Complementary DNA (cDNA)**

RNA was converted to double stranded (ds) cDNA using the BD SMART<sup>TM</sup> PCR cDNA Synthesis Kit (*BD Biosciences Clontech Division #K1052-1*). This kit utilizes a special primer set to capture both the polyadenylated tail at the 3' end of the mature transcript, and the guanidiated cap at the 5' end. The primer extension option allows the researcher to produce full-length cDNA<sup>250</sup> from as little as 50ng total RNA, thus eliminating the need to enrich for messenger RNA (mRNA). Because it can take from 100-500ug total RNA to isolate 1ug mRNA<sup>213,234,251</sup> depending on the method, the

SMART™ technology employed during first strand synthesis significantly decreased the number of required cell culture preparations, and, therefore, the number of pigeons, needed to complete the proposed RDA experiments.

The SMART™ kit provides two options for generating cDNA, depending on the downstream applications of the product. The first protocol is specific for library construction, and the second protocol generates cDNA for other purposes. Although the second option was selected, both procedures depend on patented SMART™ (Switching Mechanism At 5' end of RNA Transcript) oligonucleotides and a lock-docking primer to capture the polyadenylated tail from the 3' end of the RNA transcript and add a deoxycytidine "anchor" to the 5' end. This SMART™ anchor is then used as an extended template for PowerScript RT™, a modified Moloney murine leukemia virus (MMLV) reverse transcriptase. In theory, only full-length clones will be produced during the first strand synthesis reaction because transcripts lacking the 5' anchor will not serve as a template for the second strand.

The SMART™ protocol was optimized to use 1ug of total RNA to generate double-stranded cDNA for the proposed RDA experiments. The first strand reaction produces unstable RNA/cDNA hybrid molecules that can be stored up to three months. This template consistently yields 12-15ug double-stranded cDNA, and the kit provides reagents for 25-second strand reactions.

Second strand synthesis reactions were performed in the MasterCycler Gradient Thermal Cycler (*Eppendorf # 5331; 115V*) according the SMART™ PCR cDNA Synthesis Kit User Manual (*Version No. PR37276*). The potential

for non-specific hybridization product is increased with every round of PCR<sup>213,249,252</sup>. Therefore, the number of PCR cycles was optimized at 23; the minimum number of amplification cycles before the amount of generated cDNA began to plateau. PCR products were extracted first with 100ul phenol-chloroform-isoamyl alcohol (PCI), then with 100ul chloroform. The aqueous phase containing the cDNA was mixed with 0.1 volume (10ul) 3M sodium acetate (pH 5.3), 1ul glycogen (10mg/ml, *Sigma #G1767*), and 2.5 volumes (250ul) of 100% molecular grade ethanol, vortexed briefly, and placed at  $-20^{\circ}\text{C}$ .

Following overnight incubation cDNA was sedimented by centrifugation at 16,000xg for 18 minutes. The pellet was washed in 70% ethanol to remove the sodium acetate and spun for an additional 2 minutes. The supernatant was removed using a vacuum with a 20uL-filtered pipet tip, and the rim of each tube was wicked dry against a clean paper towel. The purified cDNA was dissolved in 25uL of 18.2 mgOhm water, and yield was determined using spectrophotometry (A260). An aliquot (0.5ug) of cDNA was separated on a 1.2% surface tension agarose gel at 85V for 20 minutes and photographed under UV light using the MultiDoc-It Digital Imaging System. Samples to be compared using RDA must demonstrate identical size distribution<sup>234</sup>, otherwise differential products will simply reflect variations in the molecular weight of converted RNA transcripts rather than their relative expression levels. Ethidium bromide staining (0.01mg/ml) should reveal a general smear from about 300bp-7000bp, with specific intensity visible from 2000-3000bp<sup>234</sup>. Purified, analyzed cDNA was stored at  $-20^{\circ}\text{C}$  until used for RDA.

## **Representational Difference Analysis (RDA)**

As described in the Literature Review, RDA is a multi-step positive enrichment technique that couples PCR with competitive hybridization to select and amplify the transcripts that are differentially expressed between two populations of RNA. A detailed protocol for RDA was generously provided to our laboratory by Pastorian et al<sup>234</sup> and was systematically modified as needed to accommodate both the small amount of RNA produced by cultured aortic cells and the inevitable sample loss incurred during multiple chemical and physical manipulations<sup>235</sup>. Therefore, the procedural steps described in this section reflect the optimized protocol developed during the first successful RDA experiment and subsequently followed for all replicates.

### **Amplicon Preparation**

Five micrograms of cDNA prepared from WC and SR cultured aortic cells were matched for RDA by size distribution and digested overnight at 37°C with 44.4 units of Dpn II (10U/ul; *New England Biolabs*), a four-cutter restriction enzyme that generates cDNA fragments with a 5'GATC overhang, or "sticky end" on each strand. Following enzymatic restriction, the cDNA was extracted with PCI and chloroform as previously described, and the aqueous phase was run through Sephacryl® 300 HR Columns (*GE Healthcare/ Amersham #27-5130-01*) to remove both the restriction reagents and the cut fragments less than 100bp. The size-selected cDNA was precipitated for a minimum of three



hours (sometimes overnight) with glycogen, sodium acetate, and ethanol as described elsewhere and dissolved in 20ul Tris-EDTA (TE) buffer (pH 8.0). At this point, the cDNA population was an average size of 256bp, ranging from 100-1000bp, and ready for adapter ligation and amplification. These cDNA “amplicons” were stored no longer than two days at  $-20^{\circ}\text{C}$ .

### Primer Information

The interchangeable primer sets required for adapter ligation and PCR amplification are presented in Table 3 and were purchased from Genosys, a subdivision of Sigma, as recommended by Pastorian<sup>234</sup>.

Table 3: RDA Primer Sets

	<b>Primer Set</b>	<b>12mer (5' – 3')</b>	<b>24mer (5'- 3')</b>
<b>A</b>	Driver	GATCAATAACTA	TGACGGACCGGTTGCGTAGTTATT
<b>B</b>	Tester Round 1	GATCTTATGGCT	AGACAGTGCCGGATGTAGCCATAA
<b>C</b>	Tester Round 2	GATCCTATTGAC	ATCTCAGGGGACCTGAGGCAATAG
<b>D</b>	Tester Round 3	GATCCAGATGTA	ATACGTGCAGGCTGGTTACATCTG
<b>E</b>	Tester Round 4	GATCTTATACTA	TAACCTCGGCCCTCGGTTAGTATAA
<b>F</b>	Tester Round 5	GATCTACGTACT	TCACATCGCCCCCTATAGTACGTA

Lyophilized primers (12mers & 24mers) were dissolved in the appropriate amount of TE buffer as determined by nucleotide composition, optical density, and molecular weight provided by Genosys. The concentration of each re-solubilized primer was verified by spectrophotometric analysis (A260). Paired primer sets (A-F) consisting of 9ug of the 24mer and 4.5ug of the companion 12mer were prepared and stored at  $-20^{\circ}\text{C}$  in ready-to-use aliquots to prevent hydrolysis caused by repeated freeze-thaw steps. The primer sets were used

interchangeably throughout the RDA experiments, and care was taken to use a different 12/24mer set to prepare the Drivers and each round of Tester within a given experiment.

### Driver Preparation

One of the prepared 12/24mer primer sets (Primer Set A) was thawed and mixed with 9ul 10X ligation buffer (*New England Biolabs*) in a total volume of 45ul. Twenty microliters were allocated to the entire volume (20ul) of the cDNA amplicons and subjected to the "SLOWCOOL" program (Appendix B, p.202) in the MasterCycler Thermal Cycler. Briefly, the cDNA was incubated at 55°C for two minutes, allowing the primers to melt while the cDNA template remained intact. The temperature was slowly decreased to 8°C over a 45 minute time period to allow the 12mer and 24mer to form hydrogen bonds with each other (anneal). At the same time, the GATC nucleotide sequence at the 5' end of the 12mer is annealed to the 5' "sticky" end (CTAG) of the Dpn II digested cDNA.

Nine microliters of T4 DNA Ligase (400 Cohesive end units/ul; *New England Biolabs* #M02025) were mixed with 11ul 10X Ligation Buffer (included) and 90ul water. A fifty- microliter aliquot of this mixture was added to each tube containing the prepared cDNA and incubated at 15°C overnight. During this step, the 24mer ligates to the 5' overhang generated by Dpn II. Because ligation only occurs at the 5' ends of the molecule, the 12mer does not ligate at the

temporary bridge sites created at the 3' end of the cDNA in the previous step, and will separate from the cDNA molecule during subsequent manipulations.

Following overnight ligation, the products were incubated at 65°C for 10 minutes to inactivate the ligase and re-melt the unincorporated 12mers. The heated samples were immediately passed through HR S-300 columns as previously described to remove the ligase, buffer, and residual primers. The purified products were divided into six thin-walled tubes, for a total of 12 tubes per experiment. Taq Polymerase (5U/ul) was purchased from Promega (#M1861) and the Drivers were amplified exactly as described in the protocol obtained from Pastorian et al<sup>234</sup>.

Briefly, a PCR master mix of 250ul 10X PCR Buffer, 400ul 25mM magnesium chloride, 62.5ul 10mM dNTPs (*Promega # U151B*), 1487.5ul water, and 18ug (50ul) of the same 24mer used to ligate the adapters to the cDNA was prepared. This mix was divided evenly among the 12 PCR reactions and incubated at 72°C for 2 minutes to fully extend the cDNAs and create primer-binding sites on the 3' ends. One microliter of Taq was added to each tube. The samples were then incubated for 5 more minutes at 72°C and cycled through 25 rounds of PCR each for 45 seconds at 95°C followed by 4 minutes at 72°C<sup>234</sup>. The PCR products were further extended for 10 min at 72°C, pooled, extracted with PCI and chloroform, precipitated as previously described, and dissolved in 100ul TE buffer.

The 24mers that were ligated to the cDNA template prior to PCR were removed by restriction digest by adding 4.5ul of Dpn II (10U/ul) and 10.5ul 10X

enzyme buffer to 87ul of each Driver and incubating the reactions overnight at 37°C to regenerate the 5' sticky ends. The samples were again subjected to phase extraction with PCI and chloroform, spin column purification, and overnight precipitation. The pellets were dissolved in 50ul TE buffer, and 3ul were used to estimate the yield of amplified Drivers at A260. An additional aliquot (0.5ug) was run on a 2% agarose gel<sup>241</sup> for 50 minutes at 75V. The gel was stained with ethidium bromide as previously described, and usually showed smearing from about 150bp –1000bp, with most of the intensity below 700bp. If any cDNA was visible above 1Kb, the Drivers were not completely restricted and were not suitable for Tester preparation. Acceptable Drivers were stored at –20°C and thawed as needed.

Initial attempts to generate Drivers yielded no more than eight micrograms of cDNA. Because a minimum of 50 micrograms are required to perform the subsequent subtractive hybridization steps<sup>234</sup>, four aspects of Driver preparation were optimized for pigeon aortic cells.

Restriction Digest Parameters. The parameters suggested by Pastorian et al<sup>234</sup> were insufficient to fully restrict the pigeon cDNA. The digestion time was therefore increased from four hours to overnight. In addition, the amount of Dpn II was increased from 30 to 44.4 units.

Adapter Ligation Protocol. Prior to adaptor ligation, the RDA 12/24-mer primer sets were melted at 55°C and annealed to the restricted cDNA as the temperature was steadily decreased to 8°C during a “SLOWCOOL” step. During this time, the 12mers form a temporary bridge that allows the 24mers to

ligate to the 5' end of cDNA template when T4 DNA ligase is added. In the original protocol, "SLOWCOOL" was performed by heating an aluminum block containing the reactants to 55°C, and then moving the block to 4°C. Moving the entire reaction to a thermal cycler standardized the protocol. The ramp time is controllable in a thermal cycler, and is so quick that the length of time spent at 55°C was increased from 1 minute to 2 minutes prior to cooling in order for primers to fully melt. The rate of cooling was controlled by reducing the temperature of the thermal cycler two degrees per minute until the reaction reached 8°C.

Phase Extraction Parameters. According to Shanahan et al. 30-50% of the cDNA is lost at each phase extraction step<sup>235</sup>. RDA requires multiple phase extractions throughout the procedure, and in order to decrease the amount of sample lost during phenol and chloroform extraction steps, multiple optimization steps were performed.

First, a "back extraction" step was performed in an attempt to maximize the recovery of the cDNA<sup>249</sup>. This was accomplished by adding TE buffer to dilute the extracted cDNA, and re-extracting the aqueous phase with an equal volume of chloroform. Although slightly increasing the yield, sample was still lost. Second, the overall reaction volumes were increased to dilute the cDNA, thus minimizing the amount lost per microliter sample at the interphase. Finally, Phase Lock Gel (PLG) tubes were employed (*Eppendorf # 955154151*). The PLG tubes include an inert gel that forms a protective barrier between the

aqueous and organic phases during the centrifugation step<sup>253</sup>, allowing the cDNA to be recovered in total.

Precipitation Parameters. The original protocol precipitated the extracted cDNA with ammonium acetate (4M). According to Farrell<sup>213</sup>, the types of salt and alcohol can be varied, as long as the correct ratios are used for a given pair. Sodium acetate (3M) was used successfully during the cDNA synthesis procedures, and the same protocol was applied to the RDA products. One microliter of glycogen (10mg/mL) was also added to the precipitation reaction to assist in visualizing the cDNA pellets. Once the four steps described above (1-4) were optimized, there were no further problems with yield, and I was able to proceed with tester preparation.

#### Tester Preparation

One microgram of each Driver population was used to prepare Testers for the first round of subtractive hybridization. A second set of primers (Primer Set B) was used to perform the "SLOWCOOL" annealing step, and the 24mers were ligated to the 5'ends in a 60ul volume as described during the first step of Driver preparation. The only difference was the primer set selected. This is a small but important distinction because it ensures that that cDNAs with uncut adapters from Primer Set A will not be carried over to the hybridization step. Following overnight ligation at 15°C, the Testers were heated to 65°C for 10 minutes and immediately used in the following procedural step.

### Round One Hybridization

Thirty microliters (0.5ug) WC Tester population was transferred to a new tube. Five micrograms of SR Driver was added, and the volume was adjusted to 100ul with TE buffer. In the reciprocal experiment, 30ul (0.5ug) of SR Tester was combined with 5ug of WC Driver. Both samples were then extracted with 100ul PCI, 100ul chloroform, and precipitated for a minimum of 4 hours (sometimes overnight) at  $-20^{\circ}\text{C}$ .

The cDNA pellets are very loose at this stage, and the supernatant was decanted very carefully. The pellets were washed in 70% ethanol, dried, and dissolved in 4ul of hybridization buffer<sup>234</sup>. For each hybridization reaction, 1ul of 5M sodium chloride was added to a fresh thin walled tube and placed at  $95^{\circ}\text{C}$ . Each cDNA aliquot (4ul) was incubated at  $95^{\circ}\text{C}$  for 1 minute to denature the double stranded Tester and Driver populations. The cDNA was then added to the tubes containing sodium chloride, mixed well, covered with 20-25ul mineral oil (*Sigma #M5904*), and incubated for an additional 3 minutes at  $95^{\circ}\text{C}$ . Without removing the tubes from the thermal cycler, the temperature was decreased to  $67^{\circ}\text{C}$  and the hybridization reactions continued for 24 hours at which time they were diluted with 45ul water and stored at  $-20^{\circ}\text{C}$ .

### Amplification of Round One Hybridization Products

Following hybridization, three types of double-stranded molecules are formed. Because the Driver is added in excess (10:1) of the Tester, most of the cDNA hybrids are Driver: Driver. During hybridization, some of the Tester sequences form hydrogen bonds with the excess Driver sequences, creating

Driver: Tester hybrids of transcripts common to both WC and SR populations.

The smallest pool of molecules consist of Tester: Tester hybrids, which represent the target sequences present only in the Tester population.

Background hybrids that do not represent differentially expressed transcripts will be effectively "drowned out" (subtracted) by the exponential amplification of Tester: Tester sequences during PCR. The PCR parameters used to amplify the subtracted differences are the same as used to generate the Drivers.

Although all three types of cDNA hybrids are present at this point in the RDA procedure, they are each different at their 5' ends and so will react differently during the amplification step. The Drivers were generated with Primer Set A and subsequently digested with Dpn II to remove the adapters. Therefore, Driver:Driver hybrids have no priming site on either 5' overhang, and will not be amplified. The Testers were made with Primer set B, and the ligated adapter makes the tester template accessible for priming by 24mer B. Driver:Tester hybrids only have an adapter on one of the two strands, and will be amplified linearly. Only the Tester:Tester hybrids have a ligated primer site on the 5' end of each strand and so will be amplified exponentially.

To further enrich the final PCR products for target hybrids, an aliquot (1/10 reaction volume, or 10ul) is transferred to four fresh PCR reactions after the first 7 cycles. This dilution step replaces the need to treat the samples with mung bean nuclease by reducing the background hybrids "to such an extent that only exponentially amplifiable cDNA hybrids (Tester:Tester) will contribute



significantly to the final product”<sup>234</sup> after 20 additional rounds of PCR cycling in a total reaction volume of 210ul.

### Purification & Assessment of Round One Difference Products

Following PCR, like samples were pooled and extracted with PCI and chloroform, and precipitated at  $-20^{\circ}\text{C}$  with sodium acetate, glycogen, and ethanol as previously described for at least two hours. The samples were centrifuged at 16,000xg for 18 minutes, washed with 70% ethanol, spun for an additional two minutes, dried, and the purified Tester: Tester cDNA hybrid pellets were dissolved in 100ul TE buffer.

As with the final step in Driver preparation, the ligated 24mer adapters were removed by an overnight restriction digest with Dpn II to regenerate sticky ends. The samples were again extracted with PCI and chloroform, and the aqueous phase was run through the HR S-300 spin columns. The cDNAs were precipitated for at least two hours, centrifuged, washed, dried, dissolved in 50ul TE buffer, and labeled “Difference Product Round One” (DP1).

The yield of the amplified difference products (testers) was determined by measuring the optical density of each DP1 at 260nm, and was usually between 10-20ug as suggested by Pastorian<sup>234</sup>. An aliquot (0.5ug) was separated on a 2% agarose gel for 50 minutes at 75V and stained with ethidium bromide (0.01mg/ml). As with the Drivers, if smearing was visible above 1Kb, the cDNA restriction step was considered incomplete. A few differential bands were sometimes observed, but the real reason for electrophoretic analysis at

this point in the RDA protocol was to confirm the size distribution of the difference products from round one.

### Subsequent Rounds of RDA

The amplified difference products from the first round of subtractive hybridization (DP1) were used to make the Testers for round two. Just as in the first round, 1ug DP1 was mixed with a brand new primer set (C), and the adapters were ligated onto the 5' ends of the subtracted cDNAs overnight at 15°C. However, for each round of RDA the hybridization ratio is increased. Therefore, following a brief heating of the samples to 65°C for 10 minutes, only 50ng of the ligated 24mer C WC Tester was mixed with 5ug of the original SR Driver and vice versa, for a Driver:Tester ratio of 100:1. The cDNA mixture was co-precipitated and hybridized as in round one, and the heterogeneous hybrids were subjected to PCR using 24mer C to prime the amplification. Following the first 7 rounds of PCR, the sample was again diluted in a fresh reaction mix for an additional 20 cycles. Following PCR, the Tester:Tester hybrids were purified and assessed exactly as described in RDA step 7. There were usually a few distinct bands visible at this point in the RDA procedure. Again, 1ug of the difference products from this round (DP2) was used to prepare the Testers for round 3.

To prepare for the third round of subtractive hybridization, Primer Set D was used to ligate 1ng Tester, which was mixed with 5ug of Driver for a Driver: Tester ratio of 5000:1. To accommodate the small amount of Tester available to

form Tester:Tester hybrids, following the first 7 cycles of PCR the reactions were again diluted, but then subjected to 25 additional cycles rather than 20 as with the previous two rounds. The PCR products were purified, analyzed, and labeled "DP3". The difference products were stored at  $-20^{\circ}\text{C}$  for no more than two weeks prior to cloning.

### Cloning the RDA Difference Products

The pBlueScript II SK+ (pBS) Vector (*Stratagene #212205*) was selected for its flexibility because it contains multiple cloning sites as well as complementary binding sites for six different sequencing primers. The Dpn II generated difference products are easily cloned into the *Bam*HI site of the pBS<sup>62,237,241,254</sup> to be replicated as with a traditional plasmid vector.

To prepare the vector to accept the cDNA inserts, 5ug of the circular 3.0Kb vector was digested with 5ul *Bam*HI (5U/ul, Promega) for 1 hr at  $37^{\circ}\text{C}$  in a reaction volume of 50ul. The restricted vector was then dephosphorylated with 20 units of calf intestine alkaline phosphatase (*CIAP, Stratagene # 600015*), increasing the total volume of the reaction to 100ul. The solution was incubated at  $37^{\circ}\text{C}$  for 30 minutes, followed by an additional 15 minutes at  $68^{\circ}\text{C}$ . The product was extracted with PCI and chloroform, precipitated for a minimum of 1 hr at  $-20^{\circ}\text{C}$ , centrifuged, washed, and dried as previously described, and subsequently dissolved in 5ul TE. One microliter was used to determine the

optical density at 260nm, and 0.5ug was run on a 1.2% surface tension gel at 85V for 25 minutes.

The digested, linearized vector was visible as a single band at 3.0Kb when compared against a 10kb DNA ladder (*Fermentas #SM0311*). The prepared vector was stored at -20°C and diluted to 0.1ug/ul immediately prior to the cloning step. The difference products were also diluted to 0.1ug/ul just prior to cloning, and if stored at -20°C for longer than two weeks, an additional restriction digest with Dpn II was performed to ensure the sticky ends were still available for vector insertion.

Ligation parameters were adjusted and analyzed as suggested by the Stratagene protocol to verify the efficacy of the dephosphorylation step and to establish the optimal insert: vector ratio. It was determined that a 1:2 ratio resulted in the most efficient transformation; consequently, 0.5ul of cDNA insert (0.1ug/ul) was combined with 1.0ul (0.1ug/ul) prepared vector and ligated overnight at 4°C with 0.5ul T4 DNA Ligase (6U/ul) in a reaction volume of 10ul.

The pBS vectors with the ligated cDNA inserts were then transformed into chemically competent XLI-blue MRF- cells (*Stratagene #200230*) exactly as instructed by the product manual. The advantage of using this particular host cell is that the F- episome allows for visual identification of the bacterial cells that contain the target insert. Cells that were successfully transformed will form white colonies, and those that were not transformed will be blue. Two microliters of the ligation reaction were added to 100ul of the thawed competent cells, gently stirred with the pipet tip, and incubated on ice for 30 minutes. The

samples were “heat-shocked” by placing them in a 42°C water bath for exactly 45 seconds. This “transformation” was then diluted with 900ul of warm (42°C) SOC buffer<sup>249</sup> and shaken at 230rpm for one hour at 37°C.

An aliquot (40ul) of each transformation was diluted with 160ul warm SOC and plated on LB/agar bacterial growth plates containing 100ug/ml ampicillin (*Teknova* #A9510), 80ug/ml fresh 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside (X-gal) (*Promega*, #V3941), and 30mM isopropyl- β-D-thiogalactopyranosidase (IPTG) (*Promega*, #V3955). The dilution was allowed to seep into the agar for 15-20 minutes at room temperature at which time the plates were inverted and incubated for 18-19 hours at 37°C. To enhance the blue/white color selection, the plates then were cooled at 4°C for two hours. The remaining transformation was mixed with 333ul of 60% glycerol (*Biochemika* #49767) and stored at -80°C for subsequent plating if needed.

One hundred ninety-two white colony-forming units (cfu) from each transformation were hand picked with sterile toothpicks into 1ml 96-well plates (*VWR* # 40002-011) containing 400ul yeast-tryptone bacterial growth media (2x YT<sup>249</sup>), and 0.4ul ampicillin (50mg/ml) per well. The plates were sealed with Air-Pore tape (*Rainin* # 96-SP-STR) to allow respiration, and the clones were shaken at 180rpm for 18 hours at 37°C. The following morning, 133ul of 60% glycerol was added to each well, and the plates were sealed, vortexed, and stored at -80°C.

## Sequencing of Cloned Difference Products

Rolling Circle Amplification (RCA) was used to purify the pBlueScript SK+ vector from the XLI-blue MRF- host cells and to prepare the plasmids to be sequenced. The TempliPhi 500 Amplification Kit (*GE Healthcare/Amersham Biosciences #25-6400-50*) replaced the more traditional method of alkaline lysis to purify plasmids. In addition to requiring overnight growth of the bacterial hosts and many centrifugation steps, alkaline lysis does not produce a consistent concentration of purified plasmid across samples<sup>255</sup>. By normalizing the incubation time of the RCA reactions, sample concentrations are equivalent which is advantageous for obtaining reproducible sequence results by capillary electrophoresis<sup>255,256</sup>.

RCA reactions were performed in 96-well PCR plates (*Eppendorf # 951020303*), and the protocol was modified to use half-reaction volumes, thus doubling the number of plasmids that can be purified using the RCA kit. Briefly, a 0.5ul aliquot from each well of the glycerol stock plates generated in step F was transferred to a clean PCR plate using a multi-channel pipet. This template was mixed with 2.5ul of sample buffer, covered with strip caps, and “pulse” spun for 20 seconds (0-3867rpm) in the Beckman Coulter Allegra 25R centrifuge to collect the reactions at the bottom of the plate. The samples were denatured at 95°C for 3 minutes to release the plasmids for subsequent amplification, and cooled to 4°C. A master mix of 250ul reaction buffer and 10ul TempliPhi enzyme mix was prepared, and 2.5ul of this was added to each

denatured sample. The plates were pulse spun to collect the total volume at the bottom of each well, and the entire plate was incubated at 30°C for 16 hours.

Following the overnight isothermal amplification of the plasmids, the plates were heated for 10 minutes at 65°C, cooled to 4°C, and pulse spun to collect the condensation. The samples were diluted with 10ul molecular biology grade water and stored at -20°C until restriction digest and sequencing were complete. Restriction analysis was performed on each prepared plasmid to determine whether the plasmid was indeed released from the bacterial host, and if the isolated plasmid contained the DP insert. Although blue/white selection increases the confidence that each clone contains an insert, it is important to verify this before investing time and money in the sequencing phase of analysis. A double restriction digest was performed using *HindIII* and *XbaI*. As described previously, the pBS vector contains multiple cloning sites, and the *BamHI* site was selected to clone the RDA fragments. There is a *HindIII* restriction site (TCTAGA) 24 bp upstream from the *BamHI* site, and an *XbaI* (AAGCTT) site just 8 bp downstream. Therefore, these two enzymes were selected to cut the insert out of the plasmid for electrophoretic assessment. A master mix was prepared containing 55ul *HindIII* (10U/ul), 55ul *XbaI* (10U/ul), 11ul BSA (*Promega*), 110ul Multi Core Buffer (*Promega*), and 319 ul water. Five microliters of the restriction reaction was added to 5ul of diluted RCA product in a fresh 96-well PCR plate and incubated at 37°C for 2hours. The entire volume was mixed with 4ul gel loading buffer<sup>249</sup> and a 7ul aliquot of this mixture was run

on a 1.2% agarose gel for 30 minutes at 85V. The gel was stained with ethidium bromide (0.01ug/ml) and visualized under UV light as previously described.

For the sample to be accepted for sequencing, at least three regions of intensity should be visible on the gel. The chromosomal DNA from the bacterial host can be seen at the origin of the well, and the purified plasmid from the RCA reaction is observed as a distinct band at 3000bp. A third band represents the portion of the DP insert that was cut out of the plasmid during the restriction digest. For high-throughput purposes, if at least 88 of the 96 plasmids contained inserts, the entire plate was submitted for sequencing. Otherwise, additional clones were prepared as described in order to submit a full 96-well plate.

The final round of difference products (DP3) from the first two reciprocal RDA experiments were cloned into the XL1-Blue MRF- Vector and grown under the selection of ampicillin, X-gal, and IPTG as described. The third RDA experiment was performed in the same manner as the first two. However, the ratio between blue and white colonies was vastly different. In the first two experiments, the number of blue and white cfu was fairly evenly distributed. In the third experiment, there were 17,750 white cfu and only 3500 blue cfu for the WC. This trend was repeated with the SR, in that 22,250 white cfu grew and only 6750 blue. Rather than indicating an extremely efficient transformation, restriction digest following RCA revealed that many of the white colonies picked did NOT contain an insert. It is possible that during the cloning step, the F-episome can fall off. If that happens, the colony will remain white, even though there is no cDNA insert in the vector. Because so many of the white colonies



picked did not contain an insert, this appears to have happened in the third experiment. To reduce the number of these false positives, the Stratagene manual suggests a double antibiotic selection. Tetracycline selects specifically for the F- episome, so growing the colonies on media supplemented with both ampicillin and tetracycline, further selects for vectors containing inserts. Once double selection was implemented, misleading white colonies were no longer a problem.

The Hubbard Center for Genome Studies (HCGS) at the University of New Hampshire performed the sequencing of differentially expressed RDA transcripts on a service basis using the CEQ 8000 Genetic Analysis Capillary System (*Beckman*). The M13 Reverse primer (5' GGAAACAGCTATGACCATG 3') was selected to amplify the target inserts from the pBS vector. The results were provided in a digital file that included the individual sequences and quality scores in "fasta" format, along with the associated chromatograms.

### **Sequence Analysis and Characterization of RDA Difference Products**

It is challenging to identify differentially expressed genes from an uncharacterized genome and to establish biological relevance of the significant players. When confronted with 384 raw sequences per RDA experiment, it is important to stay focused on the research objectives. Each individual base pair position is scored, and any nucleotide can be manually eliminated or substituted in order to produce a more refined query for subsequent identification steps.

This process can be subjective if strict criteria are not established, and it is easy to get distracted with sequence analysis for its own sake. Therefore, sequence trimming and curation was kept to a minimum, and the few steps that were taken are described in the following paragraphs.

### Raw Sequence

The length of each sequence was recorded as the number of nucleotide bases. If no nucleotides were sequenced from a given clone (# bp = 0), the raw sequence was REJECTED. All remaining sequences were analyzed for vector contamination in step B.

### Vector Trimming

The raw sequences were screened for vector contamination using the "VecTrim" program available through the National Center for Bioinformatics (NCBI). This is necessary because the target sequence amplified with the M13R primer will include the flanking vector sequences. Once the vector nucleotides were identified, the first and last base pair position of the target transcript was recorded, and the adjusted sequence size calculated. For example, if the raw sequence was 748 bp in length, and the "VecTrim" program identified two regions of vector nucleotides (nt 3-75; nt 273-677), the target transcript is represented from nt 76-272, and the adjusted length would be 197bp. This truncated sequence is the "query" for step C. If the entire sequence was identified as vector (no insert), the sequence was REJECTED.

## BLASTn

The vector-trimmed sequences were queried against the NCBI non-redundant nucleotide database (BLASTn) in order to identify them by inferred sequence homology to other species<sup>257</sup>. Subject matches were accepted as statistically significant if the Expect Value (E) was  $\leq e-10$ . Orthologous sequences that met this criterion were considered to be "Transcript Tag Orthologs" and were further annotated in step G. Queries that resulted in no subject match, or a subject match with an E of  $>e-10$  were analyzed for sequence quality in step D. In practice, the sequence quality was evaluated while waiting for the BLASTn output screen. It may seem like the sequences should be evaluated for quality earlier in the analytical process, but in fact, the vector nucleotides tended to have higher quality scores than the target transcript. Analyzing the trimmed sequences prevented the acceptance of a sequence based on the vector region rather than the target.

## Sequence Quality

The CEQ 8000 Genetic Analysis system provides a modified version of standard Phred quality scores in fasta format, called Call Scores (*Beckman Coulter, A-1940A*). Each base pair "call" in a sequence is given a confidence rating on a scale of 0-100. Call Scores above 90 are generally reserved for nucleotides that have been inserted, substituted, or otherwise manually edited (*CEQ manual*). The fasta files containing the quality Call Scores were easily viewed using the Microsoft Notepad program, but in order to view the Standard

Chromatogram Format (SCF) file associated with each sequence, a special viewer (Finch TV) was downloaded from Geospiza. Finch TV (Version 1.4.0 ©2004-2006) was developed with an NIH SBIR grant (R44HG02244-02) and is distributed freely from the company's website, <http://www.geospiza.com/finchtv/>.

Evaluation of quality grids along with the chromatograms of validated sequences indicated that if "most" of the call scores were above 40, the corresponding peaks were clearly distinguished from each other. When the Call Score dipped below 40, the peaks began to have "shoulders" and the number of overlapping peaks slowly increased. These observations were too subjective to quantify. Therefore, in order to establish a consistently routine procedure, the following criteria were developed to accept or reject the quality of the query region of the sequence. If there were no Call Scores above 31, the entire sequence was REJECTED. If there were some (more than one) Call Scores above 37, the sequence was ACCEPTED.

Although crude, these cut-off values served as usable guidelines for accepting or rejecting a sequence for subsequent analysis. Query sequences that were of acceptable quality but had no hit in the BLASTn database search were further characterized in step E. Although the decision was made to not work with any sequences of borderline quality for this project, the rejected sequences were saved for future manipulations.

## TBLASTx

There is a chance that transcript tags not identified using BLASTn can be identified using the tBLASTx search tool, also available through the NCBI website<sup>258</sup>. Because of the degeneracy of the genetic code, the amino acid sequence of a protein may be conserved even though the nucleotide sequence may have diverged to the point that it is no longer homologous to the ancestral sequence. BLASTn aligns nucleotides of the query with nucleotides of subjects in the database, but tBLASTx translates all six reading frames of the nucleotide sequences of both the query and the database entries to their amino acid sequence in order to identify potential orthologs.

The E values of the tBLASTx output were more difficult to interpret than those of BLASTn because the queried sequences are one-third the size. The E decreases with sequence length, and if the same criteria used to accept BLASTn hits were used to accept tBLASTx hits, it would be possible to reject a potential match that may be biologically relevant. Therefore, the tBLASTx alignment results were scored from 0-3 based on the following criteria. One point was assigned for an E value of  $\leq e^{-10}$ . One point was assigned if the percent identity was over 50%, and the alignment was at least 12 amino acids in length. One point was assigned if there was a six amino acid region that aligned with no gaps or substitutions. Alignments receiving a score of 2-3 were categorized as "Transcript Tag Orthologs" whereas queries that resulted with no match, or received an alignment score of 0-1 were labeled "Unidentified Transcript Tags", and further characterized in step F. Unidentified transcripts

represented query sequences of acceptable quality that were not identified in the NCBI database with either the BLASTn or the tBLASTx search tool.

### Unidentified Difference Products

The unidentified transcripts from each RDA experiment were merged into one large fasta file for two types of batch analysis. The sequences were first uploaded to the HCGS BLAST Server and queried against the most recent version of the chicken genome (*May, 2006 Version 2.2*). Although no significant matches were expected because of the lack of introns in cDNA compared to genomic DNA, this step was done to determine if any of the pigeon transcripts did match to a specific chicken chromosome. The unidentified transcripts were then compared to one another for sequence homology using the Sequencher Software (Demo Version 4.7 Build 2946, ©1991-2006 Gene Codes Corporation). Assembly parameters were set for a minimum of 85% identity and a minimum overlap of 20 bases. Sequencher performed 8,382 comparisons and clustered the sequences into contigs and singletons. After categorizing the unidentified transcripts by this method, they were named and counted for statistical analysis as described (p93). Because the unidentified transcripts were in fact, differentially expressed, it was important to characterize them as completely as possible. Therefore, each sequence was translated into its six potential reading frames using the Expert Protein Analysis System (ExpASy) available at <http://www.expasy.org> and maintained by the Swiss Institute of Bioinformatics<sup>259</sup>. The frame with the longest amino acid sequence that did not

include a stop codon (TAG, TAA, TGG) was selected as the most likely open reading frame for that nucleotide sequence.

In a final attempt to characterize all transcripts as completely as possible, the contigs, singletons, and expressed sequence tags (ESTs) were BLASTed against the Tentative Consensus (TC) sequences in the Chicken Gene Index Database, currently maintained by the Dana Faber Cancer Institute (DFCI)<sup>260,261</sup>. The TC identification number was recorded for any sequence alignment that met the criterion of an E value of  $\leq e-10$ . Transcripts that did not have a TC number remained "unidentified". Transcripts with a TC number were analyzed for sequence homology to known genes. The tentative annotation assigned to each TC is scored as follows: 90-100% identity is considered a "homologue to", 70-90% identity is "similar to", and anything below 70% is "weakly similar to"<sup>260</sup>. For the purposes of this project, sequence similarity  $\geq 80\%$  was considered annotated. Fifteen transcripts that were "unidentified" in GenBank were identified using the DFCI Chicken Gene Indices and included in the overall analysis.

#### Annotation of Difference Products

Transcript orthologs identified using BLASTn, tBLASTx, and the Gene Indices were further analyzed in order to place the transcript tags into gene families or metabolic pathways that contribute to the susceptibility and/or resistance to atherosclerosis in the pigeon. A gene family is a loose, non-binding classification that relates genes that share important characteristics to

each other. For example, genes can be grouped into families according to structure, size, function, expression patterns, location, etc. and can therefore be placed in multiple categories. For the purposes of this project, it was determined that that most informative way to organize the genetic transcripts would be by cellular function. Two distinct databases were used to annotate those transcripts with known protein products. Sequences that were not translated to a specific protein, such as those matching ESTs and other clones from the Chicken Genome Project were not annotated using either software package because there was not enough descriptive information available.

The first software system used to annotate the RDA difference products was the Kyoto Encyclopedia of Genes & Genomes (KEGG) Database. This compilation of regulatory and metabolic pathways was first established in 1995 as part of the Japanese Human Genome Program, and has been continually updated as new genomes are sequenced and experimental functional data becomes available worldwide<sup>262</sup>. Version 41.1, released February 1, 2007 was freely accessed at <http://www.genome.jp/kegg>. Each molecular entry in KEGG has been assigned a unique identifier called the "KO" number and placed within a functional hierarchy based on gene orthology.

The first level of the KEGG hierarchy provides the most basic description of biological function. KEGG Level One segregates entries into five major groups: Metabolism (01100), Genetic Information Processing (01200), Environmental Information Processing (01300), Cellular Processes (01400), and Human Diseases (01500). If a single molecule participates in more than



one pathway or functional group, each unique association was mapped so as not to limit the biological interpretation of its role. Enolase is an example of a protein that fits into more than one KEGG Level One category. As an enzyme, enolase is part of the glycolytic pathway of carbohydrate metabolism. Enolase is also capable of translocation to the nucleus and acting as a transcription factor<sup>263</sup>. Using the KEGG system, enolase would be assigned both to the Metabolism Group (01100) and the Genetic Information Processing Group (01200) for Level One annotation.

The five KEGG categories of Level One are further delineated in KEGG Level Two, where general pathways specific to each functional group are assigned another unique identifier. For example, Genetic Information Processing (01200) is divided into four sub-categories: Transcription (01210), Translation (01220), Replication & Repair (01240), and Folding, Sorting, & Degradation (01230). Metabolism (01100) is divided into carbohydrate metabolism (01110), energy metabolism (01120), lipid metabolism (01130), etc. Again, if a single molecule participates in more than one pathway, its unique KO number will cross reference to each appropriate match.

KEGG Level Three divides the general pathways of Level Two into even more specific functions within that pathway. In the enolase example, carbohydrate metabolism (01110) is the Level Two designation, whereas glycolysis (00010) represents only one specific aspect of carbohydrate metabolism. Therefore, glycolysis is assigned a unique KO number at Level Three to distinguish it from other aspects of carbohydrate metabolism such as

the pentose phosphate shunt (00030). Up regulated transcripts from each pigeon breed were categorized by all three levels of the KEGG orthology system. Obviously, if an identified transcript was not represented by a KO, it was not included in the KEGG Results.

In an attempt to look at the functional relationship between the differential transcripts from as many angles as possible, a second software package was used to annotate the RDA products. Pathway Studio is a commercially available package that contains "more than 100,000 events of regulation, interaction and modification between proteins, cell processes, and small molecules"<sup>264</sup> and has an automated connection to PubMed. The 30-day trial version of Pathway Studio (Version 4.0; Ariadne Genomics, Inc. ©2002-2006) was used to complement and extend the functional information obtained from KEGG. In addition to biological process and protein "groups", Pathway Studio also provides specifics on the cellular compartment where the functional protein is known to be active. Upregulated transcripts that were represented in the database were characterized by all three systems and reported in conjunction with the KEGG results in order to present the most complete profile of differential gene expression between the White Carneau and Show Racer Pigeon that can be determined at this time.

#### Assignment of Pigeon Transcripts to the Chicken Physical Map

The average avian genome is 1.43 Gbp, ranging from 0.97 Gbp in the common pheasant to 2.16 Gbp in the ostrich<sup>265</sup>. The typical avian karyotype

includes several pairs of macro-chromosomes, and numerous micro-chromosomes<sup>266</sup>. The chicken has a haploid size of 39, whereas the pigeon haploid number is 38. Although the recombination rate of birds is three times that of mammals,<sup>267</sup> most of these occurrences are on the same chromosome, rather than alleles crossing over to a different chromosome<sup>266</sup>. The chicken genome displays fewer evolutionary rearrangements than the mouse, and it is expected that genome organization and synteny are highly conserved between avian species<sup>181,268</sup>.

In 2004, Derjusheva et al compared four species of birds using chromosome painting<sup>266</sup>. Chromosomes #1-10, Z, and 9 micro-chromosomes from Galliforme (chicken) were compared with the counterparts in Columbiforme (pigeon) and two Passerines (redwing and chaffinch). No more than two inter-chromosome rearrangements were observed in each species, and the authors concluded that all four species "have retained the chromosomal integrity of ancestral synteny". More recently, 8 macrosomes and 14 microsomes of the chicken and the quail (*coturnix japonica*) were compared using fluorescent in-situ hybridization (FISH). All of the chicken clones selected hybridized to the appropriate quail chromosome, as 49 macrosome hybrids and 20 microsome hybrids were observed<sup>269</sup>.

Although the microsomes were difficult to distinguish from one another, the authors of the FISH study noted a high degree of stability between avian karyotypes, further supporting the conclusions reached in the chromosome painting study, that intra-chromosomal arrangements are very rare across

various species of birds. Because of the conserved synteny of the avian genome, transcripts identified in the pigeon were compared to their chicken counterparts and assigned to the corresponding chromosome. Not all of the chicken chromosomes have been mapped, but as of April 2007, 15,368 genes had been placed on chromosomes 1-28, W, Z, and within the mitochondria<sup>270</sup>.

### **Statistical Analysis of Differential Expression**

The Chi Square Statistical Test for Similarity was used to demonstrate that the transcripts isolated using RDA were not simply a collection of random artifacts, but were in fact, differentially expressed. Chi Square is an appropriate test to analyze the distribution of categorical data, and has been shown to have the “best power and robustness” to detect significant changes in ESTs<sup>271</sup> such as those generated by SAGE<sup>272</sup> experiments. Therefore, once a dataset of differentially expressed (identified and unidentified) “Transcript Tags” was generated for each experiment, the results were analyzed in a pairwise contingency table to determine the significance of distribution within each individual experiment, and within the combined totals of all four experiments.

The calculations were performed using the number of times a specific transcript was expressed (Express) compared to the number of times it was not (No Express). The hypothesis (null) was that the expression level of each transcript would be identical between breeds. Ideally the total number of occurrences (SUM = express + no express) would be the total number of

clones. However, sequences that were REJECTED during previous analytical steps were not included in the total. Therefore, the SUM is different for each experiment. Acceptable statistical significance of differential expression was set at  $p < 0.05$ . All p-values were recorded in a Microsoft Excel Worksheet and sorted according to pigeon breed.

Chi Square Analysis was also used to determine the significance of the distribution of genes and copy number variation across the individual chromosomes of the chicken genome. No statistics were performed at the pathway level, because there were multiple cases where the expression of a given pathway was not different between two breeds, but the specific genes expressed therein were different.

### **Determination of False Positives**

Differentially expressed transcripts were considered to be false positives if the transcripts were expressed in both breeds at low copy number and/or were found to be up regulated in each breed in different experiments. A transcript was not a candidate gene for susceptibility/resistance to atherosclerosis if its expression was not consistent across the replicate experiments. False positives were not included in any of the analyses, because they were not different between the WC and the SR.

## CHAPTER III

### RESULTS

The raw sequence data from the four reciprocal RDA experiments are presented in Table 4 (p.96) along with the identification rate of the quality-accepted, vector-trimmed sequences derived from BLAST analysis of difference products.

A total of 1344 clones representing 672 difference products from each breed were sequenced and analyzed. Of the 880 transcripts that were identified, 165 represented unique (non-redundant) tags. Seventy-four of these unique transcripts were up regulated in the WC, and 63 were up regulated in the SR. Twenty-eight transcripts were found to be present in both breeds in varying copy number, and 18 of these were determined to be false positives.

The compiled list of difference products isolated in the four reciprocal RDA experiments and the corresponding genes are presented in Table 5a (WC), Table 5b (SR) and Table 5c (WC & SR). Copy number and the statistical significance of each transcript as determined by Chi Square Analysis are also included in Tables 5a-5c (p.97-105), along with the identification of the RDA replicate number in which each genetic transcript was present.

Table 4: Raw Sequence Data and BLAST Results

	Pigeon Clones		Raw Sequence Data				Valid, Minimal Quality Sequences BLASTed				
	Breed	Plate	Total #bp	Mean length (#bp+/- SEM)	# Accepted	%	Mean Query (#bp+/- SEM)	# Tags ID	%	# Unidentified tags	%
RDA Rep #1	WC	A	65,831	686 +/- 25	53		390 +/- 22	47		6	
		B	58,372	608 +/- 16	38		419 +/- 26	30		8	
		<b>Total:</b>	<b>124,203</b>		<b>91</b>	<b>47%</b>		<b>77</b>	<b>85%</b>	<b>14</b>	<b>15%</b>
RDA Rep #1	SR	A	58,587	610 +/- 32	61		388 +/- 24	58		3	
		B	64,303	670 +/- 18	65		358 +/- 21	59		6	
		<b>Total:</b>	<b>122,890</b>		<b>126</b>	<b>66%</b>		<b>117</b>	<b>93%</b>	<b>9</b>	<b>7%</b>
RDA Rep #2	WC	A	67,126	699 +/- 25	61		400 +/- 19	59		2	
		B	60,320	628 +/- 19	70		342 +/- 18	66		4	
		<b>Total:</b>	<b>127,446</b>		<b>131</b>	<b>68%</b>		<b>125</b>	<b>95%</b>	<b>6</b>	<b>5%</b>
RDA Rep #2	SR	A	73,126	762 +/- 24	74		357 +/- 21	66		8	
		B	63,966	666 +/- 19	77		309 +/- 18	72		5	
		<b>Total:</b>	<b>137,092</b>		<b>151</b>	<b>79%</b>		<b>138</b>	<b>91%</b>	<b>13</b>	<b>9%</b>
RDA Rep #3	WC	A	57,980	604 +/- 27	61		376 +/- 20	60		1	
		B	67,073	699 +/- 14	80		370 +/- 17	78		2	
		<b>Total:</b>	<b>125,053</b>		<b>141</b>	<b>73%</b>		<b>138</b>	<b>98%</b>	<b>3</b>	<b>3%</b>
RDA Rep #3	SR	A	66,132	689 +/- 22	74		351 +/- 18	66		8	
		B	65,442	682 +/- 31	66		367 +/- 21	60		6	
		<b>Total:</b>	<b>131,574</b>		<b>140</b>	<b>73%</b>		<b>126</b>	<b>90%</b>	<b>14</b>	<b>10%</b>
RDA Rep #4	WC	A	67,716	705 +/- 11	86		358 +/- 18	78		8	
		<b>Total:</b>	<b>67,716</b>		<b>86</b>	<b>90%</b>		<b>78</b>	<b>91%</b>	<b>8</b>	<b>9%</b>
	SR	A	69,124	720 +/- 19	84		320 +/- 16	81		3	
	<b>Total:</b>	<b>69,124</b>		<b>84</b>	<b>88%</b>		<b>81</b>	<b>96%</b>	<b>3</b>	<b>4%</b>	
Totals		<b>1344</b>	<b>905,098</b>		<b>950</b>	<b>71%</b>		<b>880</b>	<b>93%</b>	<b>70</b>	<b>7%</b>

Table 5a: Differentially Expressed Genes Exclusive to White Carneau

Upregulated Transcript Ortholog	Gene	Replication Information		# Copies WC:SR	Chi Square p=
		# Reps (4)	Rep # ID		
<b>Chi Square is Significant (p&lt;0.05)</b>					
Cleavage & Polyadenylation Specific Factor	CPSF2	2:0	3,4	16:0	0.0000
Enolase (alpha)	ENO1	2:0	1,2	30:0	0.0000
Retinol Binding Protein 7	RBP7	2:0	3,4	15:0	0.0000
Ribophorin I	RPN1	2:0	1,2	16:0	0.0000
BAC CH261-124L19	?	3:0	2,3,4	14:0	0.0001
N-acetyltransferase 13	NAT13	2:0	3,4	13:0	0.0001
Ribosomal Protein L27 (mitochondrial)	L27mt	2:0	1,2	13:0	0.0001
Proteosome 26S ATPase Subunit 2	PSMC2	2:0	1,2	11:0	0.0004
Contig 1	?	2:0	1,2	10:0	0.0006
Diacylglycerol O-acetyltransferase	DGAT2	2:0	3,4	10:0	0.0008
Dachshund Homolog 1	DACH1	2:0	3,4	9:0	0.0015
Ligatin	LGTN	2:0	3,4	8:0	0.0027
Annexin 2	ANXA2	1:0	2	7:0	0.0050
Ribosomal Protein L32	RPL32	3:0	1,3,4	6:0	0.0094
Transketolase	TKT	1:0	1	6:0	0.0094
Pro-alpha 2 Collagen Type 1	COL1A2	1:0	1	5:0	0.0179
BAC CH261-138K4	?	2:0	3,4	4:0	0.0343
TNFa Induced Protein 8	TNFaIP8	2:0	3,4	4:0	0.0343
<b>Chi Square is Not Significant (p&gt;0.05)</b>					
Contig 5	?	1:0	1	3:0	0.0629
Lactate Dehydrogenase A	LDHA	1:0	2	3:0	0.0669
Macrophage Erythroblast Attacher	MAEA	2:0	3,4	3:0	0.0669
Proteosome 26S ATPase Subunit 3	PSMC3	1:0	2	3:0	0.0669
Spondin 1	SPON1	1:0	3	3:0	0.0669
Contig 4	?	2:0	1,2	2:0	0.1348
Decorin	DCN	1:0	3	2:0	0.1348
Eukaryotic Translation Initiation Factor 4E2	EIF4E2	2:0	2,3	2:0	0.1348
Glucose Phosphate Isomerase	GPI/PGI	2:0	2,3	2:0	0.1348
HRAS Like Suppressor	HRAS-LS	1:0	2	2:0	0.1348
MGC75678	?	1:0	1	2:0	0.1348
Nardilysin	NARD	1:0	3	2:0	0.1348
Nuclear Hormone Receptor Activity	NRH	1:0	3	2:0	0.1348
Nucleoside Diphosphate Kinase	NDPK	2:0	2,3	2:0	0.1348



Table 5a: (continued):

Upregulated Transcript Ortholog	Gene	Replication Information		# Copies WC:SR	Chi Square p=
		# Reps (4)	Rep # ID		
<b>Chi Square is Not Significant (p&gt;0.05)</b>					
Pectinesterase Inhibitor	Pecinhib	1:0	3	2:0	0.1348
Phosphogluconate Dehydrogenase	PGD	1:0	2	2:0	0.1348
Reprolysin type disintegrin & metalloprotease	ADAMSTS15	1:0	1	2:0	0.1348
Ribosomal Protein L7	RPL7	2:0	2,3	2:0	0.1348
Ribosomal Protein L7a	RPL7a	2:0	1,2	2:0	0.1348
Ribosomal Protein S3	RPS3	1:0	2	2:0	0.1348
RNA Binding Motif S1	RBMS1	1:0	3	2:0	0.1348
Small Nuclear RNA LSM3	LSM3	1:0	1	2:0	0.1348
Transmembrane Protein 167	TM167	2:0	1,3	2:0	0.1348
Acetyl-coA Acetyltransferase/Thiolase	ACAT1	1:0	3	1:0	0.2906
Acyl-coA Synthetase	ACSL1	1:0	4	1:0	0.2906
Cadherin Related Neuronal Receptor	CNRco1	1:0	1	1:0	0.2906
Cartilage Associated Protein	CRTAP/CASP	1:0	2	1:0	0.2906
Chaperonin Protein theta	CCT8	1:0	2	1:0	0.2906
Chaperonin Protein zeta	CCT6A	1:0	2	1:0	0.2906
Chromosome 10 Orf 58	C10orf58	1:0	3	1:0	0.2906
Clone 38f16	?	1:0	3	1:0	0.2906
Clone BU37787	?	1:0	2	1:0	0.2906
Clone BU447803	?	1:0	2	1:0	0.2906
Collagen, alpha 2 Type 5	COL5A2	1:0	2	1:0	0.2906
Cyclin D2	CCND2	1:0	2	1:0	0.2906
Cytokine like nuclear factor n-pac	CYNF n-pac	1:0	2	1:0	0.2906
Dipeptidase M20 Family	CNDP2	1:0	2	1:0	0.2906
Eukaryotic Translation Elongation Factor 1A1	EEF1A1	1:0	3	1:0	0.2906
Exocyst Complex Component 7	EXOC7	1:0	2	1:0	0.2906
Ezrin/Villin	VIL2	1:0	2	1:0	0.2906
Fibronectin Type 3	FN3/FANK	1:0	3	1:0	0.2906
GABA A Receptor	GABA A R	1:0	2	1:0	0.2906
Heme-Oxygenase 1	HMOX1	1:0	2	1:0	0.2906
Hypothetical Protein	?	1:0	3	1:0	0.2906
Inhibitor of Kappa Light Polypeptide Enhancer in B cells	IKBKAP	1:0	2	1:0	0.2906
KIAA1432	?	1:0	3	1:0	0.2906
Mannosidase, alpha Class 2B Member 2	MAN2BA	1:0	2	1:0	0.2906
Marvel Domain Containing 2	MDC2	1:0	3	1:0	0.2906
Myosin regulator light chain L20-B	MYRN	1:0	2	1:0	0.2906
Periostin	POSTN	1:0	3	1:0	0.2906
Plastin 3	PLS3	1:0	3	1:0	0.2906
Ribosomal Protein L26	RPL26	1:0	2	1:0	0.2906
Ribosomal Protein L53 (mitochondrial)	L53mt	1:0	3	1:0	0.2906
Sec13L	Sec13L	1:0	1	1:0	0.2906
Secreted Modular Calcium Binding	SMOC2	1:0	3	1:0	0.2906
Small Inducible Cytokine E1	SCYE1	1:0	3	1:0	0.2906
Spleen Focus Forming Virus Sp1 Oncogene Activation	SFFV	1:0	1	1:0	0.2906
Vimentin	VIM	1:0	2	1:0	0.2906
White Carneau Unidentified Singlet # 1	?	1:0	1	1:0	0.2906
White Carneau Unidentified Singlet # 2	?	1:0	1	1:0	0.2906
White Carneau Unidentified Singlet # 3	?	1:0	2	1:0	0.2906
White Carneau Unidentified Singlet # 4	?	1:0	2	1:0	0.2906
White Carneau Unidentified Singlet # 5	?	1:0	2	1:0	0.2906
White Carneau Unidentified Singlet # 6	?	1:0	3	1:0	0.2906
White Carneau Unidentified Singlet # 7	?	1:0	3	1:0	0.2906
White Carneau Unidentified Singlet # 8	?	1:0	3	1:0	0.2906

Table 5b: Differentially Expressed Genes Exclusive to Show Racer

Upregulated Transcript Ortholog	Gene	Replication Information		# Copies	Chi Square
		# Reps (4)	Rep # ID	WC:SR	p=
<b>Chi Square is Significant (p&lt;0.05)</b>					
Cytochrome B	CYTB	0:4	1,2,3,4	0:36	0.0000
Fibronectin Type 1	FN1	0:2	3,4	0:21	0.0000
Fibulin 5	FBLN5	0:1	2	0:24	0.0000
Contig 2	?	0:1	3	0:12	0.0011
Smooth Muscle Cell Myosin Heavy Chain	MYH11	0:1	2	0:8	0.0072
Tropomyosin, alpha	TPM1	0:2	3,4	0:8	0.0072
Alpha 2 Actin	ACTA2	0:2	1,2	0:7	0.0119
Myosin Light Chain Kinase/Telokin	MYLK	0:2	1,2	0:6	0.0200
Proteosome 26S non-ATPase Subunit	PSMD1	0:1	4	0:6	0.0200
Coactosin	COTL1	0:1	3	0:5	0.0338
Cytochrome Oxidase I	COI	0:2	1,2	0:5	0.0338
<b>Chi Square is Not Significant (p&gt;0.05)</b>					
BAC CH261-3201	?	0:1	2	0:4	0.0578
BAC CH261-71A16	?	0:1	2	0:4	0.0578
KS5 Protein	KS5	0:1	3	0:4	0.0578
BAC CH261-20I24	?	0:1	2	0:3	0.1005
GTP Binding Protein	RAB1A	0:2	3,4	0:3	0.1005
Mariner 1 Transposase Gene	SET	0:1	1	0:3	0.1005
Proteosome 26S ATPase Subunit 4	PSMC1	0:2	3,4	0:3	0.1005
SEC61 Gamma Subunit	SEC61G	0:2	1,2	0:3	0.1005
BAC CH261-46G16	?	0:1	2	0:2	0.1802
FLJ13089	FLJ13089	0:1	3	0:2	0.1802
Fumarate Hydratase/fumarase	FH	0:1	3	0:2	0.1802
Histone Promoter Control Protein 1	HPC1	0:1	3	0:2	0.1802
Sarcolipin	SLN	0:1	2	0:2	0.1802
Contig 3	?	0:1	2	0:2	0.1864
Actin Related Protein 10	ACTR10	0:1	1	0:1	0.3435
Actin Related Protein 3	ACTR3	0:1	3	0:1	0.3435
Activin (TGF-B Receptor 1)	ACVR1	0:1	3	0:1	0.3435
Adenosylmethionine Carboxylase	AMD1	0:1	3	0:1	0.3435
Aprataxin	APTX	0:1	4	0:1	0.3435
Arrestin Domain	ARRB	0:1	1	0:1	0.3435
BAC CH261-10P22	?	0:1	2	0:1	0.3435
BAC CH261-15P9	?	0:1	2	0:1	0.3435
BAC CH261-16J12	?	0:1	1	0:1	0.3435
BAC CH261-187N23	?	0:1	1	0:1	0.3435
BAC CH261-93G5	?	0:1	2	0:1	0.3435
C3a Anaphylatox	C3AR	0:1	1	0:1	0.3435
Calnexin	CANX	0:1	1	0:1	0.3435
Centromere Protein	CENP-C	0:1	2	0:1	0.3435
chEST380b13	?	0:1	1	0:1	0.3435
Clone BX268708	?	0:1	3	0:1	0.3435
Clone Q553U2	?	0:1	3	0:1	0.3435
FK506 Binding Protein 9	FKBP9	0:1	2	0:1	0.3435
H+/K+ ATPase Subunit B	ATP4B	0:1	2	0:1	0.3435
Heat Shock Protein 40	DNAJ2	0:1	1	0:1	0.3435
High Mobility Group 14	HMG14	0:1	1	0:1	0.3435
High Mobility Group 2a	HMG2a	0:1	3	0:1	0.3435
Josephin Domain Containing 3	JOSD3	0:1	1	0:1	0.3435

Table 5b (continued):

Upregulated Transcript Ortholog	Gene	Replication Information		# Copies WC:SR	Chi Square p=
		# Reps (4)	Rep # ID		
<b>Chi Square is Not Significant (p&gt;0.05)</b>					
Microsatellite 2G	?	0:1	2	0:1	0.3435
Microtubule Associated Serine /Threonine Kinase	MAST4	0:1	1	0:1	0.3435
Myoglobin Gene	MB	0:1	2	0:1	0.3435
Nucleoporin	NUP37	0:1	1	0:1	0.3435
Phosphoglucomutase 5	PGM5	0:1	1	0:1	0.3435
Ribonuclease/Angiogenin Inhibitor 1	RNH1	0:1	1	0:1	0.3435
ribosomal DNA repeat	rDNA R	0:1	2	0:1	0.3435
RIKEN 3110009E18	?	0:1	1	0:1	0.3435
RIKEN 6030443007	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 1	?	0:1	1	0:1	0.3435
Show Racer Unidentified Singlet # 2	?	0:1	1	0:1	0.3435
Show Racer Unidentified Singlet # 3	?	0:1	1	0:1	0.3435
Show Racer Unidentified Singlet # 4	?	0:1	1	0:1	0.3435
Show Racer Unidentified Singlet # 5	?	0:1	1	0:1	0.3435
Show Racer Unidentified Singlet # 6	?	0:1	1	0:1	0.3435
Show Racer Unidentified Singlet # 7	?	0:1	1	0:1	0.3435
Show Racer Unidentified Singlet # 8	?	0:1	1	0:1	0.3435
Show Racer Unidentified Singlet # 9	?	0:1	1	0:1	0.3435
Show Racer Unidentified Singlet # 10	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 11	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 12	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 13	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 14	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 15	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 16	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 17	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 18	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 19	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 20	?	0:1	2	0:1	0.3435
Show Racer Unidentified Singlet # 21	?	0:1	3	0:1	0.3435
Show Racer Unidentified Singlet # 22	?	0:1	3	0:1	0.3435
SP1 Transcription Activator cofactor	SP1 TFC	0:1	1	0:1	0.3435
Spermidine/Spermine N1-acetyltransferase	SAT	0:1	2	0:1	0.3435
Squalene epoxidase	SQLE	0:1	3	0:1	0.3435
Steroid Alpha Reductase	SRD5A2L2	0:1	2	0:1	0.3435
Ubiquitin	UBN1	0:1	2	0:1	0.3435
Ubiquinone/NADH Dehydrogenase	NDUFA10	0:1	3	0:1	0.3435
Zinc Finger Protein 100	ZFP-100	0:1	4	0:1	0.3435
Zinc Finger Protein 183	ZFP-183	0:1	4	0:1	0.3435

Table 5c: Differentially Expressed Genes Found in WC and SR

Upregulated Transcript Ortholog	Gene	Replication Information			# Copies WC:SR	Chi Square p=
		# Reps (4)	Rep # ID			
			WC	SR		
<b>Chi Square is Significant (p&lt;0.05)</b>						
Clone WAG-65N20	?	3:1	1,3,4	1	36:1	0.0000
Lumican/Keratan Sulfate PG	LUM	2:2	1,3	1,2	2:47	0.0000
Ribosomal Protein L3	RPL3	1:2	3	3,4	5:69	0.0000
Chemokine Ligand 12	CLXL12	3:1	1,2,3	4	13:1	0.0006
NADH Subunit 4	ND4	1:4	3	1,2,3,4	1:12	0.0040
Beta Actin	ACTB	3:1	1,2,3	1	7:1	0.0221
16S Ribosomal RNA	?	1:2	3	1,2	1:7	0.0480
<b>Chi Square is Not Significant (p&gt;0.05)</b>						
Cytochrome Oxidase II	COII	4:4	1,2,3,4	1,2,3,4	11:25	0.0559
Eukaryotic Translation Initiation Factor 4A2	EIF4A2	2:2	3,4	1,2	6:16	0.0574
Aldehyde Dehydrogenase Isoform E3	ALDH9A1	1:1	1	4	5:1	0.0758
Open Reading Frame 45 Chromosome 20	C20orf45	1:2	3	3,4	1:6	0.0794
SEC 61 Alpha Subunit	SEC61A	3:1	1,2,3	1	6:2	0.1162
Peroxiredoxin I	PRDX1	2:4	2,3	1,2,3,4	8:17	0.1191
Proteasome Maturation Protein	POMP	1:2	3,4	1,3,4	4:8	0.3307
Ribosomal Protein S8	RPS8	1:2	2	1,2	1:3	0.3715
Solute Carrier 25 A5/A6	SLC25A6	2:1	2,3	1	5:3	0.3882
Eukaryotic Translation Initiation Factor 4G2	EIF4G2	1:1	4	1	2:1	0.5002
Matrix GLA Protein	MGP	1:1	3	1	2:1	0.5002
Ribosomal Protein S3A	RPS3A	1:1	2	3	2:1	0.5002
Lactate Dehydrogenase B	LDHB	4:4	1,2,3,4	1,2,3,4	18:23	0.6595
NADH Subunit 1	ND1	2:3	1,2	1,2,4	3:4	0.8147
ATPase 6/8 subunits	ATPase6/8	2:1	3,4	1	2:2	0.9125
Fatty Acid Binding Protein 4	FABP4	1:1	3	1	2:2	0.9125
Prohibitin 2	RCJMB04	1:1	1	3	2:2	0.9127
Chaperonin Protein CCT5	CCT5	1:1	4	4	1:1	0.9381
Eukaryotic Translation Initiation Factor 3S	EIF3S	1:1	2	4	1:1	0.9381
Nuclear Phosphoprotein 32	ANP32A	1:1	1	4	1:1	0.9381
Ribosomal Protein S6	RPS6	1:1	2	4	1:1	0.9381

The most consistently differentially expressed transcript codes for cytochrome b (CYTB), which was exclusively up regulated in the SR in all four experimental repetitions (Table 5b). Cytochrome oxidase II (COII) was also found in all four experiments (Table 5c). However, it was expressed in both breeds, although in each case, relative copy number was greater in the SR. Chemokine Ligand 12 (CXCL12) and beta actin (ACTB) were up regulated in

the WC in three comparisons (Table 5c). In contrast, alpha actin (ACTA2) was up regulated in the SR in two of the comparisons (Table 5b). The fact that ACTB and CXCL12 were each found once in the SR illustrates the capacity of RDA to identify relative differences in copy numbers of the same transcript. A number of transcripts were found only in one copy, in one experiment, resulting in a non-significant contribution to the overall genetic difference between breeds. However, knowing that RDA can isolate rarely expressed transcripts, all difference products were analyzed in terms of their metabolic pathway and biological process. Of the 165 unique tags, 124 were entered into the KEGG metabolic hierarchy, and 84 were successfully placed within this system. The distribution of general biological functions across KEGG Level One is presented in Table 6.

Table 6: General Biological Functions (KEGG Level 1)

General Biological Functions KEGG Level 1 Category	White Carneau				Show Racer			
	# genes	% genes	# copies	% copies	# genes	% genes	# copies	% copies
Metabolism	19	23.5%	101	33.0%	13	21.0%	110	31.4%
Genetic Information Processing	16	19.8%	53	17.3%	10	16.1%	91	26.0%
Environmental Information Processing	11	13.6%	36	11.8%	9	14.5%	37	10.6%
Cellular Processes	9	11.1%	31	10.1%	10	16.1%	43	12.3%
Human Diseases	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Not Placed	26	32.1%	85	27.8%	20	32.3%	69	19.7%
<b>Totals:</b>	<b>81</b>	<b>100%</b>	<b>306</b>	<b>100%</b>	<b>62</b>	<b>100%</b>	<b>350</b>	<b>100%</b>

At this level of analysis, there are no obvious differences between WC and SR aortic SMC. The number of unique genes is not significantly different from each other, but there is no information regarding which genes are actively participating in the different levels of Biological Functions. In order to observe

the actual genetic differences at the cellular level, it is necessary to look at the specific types of metabolism in Level 2 of the KEGG system (Table 7).

Table 7: Types of Metabolism (KEGG Level 2)

Types of Metabolism (KEGG Level 2)	Differentially Expressed Genes							
	White Carneau				Show Racer			
	# genes	% genes	# copies	% copies	# genes	% genes	# copies	% copies
Amino acid metabolism	1	1.1%	13	4.0%	1	1.4%	1	0.2%
Biosynthesis of Secondary Metabolites	0	0.0%	0	0.0%	2	2.9%	2	0.5%
Carbohydrate Metabolism	6	6.8%	45	13.8%	2	2.9%	3	0.7%
Cell adhesion molecules	0	0.0%	0	0.0%	1	1.4%	21	5.1%
Cell Communication	4	4.5%	14	4.3%	4	5.7%	35	8.6%
Cell Growth and death	2	2.3%	2	0.6%	0	0.0%	0	0.0%
Cell Motility	2	2.3%	8	2.4%	6	8.6%	37	9.0%
Chaperones & folding catalysts	0	0.0%	0	0.0%	2	2.9%	2	0.5%
CHO/LIP/AA/XEN Metabolism	2	2.3%	6	1.8%	1	1.4%	1	0.2%
Endocrine System	3	3.4%	3	0.9%	0	0.0%	0	0.0%
Energy Metabolism	1	1.1%	1	0.3%	5	7.1%	55	13.4%
Folding, Sorting & Degradation	4	4.5%	16	4.9%	3	4.3%	10	2.4%
Glycan Biosynthesis & Metabolism	2	2.3%	17	5.2%	0	0.0%	0	0.0%
Immune System	4	4.5%	22	6.7%	4	5.7%	4	1.0%
Lipid Metabolism	2	2.3%	11	3.4%	1	1.4%	1	0.2%
Membrane Transport	1	1.1%	6	1.8%	2	2.9%	5	1.2%
Metabolism of Cofactors & Vitamins	1	1.1%	1	0.3%	0	0.0%	0	0.0%
Nucleotide metabolism	1	1.1%	2	0.6%	0	0.0%	0	0.0%
Peptidase	2	2.3%	3	0.9%	0	0.0%	0	0.0%
Protein kinases	0	0.0%	0	0.0%	1	1.4%	1	0.2%
Proteoglycan Metabolism	1	1.1%	2	0.6%	1	1.4%	47	11.5%
Signal Transduction	5	5.7%	9	2.8%	3	4.3%	3	0.7%
Signaling Molecules & Interaction	6	6.8%	24	7.3%	6	8.6%	33	8.1%
Transcription	0	0.0%	0	0.0%	3	4.3%	3	0.7%
Translation	12	13.6%	37	11.3%	2	2.9%	76	18.6%
Not Placed	26	29.5%	85	26.0%	20	28.6%	69	16.9%
<b>Totals:</b>	<b>88</b>	<b>100.0%</b>	<b>327</b>	<b>100.0%</b>	<b>70</b>	<b>100.0%</b>	<b>409</b>	<b>100.0%</b>

Subtle differences between the two breeds begin to appear at KEGG Level 2. Glycan metabolism and peptidase activity appear to be increased in the WC, in addition to genes involved in translation. In contrast, only 2 genes in the SR contribute to translation, but the overall copy number is much higher (76 in SR vs. 37 in WC). In the SR, energy metabolism is clearly up regulated, with these transcripts representing 13.4% of all tags placed versus 0.3% in the WC. In the WC, carbohydrate metabolism is up regulated, as 13.8% of the WC transcripts participated in this pathway in contrast to only 0.7% of SR tags.

There appears to be no difference in the number of genes involved in the immune system, cell communication, or signaling molecules and interaction between the two breeds. However, further analysis into the specific pathways of each of the types of metabolism reveal differences that are not apparent in the first two levels of the KEGG hierarchal system. Table 8 (p105-108) displays the individual genes that participate in each of the delineated pathways as categorized by KEGG Level 3.

Table 8: Types of Metabolism and Specific Pathways (KEGG Level 3)

Types of Metabolism (KEGG Level 2)		White Carneau		Show Racer		
Specific Pathways (KEGG Level 3)	Gene	Protein Product	Copy #	Gene	Protein Product	Copy #
<b>Amino acid metabolism</b>						
Methionine Metabolism						
Urea Cycle	NAT13	N-acetyltransferase 13	13	AMD1	adenosylmethionine decarboxylase	1
Urea Cycle & Metabolism of Amino Groups				AMD1	adenosylmethionine decarboxylase	1
<b>Biosynthesis of Secondary Metabolites</b>						
Terpenoid biosynthesis				SRD5A2L2	steroid alpha reductase	1
				SQLE	squalene epoxidase	1
<b>Carbohydrate Metabolism</b>						
Citric Acid cycle (TCA)				FH	fumarate hydratase/fumarase	2
Glycolysis/Gluconeogenesis	ENO1 GPI/PGI LDH A	enolase glucose phosphate isomerase lactate dehydrogenase A	30 2 3	PGM5	phosphoglucomutase 5	1
Pentose & Glucouronate interconversion	Pectinhib	pectinesterase inhibitor	2			
Pentose phosphate pathway	GPI/PGI PGD TKT	glucose phosphate isomerase phosphogluconate dehydrogenase transketolase	2 2 6	PGM5	phosphoglucomutase 5	1
Pyruvate metabolism	LDHA	lactate dehydrogenase A	3			
Starch & Sucrose Metabolism	Pectinhib	pectinesterase inhibitor	2			
<b>Cell adhesion molecules</b>						
ECM Components				FN1	fibronectin type 1	21
<b>Cell Communication</b>						
Adherens Junction	ACTB	beta actin	7	ACTA2 ACTB	alpha actin beta actin	7 1
Focal Adhesions	CCND2 COL5A2 COL1A2 ACTB	cyclin D2 collagen, alpha 2 type 5 collagen, alpha 2 type 1 beta actin	1 1 5 7	ACTA2 FN1 MYLK ACTB	alpha actin fibronectin type 1 myosin light chain kinase beta actin	7 21 6 1
Tight Junction	ACTB	beta actin	7	ACTA2 ACTB	alpha actin beta actin	7 1
<b>Cell Growth and death</b>						
Apoptosis	CCND2	cyclin D2	1			
Cell Cycle	IKBKAP	inhibitor of kappa light polypeptide enhancer	1			
<b>Cell Motility</b>						
Regulation of Actin Cytoskeleton	EZR/VIL2  ACTB	ezrin/villin 2  beta actin	1  7	ACTA2 ACTR 10 ACTR 3 FN1 MYLK ACTB	alpha actin actin related protein 10 actin related protein 3 fibronectin type 1 myosin light chain kinase beta actin	7 1 1 21 6 1



Table 8 (continued)

	White Carneau		Show Racer	
	Gene	Protein Product	Gene	Protein Product
<b>Types of Metabolism (KEGG Level 2)</b>				
<b>Specific Pathways (KEGG Level 3)</b>				
<b>Chaperones &amp; folding catalysts</b>				
Peptidyl Prolyl isomerase			FKBP9	FK506 binding protein 9
Endoplasmic Reticulum			CANX	calnexin
<b>CHOLIP/AAXEN Metabolism</b>				
General Metabolism	ALDH9A1 ACAT1	aldehyde dehydrogenase E3 acetyl CoA acetyltransferase/thiolase	ALDH9A1	aldehyde dehydrogenase E3
<b>Endocrine System</b>				
Adipocyte Signaling Pathway	IKBKAP	inhibitor of kappa light polypeptide enhancer		
Insulin Signaling Pathway	EIF4E2 EXOC7	eukaryotic translation initiation factor 4E2 exocyst complex component 7		
<b>Energy Metabolism</b>				
Oxidative Phosphorylation	ND4	NADH subunit 4	ATP4B COI CYTB ND4 NDUFA10	H+/K+ ATPase subunit B Cytochrome oxidase 1 cytochrome b NADH subunit 4 NADH dehydrogenase
<b>Folding, Sorting &amp; Degradation</b>				
Protein Folding and Associated Processes	CCT6A CCT8	chaperonin zeta chaperonin theta	DNAJ2	heat shock protein 40
Proteosome	PSMC2 PSMC3	ATPase Subunit 2 ATPase subunit 3	PSMC1 PSMD1	ATPase subunit 4 Non-ATPase subunit
<b>Glycan Biosynthesis &amp; Metabolism</b>				
N-Glycan Biosynthesis	RPN1	ribophorin I		
N-Glycan Degradation	MAN2BA	mannosidase, alpha, Class B		
<b>Immune System</b>				
Antigen processing & presentation			CANX	calnexin
B cell receptor pathway	IKBKAP	inhibitor of kappa light polypeptide enhancer		
Complement & coagulation cascades			C3AR1	C3a anaphylatoxin receptor
Leukocyte transendothelial migration	ACTB CXCL12 EZR/VIL2	beta actin chemokine ligand 12 ezrin/villin 2	ACTB CXCL12	beta actin chemokine ligand 12
T cell receptor pathway	IKBKAP	inhibitor of kappa light polypeptide enhancer		
Toll-like receptor pathway	IKBKAP	inhibitor of kappa light polypeptide enhancer		
<b>Lipid Metabolism</b>				
Biosynthesis of steroids				
Fatty acid biosynthesis	ACSL1	acyl CoA synthetase		
Glycerol lipid metabolism	DGAT2	diacylglycerol O-acetyltransferase	SQLE	squalene epoxidase

Table 8 (continued)

	Types of Metabolism (KEGG Level 2)		White Carneau		Show Racer	
	Gene	Protein Product	Gene	Protein Product	Gene	Protein Product
Membrane Transporters						
Other Transporters						
	SEC61A	SEC 61 protein alpha subunit	SEC61A SEC61G	SEC 61 protein alpha subunit SEC 61 protein gamma subunit	SEC61A SEC61G	SEC 61 protein alpha subunit SEC 61 protein gamma subunit
Metabolism of Cofactors & Vitamins						
Porphyrin Metabolism	HMOX1	heme oxygenase 1				
Nucleotide metabolism						
Purine Metabolism	NDPK	nucleoside diphosphate kinase				
Pyrimidine Metabolism	NDPK	nucleoside diphosphate kinase				
Peptidase						
Metallopeptidases (M20)	CNDP2 NARD	dipeptidase, M20 family nardiysin				
Protein Kinases						
Serine/Threonine Kinase						
Proteoglycan Metabolism						
Small leucine rich (SLRP) Class II	LUM	lumican			MAST4 LUM	microtubule associated S/T kinase lumican
Signal Transduction						
Jak/Stat Signaling Pathway	CCND2	cyclin D2				
MAPK Signaling Pathway					ARRB	arrestin domain
mTOR Signaling Pathway	EIF4E2	eukaryotic translation initiation factor 4E2				
Notch Signaling Pathway	ADAMSTS15	reprolysin type disintegrin & metalloprotease				
TGF-beta Signaling Pathway	DCN SPON1	decorin spodin			ACVR1 Sp1 TFC	activin/TGFB receptor 1 Sp1 Transcription cofactor
Wnt Signaling Pathway	CCND2	cyclin D2				
Signaling Molecules & Interaction						
Calcium Signaling Pathway						
CAM Ligands					MYLK FN1	myosin light chain kinase fibronectin type 1
Cytokine-Cytokine Receptor Interaction	CXCL12	chemokine ligand 12				
					CXCL12 ACVR1	chemokine ligand 12 activin/TGFB receptor 1
ECM-Receptor Interaction	COL5A2 COL1A2 SPON1	collagen, alpha 2 type 5 collagen, alpha 2 type 1 spodin			FN1	fibronectin type 1
GTP Binding Proteins						
Neuroactive Ligand-Receptor Interaction	COPG GABA A Rec	coatomer protein complex gamma GABA A Receptor			RAB1A C3AR1	C3a anaphylatox receptor

Table 8 (continued)

Types of Metabolism (KEGG Level 2)		White Carneau		Show Racer		
Specific Pathways (KEGG Level 3)	Gene	Protein Product	Copy #	Gene	Protein Product	Copy #
Transcription						
Transcription factors				HMG2a HMG-14	high mobility group 2a high mobility group 14	1 1
Other Transcription related proteins				RNH1	ribonuclease/angiotensin inhibitor 1	1
Translation						
Translation initiation factors	EEF1A1 EIF4E2	eukaryotic translation elongation factor 1A1 eukaryotic translation initiation factor 4E2	1 1			
Ribosome	L27mt L53mt RPL26L1 RPL3 RPL32 RPL7 RPL7a RPS3 16S rRNA	mitochondrial ribosomal protein L27 mitochondrial ribosomal protein L53 ribosomal protein L26 like 1 ribosomal protein L3 ribosomal protein L32 ribosomal protein L7 ribosomal protein L7a ribosomal protein S3 16S ribosomal RNA (unresolved)	13 1 1 5 6 2 2 2 1	RPL3	ribosomal protein L3	69
Other translation proteins	16S rRNA RBMS1	16S ribosomal RNA (unresolved) RNA binding motif	1 2	16S rRNA	16S ribosomal RNA (unresolved)	7
Not Placed	26		85	20		69

From Table 8, it becomes clear that although each breed expresses the same number of genes in Cell Communication and the Immune System, very different messages are being transcribed. For example, Adherens Junction and Tight Junction in the WC are represented by ACTB, a marker of the synthetic phenotype. The Focal Adhesion pathway in the SR is represented by ACTA2 and myosin light chain kinase (MYLK), both of which are commonly associated with the contractile phenotype. In contrast, Focal Adhesion in the WC was comprised of collagen (COL1A2, COL5A2), cyclin D2 (CCND2), and ACTB. Phenotypic modifications between breeds are also apparent in the Cell Motility and Signaling Molecules and Interaction categories. Ezrin (VIL2), ACTB, and spondin (SPON) are markers of a synthetic state in the WC cells, whereas actin related proteins 3 & 10 (ACTR3, ACTR10), ACTA2, MYLK, and fibronectin type 1 (FN1) are evidence of a more contractile phenotype in the SR aortic cells. Proteoglycan synthesis is increased in the SR as evidenced by lumican (LUM), but glycosylation itself is up regulated in the WC – indicated by ribophorin 1 (RPN1). CXCL12 participates in both Signaling Molecules and Interaction and the Immune System. As an immune factor, CXCL12 is responsible for leukocyte transendothelial migration and it functions as a signal in Cytokine-Cytokine-Receptor Interactions, both of which are up regulated in the WC.

Ribosome biogenesis and protein translation is clearly up regulated in the SR, but closer inspection of the ribosomal proteins (RP) reveals that all of the expression in the SR is represented by RPL3, whereas in the WC, 8 unique ribosomal proteins are expressed, representing transcripts from both the large (L)

and small (S) subunits. Individual subunits of the 26S proteasome are also differentially expressed between the two breeds, with PSMC2 and PSMC3 proteasome ATPase subunits present in the WC, whereas PSMC1 and PSMD1 subunits were exclusive to the SR.

Differences in energy metabolism observed earlier in the hierarchy are also seen in KEGG Level 3. Oxidative phosphorylation is up regulated in the SR, represented by CYTB, cytochrome oxidase I (COI), NADH subunit 4 (ND4), NADH dehydrogenase (NDUFA10), and H<sup>+</sup>/K<sup>+</sup> ATPase subunit b (ATP4B). Glycolysis is up regulated in the WC, represented by enolase (ENO1), glucose-phosphate isomerase (GPI), and lactate dehydrogenase subunit A (LDHA). Carbohydrate metabolism is also over represented in the WC by the pentose phosphate shunt. Coupled with the increase in fatty acid biosynthesis and glycerol lipid metabolism observed in the WC, it appears that synthesis reactions are favored in the susceptible breed, whereas oxidation is favored by the resistant SR cells.

In contrast to KEGG, which categorizes annotated genes in a hierarchal reference to each other, Pathway Studio employs three distinct lenses to group protein products: Functional Cell Compartment, Types of Metabolism, and Biological Process. One hundred twelve transcripts were placed in at least one of the three Pathway Studio annotation categories. In the Functional Cell Compartments category, the cell is divided into 13 distinct regions. The distribution of WC and SR transcripts across the cellular landscape is presented

in Table 9. The specific genes composing each functional cell compartment are listed individually in Table 10 (p112-114).

Table 9: Functional Cell Compartment of WC and SR Differentially Expressed Genes (Pathway Studio)

Cell Compartment	White Carneau				Show Racer			
	# genes	% genes	# copies	% copies	# genes	% genes	# copies	% copies
Actin Cytoskeleton	0	0.0%	0	0.0%	2	2.9%	2	0.6%
Brush Border	1	1.2%	1	0.3%	0	0.0%	0	0.0%
Cell Surface	1	1.2%	2	0.6%	0	0.0%	0	0.0%
Cytoplasm	31	38.3%	140	40.0%	26	37.7%	160	44.1%
Cytoskeleton	1	1.2%	1	0.3%	1	1.4%	5	1.4%
Endoplasmic Reticulum	4	4.9%	28	8.0%	5	7.2%	10	2.8%
Extracellular	12	14.8%	33	9.4%	4	5.8%	93	25.6%
Golgi	0	0.0%	0	0.0%	1	1.4%	3	0.8%
Lysosome	1	1.2%	1	0.3%	0	0.0%	0	0.0%
Membrane	1	1.2%	8	2.3%	0	0.0%	0	0.0%
Mitochondria	4	4.9%	6	1.7%	5	7.2%	55	15.2%
Nucleus	4	4.9%	41	11.7%	10	14.5%	14	3.9%
Plasma Membrane	4	4.9%	15	4.3%	8	11.6%	9	2.5%
Not Placed	17	21.0%	74	21.1%	7	10.1%	12	3.3%
<b>Totals:</b>	<b>81</b>	<b>100.0%</b>	<b>350</b>	<b>100.0%</b>	<b>69</b>	<b>100.0%</b>	<b>363</b>	<b>100.0%</b>

Table 10: WC and SR Genes Expressed by Functional Cell Compartment (Pathway Studio)

Cell Compartment	White Carneau		Show Racer	
	Gene	Protein Product	Gene	Protein Product
Actin Cytoskeleton			ACTR 10	actin related protein 10
			ACTR 3	actin related protein 3
Brush Border	PLS3	plastin		
Cell Surface	NARD	nardilyisin		
Cytoplasm	16S rRNA	16S ribosomal RNA (unresolved)	16S rRNA	16S ribosomal RNA (unresolved)
	ACTB	beta actin	ACTB	beta actin
	ALDH9A1	aldehyde dehydrogenase E3	ALDH9A1	aldehyde dehydrogenase E3
	RPL3	ribosomal protein L3	RPL3	ribosomal protein L3
	CNDP2	dipeptidase M20 family	ACTA2	alpha actin
	EEF1A1	eukaryotic translation elongation factor 1A1	MYH11	SMC myosin heavy chain
	EXOC7	exocyst complex component 7	MYLK	myosin light chain kinase
	GPI/PGI	glucose phosphate isomerase	PSMD1	proteasome, non ATPase subunit (lid)
	IKBKAP	inhibitor kappa light polypeptide enhancer	SAT	spermidine/spermine N1-acetyltransferase
	LDH A	lactate dehydrogenase subunit A	TPM	tropomyosin
	NAT13	N-acetyltransferase 13	PSMC1	proteasome, ATPase subunit 4
	PGD	phosphogluconate dehydrogenase	RNH1	ribonuclease/angiogenin inhibitor
	RPL26L1	ribosomal protein L26 like 1	ACTR 10	actin related protein 10
	RPL32	ribosomal protein L32	ACTR 3	actin related protein 3
	RPL7	ribosomal protein L7	COTL1	coactosin
	RPL7a	ribosomal protein L7a	AMD1	adenosylmethionine decarboxylase
	RPS3	ribosomal protein S3	SET	mariner 1 transposase gene
	TKT1	transketolase	DNAJA2	heat shock protein 40
	TNFAIP8	TNFA induced protein 8	PGM5	phosphoglucomutase 5
	VIM	vimentin	CANX	calnexin
	ANXA2	annexin 2	FKBP9	FK 506 binding protein 9
	NDPK	nucleoside diphosphate kinase	FN1	fibronectin type 1
	NARD	nardilyisin	APTX	aprataxin
	MAN2B2	mannosidase, alpha Class 2B	ZFP-100	zinc finger protein 100

Table 10 (continued)

Cell Compartment	White Carneau		Show Racer	
	Gene	Protein Product	Gene	Protein Product
Cytoplasm	SPON1	spondin	ZFP-183	zinc finger protein 183
	LGTN	ligatin	RAB1A	GTP binding protein
	ENO1	enolase		
	CCND2	cyclin D2		
	RPN1	ribophorin 1		
	FN3/FANK1	fibronectin type 3		
	PSMC3	proteasome, ATPase subunit 3		
	PLS3	plastin	COTL1	coactosin
	DGAT2	diacylglycerol O-acetyltransferase	SEC61G	SEC 61 protein gamma subunit
	SEC13L	SEC 13 Protein L	RAB1A	GTP binding protein
Endoplasmic Reticulum	HMOX1	heme oxygenase	SLN	sarcolipin
	RPN1	ribophorin 1	CANX	calnexin
			FKBP9	FK 506 binding protein 9
			CXCL12	chemokine ligand 12
Extracellular	CXCL12	chemokine ligand 12	LUM	lumican
	LUM	lumican	FN1	fibronectin type 1
	COL1A2	collagen, alpha 2 type 1	FBLN5	fibulin 5
	COL5A2	collagen, alpha 2 type 5		
	CRTAP	cartilage associated protein		
	DCN	decorin		
	SCYE1	small inducible cytokine		
	SMOC2	secreted modular calcium binding		
	POSTN	perostin		
	FN3/FANK1	fibronectin type 3		
Golgi	NDPK	nucleoside diphosphate kinase		
	SPON1	spondin		
	MAN2B2	mannosidase, alpha Class 2B	RAB1A	GTP binding protein
Lysosome				



Table 10 (continued)

Cell Compartment	Gene	White Carneau Protein Product	Copy #	Gene	Show Racer Protein Product	Copy #
Membrane	LGTN	ligatin	8			
	ND4	NADH subunit 4	1	ND4	NADH subunit 4	12
Mitochondria	ACSL1	acyl CoA synthetase	1	CO1	cytochrome oxidase subunit 1	5
	PSMC3	proteasome, ATPase subunit 3	3	CYTB	cytochrome b	36
	ACAT1	acetyl CoA acetyltransferase/thiolase	1	NDUFA10	NADH dehydrogenase	1
				AMD1	adenosylmethionine decarboxylase	1
Nucleus	SFFV	spleen focus forming virus	1	HMG-14	high mobility group 14	1
	DACH1	dachshund homolog 1	9	HMG2a	high mobility group 2a	1
	ENO1	enolase	30	NUP37	nucleoporin	1
	CCND2	cyclin D2	1	APTX	aprataxin	1
				CENP-C	centromere protein	1
				ZFP-100	zinc finger protein 100	1
				ZFP-183	zinc finger protein 183	1
Plasma Membrane				PSMC1	proteasome, non ATPase subunit (lid)	3
				RNH1	ribonuclease/angiogenin inhibitor	1
				SET	matriner 1 transposase gene	3
	SEC61A	SEC 61 protein alpha subunit	6	SEC61A	SEC 61 protein alpha subunit	2
	VIL2	ezrin/villin 2	1	DNAJA2	heat shock protein 40	1
	HMOX1	heme oxygenase	1	ACVR1	activin/TGFB receptor 1	1
	ANXA2	annexin 2	7	ARRB	arrestin domain	1
			ATP4B	H+/K+ ATPase subunit B	1	
			C3AR1	C3a anaphylatox receptor	1	
			SQLE	squalene epoxidase	1	
			PGM5	phosphoglucomutase 5	1	
Not Placed	17		74	7		12
Totals:			350			363

Although the number of genes actively expressed in the mitochondria does not appear different between WC and SR ( $n= 4$ ,  $n= 5$  respectively), transcript number and individual gene identity reveal significant differences in oxidative phosphorylation. This metabolic difference was also revealed in the KEGG system, but due to the different placement schemes the percentages are slightly different.

There were 12 different genetic products expressed in the extracellular compartment in the WC, and only 4 in the SR. However, copy number variation analysis between the two breeds indicates the polygenic contribution from the WC accounts for only 9.4% of overall expression, whereas one-third of the number of genes in the SR account for 25.6% of total expression. The opposite scenario is true in the nucleus, where 4.9% of the genes expressed in the WC constitute 11.7% of the overall expression. In contrast, 14.5% of the genes actively expressed in the SR compose only 3.9% of the total message.

The types of differential metabolism generated by Pathway Studio (Table 11, p116) are comparable to those generated by KEGG. Protease activity is up regulated in the WC, which corresponds to the increase in peptidases observed in KEGG. One major difference between the two annotation systems is that Pathway Studio categorizes the electron transport chain as a general "transporter" where KEGG specified oxidative phosphorylation as a distinct type of metabolism. Glycolysis, amino acid metabolism, and the enzymes of lipid biosynthesis are all encompassed by the umbrella "Metabolic Enzymes" in Pathway Studio.

Table 11: Types of Metabolism (Pathway Studio)

Type of Metabolism (Pathway Studio Group)	White Carneau				Show Racer			
	# genes	% genes	# copies	% copies	# genes	% genes	# copies	% copies
Acetylases	2	2.9%	14	5.1%	1	1.9%	1	0.3%
Cytoskeleton	4	5.8%	10	3.6%	6	11.1%	26	8.3%
Deglycosylase	1	1.4%	1	0.4%	0	0.0%	0	0.0%
ECM	8	11.6%	14	5.1%	2	3.7%	68	21.6%
Extracellular proteins	0	0.0%	0	0.0%	1	1.9%	1	0.3%
Glycosyltransferases	1	1.4%	16	5.8%	0	0.0%	0	0.0%
Kinases	2	2.9%	5	1.8%	3	5.6%	10	3.2%
Ligands	3	4.3%	16	5.8%	2	3.7%	25	7.9%
Metabolic enzymes	6	8.7%	18	6.5%	4	7.4%	4	1.3%
Methyltransferase	0	0.0%	0	0.0%	1	1.9%	3	1.0%
Ph	1	1.4%	1	0.4%	0	0.0%	0	0.0%
Proteases	2	2.9%	3	1.1%	0	0.0%	0	0.0%
Receptors	1	1.4%	8	2.9%	2	3.7%	2	0.6%
Transcription Factors	4	5.8%	41	14.9%	6	11.1%	8	2.5%
Transporters	2	2.9%	7	2.5%	7	13.0%	60	19.0%
Not Placed	32	46.4%	121	44.0%	19	35.2%	107	34.0%
<b>Totals:</b>	<b>69</b>	<b>100.0%</b>	<b>275</b>	<b>100.0%</b>	<b>54</b>	<b>100.0%</b>	<b>315</b>	<b>100.0%</b>

The individual genes composing each category are listed in Table 12 (p117-118). From Table 12 it becomes obvious that the differences observed in the nucleus (Table 10) correlate with the specific transcription factors that differ between the two breeds. For example, dachshund homologue 1c (DACH1), ENO1, inhibitor of kappa light polypeptide enhancer in B cells (IKBKAP), and spleen focus forming virus (SFFV) are regulating transcriptional events in the WC, whereas high mobility groups (HMG14, HMG2a) and zinc finger proteins (ZFP-100, ZFP-183) are active in the SR.

Table 12: WC and SR Genes Expressed by Types of Metabolism (Pathway Studio)

Type of Metabolism (Pathway Studio Group)	White Carneau			Show Racer		
	Gene	Protein Product	Copy #	Gene	Protein Product	Copy #
Acetylases	ACAT1	acetyl CoA acetyltransferase/thiolase	1	SAT	spermidine/spermine N1-acetyltransferase	1
	NAT13	N-acetyltransferase 13	13			
Cytoskeleton	ACTB	beta actin	7	ACTB	beta actin	1
	PLS3	plastin	1	ACTA2	alpha actin	7
	VIL2	ezrin/villin 2	1	ACTR3	actin related protein 3	1
	VIM	vimentin	1	ACTR10	actin related protein 10	1
				MYH11	SMC myosin heavy chain	8
			TPM1	tropomyosin	8	
Deglycosylase	MAN2B2	mannosidase				
ECM	LUM	lumican	2	LUM	lumican	47
	COL1A2	collagen, alpha 2 type 1	5	FN1	fibronectin type 1	21
	COL5A2	collagen, alpha 2 type 5	1			
	CRTAP	cartilage associated protein	1			
	DCN	decorin	2			
	FN3	fibronectin type 3	1			
	POSTN	perlestin	1			
	SMOC2	secreted modular calcium binding	1			
				MB	myoglobin gene	1
Extracellular proteins	RPN1	ribophorin 1	16			
Glycosyltransferases	NDPK	nucleoside diphosphate kinase	2	MYLK	myosin light chain kinase	6
	PSMC3	proteasome ATPase subunit 3	3	PSMC1	proteasome non-ATPase subunit (fid)	3
				ACVR1	activin/TGFB receptor 1	1
Kinases	CXCL12	chemokine ligand 12	13	CXCL12	chemokine ligand 12	1
	GPI/PGI	glucose phosphate isomerase	2	FBLN5	fibulin 5	24
	SCYE1	small inducible cytokine	1			

Table 12 (continued)

Type of Metabolism (Pathway Studio Group)	Gene	White Carneau Protein Product	Copy #	Gene	Show Racer Protein Product	Copy #
Metabolic enzymes	DGAT2	diacylglycerol O-acetyltransferase	10	AMD1	adenosylmethionine decarboxylase	1
	GABA R	GABA Receptor	1	FKBP9	FK 506 binding protein 9	1
	HMOX1	heme oxygenase	1	PGM5	phosphoglucomutase 5	1
	LDHA	lactate dehydrogenase	3	SQLE	squalene epoxidase	1
	PGD	phosphogluconate dehydrogenase	2			
	ACSL1	acyl CoA synthetase	1			
Methyltransferase Ph	VIL2	ezrin/villin2		SET	mariner 1 transposase gene	3
Proteases	NARD	nardilysin	2			
	CNDP2	dipeptidase M20 family	1			
	LGTN	ligatin	8	ACVR1	activin/TGFB receptor 1	1
Receptors				C3AR1	C3a anaphylatox receptor	1
				APTX	aprataxin	1
Transcription Factors	DACH1	dachshund homolog 1	9	HMG14	high mobility group 14	1
	ENO1	enolase	30	HMG2a	high mobility group 2a	1
	IKBKAP	inhibitor kappa light polypeptide enhancer	1	ZFP-100	zinc finger protein 100	1
	SFFV	spleen focus forming virus	1	ZFP-183	zinc finger protein 183	1
				RAB1A	GTP binding protein	3
				ATP4B	H+/K+ ATPase subunit b	1
Transporters				CO1	cytochrome oxidase 1	5
				CYTB	cytochrome b	36
	ND4	NADH subunit 4	1	ND4	NADH subunit 4	12
				NDUFA10	NADH dehydrogenase	1
	SEC61A	SEC 61 protein alpha subunit	6	SEC61A	SEC 61 protein alpha subunit	2
				SEC61G	SEC 61 protein gamma subunit	3

Consistent with the gene expression profile reported for the extracellular component of each breed (Table 10, p112-114), the extracellular matrix metabolism is also quite different (Table 12, p117-118). Although 7 genes constituting a synthetic phenotype are up regulated in the WC, this represents only 5.1% of the total expression. In contrast, only 2 genes are expressed in the SR ECM, but their activity constitutes 21.6% of the overall metabolism, indicating a major difference between the two breeds.

Perhaps the most informative of the Pathway Studio analyses is the breakdown of the actual biological processes in the cell. Table 13 (p.120-123) presents the distribution of involved genes and copy number variation as a percentage of the total cellular process. The contribution of specific genes to each category is listed in Table 14 (p.124-127). Cell adhesion appears to be up regulated in the WC, with the expression of COL1A2, COL5A2, SPON and periostin (POSTN) reflective of a synthetic phenotype. In contrast, Cytoskeletal Organization is up regulated in the SR, where MYLK, SMC myosin heavy chain (MYH11), and ACTA2 comprise a contractile phenotype. Corresponding to the greater ATP needs of a contractile cell, oxidative phosphorylation is also up regulated in the SR. In the WC, glycolysis was again up regulated, revealing a stark contrast in energy metabolism that is a consistent theme throughout the analyses.

Table 13: Types of Biological Processes (Pathway Studio)

Biological Process (GO) (Pathway Studio Group)	White Carneau			Show Racer		
	# genes	% genes	# copies	# genes	% genes	# copies
Acute Phase Response (wound)	1	0.9%	1	0	0.0%	0
Aldehyde & Carnitine Metabolism	1	0.9%	5	1	1.4%	1
Amino acid Phosphorylation	1	0.9%	1	1	1.4%	1
Anaerobic Glycolysis	1	0.9%	3	0	0.0%	0
Angiogenesis	1	0.9%	7	0	0.0%	0
Anti-apoptosis	1	0.9%	4	0	0.0%	0
Apoptosis	1	0.9%	2	0	0.0%	0
Carbohydrate Metabolism	0	0.0%	0	1	1.4%	1
Cell Adhesion	4	3.7%	10	1	1.4%	1
Cell Death	0	0.0%	0	1	1.4%	1
Cell Matrix Adhesion	0	0.0%	0	1	1.4%	24
Cell Motility	1	0.9%	7	2	2.7%	2
Cell Proliferation	1	0.9%	2	0	0.0%	0
Collagen Fibril Organization	1	0.9%	2	1	1.4%	47
Cytoskeletal Organization	1	0.9%	7	4	5.4%	22
Development	1	0.9%	3	0	0.0%	0
DNA Dependent Regulation of Transcription	0	0.0%	0	1	1.4%	1
DNA Packaging	0	0.0%	0	1	1.4%	1
DNA Repair	0	0.0%	0	1	1.4%	1
DNA Unwinding	1	0.9%	3	0	0.0%	0
ECM Organization & Biosynthesis	1	0.9%	2	0	0.0%	0
Electron Transport	0	0.0%	0	1	1.4%	1
Exocytosis	1	0.9%	1	0	0.0%	0
Fat Body Development	1	0.9%	10	0	0.0%	0
Fibrinolysis	1	0.9%	7	0	0.0%	0
G1/S Transition	1	0.9%	1	0	0.0%	0
Generation of Precursor Metabolites & Energy	0	0.0%	0	1	1.4%	1

Table 13 (continued)

Biological Process (GO) (Pathway Studio Group)	White Carneau			Show Racer		
	# genes	% genes	# copies	% genes	# genes	% copies
Glycolysis	2	1.9%	36	8.3%	0	0.0%
Glycolysis/GNG	1	0.9%	2	0.5%	0	0.0%
G-protein Regulation	0	0.0%	0	0.0%	1	1.4%
Immune Response	3	2.8%	16	3.7%	1	1.4%
Inflammatory Response	1	0.9%	1	0.2%	1	1.4%
Intermediate Filament Based Processes	1	0.9%	1	0.2%	0	0.0%
Intracellular Protein Transport	2	1.9%	2	0.5%	1	1.4%
Intracellular Signaling Cascade	0	0.0%	0	0.0%	1	1.4%
Intracellular Transport	1	0.9%	8	1.8%	1	1.4%
Ion Transport	0	0.0%	0	0.0%	1	1.4%
Lipid/ Fatty Acid Metabolism	1	0.9%	1	0.2%	0	0.0%
Lipoprotein Biosynthesis	1	0.9%	1	0.2%	0	0.0%
Macrophage & Lymphocyte Differentiation	1	0.9%	1	0.2%	0	0.0%
Maintenance of Polarity	1	0.9%	1	0.2%	0	0.0%
Mannose Metabolism	1	0.9%	1	0.2%	0	0.0%
Metabolism	1	0.9%	13	3.0%	0	0.0%
Microtubule Based Movement	0	0.0%	0	0.0%	1	1.4%
Morphogenesis	0	0.0%	0	0.0%	1	1.4%
Muscle Contraction	0	0.0%	0	0.0%	1	1.4%
Muscle Development	0	0.0%	0	0.0%	1	1.4%
Nucleoside Metabolism	0	0.0%	0	0.0%	1	1.4%
Nucleotide Metabolism	1	0.9%	2	0.5%	0	0.0%
Ossification	1	0.9%	1	0.2%	0	0.0%
Oxidative Phosphorylation	1	0.9%	1	0.2%	3	4.1%
Oxygen Transport	0	0.0%	0	0.0%	1	1.4%
Pentose Phosphate Shunt	1	0.9%	2	0.5%	0	0.0%
Phosphate Transport	1	0.9%	1	0.2%	0	0.0%



Table 13 (continued)

Biological Process (GO) (Pathway Studio Group)	White Carneau			Show Racer		
	# genes	% genes	# copies	# genes	% genes	# copies
Phospholipid Metabolism	1	0.9%	1	0	0.0%	0
Protein Amino Acid Glycosylation	1	0.9%	16	0	0.0%	0
Protein Biosynthesis	7	6.5%	19	2	2.7%	76
Protein Catabolism	2	1.9%	14	1	1.4%	3
Protein Deglycosylation	1	0.9%	1	0	0.0%	0
Protein Folding	0	0.0%	0	3	4.1%	3
Protein Secretion	1	0.9%	6	2	2.7%	3
Proteolysis	3	2.8%	14	1	1.4%	1
Regulation of Cell Cycle	1	0.9%	1	1	1.4%	1
Regulation of Cell Migration	1	0.9%	13	1	1.4%	1
Regulation of Cell Shape	1	0.9%	1	0	0.0%	0
Regulation of Growth	1	0.9%	6	1	1.4%	24
Regulation of Muscle Contraction	0	0.0%	0	1	1.4%	8
Regulation of Progression through Cell Cycle	0	0.0%	0	1	1.4%	6
Regulation of Transcription	3	2.8%	40	3	4.1%	5
Regulation of Translation	1	0.9%	1	0	0.0%	0
Regulator of IKAP/NFK Cascade	1	0.9%	1	0	0.0%	0
Retinol metabolism	1	0.9%	15	0	0.0%	0
Ribosome Biogenesis	7	6.5%	19	2	2.7%	76
RNA Catabolism	0	0.0%	0	1	1.4%	1
Serine/Threonine Kinase Signaling	1	0.9%	5	1	1.4%	1
Signal Transduction	1	0.9%	1	2	2.7%	2
Skeletal Development	1	0.9%	1	0	0.0%	0
Smooth Muscle Contraction	0	0.0%	0	1	1.4%	8
Spermine Catabolism	0	0.0%	0	1	1.4%	1
Spermine/Spermidine Biosynthesis	0	0.0%	0	1	1.4%	1
Sterol Biosynthesis	0	0.0%	0	1	1.4%	1

Table 13 (continued)

Biological Process (GO) (Pathway Studio Group)	White Carneau			Show Racer		
	# genes	% genes	# copies	% genes	# genes	% copies
TAG Metabolism	1	0.9%	10	2.3%	0	0.0%
TCA intermediate	1	0.9%	3	0.7%	0	0.0%
TGF Signaling Pathway	0	0.0%	0	0.0%	1	0.2%
Thiamine metabolism	1	0.9%	6	1.4%	0	0.0%
Translation Elongation	1	0.9%	1	0.2%	0	0.0%
Translation Initiation	1	0.9%	8	1.8%	0	0.0%
Transmembrane receptor	1	0.9%	1	0.2%	0	0.0%
Transport	4	3.7%	25	5.8%	3	1.1%
Unknown Biological Process	1	0.9%	1	0.2%	1	1.1%
Not Placed	17	15.7%	35	8.1%	8	10.8%
<b>Totals:</b>	<b>108</b>	<b>100.0%</b>	<b>434</b>	<b>100.0%</b>	<b>74</b>	<b>100.0%</b>
					<b>464</b>	<b>100.0%</b>

Table 14: WC and SR Genes Expressed by Biological Process (Pathway Studio)

Biological Process (GO)		White Carneau		Show Racer		
	Gene	Protein Product	Copy #	Gene	Protein Product	Copy #
Acute Phase Response (wound)	FN3	fibronectin type 3	1			
Aldehyde & Carnitine Metabolism	ALDH9A1	aldehyde dehydrogenase E3	5	ALDH9A1	aldehyde dehydrogenase E3	1
Amino acid Phosphorylation	IKBKAP	inhibitor kappa polypeptide enhancer	1	MAST4	microtubule associated serine/threonine kinase	1
Anaerobic Glycolysis	LDHA	lactate dehydrogenase subunit A	3			
Angiogenesis	ANXA2	annexin 2	7			
Anti-apoptosis	TNFaip8	TNFa induced protein 8	4			
Apoptosis	NDPK	nucleoside diphosphate kinase	2			
Carbohydrate Metabolism				PGM5	phosphoglucomutase 5	1
	COL1A2	collagen, alpha 2 type 1	5	PGM5	phosphoglucomutase 5	1
	COL5A2	collagen, alpha 2 type 5	1			
Cell Adhesion	SPON	spondin	3			
	POSTN	perostin	1			
Cell Death				APTJ	aprataxin	1
Cell Matrix Adhesion				FBLN5	fibulin 5	24
Cell Motility	ACTB	beta actin	7	ACTB	beta actin	1
				ACTR3	actin related protein 3	1
Cell Proliferation	NARD	nardilysin	2			
Collagen Fibril Organization	LUM	lumican	2	LUM	lumican	47
	ACTB	beta actin	7	ACTB	beta actin	1
				MYH11	SMC myosin heavy chain	8
				MYLK	myosin light chain kinase	6
				ACTA2	alpha actin	7
Cytoskeletal Organization						
Development	SPON	spondin	3			
DNA Dependent Regulation of Transcription				HMG2a	high mobility group 2a	1
DNA Packaging				HMG14	high mobility group 14	1
DNA Repair				APTJ	aprataxin	1
DNA Unwinding	PSMC3	proteasome, ATPase subunit 3	3			
ECM Organization & Biosynthesis	DCN	decorin	2			
Electron Transport				SQLE	squalene epoxidase	1
Exocytosis	EXOC7	exocyst complex component 7	1			
Fat Body Development	DGAT2	diacylglycerol O- acyltransferase	10			
Fibrinolysis	ANXA2	annexin 2	7			
G1/S Transition	CCND2	cyclin D2	1			
Generation of Precursor Metabolites & Energy				NDUFA10	NADH dehydrogenase	1

Table 14 (continued)

Biological Process (GO) (Pathway Studio Group)	White Carneau		Show Racer	
	Gene	Protein Product	Gene	Protein Product
Glycolysis	ENO1	enolase		
	TKT	transketolase		
	GPI/PGI	glucose phosphate isomerase		
Glycolysis/GNG				
G-protein Regulation				
Immune Response	CXCL12	chemokine ligand 12	ARRB	arrestin domain
	GPI/PGI	glucose phosphate isomerase	CXCL12	chemokine ligand 12
Inflammatory Response	IKBKAP	inhibitor kappa polypeptide enhancer		
	SCYE1	small inducible cytokine	C3AR1	C3a anaphylatoxin receptor
Intermediate Filament Based Processes	VIM	vimentin		
Intracellular Protein Transport	EXOC7	exocyst complex component 7	SEC61G	SEC 61 protein gamma subunit
	SEC13L	SEC 13 protein subunit L		
Intracellular Signaling Cascade				
Intracellular Transport	LGTN	ligatin	MAST4	microtubule associated serine/threonine kinase
Ion Transport			RAB1A	GTP Binding Protein
Lipid/ Fatty Acid Metabolism	ACSL1	acyl CoA synthetase	ATP4B	H+/K+ ATPase subunit B
Lipoprotein Biosynthesis	ACAT1	acetyl CoA acetyltransferase/thiolase		
Macrophage & Lymphocyte Differentiation	SP11	spleen focus forming virus		
	VIL2	ezrin/villin 2		
Maintenance of Polarity	MAN2B2	mannosidase, alpha 2 Class B		
Mannose Metabolism	NAT13	N-acetyltransferase 13		
Metabolism				
Microtubule Based Movement			ACTR10	actin related protein 10
Morphogenesis			PGM5	phosphoglucomutase 5
Muscle Contraction			ACTA2	alpha actin
Muscle Development			MB	myoglobin gene
Nucleoside Metabolism			NDUFA10	NADH dehydrogenase
Nucleotide Metabolism	NDPK	nucleoside diphosphate kinase		
Ossification	COL5A2	collagen, alpha 2 type 5		
	ND4	NADH subunit 4	ND4	NADH subunit 4
Oxidative Phosphorylation			CO1	cytochrome oxidase 1
			CYTB	cytochrome b
Oxygen Transport			MB	myoglobin gene
Pentose Phosphate Shunt	PGD	phosphoglucomate dehydrogenase		
Phosphate Transport	COL5A2	collagen, alpha 2 type 5		
Phospholipid Metabolism	HMOX1	heme oxygenase		

Table 14 (continued)

Biological Process (GO) (Pathway Studio Group)	White Carneau		Show Racer	
	Gene	Protein Product	Gene	Protein Product
Protein Amino Acid Glycosylation	RPN1	ribophorin I		
	16S rRNA	16S ribosomal RNA (unresolved)	16S rRNA	16S ribosomal RNA (unresolved)
	RPL3	ribosomal protein L3	RPL3	ribosomal protein L3
Protein Biosynthesis	RPL26L1	ribosomal protein L26 like 1		
	RPL32	ribosomal protein L32		
	RPL7	ribosomal protein L7		
	RPL7a	ribosomal protein L7a		
	RPS3	ribosomal protein S3		
Protein Catabolism	PSMC2	proteasome, ATPase subunit 2	PSMC1	proteasome, ATPase subunit 4
	PSMC3	proteasome, ATPase subunit 3		
Protein Deglycosylation	MAN2B2	mannosidase, alpha 2 Class B		
Protein Folding			FKBP9	FK 506 binding protein 9
			CANX	calnexin
Protein Secretion			DNAJ2	heat shock protein 40
	SEC61A	SEC 61 protein alpha subunit	SEC61A	SEC 61 protein alpha subunit
Proteolysis			CANX	calnexin
	NARD	nardilysin	PGM5	phosphoglucomutase 5
Regulation of Cell Cycle	PSMC2	proteasome, ATPase subunit 2		
	CNDP2	dipeptidase M20 family		
Regulation of Cell Migration	CCND2	cyclin D2	DNAJ2	heat shock protein 40
	CXCL12	chemokine ligand 12	CXCL12	chemokine ligand 12
Regulation of Cell Shape	VIL2	ezrin/villin 2		
Regulation of Growth	TKT	transketolase	FBLN5	fibulin 5
Regulation of Muscle Contraction			TPM	tropomyosin
			PSMD1	proteasome, non ATPase subunit (lid)
Regulation of Transcription	SFFV	spleen focus forming virus	RAB1A	GTP Binding Protein
	DACH1	dachshund homolog 1	ZFP-100	zinc finger protein 100
	ENO1	enolase	ZFP-183	zinc finger protein 183
Regulation of Translation	EEF1A1	eukaryotic translation elongation factor 1A1		
Regulator of IKAP/NFK Cascade	HMOX1	heme oxygenase		
Retinol metabolism	RBP7	retinol binding protein 7		

Table 14 (continued)

Biological Process (GO) (Pathway Studio Group)	White Carneau			Show Racer		
	Gene	Protein Product	Copy #	Gene	Protein Product	Copy #
Ribosome Biogenesis	16S rRNA	16S ribosomal RNA (unresolved)	1	16S rRNA	16S ribosomal RNA (unresolved)	7
	RPL3	ribosomal protein L3	5	RPL3	ribosomal protein L3	69
	RPL26L1	ribosomal protein L26 like 1	1			
	RPL32	ribosomal protein L32	6			
	RPL7	ribosomal protein L7	2			
RNA Catabolism	RPL7a	ribosomal protein L7a	2			
	RPS3	ribosomal protein S3	2			
Serine/Threonine Kinase Signaling	COL1A2	collagen, alpha 2 type 1	5	RNH1	ribonuclease/angiogenin inhibitor	1
Signal Transduction	SCYE1	small inducible cytokine	1	ACVR1	activin/TGFB receptor 1	1
Skeletal Development	POSTN	perostin	1	C3AR1	C3a anaphylatox receptor	1
Smooth Muscle Contraction				ACVR1	activin/TGFB receptor 1	1
Spermine Catabolism						
Spermine/Spermidine Biosynthesis				MYH11	SMC myosin heavy chain	8
Sterol Biosynthesis				SAT	spermidine/spermine N1-acetyltransferase	1
TAG Metabolism	DGAT2	diacylglycerol O-acyltransferase	10	AMD1	adenosylmethionine decarboxylase	1
TCA Intermediate	LDHA	lactate dehydrogenase subunit A	3	SOLE	squalene epoxidase	1
TGF Signaling Pathway						
Thiamine metabolism	TKT	transketolase	6	ACVR1	activin/TGFB receptor 1	1
Translation Elongation	EEF1A1	eukaryotic translation elongation factor 1A1	1			
Translation Initiation	LGTN	ligatin	8			
Transmembrane Receptor	FN3	fibronectin type 3	1			
Transport	SEC61A	SEC 61 protein alpha subunit	6	SEC61A	SEC 61 protein alpha subunit	2
	VIM	vimentin	1	NUP37	nucleoporin	1
	RBP7	retinol binding protein 7	15	SLN	sarcolipin	2
	PSMC3	proteasome, ATPase subunit 3	3			
Unknown Biological Process	SMOC2	secreted modular calcium binding	1	COTL1	coactosin	5
Not Placed	17		35	8		50

There were more ribosomal proteins located within the KEGG database (n=9) than were found in Pathway Studio (n=7). However, those that were identified in each annotation system were placed in the same manner. Genes involved in the biological process of Protein Biosynthesis were identical to those representing Ribosome Biosynthesis (Table 14, p122). In the KEGG hierarchy the same list of genes representing general levels of Translation were the same as those representing the specific pathway of Ribosomes (Table 8, p105).

Pigeon Transcripts that had an orthologous gene in the chicken were mapped against the chicken genome. Not all of the chicken chromosomes have been resolved, but as of April 2007, 15,368 genes had been placed on chromosomes 1-28, W, Z, and those within the mitochondria<sup>268</sup>. Figure 3 (p129) and Figure 4 (p130) compare the total number of genes that have been placed on each chicken chromosome to the corresponding genes differentially expressed by each pigeon breed.

Although incomplete, the chicken genome provides a null expectation for pigeon gene expression. If the expression differences between breeds were truly random, it would be expected that the number of genes expressed per chromosome would be proportional to chromosome size. Table 15 (p131) presents the Chi Square analysis of the WC compared to the chicken, the SR compared to the chicken, and the WC compared to the SR. A p-value < 0.05 indicates that the number of genes expressed on a given chromosome is different from the null expectation.

Figure 3: Comparison of WC, SR, and Chicken (Gg) Gene Count Per Chicken Chromosome

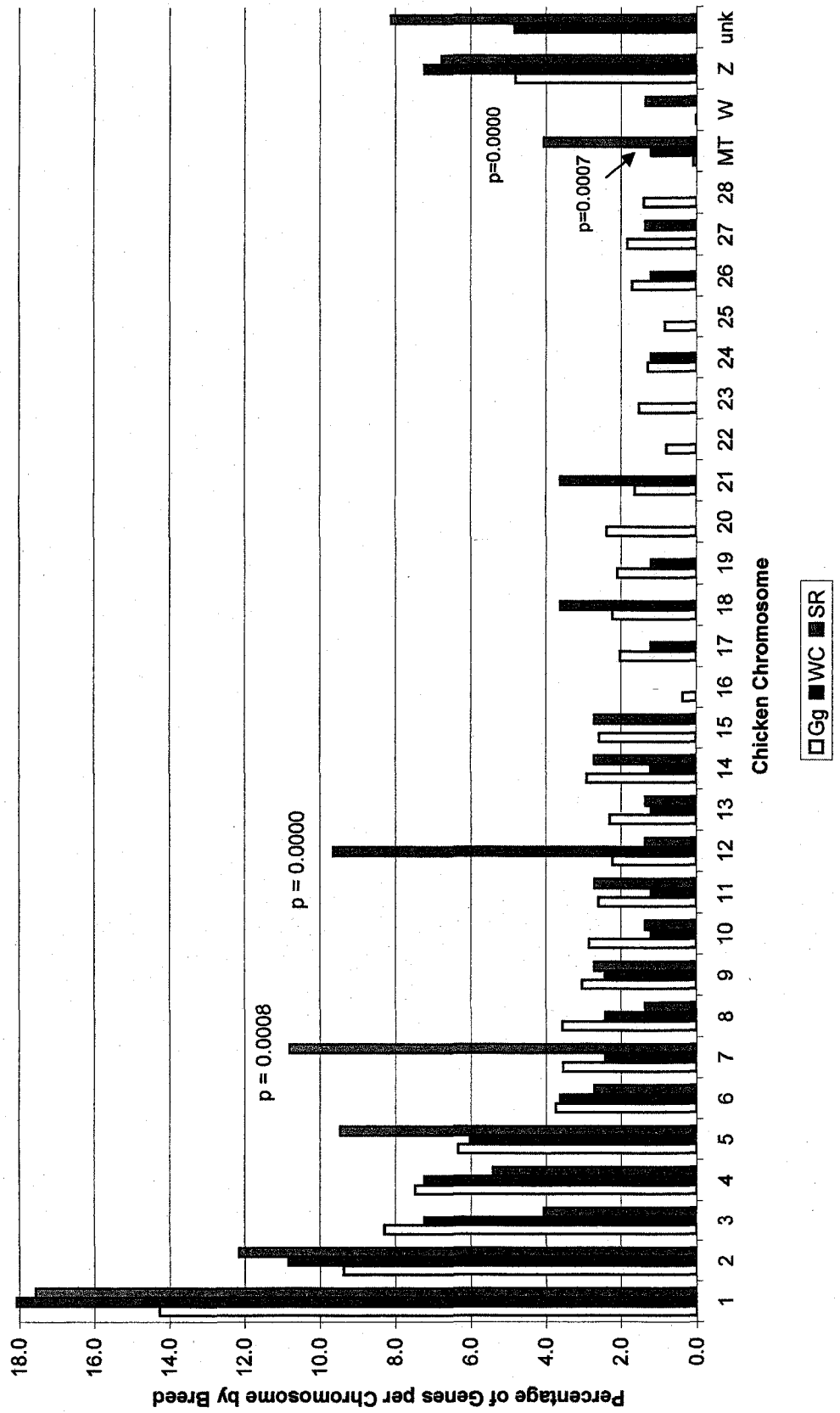




Figure 4: Comparison of WC and SR Gene Count Per Chicken Chromosome

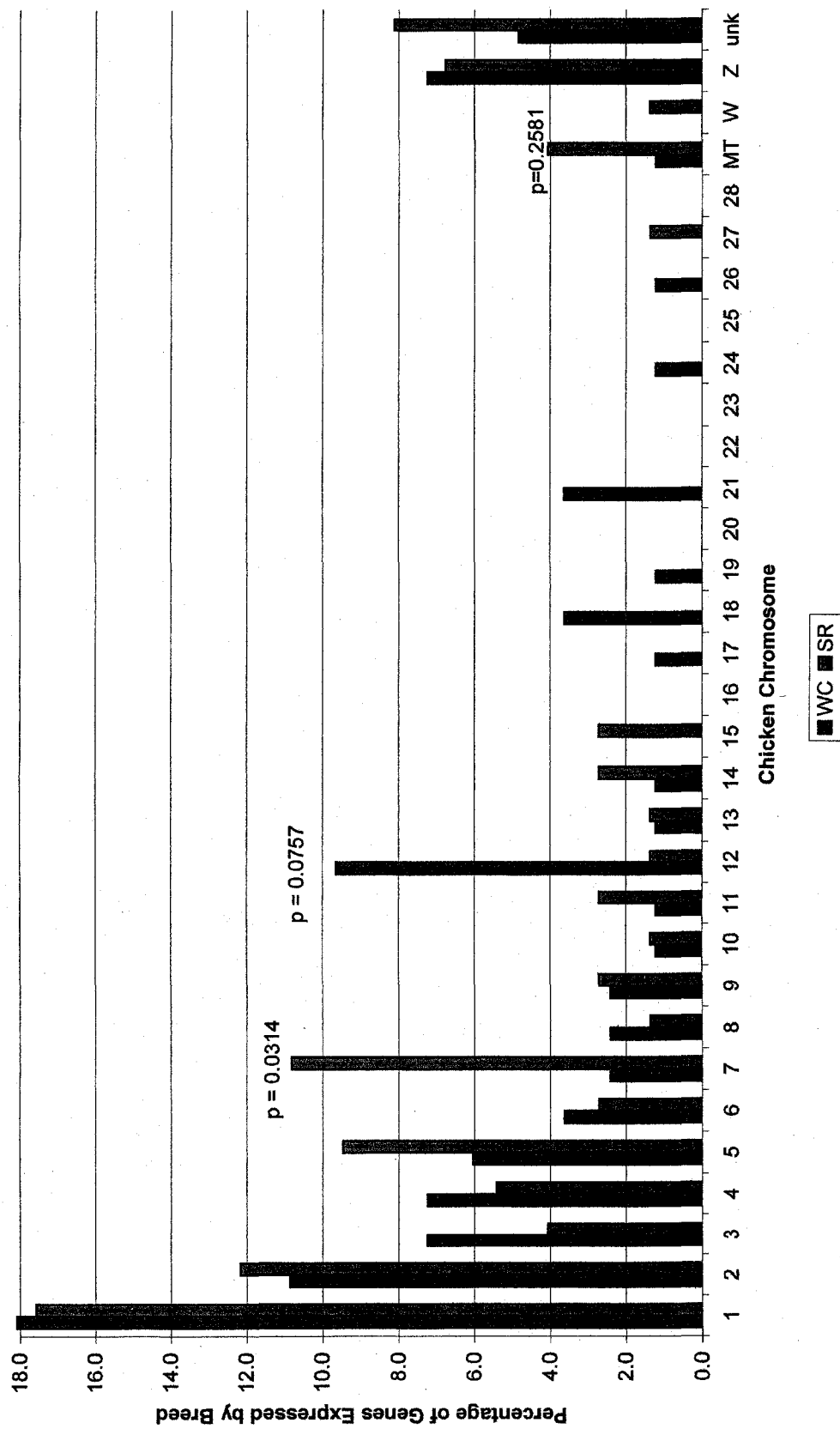


Table 15: Comparison of WC, SR, and Chicken (Gg) Gene Count per Chicken Chromosome

Gg	Chicken vs. White Carneau			Chicken vs. Show Racer			White Carneau vs. Show Racer		
	#Gg	# WC	p=	#Gg	#SR	p=	# WC	#SR	p=
1	2192	15	0.3226	2192	13	0.4177	15	13	0.9343
2	1440	9	0.6461	1440	9	0.4113	9	9	0.7957
3	1272	6	0.7296	1272	3	0.1897	6	3	0.3930
4	1149	6	0.9318	1149	4	0.4989	6	4	0.6405
5	973	5	0.9087	973	7	0.2708	5	7	0.4187
6	574	3	0.9539	574	2	0.6401	3	2	0.7454
7	543	2	0.5799	543	8	0.0008	2	8	0.0314
8	547	2	0.5726	547	1	0.3057	2	1	0.6287
9	465	2	0.7437	465	2	0.8714	2	2	0.8964
10	438	1	0.3683	438	1	0.4390	1	1	0.9349
11	398	1	0.4276	398	2	0.9514	1	2	0.4938
12	341	8	0.0000	341	1	0.6129	8	1	0.0757
13	353	1	0.5072	353	1	0.5877	1	1	0.9349
14	448	1	0.3549	448	2	0.9137	1	2	0.4938
15	395	0	0.1390	395	2	0.9428	0	2	0.1317
16	56	0	0.5817	56	0	0.6029	0	0	na
17	311	1	0.5968	311	0	0.2164	1	0	0.3435
18	342	3	0.3930	342	0	0.1944	3	0	0.0987
19	322	1	0.3361	322	0	0.2083	1	0	0.3435
20	364	0	0.1559	364	0	0.1803	0	0	na
21	251	3	0.1569	251	0	0.2677	3	0	0.0987
22	123	0	0.4132	123	0	0.4397	0	0	na
23	235	0	0.2563	235	0	0.2837	0	0	na
24	198	1	0.9463	198	0	0.3257	1	0	0.3435
25	129	0	0.4019	129	0	0.4287	0	0	na
26	260	1	0.7313	260	0	0.2591	1	0	0.3435
27	280	0	0.2146	280	1	0.7625	0	1	0.2880
28	215	0	0.2779	215	0	0.3055	0	0	na
Mt	13	1	0.0007	13	3	0.0000	1	3	0.2581
W	2	0	0.9172	2	1	0.0000	0	1	0.2880
Z	739	6	0.3047	739	5	0.7620	6	5	0.9079
unk	na	4	na	na	6	na	4	6	0.3966
	<b>15,368</b>	<b>83</b>		<b>15,368</b>	<b>74</b>		<b>83</b>	<b>74</b>	

In general, the number of genes expressed was, in fact, proportional to the expected number of genes per chromosome. However, it is clear from Figure 3 (p129) that significant differences exist at Chromosome #7 (Gg7) between the chicken and the SR pigeon ( $p=0.0008$ ), and Chromosome 12 (Gg12) between

the chicken and the WC (0.0000). Figure 4 (p130) shows that although the increased copy number observed on Gg7 in the SR is also significant ( $p=0.0314$ ) when compared directly to the WC, this is not the case with Gg12. Although increased expression of genes located on Gg12 in the WC is still observed when compared to the SR, the trend is not significant ( $p=0.0757$ ).

In addition to looking at the relationship between the number of pigeon genes expressed per chromosome, the copy number variation between chromosomes was also analyzed. The chicken was not used as a null expectation, because the expression levels in the chicken were not measured. Therefore, the WC and SR were compared only to each other. Figure 5 (p133) presents the differential expression of transcripts observed between the WC ( $n=346$ ) and SR ( $n=350$ ) cultured aortic cells, and the statistics are reported in Table 16 (p134)..

Figure 5: Comparison of WC and SR Transcript Count Per Chicken Chromosome

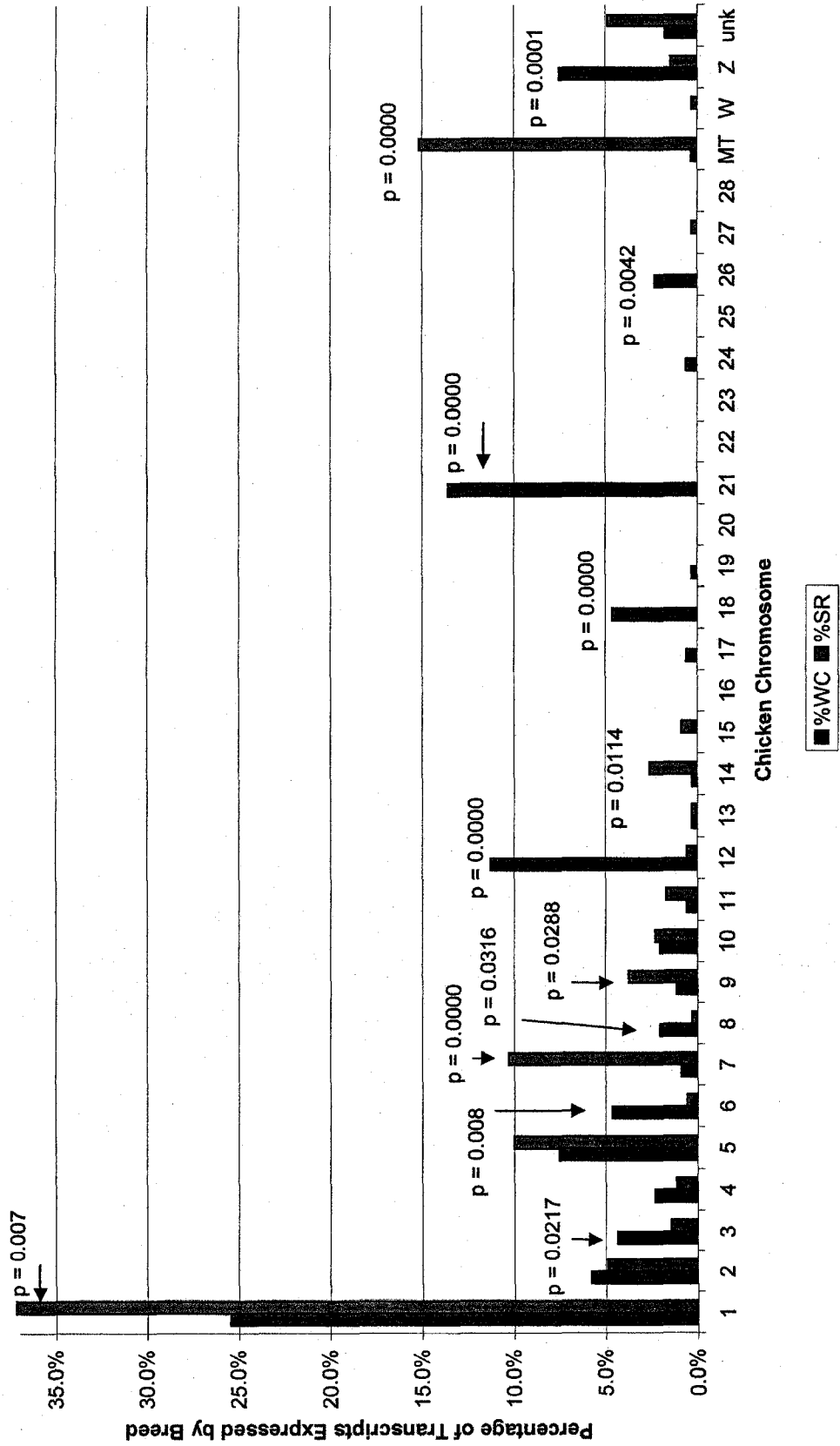


Table 16: Comparison of WC and SR Transcript Count Per Chicken (Gg) Chromosome

Gg	WC	%WC	p=	SR	%SR
1	88	25.4%	0.0007	131	37.4%
2	20	5.8%	0.5873	17	4.9%
3	15	4.3%	0.0217	5	1.4%
4	8	2.3%	0.2361	4	1.1%
5	26	7.5%	0.2463	35	10.0%
6	16	4.6%	0.0008	2	0.6%
7	3	0.9%	0.0000	36	10.3%
8	7	2.0%	0.0316	1	0.3%
9	4	1.2%	0.0288	13	3.7%
10	7	2.0%	0.8115	8	2.3%
11	2	0.6%	0.1597	6	1.7%
12	39	11.3%	0.0000	2	0.6%
13	1	0.3%	0.9935	1	0.3%
14	1	0.3%	0.0114	9	2.6%
15	0	0.0%	0.0844	3	0.9%
16	0	0.0%	na	0	0.0%
17	2	0.6%	0.1543	0	0.0%
18	16	4.6%	0.0000	0	0.0%
19	1	0.3%	0.3142	0	0.0%
20	0	0.0%	na	0	0.0%
21	47	13.6%	0.0000	0	0.0%
22	0	0.0%	na	0	0.0%
23	0	0.0%	na	0	0.0%
24	2	0.6%	0.1543	0	0.0%
25	0	0.0%	na	0	0.0%
26	8	2.3%	0.0042	0	0.0%
27	0	0.0%	0.3197	1	0.3%
28	0	0.0%	na	0	0.0%
MT	1	0.3%	0.0000	53	15.1%
W	0	0.0%	0.3197	1	0.3%
Z	26	7.5%	0.0001	5	1.4%
unk	6	1.7%	0.0212	17	4.9%
	<b>346</b>	<b>100.0%</b>		<b>350</b>	<b>100.0%</b>

Highly significant ( $p=0.0000$ ) differences in transcript numbers per chromosome between breeds are observed at Gg7, Gg12, Gg18, Gg21, and the mitochondria.

The mitochondrial difference was not evident in the previous chromosome analyses because the number of genes differentially expressed was not

statistically significant. In contrast, Figure 5 (p133) clearly demonstrates the up regulation of mitochondrial activity in the SR, which is consistent with the annotation results demonstrating elevated expression of oxidative phosphorylation relative to the WC. Other chromosomes (Gg1, Gg3, Gg6, Gg26, and GgZ) also demonstrate significant differences in gene expression patterns, and will be considered in the final analysis.

As stated previously, transcripts that were successfully identified in the NCBI or DFGI Databases, but were not associated with a known protein product were not annotated. However, they were examined to determine the most likely reading frame and corresponding amino acid sequence. Table 17 (p136) presents the amino acid sequences for 22 of the transcripts matching the WAG 65N20 clone, the most highly expressed differential transcript in the WC. From Table 17, it appears that the consensus sequence for the WAG-65N20 clone is: >SLTTIFTNVS GKPVEDAESV SEHSTIVRAK SSCRSQHNM SQRLCFLGLA LCNVAVSHNTLPLRLPDVAA//. This peptide consists of 70 amino acids, with an estimated molecular weight of 7580Da and an isoelectric point of 7.78. It does not have a match in any of the current databases, so its cellular function remains unknown at this time. Table 18 (p137) presents the most likely amino acid sequence for all of the remaining clones that were identified using the databases, but did not have an annotation product.

Table 17: Predicted Amino Acid Sequences of WAG-65N20 Clones

Plate ID	#bp	Off	#aa	MW	PI	Predicted Amino Acid Sequences of WAG-65N20 Ortholog Transcripts
136SRA07	606	2	101	11768	11.08	FRDSISKVLKQLRFLCHLSIRSCYRLPSVRRSLTIFTNISRKPYEDAERVSESTIV RAKSSORSQCHHNMSCRFPLQLALRTXSHHTPSLRLLMCA S
136NCA03	292	4	70	7611	7.78	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCA01	292	4	70	7611	7.78	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCAB01	294	5	92	10120	9.08	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA LAKSKQNPQMDKVSFSLVMD
224NCA05	292	6	59	6196	7.86	ATTSGSLKGGKVLQETATLQSGLESKAFATCODAVTORSFLPHNSTVSRDALQLIWF A
224NCA05	575	5	100	10869	11.52	RIISGLRNLGKGAIGLRLFLITFGWPKGNMFGKLVPCVQYQYTONHVARVYTLRGLGTPPPRGST GRIRLEYSKPKDPPGGFFSLSSKINST
224NCAE01	284	4	70	7645	7.78	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCAF01	292	1	85	9482	9.19	STRDGLLETLHMOGLFCSLLAKLQDQVLLKGGQOKLFRCKVIRGLESKASATCODAVTORSFLPSCQYCV QRRSLRFHM/CLKHL
224NCAF10	292	1	70	7611	7.78	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCA02	700	2	85	9608	8.91	STRDGLLETLHMOGLFCSLLAKLQDQVLLKGGQOKLFRCKVIRGLESKALATCODAVTOR SFLPSCQYCVQRRSLRFHM/CLKHL
224NCBA02	292	4	70	7612	6.86	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBA07	680	2	94	11292	12.01	RYRLQFPNRLHNPGRAPTRILPTIRSFPPSOSYRISNSHRYCRRRSTILLSPLFRSKNCRNRT TKPNIHMPTRNLLWMLRNLAG
224NCBA12	292	1	70	7611	7.78	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBE04	292	1	70	7597	9.06	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBE07	291	4	71	7805	8.84	SLTTIFTNMPGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBD01	292	4	70	7540	7.72	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBE04	292	1	70	7584	7.78	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBE12	292	1	70	7583	7.78	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBF09	292	1	70	7673	7.77	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBF11	294	1	70	7445	8.77	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBG03	293	4	70	7611	7.78	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBG07	292	4	70	7611	7.78	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBG11	445	4	70	7595	7.98	GLTTIFTNMTGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBG12	292	4	70	7509	7.78	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA
224NCBH07	295	1	93	10476	9.27	SLTTIFTNMSGKPYEDAESVSESTIVRAKSSORSQCHHNMSCRLQFLGLALONAVSNTLPLRLPDVAA PPFRISACORLRKACKTKTHKCKLNP SLVMD

From Table 18, it can be observed that some of the nucleic acid sequences that matched multiple BAC clones represent the same peptide fragment when translated. The regions that matched BAC CH261-3201, -93G5, -71A6, and -15P9 begin with >REVLLPYST LVRPHLEYCV//. There is variation in the codon usage within the larger peptide, explaining why they matched different genomic clones. However, if these sequences had been clustered and reported as one match, the transcript itself would have been significant by Chi Square analysis. The functional pathway of this transcript is not known, but, because some version





Seventy transcripts were not identified in either database. Cluster analysis revealed 35 unique sequences; 11 of which were up regulated in the WC, and 24 that were up regulated in the SR. The most likely open reading frames (ORF) for each of these unidentified, but differentially expressed transcript tags were determined and are presented in Table 19a and 19b.

Table 19a: Predicted Amino Acid Sequences of Unidentified WC Transcripts

ID	#bp	#aa	MW	PI	Predicted Amino Acid Sequences of Unidentified White Carneau Transcripts
C1	235	54	5658	7.80	CVVSCNHGCD KCCSLSLILP AWFTRNPAS GCLSTSVDG LENSRLPPK QNHQ
C4	217	69	7664	8.39	CSCARRFSQT ADACAYGTTT C FNGLASVGH WWSWELDKVS FKHYWKTEES LARAPACAGNERLSSLQGG
C5	434	78	8419	10.41	VTDAAVLAGE SQYSRYLCMW SHLLSMGWRQ RANTGGAGSS INVSSQRTTG KQKQSPVHSS PRPAAGNETD SLHFPRAL
WC1	315	103	11096	5.49	SCSSGGGREVLQVLSLRVLN ECSCIGMMVF CSVFSCRESG PFPSSDSLW EQQMPSLGGS DKKPSQPLW QSFVPDWGHC ATEISLPIV SAWVAIRPH QAY
WC2	519	74	8750	10.17	LETSGFLSR RSRHCNVTXE HXIRYXVXDX XELRYKRVIS NNHXCIWVXX RRTYXGARNXX TPYSXXGPRX XDGS
WC3	56	18	1958	5.24	SQFVSWLPSL LLLTLGPS
	56	18	1784	5.49	VSLLAGFLPS SLLLAPV
	56	18	1944	9.52	TGAKSKEEEG RKPANKLT
WC4	104	34	4048	9.87	CALYKAWLRE HQTASKHLK LLNSFKPGLN PMPW
WC5	410	96	11107	9.43	DLVLTISIRIL ARSREFDELV IHDRCLLTF GIRDIVGMDV AGTYCSSQDW QRRTGITRLOKVTQVCRGT FVFFLMGIPL KRPTFGWQV VPKXW
WC6	593	66	7479	10.22	CDYKQLTPRL KNKTNIDIGMD EELCNSPPPR GPWAKEKLRN SORPYSNTGR XSRGGARVTQ IRPLLG
WC7	473	88	9666	9.17	KNPAVWFRD SSITGMLHKS RIKTCLDSTR AFLALERPLL KAGAGQTPW CSYCRWCTKPKDIASVLDL NSYGMHNLV AVMSLNV
WC8	241	48	5193	9.60	VPQLQWVS VAKKPTNSYL SVLELISLL ALKALQSTR SSGYKYGE

Table 19b: Predicted Amino Acid Sequences of Unidentified SR Transcripts

Tag	#bp	#aa	MW	PI	Predicted Amino Acid Sequences of Unidentified Show Racer Transcripts
C2	184	40	4595	6.73	TKAVLMSHEK HPYGFEGSDC PALNCLPFLM NOANHCKMFL
	234	40	4581	10.06	CRAVTLKKPI RMFLMVHKNC FGLEGNGLK IPMAKFFNPP
C3	223	74	8533	9.80	KTITQSNBPH KPLPQTAGNC CFPQLGTRVW GSYLLPPKLM ANSWRYLDFS LFFFSRQTI FFSNITFNLT PHVH
SR1	253	54	6128	7.84	FGLQWDLGCR SYMNLITCHS VAHQNHCGLE CKTYICFFSN GFEIAKRLNG SLL
SR2	256	54	6128	7.84	FGLQWDLGCR SYMNLITCHS VAHQNHCGLE CKTYICFFSN GFEIAKRLNG SLL
SR3	308	52	5538	8.06	FAHSGTESAP QVDTTSLTVP KRTSDCHSPR RAALHASDKF KPSPVGNNE SC
SR4	337	46	8278	9.30	AVSRRICMGG QLSPLRKLQ QETQNNPRLF YASARLNKSE QSCSSTLLGV SGLGHEYLCPMSSLFAPVYP YSLLQS
SR5	151	44	5038	6.75	KMELFSSGCN RLMTFVFCP VTTRDSQTVT PRAFELHQQL EFAG
SR6	297	86	8967	5.01	IGTSLCLSCP GQREGDIPA YCASSKLMEE GTRLSVQSLI PSPACLSSGN SRPSGFLNIGIVQVGTAMLL SCMLCSYEH LPLCDSS
SR7	262	36	4183	8.60	PNTSLIVDL SRCLGRTKNR TNIPPEISSE PQIIF
SR8	286	54	5838	7.14	CVRLLTLQH PALHHLSPSP ATQNRHDLNS LLSPAEEGAV ARSSLCSHVL FSVL
SR9	266	47	5588	8.36	MOVLAMLIQD NTHLFTFWQL NLRTCPVHYR WLHAGLLCLF HGMFLES
SR10	302	72	7581	9.05	ASPSLSCPC HYPPALHSPG WPCCPAAVSL RKNRTRICFA VKIAGAVPFF VAVGCGSDRLLPFCVCP LK
SR11	450	73	8028	6.43	YHVCICSELG EMCGSVSSD VPPHWPASSR GICGLSPRSP LFTQITCHRK FSLFLCESVSGATALRVWFVS EVH
SR12	520	102	11248	5.55	QATENCFLW IQQSPASHS LLSCLFRVRT TLFYEDSNE CKLPINPKAV PSNFPFHSQVPCPSMDQAA RKAGAPLSS NLELLLTEN NSCFSPSC
SR13	373	87	9948	8.39	KECVLLCKDR HYLQGRKPC AWGDFALQV PTLVPERFIE SLSTLPAADR FTTPDLTPEPEGRFWINCOI RPSGRIDTFT HTKGRVY
SR14	575	64	7605	9.39	FWATLHACSV RPRNVIVQLS YPTAMYCDNL KLENHLYHL VNTKVWENKQ YLAHVLLRVPNKIQ
SR15	299	77	8914	8.71	KKLGFSSVVP LISLTLHYV PDGPHCLIKW WLSTSSFLQ FYNVTKLEEN TYLLSSRMVFCSTLLAKLGF WTLPEMT
SR16	294	71	7661	7.77	NGTQFPVSCD LSFATDLKC EPLKEIKEPR CIYPCSTIRY HCDSTSEAG VLAAKACSA A RFNTRVWIGA S
SR17	352	52	5799	10.17	VSVYFRLSAF NLHYVRPSLL LLSLSTGHPT TDQLLHKLFL TGQKPTVSK SS
	352	52	6025	10.05	MCHGRDVSQS KSTRFTYRWF LNPSEQLVQ QLVSSGMACG QRQKQAWSH IV
SR18	330	63	6920	7.95	EHGAASEASP PGHSHAVSKI SACITRQSL QYQAIGAMFQ WLLQDYAKVK VGSYCPVFCSLYC
SR19	124	41	4228	4.35	DVGDFFSSRL FPAFCSSVPS PAGLAARHQL GDYQPAEISS P
SR20	270	85	9468	10.02	SSAAASPQPS RYSGVFFLSF LFIFFPQQKC VDILQDLFQ AYVSVSPVFG MRKVTPHSPWPFKLGKSNL LPLLHHLGQR QGFIP
SR21	343	100	11687	9.72	YIHEIKFLYL GIFFCKRMF QYRFGGWSK APQNAVLSAS LLSHPVSRQ GDVRTTETLL LKDKSILSN RFTYQLVLR DELELHFSKN TVFPCSSWV
SR22	734	71	7518	9.23	GGLTGRHXAY GXDXXTIGAF KADTKRDLXS HDVSNNGNFX LKERTSVTKV EXTVDNHNHPPPLGGLGVKVR G

Finally, because the reproducibility of RDA has not been reported in the literature, experimental replications 3 & 4 were subsequently analyzed as a data subset because these two RDA experiments were conducted on the same pools of cells. Two plates of clones (n=192) were analyzed in RDA #3 (Plate A, B), and only 1 plate (Plate A) was analyzed in RDA #4 (n=96). The number of transcripts that were duplicated and different between the experiments was counted. Transcripts that were differentially expressed between replicates were further categorized as to whether the transcript was missed or was additive within the comparison. In order to determine how much information is gained by performing additional RDA experiments versus characterizing more clones per individual RDA experiment, three types of comparisons were made and are reported in Table 20 (p140).

The first comparison reflects the plate-to-plate variation that randomly occurs within a single experiment (RDA #3). From Table 20, it appears that approximately 7.3% of the observed variation between plate A and plate B can be explained by the random nature of colony picking. This means that for any given set of 96 clones, one clone will match another clone in the data set 92.8% of the time. Of the total variance, 4.1% represented genes that were missed in the second plate, compared to 3.2% of the genes that were new. The second comparison examined the variation between two given plates resulting from separate RDA experiments. An experimental average of 10.2% variance was observed between RDA #3 and RDA #4. Because 7.3% can be explained by

random picking, at least 2.9% of the variation between the two plates can be explained by the difference in the RDA replicate itself.

Table 20: Technical Replication Analysis

Comparison #1	% Duplicated	% Different	% Missed	% New
<b>3A vs. 3B</b>				
WC	90.9	9.1	5.5	3.6
SR	94.6	5.4	2.7	2.7
<b>Mean:</b>	<b>92.8</b>	<b>7.3</b>	<b>4.1</b>	<b>3.2</b>
<b>Comparison # 2a</b>				
<b>3A vs. 4A</b>				
WC	85.4	14.6	7.3	7.3
SR	94.6	5.4	2.7	2.7
<b>Comparison # 2b</b>				
<b>3B vs. 4A</b>				
WC	92.7	7.3	5.5	1.8
SR	86.5	13.5	8.1	5.4
<b>Mean:</b>	<b>89.8</b>	<b>10.2</b>	<b>5.9</b>	<b>4.3</b>
<b>Comparison # 3a</b>				
<b>3A,B vs. 3A</b>				
WC	98.2	1.8	0.0	1.8
SR	97.3	2.7	0.0	2.7
<b>Comparison # 3b</b>				
<b>3A,B vs. 3B</b>				
WC	94.5	5.5	0.0	5.5
SR	100.0	0.0	0.0	0.0
<b>Mean:</b>	<b>97.5</b>	<b>2.5</b>	<b>0.0</b>	<b>2.5</b>

The third comparison looks at the combined data obtained from two plates representing one experiment (n=192) to the data obtained from either plate alone (n=96). By analyzing 192 clones from an individual experiment rather than 96, the rate of discovery is increased by 2.5%.

## CHAPTER IV

### DISCUSSION

#### **Biological Relevance of Results to Atherogenesis**

##### Determination of Candidate Genes Contributing to Pigeon Atherosclerosis

Multiple perspectives were used to prioritize the contribution of differentially expressed genes to atherosclerotic susceptibility and/or resistance in the pigeon. In the first analysis, individual genes were divided into four tiers of candidacy based on copy number variation and reproducibility as presented in Table 5a-5c (p97). In order to qualify for placement in the first tier, genetic transcripts needed to be differentially expressed in at least two experiments, and be present in high enough copy number to be statistically significant by Chi Square Analysis ( $p \leq 0.05$ ). For the second tier of candidacy, one of two criteria needed to be met. First, transcripts needed to be differentially expressed in at least two experiments, as in Tier 1, but not present in high enough copy to be statistically significant by Chi Square Analysis ( $p \geq 0.05$ ). Second, transcripts were differentially expressed in only one experiment, but were present at high enough levels to be statistically significant by Chi Square Analysis ( $p \leq 0.05$ ). Although each criterion reflects a different perspective, they were coupled in the same tier to avoid the temptation of placing emphasis of copy number over

biological reproducibility, and vice versa. In Tier 3, transcripts were differentially expressed in one experiment, the copy number was >1, yet it was not high enough to be statistically significant by Chi Square Analysis ( $p \geq 0.05$ ). Tier 4 consisted of genes that were isolated one time in one experiment, and were not considered to be major candidates from this perspective. The top candidate genes represented in the first three tiers are presented in Table 21.

Table 21: Candidate Genes Determined by Copy Number and Reproducibility

White Carneau (Susceptible)		Show Racer (Resistant)	
Gene	Protein Product	Gene	Protein Product
<b>Tier 1 Candidate Genes</b>			
ACTB	beta actin	ACTA2	alpha actin
BAC-124L19	unknown	CO1	cytochrome oxidase subunit 1
BAC-138K4	unknown	CYTB	cytochrome b
Contig 1	unknown	FN1	fibronectin type 1
CPSF2	cleavage & polyadenylation specific factor 2	LUM	lumican
CXCL12	chemokine ligand 12	MYLK	myosin light chain kinase
DACH1	dachshund homolog 1c	ND4	NADH subunit 4
DGAT2	diacylglycerol O-acetyltransferase	RPL3	ribosomal protein L3
ENO1	alpha enolase	TPM1	alpha tropomyosin
L27mt	ribosomal protein L27, mitochondrial	16S rRNA	unresolved
LGTN	ligatin		
NAT13	N-acetyltransferase 13		
PSMC2	proteasome, ATPase subunit 2		
RBP7	retinol binding protein		
RPL32	ribosomal protein L32		
RPN1	ribophorin 1		
TNFaIP8	TNFa induced protein 8		
WAG-65N20	unknown		
<b>Tier 2 Candidate Genes</b>			
ANXA2	annexin II	Contig 2	unknown
COL1A2	collagen, alpha 2 type 1	COTL1	coactosin
Contig 4	unknown	FBLN5	fibulin 5
EIF4E	eukaryotic translation initiation factor 4E	MYH11	SMC myosin heavy chain
GPI	glucose phosphate isomerase	PSMC1	proteasome, ATPase subunit 4
HRASLS	HRAS like suppressor	PSMD1	proteasome, non ATPase subunit
MAEA	macrophage erythrocyte attacher	RAB1A	GTP binding protein
NDPK	nucleoside diphosphate kinase	SEC61G	sec 61, gamma subunit
PSMC3	proteasome, ATPase subunit 3		
RPL7	ribosomal protein L7		
RPL7a	ribosomal protein L7a		
SEC61A	sec 61 protein, alpha subunit		
TKT1	transketolase		
TM167	transmembrane protein 167		

Table 21 (continued)

White Carneau (Susceptible)		Show Racer (Resistant)	
Gene	Protein Product	Gene	Protein Product
<b>Tier 3 Candidate Genes</b>			
ADAMSTS15	reprolysin type disintegrin & metalloprotease	BAC-20I24	unknown
ALDH9A1	aldehyde dehydrogenase isoform E3	BAC-32012	unknown
Contig 5	unknown	BAC-46G16	unknown
DCN	decorin	BAC-71A6	unknown
LDHA	lactate dehydrogenase subunit A	C20orf45	unknown
LSm3	small nuclear RNA LSm3	Contig 3	unknown
MGC75678	unknown	FH	fumarate hydratase
NARD1	nardilysin	FLJ13089	unknown
NRH	nuclear receptor hormone	HPC1	histone promoter control 1
Pectnhib	pectinesterase inhibitor	KS5	unknown
PGD	phosphoglucose dehydrogenase	SET	mariner 1 transposase
PSMC3	proteasome, ATPase subunit 3	SLN	sarcolipin
RBMS1	RNA binding motif S1		
RPS3	ribosomal protein S3		
SPON1	spondin 1		

In the second analysis, differentially expressed genes were prioritized according to their physical loci. This analytical step was performed because if differentially expressed genes are clustered together on a single chromosome, they are more likely to be inherited together and possibly respond to a common transcription factor. Individual chromosomes containing significantly more transcriptionally active genes than the corresponding chromosome in the opposite breed were identified using Chi Square Analysis. The graphic results of these comparisons were presented in Figure 3, 4, and 5 (p129, 130, 133) and formed the basis for the tiered ranking system presented in Table 22 (p145). In order to be included in Tier 1, differentially expressed genes must reside on a chromosome that was determined to be “up regulated” in all three statistical comparisons. Tier 2 is comprised of genetic transcripts residing on chromosomes that were “up regulated” in two of the three comparisons. In addition, mitochondrial genes were included in Tier 2 because at least one-third

of them were differentially expressed. This difference was not apparent until the third comparison because of the small number of genes coded by the mitochondria (n=13). Finally, differentially expressed genes were placed in Tier 3 if the chromosomes on which they reside were “up regulated” in one of the three comparisons, and was not W or Z.

Sex-determinant chromosomes were not considered as candidate genes for atherosclerosis for two reasons. First, breeding studies have indicated that disease susceptibility in the WC pigeon is inherited as an autosomal trait, and there is no difference between males and females, only between WC and SR. Second, the number of genes expressed on the W and Z chromosomes is dependent on the gender of the embryos used to culture the aortic cells. This information is usually recorded for *in-vivo* experiments, as once the birds are six weeks of age, gender can be determined during autopsy. However, it is not known how many males (ZZ) and females (ZW) are represented in the cell culture pools used for RDA. Therefore, any conclusion regarding the impact of either chromosome on phenotype is inappropriate. In the future, it is recommended that a blood sample be taken from each embryo sacrificed for cell culture in order to determine the gender<sup>273</sup> and the expected number of alleles per biological pool.

Table 22: Candidate Genes Determined by Chromosomal Position

Gg	Gene	Protein Product	Gene	Protein Product
		<b>White Carneau (Susceptible)</b>	<b>Show Racer (Resistant)</b>	
<b>Tier 1 Candidate Genes</b>				
7	COL5A2	collagen, alpha 2 type 5	ACTR3	actin related protein 3
	RBMS1	RNA binding motif S1	ACVR1	activin/TGFB receptor 1
			FN1	fibronectin type 1
			KS5	unknown
			MYLK	myosin light chain kinase
			NDUFA10	NADH dehydrogenase
			RIKEN 3110009E18	unknown
		ZFP-100	zinc finger protein 100	
<b>Tier 2 Candidate Genes</b>				
12	clone 38f16	unknown		
	COPG	coatomer protein complex gamma		
	LSm3	U6 small nuclear RNA LSm3		
	RPL32	ribosomal protein L32		
	RPN1	ribophorin 1		
	SEC61A	Sec 61 alpha subunit		
	SEC13L	Sec 13L		
	TKT	transketolase		
mt			CYTB	cytochrome B
			CO I	cytochrome oxidase I
			ND4	NADH subunit 4
<b>Tier 3 Candidate Genes</b>				
21	ENO1	enolase, alpha		
	PGD	phosphoglucose dehydrogenase		
	RBP7	retinol binding protein 7		
18	EXOC7	exocyst complex component 7		
	L27mt	mitochondrial RPL27		
	NDPK	nucleoside diphosphate kinase		
9	EIF4E2	eukaryotic translation initiation factor	ACTA2	alpha 2 actin
	HRASLS	HRAS like suppressor	PSMD1	proteasome regulatory subunit
15			chEST380b13	unknown
			FLJ13089	unknown
26	LGTN	ligatin		
5	CPSF2	cleavage & polyadenylation factor 2	ACTR10	actin related protein 10
	LDHA	lactate dehydrogenase subunit A	BAC-16J12	unknown
	SPON	spondin	BAC-20I24	unknown
	PSMC3	proteasome, ATPase subunit 3	HPC1	histone promoter complex 1
	SFFV	spleen focus forming virus	PSMC1	proteasome, ATPase subunit 3
			RNH1	ribonuclease/angiogenesis inhibitor
		FBLN5	fibulin 5	
8	ALDH9A1	aldehyde dehydrogenase isoform E3		
	NARD1	nardilysin		



From Table 22, chromosomes C7, C12, and the mitochondria occupy the top two tiers. This suggests that the genes expressed at these loci are potentially involved in the atherosclerotic phenotype, and should be considered in the final analysis regardless of copy number. The transcripts expressed from Gg1 and Gg3 were placed in Tier 3 – however, the total number of genes residing at these loci was too extensive to record in Table 22. Therefore, a complete list of all the genes that were physically placed on theoretical chromosomes is presented in Appendix D (Table D.31).

Finally, KEGG and Pathway Studio were used to annotate the identified transcripts. As stated in the Methods, Chi Square analysis was not performed at the pathway level. This is because there are obvious differences in gene expression within a pathway between the two breeds that would not be detected. For example, Table 7 (p103) reveals no difference between the WC and SR in Signaling Molecules & Interaction. In the WC, 6.8% of the differentially expressed genes represented this category, and 7.3% of total transcripts. In the SR, 8.6% of genes fit, representing 8.1% of total transcript number. This comparison is not statistically significant. However, when the individual genes within Signaling Molecules & Interaction are examined, it becomes evident that very different signals are being sent out from developing WC and SR aortic smooth muscle cells.

Five general themes of differentially expressed pathways were consistently observed between the WC and SR aortic cells and are presented in Table 23 (p147). These themes will be subsequently discussed in terms of their

potential contribution to atherosclerotic susceptibility and resistance in the pigeon. Briefly, dramatic differences were observed in energy metabolism, phenotypic markers, protein metabolism & modification, the immune response, and in the types of signals, transcription factors, and receptors expressed. The subcategories in each column of Table 23 represent the KEGG and Pathway Studio groups that contributed to each major theme.

Table 23: Five Major Themes as Determined by KEGG and Pathway Studio

Energy Metabolism	Phenotype Markers	Signaling & Receptors	Protein Modification & Metabolism	Immune Response
Glycerol Lipid Metabolism	Cell adhesions	Calcium binding	Amino acid metabolism	Acute phase response
Glycolysis	Cell matrix adhesions	Nucleus component	N-Glycan metabolism	Antigen presenting
Oxidative Phosphorylation	Cell motility	Receptors	Proteasome	B cell receptor pathway
Pentose Phosphate Shunt	Collagen fibril organization	Signal Transduction	Protein catabolism	Leukocyte transendothelial migration
TCA cycle	Cytoskeletal metabolism	Signaling molecules & interactions	Protein folding	Toll receptor pathway
	Cytoskeletal organization	TGFB signaling	Protein kinase	
	ECM metabolism	Transcription Factors	Protein secretion	
	Extracellular component		Protein/ribosome biosynthesis	
	Focal adhesions		Proteolysis	
	Intermediate filament metabolism			
	Proteoglycan metabolism			
	Regulation of actin cytoskeleton			
	Smooth muscle contraction			

A scoring system was developed in order to compile one list of candidate genes derived from all three levels of analysis. To accomplish this objective, a genetic transcript was assigned one point for each contribution to one of the five thematic pathways listed in Table 23. For the candidate genes that were ranked according a system of tiers (Table 21, 22), points were distributed as follows: Tier 1= 3 pts, Tier 2 = 2 pts, Tier 3 = 1 pt, and Tier 4 = 0 pts. Although imperfect, categorizing the 165 unique and differentially expressed genes according to these criteria provided a comprehensive framework for prioritizing the contribution of candidate genes to the susceptible or resistant phenotype in

pigeons. Table 24 (p149) presents the top candidate genes in descending order of score, excluding transcripts that scored less than 3 points.

There are 22 difference products that scored at least 5 points, indicating their relevance in at least two of the analytical perspectives described. The “top 22” candidate genes represented each of the five pathways: phenotype marker (n=6), energy metabolism (n=6), signaling molecules (n=9), protein metabolism (n=7) and the immune response (n=2). The genes involved in protein metabolism were evenly divided between those active in the 26S proteasome (n=3) and the ribosomal proteins (n=3). Since 22 genes represent 30 distinct pathway functions, it is clear that some of the candidates have multiple functions. These genes are of special interest because they represent a potential link between the pathways, and thus are a potential mechanism to connect seemingly divergent events in atherogenesis.

This select group includes MYLK, which is a marker of the contractile phenotype, a signaling molecule, and, as a kinase, plays a role in protein modification in the SR. Myosin light chain kinase was placed on Gg7, which displayed a significant increase in both the number of genes and in the copy number expressed when compared against the WC and the chicken. In addition, MYLK was important in its own right, as it was isolated in two experiments, and in high enough copy number to be considered significant by Chi Square Analysis. Beta actin, in the WC, is a synthetic phenotypic marker and a signaling molecule, but its third role is in the immune system. The presence of COL5A2 also marks a synthetic phenotype in the WC, and,

Table 24: Top Candidate Genes Determined from Multiple Perspectives

Gene	Protein Product	Level 1	Level 2	Energy Metabolism	Phenotype Markers	Signaling Molecules	Protein/AA Metabolism	Immune Response	Total
MYLK	myosin light chain kinase	3	3	0	1	1	1	0	9
FN1	fibronectin type 1	3	3	0	1	1	0	0	8
ACTB	beta actin	3	0	0	1	1	0	1	6
CXCL12	chemokine ligand 12	3	1	0	0	1	0	1	6
ENO1	alpha enolase	3	1	1	0	1	0	0	6
LUM	lumican	3	1	0	1	0	1	0	6
RPL32	ribosomal protein L32	3	2	0	0	0	1	0	6
RPN1	ribophorin 1	3	2	0	0	0	1	0	6
TKT1	transketolase	2	2	1	0	1	0	0	6
CO1	cytochrome oxidase subunit 1	3	2	1	0	0	0	0	6
CYTB	cytochrome B	3	2	1	0	0	0	0	6
ND4	NADH subunit 4	3	2	1	0	0	0	0	6
ACTA2	alpha actin	3	1	0	1	0	0	0	5
COL5A2	collagen, alpha 2 type 5	0	3	0	1	1	0	0	5
DACH1	dachshund homolog 1c	3	1	0	0	1	0	0	5
DGAT2	diacylglycerol O-acetyltransferase	3	1	1	0	0	0	0	5
L27mt	ribosomal protein L27, mitochondrial	3	1	0	0	0	1	0	5
LGTN	ligatin	3	1	0	0	1	0	0	5
NAT13	N-acetyltransferase 13	3	1	0	0	0	1	0	5
PSMC2	proteosome, ATPase subunit 2	3	1	0	0	0	1	0	5
RPL3	ribosomal protein L3	3	1	0	0	0	1	0	5
SEC61A	Sec 61 alpha subunit	2	2	0	0	0	1	0	5
ACTR3	actin related protein 3	0	3	0	1	0	0	0	4
ACVR1	activin/TGFB receptor 1	0	3	0	0	1	0	0	4
ANXA2	annexin II	2	0	0	0	1	1	0	4
COL1A2	collagen, alpha 2 type 1	2	0	0	1	1	0	0	4
DCN	decorin	1	1	0	1	1	0	0	4
EIF4E2	eukaryotic translation initiation factor 4E	2	1	0	0	1	0	0	4
GPI	glucose phosphate isomerase	2	0	1	0	0	0	1	4
KS5	unknown	1	3	0	0	0	0	0	4
MYH11	SMC myosin heavy chain	2	1	0	1	0	0	0	4
NDUFA10	NADH dehydrogenase	0	3	1	0	0	0	0	4
PSMD1	proteosome, non ATPase subunit	2	1	0	0	0	1	0	4
RAB1A	GTP binding protein	2	1	0	0	1	0	0	4
RBMS1	RNA binding motif S1	1	3	0	0	0	0	0	4
RBP7	retinol binding protein	3	1	0	0	0	0	0	4
TNFAIP8	TNFA induced protein 8	3	0	0	0	0	0	1	4
TPM1	alpha tropomyosin	3	0	0	1	0	0	0	4
ZFP-100	zinc finger protein 100	0	3	0	0	1	0	0	4
ADAMTS15	reprolysin type disintegrin & metalloprotease	1	0	0	0	1	1	0	3
CANX	calnexin	0	0	0	0	1	1	1	3
C3AR	C3a anaphylatox receptor	0	1	0	0	1	0	1	3
CCND2	cyclin	0	1	0	1	1	0	0	3
COPG	coatomeer protien complex gamma	0	2	0	0	1	0	0	3
COTL1	coactosin	2	0	0	1	0	0	0	3
CPSF2	cleavage & polyadenylation specific factor 2	3	0	0	0	0	0	0	3
FBLN5	fibulin 5	2	0	0	0	1	0	0	3
FN3	fibronectin type 3	0	1	0	1	0	0	1	3
HRASLS	HRAS like suppressor	2	1	0	0	0	0	0	3
IKBKAP	inhib. Kappa light polypeptide enhancer	0	0	0	0	1	1	1	3
LSM3	small nuclear RNA LSM3	1	2	0	0	0	0	0	3
MAEA	macrophage erythrocyte attacher	2	0	0	0	0	0	1	3
NARD1	nardilysin	1	1	0	0	0	1	0	3
NDPK	nucleoside diphosphate kinase	2	1	0	0	0	0	0	3
NRH	nuclear receptor hormone	1	1	0	0	1	0	0	3
PGD	phosphoglucose dehydrogenase	1	1	1	0	0	0	0	3
PSMC1	proteosome, ATPase subunit 4	2	0	0	0	0	1	0	3
RIKEN 3110009E18	unknown	0	3	x	x	0	x	0	3
RPL7	ribosomal protien L7	2	0	0	0	0	1	0	3
RPL7a	ribosomal protien L7a	2	0	0	0	0	1	0	3
RPS3	ribosomal protein S3	1	1	0	0	0	1	0	3
SEC13L	Sec 13L	0	2	0	0	0	1	0	3
SEC81G	sec 81, gamma subunit	2	0	0	0	0	1	0	3
SPON1	spondin 1	1	0	0	1	1	0	0	3
VIL2	villin 2/ ezrin	0	1	0	1	0	0	1	3

like ACTB and MYLK, can function as a signaling molecule. This particular collagen isoform is also located on the same chromosome as MYLK (Gg7). Enolase is a glycolytic enzyme that normally functions in the cytoplasm as part of glycolysis. However, a shortened transcript codes for tau-crystallin enolase, which has been reported to translocate to the nucleus and act as a transcription factor<sup>263,264</sup>. Finally, the chemokine ligand CXCL12, also known as stromal derived factor 1 (SDF1), plays a major role in the immune system by interacting with many types of cytokines, but also has other signaling functions, many of which are not yet understood.

#### The Five Major Themes Associated with Atherogenesis in the Pigeon

Many of the differentially expressed genes identified in this set of experiments correlate with biochemical pathways and morphological events previously documented in the pigeon. Differences in proteoglycan metabolism, myofilament organization, mitochondrial function, and ATP usage patterns reported in the literature were observed at the genetic level in the current study. In addition, this is the first study to report an immune response in the spontaneous (non-induced) model. The immune factors expressed by the WC precede actual foam cell formation, and the SR cells grown under parallel conditions do not express these same factors.

Theme One: Energy Metabolism. The most striking difference observed between the WC and SR was energy metabolism. The SR cells clearly relied on oxidative phosphorylation, whereas the WC cells generated most of their ATP

from glycolysis and the pentose phosphate shunt, bypassing the mitochondrial reactions altogether. This finding corroborates previous studies that demonstrated mitochondrial abnormalities<sup>145,184</sup>, increased glycolysis<sup>173</sup> and decreased oxidative phosphorylation<sup>170</sup> in the WC relative to the SR. Hajjar and Smith<sup>168</sup> showed that WC pigeons lacked the necessary regulation of NADH transhydrogenation. Inappropriate amounts of NADH in the cell would impact the NADPH/NADH ratio, which, among other cellular events, regulates the synthesis of triacylglycerols (TAG) and fatty acids. Because of this, the authors suggested that a defect in transhydrogenation could explain the enhanced lipid biosynthesis in the WC.

In the current study, WC aortic cells expressed diacylglycerol transferase (DGAT2), a metabolic enzyme that facilitates the final step in TAG synthesis. In further concordance with Hajjar & Smith, both NDUFA10 and ND4 were up regulated in SR, suggesting that the SR smooth muscle cells have maintained their ability to regulate NADH transhydrogenation. Also, up regulated in the SR, were CTYB and COI, providing evidence that deficiency of these protein components of the electron transport chain may be involved in the mitochondrial dysfunction observed in the WC.

The primary function of a mature SMC is contraction, so by necessity this cell type requires more ATP than its non-contractile muscle counterparts. To accomplish this, energy production in muscle cells is compartmentalized in order to optimize ATP delivery to the cellular ATPases<sup>274</sup>. Weiss et al. observed that ATP generated by oxidative phosphorylation was used for

myofibril contraction, whereas glycolysis-produced ATP is used by the membrane and transport ATPase complexes<sup>274</sup>. If this division of labor also occurs in the pigeon, insufficient oxidative phosphorylation could account for the lack of contractile elements identified in the WC. The WC aortic cell may try to compensate by up regulating glycolysis, but this pathway does not produce nearly enough ATP for contraction. In that case, excess ATP could be re-directed into storage, first in triacylglycerols as suggested by DGAT2 expression in the WC, and later in cholesterol esters, as the literature reports. The excess ATP could be used to drive untimely phosphorylation reactions, leading to various abnormalities in cell signaling and enzyme activation states.

The precise reason as to why glycolysis is up regulated in the WC is not clear from these studies. It could be a compensatory measure or a response to a marginally aerobic cell. In either case, the dual effects of one of the glycolytic enzymes identified, enolase, has already been discussed. Glucose phosphate isomerase (GPI) and lactate dehydrogenase A (LDHA), were also identified in the WC, and play a role in cell signaling. An important distinction to note is that the enzymes that participate in oxidative phosphorylation do not signal. The function of the electron transport chain (ETC) is to produce ATP, which in a SMC, is believed to specifically target contraction. One observation that supports the idea that the ATP produced by the ETC is coupled to contraction is that in the SR cell, along with the increased expression of ETC enzymes was the concurrent up regulation of many phenotypic markers denoting a contractile cell.

## Theme Two: Phenotypic Modification. Synthetic and contractile

phenotypes are on opposite ends of a continuum, yet cells can exhibit both features at the same time<sup>50</sup>. This occurs during vascular development, which is considered the “maturation” phase<sup>52</sup>. The phenotype can also switch if the adult tissue is injured, and the vessels need to be repaired. Under these conditions a contractile cell will start expressing synthetic transcripts again<sup>67</sup>. In reality, the phenotype of a SMC population seems to be the net result of the expression ratio of synthetic to contractile elements.

The aortic cells compared in this study were harvested on day 7, which is just prior to actual foam cell development in the WC. At the same time, the resistant SR SMC are moving from the “Maturation Phase” to the “Contractile Phase” as described by Owens et al<sup>52</sup>. This smooth transition is not occurring in susceptible cells because the WC are losing their myofilament structure<sup>145</sup>, and do not fully reach the contractile phase. This initial disorganization of the cytoskeleton in the WC has been reported to occur *prior to* foam cell development in the WC<sup>275</sup> and cell proliferation in culture<sup>276</sup>. Studies in mice have demonstrated that a dramatic reduction in ACTA2, alpha tropomyosin (TPM1), smooth muscle myosin heavy chain (MYH11), and calponin<sup>45</sup> accompanies foam cell formation.

Among the many contractile markers found in the SR, MYLK was the most consistently expressed and scored the most points in the compiled ranking system. Telokin, a specialized protein within the MYLK gene, was also identified. Telokin stabilizes un-phosphorylated myosin filaments so the kinase



domain of the protein can perform its function. It is also believed to play a role in SMC relaxation and calcium sensitivity<sup>61</sup>. Telokin is highly specific to SMC, but is not used as a phenotypic marker because of its low expression levels in the aorta and coronary arteries. The fact that telokin was isolated in the SR aortic cells using RDA demonstrates the sensitivity of the method.

The enzymatic action of MYLK is to phosphorylate myosin regulatory light chain (MYRN)<sup>277</sup>. One copy of MYRN was found in the WC, but it must be phosphorylated in order to then phosphorylate the myosin heavy chain (MYH11). MYH11 performs the actual mechanics of contraction, and is calcium and calmodulin dependent<sup>264</sup>. The status of MYH11 phosphorylation during the early stages of SMC differentiation has been investigated. Despite the marked presence of myosin phosphatase (MP) in the SMC, the myosin itself was heavily phosphorylated<sup>278</sup>. When the action of MYLK was inhibited, even partially, the differentiation process was slowed, or even arrested. The authors were not sure if the inhibitors blocked MYLK itself, or if it blocked the calcium release.

If MYLK were either deficient or inefficient in the WC, ATP generated from oxidative phosphorylation would be unable to support muscle contraction. In response to the decreased energetic needs of the cell, the electron transport chain would down regulate to basal levels, making it appear to be up regulated in the SR. The prolonged lack of contraction may cause (by default) a phenotypic switch, turning on synthetic transcripts and halting differentiation. The SR appears to proceed through the maturation phase into differentiation,

evidenced by the presence of both beta and alpha actin. Two copies of ACTB and 7 copies of ACTA2 suggest that the SR cells were harvested during their phenotypic transition. However, at the same stage of development, the WC only expresses ACTB, evidence that differentiation is not moving forward smoothly. Differences in the alpha/beta actin ratio during lesion formation have also been observed in chickens. It was determined that there was more alpha actin in the aortic SMC tissue than in the developing plaque from the same group of birds<sup>279</sup>.

Alpha actin is a major constituent of the contractile apparatus, and the concentration of ACTA2 and TPM1 in SMC is twice that of skeletal muscles<sup>277</sup>. Unlike ACTB, ACTA2 does not have a double role in cell signaling. Perhaps as the expression of ACTA2 decreases, the parallel increase in ACTB expression signals that the cell is losing contractility. This idea is supported by RDA experiments conducted in 1-3 day old pigeons<sup>280</sup> where ACTA2 and FBLN5 were identified in the WC, suggesting that differentiation is initiated, but then stops. It is possible that the reason ACTB has evolved a secondary immune function is to communicate the presence of a non-muscle phenotype in a muscle cell.

A second characteristic of a synthetic cell type is the development of an extensive extracellular matrix. The ECM is an integrated structure of proteoglycans and collagen fibrils, and the distribution of these molecules and the ligands they bind are believed to determine the integrity of the matrix in the face of metalloproteases and immune assault. The small, leucine rich

proteoglycan, lumican (a keratin sulfate GAG), was found to be significantly up regulated in the SR, whereas in the WC, decorin, alpha collagen type 1 (COL1A2), and collagen type 5 (COL5A2) were dominant proteins in the ECM. Pro alpha collagen, elastin, and gamma actin have been identified in the WC aorta *in vivo*<sup>281</sup>, along with an increase in chondroitin 6 sulfate (C6S). Increases in COL1A2 and ACTB have also been observed in apoE null mice<sup>282</sup>. Enzymes related to C6S were not identified in the current study and the finding of lumican instead, and in the SR, was a surprise. Lumican has been reported to be present in advanced human plaques, and is believed to contribute to the matrix degradation and collagen fibrillogenesis that was also observed in the coronary artery<sup>104,283</sup>. However, lumican was one of eight proteins that were down regulated in response to mechanically induced injury<sup>284</sup>. The authors provided no rationale for the loss of lumican expression in damaged cells, as it was not their research target, and possibly because it could not be explained in the light of conventional wisdom, which states that lumican is associated with an increase in atherosclerosis, not vice versa.

One of the advantages of conducting a discovery-based versus a hypothesis-driven experiment is exemplified by the unexpected identification of lumican as a differentially expressed gene. This is the first study to report the presence of lumican in either pigeon breed, and the increased expression of keratin sulfate in the resistant SR. Most proteoglycan research described for the pigeon has focused on chondroitin sulfate (C6S) in the WC, as it is observed in susceptible regions of the WC aorta as early as six weeks of age. The SR is

commonly used as a control for these studies, as the accumulation of C6S is unique to the WC aorta. However, as genetic transcripts specific for C6S proteoglycans were not identified in this study, the accumulation of C6S in the WC could be a consequence of lumican *not* being expressed. It is possible that lumican confers resistance to the SR by a mechanism that has not yet been described.

Support for the protective nature of lumican comes from the cancer field, where studies have shown that lumican decreased tumor growth and the invasion of melanoma cells. In addition, lumican appeared to increase the apoptosis of malignant cells, causing the authors to suggest that lumican might be “considered an anti-tumor factor from the ECM”<sup>285</sup>. Lumican is also known to bind collagen, thus regulating the interfibrillar spacing<sup>264</sup>. In this capacity, lumican is thought to regulate cell growth and circumference.

As “rounding of the cell” is one of the degenerative events seen in the WC<sup>145,184</sup>, perhaps the appropriate conformation for the SMC is not possible without the framework provided by lumican. Lumican deficient mice are characterized by a very loose connective tissue and collagen structure that parallels Ehlers-Danlos syndrome in humans<sup>286</sup>. Finally, recent studies in the field of arthritis have shown that in a cell containing multiple proteoglycans, one of the metalloproteases, MMP-13, preferentially cleaved fibromodulin and biglycan rather than lumican<sup>287</sup>. This suggests that the presence of lumican in the ECM of the SR aortic cells may cause them to be more resistant to cellular proteases. This type of protection would prevent the matrix from binding and

retaining circulating lipoproteins and cytokines<sup>82</sup>. This possibility is further supported by the fact that both proteases and chemokine ligands were up regulated in the WC cells.

In addition to the differences reported for MYLK, lumican, and the actin isomers, many other phenotypic markers were differentially expressed between the two breeds. In the WC, synthetic proteins such as plastin, periostin, vimentin, spondin, ezrin, and fibronectin type 3 were identified. In the SR, actin related proteins 3 & 10; coactosin, fibronectin type 1, and fibulin 5 were expressed. Interestingly, FBLN5 null mice have displayed exaggerated vascular remodeling accompanied by a thickened intima, revealing a potential role of FBLN5 as an inhibitor of VSMC proliferation and migration<sup>288</sup>. The ability of FBLN5 to suppress tumor growth has also been observed<sup>289</sup>, and, unlike FBLN1, it is down regulated in most cancer cells<sup>290</sup>. In humans, aberrant versions of FBLN5 have been found to actually contribute to tumor development. Finally, as a secretory protein, FBLN5 is believed to mediate cell-cell and cell-matrix interactions. The overall distribution of phenotypic markers in this study clearly demonstrates that the WC and SR cells are not following the same pattern of SMC maturation and differentiation.

Theme Three: Immune Response. The immune response has not been reported as a factor in the spontaneous pigeon model of atherogenesis. Macrophage cells and leukocytes are well documented in the cholesterol-fed pigeon model, where activation is assumed to occur in response to the unnatural diet. In one of these studies, Denholm & Lewis isolated a monocyte

chemoattractant produced by the WC *in vivo* prior to foam cell development<sup>159</sup>, but did not clarify what cell type was expressing the signal. The current study demonstrates that SMC can, and do, signal for an immune response.

Mouse aortic SMC have been observed to become “macrophage-like”. The authors termed this phenomena “Transdifferentiation”, and suggested that the inflammatory changes observed *in vivo* “may not be a simple consequence of cholesterol accumulation”<sup>45</sup>. As there were no major differences in the genes controlling cholesterol metabolism between the cell types of the two pigeons, it appears that something other than lipid accumulation is also signaling the immune system in the WC pigeon.

Chemokine ligand 12 (CXCL12) was highly expressed in the WC aortic cells, yet no receptors were identified. This finding suggests that the cell types that express receptors for, and normally respond to, CXCL12 were not present in the culture system. Further investigation is required *in vivo* to fully characterize this ligand-receptor partnership. Microarray analysis of 1-3 day old WC and SR aortic tissue revealed the increased expression of one type of chemokine receptor in the WC, CXCR4<sup>291</sup>. In a recent description of 23 chemokines and 14 receptors confirmed to function in the chicken, CXCR4 was identified as the putative receptor for CXCL12<sup>292</sup>. A non-specific chemokine receptor has been identified in advanced human lesions, along with Tumor necrosis factor alpha induced protein 2 (TNFaIP2)<sup>104</sup>.

Although expression of TNFa itself was not detected, traces of its activity were present in the WC. The non-specific responder TNFaIP8 was up

regulated, and is reported to have anti-apoptosis activity<sup>264</sup>. The ADAMS family of matrix proteases is also induced by TNF $\alpha$ <sup>82</sup>, and ADAMSTS15 was isolated from the WC cells. In addition, CXCL12 is known to respond to TNF $\alpha$ , along with lipopolysaccharide and interleukin-1, neither of which was identified. Recent proteomic studies<sup>293</sup> have also detected a TNF $\alpha$  induced protein in the WC pigeon.

The immune response in the WC cells was also represented by macrophage erythrocyte attacher protein (MAEA), spleen focus forming virus (SFFV), inhibitor of kappa polypeptide enhancer in B cells (IKBKAP), ezrin, ACTB, and glucose phosphate isomerase (GPI). Two proteins were found to be associated with the inflammatory response: the anaphylatox receptor (C3AR1) in the SR, and a small inducible cytokine (SCYE1) in the WC. The catalyst for the immune response in the WC is not apparent from these experiments, nor is the relationship between the responders. One possibility is that the presence of ACTB in a non-contractile cell may be considered foreign, eliciting a response. However, other differences in signaling are simultaneously observed between the WC and the SR, so THE causative signal remains elusive.

Theme Four: Signaling, Receptors, and Transcription Factors. Cytokine binding and other cellular events induce a myriad of signaling cascades. Of special interest in the WC are the "signal transducers and activators of transcription" (STAT). One member of the STAT family, STAT4, may play a role in VSMC proliferation and has recently been demonstrated to be up regulated in cholesterol fed WC pigeons *in vivo*<sup>294</sup>. In the current study, the

STAT signaling pathway was represented in WC aortic cells by cyclin D2 (CCND2). The cyclins regulate the mitotic cell cycle, and CCND2 is expressed as the cell moves into its replication phase<sup>63</sup>. In a recent genome wide association (GWA) study, two types of cyclin dependent kinases (CDK) were found to be an independent predictor of coronary artery disease (CAD) in humans<sup>295</sup>.

The GWA consortium screened 17,000 individuals with 500,000 markers of seven human diseases. Within the susceptible loci, both CDKs were differentially expressed between those individuals with CAD and the experimentally matched controls. This was the first study to implicate cyclins in heart disease, a risk factor not currently considered during routine screening tests. The fact that CCND2 was up regulated in the susceptible WC aortic cells is very exciting in light of this new discovery, as both of the CDKs identified in humans are regulated by CCND2<sup>262</sup>, and can be studied in the spontaneous pigeon model.

The transforming growth factor beta (TGFB) signaling pathway was also differentially expressed between the WC and SR. Decorin and spondin were isolated in the WC, whereas activin (ACVR1) and Sp1 were expressed in the SR. Sp1 is a transcription factor necessary for SMC differentiation, and may be responsible for successful differentiation in resistant aortic cells. Since decorin and spondin are synthetic proteins, and also respond to TGFB, it appears that TGFB may be involved in the phenotypic differences observed between the WC and SR. It was recently demonstrated by Wagner et al. that TGFB reduced



growth in SR cells (decreased proliferation), but increased growth in WC cells from the comparable site in the celiac region<sup>296</sup>.

TGFB has historically been accused of a pro-atherogenic role because of its ability to promote fibrosis and inhibit endothelial regeneration<sup>297</sup>. More recent studies have shown that TGFB limits atherosclerosis by modulating a number of processes, including lipid accumulation and the inflammatory response. At this point, the mechanism by which TGFB contributes to atherosclerosis remains to be clarified, as there are widely varying dose-dependent effects even within the same experimental system. Activin is a type 1 TGFB receptor, and was expressed in the SR aortic cells. Although many transcripts that are induced by TGFB were identified this study, including fibulin 5, and periostin, the specific ligand for activin in the SR was not determined.

A different receptor, ligatin, was up regulated in the WC. Ligatin has been shown to bind glycoproteins to the cell surface<sup>298</sup> in a calcium dependent manner<sup>299</sup>, possibly protecting itself from trypsin digest. It also retains phosphorylated sugars, allowing hydrolases to bind and degrade the glycosyl groups from core glycoproteins as needed. Histochemical studies have reported a "fuzzy ring" around WC aortic cells prior to foam cell development, but this phenomenon has not been investigated<sup>300</sup>. Ligatin receptor is able to embed in the hydrophobic cell membrane because it contains a domain that binds covalently to palmitate<sup>301</sup>. This association with palmitate suggests a mechanistic link with one of the traditional risk factors, saturated fat. Palmitate is the 16-carbon saturated fat most commonly handled by the body, as it is

prevalent in the American diet and is the point that *de novo* fatty acid synthesis stops. Either of these scenarios would ensure ample palmitate in the cell membrane, allowing more ligatin receptors to embed, thus increasing the retention of extracellular glycoproteins. In addition, ligatin also functions as a transcription factor, suggesting a direct connection between membrane composition and gene expression.

It was very interesting to look at the types of transcription factors that were identified. Subjectively, the SR seemed to express what might be considered "normal" transcription factors (TF) such as the high mobility group proteins (HMG14, HMG2a), and zinc finger proteins (ZFP-100, ZFP-183). HMG-14 specifically facilitates gene transcription by maintaining the chromatin in an accessible conformation. In contrast, the transcription factors isolated from the WC did not seem "normal".

Abnormal transcription in the WC is best exemplified by SFFV. In the current study SFFV was identified in one experiment, ranking it rather low as a candidate gene for susceptibility. However, it has also been shown to be expressed in the 1-3 day old WC *in vivo*<sup>291</sup>. Many studies in chickens have looked at the effects of Herpes, Marek's disease, and Chlamydia on atherosclerosis, but these have not been found to play a causative role in pigeon atherogenesis. Yet, the identification of SFFV in the WC keeps the virus theories alive, because the initiating genetic factor for WC susceptibility is not evident.

In addition to SFFV, tau-crystalline enolase, an isomer of ENO1 that acts like a transcription factor was also active in the WC aortic cells. ENO1, along with GPI, and LDHA have demonstrated transcription activity<sup>263</sup>, and all of these enzymes were up regulated in the WC. It may not have been so surprising that these glycolytic enzymes were expressed at such high levels, but coupled with the fact that these WC cells exhibited no oxidative phosphorylation, the communication between glycolysis and the nucleus seems excessive. LDHA and ENO1 were two of the substrates first identified for the oncogene tyrosine kinase v-src., a compound highly investigated for pharmaceutical purposes in cancer<sup>263</sup>. It would be interesting to know if individuals undergoing this type of chemotherapy demonstrate a decreased risk for atherosclerosis. LDHA can bind to single-stranded DNA (ssDNA) and destabilize the helix. NADH has been shown to inhibit the binding of LDHA to ssDNA. However, given the lack of regulation of the NADPH/NADH ratio in the WC<sup>168</sup>, LDHA could be a de-stabilizing factor in the susceptible breed. This idea is supported by the fact that LDHA was exclusive to the WC; whereas both WC and SR expressed LDHB.

Finally, the over-expression of ENO1 and LDHA was coupled with GPI in the WC cells. GPI is one of the enzymes secreted by cancerous cells to promote proliferation, but the mechanism of this response is not completely clear<sup>168</sup>. However, all of these glycolytic enzymes were up regulated in the WC aortic cells, and their dual role as transcription factors may be relevant in the susceptible phenotype. Inappropriate activation of transcription can have

innumerable effects. If, as is in the case of v-src, a glycolytic enzyme such as enolase is capable of turning on an oncogene: one of two things could happen: The oncogene can in turn activate many genes in multiple signaling pathways to achieve the simultaneous overproduction of growth factors and associated proteins. Or, the activated oncogene can overproduce just one co-activator that coordinately controls expression of all necessary genes for a malignant cell to outstrip its peers in growth<sup>209</sup>. Both of these scenarios may help to explain why so many genes were found to be differentially expressed between the two breeds, when susceptibility to atherosclerosis in the pigeon is caused by a single gene.

In addition to oncogenes, nuclear lipid second messengers also effect cell proliferation by activating many different genes<sup>302</sup>. A non-specific nuclear hormone receptor (NHR) was identified in the WC, as was the increased expression of retinol binding protein (RBP). Retinoic acid is required for many types of NHR activation, and these results may suggest coordinate regulation by a single genetic response element or promoter that impacts downstream gene expression.

Theme Five: Protein Modification & Metabolism. Ribosomal proteins are often considered to be false positives in RDA experiments because they are expressed in such high copy number compared to other cellular products. In the current study, 3 out of 18 false positives were transcribed for the small subunit (S). However, the differential expression of many other ribosomal genes in the pigeon aortic cells was evident in multiple analyses, and appears to be of

significance in this model. Ultrastructural examination of WC aortic cells harvested between 6-8 days indicates a loss of ribosomes<sup>145</sup>. It is possible that the WC is attempting to compensate for ribosomal degradation by increased synthesis. It is suspect that so many transcripts of RPL3 were isolated in the SR, especially since they all came from the same biological pool of birds. It could be an example of PCR preferential amplification, or it could be a true reflection of increased ribosome and protein synthesis in the SR. However, the excessive expression of RPL3 in the SR cannot be explained by the current data, and the difference in distribution of ribosomal proteins between the WC and SR aortic cells will need to be further investigated before a conclusion can be reached.

Protein catabolism was up regulated in the WC in both the expression of proteasome enzymes, and in the expression of non-specific proteases. Proteasomes are distributed throughout eukaryotic cells at a high concentration and cleave peptides in an ATP/ubiquitin dependent manner<sup>264</sup>. The 26S proteasome degrades cytoplasmic and nuclear proteins<sup>303</sup> that are not targeted to the lysosome. One of the 26S ATPase subunits, (PSMC2) has been identified in advanced human atherosclerotic plaques<sup>102</sup>. This subunit was also up regulated in the WC aortic cells, along with PSMC3. In contrast, PSMC1, and PSMD1, the non-ATPase regulatory subunit, were up regulated in the SR. It was interesting that the proteasome maturation protein (POMP) was determined to be a false positive in these experiments, as it seems that different proteasome activities are occurring between the two breeds. This discrepancy

is possibly explained by the presence of ribophorin 1 (RPN1) in the WC, which directly connects the rough endoplasmic reticulum (RER) to the 26S proteasome base<sup>286</sup>, and may be responsible for determining the types of proteins that will be degraded.

Ribophorin is believed to perform a regulatory function in the ER, by tagging improperly folded proteins for degradation by the proteasome. RPN1 is one of many subunits of the oligosaccharide transferase (OST) complex<sup>304</sup> and functions to mediate the transfer of "high mannose oligosaccharides from a dolichol carrier to suitable (asparagine) acceptor sites in nascent chains"<sup>305</sup>. The OST is closely associated with the ER translocon (SEC61) for secretion from the ER. The SEC61 complex is the major site for the integration of both single spanning and transmembrane proteins<sup>306</sup>. As with the ATPase subunits of the proteasome, the SEC61 subunits were also differentially expressed between breeds. SEC61a was present in each breed, but up regulated in the WC, whereas SEC61G was exclusive to the SR. However, the distinct function of each of the SEC61 subunits is not yet understood. Recent experiments have isolated RPN1 as part of the lysosomal membrane protein fraction. Therefore, its glycosyl transferase activity is not restricted to the ER<sup>307</sup> and it is possible that RPN1 also targets proteins to for degradation by the lysosomes, although this remains to be seen.

Although protein glycosylation was clearly up regulated in the WC, and one de-glycosylase enzyme was also found, the major proteoglycan identified in either cell type was lumican in the SR. Although this specific finding was

unexpected, similar observations have been made in WC pigeons at all ages over six months. In one of these experiments, Edwards & Wagner<sup>308</sup> demonstrated less proteoglycan production, but more proteoglycan degradation in the WC than the SR. The authors proposed that the increase of C6S observed in susceptible pigeons may be a result of longer chain length and/or increased hydrolysis, rather than increased synthesis of a core protein such as versican or aggrecan. The glycosylation of native proteins and the subsequent degradation of non-relevant glycoproteins is a very intriguing pathway to investigate in the spontaneous pigeon model, especially in light of the varying resistance of proteoglycans to proteases, and the striking differences observed within the proteasome complex itself.

Finally, the degradation of the ECM by metalloproteases (MMPs) has been well documented in the literature, mostly in terms of plaque stability because of the signaling relationship between MMPs and the immune system. Although investigating an entirely different stage of the disease in this study, proteases were identified in the susceptible aortic cells, suggesting that the degradation begins very early in the pathogenesis. A Disintegrin And Metalloprotease Thrombospondin (ADAMSTS15), nardilysin (NARD1), and a non-specific dipeptidase (CNDP2) were all up regulated in the WC cells. Nardilysin seems to be able to increase cell migration and proliferation by binding to heparin bound growth factors<sup>309</sup>. However, peptidases did not represent the dominant pathway in the WC, and as ECM degradation had not

yet occurred, proteolysis is not believed to be a primary or initiative factor in atherosclerotic susceptibility in the pigeon.

The five major biological themes identified in the RDA experiments represent intermediate phenotypes<sup>214</sup> or sub-phenotypes<sup>46</sup> that are common to many types of disease. For example, some of the genetic transcripts identified in the WC cells susceptible to atherosclerosis were understood only in terms of their effects in cancer or arthritis. It may seem counterintuitive that so many genes were differentially expressed in a monogenic phenotype. But the development of atherosclerosis is gradual, with many cellular changes occurring simultaneously, all of which depend on the transcription of DNA and processing instructions. Overall, there were very few transcripts identified in this study that were capable of cross-talk between the pathways, and these interactions are a logical place to begin integrating the diverse theories of atherogenesis into one coherent story.

Genetic transcripts representing oxidative phosphorylation and contractility were significantly up regulated in the SR cells. In contrast, the synthetic phenotype was indicated in the WC, as well as increased glycolysis. Energy metabolism is tightly coupled to muscle contraction in aortic cells. Therefore, a genetic defect in this partnership might determine susceptibility and/or resistance to atherosclerosis in the pigeon. It is not clear from the RDA results whether myosin light chain kinase (MYLK) cannot perform its role in contraction in WC cells because it does not have the energy, or it is dysfunctional. The transport of ATP between the mitochondria and the



contractile apparatus of the cell does not appear to be implicated in pathogenesis because the expression of one type of transporter, SCL25A5, was not differentially expressed between the two breeds, and no others were identified.

#### Hypotheses of Human Atherogenesis Supported by Pigeon RDA Results

As most of the more common theories of human atherogenesis were described in detail in the Literature Review, the background will not be repeated here. The hypothesis of phenotypic modification received the most supporting evidence in this study. The resistant SR cells were clearly expressing the normal contractile phenotype of a SMC, whereas the WC cells were in the synthetic phase. It remains unclear if contractility does not develop in WC because of inefficient ATP production, a dysfunction in MYLK, or a deficiency of lumican. Regardless, the arrested differentiation observed in the WC aortic cells is one of the first events in pathogenesis.

The presence of active proteases in the WC provides a plausible mechanism to support the response to retention hypothesis. Lumican has been shown to be more resistant to MMP13 degradation. Without the protection of lumican, the matrix between WC cells is vulnerable to degradation, rendering more hydroxyl groups available to bind lipid and inflammatory factors, leading to a destructive cycle. The response to retention hypothesis is also supported by the expression of ligatin and ribophorin in the WC. Ligatin is a cell membrane receptor that actually binds glycoproteins to the cell surface, providing an additional mechanism for retention. Ribophorin glycosylates proteins using a

dolichol carrier, and if these by-products accumulate in the cell, they can aggregate and provide clusters of binding sites for infiltrating LDL or aberrant signaling molecules.

The hypothesis of mitochondrial dysfunction was also supported in this study. The fact that the WC displays no oxidative phosphorylation, and an over reliance on glycolysis makes this hypothesis appealing. It seems fairly clear that the ATP generated in the mitochondria is effectively coupled to muscle contraction in the SR, and not in the WC. Cytochrome b is strongly implicated as a resistance factor, but it cannot be deleted or completely disabled in the WC, or the birds would not survive to reproduce and transmit the susceptible allele. However, if not enough CYTB was produced, or the complex did not transport electrons efficiently, the WC may not be able to make enough ATP to meet the high energy needs of muscle contraction. Unable to perform its fully differentiated function, the WC SMC may slowly revert to a more synthetic and less differentiated phenotype. However, this scenario would also take place if MYLK were deficient, so it remains unclear whether phenotypic modification causes mitochondrial dysfunction, or vice versa.

The response to injury hypothesis was not directly supported or refuted in this study, as injurious perturbation was not a controlled variable in these RDA experiments. The inflammatory response was not invoked, but the WC cells were clearly attempting to signal the immune system with CXCL12, ACTB, and IKBKAP. However, it is not clear what the immune system was being asked to respond to. Experiments need to be designed to directly explore the immune

response to phenotypic change in the SMC, and be conducted in a non-induced model of atherogenesis to remove ambiguity.

Finally, although lipid metabolism was slightly elevated in the WC with the expression of DGAT2, there was no direct evidence that the substrate used for triacylglycerol synthesis was delivered by infiltration. Therefore the Lipid Infiltration Hypothesis was not supported by this study. In fact, given the lack of TCA enzymes and electron transporters in the WC, it is more likely that the anabolic substrate for TAG synthesis was derived from incomplete beta-oxidation, forcing fatty acids into synthesis and storage.

### **The Chicken as a Comparative Genomics Resource for the Pigeon**

Despite the vast difference in genome size, the human genome is closer to that of the chicken than of the mouse<sup>268</sup>. Therefore, comparative avian genomics are likely to provide valuable insight into gene regulation, splicing mechanisms, and other types of functional divergence relevant to human disease. The recent publication of the *Gallus gallus* genome contributed greatly to the 93% identification rate of sequenced clones generated in the current study.

One hundred forty five of the 167 unique genetic transcripts isolated in the pigeon were orthologous to the chicken. The categorical exception was that the mitochondrial genes expressed in the pigeon were more closely related to birds such as the oriental stork, greater flamingo, and turkey vulture. In fact, the

SR transcript of CYTB matched over 100 other avian species before aligning to the chicken. The ND4 gene (NADH subunit 4) was used to analyze the phylogenetic relationship between the pigeon, chicken, and various other birds that were matched during the BLAST analyses. Table 25 lists the avian species selected and their nucleotide-coding region of the ND4 gene within the complete mitochondrial genome (GenBank).

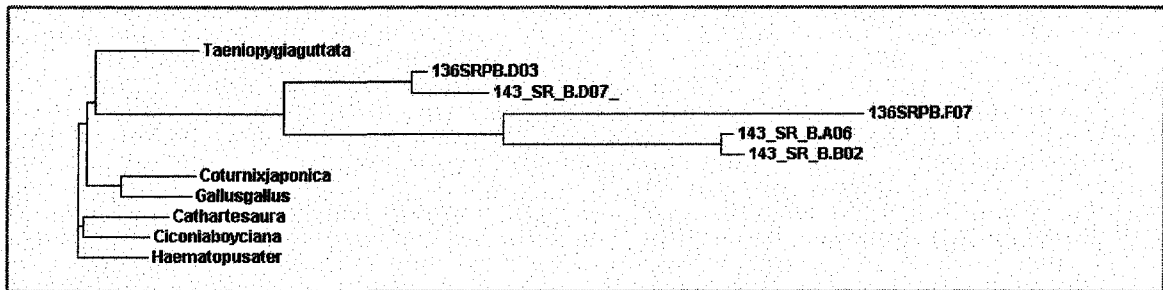
Table 25: Coding Region of Avian ND4 Gene within Mitochondrion

Avian Species	Common name	NADH4
<i>Cathartes aura</i>	Turkey vulture	10,271-11,648
<i>Ciconia boyciana</i>	Oriental stork	12,317-13,694
<i>Coturnix japonica</i>	Common quail	11,405-12,782
<i>Haematopus ater</i>	Blackish oystercatcher	10,218-11,595
<i>Gallus gallus</i>	Red jungle fowl	11,496-12,873
<i>Taeniopygia guttata</i>	Zebra finch	10,278-11,655

CLUSTALW was used to align the six nucleotide sequences from GenBank with five different ND4 transcript tags isolated from the Show Racer in the current experiments. The resultant Phylogram is depicted in Figure 6 (p174)

From Figure 6, it can be seen that the chicken is most closely related to the quail, as are the turkey vulture and oystercatcher. The SR transcripts are obviously related to each other, although not congeneric, and are in fact, more homologous to the zebra finch than the chicken. This corresponds with a phylogenetic tree generated using 12S/16S rRNA and reported by Van Tuinenn

Figure 6: Avian ND4 Phylogram (CLUSTALW)



produced by Pimentel-Smith et al<sup>311</sup> using a proteoglycan gene, the chicken was very closely related to the guinea fowl, quail, and turkey, demonstrating approximately 97% homology at the nucleotide level. In contrast, the pigeon and chicken shared only 30% homology.

These phylogenetic comparisons reveal that despite the observed synteny within avian species, there is still a good deal of nucleotide diversity at the level of the individual genes. From the results of the current study, it would appear that the majority of this diversity is mitochondrial, as 147 out of 165 transcripts were identified using the chicken, and none represented this organelle. Mitochondrial comparisons between the pigeon and the chicken are therefore not appropriate, as limited homology is expected. However, based on the NADH4 transcripts characterized from the SR pigeon, the zebra finch will help bridge this gap. The zebra finch genome project is well underway at the Arnold lab at UCLA and, with continual annotation updates to the chicken genome, will provide an invaluable resource for further pigeon studies that remain dependent on comparative avian genomics.

## Limitations to Data Interpretation

### Inherent Limitations of RDA

Differentially expressed RNA transcripts will only be captured using the described method if the cDNA product contains at least two internal Dpn II restriction sites. Therefore, although finding a transcript is evidence that the gene is active, the reverse is not true<sup>271</sup>. Lack of identification does not mean the gene is not present; it could be that a biologically relevant and differentially expressed transcript was isolated in the RNA extraction, but did not have an amplifiable restriction fragment. This is expected to be the case in approximately 10-15% of mammalian genes<sup>312</sup>. Until the restriction profile is known for the avian counterparts, it can be assumed that only 85-90% of the target genes will be identified using RDA in the pigeon. One way to increase the number of cDNA transcripts in the final difference products is to repeat the experiment with a different four-cutter enzyme. In addition, if the same restriction site is included in the primers used to generate cDNA as is used to generate amplicons; the potential to capture amplifiable fragments increases.

The number of difference products isolated does not directly correlate to the original copy number of the RNA transcript. This is because the number of Dpn II sites and the efficiency of each restriction step determine the final number of fragments produced. In some cases, such as SMC myosin heavy chain, multiple restriction fragments, representing distinct regions of the original transcript were isolated. This can be confusing because it is not clear whether

the sequence tags came from one full-length transcript of myosin heavy chain, or from multiple copies of myosin heavy chain that happened to be randomly picked. In contrast, only one fragment was isolated from ribophorin, yet this one fragment was picked 16 times. To avoid over emphasizing the influence of one gene, restraint is required when analyzing copy number variation with Chi Square Analysis. It is possible that differences in copy number are a result of preferential amplification of select transcripts during PCR<sup>230</sup>. It is also possible that observed variation in copy number reflects general differences in RNA stability, transcription, and translation levels of diverse proteins unrelated to the disease. On the other hand, the number of transcripts identified might be a true reflection of the atherosclerotic susceptible/resistant phenotype in the pigeon.

Finally, the fact that a gene is present in low copy number does not mean it is less important than a gene present at high copy levels. One of the advantages of RDA is the ability to capture rare cellular messages. It is possible that one genetic transcript could be responsible for eliciting a multitude of responses in other genes and on other chromosomes. In addition, the Chi Square statistical test fails to recognize real biological differences when the sample number is less than five<sup>271</sup>. For these reasons, although copy number was recorded, and was an important aspect of the analyses, it was not the only perspective used to determine genetic differences between the WC and SR.

## Experimentally Introduced Limitations

Although every effort was made to keep the parameters consistent throughout all of the RDA experiments, the availability of cell culture material and the changes made in antibiotic selection could impact repeatability. Table 26 presents the experimental variables potentially impacting gene expression in the WC and SR aortic cells.

Table 26: Cell Culture and RDA Experimental Parameters

Cell Culture Parameters	RDA Rep #1		RDA Rep #2		RDA Rep #3		RDA Rep #4	
	WC	SR	WC	SR	WC	SR	WC	SR
RNA Extraction Exp#	135	117	133	141	149	141	149	141
Number of Embryos	5	nr	nr	nr	nr	nr	nr	nr
Days Grown	8	7	7	7	7	7	7	7
Relative Cell Growth	Excellent	Very Good	Very Good	Excellent	Excellent	Excellent	Excellent	Excellent
Harvest Date	6/6/04	9/22/04	6/1/04	6/29/04	9/30/05	6/29/04	9/30/05	6/29/04
BSS Lot	6/2/04	nr	4/3/04	6/22/04	9/9/04	6/22/04	9/9/04	6/22/04
CEE50 Lot	6/5/03	nr	6/5/03	6/5/03	8/5/04	6/5/03	8/5/04	6/5/03
Medium	5/24/02	nr	5/24/02	6/22/04	6/22/04	6/22/04	6/22/04	6/22/04
Horse Serum Lot	5/23/03	nr	5/23/03	5/23/03	6/17/04	5/23/03	6/17/04	5/23/03
1% Ethanol Present?	Yes	No	No	No	No	No	No	No
RDA Parameters	RDA Rep #1		RDA Rep #2		RDA Rep #3		RDA #4	
	WC	SR	WC	SR	WC	SR	WC	SR
Driver Yield (ug)	41.75	30.75	63.75	46.75	36.08	37.57	40.95	49.72
DP1 Hybridization Time (hrs)	26	26	25	25	24	24	24	24
DP1 Yield (ug)	22.25	33.00	22.75	29.00	6.00	12.90	16.65	14.15
DP2 Hybridization Time (hrs)	25	25	23	23	23.5	23.5	23.5	23.5
DP2 Yield (ug)	23.25	14.75	24.25	19.25	17.50	16.25	15.50	18.75
DP3 Hybridization Time (hrs)	25	25	26.5	26.5	24.5	24.5	24.5	24.5
DP3 Yield (ug)	36.70	28.90	30.00	32.75	21.50	7.00	10.25	13.25
Antibiotic Selection	AMP	AMP	AMP	AMP	AMP/TET	AMP/TET	AMP/TET	AMP/TET
White cfu/mL	44,000	60,000	41,000	18,000	nr	nr	nr	nr

It is critical that the two cell populations be grown under identical conditions and harvested at similar densities<sup>313</sup>. This is because the number of cells will impact the number of alleles expressed at any given moment in time. However, *in vivo*, the WC aorta is hypocellular compared to the SR, having a larger lumen<sup>281,296</sup>. For this reason, it seemed more important to harvest the cells at similar points in development, so RNA was extracted just as the cells



formed a monolayer. Typically this monolayer forms in seven days, but in the WC cells used for RDA #1, 8 days were required.

Tissue samples should also be as closely matched as possible, and when experiments require the processing of cells grown *in vitro*, it is best to split a single culture for selected treatment to generate the driver and tester, rather than using separate cultures<sup>313</sup>. Our lab routinely follows this guideline when investigating different stages or enzyme effects within one breed. However, by necessity, WC and SR cultures are prepared at separate times so their phenotypes remain distinct. It is clear from Table 26 that none of the cultured cells used for the RDA comparisons were perfectly matched. RDA Experiment #2 appears to be the best matched in terms of the relative growth of the cells and the dates the media reagents were made, but the impact of this on the gene expression not known.

Although changing environmental conditions may have impacted the overall gene expression profile, there were some transcripts, such as cytochrome B, cytochrome oxidase II, and LDHB that were consistently isolated in all four RDA experiments. Because atherosclerosis in pigeons is an autosomal recessive trait, it was assumed that variations in gene expression *between* experiments were a result of the different environments, rather being a function of the pathology itself.

## Reproducibility & Efficacy of RDA for Identifying Differentially Expressed Genes in the Pigeon

One technical replicate (Table 20, p138) was performed in order to guide the direction of future experiments. For a proper statistical comparison to be made, the same number of clones from each experiment should have been sequenced and characterized. However, determining the reproducibility of the RDA method was not one of the primary research objectives. From the few comparisons that were made, it appears that approximately 2.9% more genes were identified by doing a technical replicate, which is similar to what was gained by doubling the number of clones analyzed in one experiment (2.5%). Therefore, there is no advantage to performing technical replicates from the same biological sample. In fact, one author has stated that generating excess technical information is "akin to studying the difference in heights of the two sexes by repeatedly measuring one man and one woman"<sup>314</sup>. Resources would be better spent either maximizing what can be learned from one individual experiment by picking more colonies, performing biological repetitions, or repeating the experiment with a primer set designed around an alternate restriction enzyme.

## CONCLUSIONS

Genes representing several biochemical pathways were distinctly different between the aortic cells from susceptible (WC) and resistant (SR) pigeons. These differences include:

1. Cells from the susceptible aortas express genes controlling a synthetic phenotype while those from resistant aortas produce a contractile phenotype.
2. Susceptible aortic cells derive their energy from glycolysis and the pentose phosphate shunt in contrast to resistant aortic cells, which rely on oxidative phosphorylation. This indicates that the beta-oxidation of lipids is limited in the WC, while lipid synthesis is enhanced.
3. Resistant aortic cells appear to be more effective at organizing and maintaining an intact extracellular matrix and internal cytoskeleton, whereas susceptible cells do not develop a supportive structural foundation.
4. In general, proteasome degradation is enhanced in susceptible cells while protein synthesis predominates in the resistant aortic cells. There are also dramatic differences between the individual ribosomes being used by each cell type for protein translation.
5. Although the synthesis of core proteoglycans was increased in the resistant cells, protein glycosylation itself predominated in the susceptible cells. In addition, the susceptible cells expressed a unique glycoprotein receptor/transcription factor.
6. Consistent with the above generalizations, genes expressed by susceptible cells in the nucleus and endoplasmic reticulum are different than those of resistant cells. This suggests that there are cellular differences at the level of transcription and translation.

These conclusions support the atherogenic hypotheses of phenotypic modification and mitochondrial dysfunction of smooth muscle cells, and, to a lesser degree, the response to retention hypothesis as contributing factors in pigeon atherosclerosis. A causative role of inflammation was not supported by these experimental results. However, this is the first study to observe an immune response in a non-induced, non-hyperlipidemic model of atherosclerosis, and further studies are needed to determine the cellular triggers that are eliciting the immune response. The current results do suggest that the decreasing ratio of alpha/beta actin that precedes foam cell development in the WC may be enough to signal an immune response.

Because oxidative phosphorylation is so tightly coupled to smooth muscle cell contraction, and the time scale of in-vitro cell growth is so compressed, it is not obvious whether insufficient ATP synthesis is preventing the WC aortic cell from performing its contractile function, causing differentiation to slow down; or, if the lack of functional contractile elements are causing the mitochondrial synthesis of ATP to down-regulate. In order to determine the causative sequence of events, parallel RDA experiments are being conducted on the celiac bifurcation in one-day and six-week old pigeons of each breed. The pending results of these experiments should facilitate the placement of differentially expressed genes in a chronological context during atherogenesis.

## LIST OF REFERENCES

1. Funke, H. & Assmann, G. Strategies for the assessment of genetic coronary artery disease risk. *Curr Opin Lipidol* **10**, 286-291 (1999).
2. Fenton, J. I. & Hord, N. G. Stage matters: choosing relevant model systems to address hypotheses in diet and cancer chemoprevention research. *Carcinogenesis* **27**, 893-902 (2006).
3. Bassiouny, H., Zarins, C., Kadowaki, M. & Glagov, S. Hemodynamic stress and experimental aortoiliac atherosclerosis. *J Vasc Surg* **19**, 426-434 (1994).
4. Kjaernes, M., Svindland, A., Walloe, L. & Wille, S. Localization of early atherosclerotic lesions in an arterial bifurcation in humans. *Acta Pathol Microbiol Scand [A]* **89**, 35-40 (1981).
5. Stary, H. Evolution and progression of atherosclerotic lesions in coronary arteries of children and young adults. *Arterioscler. Thromb. Vasc. Biol.* **9**, 19-32 (1989).
6. Napoli, C., de Nigris, F., Welch, J.S., Calara, F.B., Robert O. Stuart, R.O. et al. Maternal hypercholesterolemia during pregnancy promotes early atherogenesis in LDL receptor-deficient mice and alters aortic gene expression determined by microarray. *Circulation* **105**, 1360-1367 (2002).
7. Getz, G. S. When is atherosclerosis not atherosclerosis? *Arterioscler. Thromb. Vasc. Biol.* **20**, 1694 (2000).
8. Munro, J. & Cotran, R. The pathogenesis of atherosclerosis: atherogenesis and inflammation. *Lab Invest* **58**, 249-261 (1988).
9. Wissler, R., Hiltcher, L., Oinuma, T. & Group, P. R. *The Lesion of Atherosclerosis in the Young: From Fatty Streaks to Intermediate Lesions* (eds. Fuster, V., Ross, R. & Topol, E.) Lippincott-Raven Publishers, Philadelphia, 1996.
10. Balis, J. U., Haust, M. D. & More, R. H. Electron-microscopic studies in human atherosclerosis: cellular elements on aortic fatty streaks. *Exp Mol Path* **3**, 511-525 (1964).
11. Ross, R. & Glomset, J. A. Atherosclerosis and the arterial smooth muscle cell. *Science* **180**, 1332-1339 (1973).
12. Mosse, P., Campbell, G., Wang, Z. & Campbell, J. Smooth muscle phenotypic expression in human carotid arteries. I. Comparison of cells from diffuse intimal thickenings adjacent to atheromatous plaques with those of the media. *Lab Invest* **53**, 556-562 (1985).
13. Mosse, P., Campbell, G. & Campbell, J. Smooth muscle phenotypic expression in human carotid arteries. II. Atherosclerosis-free diffuse intimal thickenings compared with the media. *Arterioscler. Thromb. Vasc. Biol.* **6**, 664-669 (1986).

14. Gabbiani, G., Kocher, O. & Bloom, W. S. Actin expression in smooth muscle cells of rat aortic intimal thickening, human atheromatous plaque, and cultured rat aortic media. *J. Clin Invest* **73**, 148-152 (1984).
15. Katsuda, S. & Okada, Y. Vascular smooth muscle cell migration and extracellular matrix. *J Atheroscler Thromb* **1 Suppl 1**, S34-S38 (1994).
16. Adelman, S. & St. Clair, R. Lipoprotein metabolism by macrophages from atherosclerosis-susceptible White Carneau and resistant Show Racer pigeons. *J. Lipid Res.* **29**, 643-656 (1988).
17. Strong, J. P., Malcom, G.T., McMahan, C.A., Tracy, R.E., Newman W.P. III, et al. Prevalence and extent of atherosclerosis in adolescents and young adults: implications for prevention from the pathobiological determinants of atherosclerosis in youth study. *JAMA* **281**, 727-735 (1999).
18. Xu, Q. Mouse Models of arteriosclerosis: from arterial injuries to vascular grafts. *Am. J. Pathol.* **165**, 1-10 (2004).
19. Knowles, J.W. & Maeda, N. Genetic modifiers of atherosclerosis in mice. *Arterioscler. Thromb. Vasc. Biol.* **20**, 2336-2345 (2000).
20. Zhang, S., Reddick, R., Piedrahita, J. & Maeda, N. Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E. *Science* **258**, 468-471 (1992).
21. Nakashima, Y., Fujii, H., Sumiyoshi, S., Wight, T. N. & Sueishi, K. Early human atherosclerosis: accumulation of lipid and proteoglycans in intimal thickenings followed by macrophage infiltration. *Arterioscler. Thromb. Vasc. Biol.* **27**, 1159-1165 (2007).
22. VanderLaan, P. A. & Reardon, C. A. Thematic review series: the immune system and atherogenesis. The unusual suspects: an overview of the minor leukocyte populations in atherosclerosis. *J. Lipid Res.* **46**, 829-838 (2005).
23. Gibbons, G. H., Liew, C.C., Goodarzi, M.O., Rotter, J.I., Hsueh, W.A. et al. Genetic markers: progress and potential for cardiovascular disease. *Circulation* **109**, 47-58 (2004).
24. Ridker, P. New clue to an old killer: inflammation and heart disease. *Nutrition Action* **27**, 3-5 (2000).
25. Tsimikas, S., Willerson, J. T. & Ridker, P. M. C-Reactive protein and other emerging blood biomarkers to optimize risk stratification of vulnerable patients. *J. Am. Coll. Cardiol.* **47**, 19-31 (2006).
26. Shao, B., Oda, M., Oram, J. & Heinecke, J. Myeloperoxidase: an inflammatory enzyme for generating dysfunctional high density lipoprotein. *Curr Opin Cardiol* **21**, 322-328 (2006).
27. Visvikis-Siest, S. & Marteau, J. Genetic variants predisposing to cardiovascular disease. *Curr Opin Lipidol* **17**, 139-151 (2006).
28. Davies, P. Vascular cell interactions with special reference to the pathogenesis of atherosclerosis. *Lab Invest* **55**, 5-24 (1986).
29. Ross, R. The pathogenesis of atherosclerosis--an update. *N. Engl. J. Med.* **314**, 488-500 (1986).

30. Steinberg, D. A critical look at the evidence for the oxidation of LDL in atherogenesis. *Atherosclerosis* **131**, S5-S7 (1997).
31. Wustner, D., Mondal, M., Tabas, I. & Maxfield, F. Direct observation of rapid internalization and intracellular transport of sterol by macrophage foam cells. *Traffic* **6**, 396-412 (2005).
32. Khalil, M. F., Wagner, W. D. & Goldberg, I. J. Molecular interactions leading to lipoprotein retention and the initiation of atherosclerosis. *Arterioscler Thromb Vasc Biol* **24**, 2211-2218 (2004).
33. Niinikoski, J., Heughan, C. & Hunt, T. Oxygen tensions in the aortic wall of normal rabbits. *Atherosclerosis* **17**, 353-359 (1973).
34. Bernal-Mizrachi, C., Gates, A.C., Weng, S., Imamura, T., Knutsen, R.H. et al. Vascular respiratory uncoupling increases blood pressure and atherosclerosis. *Nature* **435**, 502-506 (2005).
35. Chobanian, A. V. & Manzur, F. Metabolism of lipid in the human fatty streak lesion. *J. Lipid Res.* **13**, 201-206 (1972).
36. Williams, K. J. & Tabas, I. The response-to-retention hypothesis of early atherogenesis. *Arterioscler Thromb Vasc Biol* **15**, 551-561 (1995).
37. O'Brien, K. D., Olin, K.L., Alpers, C.E., Chiu, W., Ferguson, M. et al. Comparison of apolipoprotein and proteoglycan deposits in human coronary atherosclerotic plaques: colocalization of biglycan with apolipoproteins. *Circulation* **98**, 519-527 (1998).
38. Skalen, K., Gustafsson, M., Rydberg, E.K., Hulten, L.M., Wiklund, O. et al. Subendothelial retention of atherogenic lipoproteins in early atherosclerosis. *Nature* **417**, 750-754 (2002).
39. Pentikainen, M.O., Öörni, K., Lassila, R. & Kovanen, P.T. et al. Human arterial proteoglycans increase the rate of proteolytic fusion of low density lipoprotein particles. *J. Biol. Chem.* **272**, 7633 (1997).
40. Reardon & Getz. Mouse models of atherosclerosis. *Curr Opin Lipidol* **12**, 167-173 (2001).
41. Libby, P., Ridker, P. M. & Maseri, A. Inflammation and atherosclerosis. *Circulation* **105**, 1135-1143 (2002).
42. Steinberg, D. Atherogenesis in perspective: hypercholesterolemia and inflammation as partners in crime. *Nat Med* **8**, 1211-1217 (2002).
43. Hansson, G. & Libby, P. The immune response in atherosclerosis: a double-edged sword. *Nat Rev Immunol* **6**, 508-519 (2006).
44. Rong, J.X., Shen, L., Chang, Y.H., Richters, A., Hodis H.N. et al. Cholesterol oxidation products induce vascular foam cell lesion formation in hypercholesterolemic New Zealand white rabbits. *Arterioscler. Thromb. Vasc. Biol.* **19**, 2179-2188 (1999).
45. Rong, J. X., Shapiro, M., Trogan, E. & Fisher, E. A. Transdifferentiation of mouse aortic smooth muscle cells to a macrophage-like state after cholesterol loading. *PNAS* **100**, 13,531-13,536 (2003).

46. Loscalzo, J. *Molecular Mechanisms of Atherosclerosis*. Taylor & Francis, London, New York, 2005.
47. Tedgui, A. & Mallat, Z. Cytokines in atherosclerosis: pathogenic and regulatory pathways. *Physiol Rev* **86**, 515-581 (2006).
48. Benditt, E. P. Evidence for a monoclonal origin of human atherosclerotic plaques and some implications. *Circulation* **50**, 650-652 (1974).
49. Virmani, R., Kolodgie, F. D., Burke, A. P., Farb, A. & Schwartz, S. M. Lessons from sudden coronary death: a comprehensive morphological classification scheme for atherosclerotic lesions. *Arterioscler. Thromb. Vasc. Biol.* **20**, 1262-1275 (2000).
50. Schwartz, S., Campbell, G. & Campbell, J. Replication of smooth muscle cells in vascular disease. *Circ. Res.* **58**, 427-444 (1986).
51. Thyberg, J., Hedin, U., Sjolund, M., Palmberg, L. & Bottger, B. Regulation of differentiated properties and proliferation of arterial smooth muscle cells. *Arterioscler. Thromb. Vasc. Biol.* **10**, 966-990 (1990).
52. Owens, G. K. in *Atherosclerosis and Coronary Artery Disease* (eds. Fuster, V., Ross, R. & Topol, E.) Lippincott-Raven, Philadelphia, 1996.
53. Schwartz, S. M. & Murry, C. Proliferation and the monoclonal origins of atherosclerotic lesions. *Ann Rev Med* **49**, 437-460 (1998).
54. Doherty, T. M., Shah, P. K. & Rajavashisth, T. B. Cellular origins of atherosclerosis: towards ontogenetic endgame? *FASEB J* **17**, 592-597 (2003).
55. Zalewski, A., Shi, Y. & Johnson, A. G. Diverse origin of intimal cells: smooth muscle cells, myofibroblasts, fibroblasts, and beyond? *Circ. Res.* **91**, 652-655 (2002).
56. Hiltunen, M. O. Turunen, M.P., Häkkinen, T.P., Rutanen, J., Hedman, M. et al. DNA hypomethylation and methyltransferase expression in atherosclerotic lesions. *Vasc Med* **7**, 5-11 (2002).
57. Kaneda, A., Takai, D., Kaminishi, M., Okochi, E. & Ushijima, T. Methylation-sensitive representational difference analysis and its application to cancer research. *Ann. N.Y. Acad. Sci.* **983**, 131-141 (2003).
58. Harris, J. D., Graham, R., Schepelmann, S., Stannard, A.K., Roberts, M.L. et al. Acute regression of advanced and retardation of early aortic atheroma in immunocompetent apolipoprotein-E (apoE) deficient mice by administration of a second generation [E1-, E3-, polymerase-] adenovirus vector expressing human apoE. *Hum. Mol. Genet.* **11**, 43-58 (2002).
59. Malinow, M. Experimental models of atherosclerosis regression. *Atherosclerosis* **48**, 105-118 (1983).
60. Hadjiisky, P., Bourdillon, M. & Grosogoeat, Y. [Natural history of the regression of atherosclerosis: from animal models to men]. *Arch Mal Coeur Vaiss* **81**, 1411-1417 (1988).
61. Owens, G. K., Kumar, M. S. & Wamhoff, B. R. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiol Rev* **84**, 767-801 (2004).



62. Tyson, K. L., Weissberg, P. L. & Shanahan, C. M. Heterogeneity of gene expression in human atheroma unmasked using cDNA representational difference analysis *Physiol Genomics* **9**, 121-130 (2002).
63. Gizard, F., Amant, C., Barbier, B., Belloc, S., Robillard, R. et al. PPAR $\alpha$  inhibits vascular smooth muscle cell proliferation underlying intimal hyperplasia by inducing the tumor suppressor p16INK4a. *J. Clin. Invest.* **115**, 3228-3238 (2005).
64. Campbell, J., Popadyne, L., Nestel, P. & Campbell, G. Lipid accumulation in arterial smooth muscle cells. Influence of phenotype. *Atherosclerosis* **47**, 279-295 (1983).
65. Worth, N., Rolfe, B., Song, J. & Campbell, G. Vascular smooth muscle cell phenotypic modulation in culture is associated with reorganisation of contractile and cytoskeletal proteins. *Cell Motil Cytoskeleton* **49**, 130-145 (2001).
66. Worth, N. F., Campbell, G. R. & Rolfe, B. E. A role for rho in smooth muscle phenotypic regulation. *Ann. N.Y. Acad. Sci.* **947**, 316-322 (2001).
67. Shanahan, C. M. & Weissberg, P. L. Smooth muscle cell heterogeneity: patterns of gene expression in vascular smooth muscle cells In Vitro and In Vivo. *Arterioscler. Thromb. Vasc. Biol.* **18**, 333-338 (1998).
68. Hendrix, J. A., Wamhoff, B.R., McDonald, O.G., Sinha, S., Yoshida, T. et al. 5' CArG degeneracy in smooth muscle  $\alpha$ -actin is required for injury-induced gene suppression in vivo. *J. Clin. Invest.* **115**, 418-427 (2005).
69. Suzuki, T., Nagai, R. & Yazaki, Y. Mechanisms of transcriptional regulation of gene expression in smooth muscle cells. *Circ Res* **82**, 1238-1248 (1998).
70. Ordoas, J. & Shen, A. Genetics, the environment, and lipid abnormalities. *Curr Cardiol Rep* **4**, 508-513 (2002).
71. Lichter, P. New Tools in Molecular Pathology. *J. Mol Diag* **2**, 171-173 (2000).
72. Scheckhuber, C. Mitochondrial dynamics in cell life and death. *Sci. Aging Knowl. Environ.* **2005**, 36 (2005).
73. Palinski, W. & Napoli, C. The fetal origins of atherosclerosis: maternal hypercholesterolemia and cholesterol-lowering or antioxidant treatment during pregnancy influence in utero programming and postnatal susceptibility to atherogenesis. *FASEB J* **16**, 1348-1360 (2002).
74. Galton, D. J. & Ferns, G. A. A. in *Genetic factors in atherosclerosis: approaches and model systems* (ed. Sparkes, S.R.) 95-109 Karger & Basel, 1989).
75. Stein, O., Thiery, J. & Stein, Y. Is there a genetic basis for resistance to atherosclerosis? *Atherosclerosis* **160**, 1-10 (2002).
76. Garcia, C. K., Wilund, K., Marcello Arca, M., Zuliani, G., Fellin, R. et al. Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science* **292**, 1394-1398 (2001).
77. Attie, A. D. Atherosclerosis modified. *Circ Res* **89**, 102-104 (2001).
78. Goldstein, J. L. & Brown, M. S. The cholesterol quartet. *Science* **292**, 1310-1312 (2001).

79. Brown, M., Kovanen, P. & Goldstein, J. Regulation of plasma cholesterol by lipoprotein receptors. *Science* **212**, 628-635 (1981).
80. Fielding, C. in *Biochemical and Physiological Aspects of Human Nutrition* (ed. Stipanuk, M.) WB Saunders Company, Ithaca, 2000.
81. Gurr, M. I. Dietary lipids and coronary heart disease: old evidence, new perspective. *Prog Lipid Res* **31**, 195-243 (1992).
82. Wight, T. N. & Merrilees, M. J. Proteoglycans in atherosclerosis and restenosis: key roles for versican. *Circ. Res.* **94**, 1158-1167 (2004).
83. Lohse, P., Maas, S., Lohse, P., Sewell, A.C., van Diggelen, O.P. et al. Molecular defects underlying Wolman disease appear to be more heterogeneous than those resulting in cholesteryl ester storage disease. *J Lipid Res* **40**, 221-228 (1999).
84. Kuriyama, M., Yoshida, H., Suzuki, M., Fujiyama, J. & Igata, A. Lysosomal acid lipase deficiency in rats: lipid analyses and lipase activities in liver and spleen. *J Lipid Res* **31**, 1605-1642 (1990).
85. Pàgani, F., Garcia, R., Pariyarath, R., Stuani, C., Gridelli, B. et al. Expression of lysosomal acid lipase mutants detected in three patients with cholesteryl ester storage disease. *Hum Mol Gen* **5**, 1611-1617 (1996).
86. Blanchette-Mackie, E.J., Dwyer, N.K., Amende, L.M., Kruth, H.S., Butler, J.D. et al. Type-C Niemann-Pick disease: low density lipoprotein uptake is associated with premature cholesterol accumulation in the golgi complex and excessive cholesterol storage in lysosomes. *PNAS, USA* **85**, 8022-8026 (1988).
87. Seals, D. F. & Courtneidge, S. A. The ADAMs family of metalloproteases: multidomain proteins with multiple functions. *Genes & Amp Dev.* **17**, 7-30 (2003).
88. Arndt, M., Lendeckel, U., Röcken, C., Nepple, K., Wolke, C. et al. Altered expression of ADAMs (a disintegrin and metalloproteinase) in fibrillating human atria. *Circulation* **105**, 720-725 (2002).
89. Santamarina-Fojo, S., Remaley, A. T., Neufeld, E. B. & Brewer, H. B., Jr. Regulation and intracellular trafficking of the ABCA1 transporter. *J. Lipid Res.* **42**, 1339-1345 (2001).
90. Faber, B., Heeneman, S., Daemen, M. & Cleutjens, K. Genes potentially involved in plaque rupture. *Curr Opin Lipidol* **13**, 545-552 (2002).
91. Tall, A., Breslow, J. L. & Rubin, E. M. in *The metabolic and molecular bases of inherited disease* (eds. Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) McGraw-Hill, New York, 2001.
92. Moghadasian, M. H., Frohlich, J. J. & McManus, B. M. Advances in Experimental Dyslipidemia and Atherosclerosis. *Lab. Invest.* **81**, 1173-1183 (2001).
93. Online Mendelian Inheritance in Man  
<http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM>
94. Smith, J.D., Bhasin, J.M., Baglione, J., Settle, M., Xu, Y. et al. Atherosclerosis susceptibility loci identified from a strain intercross of apolipoprotein E-deficient mice via a high-density genome scan. *Arterioscler. Thromb. Vasc. Biol.* **26**, 597-603 (2006).

95. Cohen, J. C. & Zannis, V. I. Genes affecting atherosclerosis. *Curr Opin Lipidol* **12**, 93-95 (2001).
96. Desvergne, B., Michalik, L. & Wahli, W. Transcriptional regulation of metabolism. *Physiol Rev* **86**, 465-514 (2006).
97. Hartman, J. L., IV, Garvik, B. & Hartwell, L. Principles for the buffering of genetic variation. *Science* **291**, 1001-1004 (2001).
98. Peltonen, L. & McKusick, V. A. Genomics and medicine: dissecting human disease in the postgenomic era. *Science* **291**, 1224-1229 (2001).
99. Warden, C. H. & Fidler, J. S. Integrated methods to solve the biological basis of common disease. *Meth Enzymol* **13**, 347-357 (1997).
100. Allayee, H., Ghazalpour, A. & Lusis, A. J. Using mice to dissect genetic factors in atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **23**, 1501-1509 (2003).
101. Forcheron, F., Legedz, L., Chinetti, G., Feugier, P., Letexier, D. et al. Genes of cholesterol metabolism in human atheroma: overexpression of perilipin and genes promoting cholesterol storage and repression of ABCA1 expression. *Arterioscler. Thromb. Vasc. Biol.* **25**, 1711-1717 (2005).
102. Hiltunen, M., Tuomisto, T.T., Niemi, M., Brasen, J.H., Rissanen, T.T. et al. Changes in gene expression in atherosclerotic plaques analyzed using DNA array. *Atherosclerosis* **165**, 23-32 (2002).
103. Shanahan, C. M., Carey, N. R., Osbourn, J. K. & Weissberg, P. L. Identification of osteoglycan as a component of the vascular matrix. *Arterioscler, Thromb, Vasc Biol* **17**, 2437-2447 (1997).
104. Archacki, S.R., Angheloiu, G., Tian, X.L., Tan, F.L., DiPaola, N. et al. Identification of new genes differentially expressed in coronary artery disease by expression profiling. *Physiol Genomics* **15**, 65-74 (2003).
105. Faber, B., Cleutjens, K., Niessen, R., Aarts, P., Boon, W. et al. Identification of genes potentially involved in rupture of human atherosclerotic plaques. *Circ Res* **89**, 547-554 (2001).
106. Papaspyridonos, M., Smith, A., Burnand, K.G., Taylor, P., Padayachee, S. et al. Novel candidate genes in unstable areas of human atherosclerotic plaques. *Arterioscler. Thromb. Vasc. Biol.* **26**, 1837-1844 (2006).
107. Adams, L. D., Geary, R. L., Li, J., Rossini, A. & Schwartz, S. M. Expression profiling identifies smooth muscle cell diversity within human intima and plaque fibrous cap: loss of RGS5 distinguishes the cap. *Arterioscler. Thromb. Vasc. Biol.* **26**, 319-325 (2006).
108. Suckling, K.E. & Jackson, B. Animal models of human lipid metabolism. *Prog Lipid Res* **32**, 1-24 (1993).
109. Breslow, J. L. Genetic differences in endothelial cells may determine atherosclerosis susceptibility. *Circulation* **102**, 5-6 (2000).
110. Overturf, M., Smith, S.A., Gotto, A.M. Jr., Morrisett, J.D., Tewson, T. et al. Dietary cholesterol absorption, and sterol and bile acid excretion in hypercholesterolemia-resistant white rabbits. *J. Lipid Res.* **31**, 2019-2027 (1990).

111. Khosla, P. & Sundram, K. Effects of dietary fatty acid composition on plasma cholesterol. *Prog Lipid Res* **35**, 93-132 (1996).
112. Pitman, W. A., Hunt, M. H., McFarland, C. & Paigen, B. Genetic analysis of the difference in diet-induced atherosclerosis between the inbred mouse strains SM/J and NZB/BINJ. *Arterioscler. Thromb. Vasc. Biol.* **18**, 615-620 (1998).
113. Cullen, P., Baetta, R., Bellosta, S., Bernini, F., Chinetti, G. et al. Rupture of the atherosclerotic plaque: does a good animal model exist? *Arterioscler. Thromb. Vasc. Biol.* **23**, 535-542 (2003).
114. Shi, W., Brown, M.D., Wang, X., Wong, J., Kallmes, D.F. et al. Genetic backgrounds but not sizes of atherosclerotic lesions determine medial destruction in the aortic root of apolipoprotein E-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **23**, 1901-1906 (2003).
115. Rader, D. J. & Pure, E. Genetic susceptibility to atherosclerosis: insights from mice. *Circ Res* **86**, 1013-1015 (2000).
116. Sigmund, C. D. Viewpoint: are studies in genetically altered mice out of control? *Arterioscler. Thromb. Vasc. Biol.* **20**, 1425-1429 (2000).
117. Colinayo, V., Qiao, J.H., Wang, X., Krass, K.L., Schadt, E. et al. Genetic loci for diet-induced atherosclerotic lesions and plasma lipids in mice. *Mamm Genome* **14**, 464-471 (2003).
118. Lusis, A. J., Fogelman, A. M. & Fonarow, G. C. Genetic basis of atherosclerosis: part I: new genes and pathways. *Circulation* **110**, 1868-1873 (2004).
119. Zhang, S., Reddick, R., Burkey, B. & Maeda, N. Diet-induced atherosclerosis in mice heterozygous and homozygous for apolipoprotein E gene disruption. *J. Clin. Invest.* **94**, 937-945 (1994).
120. Yutzey, K. E. & Robbins, J. Principles of genetic murine models for cardiac disease. *Circulation* **115**, 792-799 (2007).
121. Watanabe, Y., Ito, T. & Shiomi, M. The effect of selective breeding on the development of coronary atherosclerosis in WHHL rabbits. An animal model for familial hypercholesterolemia. *Atherosclerosis* **56**, 71-79 (1985).
122. Shiomi, M., Ito, T., Yamada, S., Kawashima, S. & Fan, J. Development of an animal model for spontaneous myocardial infarction (WHHLMI Rabbit). *Arterioscler. Thromb. Vasc. Biol.* **23**, 1239-1244 (2003).
123. Weigensberg, B., Lough, J. & More, R. Modification of two types of cholesterol atherosclerosis in rabbits by blocking lipoprotein lysine epsilon-amino groups. *Atherosclerosis* **57**, 87-98 (1985).
124. Lusis, A. Atherosclerosis. *Nature* **407**, 233-241 (2000).
125. Inaba, T., Yamada, N., Gotoda, T., Shimano, H., Shimada, M. et al. Expression of M-CSF receptor encoded by c-fms on smooth muscle cells derived from arteriosclerotic lesion. *J. Biol. Chem.* **267**, 5693-5699 (1992).
126. Mozes, G., Mohacsi, T., Gloviczki, P., Menawat, S., Kullo, I. et al. Adenovirus-mediated gene transfer of macrophage colony stimulating factor to the arterial wall In Vivo. *Arterioscler. Thromb. Vasc. Biol.* **18**, 1157-1163 (1998).

127. Julien, P., Downar, E., & Angel, A. Lipoprotein composition and transport in the pig and dog cardiac lymphatic system. *Circ. Res.* **49**, 248 - 254 (1981).
128. Scott, R., Kim, D. & Schmee, J. Endothelial and lesion cell growth patterns of early smooth-muscle cell atherosclerotic lesions in swine. *Arch Pathol Lab Med* **109**, 450-453 (1985).
129. Larcher, O. Racueil de med veter. *Der Tierarzt* **13** (1874), in Siller, W.G. Spontaneous Atherosclerosis in Fowl. *Comparative atherosclerosis: the morphology of spontaneous and induced atherosclerotic lesions and animals in relation to disease.* (eds. Roberts, J. & Strauss, R.) Harper & Row, New York, 1965.
130. Yamagiwa, K. & Adachi, O. Uber die Atherosklerose bei Huhren. *Verhandl D. Japan. Path* **4**, 55-60 (1914), in Siller, W.G. Spontaneous Atherosclerosis in Fowl. *Comparative atherosclerosis: the morphology of spontaneous and induced atherosclerotic lesions and animals in relation to disease.* (eds. Roberts, J. & Strauss, R.) Harper & Row, New York, 1965.
131. Clarkson, T. B., Prichard, R. W., Netsky, M. G. & Lofland, H. B. Atherosclerosis in pigeons: its spontaneous occurrence and resemblance to human atherosclerosis. *AMA Arch Path* **68**, 143-147 (1959).
132. Herndon, C. N., Goodman, H. O., Clarkson, T. B., Lofland, H. B. & Prichard, R. W. Atherosclerosis resistance and susceptibility in two breeds of pigeons. *Genetics* **47**, 958 (1962).
133. Cornhill, J., Levesque, M. & Nerem, R. Quantitative study of the localization of sudanophilic coeliac lesions in the White Carneau pigeon. *Atherosclerosis* **35**, 103-110 (1980).
134. Qin, Z. & Nishimura, H. Ca<sup>2+</sup> signaling in fowl aortic smooth muscle increases during maturation but is impaired in neointimal plaques. *J. Exp. Biol.* **201**, 1695-1705 (1998).
135. St Clair, R. Metabolic changes in the arterial wall associated with atherosclerosis in the pigeon. *Fed Proc* **42**, 2480-2485 (1983).
136. St Clair, R. The contribution of avian models to our understanding of atherosclerosis and their promise for the future. *Lab Animal Sci* **48**, 565-568 (1998).
137. Prichard, R., Clarkson, T., Goodman, H. & Lofland, H. Aortic atherosclerosis in pigeons and its complications. *Arch Pathol* **77**, 244-257 (1964).
138. Wagner, W. Risk factors in pigeons genetically selected for increased atherosclerosis susceptibility. *Atherosclerosis* **31**, 453-463 (1978).
139. Wagner, W. D., Conner, J. & Labutta, T. Blood pressure in atherosclerosis-susceptible and -resistant pigeons. *Proc. Soc. Exp Biol Med* **162**, 101-104 (1979).
140. Hadjiisky, P., Bourdillon, M. & Grosogoeat, Y. [Experimental models of atherosclerosis. Contribution, limits and trends]. *Arch Mal Coeur Vaiss* **84**, 1593-1603 (1991).
141. Cornhill, J., Akins, D., Hutson, M. & Chandler, A. Localization of atherosclerotic lesions in the human basilar artery. *Atherosclerosis* **35**, 77-86 (1980).

142. Wagner, W. D., Clarkson, T. B., Feldner, H. B., Lofland, H. B. & Prichard, R. W. The development of pigeon strains with selected atherosclerosis characteristics. *Exp Mol Path* **19**, 304-319 (1973).
143. St. Clair, R., Leight, M. & Barakat, H. Metabolism of low density lipoproteins by pigeon skin fibroblasts and aortic smooth muscle cells. comparison of cells from atherosclerosis- susceptible and atherosclerosis-resistant pigeons. *Arterioscler. Thromb. Vasc. Biol.* **6**, 170-177 (1986).
144. Santerre, R., Wight, T., Smith, S. & Brannigan, D. Spontaneous atherosclerosis in pigeons. A model system for studying metabolic parameters associated with atherogenesis. *Am. J. Pathol.* **67**, 1-22 (1972).
145. Cooke, P. & Smith, S. C. Smooth muscle cells: source of foam cells in atherosclerotic White Carneau pigeons. *Exp Mol Path* **8**, 171-189 (1968).
146. Barakat, H. & St. Clair, R. Characterization of plasma lipoproteins of grain- and cholesterol-fed White Carneau and Show Racer pigeons. *J. Lipid Res.* **26**, 1252-1268 (1985).
147. Fronek, K. & Alexander, N. Genetic difference in the sympathetic nervous activity and susceptibility to atherosclerosis in pigeon. *Atherosclerosis* **39**, 25-33 (1981).
148. Smith, S. & Smith, E. C. Unpublished Observations.
149. Randolph, R., Smith, B. & St. Clair, R. Cholesterol metabolism in pigeon aortic smooth muscle cells lacking a functional low density lipoprotein receptor pathway. *J. Lipid Res.* **25**, 903-912 (1984).
150. Randolph, R. & St. Clair, R. Pigeon aortic smooth muscle cells lack a functional low density lipoprotein receptor pathway. *J. Lipid Res.* **25**, 888-902 (1984).
151. Smith, S. C. & Smith, E. C. Unpublished Data.
152. Schulz, J., Bermudez, A., Tomlinson, J., Firman, J. & He, Z. Blood plasma chemistries from wild mourning doves held in captivity. *J. Wild. Dis.* **36**, 541-545 (2000).
153. Siekert, R., Dicke, B., Subbiah, M. & Kottke, B. Cholesterol balance in atherosclerosis-susceptible and atherosclerosis-resistant pigeons. *Res Commun Chem Pathol Pharmacol* **10**, 181-184 (1975).
154. Subbiah, M. & Connelly, P. Effect of dietary restriction on plasma cholesterol and cholesterol excretion in the White Carneau pigeon. *Atherosclerosis* **24**, 509-513 (1976).
155. Hulcher, F. & Margolis, R. Rate-limiting, diurnal activity of hepatic microsomal cholesterol-7 alpha-hydroxylase in pigeons with high serum cholesterol. *Biochim Biophys Acta* **712**, 242-249 (1982).
156. Nicolosi, R. J., Santerre, R. F. & Smith, S. C. Lipid accumulation in muscular foci in White Carneau and Show Racer pigeon aortas. *Exp Mol Path* **17**, 29-37 (1972).
157. Jerome, W. & Lewis, J. Early atherogenesis in White Carneau pigeons. II. ultrastructural and cytochemical observations. *Am. J. Pathol.* **119**, 210-222 (1985).

158. Gosselin. A morphological and ultrastructural study of spontaneous and cholesterol aggravated atherosclerosis in susceptible and resistance pigeons. M.S. Thesis in Animal Science. University of New Hampshire, Durham, 1979.
159. Denholm, E. & Lewis, J. Monocyte chemoattractants in pigeon aortic atherosclerosis. *Am. J. Pathol.* **126**, 464-475 (1987).
160. Jerome, W. & Lewis, J. Early atherogenesis in White Carneau pigeons. I. leukocyte margination and endothelial alterations at the celiac bifurcation. *Am. J. Pathol.* **116**, 56-68 (1984).
161. Langelier, M., Connelly, P. & Subbiah, M. Plasma lipoprotein profile and composition in White Carneau and Show Racer breeds of pigeons. *Can J Biochem* **54**, 27-31 (1976).
162. Jones, N., Jerome, W. & Lewis, J. Pigeon monocyte/macrophage lysosomes during beta VLDL uptake. induction of acid phosphatase activity. a model for complex arterial lysosomes. *Am. J. Pathol.* **139**, 383-392 (1991).
163. in. Irradiation and atherosclerosis in pigeons. *Nutr Rev* **24**, 178-179 (1966).
164. Kauniz, H. Cholesterol and repair processes in arteriosclerosis. *Lipids* **13**, 373-374 (1977).
165. Goodman, H. O. & Herndon, C. N. Genetic aspects of atherosclerosis in pigeons. *Fed Proc* **22**, Abstract # 1336 (1963).
166. Curwen, K. & Smith, S. Aortic glycosaminoglycans in atherosclerosis-susceptible and -resistant pigeons. *Exp Mol Path* **27**, 121-133 (1977).
167. Hajjar, D., Wight, T. & Smith, S. Lipid accumulation and ultrastructural change within the aortic wall during early spontaneous atherogenesis. *Am J Pathol* **100**, 683-706 (1980).
168. Hajjar, D. & Smith, S. Focal differences in bioenergetic metabolism of atherosclerosis-susceptible and -resistant pigeon aortas. *Atherosclerosis* **36**, 209-222 (1980).
169. Santerre, R., Nicolosi, R. & Smith, S. Respiratory control in preatherosclerotic susceptible and resistant pigeon aortas. *Exp Mol Path* **20**, 397-406 (1974).
170. Hajjar, D., Farber, I. & Smith, S. Oxygen tension within the arterial wall: relationship to altered bioenergetic metabolism and lipid accumulation. *Arch Biochem Biophys* **262**, 375-380 (1988).
171. Sweetland, R. Lysosomal acid lipase activity in atherosclerosis susceptible and resistant pigeon aortas. M.S. Thesis in Animal Science. University of New Hampshire, Durham, 1999.
172. Fastnacht, C., Smith, S. C. & Smith, E. C. in *26th Lofland Conference on Arterial Wall Metabolism* (Winston-Salem, North Carolina, 1993).
173. Zempenyi, T. & Rosenstein, A. Arterial enzymes and their relation to atherosclerosis in pigeons. *Exp Mol Path* **22**, 225-241 (1975).
174. Subbiah, M., Schweiger, E., Deitmeyer, D., Gallon, L. & Sinzinger, H. Prostaglandin synthesis in aorta of atherosclerosis susceptible and atherosclerosis resistant pigeons. *Artery* **8**, 50-55 (1980).

175. Wight, T. Proteoglycans in pathological conditions: atherosclerosis. *Fed Proc* **44**, 381-385 (1985).
176. Edwards, I. J., Xu, H., Obunike, J. C., Goldberg, I. J. & Wagner, W. D. Differentiated macrophages synthesize a heparan sulfate proteoglycan and an oversulfated chondroitin sulfate proteoglycan that bind lipoprotein lipase. *Arterioscler. Thromb. Vasc. Biol.* **15**, 400-409 (1995).
177. Wight, T. Differences in the synthesis and secretion of sulfated glycosaminoglycans by aorta explant monolayers cultured from atherosclerosis-susceptible and -resistant pigeons. *Am. J. Pathol.* **101**, 127-142 (1980).
178. Wagner, W., Edwards, I., St Clair, R. & Barakat, H. Low density lipoprotein interaction with artery derived proteoglycan: the influence of LDL particle size and the relationship to atherosclerosis susceptibility. *Atherosclerosis* **75**, 49-59 (1989).
179. Tovar, A. M. F., Cesar, D. C. F., Leta, G. C. & Mourao, P. A. S. Age-related changes in populations of aortic glycosaminoglycans : species with low affinity for plasma low-density lipoproteins, and not species with high affinity, are preferentially affected. *Arterioscler. Thromb. Vasc. Biol.* **18**, 604-614 (1998).
180. Brannigan, D. Reproductive behavior and squab development in atherosclerosis-susceptible White Carneau and atherosclerosis-resistant show racer pigeons. PhD. Dissertation in Zoology. University of New Hampshire, Durham, 1973.
181. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716 (2004).
182. Smith, S. C., Smith, E. C. & Taylor, R. L., Jr. Susceptibility to spontaneous atherosclerosis in pigeons: an autosomal recessive trait. *J. Hered.* **92**, 439-442 (2001).
183. Smith, S. C., Strout, R. G., Dunlop, W. R. & Smith, E. C. Fatty acid composition of cultured aortic smooth muscle cells from White Carneau and Show Racer pigeons. *J. Atheroscler Res* **5**, 379-387 (1965).
184. Wight, T., Cooke, P. & Smith, S. An electron microscopic study of pigeon aorta cell cultures. cytodifferentiation and intracellular lipid accumulation. *Exp Mol Path* **27**, 1-18 (1977).
185. Hungerford, J., Owens, G., Argraves, W. & Little, C. Development of the aortic vessel wall as defined by vascular smooth muscle and extracellular matrix markers. *Dev Biol* **178**, 375-392 (1996).
186. Holifield, B., Helgason, T., Jemelka, S., Taylor, A., Navran, S. et al. Differentiated vascular myocytes: are they involved in neointimal formation? *J. Clin. Invest.* **97**, 814-825 (1996).
187. Gown, A., Vogel, A., Gordon, D. & Lu, P. A smooth muscle-specific monoclonal antibody recognizes smooth muscle actin isozymes. *J. Cell Biol.* **100**, 807-813 (1985).
188. McDonald, O. G., Wamhoff, B. R., Hoofnagle, M. H. & Owens, G. K. Control of SRF binding to CArG box chromatin regulates smooth muscle gene expression in vivo. *J. Clin. Invest.* **116**, 36-48 (2006).



189. Campbell, G. & Campbell, J. Smooth muscle phenotypic changes in arterial wall homeostasis: implications for the pathogenesis of atherosclerosis. *Exp Mol Path* **42**, 139-162 (1985).
190. Meir, K. S. & Leitersdorf, E. Atherosclerosis in the apolipoprotein E-deficient mouse: a decade of progress. *Arterioscler. Thromb. Vasc. Biol.* **24**, 1006-1014 (2004).
191. Cullen, P., Lorkowski, S. (eds). *Analysing Gene Expression*. Wiley-VCH, Weinheim, Germany, 2003.
192. Frossard, P. M. & Vinogradov, S. in *Genetic Factors in Atherosclerosis: Approaches and Model Systems* (eds. Lusis, A. J. & Sparkes, S. R.) Karger & Basel, 1989).
193. Bijmens, A., Lutgens, E., Ayoubi, T., Kuiper, J., Horrevoets, A.J. et al. Genome-wide expression studies of atherosclerosis. critical issues in methodology, analysis, and interpretation of transcriptomics data. *Arterioscler. Thromb. Vasc. Biol.* **26**, 1226-1235 (2006).
194. Chen, J., Sun, M., Lee, S., Zhou, G., Rowley J.D. et al. Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *PNAS* **99**, 12,257-12,262 (2002).
195. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
196. Venter, J. C., Adams, M. D., Myers, E. W. & Li, P. W. The sequence of the human genome. *Science* **291**, 1304-1351 (2001).
197. Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493-521 (2004).
198. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562 (2002).
199. Schieffer, B. & Drexler, H. ACE gene polymorphism and coronary artery disease. *Arterioscler Thromb Vasc Biol* **20**, 281-282 (2000).
200. Sing, C. F., Stengard, J. H. & Kardia, S. L. R. Genes, environment, and cardiovascular Disease. *Arterioscler. Thromb. Vasc. Biol.* **23**, 1190-1196 (2003).
201. Gura, T. Can SNPs deliver susceptibility genes? *Science* **293**, 593-595 (2001).
202. Arnett, D. K. for the Writing Group. Summary of the American Heart Association's scientific statement on the relevance of genetics and genomics for prevention and treatment of cardiovascular disease. *Arterioscler. Thromb. Vasc. Biol.* **27**, 1682-1686 (2007).
203. Carden, L. Delivering new disease genes. *Science* **314**, 1403-1405 (2006).
204. Pollex, R. L. & Hegele, R. A. Complex trait locus linkage mapping in atherosclerosis: time to take a step back before moving forward? *Arterioscler. Thromb. Vasc. Biol.* **25**, 1541-1544 (2005).
205. Fields, S. Proteomics in genomeland. *Science* **291**, 1221-1224 (2001).

206. Shin, C. & Manley, J. L. Cell signaling and the control of pre-mRNA splicing. *Nat Rev Mol Cell Biol* **5**, 727-738 (2004).
207. Varga, R., Eriksson, M., Erdos, M.R., Olive, M., Harten, I. et al. Progressive vascular smooth muscle cell defects in a mouse model of Hutchinson-Gilford progeria syndrome. *PNAS* **103**, 3250-3255 (2006).
208. Davies, K. Decoding the genetics of common disease. *Bio IT World*, 24 (2006).
209. O'Malley, B. W. Molecular biology: little molecules with big goals. *Science* **313**, 1749-1750 (2006).
210. Strohman, R. Maneuvering in the complex path from genotype to phenotype. *Science* **296**, 701-703 (2002).
211. Pennisi, E. Genomics comes of age. *Science* **290**, 2220-2221 (2000).
212. Hardison, R. C. Primer: Comparative Genomics. *PLoS Biol* **1**, E58 (2003).
213. Farrell, R. E. *RNA Methodologies: A Laboratory Guide for Isolation and Characterization* Academic Press, New York, 1999.
214. Arnett, D.K., Baird, A.E., Barkley, R.A., Basson, C.T., Boerwinkle, E. et al. Relevance of genetics and genomics for prevention and treatment of cardiovascular disease: a scientific statement from the American Heart Association council on epidemiology and prevention, the stroke council, and the functional genomics and translational biology interdisciplinary working group. *Circulation* **115**, 2878-2901 (2007).
215. Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720-1730 (1999).
216. Pennisi, E. Genomics: DNA study forces rethink of what it means to be a gene. *Science* **316**, 1556-1557 (2007).
217. Cunningham, B. A. Assessing differential gene expression. *The Scientist* **15**, 27 (2001).
218. Schadt, E. E. & Lum, P. Y. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. reverse engineering gene networks to identify key drivers of complex disease phenotypes. *J. Lipid Res.* **47**, 2601-2613 (2006).
219. Tegner, J., Skogsberg, J. & Bjorkegren, J. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. multi-organ whole-genome measurements and reverse engineering to uncover gene networks underlying complex traits. *J. Lipid Res.* **48**, 267-277 (2007).
220. Adam, P. J. in *Vascular Disease: Molecular Biology & Gene Therapy Protocols* 99-109 Humana Press, Totowa, NJ, 2000.
221. Zvara, A., Hackler, L., Nagy, Z., Micsik, T. & Puskas, L. New molecular methods for classification, diagnosis and therapy prediction of hematological malignancies. *Pathol Oncol Res* **8**, 231-240 (2002).
222. Service, R.F. The Race for the \$1000 genome. *Science* **311**, 1544-1546 (2006).
223. Ding, C. & Cantor, C. Quantitative analysis of nucleic acids--the last few years of progress. *J Biochem Mol Biol* **37**, 1-10 (2004).

224. Burge, C. Chipping away the transcriptome. *Nat Gen* **27**, 232-234 (2001).
225. Lisitsyn, N., Lisitsyn, N. & Wigler, M. Cloning the differences between two complex genomes". *Science* **259**, 946-951 (1993).
226. Everts, R., Versteeg, S.A., Renier, C., Vignaux, F., Groot, P.C. et al. Isolation of DNA markers informative in purebred dog families by genomic representational difference analysis (gRDA). *Mamm Genome* **11**, 741-747 (2000).
227. Vorster, B., Kunert, K. & Cullis, C. Use of representational difference analysis for the characterization of sequence differences between date palm varieties. *Plant Cell Rep* **21**, 271-275 (2002).
228. Hubank, M. & Schatz, D. G. Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucl Acids Res.* **22**, 5640-5648 (1994).
229. O'Neill, M. & Sinclair, A. Isolation of rare transcripts by representational difference analysis. *Nucl Acids Res.* **25**, 2681-2682 (1997).
230. O'Hara, E., Williams, M.B., Rott, L., Abola, P., Hansen, N. et al. Modified representational difference analysis: isolation of differentially expressed mRNAs from rare cell populations. *Anal Biochem* **336**, 221-230 (2005).
231. Sung, B., Jung, K.J., Song, H.S., Son, M.J., Yu, B.P. et al. cDNA representational difference analysis used in the identification of genes related to the aging process in rat kidney. *Mech Ageing Dev* **126**, 882-891 (2005).
232. Dinel, S., Bolduc, C., Belleau, P., Boivin, A., Yoshioka, M. et al. Reproducibility, bioinformatic analysis and power of the SAGE method to evaluate changes in transcriptome. *Nucl. Acids Res.* **33**, e26 (2005).
233. Remaley, A., Schumacher, U., Amouzadeh, H., Brewer, H., Jr & Hoeg, J. Identification of novel differentially expressed hepatic genes in cholesterol-fed rabbits by a non-targeted gene approach. *J. Lipid Res.* **36**, 308-314 (1995).
234. Pastorian, K., Hawel, L. & Byus, C. Optimization of cDNA representational difference analysis for the identification of differentially expressed mRNAs. *Anal Biochem* **283**, 89-98 (2000).
235. Tyson, O. & Shanahan, C. M. Use of cDNA representational difference analysis to identify disease specific genes in human atherosclerosis. in *Methods in Molecular Medicine*. Humana Press, Totawa, NJ, 2000.
236. Oba-Shinjo, S., Bengtson, M.H., Winnischofer, S.M., Colin, C., Vedoy, C.G. et al. Identification of novel differentially expressed genes in human astrocytomas by cDNA representational difference analysis. *Brain Res Mol Brain Res* **140**, 25-33 (2005).
237. Andersson, T., Borang, S., Larsson, M., Wirta, V., Wennborg, A. et al. Novel candidate genes for atherosclerosis are identified by representational difference analysis-based transcript profiling of cholesterol-loaded macrophages. *Pathobiology* **69**, 304-314 (2001).
238. Ishikawa, S., Egami, H., Kurizaki, T., Akagi, J., Tamori, Y. et al. Identification of genes related to invasion and metastasis in pancreatic cancer by cDNA representational difference analysis. *J Exp Clin Cancer Res* **22**, 299-306 (2003).

239. Bowler, L. Representational difference analysis of cDNA. *Methods Mol Med* **94**, 49-66 (2004).
240. Kim, S., Zeller, K., Dang, C., Sandgren, E. & Lee, L. A strategy to identify differentially expressed genes using representational difference analysis and cDNA arrays. *Anal Biochem* **288**, 141-148 (2001).
241. Linder, K., Arner, P., Flores-Morales, A., Tollet-Egnell, P. & Norstedt, G. Differentially expressed genes in visceral or subcutaneous adipose tissue of obese men and women. *J. Lipid Res.* **45**, 148-154 (2004).
242. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**, 156-159 (1987).
243. Gruffat, D., Piot, C., Durand, D. & Bauchart, D. Comparison of four methods for isolating large mRNA: apolipoprotein B mRNA in bovine and rat livers. *Anal Biochem* **242**, 77-83 (1996).
244. Wilfinger, W. W. Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *Biotechniques* **22**, 474-481 (1997).
245. Manchester, K. L. Use of UV methods for measurement of protein and nucleic acid concentrations. *Biotechniques* **20**, 968-970 (1996).
246. Rapley, R. & Heptinstall, J. in *RNA Isolation and Characterization Protocols* (eds. Rapley, R. & Manning, D.) Humana Press, Totowa, NJ, 1998.
247. Anderson, J. L. From RNA to DNA, the construction of DNA libraries from pigeon aortic tissue: an evaluation of methods. M.S. Thesis in Animal and Nutritional Sciences, University of New Hampshire, Durham, 2003.
248. Carninci, P. & Hayashizaki, Y. High-efficiency full-length cDNA cloning. *Meth Enzymol* **303**, 19-44 (1999).
249. *Molecular Cloning: A Laboratory Manual* (ed. Sambrook) Cold Harbor Press, Cold Harbor, 2001.
250. Seth, D. Gorrell, M.D., McGuinness, P.H., Leo, M.A., Lieber, C.S. et al. SMART amplification maintains representation of relative gene expression: quantitative validation by real time PCR and application to studies of alcoholic liver disease in primates. *J Biochem Biophys Methods* **55**, 53-66 (2003).
251. Efimov, V. A., Buryakova, A. A. & Chakmakhcheva, O. G. Synthesis of polyacrylamides *N*-substituted with PNA-like oligonucleotides mimics for molecular diagnostic applications. *Nucl Acids Res* **27**, 4416-4426 (1999).
252. Gwynne, P. & Heebner, G. PCR and cloning: a technology for the 21st Century. *Science*, 1073-1083 (2000).
253. Murphy, N. & Hellwig, R. Improved nucleic acid organic extraction through use of a unique gel barrier material. *Biotechniques* **21**, 934-939 (1996).
254. Monk, D., Smith, R., Arnaud, P., Preece, M.A., Stanier, P. et al. Imprinted methylation profiles for proximal mouse chromosomes 11 and 7 as revealed by methylation-sensitive representational difference analysis. *Mamm Genome* **14**, 805-816 (2003).

255. Reagin, M. J., Giesler, T.L., Merla, A.L., Resetar-Gerke, J.M., Kapolka, K.M. et al. TempliPhi: A sequencing template preparation procedure that eliminates overnight cultures and DNA purification. *J Biomol Tech* **14**, 143-148 (2003).
256. Demidov, V. Rolling-circle amplification in DNA diagnostics: the power of simplicity. *Expert Rev Mol Diagn* **2**, 542-548 (2002).
257. Altschul, S., Gish, W., Miller, W., Myers, M. & Lipman, D. Basic local alignment search tool. *J Mol Biol* **215**, 403 - 410 (1990).
258. Altschul, S., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* **25**, 3389 - 3402 (1997).
259. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucl Acids Res.* **31**, 3784-3788 (2003).
260. Gene Indices. Dana Farber Cancer Institute, Boston MA, 2007.
261. Lee, Y., Tsai, J., Sunkara, S., Karamycheva, S., Pertea, G. et al. The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucl. Acids Res.* **33**, D71-D74 (2005).
262. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh M. et al. From genomics to chemical genomics: new developments in KEGG. *Nucl Acids Res.* **34**, D354-D357 (2006).
263. Kim, J. & Dang, C. Multifaceted roles of glycolytic enzymes. *Trends Biochem Sci* **30**, 142-150 (2005).
264. Nikitin, A., Egorov, S., Daraselina, N. & Mazo, I. Pathway studio--the analysis and navigation of molecular networks. *Bioinformatics* **19**, 2155-2157 (2003).
265. Gregory, T. Animal Genome Size Database. <http://www.cbs.dtu.dk/databases/DOGS/index.php>, (2005).
266. Derjusheva, S., Kurganova, A., Habermann, F. & Gaginskaya, E. High chromosome conservation detected by comparative chromosome painting in chicken, pigeon and passerine birds. *Chromosome Res* **12**, 715-723 (2004).
267. Burt, D. W., Bumstead, N., Bitgood, J. J., Ponce de Leon, F. A. & Crittenden, L. B. Chicken genome mapping: a new era in avian genetics. *Trends in Genetics* **11**, 190-194 (1995).
268. Burt, D., Bruley, C., Dunn, I.C., Jones, C.T., Ramage, A. et al. The dynamics of chromosome evolution in birds and mammals. *Nature* **402**, 411-413 (1999).
269. Kayang, B., Fillon, V., Inoue-Murayama, M., Miwa, M., Leroux, S. et al. Integrated maps in quail (*Coturnix japonica*) confirm the high degree of synteny conservation with chicken (*Gallus gallus*) despite 35 million years of divergence. *BMC Genomics* **7**, 101 (2006).
270. Hubbard, T.J.P., Aken, B.L., Beal, K., Ballester, B., Caccamo, M. et al. Ensembl 2007. *Nucl. Acids Res.* **35**, D610-D617 (2007).

271. Man, M. Z., Wang, X. & Wang, Y. POWER\_SAGE: comparing statistical tests for SAGE experiments. *Bioinformatics* **16**, 953-959 (2000).
272. Romualdi, C., Bortoluzzi, S. & Danieli, G. A. Detecting differentially expressed genes in multiple tag sampling experiments: comparative evaluation of statistical tests. *Hum. Mol. Gen.* **10**, 2133-2141 (2001).
273. Itoh, Y., Suzuki, M., Ogawa, A., Munechika, I., Murata, K. et al. Identification of the sex of a wide range of carinatae birds using PCR primer sets selected from chicken EEO.6 and its related sequences. *J. Hered* **92**, 315-321 (2001).
274. Weiss, J. N., Yang, L. & Qu, Z. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. network perspectives of cardiovascular metabolism. *J. Lipid Res.* **47**, 2355-2366 (2006).
275. Hansen, J. Spontaneous atherosclerosis: an ultrastructural study in the White Carneau pigeon. *Virchows Arch A Pathol Anat Histol* **375**, 147-157 (1977).
276. Owens, G., Vernon, S. & Madsen, C. Molecular regulation of smooth muscle cell differentiation. *J Hypertens Suppl* **14**, S55-S64 (1996).
277. Berne, R.M., Levy, M.N., Koepfen, B.M. (eds) *Physiology 5<sup>th</sup> Edition*. Mosbey, St. Louis, Missouri, 2004.
278. Wu, Y., Erdodi, F., Muranyi, A., Nullmeyer, K.D., Lynch, R.M. et al. Myosin phosphatase and myosin phosphorylation in differentiating C2C12 cells. *J Muscle Res Cell Motil* **24**, 499-511 (2003).
279. Kuykindoll, R., Nishimura, H., Thomason, D. & Nishimoto, S. Osteopontin expression in spontaneously developed neointima in fowl (*Gallus gallus*). *J. Exp. Biol.* **203**, 273-282 (2000).
280. Smith, S. C., Smith, E. C. & Anderson, J. L. RDA of WC and SR pigeon aortas from 1-3 day-old squabs. (Manuscript in Preparation).
281. Boyd, C. D., Song, J.Y., Kniep, A.C., Park, H.S., Fastnacht, C. et al. A restriction length polymorphism in the pigeon pro alpha 2(1) collagen gene: lack of allelic association with an atherogenic phenotype in pigeons genetically susceptible to the development of spontaneous atherosclerosis. *Conn Tiss Res* **26**, 187-197 (1991).
282. Zhao, B. Activation of the unfolded response occurs at all stages of atherosclerotic lesion development in apolipoprotein E null deficient mice. *Circulation* **111**, 1814-1821 (2005).
283. Onda, M., Ishiwata, T., Kawahara, K., Wang, R., Naito, Z. et al. Expression of lumican in thickened intima and smooth muscle cells in human coronary atherosclerosis. *Exp Mol Path* **72**, 142-149 (2002).
284. Feng, Y., Yang, J.H., Huang, H., Kennedy, S.P., Turi, T.G. et al. Transcriptional profile of mechanically induced genes in human vascular smooth muscle cells. *Circ. Res.* **85**, 1118-1123 (1999).
285. Troup, S. Njue, N., Kliewer, E.V., Parisien, M., Roskelley, C. et al. Reduced Expression of the Small Leucine-rich Proteoglycans, Lumican, and Decorin Is Associated with Poor Outcome in Node-negative Invasive Breast Cancer. *Clin. Cancer Res.* **9**, 207-214 (2003).

286. OMIN Database, John Hopkins University, (1966-2005).
287. Monfort, J., Tardif, G., Reboul, P., Mineau, F., Roughley, P. et al. Degradation of small leucine-rich repeat proteoglycans by matrix metalloprotease-13: identification of a new biglycan cleavage site. *Arthritis Res Ther* **8**, R26 (2006).
288. Spencer, J. A., Hacker, S.L., Davis, E.C., Mecham, R.P., Knutsen, R.H. et al. Altered vascular remodeling in fibulin-5-deficient mice reveals a role of fibulin-5 in smooth muscle cell proliferation and migration. *PNAS* **102**, 2946-2951 (2005).
289. Schiemann, W. P., Blobbe, G. C., Kalume, D. E., Pandey, A. & Lodish, H. F. Context-specific Effects of Fibulin-5 (DANCE/EVEC) on Cell Proliferation, Motility, and Invasion. Fibulin-5 is induced by transforming growth factor-beta and affects protein kinase cascades. *J. Biol. Chem.* **277**, 27,367-27,377 (2002).
290. Argraves, W., Greene, L., Cooley, M. & Gallagher, W. Fibulins: physiological and disease perspectives. *EMBO Rep* **4**, 1127-1131 (2003).
291. Smith, S. C., Taylor, R. L. Jr., Shine, R., Anderson, J. & Ashwell, C. Microarray analysis of WC and SR aortas from 1-3 day-old squabs. (Manuscript in Preparation).
292. Wang, J., Adelson, D.L., Yilmaz, A., Sze, S.H., Jin, Y. et al. Genomic organization, annotation, and ligand-receptor inferences of chicken chemokines and chemokine receptor genes based on comparative genomics. *BMC Genomics* **6**, 45 (2005).
293. Smith, S. C., Smith, E. C., Gilman, M. L., Taylor, R. L. Jr. & Anderson, J. Differentially expressed soluble proteins in aortic cells from atherosclerosis-susceptible & resistant pigeons. (Manuscript in Preparation).
294. Guo, F., Zarella, C. & Wagner, W. STAT4 and the proliferation of artery smooth muscle cells in atherosclerosis. *Exp Mol Path* **81**, 15-22 (2006).
295. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).
296. Wagner, W., Guo, F. & Jokinen, M. Artery regional properties and atherosclerosis susceptibility. *Life Sci* **80**, 299-306 (2007).
297. Singh, N. & Ramji, D. The role of transforming growth factor-beta in atherosclerosis. *Cytokine Growth Factor Rev* **17**, 487-499 (2006).
298. Marchase, R., Koro, L., Kelly, C. & McClay, D. Retinal ligatin recognizes glycoproteins bearing oligosaccharides terminating in phosphodiester-linked glucose. *Cell* **28**, 813-820 (1982).
299. Marchase, R., Koro, L., Kelly, C. & McClay, D. A possible role for ligatin and the phosphoglycoproteins it binds in calcium-dependent retinal cell adhesion. *J Cell Biochem* **18**, 461-468 (1982).
300. Smith, E. C., Smith, S. C. & Anderson, J. Review of spontaneous atherosclerosis in the susceptible pigeon. *Avian & Poultry Biol Rev* (Manuscript in Preparation).
301. Jakoi, E., Ross, P., Ping Ting-Beall, H., Kaufman, B. & Vanaman, T. Ligatin: a peripheral membrane protein with covalently bound palmitic acid. *J. Biol. Chem.* **262**, 1300-1304 (1987).

302. Ushio-Fukai, M. Nuclear phospholipase D1 in vascular smooth muscle: specific activation by G protein-coupled receptors. *Circ. Res.* **99**, 116-118 (2006).
303. Low, P., Hastings, R.A., Dawson, S.P., Sass, M., Billett, M.A. et al. Localisation of 26S proteasomes with different subunit composition in insect muscles undergoing programmed cell death. *Cell Death Differ* **7**, 1210-1217 (2000).
304. Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16**, 47-62 (2006).
305. Ermonval, M., Kitzmuller, C., Mir, A. M., Cacan, R. & Ivessa, N. E. N-glycan structure of a short-lived variant of ribophorin I expressed in the MadIA214 glycosylation-defective cell line reveals the role of a mannosidase that is not ER mannosidase I in the process of glycoprotein degradation. *Glycobiology* **11**, 565-576 (2001).
306. Wilson, C. M., Kraft, K., Duggan, C., Ismail, N., Crawshaw, S.G. et al. Ribophorin I associates with a subset of membrane proteins after their integration at the Sec61 translocon. *J. Biol. Chem.* **280**, 4195-4206 (2005).
307. Bagshaw, R. D., Mahuran, D. J. & Callahan, J. W. A Proteomic analysis of lysosomal integral membrane proteins reveals the diverse composition of the organelle. *Mol. Cell. Proteomics* **4**, 133-143 (2005).
308. Edwards, I. & Wagner, W. Distinct synthetic and structural characteristics of proteoglycans produced by cultured artery smooth muscle cells of atherosclerosis-susceptible pigeons. *J. Biol. Chem.* **263**, 9612-9620 (1988).
309. Nishi, E., Prat, A., Hospital, V., Elenius, K. & Klagsbrun, M. N-arginine dibasic convertase is a specific receptor for heparin-binding EGF-like growth factor that mediates cell migration. *EMBO J* **20**, 3342-3350 (2001).
310. Tuinen, M. v., Sibley, C. G. & Hedges, S. B. The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. *Mol. Biol. Evol.* **17**, 451-457 (2000).
311. Pimentel-Smith, G., Shi, L., Drummond, P., Tu, Z. & Smith, E. Amplification of sequence tagged sites in five avian species using heterologous oligonucleotides. *Genetica* **110**, 219-226 (2000).
312. Unneberg, P., Wennborg, A. & Larsson, M. Transcript identification by analysis of short sequence tags—influence of tag length, restriction site and transcript database. *Nucl. Acids Res.* **31**, 2217-2226 (2003).
313. Hubank, M. & Schatz, D. cDNA representational difference analysis: a sensitive and flexible method for identification of differentially expressed genes. *Meth Enzymol* **303**, 325-349 (1999).
314. Churchill, G. Fundamentals of experimental design for cDNA microarrays. *Nat Genet* **32**, 490-495 (2002).



## APPENDICES

## **APPENDIX A: ANIMAL RESEARCH DOCUMENTATION**

**UNH Animal Care and Usage Committee Approval #050601**

**This study involved the use of pigeon aortic cells prepared by the Dissertation director. No individual approval was required because I did not participate in either the care of the pigeons or the acquisition of pigeon tissue used for cell culture (communication with Julie Simpson, Manager, Research Conduct & Compliance Services, Office of Sponsored Research).**

## APPENDIX B: SELECTED PROTOCOLS

**Table B.27 SLOWCOOL (Adapter Ligation Protocol)**

<b>Preheat Thermal Cycler to 55°C</b>		
<b>Step</b>	<b>Temp (°C)</b>	<b>Time</b>
1	55	2 min
2	53	1 min
3	51	1 min
4	49	1 min
5	47	1 min
6	45	1 min
7	43	1 min
8	42	1 min
9	40	1 min
10	38	1 min
11	37	1 min
12	36	1 min
13	35	1 min
14	34	1 min
15	33	1 min
16	32	1 min
17	31	1 min
18	30	1 min
19	29	1 min
20	28	1 min
21	27	1 min
22	26	1 min
23	25	1 min
24	24	1 min
25	23	1 min
26	22	1 min
27	21	1 min
28	20	1 min
29	19	1 min
30	18	1 min
31	17	1 min
32	16	1 min
33	15	1 min
34	14	1 min
35	13	1 min
36	12	1 min
37	11	1 min
38	10	1 min
39	9	1 min
40	8	HOLD

APPENDIX C: RAW DATA TABLES

Table C.28 Differential Gene Expression in Cultured Pigeon Aortic Cells (Combined Totals)

Upregulated Pigeon Ortholog	Copy #		Chicken Map			GenBank Accession Numbers (* DFGI, not NCBI)
	WC	SR	Gg	Locl	Gga	
16S rRNA (unresolved)	1	7	unk	na	na	AF173568/AY274070.1
ACAT1 (acetyl-coA acetyltransferase/thiolase)	1	0	1	418968	3173	DQ217370.1
ACSL1 (acyl-coA synthetase)	1	0	4	422547	18942	NM_001012578.1
ACTA2 (alpha 2 actin)	0	7	9	423787	4530	NM_001031229.1
ACTB (beta actin)	7	1	2	396526	8939	NM_205518.1/DQ207609.1/X00182.1
ACTR 10 (actin related protein)	0	1	5	423543	5464	NM_001006492.1
ACTR 3 (actin related protein)	0	1	7	374197	4076	NM_204307.1
ACVR1 (activin/TGFB Receptor 1)	0	1	7	395246	2875	NM_204560.1
ADAMTS15 (reprolysin type disintegrin & metalloprotease)*	2	0	24	419733	18773	XM_417874*
ALDH9A1 (aldehyde dehydrogenase E3)	5	1	8	424405	19278	XM_422248.1
AMD1 (adenosylmethionine decarboxylase)	0	1	3	421755	8864	NM_001012569.1
ANP32A (nuclear phosphoprotein 32)	1	1	10	415562	13468	XM_413932.1
ANXA2 (annexin A2)	7	0	10	396297	3641	NM_205351.1/NM_174716.1/X53334.1
APTX (aprataxin)	0	1	Z	395173	9488	XM_429199.2
ARRB (arrestin domain)	0	1	Z	427107	17050	XM_424699.1
ATP4B (H+/K+ ATPase Subunit B)	0	1	1	386581	884	NM_204418.1
ATPase 6/8 (ATPase subunits 8 & 6)	2	2	mt	807646	na	AY092583.1/AB238222.1
BAC CH261-10P22	0	1	2	na	na	AC145979.4
BAC CH261-124L19	14	0	Z	na	na	AC187592.3
BAC CH261-138K4	4	0	Z	na	na	AC189018
BAC CH261-15P9	0	1	3	na	na	AC146336.1
BAC CH261-16J12	0	1	5	na	na	AC145958.3
BAC CH261-20I24	0	3	5	na	na	AC172371.3
BAC CH261-187N23	0	1	W	na	na	AC186812.3

Table C.28 (continued)

Upregulated Pigeon Ortholog	Copy #		Chicken Map		GenBank Accession Numbers (* DFGI, not NCBI)
	WC	SR	Gg	Gga	
BAC CH261-46G16	0	2	unk	na	AC175318.4
BAC CH261-3201	0	4	2	na	AC145931.4
BAC CH261-71A16	0	4	1	na	AC163706
BAC CH261-93G5	0	1	Z	na	AC187352.1
C10orf58	1	0	6	na	DQ216459.1/Q9BRX8*
C20orf45	1	6	unk	419310	NM_00103866.1
C3AR1 (C3a anaphylatox)	0	1	1	418198	NM_001030769
CANX (calnexin)	0	1	13	416288	NM_001030620.1
CCND2 (cyclin D2)	1	0	1	374047	NM_204213.1
CCT5	1	1	2	420930	NM_001012563.1
CCT6A	1	0	19	417541	NM_001008216.1
CCT8	1	0	1	418486	AY393846.1
CENPC1 (centromere protein)	0	1	4	395922	AB042324.1
Clone 38f16	1	0	12	415891	XR_027022.1
chEST380b13	0	1	15	na	CR388671.1
Clone BU37787*	1	0	unk	na	TC_BU37787*
Clone BU447803*	1	0	unk	na	TC_BU447803*
Clone BX268708*	0	1	unk	na	TC_BX268708*
Clone Q553U2*	0	1	4	na	TC_Q553U2*
CNDP2 (dipeptidase M20 family)	1	0	2	42103	NM_001006385.1
CNRC01 (cadherin-related neuronal receptor c01)*	1	0	1	497233	CR386528.1/Q6R0H9*
CO I (cytochrome oxidase subunit a)	0	5	mt	807639	AY666494.1/AF279733.1
CO II (cytochrome oxidase subunit a3)	11	25	mt	807635	na
COPG (coatermer protein complex gamma)	1	0	12	416014	DQ215363.1
COTL1 (Coactosin)	0	5	11	768420	XM_001231327.1
COL5A2 (collagen, alpha-2 type V)	1	0	7	423986	XM_421846.1
COL1A2 (collagen, alpha-2 type I)	5	0	2	420564	XM_418665
CPSF2 (cleavage & polyadenylation specific factor 2)	16	0	5	423416	NM_001031208

Table C.28 (continued)

Upregulated Pigeon Ortholog	Copy #			Chicken Map			GenBank Accession Numbers (* DFGI, not NCBI)
	WC	SR	Gg	Loci	Gga		
CRSP2 (Sp1 TF activate cofactor)	0	1	1	418572	39400	XM_416776.1	
CRTAP (cartilage associated protein) aka "CASP"	1	0	2	395992	4211	X97607.1	
CXCL12 (stromal cell-derived factor 1)	13	1	6	395180	9513	CR353382.1	
CYTB (cytochrome b)	0	36	mt	807641	na	AY509674.1	
CYNF n-pac (cytokine-like nuclear factor n-pac)	1	0	14	426988	15811	AJ719657.1/NM_001006572	
DACH1 (dachshund homolog-1c)	9	0	1	373935	79	XM_001082371.1/XM_001082371/XM_417018.2	
DCN (decorin)	2	0	1	417892	1719	NM_001030747.1/P28675*	
DGAT2 (Diacylglycerol O-acetyltransferase)	10	0	3	421309	37973	XM_419374	
DNAJA2 (Hsp40)	0	1	11	415744	20231	NM_001005841.1	
EEF1A1 (translation elongation factor)	1	0	3	373963	34328	NM_204157.2	
EIF3S	1	1	2	420288	4721	NM_001030951.1	
EIF4A2	6	16	9	395232	4580	NM_204549.1/AJ720280.1	
EIF4E2	2	0	9	424941	1103	CR385832.1/XM_422748.2	
EIF4G2 (NAT1;DAP5)	2	1	5	395905	4138	AB096098.1	
ENO1 (enolase, alpha)	30	0	21	396017	1383	DQ216685.1/NM_205120.1/X14195.1	
Enzyme (pectinesterase/methylesterase) Inhibitor*	2	0	6	na	na	AAF34827.1*	
EXOC7(exocyst complex component 7)	1	0	18	417361	9898	NM_001012802.1	
Ezrin/VIL2	1	0	3	395701	4017	NM_204885.1	
FABP4	2	2	2	374165	4939	NM_204290.1	
FBLN5 (fibulin-5 precursor/DANCE)	0	24	5	423413	10096	BX93557.1/XM_421323.1	
FH (fumarate hydratase/fumarase)	0	2	2	420969	16465	NM_001006382.1	
FKBP9 (FK506 binding protein 9)	0	1	2	395652	2663	NM_204847.1	
FLJ13089	0	2	15	416883	21386	XM_001234127.1	
FN1 (fibronectin type 1)	0	21	7	396133	3994	XM_421868.2	
FNDC1 (fibronectin type III)	1	0	3	421589	15884	XM_419627.2	
GABA A receptor	1	0	4	396173	692	NM_205245.1	
GPI/PGI (glucose phosphate isomerase)	2	0	11	415783	4883	NM_001006128.1/AJ719291.1	

Table C.28 (continued)

Upregulated Pigeon Ortholog	Copy #		Chicken Map		GenBank Accession Numbers (* DFGI, not NCBI)
	WC	SR	Gg	Gga	
HMG2a (high mobility group)	0	1	4	396232	4513 NM_205295.1
HMGN1 (high mobility group)	0	1	1	395999	1189 DQ214361
HMOX1 (heme oxygenase-1)	1	0	1	396287	2039 NM_205344.1
HPC1 (histone promoter control 1)*	0	2	5	na	19407 CR407095.1/Q6C01D1*
HSRASLS (HRAS- like suppressor )	2	0	9	424903	8333 XM_422714.1
Hypothetical Protein	1	0	3	395778	4728 XM_419664.2
IKBKAP (inhibitor of kappa light polypeptide enhancer in B cells)	1	0	Z	427375	5317 AJ720452.1
JOSD3 (Josephin Domain Containing 3)*	0	1	1	419002	9636 XM_417196*
KIAA1432	1	0	Z	427226	22909 XR_026743.1
KS5 Protein	0	4	7	395762	2152 NM_204923.1
L27mt	13	0	18	422105	3524 CR385891/XM_420108.1
L53mt	1	0	2	420196	11961 DQ213764.1
LDH A (lactate dehydrogenase subunit A)	3	0	5	396221	4398 L76362.1
LDH B (lactate dehydrogenase subunit B)	18	23	1	373997	4149 L79957.1
LGTN (ligatin)	8	0	26	419845	7117 NM_001006322.1
LUM (Lumican Precursor/keratan sulfafate PG)	2	47	1	417891	4078 XM_416135.1/AF12525.1
LSM3 (U6 snRNA-associated Sm-like protein	2	0	12	416040	4477 DQ213305.1/XM_414380.1
MAEA (macrophage erythroblast attacher)	3	0	4	426024	8825 NM_001012604.1
MAN2BA (Mannosidase alpha, Class 2B, member 2)	1	0	4	422859	21923 XM_420805.1
MAST4 (microtubule associatr serine/threonine kinase)	0	1	Z	427169	14613 BX648980
MDC2 (MARVEL domain containing)*	1	0	2	na	21047 CR388870.1/Q28D82*
MGC75678	2	0	unk	426392	8084 BX935953.2/XM_424043.2
MGP (matrix Gla protein)	2	1	1	395912	540 DQ213333.1
Microsatellite 2G	0	1	unk	na	na DQ483836.1
MLRN (Myosin regulator light chain isoform L20-B )	1	0	2	396284	1738 DQ213857.1
MYH11 (SMC Myosin Heavy Chain)	0	8	14	396211	3225 NM_205274.1
MYLK (myosin, light chain kinase); telokin	0	6	7	396445	4091 NM_205459.1/M96655.1
MB (Myoglobin gene)	0	1	1	418056	1960 AY065779.1

Table C.28 (continued)

Upregulated Pigeon Ortholog	Copy #		Chicken Map		GenBank Accession Numbers (* DFGI, not NCBI)
	WC	SR	Gg	Loci	
ND1 (NADH subunit 1)	3	4	mt	807636	na AY274070.1
ND4 (NADH subunit 4)	1	12	mt	807643	na AB026193.1/AP003324.1/DQ648776.1/AY074886.2
NARD (Nardilysin)	2	0	8	424635	8172 NM_001031284.1
NAT13 (N-acetyltransferase 13)	13	0	1	418319	8828 NM_001030778.1
NDPK (nucleoside diphosphate kinase)	2	0	18	395916	2020 CLU61287
NDUFA10 (ubiquinone/NADH dehydrogenase)	0	1	7	424032	17230 NM_001031247
NHRa (Nuclear Hormone Receptor Activity)*	2	0	1	na	na TC_Q26DB4*
NUP37 (Nucleoporin)	0	1	1	418093	12366 XM_416326.1
PGD (phosphogluconate dehydrogenase)	2	0	21	419450	1282 AJ720644.1
PGM5 (phosphoglucomutase 5)	0	1	Z	427215	31282 XM_424802.1
PLS3 (plastin 3)	1	0	4	422222	8773 NM_001006431.1
POMP (proteasome maturation factor UMP1)	4	8	1	418925	5765 XM_417119.2
POSTN (periostin)	1	0	1	395429	5154 NM_001030541.1
PRDX1 (peroxiredoxin 1)	8	17	8	424598	5204 DQ215091.1/CR353582.1/XM_001233872.1
Prohibitin 2	2	2	1	771124	34471 AJ719351.1
PSMC1 (26S ATPase 4)	0	3	5	395804	1071 DQ214549.1
PSMC2 (26S ATPase 2)	11	0	1	417716	7131 NM_001006225.1
PSMC3 (26S ATPase 3)	3	0	5	423182	4649 NM_001031190.1/XM_421107.1
PSMD1 (26S non-ATPase)	0	6	9	424926	12538 NM_001012600.1
RAB1A (GTP binding)	0	3	3	421273	34237 XM_419342.2
RBMS1 (RNA binding motif)	2	0	7	395880	4540 XM_515853.2
RBP7 (retinol binding protein)	15	0	21	419447	9386 BX929289.2
rDNA complete repeating unit*	0	1	2	na	na TC_U13369.1*
RIKEN 3110009E18*	0	1	7	424274	10291 BX934594.2/XM_422121*
RIKEN 6030443O07*	0	1	6	na	na TC_AA80679*
RNH1 (ribonuclease/angiogenin inhibitor-1)	0	1	5	423111	1426 NM_001006473.1
RPL3 (ribosomal protein L3)	5	69	1	418016	990 XM_001234707.1/NM_001006241.1
RPL26L1	1	0	13	396400	31637 BX950296.2
RPL32	6	0	12	416122	9022 BX933805.1/DQ215085.1/BX931502.2/XM_414453.1



Table C.28 (continued)

Upregulated Pigeon Ortholog	Copy #		Chicken Map		GenBank Accession Numbers (* DFGI, not NCBI)
	WC	SR	Gg	Loci	
RPL7	2	0	2	420182	4878 NM_001006345.1
RPL7a	2	0	17	417158	4080 GGRPL7AM/CR385365.1
RPS3	2	0	1	419069	34333 AJ720513.1/CR3539973.1
RPS3A	2	1	4	422477	34252 XM_420443.1
RPS6	1	1	Z	396148	4389 NM_205225.1
RPS8	1	3	8	424584	34366 CR390136.1
RPN1 (Ribophorin I)	16	0	12	416017	20472 XM_414360.1
SAT1 (spermidine/spermime N1-acetyltransferase)	0	1	1	374006	9589 DQ13896.1
SCYE1 (small inducible cytokine)	1	0	4	422533	9194 XM_420496.2
SEC13L	1	0	12	416119	12581 XM_414450.1
SEC61A1 (Sec61 alpha subunit)	6	2	12	416023	39251 CR352798.1/CR406479.1/XM_414364.2
SEC61G (Sec61 gamma subunit)	0	3	2	776639	9217 NM_014302.3
SET (Mariner 1 transposase gene)	0	3	2	776268	14150 U52077.1
SLC25A5/A6	5	3	1	374072	4827 AJ7119767.1/NM_204231.2/DQ213941.1
SLN (sarcolipin)	0	2	1	na	5977 BX935884.1
SMOC2 (secreted modular Ca binding)	1	0	3	421569	12849 XM_419600.2
SPI1 (SFFV - SP1 oncogene activation)	1	0	5	395879	1313 CR523926.1
SPON1 (spondin1)	3	0	5	395657	3330 NM_204851.1
SRD5A2L2 (steroid alpha reductase)	0	1	4	422618	12132 XM_420576.2
SQLE (squalene epoxidase)	0	1	2	420335	22304 NM_001030953.1
TM167 (transmembrane 167)	2	0	Z	770790	35453 AJ720798.1
TNFaIP8 (TNF alpha induced protein 8)	4	0	Z	427384	11736 NM_001031441
TPM1 (alpha tropomyosin)	0	8	10	396366	4108 CHKAFTR0P6
TKT (transketolase)	6	0	12	415991	21333 XM_414333.1
UBN1 (ubiquitin)*	0	1	14	416398	17271 XM_414713*
VIM (vimentin)	1	0	2	420519	9346 NM_001018076
WAG-65N20	36	1	1	na	na AC084760.2/TC284349*
ZFP-100	0	1	7	na	na DQ216879.1
ZFP-183	0	1	27	419956	11596 NM_001004396.1

Table C.29 Differential Gene Expression in Individual Experiments

Upregulated Ortholog	WC 1 n=91	SR 1 n=126	Chi Square p=	WC 2 n=131	SR 2 n=151	Chi Square p=	WC 3 n=141	SR 3 n=140	Chi Square p=	WC 4 n=86	SR 4 n=84	Chi Square p=
16S rRNA	0	3	0.1383	0	4	0.0606	1	0	0.3182	0	0	na
ACAT1	0	0	na	0	0	na	1	0	0.3182	0	0	na
ACSL1	0	0	na	0	0	na	0	0	na	1	0	0.3216
ACTA2	0	2	0.2273	0	5	0.0356	0	0	na	0	0	na
ACTB	2	1	0.3820	4	0	0.0306	1	0	0.3182	0	0	na
ACTR10	0	1	0.3943	0	0	na	0	0	na	0	0	na
ACTR3	0	0	na	0	0	na	0	1	0.3147	0	0	na
ACVR1	0	0	na	0	0	na	0	1	0.3147	0	0	na
ADAMSTS15	2	0	0.0946	0	0	na	0	0	na	-	-	na
ALDH9A1	5	0	0.0078	0	0	na	0	0	na	0	1	0.3102
AMD1	0	0	na	0	0	na	0	1	0.3147	0	0	na
ANP32A	1	0	0.2382	0	0	na	0	0	na	0	1	0.3102
ANXA2	0	0	na	7	0	0.0040	0	0	na	0	0	na
APTX	0	0	na	0	0	na	0	0	na	0	1	0.3102
ARRB	0	1	0.3943	0	0	na	0	0	na	0	0	na
ATP4B	0	0	na	0	1	0.3508	0	0	na	0	0	na
ATPase6/8	0	2	0.2273	0	0	na	1	0	0.3182	1	0	0.3216
BAC CH261-10P22	0	0	na	0	1	0.3508	0	0	na	0	0	na
BAC CH261-124L19	0	0	na	2	0	0.1276	6	0	0.0136	6	0	0.0137
BAC CH261-138K4	0	0	na	0	0	na	2	0	0.1573	2	0	0.1597
BAC CH261-15P9	0	0	na	0	1	0.3508	0	0	na	0	0	na
BAC CH261-16J12	0	1	0.3943	0	0	na	0	0	na	0	0	na
BAC CH261-20I24	0	0	na	0	3	0.1862	0	0	na	0	0	na
BAC CH261-187N23	0	1	0.3943	0	0	na	0	0	na	0	0	na
BAC CH261-46G16	0	0	na	0	2	0.1862	0	0	na	0	0	na
BAC CH261-32O1	0	0	na	0	4	0.0606	0	0	na	0	0	na
BAC CH261-71A16	0	0	na	0	4	0.1048	0	0	na	0	0	na
BAC CH261-93G5	0	0	na	0	1	0.3508	0	0	na	0	0	na

Table C.29 (continued)

Upregulated Ortholog	WC 1	SR 1	Chi Square	WC 2	SR 2	Chi Square	WC 3	SR 3	Chi Square	WC 4	SR 4	Chi Square	p=
	n=91	n=126	p=	n=131	n=151	p=	n=141	n=140	p=	n=86	n=84	p=	
C10orf58	0	0	na	0	0	na	1	0	0.3182	0	0	na	na
C20orf45	0	0	na	0	0	na	1	3	0.3104	0	3	0.0770	0.0770
C3AR1	0	1	0.3943	0	0	na	0	0	na	0	0	na	na
CANX	0	1	0.3943	0	0	na	0	0	na	0	0	na	na
CCND2	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
CCT5	0	0	na	0	0	na	0	0	na	1	1	0.9866	0.9866
CCT6A	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
CCT8	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
CENPC	0	0	na	0	1	0.3508	0	0	na	0	0	na	na
chEST380b13	0	1	0.3943	0	0	na	0	0	na	0	0	na	na
Clone 38f16	0	0	na	0	0	na	1	0	0.3182	0	0	na	na
Clone BU37787	0	0	na	1	0	0.2821	0	0	na	-	-	na	na
Clone BU447803	0	0	na	0	0	na	1	0	0.3182	-	-	na	na
Clone BX268708	0	0	na	0	0	na	0	1	0.3147	-	-	na	na
Clone Q553U2	0	0	na	0	0	na	0	1	0.3147	-	-	na	na
CNDP2	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
CNRC01	1	0	0.2382	0	0	na	0	0	na	0	0	na	na
COPG	0	0	na	0	0	na	0	0	na	1	0	0.3216	0.3216
CO I	0	3	0.1383	0	2	0.1862	0	0	na	0	0	na	na
CO II	2	8	0.1501	2	8	0.0877	5	6	0.7492	2	3	0.6307	0.6307
COTL1	0	0	na	0	0	na	0	5	0.0236	0	0	na	na
COL5A2	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
COL1A2	5	0	0.0078	0	0	na	0	0	na	0	0	na	na
CPSF2	0	0	na	0	0	na	5	0	0.0246	11	0	0.0007	0.0007

Table C.29 (continued)

Upregulated Ortholog	WC 1	SR 1	Chi Square	WC 2	SR 2	Chi Square	WC 3	SR 3	Chi Square	WC 4	SR 4	Chi Square	p =
	n=91	n=126	p =	n=131	n=151	p =	n=141	n=140	p =	n=86	n=84	p =	
CRTAP	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
CXCL12	1	0	0.2382	8	0	0.0021	4	0	0.0447	0	1	0.3102	0.3102
CYTB	0	12	0.0025	0	4	0.0606	0	10	0.0012	0	10	0.0010	0.0010
CYNF	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
DACH1	0	0	na	0	0	na	5	0	0.0246	4	0	0.0455	0.0455
DCN	0	0	na	0	0	na	2	0	0.1573	0	0	na	na
DGAT2	0	0	na	0	0	na	4	0	0.0447	6	0	0.0137	0.0137
DNAJA2	0	1	0.3943	0	0	na	0	0	na	0	0	na	na
EEF1A1	0	0	na	0	0	na	1	0	0.3182	0	0	na	na
EIF3S	0	0	na	1	0	0.2821	0	0	na	0	1	0.3102	0.3102
EIF4A2	0	12	0.0025	0	4	0.0606	3	0	0.0827	3	0	0.0841	0.0841
EIF4E2	0	0	na	1	0	0.2821	1	0	0.3182	0	0	na	na
EIF4G2	0	1	0.3943	0	0	na	0	0	na	2	0	0.1597	0.1597
ENO1	9	0	0.0003	21	0	0.0000	0	0	na	0	0	na	na
Enzyme Inhibitor	0	0	na	0	0	na	2	0	0.1573	-	-	na	na
EXOC7	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
Ezrin/VIL2	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
FABP4	0	2	0.2273	0	0	na	2	0	0.1573	0	0	na	na
FBLN5	0	0	na	0	24	0.0000	0	0	na	0	0	na	na
FH	0	0	na	0	0	na	0	2	0.1544	0	0	na	na
FKBP9	0	0	na	0	1	0.3508	0	0	na	0	0	na	na
FLJ13089	0	0	na	0	0	na	0	2	0.1544	0	0	na	na
FN1	0	0	na	0	0	na	0	18	0.0000	0	3	0.0770	0.0770
FN3	0	0	na	0	0	na	1	0	0.3182	0	0	na	na
GABA	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
GPI/PGI	0	0	na	1	0	0.2821	1	0	0.3182	0	0	na	na
HMG2a	0	1	0.3943	0	0	na	0	0	na	0	0	na	na
HMG14	0	0	na	0	0	na	0	1	0.3147	0	0	na	na

Table C.29 (continued)

Upregulated Ortholog	WC 1		SR 1		Chi Square		WC 2		SR 2		Chi Square		WC 3		SR 3		Chi Square		WC 4		SR 4		Chi Square	
	n=91	n=126	p=	n=131	n=151	p=	n=141	n=140	p=	n=86	n=84	p=	n=86	n=84	p=									
HMOX1	0	0	na	1	0	0.2821	0	0	na	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
HPC1	0	0	na	0	0	na	0	0	na	0	2	0.1544	0	0	na	0	0	na	0	0	0	0	na	na
HSRASLS	0	0	na	2	0	0.1276	0	0	na	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
Hypothetical Protein	0	0	na	0	0	na	1	0	na	0	0	0.3182	0	0	na	0	0	na	0	0	0	0	na	na
IKBKAP	0	0	na	1	0	0.2821	0	0	na	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
JOSD3	0	1	0.3943	0	0	na	0	0	na	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
KIAA1432	0	0	na	0	0	na	1	0	na	0	0	0.3182	0	0	na	0	0	na	0	0	0	0	na	na
KS5 Protein	0	0	na	0	0	na	0	4	na	0	0	0.0432	0	0	na	0	0	na	0	0	0	0	na	na
L27mt	4	0	0.0175	9	0	0.0011	0	0	na	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
L53mt	0	0	na	0	0	na	1	0	na	0	0	0.3182	0	0	na	0	0	na	0	0	0	0	na	na
LDHA	0	0	na	3	0	0.0615	0	0	na	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
LDHB	1	19	0.0004	12	1	0.0007	3	1	0.0007	3	1	0.3173	2	2	na	0	0	na	2	2	0	0	na	na
LGTN	0	0	na	0	0	na	6	0	na	0	0	0.0136	2	0	na	0	0	na	2	0	0	0	0.1597	na
LSM3	2	0	0.0946	0	0	na	0	0	na	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
LUM	1	12	0.0099	0	35	0.0000	1	0	0.0000	1	0	0.3182	0	0	na	0	0	na	0	0	0	0	na	na
MAEA	0	0	na	0	0	na	2	0	na	0	0	0.1573	1	0	na	0	0	na	1	0	0	0	na	na
MAN2BA	0	0	na	1	0	0.2821	0	0	0.2821	0	0	na	0	0	na	0	0	na	0	0	0	0	0.3216	na
MAST4	0	1	0.3943	0	0	na	0	0	na	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
MDC2	0	0	na	0	0	na	1	0	na	0	0	0.3182	0	0	na	0	0	na	0	0	0	0	na	na
MGC75678	2	0	0.0946	0	0	na	0	0	na	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
MGP	0	1	0.3943	0	0	na	2	0	na	0	0	0.1573	0	0	na	0	0	na	0	0	0	0	na	na
Microsatellite 2G	0	0	na	0	1	0.3508	0	0	0.3508	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
MYH11	0	0	na	0	8	0.0075	0	0	0.0075	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
MYLK	0	1	0.3943	0	5	0.0356	0	0	0.0356	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
MB	0	0	na	0	1	0.3508	0	0	0.3508	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
MYRN	0	0	na	1	0	0.2821	0	0	0.2821	0	0	na	0	0	na	0	0	na	0	0	0	0	na	na
ND1	2	2	0.7415	1	1	0.9196	0	0	0.9196	0	0	na	0	0	na	0	0	na	0	0	1	1	0.3102	na
ND4	0	2	0.2273	0	5	0.0356	1	3	0.0356	1	3	0.3104	0	2	na	0	0	na	0	2	2	0	0.1500	na

Table C.29 (continued)

Upregulated Ortholog	WC 1	SR 1	Chi Square	WC 2	SR 2	Chi Square	WC 3	SR 3	Chi Square	WC 4	SR 4	Chi Square	p=
	n=91	n=126	p=	n=131	n=151	p=	n=141	n=140	p=	n=86	n=84	p=	
NARD	0	0	na	0	0	na	2	0	0.1573	0	0	na	na
NAT13	0	0	na	0	0	na	7	0	0.0076	6	0	0.0137	0.0137
NDPK	0	0	na	1	0	0.2821	1	0	0.3182	0	0	na	na
NDUFA10	0	0	na	0	0	na	0	1	0.3147	0	0	na	na
NHRA	0	0	na	0	0	na	2	0	0.1573	-	-	na	na
NUP37	0	1	0.3943	0	0	na	0	0	na	0	0	na	na
PGD	0	0	na	2	0	0.1276	0	0	na	0	0	na	na
PGM5	0	1	0.3943	0	0	na	0	0	na	0	0	na	na
PLS3	0	0	na	0	0	na	1	0	0.3182	0	0	na	na
POMP	0	1	0.3943	0	0	na	1	3	0.3104	3	4	0.6761	0.6761
POSTN	0	0	na	0	0	na	1	0	0.3182	0	0	na	na
PRDX1	0	6	0.0348	7	1	0.0182	1	9	0.0097	0	1	0.3102	0.3102
Prohibitin 2	2	0	0.0946	0	0	na	0	2	0.1544	0	0	na	na
PSMC1	0	0	na	0	0	na	0	2	0.1544	0	1	0.3102	0.3102
PSMC2	10	0	0.0001	1	0	0.2821	0	0	na	0	0	na	na
PSMC3	0	0	na	3	0	0.0615	0	0	na	0	0	na	na
PSMD1	0	0	na	0	0	na	0	0	na	0	6	0.0116	0.0116
RAB1A	0	0	na	0	0	na	0	2	0.1544	0	1	0.3102	0.3102
RBMS1	0	0	na	0	0	na	2	0	0.1573	0	0	na	na
RBP7	0	0	na	0	0	na	8	0	0.0042	7	0	0.0042	0.0042
rDNA repeat	0	0	na	0	1	0.3508	0	0	na	-	-	na	na
RIKEN 3110009E18	0	1	0.3943	0	0	na	0	0	na	0	0	na	na
RIKEN 6030443O07	0	0	na	0	1	0.3508	0	0	na	-	-	na	na
RNH1	0	1	0.3943	0	0	na	0	0	na	0	0	na	na
RPL3	0	0	na	0	0	na	5	39	0.0000	0	30	0.0000	0.0000
RPL26L1	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
RPL32	4	0	0.0175	0	0	na	1	0	0.3182	1	0	0.3216	0.3216
RPL7	0	0	na	1	0	0.2821	1	0	0.3182	0	0	na	na
RPL7a	1	0	0.2382	1	0	0.2821	0	0	na	0	0	na	na

Table C.29 (continued)

Upregulated Ortholog	WC 1	SR 1	Chi Square	WC 2	SR 2	Chi Square	WC 3	SR 3	Chi Square	WC 4	SR 4	Chi Square	p=
	n=91	n=126	p=	n=131	n=151	p=	n=141	n=140	p=	n=86	n=84	p=	
RPS3	0	0	na	2	0	0.1276	0	0	na	0	0	na	na
RPS3A	0	0	na	2	0	0.1276	0	1	0.3147	0	0	na	na
RPS6	0	0	na	1	0	0.2821	0	0	na	0	1	0.3102	
RPS8	0	1	0.3943	1	2	0.6469	0	0	na	0	0	na	na
RPN1	10	0	0.0001	6	0	0.0079	0	0	na	0	0	na	na
SLN	0	0	na	0	2	0.1862	0	0	na	0	0	na	na
SAT1	0	0	na	0	1	0.3508	0	0	na	0	0	na	na
SCYE	0	0	na	0	0	na	1	0	0.3182	0	0	na	na
SEC13L1	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
SEC61A	1	1	0.8164	1	0	0.2821	3	0	0.0827	1	1	0.9866	
SEC61G	0	2	0.2273	0	1	0.3508	0	0	na	0	0	na	na
SET	0	3	0.1383	0	0	na	0	0	na	0	0	na	na
SFFV	1	0	0.2382	0	0	na	0	0	na	0	0	na	na
SLC25A5/A6	0	3	0.1383	4	0	0.0343	1	0	0.3182	0	0	na	na
SMOC2	0	0	na	0	0	na	1	0	0.3182	0	0	na	na
Sp1 TFC	0	1	0.3943	0	0	na	0	0	na	0	0	na	na
SPON1	0	0	na	0	0	na	3	0	0.0827	0	0	na	na
SRD5A2L2	0	0	na	0	1	0.3508	0	0	na	0	0	na	na
SQLE	0	0	na	0	0	na	0	1	0.3147	0	0	na	na
TM167	1	0	0.2382	0	0	na	1	0	0.3182	0	0	na	na
TNFaIP8	0	0	na	0	0	na	1	0	0.3182	3	0	0.0841	
TPM1	0	0	na	0	0	na	0	4	0.0432	0	4	0.0406	
TKT1	6	0	0.0035	0	0	na	0	0	na	0	0	na	na
UBN1	0	0	na	0	1	0.3508	0	0	na	-	-	na	na
VIM	0	0	na	1	0	0.2821	0	0	na	0	0	na	na
WAG-65N20	1	1	0.8164	0	0	na	23	0	0.0000	12	0	0.0004	
ZFP-100	0	0	na	0	0	na	0	0	na	0	1	0.3102	
ZFP-183	0	0	na	0	0	na	0	0	na	0	1	0.3102	

Table C.30 Chi Square Analysis of Unidentified Difference Products

Transcript ID	WC 1	SR 1	Chi Square	WC 2	SR 2	Chi Square	WC 3	SR 3	Chi Square	Total WC	Total SR	Chi Square	p =
	n=91	n=126	p=	n=131	n=151	p=	n=141	n=140	p=	n=363	n=417	p=	
Contig 1	8	0	0.0007	2	0	0.1276	0	0	na	10	0	na	0.0006
Contig 2	0	0	na	0	0	na	0	12	0.0004	0	12	0.0011	0.0011
Contig 3	0	0	na	0	2	0.1862	0	0	na	0	2	0.1864	0.1864
Contig 4	1	0	0.2382	1	0	0.2821	0	0	na	2	0	0.1291	0.1291
Contig 5	3	0	0.0401	0	0	na	0	0	na	3	0	0.0629	0.0629
WC1	1	0	0.2382	0	0	na	0	0	na	1	0	0.2835	0.2835
WC2	1	0	0.2382	0	0	na	0	0	na	1	0	0.2835	0.2835
SR1	0	1	0.3943	0	0	na	0	0	na	0	1	0.3505	0.3505
SR2	0	1	0.3943	0	0	na	0	0	na	0	1	0.3505	0.3505
SR3	0	1	0.3943	0	0	na	0	0	na	0	1	0.3505	0.3505
SR4	0	1	0.3943	0	0	na	0	0	na	0	1	0.3505	0.3505
SR5	0	1	0.3943	0	0	na	0	0	na	0	1	0.3505	0.3505
SR6	0	1	0.3943	0	0	na	0	0	na	0	1	0.3505	0.3505
SR7	0	1	0.3943	0	0	na	0	0	na	0	1	0.3505	0.3505
SR8	0	1	0.3943	0	0	na	0	0	na	0	1	0.3505	0.3505
SR9	0	1	0.3943	0	0	na	0	0	na	0	1	0.3505	0.3505
WC3	0	0	na	1	0	0.2821	0	0	na	1	0	0.2835	0.2835
WC4	0	0	na	1	0	0.2821	0	0	na	1	0	0.2835	0.2835
WC5	0	0	na	1	0	0.2821	0	0	na	1	0	0.2835	0.2835
SR10	0	0	na	0	1	0.3508	0	0	na	0	1	0.3505	0.3505
SR11	0	0	na	0	1	0.3508	0	0	na	0	1	0.3505	0.3505



Table C.30 (continued)

Transcript ID	WC 1		SR 1		Chi Square		WC 2		SR 2		Chi Square		WC 3		SR 3		Chi Square		Total WC		Total SR		Chi Square		
	n=91	n=126	p=	n=131	n=151	p=	n=141	n=140	p=	n=363	n=417	p=													
SR12	0	0	na	0	1	0.3508	0	0	na	0	1	0.3505	0	0	1	0.3505	0	0	0	0	1	0.3505	0	0	0.3505
SR13	0	0	na	0	1	0.3508	0	0	na	0	1	0.3505	0	0	1	0.3505	0	0	0	0	1	0.3505	0	0	0.3505
SR14	0	0	na	0	1	0.3508	0	0	na	0	1	0.3505	0	0	1	0.3505	0	0	0	0	1	0.3505	0	0	0.3505
SR15	0	0	na	0	1	0.3508	0	0	na	0	1	0.3505	0	0	1	0.3505	0	0	0	0	1	0.3505	0	0	0.3505
SR16	0	0	na	0	1	0.3508	0	0	na	0	1	0.3505	0	0	1	0.3505	0	0	0	0	1	0.3505	0	0	0.3505
SR17	0	0	na	0	1	0.3508	0	0	na	0	1	0.3505	0	0	1	0.3505	0	0	0	0	1	0.3505	0	0	0.3505
SR18	0	0	na	0	1	0.3508	0	0	na	0	1	0.3505	0	0	1	0.3505	0	0	0	0	1	0.3505	0	0	0.3505
SR19	0	0	na	0	1	0.3508	0	0	na	0	1	0.3505	0	0	1	0.3505	0	0	0	0	1	0.3505	0	0	0.3505
SR20	0	0	na	0	1	0.3508	0	0	na	0	1	0.3505	0	0	1	0.3505	0	0	0	0	1	0.3505	0	0	0.3505
WC6	0	0	na	0	0	na	1	0	0.3182	1	0	0.2835	1	0	0	0.2835	1	0	1	0	0	0.2835	1	0	0.2835
WC7	0	0	na	0	0	na	1	0	0.3182	1	0	0.2835	1	0	0	0.2835	1	0	1	0	0	0.2835	1	0	0.2835
WC8	0	0	na	0	0	na	1	0	0.3182	1	0	0.2835	1	0	0	0.2835	1	0	1	0	0	0.2835	1	0	0.2835
SR21	0	0	na	0	0	na	0	0	0.3147	0	0	0.3505	0	1	1	0.3505	0	1	0	1	1	0.3505	0	1	0.3505
SR22	0	0	na	0	0	na	0	0	0.3147	0	0	0.3505	0	1	1	0.3505	0	1	0	0	1	0.3505	0	1	0.3505

## APPENDIX D: ANNOTATION TABLES

Table D.31 Distribution of Differentially Expressed Pigeon Genes by Chicken (Gg) Chromosome Map

Chromosome	#Gg Genes	% Genome	# SR Genes	% SR of Gg	% SR	# WC Genes	% WC of Gg	% WC
1	2192	14.3	13	0.593	17.6	15	0.684	18.1
2	1440	9.4	9	0.625	12.2	9	0.411	10.8
3	1272	8.3	3	0.236	4.1	6	0.274	7.2
4	1149	7.5	4	0.348	5.4	6	0.274	7.2
5	973	6.3	7	0.719	9.5	5	0.228	6.0
6	574	3.7	2	0.348	2.7	3	0.137	3.6
7	543	3.5	8	1.473	10.8	2	0.091	2.4
8	547	3.6	1	0.183	1.4	2	0.091	2.4
9	465	3.0	2	0.430	2.7	2	0.091	2.4
10	438	2.9	1	0.228	1.4	1	0.046	1.2
11	398	2.6	2	0.503	2.7	1	0.046	1.2
12	341	2.2	1	0.293	1.4	8	0.365	9.6
13	353	2.3	1	0.283	1.4	1	0.046	1.2
14	448	2.9	2	0.446	2.7	1	0.046	1.2
15	395	2.6	2	0.506	2.7	0	0.000	0.0
16	56	0.4	0	0.000	0.0	0	0.000	0.0
17	311	2.0	0	0.000	0.0	1	0.046	1.2
18	342	2.2	0	0.000	0.0	3	0.137	3.6
19	322	2.1	0	0.000	0.0	1	0.046	1.2
20	364	2.4	0	0.000	0.0	0	0.000	0.0
21	251	1.6	0	0.000	0.0	3	0.137	3.6
22	123	0.8	0	0.000	0.0	0	0.000	0.0
23	235	1.5	0	0.000	0.0	0	0.000	0.0
24	198	1.3	0	0.000	0.0	1	0.046	1.2
25	129	0.8	0	0.000	0.0	0	0.000	0.0
26	260	1.7	0	0.000	0.0	1	0.046	1.2
27	280	1.8	1	0.357	1.4	0	0.000	0.0
28	215	1.4	0	0.000	0.0	0	0.000	0.0
29	na	na	0	na	0.0	0	na	0.0
30	na	na	0	na	0.0	0	na	0.0
31	na	na	0	na	0.0	0	na	0.0
32	na	na	0	na	0.0	0	na	0.0
33	na	na	0	na	0.0	0	na	0.0
34	na	na	0	na	0.0	0	na	0.0
35	na	na	0	na	0.0	0	na	0.0
36	na	na	0	na	0.0	0	na	0.0
37	na	na	0	na	0.0	0	na	0.0
38	na	na	0	na	0.0	0	na	0.0
39	na	na	0	na	0.0	0	na	0.0
mt	13	0.1	3	23.077	4.1	1	0.046	1.2
W	2	0.0	1	50.000	1.4	0	0.000	0.0
Z	739	4.8	5	0.677	6.8	6	0.274	7.2
Unknown	na	na	6	na	8.1	4	na	4.8
<b>Totals:</b>	<b>15,368</b>	<b>100</b>	<b>74</b>		<b>100</b>	<b>83</b>		<b>100</b>