UNIVERSITY OF CALIFORNIA, SAN DIEGO

The post-terminal differentiation fate of RNAs revealed by next-generation sequencing

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Gloria Kuo Lefkowitz

Committee in Charge:

Professor Benjamin D. Yu, Chair
Professor Richard Gallo
Professor Bruce A. Hamilton
Professor Miles F. Wilkinson
Professor Eugene Yeo

2012

The Dissertation of Gloria Kuo Lefkowitz is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2012

DEDICATION

Ma and Ba, for your early indulgence and support.

Matt and James, for choosing more practical callings.

Roy, my love, for patiently sharing the ups and downs of this journey.

EPIGRAPH

It is foolish to tear one's hair in grief,

as though sorrow would be made less by baldness.

~Cicero

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CNV.....copy number variation

DMSO.....dimethyl sulfoxide

DNA ..... deoxyribonucleic acid

EDA.....ectodysplasin-A

EDAR .....ectodysplasin-A receptor

eQTL..... expression quantitiative trait loci

FGF.....fibroblast growth factor

FGFR .....fibroblast growth factor receptor

GWAS .....genome wide association studies

HF ..... hair follicle

HS .....hair shaft

INPP5B..... Type II inositol-1,4,5-trisphosphate 5-phosphatase

MALAT-1.....metastasis associated lung adenocarcinoma transcript 1

NEAT1 ..... noncoding nuclear-enriched abundant transcript 1

NHEK .....normal human epidermal keratinocyte

Krt.....keratin

KRTAP .....keratin associated protein

MapK.....mitogen activated protein kinase

miRNA ..... microRNA

OCRL .....oculocerebrorenal syndrome of Lowe 1

qRT-PCR .....quantitative real time polymerase chain reaction

QTL ..... quantitiative trait locus

Ras ..... rat sarcoma

RNA..... ribonucleic acid

SNP..... single nucleotide polymorphism

TCHH .....trichohyalin

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge Professor Benjamin Yu for his support as the chair of my committee. I am grateful for the thoughtful mentorship and the opportunity to thrive and grow as a scientist in his lab. I would like to thank my committee members who graciously volunteered their time and expertise. I would like to acknowledge the members of the Yu Lab for their daily intellectual discussions and camaraderie. I am grateful to our many anonymous subjects for samples, especially for repeated donations during optimization of experiments.

I would like to thank my early mentors, Dr. Susan McNabb for her early vociferous support of my tentative foray into bench science, and Dr. Jon Cooper for nurturing my growth as a scientist. My early experiences in research were shared with fellow student, Dr. Christina Takanishi, who made those times much more enjoyable. I would like to note Dr. Bruce Hamilton and Carolyn Goard for encouraging me to join the program. I will reflect upon this adventure fondly.

Parts of Chapter 2, 3 and 4 appear in the published material in PLoS One, 2011. Preliminary experiments were performed by A. Mukupadhyay. Construction of the hair library was done at the Scripps Genomics Core. Bioinformatic analysis was performed by C.A. Cowing-Zitron. The dissertation author was the primary investigator and author of this material. Parts of Chapter 5 concerning the nail will be included in a manuscript currently in preparation.

VITA

2003 Bachelor of Science, Cell & Molecular Biology
Bachelor of Science, Psychology University of Washington, Seattle

2012 Doctor of Philosophy, Biomedical Sciences
University of California, San Diego

PUBLICATIONS

Lefkowitz GK, Mukhopadhyay A, Cowing-Zitron CA, Yu BD. The post-apoptotic fate of RNAs revealed by next-generation sequencing. PLoS ONE 6(11): e27603. doi:10.1371/journal.pone.0027603

Lefkowitz GK, Gleeson JG. Aurora A moonlights in neurite extension. Nat Cell Biol. 2009 Sep;11(9):1053-4.

Kuo G, Arnaud L, Kronstad-O'Brien P, Cooper JA. Absence of fyn and src causes a reeler-like phenotype. J Neurosci. 2005 Sep 14;25(37):8578-86. *Note: Cover photo*

Dewey EM, McNabb SL, Ewer J, **Kuo GR**, Takanishi CL, Truman JW, Honegger HW. Identification of the gene encoding bursicon, an insect neuropeptide responsible for cuticle sclerotization and wing spreading. Curr Biol. 2004 Jul 13;14(13):1208-13. *Note: Featured as Editor's Choice in Science 305: 575 (30 July 2004), G Chin, Ed.*

FIELDS OF STUDY

Major Field: Genetics, Development, Differentiation, Nucleic Acids, Gene Expression, Hair and Skin Biology

Studies in Transcription Factors in Neuronal Development
Professor Eric E. Turner

Studies in Aneuploidy of the Human Brain
Professor Jerold Chun

Studies in Neuronal Development and Human Genetics
Professor Joseph G. Gleeson

ABSTRACT OF THE DISSERTATION

The post-terminal differentiation fate of RNAs revealed by next-generation sequencing

by

Gloria Kuo Lefkowitz

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2012

Professor Benjamin D. Yu, Chair

Advances in RNA sequencing technology allow for potential applications in areas of personal genomics and molecular diagnostics. However, a rate-limiting step of widespread analysis of the transcriptome is the restricted access to viable tissue samples. The hair follicle is an actively regenerating mini organ and the hair shaft is the output of this organ. Here we report on the discovery of hair shafts as a rich, stable source of mRNA, microRNA, and other RNA species that are amendable to extraction and analysis by standard laboratory and next-generation sequencing techniques.

Chapter 1 discusses historical methods to assay correlations between genetic differences and phenotypic differences. Chapter 2 argues for the suitability of the hair follicle as a model organ of study. Genetic changes and other health states are known to affect the physical appearance and other properties of the hair shaft. Chapter 3

describes the demonstration of lineage specific RNA in hair shafts and our creation of a small RNA library of short sequences expressed in the human hair shaft. The library serves as a starting point to identify and predict expressed genes that may be biologically relevant. Some of the well represented RNAs are the Let-7 family of microRNAs, keratins, and keratin binding proteins. Other genes known to be specific to the cortex, medulla and cuticle of the hair are also present. Our library contains an abundance of microRNA compared with libraries made from other tissues. Chapter 4 discusses the process of cornification and compares the coverage of hair shaft RNA to that of a viable keratinocyte cell line. This also led us to investigate whether smaller RNAs were more stable than longer RNAs in the hair shaft in Chapter 5. As hair on the human scalp has an unusually long growth phase, we analyzed RNA from hair that grew out over several months. We report that hair segments representing up to a year of growth contain extractable RNA that is amendable to downstream applications. In addition, we find that another keratinzed tissue, nail, also retains RNA. Chapter 6 discusses these findings and potential applications in the field of personalized medicine and retrospective screening.

Chapter 1: The study of variation and genetics

1.1 Introduction

One of the great opportunities in the post-genomic age is the potential reality of deciphering the relationship between genomic variation and complex phenotypic variation. This will lead to greater understanding of the mechanisms of individual susceptibility to disease and contribute towards the development of personalized medicine (Collins et al., 2003). The link between genomic variation and phenotypic variation—from innocuous differences such as hair morphology to more medically relevant conditions, such as disease, lies in part, in inter-individual differences present in the DNA sequences. These may then be translated into changes of gene products such as proteins and non-coding RNA, or may affect gene regulation, such as altering the timing, localization, or duration of RNA expression levels. Understanding how these steps lead to disease may also reveal therapeutic areas of intervention.

1.2 Characterizing genomic variation

Genomes and coding regions are conserved

The first eukaryotic genomes sequenced were of the brewer's yeast, *Saccharomyces cervisiae*, and two model invertebrates, *Drosophila melanogaster* and *Caenorhabditis elegans*. Coding sequences and the core proteome are largely similar between the two invertebrates and suggest that the variation lies in gene regulation (Rubin et al, 2000). Despite the complexity observed between species, the size of most

sequenced mammalian genomes is similar, at a little over three billion bases for both mouse and human (International Human Genome Sequencing Consortium, 2004) (Mouse Genome Sequencing Consortium, 2002). Between humans and our closest relative, the chimpanzee, it is estimated that there is a 1.06% difference in genomes (Chimpanzee Sequencing and Analysis Consortium, 2005).

The human genome is composed of 3.1 billion base pairs. Only 1.1 to 1.4% accounts for the approximately 30,000 genes that code for protein. Roughly 50% of the genome consists of repetitive "junk" sequences whose function is still poorly understood. Between any two unrelated human individuals, the estimated difference ranges from 0.1 to 0.03% (International Human Genome Sequencing Consortium, 2004).

The degree of similarity seen between sequenced genomes illustrates the conservation of our basic biology. These comparisons suggest that small differences in the genomic code are responsible for the larger phenotypic variations observed between organisms and between individuals within a species. As genome sequencing became more accessible, the next major hurdle is to interpret and determine which variations are relevant to disease and diagnosis.

The genomic basis of phenotypic variation

Plato's Republic records the general practice of selective breeding of livestock with desired characteristics to obtain offspring with the same desired characteristics. This practice, he reasoned, extended to humans themselves in an early proposal for eugenics (Plato, 1998). Some of the first observations of the heritable nature of

phenotypes were formally made in the late eighteenth century and early nineteenth century by an Austrian monk, Gregor Mendel. What distinguishes Mendel's modest treatise from his contemporaries was the idea of each parent having two alleles, and transmitting either one randomly to its offspring. This concept of independent assortment and segregation of various pea traits in a discrete and mathematical pattern was published in 1865 but received little attention in his lifetime (Mendel, 1967). It was not until the turn of the century, the significance of his work was recognized and replicated by biologists in the study of humans and other organisms such as the fly.

Another concept was encompassed by the term "genetics," first coined by William Bateson to describe the "elucidation of phenomena of heritability and variation." He defines variation as "the occurrence of differences between the structure, the instincts, or other elements" between offspring and parent (Bateson 1894). While offspring are similar to parents, there will still be traits that were not observed in the parents, whether easily perceptible or not. The types of variation were further grouped into continuous traits, which changed gradually, and discontinuous traits, in which no intermediate form is observed. Discontinuous traits, which have a larger magnitude of change between individuals, are easier to observe than continuous traits. Of the human, animal and insect samples, there were specimens that had duplicated or missing segments; such as extra ribs and polydactyly, which could be variations in patterning (Bateson 1894). These definitions are useful today in considering detectable variations, even as the scope of detection has greatly increased.

The fruit fly became a prominent model to study genetics in the early 1900s in the laboratory of Thomas Hunt Morgan at Caltech. In early studies of mutations in fruit flies, the behavior of genes was further elucidated to include effects of recombination, dosage compensation between the sexes, and other genetic actions. The types of phenotypes (also termed allelomorphs) arising from mutagenesis screens were described and classified as a range of loss-of-function and gain-of-function phenotypes (Muller 1932).

A more recent example of the complexity of interactions is illustrated in the phenotype of flies lacking the neuronal adhesion molecule, fasciculin; in a wildtype background, flies develop normally, but in a background where the Abelson tyrosine kinase is also lacking, there are severe neurite extension defects (Elkins et al., 1989). Despite not physically interacting and not being functionally redundant, when both are missing, a severe developmental defect becomes unmasked.

This effect can also be seen in conditions where there should be a large degree of genetic similarity. Due to being highly inbred lab strains, individuals in model systems can be considered genetic replicates. In *C. elegans*, this is illustrated by the loss of the muscle transcription factor *tbx-9;* despite having isogenic strains grown in identical conditions, the population will invariably exhibit a range of phenotypes from wildtype to severe. The range of phenotypes can be attributed to random developmental noise and can be predicted in part by the variation in the expression of two other genes, *tbx-8* and *daf-1* that may act as buffers to the loss of *tbx-9* (Burga et al 2011).

These complications are a large issue in the study of human diseases in a genetically heterogeneous population. One illustration of this is Lowe's syndrome, a condition that encompasses multiple systems, including the central nervous system, vision, kidneys, and blood. In a majority of cases with identified mutations, the mutations have been in the gene *OCRL1* (Oculocerebrorenal syndrome of Lowe 1) on the X chromosome. Some individuals with Dent's Disease 2, a condition that primarily affects the kidneys, and may be considered a phenotypically milder version of Lowe's syndrome, also have mutations in OCRL1 (Hichri et al, 2011). This suggests that there many be different protective mechanisms, genetic or not, in place in different individuals. The multiple functions of the OCRL1 protein product, in trafficking, as a phosphatase and RhoGAP, also make elucidating the disease causing mechanisms *in vitro* or with model organisms less straightforward (Lowe 2005).

The mouse model illustrates these complications nicely. When the homolog, *Ocrl*, is deleted in mouse, none of the phenotypes observed in human were present (Janne et al., 1998). They find that *Inpp5b*, a closely related gene with similar functional domains, has different expression patterns in the mouse and likely compensates for loss of *Ocrl*. Only by deleting both genes and misexpressing the human form, *INPP5B*, in the mouse, do they find some kidney defects (Bothwell, et al., 2011). In the lines generated, the transgenic INPP5B was expressed at higher levels than endogenous Inpp5. This suggests that there is a feedback mechanism that may upregulate expression of *Inpp5b* in the absence of *Ocrl* that exists in mouse. This

compensatory mechanism may exist to a limited extent in individuals with Dent's Disease 2 and be even less effective in individuals that have Lowe's Syndrome.

The differences that cause such phenotypic variation between species and even among individuals within a species likely arise from small changes that result in more dramatic gene expression changes in strength, duration, timing, or location (Rubin et al, 2000). This variation can be in the form of small changes, such as single nucleotide polymorphisms (SNPs) which differ by only a single base pair, small repetitive repeats, insertions, inversions or deletions less than 1kb in size, to larger differences such as copy number variations (CNVs) which encompass differences of hundreds to millions of base pairs, and chromosomal rearrangements.

The most studied variation, SNPs, occur about 1 every 300 bases and there are 10 million in any individual. These are the basis for genome wide association (GWAS) studies that have been undertaken to discover the link between complex phenotypes and genomic variation. The underlying hypothesis of GWAS studies is that common variants may cause common disease. Incremental risks from multiple SNPs have been identified in disease conditions and other phenotypes such as height and BMI. There remains the problem that there are many common SNPs which in aggregate account for a small percentage of risk. While alleles related to such conditions as macular degeneration have been shown to be causative, other traits have many assocated SNPs, but overall small contribution to actual outcome (Manolio, et al., 2009).

In addition, more sensitive sequencing technology may soon elucidate how other types of variations contribute to gene expression variation and resulting phenotypes. Copy number variations (CNVs) are larger in size, usually ranging from 1kb to 1Mb in size. They can be further classified as segmental duplications or low-copy repeats, duplications, deletions, insertions, and translocations. These structural differences cover up to 12% of the genome and thus may account for millions of nucleotides of heterogeneity. These regions may encompass entire genes and regulatory elements (Redon, et al., 2006) (Wong et al., 2007). For example, one of the common detected variations resides in the amylase gene which breaks down starch (Iafrate et al., 2004). Higher number of repeats of this gene results in more enzyme being produced. Correspondingly, more repeats are found in agricultural societies as a result of positive selection (Perry et al., 2007). Current technologies to detect CNVs are still limited but they add an additional layer of complexity beyond primary DNA sequence.

The ENCODE project has annotated the human genome for functional regions, transcribed genes, *cis*-regulatory regions and other features of the genome (The ENCODE Project Consortium 2011). The degree of conservation of coding regions across species and individuals, suggests that much of the phenotypic variation is due to non-coding regions that likely have a regulatory role in gene expression.

In a screen for transcriptional enhancers, looking first in conserved regions between vertebrates, the number and action of experimentally validated enhancers have been dramatically expanded (Pennacchio et al., 2006). The species-specific

action of enhancers and promoters identified hint at human specific adaptations; these may underlie sensitive gene regulation during brain development. The validation of novel promoters also underscores the incomplete understanding of transcriptional regulation.

Genomic variations between species and individuals within species result in observable phenotypic variations such as height and weight, to more complex traits such as blood pressure, warfarin sensitivity, and disease susceptibility. As more individual genomes are sequenced and newer technologies enable more refined structural characterization, the connection between complex genetic variations and phenotypic differences may become better understood.

1.3 Linking genomic variation with expression variation

One of the classical methods of studying the effect of genetic variation on measurable phenotypes is quantitative trait linkage (QTL) mapping. Loci and functional variants for blood clotting, triglyceride levels, and body mass index have been identified and replicated in multiple studies (Almasy and Blangero, 2010). However, relatively few of these studies have progressed to the point of identifying the gene or underlying mechanism responsible for the observed phenotypes.

Ribonucleic acid (RNA) levels may also be considered as another source of variability that may be traced back to variations in DNA. Transcript and isoform abundance serve as quantitative traits and may be mapped back to genetic variations (Rockman and Kruglyak 2006). With the wide availability of microarrays, it is

possible to look at the global expression profile at a particular time or of a particular tissue sample.

Changes in expression levels, timing, and duration may be regulated by cis-regulatory elements that may also be revealed in the RNA transcripts. Gene expression variation has been found to be responsible for adaptive phenotypes. There is genome wide enrichment of positive selection and eQTLs in the Youruban HapMap as well as similar trends in the other maps examined (Kudaravalli et al., 2009). A major advantage of eQTL mapping is that differences in environmental factors may also be reflected in data (Gilad, Rifkin, and Pritchard, 2008). Studies have begun analysis of mRNA levels as a model for analyzing complex phenotypes and diseases (Cheung and Spelman, 2002).

New sequencing methods, termed "next-generation," to assay global transcripts, in particular RNA-Seq, have several improvements over microarray hybridization methods. Some of the advantages are that there is de novo sequencing by synthesis, which makes the annotation of novel isoforms, SNPs, and poly-A isoforms distinguishable; there is vastly more data generated; and there is a linear correlation between signal and abundance (Wang, Gerstein, and Snyder 2009). As yet, no clear "gold standard" for high-throughput sequencing platforms nor analysis algorithms, both commercial and publically available exists.

Expression profiling may also have applications in a medical screening. Individuals that are heterozygous for disease genes may not exhibit symptoms readily detectable by physical examination or routine medical tests. A demonstration of this is

the detection of common gene expression “markers” in lymphoblastoid cell lines of ataxia carriers. In addition to baseline differences, cells derived from asymptomatic carriers also responded differently to ionizing radiation than counterparts from non-carriers (Watts et al., 2002). This suggests that the differences in DNA lead to expression differences in RNA that may not necessarily manifest in a medically apparent manner. Screening for expression differences may uncover risk indicators of undiagnosed conditions and may allow for prophylactic treatment.

The integration of information from genotypic and phenotypic studies has been demonstrated with a large scale analysis of liver samples (Schadt et al., 2008). The liver is a source of phenotypic variation that occurs in a variety of conditions, such as diabetes, hypertension and obesity. From their samples, they find correlations potential causal relationships between SNPs in the DNA and expression levels –both cis- and trans- acting associations. The comparison between studies on human and mouse expression yields some correlations. However, perhaps what this study illustrates well is that integration of these findings with other GWAS studies—they find in common, the gene RPS26 but not ERBB3 as a susceptibility gene for Type I diabetes, and also SORT1 and CELSR2 as candidates for coronary heart disease and plasma low density levels.

As more populations are being sequenced, such as such as African bushmen and supercentenarians (individuals living past 114 years of age), the link between genomic complexity and phenotypes will be better understood (Schuster et al., 2010) (Sebastiani et al, 2012). While great progress has been made studying severe

phenotypes resulting from single mutations, other genetic interactions are less clear cut, with each allele contributing some risk; this more complicated interaction is likely what would be seen in personalized genetics. What these studies illustrate is that the complexity of potential interactions is present in the variation of expression. However, expression data comes from cell lines, biopsied tissue, or bodily fluids. As sequencing costs and time continue to decrease, genomic and transcriptomic analysis will become more prevalent. A potential future challenge will be to obtain sample tissues efficiently from large populations.

Chapter 2: Hair

2.1 The external hair

One of the most readily apparent illustrations of the genetic influence on phenotypic variation is our surface, namely, skin and hair. It is the first line of defense against the environment and, as humans settled into different climes, diverges in pigmentation, texture, and morphology. The hair is a keratinized outer appendage present on all mammals that provides warmth, protection, and tactile information.

Several properties make the hair follicle an attractive model to study biological processes. The hair follicle has long been considered a mini organ. It is self-regenerative and continuously cycling. There are also many hair follicles per individual, which lends itself well to examination of site specific effects (Schneider, Schmidt-Ulrich, and Paus et al 2009). Alterations in the appearance of hair have been considered a sign of disease or poor health (Whiting and Dy, 2006). The hair follicle is an easily accessible model to study development, genetics, cell cycling, and other aspects of biology.

2.2 Evolutionary context of hair

Recent insights have been made in the evolution of hair. The availability of genomes from the chicken and green anole lizard supplements observations made earlier from developmental comparative embryology. The common ancestor of amniotes likely had keratin genes. In mammals, these evolved to become cysteine-rich structural proteins expressed mainly in the skin, hair and nails. Keratin sequences from

reptiles and birds have some similarity to mammalian keratins and are also expressed in the same regions (Eckhart et al., 2008).

Skin evolved as a thin layer to protect against desiccation, adorned with alpha-keratinzed bumps in early terrestrial tetrapods. Basal amniotes had epidermal structure with both glands and alpha-keratinzed bumps. After synapsid/sauropsid divergence, synapsids keep glands, while sauropsids largely lose glands. Hair may have evolved from sweat glands and first functioned as a wick to draw contents to surface or for thermal insulation and regulation to enable mammals to survive harsher climate changes (Dhouailly 2009) (Morioka, 2005). The skin and claws of the green anole lizard express keratin and keratin like genes. In contrast, the claws of *xenopus laevis* likely developed later and conversely, utilize other genes (Maddin et al, 2009).

In mammals, there is great phenotypic diversity in hair types. In *mus musculus*, there are four different pelage hairs which vary in size, length and bends. The zigzag hairs, which make up roughly 70% of mouse pelage hair provides insulation, while the sparser guard hairs may function as sensors. They also have larger specialized vibrissae, or whiskers, which function in tactile perception. Other hairs, as classified by morphology and body site also exist.

In *homo sapiens*, there is a great reduction in the size and prominence of body hair, termed vellus hairs, although not a reduction in the number of hair follicles. This finer hair may have aided thermoregulation by sweating of a bipedal organism during hot temperatures (Ruxton and Wilkinson 2011). Another hypothesis is that group living conditions may have necessitated loss of hair for parasite evasion (Pagel and

Bodmer 2003). In a study of volunteers, bedbugs took longer to find a feeding spot and were more quickly detected on unshaved forearms compared to shaved arms; this difference was more pronounced in volunteers with naturally denser hair (Dean and Siva-Johty 2011). Hair is very amendable to evolutionary pressures—changes in morphology and other physical features may serve specialized functions advantageous to their environment.

In humans, individual differences in diameter, cross sectional shape, texture, composition, and strength have been well described (Franbourg et al., 2003) (Robbins, 1994). Microscopically, pigmentation, cuticle shape, degree of curl, cross-sectional shape, and other physical characteristics have been used in forensics. As with mice, the physical characteristics of human hairs also differ by body site—beard hairs generally are ribbon-like while growth and cycling of other hairs, such as eyebrows are markedly different from scalp hair. (Ogle and Fox, 1998).

2.3 Hair follicle development and cycling

The hair shaft and hair follicle represent a signaling crossroad between mesenchyme, neural crest, and ectoderm. The cells in this region interact to go through the stages of proliferation, differentiation, migration, and other basic stages of development. Hair follicle development begins in the embryonic stages. Critical for its development are the signaling events that occur between the epithelial cells and the underlying mesoderm. This pattern of induction is conserved in other organs such as the nails and teeth. Signals from the mesoderm induce a condensation of ectodermal

cells at regularly spaced distances. These cells will form the hair placode and grow downward into the epidermis to form the hair follicle while the mesenchymal cells underneath form the dermal papilla (Millar, 2002).

Hair follicles cycle through phases of growth, quiescence, shedding, and regrowth. The growth phase, termed anagen, results in the production of the hair shaft. During catagen, cells become dormant and the bulb is extruded towards the surface. During the telogen stage, the bulb atrophies and is eventually shed. New follicle growth is sometimes present and may contribute to pushing out the older hair shaft (Paus and Foitzik, 2004). There are variations in cycling that depend on site of the hair follicle, age, and health—human scalp hair has a long anagen phase and older individuals may suffer from progressively shorter anagen phases. (Robbins 1994)

2.4 Structures of the hair follicle

The hair shaft itself is made of cells that begin their differentiation in the bulb and become part of three components: an outer cuticle, the cortex, and the medulla. The cuticle makes up the outermost layer of scales, and is made of overlapping cells. The cortex is made up of keratin bundles. During cornification, EM images of keratin bundle formation in the cortex and cuticle can be observed (Morioka, 2004). The cortex consists of an arrangement of parallel keratin filaments and contains melanin granules (Robbins, 1994). The medulla at the core of the hair shaft may consist of single to multiple rows of vacuous cells. In humans scalp hair, there is variation in the

appearance of the medulla—it can be observed as a single row, a discontinuous row, or absent (Ogle and Fox, 1998).

The hair shaft is surrounded by the outer root sheath (ORS) and the inner root sheath (IRS), supporting structures of the hair follicle that do not incorporate into the hair shaft itself. The IRS can be further divided into Henle and Huxley's layers and the inner root sheath cuticle which also undergo cornification. The IRS originates from the matrix and extends halfway along the length of the hair follicle. It may act as a mold that shapes the growing hair shaft and also undergoes cornification. The ORS extends all the way to the epidermis and is hardened at the upper regions. (Morioka, 2004)

The dermal papilla is of mesenchymal origins. Its inductive properties are necessary for regrowth and regeneration of the hair follicle. It lies under the hair bulb. The sebaceous glands, arrector muscles, nerve endings, and blood vessels are structures associated with the hair follicle unit. Sebaceous glands secrete oils that contribute to the hair shaft water resistance. Arrector muscles and nerve endings provide feedback from environmental stimuli. Innervating blood vessels nourish the hair follicle. (Robbins 1994)

2.5 Genetic variations affecting hair

A combination of genetic and environmental factors affects the hair follicle and the resulting hair shaft. One of the first studies to illustrate the heritability of hair morphology showed that curly hair and wavy hair were dominant to straight hair in a group of Europeans (Davenport and Davenport 1908). Since then, many other studies

have demonstrated the genetic contribution to hair qualities such as stability, curl, and thickness (Medland, Zhu, and Martin, 2009). Some hair abnormalities are part of more pervasive conditions affecting other ectodermal structures. Other less dramatic alleles have inadvertently been discovered in animal lines established by breeders and pet fanciers.

Mutations in keratins and other structural proteins of the inner root sheath

Keratins are the most prominent of hair proteins and contribute to its shape and strength. Not surprisingly, hair keratin mutations result in hair phenotypes. Many of the keratin genes with mutations causing hair abnormalities are expressed in the inner root sheath. Studies of hair shape have noted that there is a correlation between the shape of the hair follicle and its orientation relative to the epidermis and the resulting external hair shaft (Sriwiriyanont et al., 2011)

A point mutation in the coding region of the Krt74 gene has been described in families with wooly short brittle hair (Shimomura et al., 2010). Mutations in a related keratin, Krt71, also cause hair curliness in mice and dogs as well (Runkel, et al., 2006) (Cadieu et al., 2009). It is likely that these changes in the structure of these keratins, highly expressed in the inner root sheath, may alter the shape of the hair shaft as it is being synthesized.

Studies show that SNPs in another structural gene, trichohyalin (TCHH), correlate with straight hair in European populations (Medland et al., 2009). TCHH is expressed in the inner root sheath and medulla where it acts as a cross-linker.

Granules of TCHH are exocytosed by medulla cells into the cortex during the formation of the hair shaft (Morioka 2005). Cell culture studies suggest that its transcription is enhanced by BMP4 and its transcript stabilized by laminin (Yamamoto et al., 2009).

FGF pathways

Fibroblast growth factors (FGFs) are secreted glycoproteins that bind to surface FGF receptors. These transmembrane receptors then dimerize and transphosphorylate each other. FGF signaling controls a broad range of functions; in the context of hair development, FGF5 likely regulates hair cycling. Point mutations and null mutations in the gene fibroblast growth factor 5 (FGF5) cause long silky hair in breeds of dogs, cats, and mice (Cadieu et al., 2009) (Drogemuller et al., 2007). Analysis of hair follicles in mice deficient in FGF signaling reveal that a longer anangen cycle and decrease in columns of the medulla may contribute to the observable hair phenotype (Hebert et al, 1994). An intronic SNP in fibroblast growth factor receptor 2 (FGFR2) is also weakly associated with thicker hair in Asian populations, however the molecular mechanism is unclear (Fujimoto et al., 2009).

Ras/MapK pathways

One of the key pathways downstream of fibroblast growth factor (FGF) signaling is the Ras/Raf/MapK pathway. Upon activation of the FGF receptor, the transcription factor Erk is activated, leading to transcription of its targets (Tsang and

Dawid, 2004). A notable example of signaling dysfunction is apparent in the brittle curly hair of individuals with mutations in the Ras/MapK pathways (Figure 2.2A). The spectrum of disorders, Costello syndrome, cardiofacial cutaneous (CFC) syndrome and Noonan syndrome, all result from mutations that activate the Ras/MapK pathway. Most of these mutations are *de novo* gain of function and autosomal dominant (Aoki et al., 2008). It is notable that the resulting signaling strength of these mutations, as determined by *in vitro* experiments on genetically engineered mice, correspond to the severity of various defects (Mukhopadhyay, Krishnaswami, and Yu 2011). This illustrates the less than straightforward action of genetic variation which may be dependent on unknown variables.

Hypohydrotic ectodermal dysplasia

In humans, hypohidrotic ectodermal dysplasia is a congenital condition characterized by sparse or absent hair, exocrine defects, and tooth malformations. In the X-linked form, males are affected while carrier females have milder symptoms: an overall reduction in hair follicle number, sweat glands, and occasionally missing or misshapen teeth. Mutations in ecdysoplasin (EDA), a transmembrane protein highly expressed in keratinocytes, hair follicles, and sweat glands, were first linked to X-linked hypohydrotic ectodermal dysplasia in 1996 (Kere et al., 1996). Since then, mutations in other genes that work together with EDA have been identified in inherited forms have been reported in families of Pakistani, Swedish, and Indian ancestry (Bal et al., 2007) (Lind et al., 2006) (Naeem, Muhammad, and Ahmad 2005).

EDAR signaling pathway and dysfunction

The ecdysoplasin receptor EDAR is an integral part of a defined signaling pathway that affects hair morphogenesis (Figure 2.2B). The ligand, ecdysoplasin A1 (EDA) is a transmembrane protein that belongs to the tumor necrosis family. Upon cleavage, a soluble extracelluar molecule is released and binds to its receptor, EDAR. This leads to the recruitment of the auxillary death domain containing protein EDARADD. EDAR signaling generally leads to activation and translocation of the transcription factor NFκB. Mice lacking EDA, EDAR, or EDARADD exhibit similar phenotypes. The three lines, *tabby*, *downless*, and *crinkled*, respectively, have less hair on the tails and behind the ears, tooth defects, and an absence of sweat glands (Botchkarev and Fessing, 2005).

Non-deleterious EDAR alleles

An EDAR variant, Val 370Ala, hi1ghly prevalent in the East Asian populations correlates with the thicker hair diameter. In vitro studies suggest this may be a result of differential modulation of EDAR signaling output (Fujimoto, 2008) (Mou et al., 2008). Transgenic mice overexpressing EDAR have thicker, coarser hair, as well as more branched sweat glands (Mou et al., 2008).

Strangely enough, there is evidence for selection of the EDAR allele prevalent in East Asian populations that reached fixation 100,000 years ago (Sabeti et al., 2007). The adaptive advantage of this is yet determined. Chang et al. suggest that other phenotypes seen in transgenic mice—in particular, larger and greater branching of

sweat glands may have been advantageous traits in humans colonizing cold and dry regions (2009). The potential adaptation of dentition differences, or an anti-parasitic function of thick straight hair also remain to be explored.

Non-genetic factors that affect hair morphology

Disease and nutritional states have been well documented to be related to visible alterations to hair characteristics. The general state of hair—luster, pigmentation, strength, and other gross features are related to the general health of the individual and have been used in clinical diagnosis (Whiting and Dy, 2006). Syndromic conditions such as diabetes, and systematic lupus erythromatosous have been associated with alopecia while other disorders, such as polycystic ovary syndrome are associated with hirsutism (Yell, Mbugagbaw, and Burge, 1996) (Lee and Zane, 2007).

2.6 Gene Expression in Hair Shafts as a Phenotype

In addition to the gross observable variations that correlate with genetic differences, gene expression itself is now being studied as a complex phenotype. We can consider the profile of which genes are expressed and to what extent they are expressed as an assayable phenotype. (Toung et al., 2011). However, there are notable challenges to obtaining information of diagnostic or prognostic relevance.

Owing to the large nature data sets obtained from assaying one or more transcriptomes, there will be a fair amount of expression noise—random changes in

expression that do not necessarily have an effect on the phenotype being studied. Relevant changes would be most likely contained in the tissue of interest. However, we show that there are genetic mutations and other states that cause global changes, including changes in the physical appearance of the hair shafts. This suggests that there can be organism wide expression changes in some conditions, such that differences will be present in the hair shaft. Hair shaft is also a highly accessible tissue. There is the challenge of obtaining tissue samples in a non-invasive manner. Harvesting hair for RNA extraction does not cause pain or discomfort unlike other tissues, such as blood or tissue biopsies of internal organs. Studying the hair follicle and the hair shaft as models of the effects of variation may yield novel insight into the interaction of genetic background and the environment.

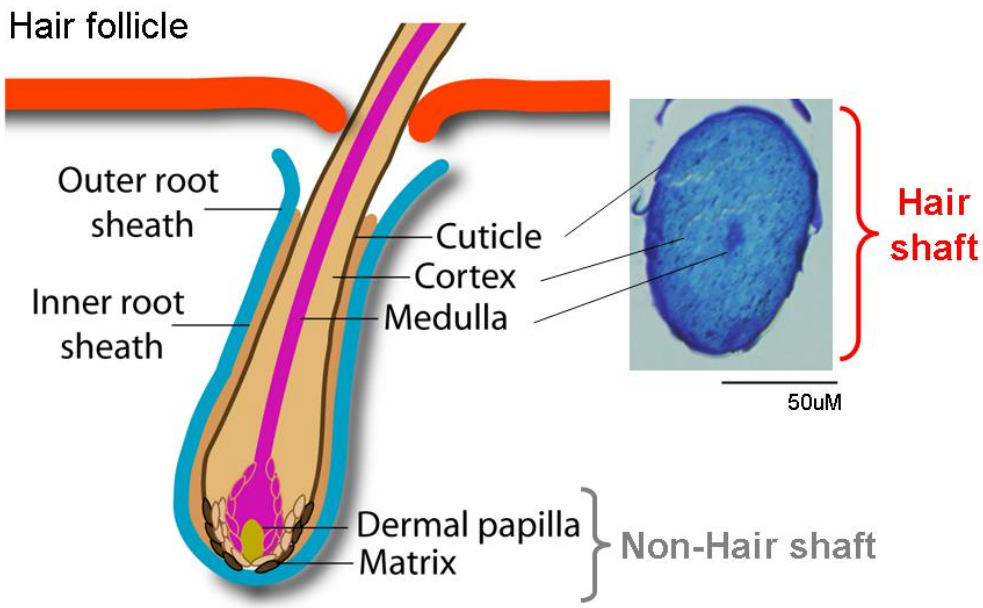

Figure 2.1: Hair Follicle

The figure on the left highlights the general features of the hair follicle which is embedded in the epidermis. Supporting structures include the outer and inner root sheaths. The cells of the matrix will receive signals from the dermal papilla. The matrix cells will divide and form the three components of the hair shaft, the cuticle, cortex, and medulla, which are arranged in concentric layers as seen with Touludine blue staining on a human hair cross section.

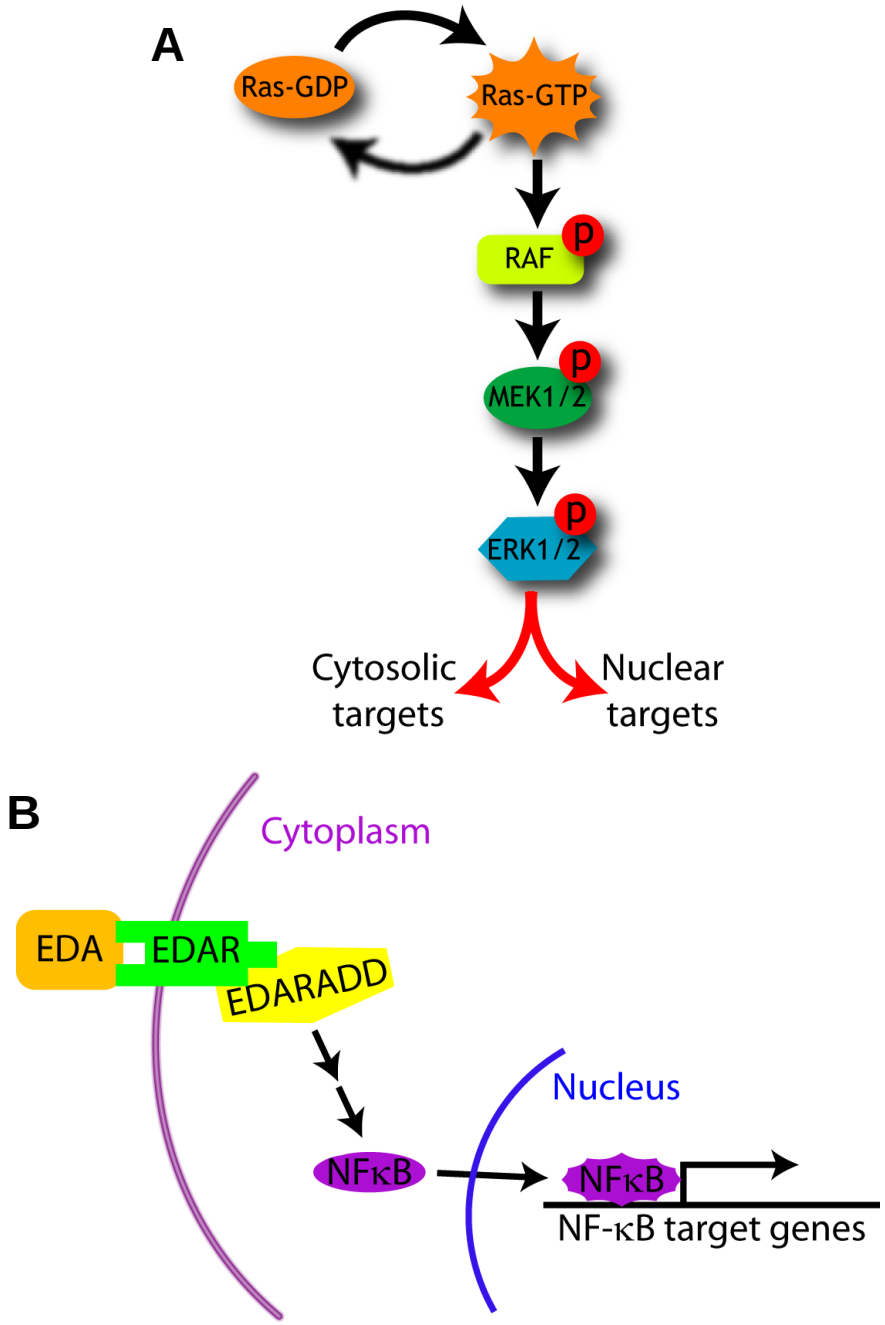

Figure 2.2: Signaling pathways affecting hair morphology
A) Mutations in the Ras signaling pathway that increase Ras signaling strength result in a spectrum of disorders, collectively categorized as the Noonan, Costello, and cardiofacial cutaneous syndromes that affect hair morphology, often resulting in short brittle hair. B) Inactivating mutations in EDA, EDAR, and EDARADD result in hypohidrotic dysplasia in humans and other animals.

Chapter 3: Hair shaft RNA and expression library

3.1 Introduction

The hair follicle produces the hair shaft in a process that incorporates terminally differentiated cells into a durable structure over time (Morioka 2005). Readily visible to the eye are the morphological changes that are affected by underlying genetic components. The external hair shaft represents the end product of a specialized programmed cell death program and may represent gene expression levels reflective of the whole organism. This led us to ask if the gene expression might also be preserved in the hair shaft.

3.2 Mouse hair shaft RNA

Cells of the matrix that will become components of the hair shaft, supporting components such as the outer root sheath, and the underlying mesenchymal dermal papilla express distinct mRNAs and microRNAs (Rendl, Lewis, and Fuchs 2005) (Yi et al, 2009). Based on these expressed on profiles, we tested whether RNAs specific to the hair follicle could also be detected in the external hair shaft and non-hair shaft transcripts excluded. As a comparison, we used whole skin, which contains all lineages of thhe hair follicle, surrounding epidermal, dermal and adipose tissue (Figure 3.1A).

Using quantitative real-time PCR (qRT-PCR), miRNAs were readily detected in the mouse hair shaft and at lower cycle numbers than skin. Relative to the control,

snoRNA-251, we find that *mir-203,* previously reported to be highly expressed in the epidermis relative to hair follicle, is ten fold less abundant in the hair shaft compared to whole skin (Yi et al, 2009). This indicates that small RNAs are sufficiently preserved in the hair shaft and detectable by qRT-PCR.

Next, we looked at mRNAs. In particular, we looked at genes known to be abundantly expressed in a cell-specific manner (Porter 2009) (Shimomura et al., 2003). Lineage specific keratins, KRT32, 34, and 35 and a central medulla gene, SH3D19, could be detected in the external hair shaft. Relative to whole skin, lineage specific genes were relatively more abundant in the hair shaft (Figure 3.1D, F). In contrast, the dermal papilla functions to regulate growth of the hair follicle but does not contribute cells to the hair shaft (Rendl, Lewis, and Fuchs 2005). The dermal papilla specific genes, BMP6 and SOX2, were significantly reduced in hair shaft compared to the whole skin (Figure 3.1 E, F). These findings indicate that detectable RNAs of three main cell types persist in the external hair of mouse and that the cornification process does not completely destroy RNAs.

We consider the possibility that residual RNA in mouse hair shaft may reflect the relatively rapid nature of hair follicle growth and turnover. A mouse hair cell undergoes one week of maturation compared to one to two months in human hair (LeBeau, Montgomery, and Brewer, 2011). The relatively short maturation time and scale of mouse hair may contribute to the persistence of cells in mouse external hair shafts rather than the lack of ribonuclease activity during and after cornification. To

address this we turned to human hair shafts, which have a longer maturation time and have been shown to be dead (Linch 2009).

3.3 Generation of a Human hair RNA-Seq library

Studies of human hair also pose some considerations. Expression studies of RNAs of the living human hair follicles have been limited and unlike laboratory mouse strains, there may be gene expression or detection biases arising from individual variation. We address these limitations by utilizing parallel sequencing of small RNAs isolated from human hair. To generate a comprehensive library of human hair RNAs and to account for some aspects of human variation, we pooled hair from the hair shafts of five individuals representing both genders, and variations in hair shape, age, and geographical origin (Table 3.1). Total RNA was extracted, pooled, ligated to adaptors designed for small RNA reads, reverse transcribed, amplified, size selected, and analyzed by Illumina based small RNA sequencing.

From this approach, 13.5 million high quality reads representing 1.2 million unique sequences were obtained. These reads were aligned to the human genome (hg18) and two small RNA libraries, miRBase release 16 and snoBase v3. A large portion of reads align to ten or more loci of the human genome due to low complexity of reads, highly repetitive sequences, or the presence of non-human RNA (Figure 3.2A). Of the reads that aligned to single loci, we found a high frequency of reads that were 22 nucleotides in length, the size of a mature miRNA (Figure 3.2B). We find that most RNA types are represented in the hair RNA library. There are messenger RNAs,

miRNAs, rRNAs, and other species and a small percentage of reads align to unannotated regions of the genome which may represent novel genes (Figure 3.2C). There were 7,193 mRNAs detected with read coverage of 10 or more, and of those, 251 mRNAs were found at 200 or more reads. For alignments that had 200 or more reads, we also identified 449 distinct miRNAs and 339 snoRNAs in the hair shaft library.

The most abundant small RNAs are members of the let-7 family (Figure 3.3A). Members of the Let-7 microRNA library are involved a large variety of essential cell functions such as cell division and differentiation (Roush and Slack 2008). The majority of mRNAs were mapped to keratin associated genes (Figure 3.3B). This family of proteins are involved in cross linking keratins and, unlike keratins, are unique to mammals (Wu, Irwin, and Zhang, 2008).

To verify that the sequences represent RNA and not contaminating DNA, we find that most of the sequences align to mRNA and other transcribed sequences (Figure 3.2C). If there were contaminating DNA, we would expect to see more unannotated or untranscribed regions which make up a larger portion of the genome. Most alignments to mRNA are in the coding orientation, as illustrated with an alignment of read orientation to clusters of genes in the KRTAP5 and KRTAP10 subfamilies (Figure 3.4). Primers used to amplify qPCR reactions were designed to cross exons when possible (Table 3.4).

We also wanted to investigate the specificity of the RNAs in the hair shaft library. Some keratins are specific to the structures of the inner root sheath which

does not contribute to the hair shaft, while others are expressed in the hair follicle in specific lineages that will eventually become the extruded hair shaft (Langbein et al, 2006) (Morioka 2005) (Figure 3.5A). We find that of these lineage specific keratins, the hair library contains most keratins specific to the hair shaft lineage, and only one of the inner root sheath (Figure 3.5B). For a comparison, we looked at neonatal human epidermal keratinocytes (NHEKs). These cells have been used as a model for skin differentiation (Pillai et. al, 2005). As such, these cells can be considered a tissue of similar origins as the hair follicle and suitable for comparison. A library of normal NHEK expression contains mRNA of both lineages (Figure 3.5C). This suggests that the more differentiated hair shafts express keratins that would exist in the developing hair follicle and while the less differentiated NHEK cells express keratins of both hair shaft and non-hair shaft lineages.

We also looked for expression of late cornificaiton genes. Filaggrin (FLG), is present, but other cornified envelope genes exclusive to skin, loricrin, involucrin, and keratinocyte differentiation-associated protein are absent. While necessary for the maturation of hair follicles, loricrin mRNA is only expressed in the IRS, infundibulum and isthmus, regions that do not become part of the mature hair shaft (de Viragh, Huber, Hohl 1994). Transglutamases and caspases, enzymes involved in initiating cornification in hair shaft keratinization are not present in the library. Involucrin and transglutamase-5 have been reported to have low expression in the hair shaft (Thibaut et al., 2005). Transglutamase-3 is hair-specific and may be essential in progressive scaffolding (Thibaut et al., 2009). Also notably absent from the hair library is

trichohyalin (TCHH), a major cross linking component in the IRS extruded as granules by the medulla of the hair shaft (Morioka 2005). These differences in protein versus transcript detection could be due to limitations of detection in constructing the RNA library such that some transcripts could be missed, or that these transcripts may have been synthesized earlier during development and then degraded by the time the hair shaft cells have completely cornified.

3.4 RNAs of the human hair shaft

In several gene annotation searches, , including a Set-Distiller batch tool, we find strong statistical associations with phenotypes and pathways affecting other organ systems. A 7,193 gene set represented the RNA in the hair shaft. In total, 377 pathways involved in environmental and pharmacologic signals and 391 pathways related to specific human disorders were associated with genes detectable in the hair shaft. Some of the relevant associations to phenotypes and compound exposure are summarized in Figure 3.6 and associations with disease are summarized in Figure 3.7. The specific associated genes present in our library are listed in Tables 3.5 and 3.6. While there are many skin related disorders that are associated with genes present in our library, what is most interesting is that conditions such as schizophrenia and renal disease with no apparent observable phenotypes reported in hair.

We wanted to compare the large number of associations found in the hair library with another set from a related source. A set of genes from a study on NHEKs consisting of roughly 11 thousand genes also yielded the same magnitude of

associations—there were some signatures that were unique to the hair shaft library, as denoted in Figure 3.6. This suggests that first, RNA present in the hair may be a marker for screening or medical testing, and second, the number of associations found in the hair library is comparable to other large data sets.

3.5 Methods

Animal and Human Specimens

All experiments and informed consents were performed and approved according to institutional guidelines established by the University of California, San Diego, Institutional Animal Care and Use Committee and the University of California, San Diego, Human Research Protection programs, Protocol ID# 091646. Adult mice (P20) were from a C57BL6 mixed background. Mouse hair was cut with scissors and visually inspected for the absence of skin fragments and hair bulbs.

The procedure and future use of samples were described to human donors. A copy of the signed consent form and a patient bill of rights were given to donors. Human hair was cut with scissors approximately 5 mm from the scalp. The cut hair was inspected by microscope and hair bulbs were removed from human hair samples. Human samples were de-identified and pooled.

Hair was washed once in 70% ethanol, and three times in DNase-free, RNase-free, molecular grade water. Hair was cut into segments, transferred to tubes, weighed,

and stored in a cool dry area protected from light. Stored hair was kept at -80℃ in RNAlater (Qiagen).

NHEK cells (Cat No C-001-5C, Life Technologies) were grown in EpiLife culture media (Cascade Biologics, Portland, OR) containing 0.06 mM calcium and defined growth supplement.

RNA extraction and isolation

We used two methods for RNA extraction. During initial experiments in mouse hair and human hair shaft RNA sequencing, Trizol was utilized with 1.0mm zirconia beads (BioSpec Products, Inc.). Subsequently, improved extraction was obtained with a modified lysis buffer of 7 M urea, 1% SDS, 0.35M NaCl, 10mM EDTA, 10 mM Tris-HCl at 55℃ (Gough, 1988). To facilitate breakdown of hair shafts, 100mM DTT was added and samples were shaken with 1.0mm zirconia beads (BioSpec Products, Inc.) during incubation. Following lysis, twice the volume of Trizol (Life Technologies) was added and RNA was extracted according to manufacturer's instructions. Briefly, 200ul of chloroform was added per 1ml of Trizol reagent. The supernatant was transferred to a new tube and an equal volume of isopropanol with sodium acetate and glycogen was added to precipitate RNA. The pellet was spun down and washed twice in 70% ethanol. The pellet was suspended in DNAse I buffer (New England Biolabs) and DNA was removed with a 30 mintue DNase digestion at 32℃ followed by addition of 10mM EDTA and 10 minute 65℃ heat inactivation. The sample was precipitated again in EtOH, washed with 70%

ethanol, and resuspended in water. RNA concentration and purity was determined with a Nanodrop 2000 Spectrophotometer (Thermo).

Similar buffers containing urea and detergents have been used to dissolve and extract hair protein (Winter et al., 1997). The effect of urea on protein is to destabilize disulfide bonds, one of the major components of keratinzed hair. In a concentration and pH dependent manner, urea also acts to inactivate major RNAses (Pace, Laurents, and Thompson, 1990). Urea is commonly used in gels as a stabilizing factor when resolving RNA (Locker 1979). These factors may have made the simple buffer containing urea at a neutral pH more suitable to break down hair fibers and retrieve RNA. However, it was also necessary to test that this method of extraction would yield RNA comparable to RNA from a standard method.

We used 6mm skin punches from mouse skin to compare the efficiency of RNA extraction between our method and Trizol. To determine if the addition of DTT was a factor in improving yields, we also added the same concentration of DTT to skin samples disassociated in Trizol. We found that the yields and quality of RNA was similar in the three conditions. Table 3.3 shows the concentrations of each RNA extraction resuspended in a final volume of 30ul, the average yield, and 260/280 ratios of three separate experiments. The yield and quality of RNA does not have significant variation between the methods. While there is a clear advantage to using the method described here for RNA from hair shafts, this suggests that both methods work similarly for ore conventional tissue.

Real-time PCR analysis for mRNAs, snoRNAs, and microRNAs

Primers for mRNA sequences were designed with Primer3. cDNA was made with Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Fermentas/Thermo) and real time PCR was performed with Maxima SYBR Green (Fermentas/Thermo). For small RNA analysis, cDNA was made with SABiosciences RT² miRNA First Strand Kit (Catalog No. 331401). Real-time PCR was performed with RT² SYBR® Green qPCR Mastermixes and small RNA primers from SABiosciences. All real time PCR were performed on a Roche 480 Lightcycler. Selected reactions were visualized on a 2% agarose gel with ethidium bromide. Oligo sequences used are listed in Table 3.4. Specificity of amplicons and potential DNA contamination were tested by melting curve analysis and control reactions containing only RNA as template (termed RT-Neg) respectively.

Preparation of RNA library

Fifty human hair shafts, equivalent to 30mg, were pooled to prepare total RNA. A yield of one ug of total RNA was isolated from pooled hair and used to generate a small RNA library. Preparation of the library was performed according to Illumina protocols for miRNA-Seq (v1.5.0). Briefly, 1 μg total RNA is ligated onto 5´-riboadenylated-3´-DNA adapter with 3 units of T4 RNA Ligase 2 (truncated) (New England Biolabs) in 1X T4RNL2 truncated Reaction Buffer with 8 mM MgCl2 and 20 units RNase OUT™ and incubated for 20ºC for 1 hour, followed by ligation to an unmodified RNA adapter with T4 RNA Ligase. This product is then primed with the

Illumina RT primer and reverse transcribed with Superscript II and the cDNA product is amplified in 18 cycles of PCR. PCR product is purified on 10% TBE PAGE gel and excised and eluted from the gel using standard crush and soak procedures, and finally ethanol precipitated using glycogen as carrier reagent.

Sequence Analysis

Sequencing was carried out on an Illumina Genome Analyzer IIx. Raw RNA-Seq data was generated in FASTQ format by the Genome Analyzer Pipeline. Quality scores were offset by 64 following Solexa-1.3+ standards. Preprocessing of the FASTQ data was done with the FASTX Toolkit package. Adaptor sequences were trimmed off and reads were required to have a minimum of 90% of bases with a quality score of 20 or higher. The library was then aligned with Bowtie to the human genome (NCBI36/HG18) from the UCSC Genome Browser (Langmead et al, 2009). Non-canonical and unscaffolded chromosomes were excluded.

Reads were aligned to mirBase release 16 and snoRNABase (Kozomara and Griffiths-Jones, 2011) (Lestrade and Weber, 2006). Alignments to both mature and pre-miRNA sequences in mirBase were counted. Reference sequences for both miRBase and snoRNAbase were artificially extended with 20 adenoside nucleotides on both ends to allow for alignment of alternatively spliced small RNA. Post-processing of reads was done with Samtools and BEDTools to generate expression counts (Li et al, 2009) (Quinlan and Hall 2010). Functional and compound associations were performed using Set-Distiller batch tool and Mouse Genome

Informatics (Stelzer et al., 2009) (Blake et al., 2011). Processing was done on the Triton Resource at the San Diego Supercomputer Center.

Portions of Chapter 3 are a reprint of the material as it appears in PLoS One 2011. Lefkowitz, GK, Mukupadhyay A, Cowing-Zitron, CA, Yu, BD. The dissertation/thesis author was the primary investigator and author of this paper.

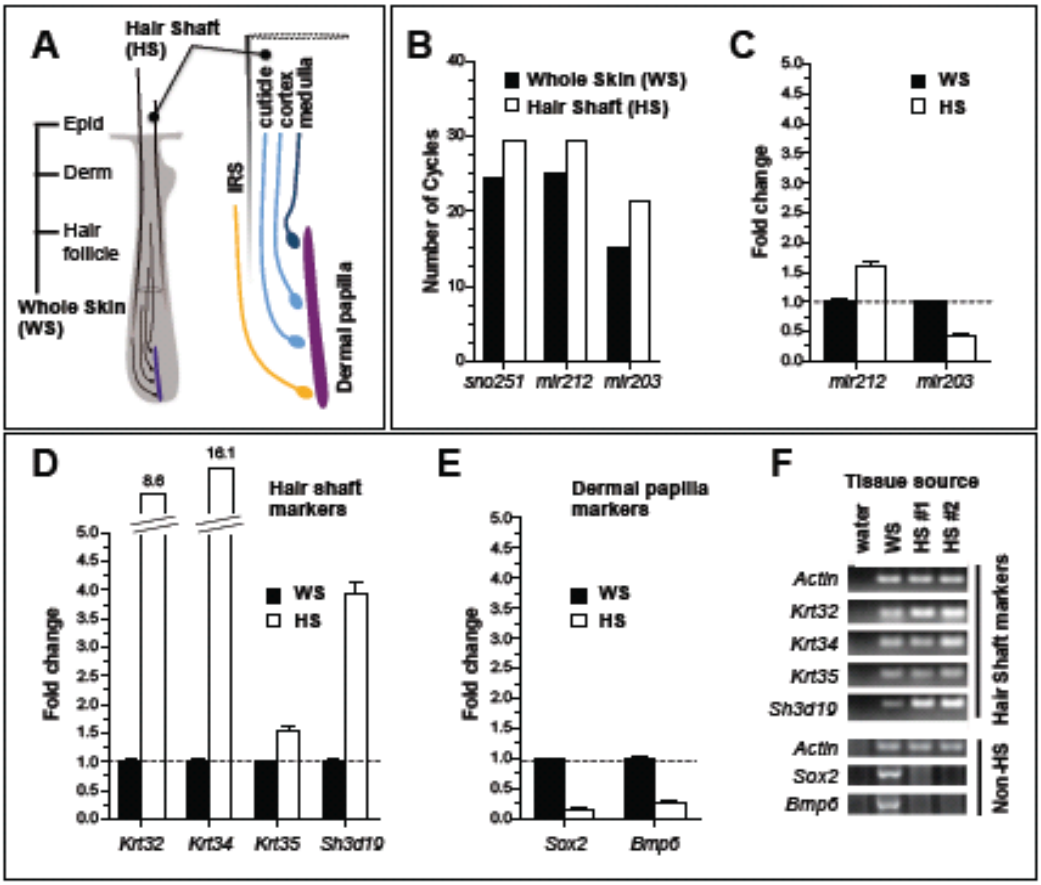

Figure 3.1: Mouse hair RNA

A) Schematic of hair follicle and lineages B) Cycle numbers of small RNAs in mouse hair and mouse skin C) Relative expression of mir-212 and mir-203 with sno251 as reference gene. Relative expression of hair shaft lineage genes (D) and dermal papilla genes (E) with beta-actin as reference gene. Products visualized on a gel (F) correspond with abundance differences seen by qRT-PCR.

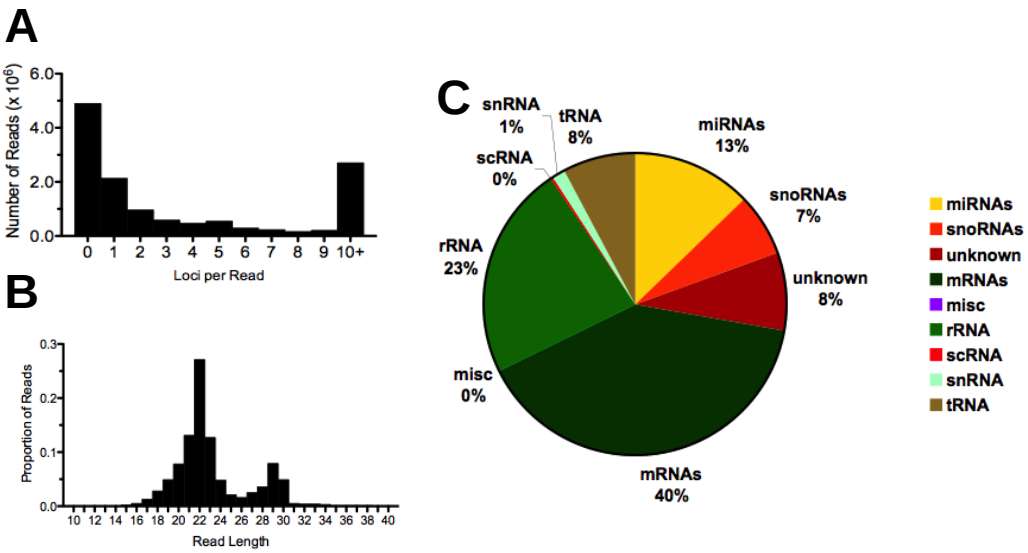

Figure 3.2: Summary of human hair shaft reads

A) Reads map to Hg18 arranged by loci. B) Histogram of unique reads plotted by read length. C) Many types of RNA are represented in the library, the most abundant being those classified as mRNAs.

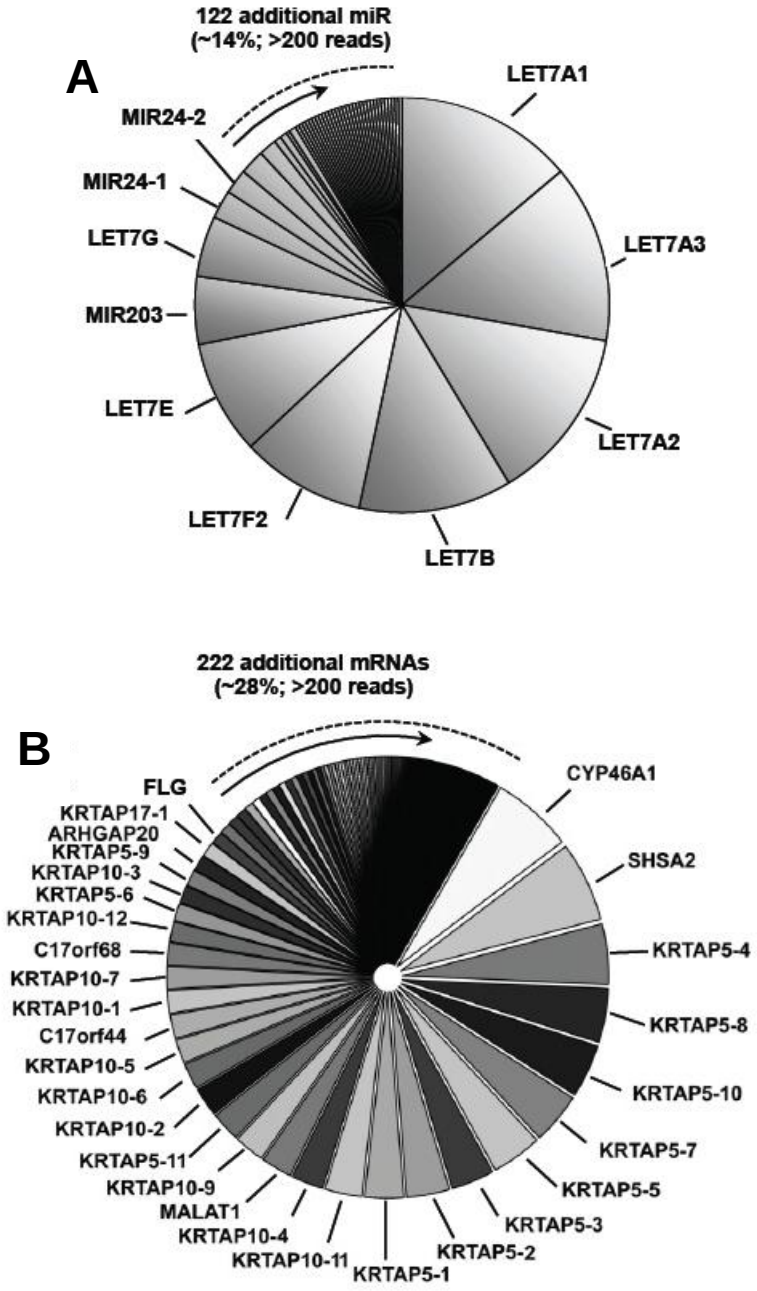

Figure 3.3: Complexity of human hair shaft reads
Reads that map to microRNA (A) and mRNAs (B) and arranged by abundance.

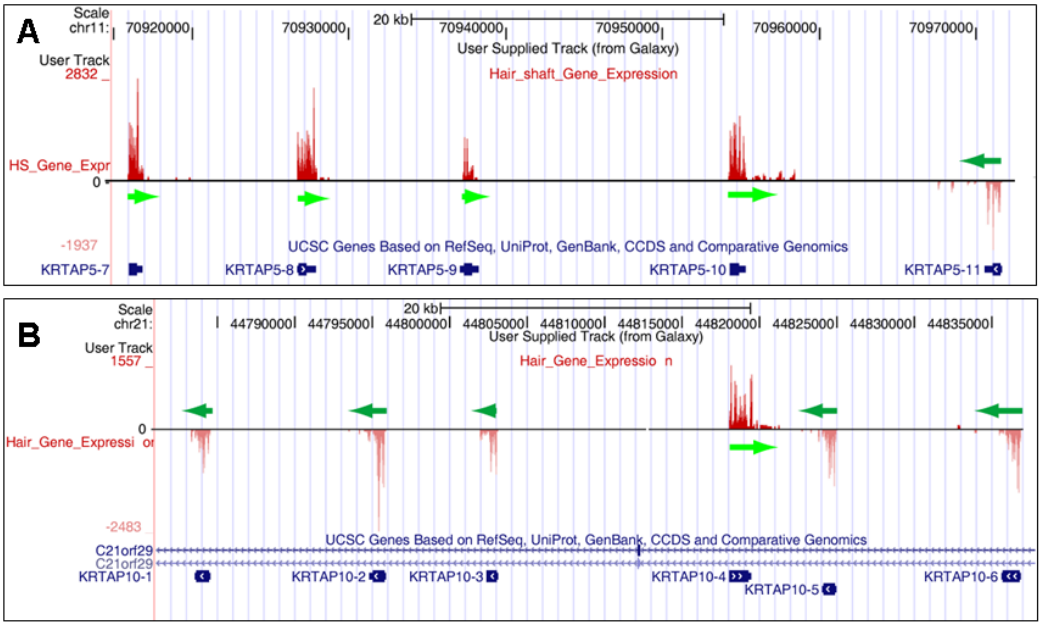

Figure 3.4: Orientation of hair library reads

Clusters of genes of the KRTAP5 family (A) and the KRTAP10 family (B) with reads that map to the (+) strand represented in red or on the minus (-) strand represented in pink. Note that reads are largely absent from intergenic regions and align in the direction of mRNA.

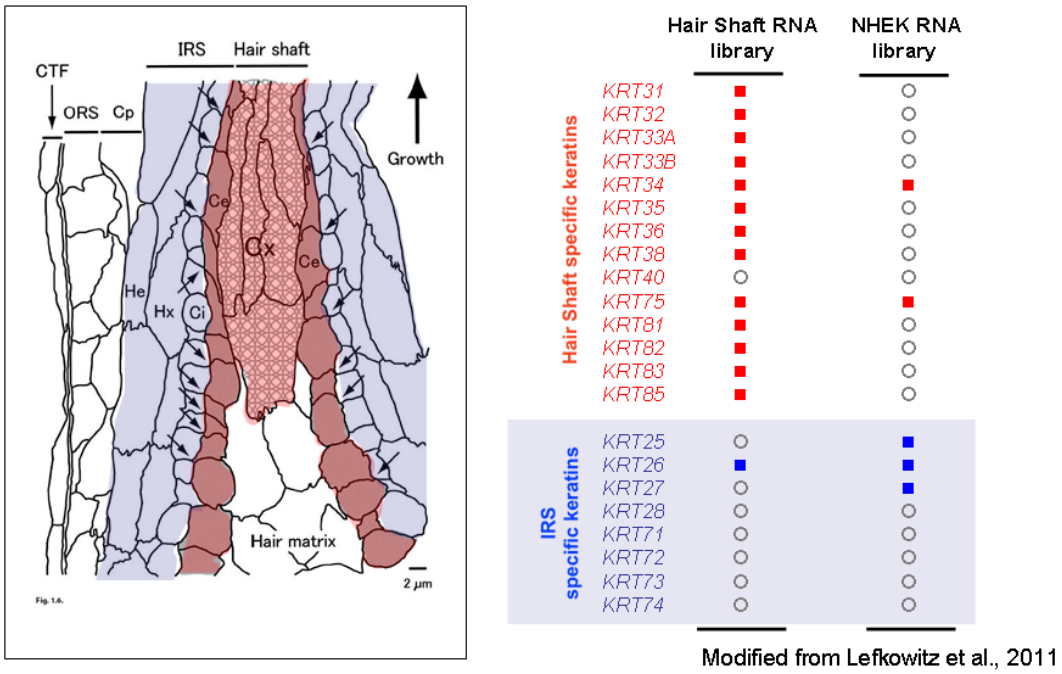

Figure 3.5: The cellular source of hair shaft RNA

The figure on the left is modified from Morioka, 2005 depicting cells in the developing hair follicle. The cells that are supporting structures of the hair shaft are shaded in blue. The cells that will divide, migrate upwards, and undergo terminal differentiation to become the mature hair shaft are shaded in red.

On the right are listed keratins that are demonstrated to have specific expression in hair shaft (red) or inner root sheath (blue). The presence of keratins in the hair library is indicated by a filled square, while absence is indicated by an empty circle. The presence or absence of these keratins was also examined in the library of a keratinocyte cell line.

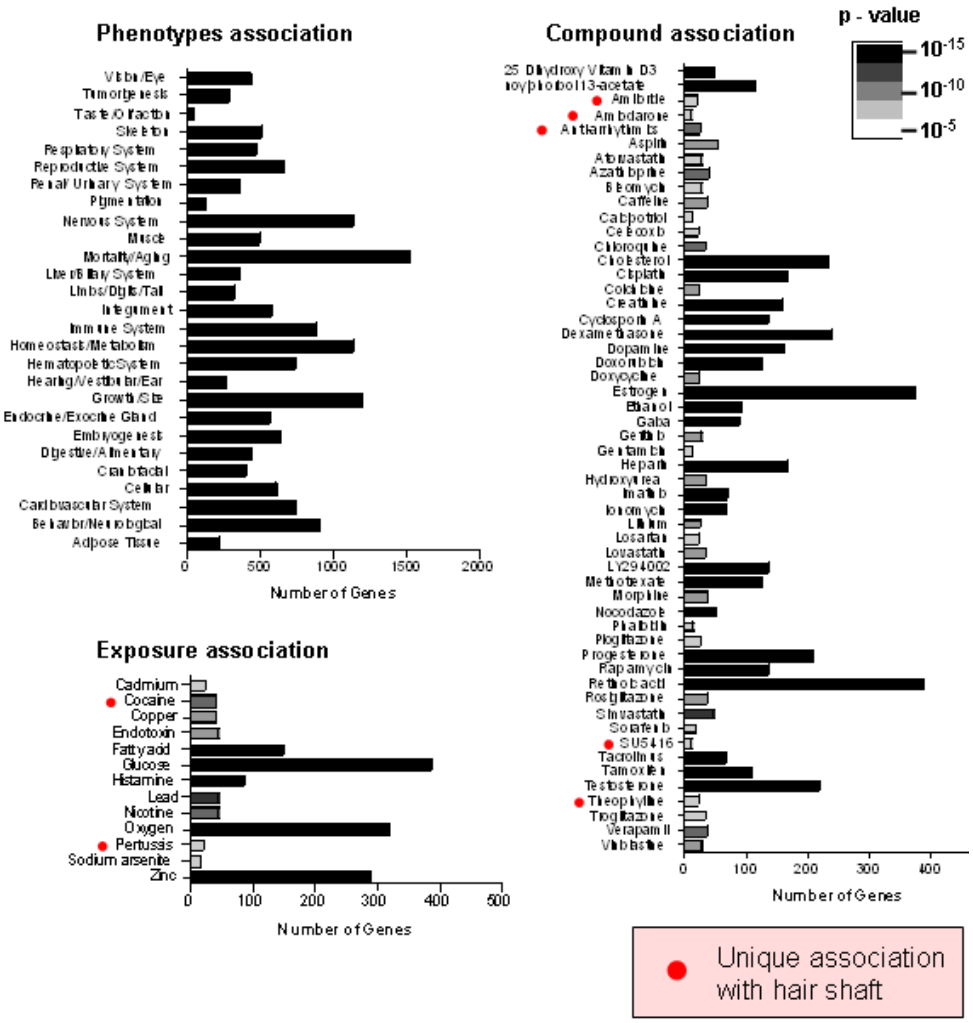

Figure 3.6: Overlapping and Unique Associations in Phenotypes, Exposure and Compounds

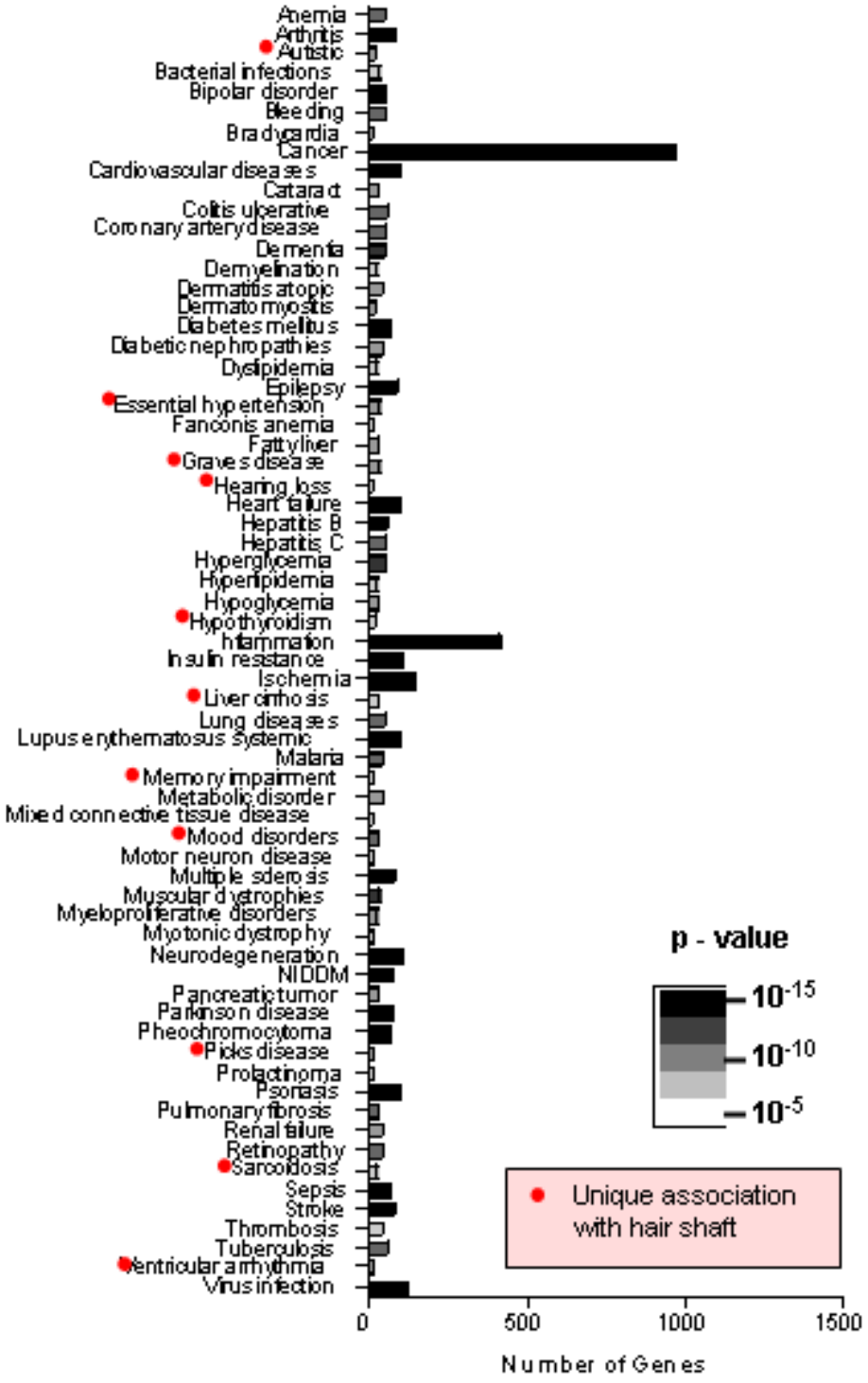

Figure 3.7: Overlapping and Unique Disorder Associations

Table 3.1: Demographic of individuals used in hair shaft small RNA library

Gender	Age (years)	Country of Origin	Hair Morphology
M	31	India	Wavy
M	40	Korea	Straight
F	21	USA	Straight
F	36	India	Wavy
F	30	China	Straight

Table 3.3: Comparison of RNA from modified hair protocol and Trizol
Three separate experiments performed and summarized below. Fresh six mm punches of mouse skin were used in each condition.

Lysis Condition	RNA Conc. (ng/ul)	Average (ng/ul) ± SD	A260	A280	260/280
Trizol	83.9	78.1 ± 8.7	2.097	0.969	2.17
	68.1		1.703	0.816	2.09
	82.3		2.058	0.948	2.17
Trizol + 100uM DTT	69.6	101.9 ± 36.4	1.740	0.810	2.15
	94.6		2.366	1.125	2.10
	141.4		3.534	1.726	2.05
Modified Protocol	156.9	137.5 ± 31.8	3.922	1.900	2.06
	154.8		3.870	1.931	2.00
	100.8		2.520	1.283	1.96

Table 3.4: Oligonucleotides

Abbreviations: Mm: *Mus musculus*; Hs: *Homo sapiens;*
O: Orientation, F: forward, R: reverse.

Species	Gene	O	Exon	Sequence
Mm	Krt34	F	6/7	GGAGAGCGAGGACTGCAACCTC
		R	7	ACGCTTTGAGCTGCCGCAAGG
Mm	Krt32	F	6	GGCCTGCTGGAGAGTGAGGACAG
		R	7	CACAGACGGTGCGGGATACCCC
Mm	Krt35	F	1	TGCTGTGGCCATGGCTTCCAAA
		R	1	GAGTGGATCCCCCTCCGGCTT
Mm	Bmp6	F	6	CTGCGCACCAACCAAACTGAA
		R	6	TTGGGGGAGGCGAACATTAGGTA
Mm	Sh3d19	F	1	GATGACGTATTGCCCACCTC
		R	2	GTTTGGCTGGGATTCGTATT
Mm	Sox2	F	1	TACTGGCAAGACCGTTTTCGTG
		R	1	CTCGGCAGCCTGATTCCAATA
Mm	ActB	F	1	TTTGCAGCTCCTTCGTTG
		R	2	CTTTGCACATGCCGGA
Hs	Snord60	F	1	GCACGTGCAGTTTTCATACG
		R	1	TGTGATGAATTGCTTTGACTTCT
Hs	Krtap5-8	F	1	AAGGGCCCAGGCTCAGGGAG
		R	1	GGAGCAGGTGAGAGGGAGGTGT
Hs	Krtap5-3	F	1	AGCCACAAGAACCGCAGCCC
		R	1	GCTGCTCCTCCAGCTGTGGC
Hs	Krtap5-3	F	1	TGAGGAGCAGCAGCAGGGCT
		R	1	GGGCTGCGGTTCTTGTGGCT
Hs	Krtap5-7	F	1	GGCTCTGGACTCAGGTCTCA
		R	1	GGACCCTGAGCAGTGGTTT
Hs	Krtap10-4	F	1	GCAAGACTGTCTGCTGCAAG
		R	1	AGCATGAAGAATCCCCACAG
Hs	Gapdh	F	6	TGCACCACCAACTGCTTAGC
		R	6/7	GGCATGGACTGTGGTCATGAG

Table 3.5: Genetic Association Database transcripts in hair

Gene Symbol	Disease Association	OMIM / PMID*
ABCA13	Schizophrenia, Bipolar, Depression susceptibility	607807
ATCAY	Cayman Ataxia	608179
BEAN	Spinocerebellar ataxia 31	612051
BHLHE41	Short Sleep Syndrome	606200
CNO	Hermansky-Pudlak Syndrome	203300
COL7A1	Dystrophic epidermolysis bullosa (Hallopeau–Siemens)	226600
CYP46A1	Alzheimer's disease risk	17192785
DNMT3B	Immunodeficiency-centromeric instability-facial anomalies	242860
DSP	Dilated cardiomyopathy with woolly hair and keratoderma	605676
	Arrhythmogenic right ventricular dysplasia/cardiomyopathy	607450
FLG	Ichthyosis vulgaris, atopic dermatitis	146700, 603165
KRT14	Epidermolysis Bullosa Simplex	131760, 601001
KRT16	Pachonychia congenita	167210
KRT5	Epidermolysos Bullosa Simplex	131760, 601001
KRT6A	Pachonychia congenita	167210
KRT6B	Pachonychia congenita	167210
KRT81	Monilethrix	158000
KRT83	Monilethrix	158000
KRT86	Monilethrix	158000
LAMB2	Congenital nephrotic syndrome (Pierson Syndrome)	609049
LEMD2	Carbamazepine adverse reactions	16538176*
MLL2	Kabuki Syndrome	147920
PIP5K1B	Chronic renal disease susceptibility	20383146*
PSORS1C2	Psoriasis susceptibility	177900
RDH8	Macular Degeneration, Age-Related 2	153800
RMRP	Catilage-Hair hypoplasia	250250
SORT1	Coronary artery disease	613589, 21378990
	Dyslipidemia	613589
SOS1	Noonan syndrome	163950
SURF1	Leigh syndrome (Infantile subacute necrotizing encephalopathy)	256000
TPM3	Nemaline myopathy	609284

Table 3.6: Genes associated with compound exposure

Gene Symbol	Novoseek Compounds
AHNAK	Calcium
BHLHE41	VEGF
BOK	Progesterone, Estradiol, Thymidine, Oxygen, Estrogen, Retinoic Acid, VEGF
CDK6	Indole-3-Carbinol, 3,3'-Diindolylmethane, Silicon Phthalocyanine, EB 1089, Fisetin, Silymarin, Roscovitine, 17-(Allylamino)-17-Demethoxygeldanamycin, Indole, Lovastatin, Thymidine, EGCG, Butyrate, Threonine, Rottlerin, Flavonol, Arsenite, Berberine, Monensin, Flavone, Gefitinib, Tyrosine, Apigenin, Trastuzumab, Resveratrol, Herbimycin A, 1,25 Dihydroxy Vitamin D3, Phorbol 12,13-Dibutyrate, TGF Beta1, Retinoic Acid, LY294002, Estrogen, Calcium, Oxygen, Proline, Ionomycin, Alanine, Progestin, Cisplatin, Tamoxifen, Serine, Dihydrotestosterone, Valine, Dexamethasone, Rapamycin
COL20A1	Hydroxylysine, Hydroxyproline, Gold, Tyrosine
COL7A1	Arginine, Valine, Serine, Cysteine, Chloramphenicol
CYP46A1	24S-hydroxy-cholesterol, Lathosterol, Cholesterol
DAPK2	Threonine, Serine
DNMT3B	5-aza-2'deoxycytidine, 5-Methylcytosine, Cytosine, Zebularine, Arsenite, Folate, Tamoxifen, Homocysteine, Oligonucleotide, Zinc, Estrogen
DSP	Asulam, N,N-Dimethylformamide, TGF Beta1, Ryanodine, Retinoic Acid, Calcium, Sucrose, Steroid, Progesterone, Tyrosine, 12-OTetradecanoylphorbol 13-Acetate, Serine, Acetone, Thapsigargin, DMSO
FLG	Citrulline, Retinoic Acid, Calcipotriol, Dithranol, Retinoid, Mecamylamine, Serine, Ceramide, Arginine, Calcium, GF 109203x, Allergens, Histidine, Lipid, Arsenate, Bromodeoxyuridine, Hematoxylin, Nitric Oxide, Nickel, Fibrinogen, Steroid, Latex, Tyrosine, Urea, Nicotine, Cysteine, Calcitriol, Infliximab, Salicylic Acid, Ascorbic Acid, SU5402, SU6668, Tyrosine, Heparan Sulfate, Heparin, VEGF, GNRH Phosphotyrosine, Apicidin, Sodium Chlorate, Imatinib, SU5416, Pd 98,059, Saporin, Suramin, Chlorate, Thymidine, SB 203580, Suberoylanilide Hydroxamic Acid, Dextran Sulfate, Histamine, Phenylalanine, Agar, Steroid, Calcium, LY294002, Glutamate, Lactate, Hydrogen, Lipid, Nitric Oxide, Glucose, Ribonucleic Acid, Lysine, Oligonucleotide, Nacl, Pyruvate, Atp, Forskolin, Fibrinogen, Chondroitin Sulfate, Progesterone, Threonine, Zinc, Glyceraldehyde 3-Phosphate,

Table 3.6: Genes associated with compound exposure (continued)

Gene Symbol	Novoseek Compounds
FLG (con't)	Phosphatidylinositol, Bromodeoxyuridine, 12-O-Tetradecanoylphorbol 13-Acetate, Methionine, Leucine, Estrogen, Adenylate, Genistein, Testosterone, Polysaccharide, Serine, Retinoic Acid, Cysteine, Dopamine, Paclitaxel, H2O2
HSP90AA1	Geldanamycin, 17-(Allylamino)-17-Demethoxygeldanamycin, Radicicol, 17-Amino-17-Demethoxygeldanamycin, 17-DMAG, LBH-589, Novobiocin, Molybdate, Steroid, Herbimycin A, Suberoylanilide Hydroxamic Acid, Atp, 5-Methoxy-1,2-Dimethyl-3-(4-Nitrophenoxymethyl)Indole-4,7-Dione, Dimethyl Pimelimidate, Rifabutin, Sodium Molybdate, Tyrosine, Amp-Pnp, Coumermycin, Lactacystin, Threonine, Progesterone, MG 132, Trastuzumab, Rapamycin, Depsipeptide, Tacrolimus, Sodium Arsenite, Quinone, Estrogen, Adp, Tcdd, Bortezomib, Imatinib, Serine, Arsenite, Phosphatidylinositol, VEGF Hydroxamate, Sulforhodamine B, 4-Hydroxytamoxifen, Methylmethanethiosulfonate, Nitric Oxide, Chlorothalonil, Nitrit, Pyrazole, LY294002, ATPgammas, Okadaic Acid, Hydroquinone, Cisplatin, Curcumin, Nonidet-P40, N-Acetylcysteine, Testosterone, Oxime, Coumarin, Mifepristone, Calpeptin, Leptomycin B, Cycloheximide, Quercetin, Citrate, Glyceraldehyde 3-Phosphate, Purine, Tungstate, Nocodazole, Polyacrylamide, Dexamethasone, Cysteine, Doxorubicin, Arachidonic Acid, Glutamate, Tamoxifen, Adenylate, Paclitaxel, Lactate, Glutamine, Morphine, Colcemid, Heparin, Alanine, Amphotericin B, Cytarabine, Oxaliplatin, Arginine, PGE2, N-Ethylmaleimide, Atorvastatin, Prostacyclin, Genistein, Anti-Fungal, Spironolactone, Butyrate, Corticosterone, 4-Hydroxynonenal, Neomycin, 5-Fluorouracil, Retinoic Acid, Glycerol, Etoposide, Dihydrotestosterone, Adenine, Carbamazepine, Camptothecin, Docetaxel, Sb 203580, Erlotinib, Thapsigargin, H2O2, Deae, Cyclosporin A, Vitamin D
KLF13	RANTES, Zinc, Proline, Alanine
LAMA5	Heparan Sulfate, Carbohydrates, Heparin, VEGF
LAMB2	Tripeptide, Acetylcholine, Heparin, Calcium
LATS2	Threonine, Serine
MYST4	Zinc
PABPC1	pABC, Adenylate
PIP5K1B	Phosphoinositide, Lipid
PNOC	Nalbzoh, Naloxone, Tertiapin, Etorphine, Damgo, Naltrindole,

Table 3.6: Genes associated with compound exposure (continued)

Gene Symbol	Novoseek Compounds
PNOC (con't)	[3h]Diprenorphine, Carbetapentane, Norbinaltorphimine, Octadecaneuropeptide, Morphine, Opiate, Npff, Dpdpe, Naltrexone, Rimcazole, Buprenorphine, Tetrapeptide, Anandamide, Aconitine, Capsaicin, Alpha-Aminoisobutyric Acid, Amphetamine, Piperidine, Trifluoroacetic Acid, Baclofen, Guanethidine, Cyclic Amp, Cocaine, Sufentanil, Gaba, Norepinephrine, Forskolin, Atropine, Hexamethonium, 5-Hydroxytryptamine, Dopamine, Nmda, Glutamate, Adenylate, Potassium, Lipid, Indole, Gtp, Sodium, Prostaglandin, Nitric Oxide, Steroid, Calcium, Progesterone, Tyrosine, Amide, Superoxide, Alanine, Estrogen, Pge2, Histamine, Acetylcholine
PPP1R15A	O6-Benzylguanine, Thapsigargin, Okadaic acid, Methylmethanesulfonate, Tunicamycin, Peroxynitrite, Cisplatin, Threonine, Carboplatin, Lactate, Methionine, Serine, VEGF
RAP1A	GTP, GDP, Guanosine, Forskolin, Cyclic Amp, Geranylgeranyl Pyrophosphate, Tyrosine, Mevalonate, Perillyl Alcohol, Guanine, LY294002, Calcium, Asparagine, Risedronate, Phorbol, Rolipram, Phosphatidylinositol, Farnesyl Diphosphate, Adenylate, Nadph, GF 109203X, Prostacyclin, IBMX, Threonine, FMLP, Serine, Nocodazole, Phosphoinositide, Ionomycin, Isoproterenol, CGMP, Lipid, PGE2, 12-OTetradecanoylphorbol, 13-Acetate, Oxygen, Agar, Nitric Oxide, Superoxide, Carbachol, Arginine, PGE1, VEGF, Glutamine, NMDA
RDH8	Vitamin A
RFK	Riboflavin, Flavin Mononucleotide, Flavin-Adenine Dinucleotide, Flavin, MGADP, ATP
RPL32	Glyceraldehyde 3-phosphate
S100A3	Calcium, Zinc
SCD	Stearoyl-Coa, Sterculic Acid, Fatty Acid, Palmitoleate, Linoleic Acid, Oleic Acid, Stearic Acid, Palmitoleic Acid, Sterol, Acetyl-Coa, Lipid, Palmitate, 3-Hydroxy-3-Methylglutaryl-Coa, Alpha-Linolenic Acid, Triacylglycerol, Docosahexaenoic Acid, Acyl-Coa, Cholesterol, Arachidonic Acid, Glycerol 3-Phosphate, Ceramides, Phospholipid, Oxo, Eicosapentaenoic Acid, Methane, Rosiglitazone, Troglitazone, Glucose, Cholesterol Ester, Farnesyl Diphosphate, IBMX, LY294002, Dexamethasone, ATP, Vitamin-E, Oxygen, P003, Iron, Phosphatidylinositol,

Table 3.6: Genes associated with compound exposure (continued)

Gene Symbol	Novoseek Compounds
SLC15A2	Glycylsarcosine, Tripeptide S, Valacyclovir, 5-Aminolevulinic Acid, Tripeptide, Ganciclovir, Nucleoside, Histidine
SORT1	Tyrosine, Glutamate
SOS1	Tyrosine, Menadione, Phosphatidylinositol, GTP, Serine, Lipid
SURF1	Lactate, Pyruvate, Bicarbonate, Oxygen
TBC1D4	Phosphatidylinositol, Glucose
TPM3	Tyrosine, Calcium
TRPV2	Capsaicin, Calcium
UBC	Finrozole, Porphobilinogen, Dexamethasone

Chapter 4: RNA preservation

4.1: Introduction

We have demonstrated that human hair shaft contains RNA that is reflective of the hair shaft lineage and non-hair genes that may be of medical relevance. This leads to the question of how the processes of terminal differentiation in the hair shaft might be favorable to the preservation of RNA, a molecule widely not considered to be stable. There are aspects of hair shaft cornification that markedly differ from the processes of living cultured cells.

4.2: Cornification

A specialized programmed cell death termed cornification occurs as hair follicle cells terminally differentiate into part of the hair shaft. Cornification is a highly organized and regenerative process that produces the cornified envelope in the skin and external hair. In the hair shaft, extensive crosslinking of structural proteins by transglutamases, ensures that the cell will become part of a growing biomaterial resistant to degradation. Hair shafts have been found well-preserved in mummified and frozen humans several centuries old (Rasmussen et al., 2010). Ultrastructural imaging captures the stages of cornification.

During this process, the nuclear envelope breaks down, DNA is cleaved, and cytoplasm is lost (Morioka 2005). In contrast to the rapid clearance of apoptotic cells,

the outermost cornified layer of skin is shed within two weeks and hair shafts of the scalp may be shed after three to six years of growth.

Apoptosis-like qualities

There are notable similarities and differences between apoptosis and cornification. Both are programmed death programs that utilize caspases and other enzymes and both are coordinated processes essential in the hair follicle. Apoptosis is a type of programmed cell death important for normal cell turnover, development and aging. During apoptosis, cells undergo distinct morphological changes including nuclear and chromatin condensation, DNA fragmentation, organelle shrinkage induced by a cascade of molecular signals (Vaux and Korsmeyer 1999) (He, Lu and Zhou, 2009). The apoptotic cell breaks down into membrane enclosed fragments that are engulfed by phagocytes, preventing an imflammatory immune response. Dysfunctions involving loss of apoptosis may lead to scarring, fibrosis, and cancer.

Apoptosis in the hair and skin occurs in the hair follicle and in the basal lamina of the skin. During development and the normal cycling of the hair follicle, high resolution electron micrscopy images reveal apoptotic cells in most compartments of the hair follicle (Mager et al., 2001). The result of this includes the creation of localized proliferative zones, formation the hair canal, separation of the hair shaft from the follicular wall as it exits, and creation of a protective barrier around the hair canal (Chang, Tsai, and Yu, 2005). Much of what is known about the cell death is based on classical apoptosis studies. However, many variations on this process exist.

Caspase activity in apoptosis and cornification

Caspases are cysteine proteases that cleave their targets at aspartic residues. The requirement for caspase activity during apoptosis was first mapped out in *Caenorhabditis elegans* genetic screen for mutants with egg laying defects. CED-3 shows similarity to human and murine interleukin-1β-converting enzyme and to the product of the mouse nedd-2 gene. The proteolytic cleavage of the pro-enzyme form of CED-3 activates its caspase activity (Yuan et al., 1993). There are only four caspases identified in *c. elegans* and only two are involved in apoptosis; there are more mammals that have other functions and some specialization. In mammals, the initiator caspases (2, 8, 9, and 10) have specificity for other caspases, while effector caspases (3, 6, and 7) have more generalized substrates which elicit cell death and characteristic morphological features (Lippens et al., 2009).

Caspase 14 is specialized for cornification in the skin and hair. It is expressed in the mammalian skin and hair—specifically in the inner root sheath which molds the hair shaft and so far, no homolog has been identified in birds and reptiles (Alibardi et al, 2005). It has been demonstrated that caspase-14 is involved in the formation of the stratum corneum, the soft flexible layer that protects against water loss and UV damage (Denecker et al., 2007). However, unlike apoptosis, where a caspase cascade plays a central role, other proteases, in particular transglutamases, are prominently involved in cornification (Candi et al., 2005).

4.3: DNA fragmentation occurs during cornification

During apoptosis, caspase activated DNase (CAD) is cleaved by caspase-3 from its inhibitor ICAD, enters the nucleus and cleaves chromosomal DNA (Enari et al., 1998). Other nucleases further break down DNA including DNase gamma, which has apoptosis specific activity (Shiokawa and Tanuma, 2001). The main DNase active the skin and hair during cornification is part of the same family. DNase1L2 is skin specific and degrades DNA in the stratum corneum and in hair. Loss of DNase1L2 results in retention of nuclear DNA in upper terminal layers of the skin and in the mature hair shaft. Decreased expression of DNase1L2 was observed in human lesion samples with known keratinocyte differentiation defects (Fischer et al., 2007). In the mouse model, no major phenotype was observed in mouse skin; in contrast, hair retaining DNA was more brittle and fragile (Fischer et al., 2011). While DNA degradation is a necessary step during cornification, the role or occurrence of RNA degradation during terminal cell death programs have not been well elucidated.

4.4 RNA degradation during cell death and life

During apoptosis, mRNAs are known to be rapidly degraded during the early stages, with half lives ranging from ½ to 3 hrs after induction. Cell culture studies suggest that other hallmarks of apoptosis (annexin V binding, poly-ADP-ribose polymerase processing, and DNA degradation) are on different pathway than RNA degradation (Del Prete et al., 2002).

The 28S subunit of ribosomal RNA is cleaved at specific sites resulting in distinct bands and its process is also decoupled from DNA degradation, and from caspase and p53 activities (King et al., 2000). However this is not necessarily indicative of apoptosis; this characteristic cleavage of 28S shows cells specificity (Samali et al., 1997).

RNase L and endonuclease G are ribonucleases known to be activated during apoptosis. Endonuclease G, a DNase/RNAse, is released from the mitrochondria following induction of the apoptotic program and has a preference for single stranded nucleic acids (Kalinowska et al 2005). RNase L is activated upon inhibition of mRNA translation by the translation inhibition; however, lack of RNAse L does not prevent apoptosis (Naito et al., 2009). The initial ribonucleases and other activating events required for RNA degradation during cell death are not established.

RNA turnover is a normal occurrence in living cells and plays an important regulatory role for gene translation. Deadenylation, or removal of the poly A tail of mRNAs, is a common initiation step in many RNA decay pathways followed by decapping of the 5' end and cleavage by endonucleases and exonucleases (Chen and Shyu 2011). Other RNA species also are detected by the cell's surveillance mechanisms and undergo regulated degradation. In both mammalian and yeast models, intronic sequences are spliced out by the spliceosome, a large protein complex made up of small nuclear RNAs and proteins. Intronic regions of pre-mRNA are joined into a lariat during splicing, linearized then degraded by exonucleases (Moore 2002). The half lives of introns can range to almost half an hour in mammalian cells

(Clement, Maiti, and Wilkinson, 2001). Incorrect transcripts, with premature stop codons are bound by proteins that signal to degrade and prevent the production of aberrant proteins (Shyu, Wilkinson, and van Hoof 2008). The complementary star sequence of miRNAs, once discarded from the Argonaut complex, is rapidly degraded (Kai and Pasquinelli 2010). This general phenomenon of rapid and efficient detection and destruction of unwanted RNA may be necessary to not disrupt the translation of other genes, as miRNA and viral RNAs are known to do (Houseley and Tollervey 2009).

4.5: RNA coverage of viable cells and external hair

The multiple mechanisms that exist to degrade RNA in living cells suggests that the existence of unwanted RNA would be short-lived. We would thus expect that the coverage of RNA Seq in a living cell to be largely restricted to functional RNA sequences rather than random discarded fragments. Knowing that the degradation process of mRNAs tends to be fairly rapid, we would also expect to see fairly unbiased coverage of exonic sequences.

On the other hand, RNA left out to break down free of a living cell's natural enzymatic processes may show a different pattern of gene coverage. Once caveat of is that we do not have at our disposal a large population of viable precursor hair shaft cells for comparison. A comparison with expression in cultured keratinocytes, NHEK, still suggests that coverage of gene regions is similar, although individual mRNAs vary. We find that RNA coverage of the 5' UTR, (coding) exonic regions, intronic

regions, and 3’ UTR is relatively similar between that of hair shaft and of viable keratinocyte cells (Figure 4.1). This suggests that RNA in the hair shaft persists after cornification and it is not passively worn down.

It should be noted that these gene regions, with exception of the defined 2kbp intragenic regions beyond the 5’ and 3’ UTRs represent areas of different lengths (Table 4.1). Of the defined regions, 5’UTRs are on average 439 bases long, while introns are 75,589 bases long. When the length of average segments is taken into account, by calculating the number of reads per base in each segment, 5’ UTRs and exons have the highest read coverage overall, followed by 3’ UTRs, then introns (Table 4.1, Figure 4.1B). The region type that varies the most in length are introns which ranges from fifteen to one million bps in length (Table 4.1). The lengths of all segment types overlap to some extent although intronic segments tend to be the longest (Figure 4.2). Shorter segments tend to have more reads per base (Table 4.2).

4.6: Genes of interest in the hair shaft library

We had unexpectedly found that keratinocyte associated protein (KRTAP) genes made up over half of the total mRNA reads (Figure 3.3). This mammalian specific gene family can be divided into a group of ultra high/high cysteine and high glycine/tyrosine and further sorted into 27 subfamilies based on phylogeny (Rogers and Schweizer, 2005) (Wu, Irwin, and Zhang, 2009).

The KRTAP genes present in the human hair shaft library are summarized in Table 4.3. These represent 52 genes from ten subfamilies that are spread among three

chromosomes. KRTAP19 belongs to the high glycine/tyrosine subfamily; the other KRTAP genes present are part of the ultra high/high cysteine families. These genes span around one thousand base pairs in length. With exception of KRTAP10-1 and KRTAP 10-10, KRTAP genes present in the library have homologous genes in the *mus musculus* and *pan troglodytes* genomes.

Size polymorphisms reported in KRTAP genes have been characterized in the repetitive coding regions, but the phenotypic result of these variations is unclear (Shimomura et al, 2002) (Kariya, Shimomura, and Ito, 2005). In contrast, of the most abundant KRTAP genes, there are 3'UTR isoforms that are longer than the RefSeq annotations. This extension beyond annotated borders suggests that one factor promoting the stability of RNAs in hair is a longer 3'UTR. The cDNA and EST libraries from which REFSEQ is based are not of the same source as hair shafts and as such, the KRTAP isoforms in our library may represent hair-specific isoforms.

Alternative poly A signals are estimated to account for over half of all mRNA isoforms; this feature affects the stability of the transcript and may alter gene expression kinetics (Proudfoot 2011). In the hair shaft, it is plausible that the alternative poly-A site adds another layer of tissue specific regulation. Alternate 3' UTRs may have additional binding motifs that proteins, other RNAs may interact with. This may ultimately result in increased longevity and hence abundance of these transcripts in the hair library.

Of the microRNAs, it has not escaped our notice that members of the let-7 family are the most prominent. The proteins of the LIN28 family directly suppress

LET-7 by interfering with LET-7 processing (Newman and Hammond, 2010). As the levels of LET-7 miRNA and LIN-28 are inversely related in other tissues, such as ES cells undergoing differentiation, LIN28 RNA has not been detected in the hair shaft library. Not surprisingly, other genes involved in suppressing differentiation and maintaining a stem cell state (MYC, OCT4, SOX2, NANOG) are also absent.

Also of note is the presence of mir-203 in the hair shaft library. It has previously been reported to be essential for the differentiation during skin morphogenesis in mouse and human skin (Yi et al., 2008). As the hair is a terminally differentiated tissue, we can imagine that its function to repress 'stemness' may occur here. However, the two proteins shown experimentally to interact with mir-203 during this process, p63 and SOCS-3, are absent in the hair library. Mir-203 may have other partners and other roles during hair follicle maturation.

Other prominent RNAs in our family may have as yet undiscovered regulatory roles—for example, metastasis associated lung adenocarcinoma transcript 1 (MALAT-1) is a long non-coding RNA located on chromosome 11 (Figure3.3B and Figure 4.3A). It is known to be associated with metastasis in lung cancer and in vitro studies demonstrate impaired motility and downregulation of the expression of motility genes upon knockdown of MALAT-1 (Tano et al., 2010). In a large scale analysis, it is found to have mouse homolog, five isoforms, histone modifications, and expression in many tissues, the most abundant being brain, as shown in Figure 4.3B (Cabili et al., 2011). It also may function as an enhancer (Ernst et al, 2011). Of the neighboring genes, another non-coding RNA, nuclear enriched abundant transcript 1 (NEAT1),

alternatively known as LINC00084, is also expressed in the hair library (Figure 4.3C). The proximity of these two non-coding RNAs may suggest that they coordinately act as enhancers during terminal differentiation of hair shaft cells. Another abundant sequence, C17orf44, likely represents another long non-coding RNA, LINC00324 on chromosome 17p (Figure 3.3B). The presence and abundance of MALAT-1 and LINC00324 in our library of non-cancerous tissue suggests that there is a role for non-coding RNA in translational control or post-translational regulation during terminal differentiation.

Other genes and their signaling pathways remain to be explored. Plucked hair follicles have been used for RNA analysis. Genes that are differentially expressed between disease states, such as psoriasis have been examined (Aubert et al, 2010). The hair follicle is a more heterogeneous tissue than hair shaft, such that RNA from the surrounding epidermis and the layers could contribute to the RNA extracted in hair follicles. Many of these genes comprise signaling networks where localization, duration, and timing of expression are relevant. Such is the case with sonic hedgehog signaling, where a cell could respond to the ligand secreted by other cells which diffuses out in a gradient fashion, while also responding to the ligand it produces itself (Harfe et al., 2004). Subtractive comparison of gene expression between the plucked hair follicle and hair shaft may provide localization data related to gene expression variations.

In summary, we demonstrate that the RNA retained in hair after cornification is representative of coding genes. The gene regions present are comparable to those of

a viable cell, suggesting that the reads in the hair library are not intermediate fragments leftover from an active degradation process. This library provides a framework for further exploration of the process of terminal differentiation and discovery of non-invasive biomarkers. Keratin associated proteins are highly abundant in the hair library, suggesting a larger role than previously considered during cornification. Non coding RNAs such as members of the LET-7 family and lincRNAs were also highly abundant and warrant future study. The end product of cornification is a hardened stable proteinaceous matrix. This process may have a heavier dependence on the concerted interactions of multiple types of RNA than previously appreciated. The localization and strength of signaling differences may be reasoned from comparisons of RNA in hair shaft versus hair follicle.

4.7 Methods

For NHEK to hair shaft RNA library comparisons, representative coverages of intragenic regions in REFSEQ were compared. Each gene in REFSEQ was segmented into annotated 5′-UTR, introns, exons, and 3′-UTR regions; these regions were further broken up into windows of 5% of the length of the region. Intergenic regions were defined as 2kbp beyond annotated 5’ and 3’ ends. The hair shaft library was aligned to these regions, and the total coverage of all 5% windows for each region type in each gene was plotted, generating a summary of overall coverage patterns in intragenic regions.

Portions of Chapter 4 are a reprint of the material as it appears in PLoS One 2011. Lefkowitz, GK, Mukupadhyay A, Cowing-Zitron, CA, Yu, BD. Bioinformatic analysis was performed by BDY and CAC. The dissertation/thesis author was the primary investigator and author of this paper.

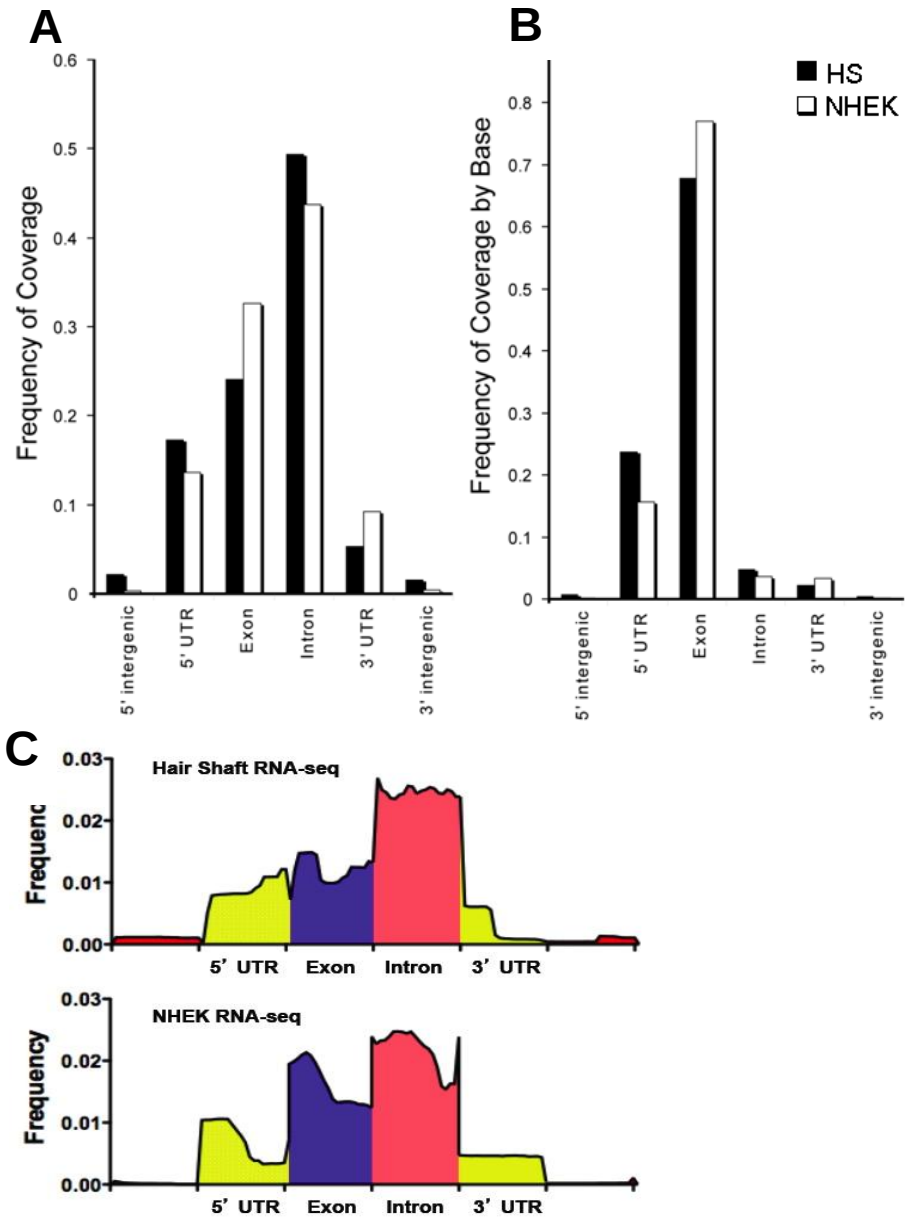

Figure 4.1: Comparison of RNA read coverage between hair shaft and NHEKs
Reads were classified as by gene region and the relative frequency of coverage of gene region is shown in (A) and the relative frequency per base is shown in (B) for the hair and NHEK library. The alignments from (A) were normalized, then divided into twenty windows per gene region and mapped in (C).

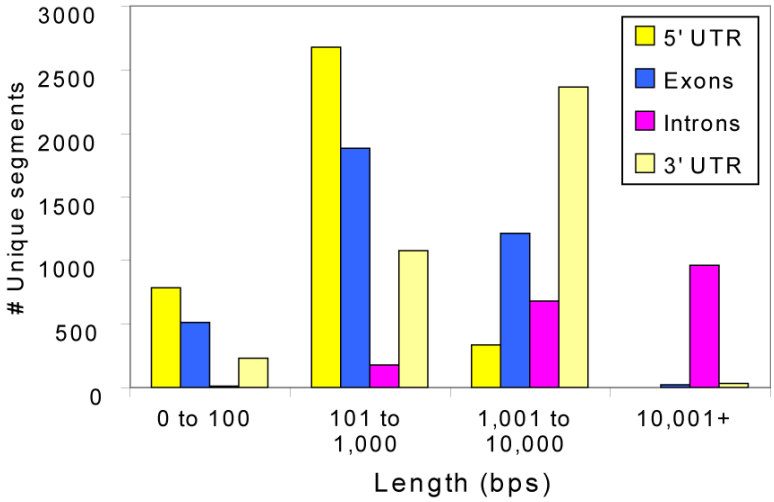

Figure 4.2: Gene regions represented in the hair library
The regions that hair library reads represent are classified by length. This shows that the introns presented in the library, on average, tend to be much longer than translated regions (5'UTR, 3' UTR, and exons).

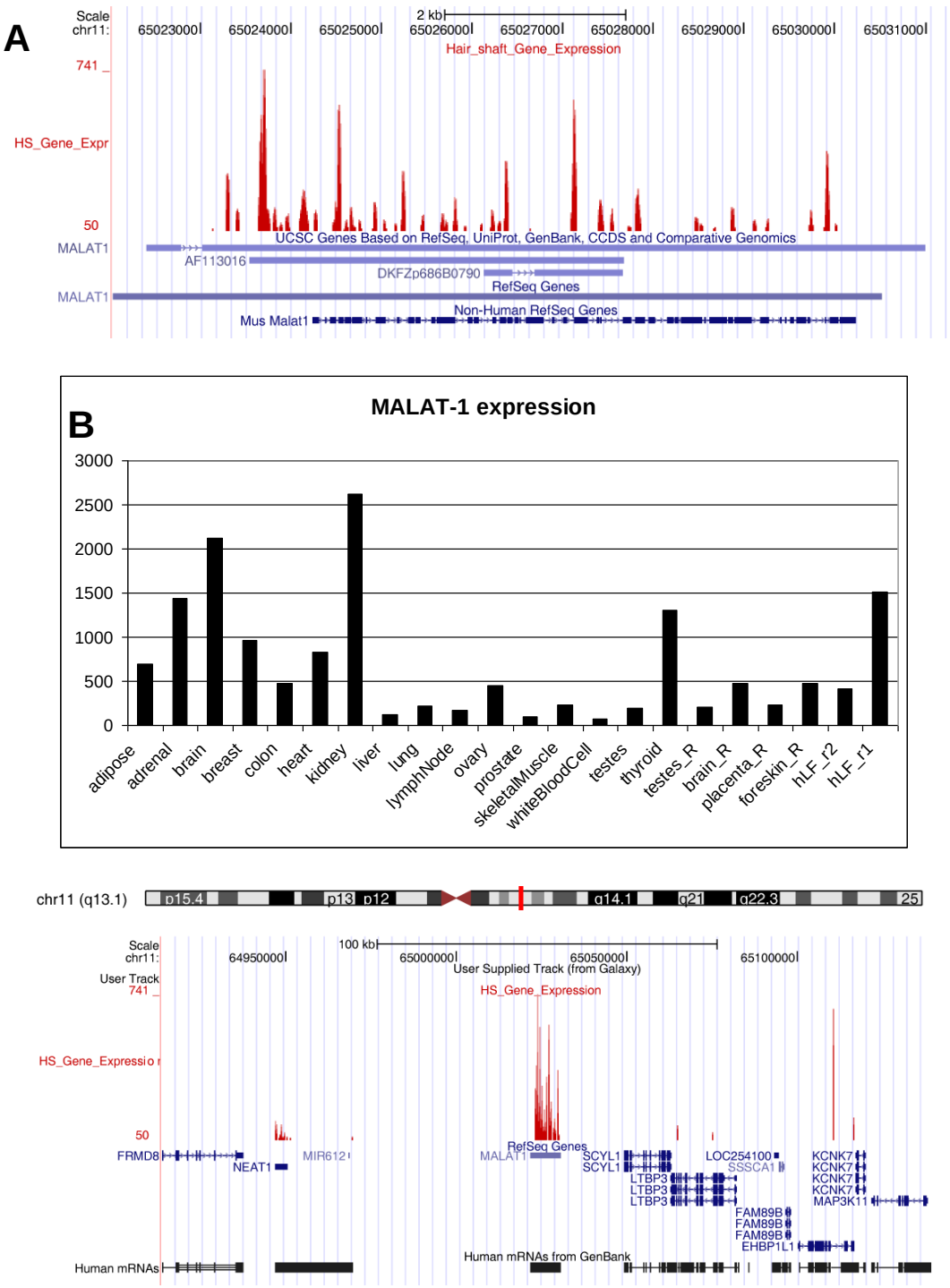

Figure 4.3: MALAT-1, a long non-coding RNA

A) Schematic of reads that map to MALAT-1, a largely uncharacterized lncRNA, in the hair library. B) Relative expression of MALAT-1 in various tissues based on RNA-Seq data from 24 tissues (from Cabili et al., 2011). C) Location of MALAT-1 and neighboring genes on chromosome 11.

Table 4.1: Read coverage across gene regions

This lists the regions as defined by UCSC refFlat descriptions used in Figures 4.1 and 4.2. Each gene region type was then divided into twenty windows and displayed in Figure 4.2. The range of length variation of gene regions present in the library and the resulting average window are shown.

Gene Region	Min (bps)	Max (bps)	Average length (bps)	Average window (bps)	Reads per base
Exon	27	59,461	1,307	65	16.8
Intron	15	1,043,911	75,589	3779	0.3
3’ UTR	3	59,461	2,261	113	7.0
5’ UTR	1	37,027	4,39	22	9.7

Table 4.2: Reads per region based on segment length
Gene regions were further classified by length and read abundance compared.

Segment type	Range (bps)	Average length (bps)	Reads per bp	Unique segments
3' UTR	0 to 100	75.22	151.05	225
	101 to 1,000	494.93	0.79	1077
	1,001 to 10,000	2978.41	0.016	2360
	10,001+	15000.26	0.0012	27
Exons	0 to 100	79.83	153.95	510
	101 to 1,000	331.30	2.36	1879
	1,001 to 10,000	3020.89	0.45	1213
	10,001+	17580.60	0.03	20
Intron	0 to 100	73.88	3.98	8
	101 to 1,000	492.27	1.95	180
	1,001 to 10,000	4586.39	1.14	681
	10,001+	206660.52	0.02	958
5' UTR	0 to 100	67.61	56.70	783
	101 to 1,000	305.32	1.25	2677
	1,001 to 10,000	2464.14	1.06	332
	10,001+	16492.60	0.0059	5

Table 4.3: Keratin associated proteins present in the human hair shaft library
Boldface indicates reads that extend past REFSEQ 3' UTR annotations.
Abbreviations: FPKM: fragments per kilobase mapped; Chr: chromosome

Gene	FPKM	Chr	Gene	FPKM	Chr
KRTAP5-4	21063	11	KRTAP12-2	2022	21
KRTAP5-8	18948	11	**KRTAP19-1**	1871	21
KRTAP5-10	18718	11	KRTAP9-8	961	17
KRTAP5-7	17881	11	KRTAP19-3	937	21
KRTAP5-5	17709	11	KRTAP9-3	911	17
KRTAP5-3	15045	11	KRTAP9-4	889	17
KRTAP5-2	14959	11	KRTAP9-9	648	17
KRTAP5-1	13637	11	KRTAP4-12	533	17
KRTAP10-11	13007	21	KRTAP24-1	508	21
KRTAP10-4	11224	21	KRTAP4-8	495	17
KRTAP10-9	10635	21	KRTAP4-11	458	17
KRTAP5-11	10398	11	KRTAP4-9	456	17
KRTAP10-2	10129	21	KRTAP3-3	416	17
KRTAP10-6	9075	21	KRTAP3-3	416	17
KRTAP10-5	8352	21	KRTAP4-7	374	17
KRTAP10-1	7971	21	KRTAP4-4	348	17
KRTAP10-7	7863	21	KRTAP4-3	333	17
KRTAP10-12	6925	21	KRTAP1-5	309	17
KRTAP5-6	6163	11	KRTAP1-5	299	17
KRTAP10-3	6041	21	KRTAP19-5	277	21
KRTAP5-9	5552	11	KRTAP3-2	249	17
KRTAP17-1	5151	17	KRTAP3-2	249	17
KRTAP10-10	3848	21	KRTAP1-1	243	17
KRTAP10-8	3433	21	KRTAP4-2	238	17
KRTAP10-7	2561	21	KRTAP4-5	225	17
KRTAP12-1	2469	21	KRTAP26-1	205	21

Chapter 5: RNA preservation in keratinized tissue

5.1: Introduction

Hair is a stable biomaterial which has been well preserved in ancient samples. Various structural characteristics of a 4,000 year old human hair were still discernible, such as the cuticle scales, pigmentation, and evidence of parasites boring into the cortex (Rasmussen et al., 2010). We previously provide evidence that the RNA of hair is also preserved. This preservation may be due to the high degree of crosslinking, lack of active RNases, or other unknown properties of the cornified hair shaft.

5.2: Scalp hair

Unlike the hair found on most other mammals and on other regions of the human body, scalp hair has the unusual property of a long growth phase, lasting from three to six years and up to ten years in one documented case (Robbins, 1994). One evolutionary hypothesis is that long scalp hair might function as a trait undergoing sex selection. Mesko and Bereczkei found that European men generally found images of women more attractive when they had long or medium length hair (2004). This feature may also function as a social sign of status and well being, as long hair is costly to grow and maintain (Synott, 1987). Much like colorful plumage on male birds, long scalp hair may indicate an absence of parasites, good genes and good health (Hamilton and Zuk 1982). The gain of hair shaft abundance on the head may be a result of an adaptive survival benefit, such as protection from the sun as bipedalism developed or

as trait undergoing sexual selection (Pagel and Bodmer 2003). Whatever the selective pressure, the long growth phase of human scalp hair makes it uniquely suitable for retrospective analysis.

To test the extent of hair RNA preserved in these aged samples, we compared sequential segments of the same hair shafts. Scalp hair was taken from two individuals and aligned from bulb to tip. The bulb segment, measuring from the tip to 0.5 cm out, was removed to avoid contamination with other tissue. Sequential 2.5 cm segments, each representing approximately two months of growth were taken.

First we looked at yields of RNA from each segment of hair. We expect that degradation of RNA post-cornification or after exiting the scalp will, combined with exposure to environmental factors will result in gradually lower RNA yields. As shown in Figure 5.1A, we find that RNA can be extracted in later segments representing up to a year of aging.

We next compared detection by qRT-PCR of specific genes present in the hair RNA library. At increasing distances, higher cycle numbers were required to detect both miRNA and mRNA target genes via qRT-PCR, indicating decreased abundance of specific RNA transcripts. We verified the specificity of RNA products by melting curve analysis (Figure 5.1C) and gel electrophoresis. From this we can estimate a rough timeline of the decay of RNA. We find a two-fold reduction of mRNA and miRNAs occurs over the length of 0.92±0.11 cm and 0.81±0.16 cm respectively. This suggests that RNA remains stable during hair formation and persists even months after cornification.

5.3: Testing RNA retention of other cornified tissues

Other tissues undergo similar cornification steps. The claw or nail is another ectodermal appendage that is also highly keratinzed. Unlike hair cycling which has a quiescent or resting phase, nails continually grow from the nail matrix. Claws and nails function in climbing, grasping, and walking. Among primates, humans have a single layer of nail on all their digits while others primates have some combination of nails, toilet claws or claws that have one or two distinct layers (Soligo and Müller 1999).

The nail grows roughly one cm per month. There are some structures that are homologous to those of the hair follicle. The eponychium, also known as the cuticle, is a thick layer of skin surrounding the nail. Underneath, is the matix, from which cells that make up the keratinized nail grow. The matrix is sometimes visible as the lunula, or moon, in the nail plate. The nail bed is the supporting tissue underneath the growing nail plate. Like hair, nails have been monitored for health assessment and are visibly altered in certain disease states, notably skin conditions such as psoriasis and eczema and more general conditions such as autoimmune disorders (Barth, Telfer, and Dawber 1988).

As discussed in Chapter 3 and illustrated by Figure 3.4, the hair library contains many of the keratins found *in situ* in developing hair follicles which are also detectable by proteomic analysis. Immunohistochemistry and *in situ* analysis has shown that many of the hair keratins are also present in the nail (Perrin 2007). Shotgun

proteomic analysis of hair shaft and nail proteins have identified many common proteins. Keratins and keratin associated proteins are the most abundant of proteins identified and are present in both structures; other prominent shared proteins have membrane and junctional functions (Rice et al., 2010) (Lee, Rice, and Lee 2006). Table 5.1 compares published proteomic proteins in hair and nail from two studies and in our hair library. Of the 71 proteins identified in hair, there are 17 proteins also present in nail protein extracts. There were also 29 proteins identified in nail that are absent in hair. However, when compared to our hair library, we also do not find a large overlap. This highlights the possibility that mass spectrophotometry analysis may have different technical limitations than RNA analysis. It is also possible that RNA-Seq analysis of cornified tissue may be biased against detection of mRNA of structural proteins that are translated early on in cornification may not be well represented in mRNA but the protein would be detected.

We find that nail RNA can also be extracted by methods used for hair. The yield of RNA per ug of nail samples from five healthy anonymous individuals is summarized in Table 5.2A. The quality of RNA extracted is comparable to RNA extracted from other tissues using the method described in Chapter 3 as shown by the A260/280 ratio in Table 5.2. The absorbance reading in Figure 5.1A has a single peak at 260 nm which shows that the RNA from nail is free from contaminants such as phenol, Tris, or alcohol. The range of RNA extracted per ug of nail varies between samples. This may be due to variations in sample preparation or other individual

factors such as average nail growth, age or gender of donor, and exposure which were may be explored in future studies.

Nail RNA is also useable for analysis by qRT-PCR. We assayed for GAPDH and KRT34, a common keratin gene in three separate nail samples and as with previous experiments described in Chapter 3, compared this to the keratinocyte cell line, NHEK. Nail cycle numbers for GAPDH and KRT34 were 26.2±0.26 and 23.3±0.10 respectively and NHEK cycle numbers for GAPDH and KRT34 were 30.1 ± 1.36 and 27.7 ± 0.40 as displayed in Figure 5.2B. Melting curve analysis shows that the product is specific in both samples (Figure 5.2C). This suggests that retention of RNA is a common property of keratinized tissue. This raises the question of whether RNA might also be retained in other tissues or if unique shared properties of hair and nail make RNA retention possible.

One of the advantages of using hair or nail as a source of screening is that the RNA is still present and storage of hair for future extraction has mainly been at room temperature. As mentioned in Chapter 3, murine hair shafts cells undergo a week of differentiation before exiting the skin (LeBeau, Montgomery, and Brewer, 2011). We wanted to contrast this RNA retention with biological tissue commonly used for expression analysis, the liver and skin. Conventional practices to preserve the integrity of RNA from tissue often begin with rapid preservation of the tissue. This can be accomplished by flash freezing with liquid nitrogen or immersing the tissue in a larger volume of a preservative, such as RNALater, upon dissection. According to studies on RNA integrity, tissues immersed in RNALater will be stable for up to a week stored at

4 degrees C, and longer if frozen (Kasahara et al, 2006). This would suggest that RNA would be lost in most tissues without some method of preservation.

To test this, a simple experiment was carried out. Lobes of liver and sections of skin were removed from animals and divided into sets of tubes. The first set was immediately placed in RNAlater (Qiagen), while the two other sets were left at room temperature. RNA was extracted as described in Chapter 3. The yield of RNA per ug of starting material are shown in Figure 5.3. Surprisingly, there is RNA retained in biological tissues kept at ambient temperature for a week. However, in samples kept at a longer period, two weeks, there is much less RNA. This suggests that the biological process of RNA degradation is not rapid in dead tissues, however RNA does degrade more so in these tissues than in hair. The variation seen in RNA yield from immediately preserved samples may come from differences in tissue intactness or the extent of vascularization or water content of the tissues. It can be imagined that an intact piece of tissue would have protected RNA in the center of its mass and that blood might contain endonucleases and proteases that were not accounted for. The quality of this RNA for downstream applications remains to be tested. The main finding suggests that the RNA breakdown mechanism of dying tissue is not rapid, yet still occurs in tissue other than hair or nail.

The fact that RNA remains in hair shaft and nail raises questions about the biology of terminal differentiation and cornification. RNA degradation in living cells is quite efficient and tightly regulated by enzymes and other RNA binding proteins that are present in the hair library. Other cell death programs are reported to oxidize

and degrade RNA fairly quickly (Shan, Chang, and Lin 2007) (Del Prete, et al., 2002). In other specialized cells such as red blood cells, mRNA exists after eunucleation for continued protein synthesis and cell division (Schwertz et al., 2010).

To test for transcriptional activity, hair shafts and NHEK cells were treated with an irreversible transcription inhibitor, actinomycin D, at a dose of 5ug/ml for two hours in a 37^{o} incubator. We then assayed the RNA for GAPDH and KRT5. While there was an observable decrease in mRNA abundance of NHEK cells used for comparison, there was no observable difference between hair samples treated with the drug or vehicle (Figure 5.4). This preliminary experiment suggests that there is no obvious transcriptional activity. However, there are other factors to consider. Transcriptional activity may be present, but at such a low rate relative to NHEK cells to render it undetectable. The drug may not penetrate into the highly crosslinked fibers of the hair and may need a carrier. Other measures of transcriptional activity, such as incubation with radiolabeled nucleotides, may be a possible future step. While this does not rule out transcriptional or translational activity, it does suggest more sensitive methods should be used to study the role of RNA retention in keratinized tissue.

In summary, we find that RNA is preserved in keratinized tissue and not subject to immediate degradation following cornification. In hair, RNA is detectable even out to extended lengths representing six months of growth and the half lives of these mRNAs may be calculated. Another keratinized tissue, the nail or claw, also

retain RNAs and may be amendable to similar study. These findings raise the possibility of retrospective screening using these tissues.

Portions of Chapter 5 are a reprint of the material as it appears in PLoS One 2011. Lefkowitz, GK, Mukupadhyay A, Cowing-Zitron, CA, Yu, BD. Bioinformatic analysis was performed by CAC. The dissertation/thesis author was the primary investigator and author of this paper.

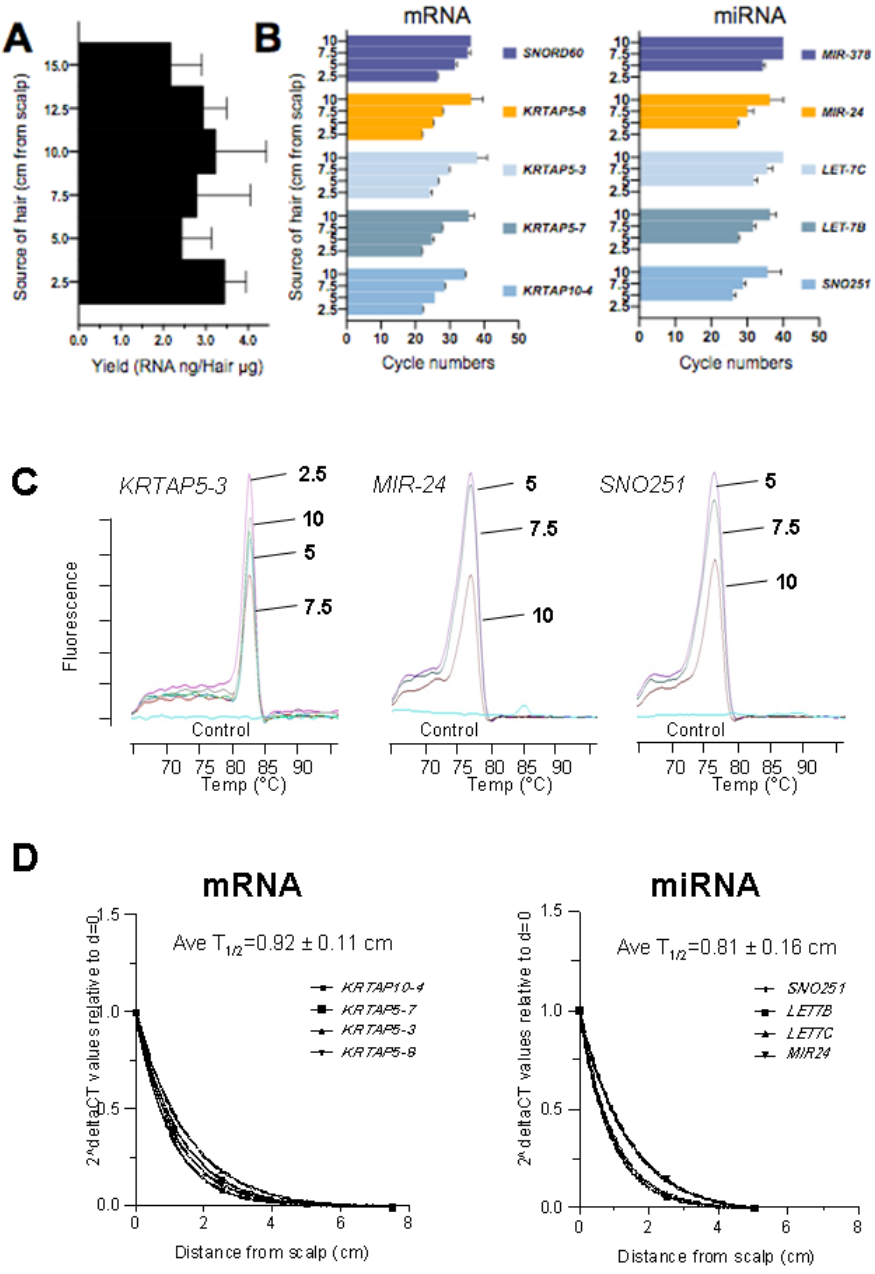

Figure 5.1: Sequential segments of hair contain RNA

A) Average yield of RNA in sequential segments of hair. B) Cycle number of selected genes in sequential segments of hair increase, indicating less amplifiable targets in older segments. C) Melting peaks of segments compared with controls indicate a specific product being made. D) The half life of transcripts is estimated by fitting cycle numbers onto an exponential curve.

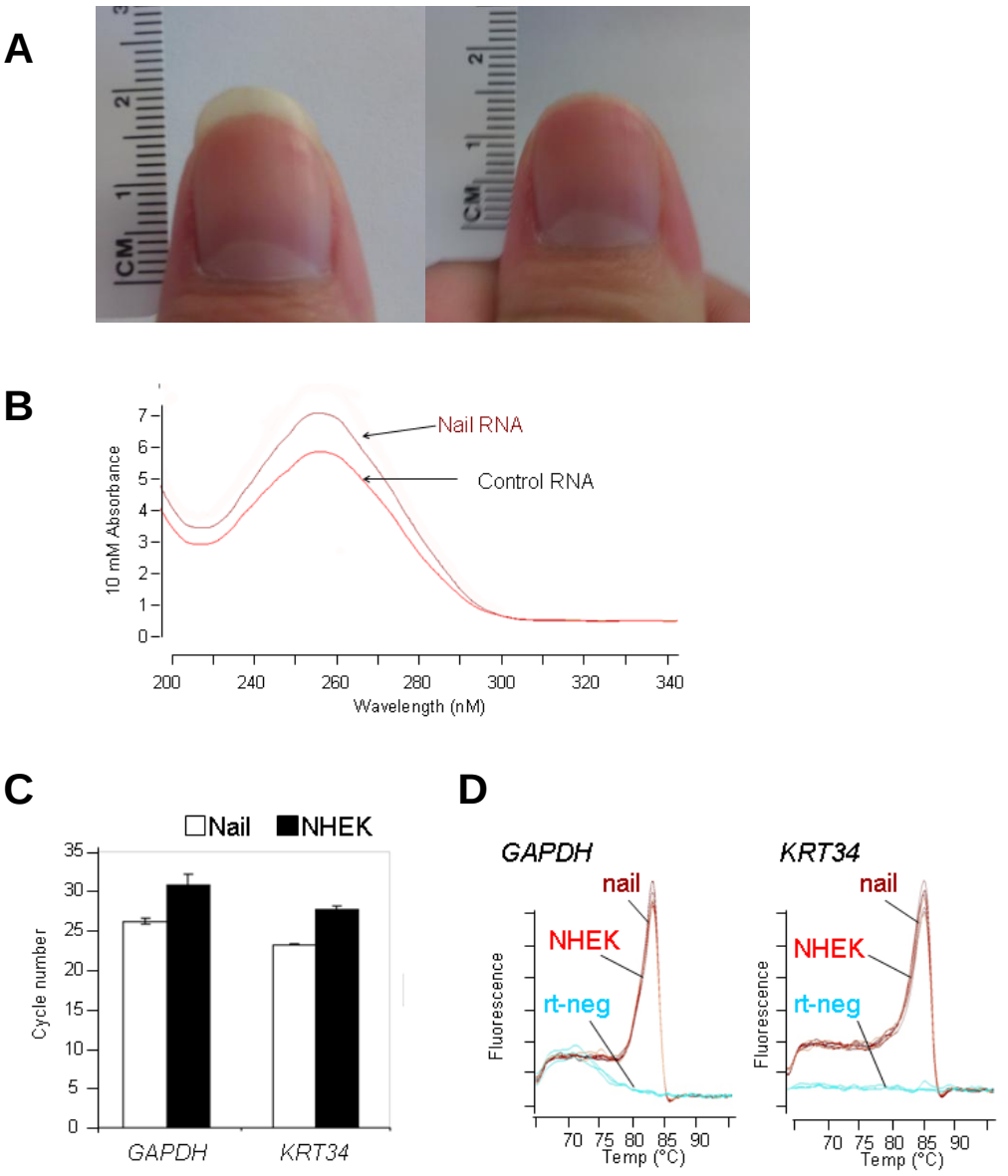

Figure 5.2: Nail expression analysis

A) The average thumbnail before(left) and after (right) harvesting for RNA purification. B) Absorbance reading of nail RNA compared to control RNA from NHEK. C) Cycles at which GAPDH and KRT34 products amplify are comparable between nail and NHEK samples. D) Melting curve analysis of qPCR products.

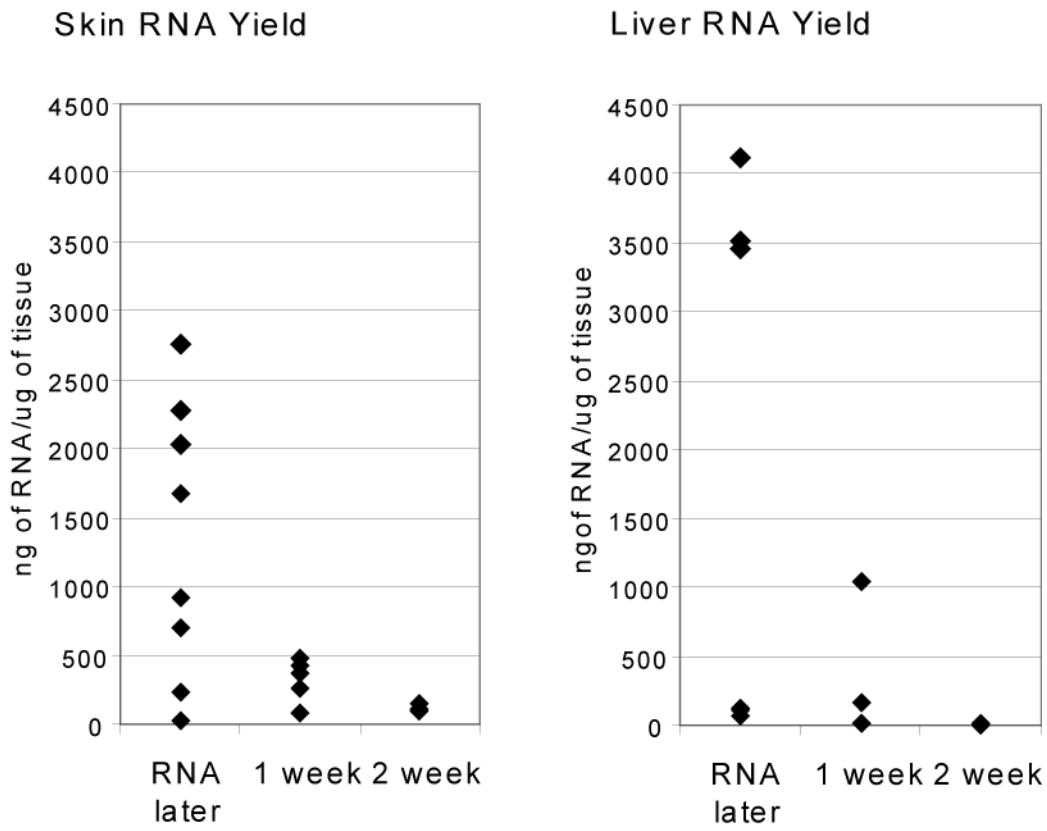

Figure 5.3: Liver and skin RNA
Yields of RNA per ug of original sample weight from mouse skin (A) and liver (B) samples. These represent two sets of experiments; each preservation condition was tested in three to six samples per tissue examined.

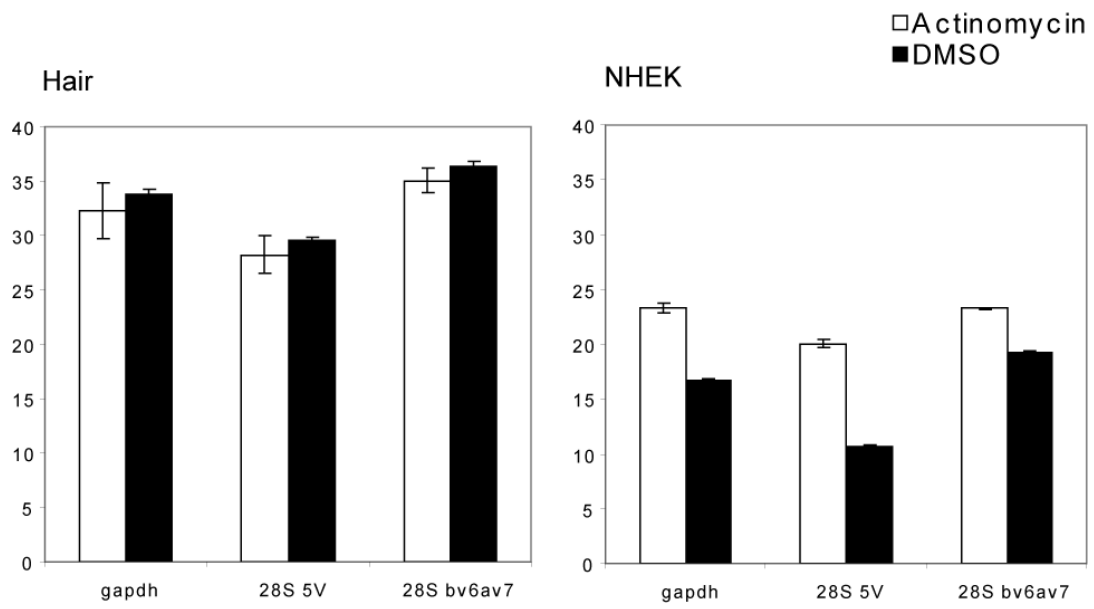

Figure 5.4: Actinomycin treatment of hair and keratinocytes
Cycle numbers were averaged from five hair samples each incubated with actinomycin or DMSO. As a positive control, NHEK cells (keratinocytes) were treated in parallel. While higher cycle numbers indicate that transcription is inhibited in NHEK cells treated with actinomycin compared with cells treated with the vehicle, DMSO, no such reduction is apparent in hair samples examined.

Table 5.1: List of common proteins in hair and nails
Boldface denotes mRNA present in hair library.
Italics denotes protein present in nail.

Symbol	Symbol	Symbol
ACTB	HIST1H4E	**KRTAP26-1**
ACTG1	HIST1H4F	*LAP3*
AIM1	HIST1H4H	LGALS3
ALDH2	HIST1H4I	LMNA
ANXA2	HIST1H4J	*LRRC15*
ANXA2P2	HIST1H4L	**LYG2**
ATP5B	HIST2H4A	MDH2
BLMH	HIST2H4B	NEU2
CALML3	HIST4H4	**PABPC1**
CLTC	**HSD17B4**	**PADI3**
CTNNB1	HSP90AB1	PKM2
CTSD	HSPA2	*PKP1*
DSP	**HSPA5**	**PLCD1**
EEF1G	**HSPA8**	**PLEC1**
EEF2	*KRT32*	*PRDX6*
EIF4A1	*KRT33B*	RPLP2
ENO1	*KRT85*	S100A3
GAPDH	*KRT86*	*SFN*
GDI2	**KRTAP10-11**	**TGM3**
GSTP1	**KRTAP10-12**	TPI1
HIST1H4A	*KRTAP11-1*	TUBA1A
HIST1H4B	**KRTAP12-2**	TUBB2A
HIST1H4C	KRTAP2-2	*VSIG8*
HIST1H4D	*KRTAP2-4*	

Table 5.2: Nail RNA Yields

ID	Source	Nail weight (g)	Yield (ng/ul)	Ng of RNA per ug of nail	A260/280
1	Fingers	0.012	473.8	1204.6	1.86
2	Fingers	0.014	1511.2	3238.3	1.80
3	Fingers (left)	0.035	103.7	89.8	1.52
3	Fingers (right)	0.032	377.8	357.0	1.80
4	Fingers	0.033	84.4	75.8	1.53
5	Fingers	0.080	135.6	50.9	1.64
1	Toes	0.021	85.1	121.0	1.54
2	Toes	0.098	308.1	94.8	1.89
3	Toes	0.108	251.5	70.0	1.76
4	Toes	0.046	111.2	73.0	1.57
5	Toes	0.109	135.6	37.4	1.80

Chapter 6: Conclusions, future directions, and implications

Summary of project

This project has demonstrated that RNA exists and is preserved in hair. We have shown that the RNA coverage is similar to that of living cells. The RNA present in hair persists and is detectable after up to a year of growth. We have tested another keratinized tissue, nail, and found it is also present. This highlights a novel difference between terminal differentiation and apoptosis. We find that of the transcripts present in the hair shaft are representative of genes known to be expressed in the hair shaft lineage. More intriguingly, we find that there were roughly seven thousand mRNA genes present in the hair library. This large gene set contained genes associated with diseases, compound exposure phenotypes, and various signaling pathways.

6.2: Feasibility of application in population studies

Other sources of RNA, in particular, blood or saliva, require trained specimen collectors, specialized collection tools, and storage and shipping methods. Hair and nail are non-invasive tissues that can be collected and sent by volunteer donors directly. As these tissues are handled, trimmed, and clipped during the normal course of grooming, individual donors would be more likely to handle proper collection and shipment protocols with minimal instruction and no need of an intermediate technician. These samples may be sent via regular mail with no need of specialized packaging or expedited shipping since RNA will be stably preserved, thus a more cost

effective alternative to other shipped tissues. Hair and nail samples may be feasibly solicited successfully from donors on a large scale, nationally or globally.

6.3: Personalized medical screening

As argued in chapter 1, the integration of genotypic and phenotypic data is essential to understanding the mechanisms that cause individual variation in human health. Many of the genes found in the hair shaft library may be indicative of underlying cryptic phenotypes. Hair and nail RNA as a source of eQTL data is promising—there are many disease and phenotype markers present in the initial library. The next step would be to characterize expression levels of candidate genes and screen for biomarkers that associate with disease that are retained in hair RNA.

The predictive value of gene expression analysis of the hair shaft remains to be explored. Some genes of interest may play a role in individual variation in their response to a treatment, perhaps affecting drug or toxin metabolism. The large number of genes associated with compounds (Figure 3.6) suggests that screening could be performed before, during, and after a course of treatment.

6.4: Retrospective screening

One of the properties of nail and hair is their continual growth during the life of an organism. This allows for the possibility of retrospective testing. Sociocultural trends in scalp hair maintenance and grooming make it likely that even a male sporting a short haircut, such as those used in the military, will retain hair representing at least

one week of growth. In this way, we can imagine that when an individual comes into a clinic with a medical problem, the hair and nail may retain information about the past state and may aid in diagnosis and provide a means to assay the state of the individual retrospectively. Two areas of study where this would be useful are pediatrics and forensics.

Newborn babies contain hair and nail that have developed *in utero.* Collecting information about the prenatal environment has involved amniocentesis, an invasive procedure that may introduce contaminants into the placenta and harm the developing fetus, or has relied on sorting out relatively diluted fetal signals from maternal blood. In addition, there is also the possibility that the need to assay the placental environment or fetal development would not be known until after birth. Harvesting the laguno hair or fingernails may provide clues to the past.

One set of conditions where this may be relevant are autism spectrum disorders. Symptoms of autism often only become apparent in late stages, when language or social development is delayed; however, early diagnosis and therapies may be associated with better outcomes (Warren et al., 2011). A screen of autism associated genes present in laguno hair may some day predict the risk of autism and provide an earlier opportunity to begin treatment.

This project's findings may also have some utility in the identification of individuals or in the analysis of the state of health of a deceased individual. Past methods have utilized DNA to genotype. However, one of the drawbacks to this is that mechanisms to degrade DNA exist, leading to a preference to genotype mitochondrial

DNA, or short tandem repeats for its relatively high copy numbers (Szabo et al., 2011). Being able to measure RNA of a deceased individual may offer clues into the general state of the individual and may even help identify the cause of death (Bauer 2007). Hair and nail, which would retain RNA for up to six months with no need for specialized storage, would be valuable in such a case.

Another potential avenue of investigation is the rate and nature of RNA decay. Much work has been focused on ancient DNA, and studies have found that there is a pattern to common nucleotide substitutions. Deamination commonly results in the conversion of cytosine to uracil, which is then read as thymine, or in short, a C →T substitution (Hofreiter et al., 2001). As discussed in Chapter 5, we find that while the yield of RNA from sequential segments does not change noticeably, the abundance of mRNAs and small RNAs decreases. While DNA of ancient samples may be thought of as a fairly one dimensional snapshot of a long-deceased organism, RNA decay observed from still living specimens may be a factor of the environmental stresses or insults. It would be imaginable that should the ectodermal exterior be exposed to harsher environmental conditions, that the RNA would exhibit more degradation and have more substitution errors. Differences in the extent of RNA substitutions may be seen in hair and claws of domesticated versus feral animals or even of the same animal harvested over different living conditions.

6.5: Discovery of novel transcripts or novel functions of transcripts

RNA profiling of the human hair shaft with NGS techniques has led to identification of over 1.2 million sequences that represent mRNAs and many non-coding RNAs. The function of some RNAs is unknown. This template can now be used as the basis to study development, disease, and other biological processes with current experimental methods, such as in situ hybridization or microarray hybridization, which require an *a priori* approach. Targeted assays for RNAs and the protein products of RNAs in the library may be used to detect protein-RNA or RNA-RNA complexes that may play a role in hair follicle biology.

6.6: Localization and developmental cues

Another possible use for hair and nail RNA, is to look for signaling cues related to localization. Differences in signaling gradients specify the anterior-posterior and rostral-caudal axes during development (Rinn et al., 2008). The signals to make a thumb or pinky may be retained as differences in expression of genes. Morphologically, the nails are thicker on the thumb than the pinky and thicker on the toes than fingers. Body hair morphology also varies by location within an individual. Some hairs, such as thicker axillary hairs, are responsive to hormonal signals. Genes responsive to testosterone, estrogen, and progesterone are present in the hair library (Figure 3.6). In cases of hormone signaling, there are conditions that induce site specific loss and gain of hair. Polycystic ovary syndrome is an interesting condition where affected women may experience hair loss on the scalp, termed androgenic alopecia, and excess hair growth on other regions of the body such as the face (Lee

and Zane 2007). At least one keratin has been identified as specific to beard hairs with an upstream androgen receptor binding site (Rutberg et al., 2006). Skin biopsies of individuals with PCOS suggest that there is greater activity of an enzyme that converts testosterone that correlates to higher levels of testosterone (Skałba et al., 2006). The profile of the differentially affected hairs may shed light on underlying localization cues or hormone responsiveness of these follicles to either enhance or retard hair shaft growth.

While early developmental stages are constrained, as evidenced by the similarities of developing embryos across species, hair varies vastly between body sites and between individuals. As studies on the EDAR and ERAR-related alleles, discussed in Chapter 2, have shown, a major question this raises is whether the variation seen is directly because of an adaptive advantage of different hair morphologies or if variations in phenotype might occur as passive non-lethal hitchhikers in evolution. This may lead to future studies of expression differences that clearly affect hair and other systems. The function of the EDAR gene, as described in Chapter 2, is an example of mutations that clearly affect hair follicles, and can be tested in cell culture and animal models. More careful examination of the observable products of EDAR mutation, which may include a more thorough analysis of hair shaft RNA, could involve assaying expression of downstream targets of prolonged or enhanced EDAR signaling, such as NFκB or identifying new ones.

6.7: RNA as clues to tensile strength and other physical properties of hair

It has long been observed that hair is a strong stable biomaterial. These last surviving messages may lend insight into the formation of stable structures. From our preliminary findings, it is plausible that the mammalian specific KRTAP proteins are more important than previously thought. In addition to hair, we find that nail also retains RNA. This suggests that ectodermal tissues may function as a signaling sink for RNA during maturation. This then raises the question of whether RNA retention could be a property of mainly mammals or if the ectodermal appendages of other species would also retain RNA.

One wonders if the high composition of keratin and the resulting disulfide crosslinking may contribute to RNA stability. A suitable comparison would be the claws of green anole lizard and the clawed frog *xenopus laevis*. As mentioned in chapter 2, lizard claws may be homologous keratinized structures while frog claws evolved independently utilizing other genes for strength and hardness (Eckhart et al., 2008) (Maddin et al, 2009). It is plausible to test if the high disulfide bonds in keratinzed tissue is a unique or critical factor in preserving RNA or of other common features of these claws are more important for RNA preservation. It would be interesting to look at other terminally differentiated tissues that are not keratinized, such as the lens of the eye, bone and cartilage.

We could extend this analysis to hardened structures in other more distantly related species. The silk of spiders and silkworms are similar but spider silk is of commercial interest due to its much stronger tensile strength. Largely due to the unsuitability of spiders to domestication, efforts to use spider silk have focused on

introducing spider genes into other, more docile animals, such as goats and silkworms. Comparison of the RNA retained in silk from spiders and silkworms may provide insight into which genes are needed for strength. Expression analysis has found more genes, including less abundantly expressed and non coding RNAs, that proteomic studies may miss due to technical limitations.

6.8: Biological significance of RNA retention

The preservation of RNA in hair shaft and nail raises questions about the biology of terminal differentiation and cornification. We find that RNA can be extracted and amplified from about 50 strands of hair. The amplicons used on our hair RT-qPCR analysis were roughly 100 base pairs in length, suggesting that RNA fragments were at least this long. RNA degradation in living cells is quite efficient and tightly regulated by enzymes and other RNA binding proteins that are present in the hair library. Other cell death programs tend to degrade RNA fairly quickly. The lack of an activated targeted RNA degradation program hints at a function for RNA.

In other specialized cells such as red blood cells, mRNA exists after eunucleation for continued protein synthesis and cell division (Schwertz et al., 2010). It would be interesting to speculate that the preservation of RNA might have some function in hair and nail as well, such as for transcription or as part of a RNA-protein complex, promoting the stability of the cornified hair shaft through non-translational and non-transcriptional mechanisms, or if it was simply a matter of neglect.

6.9: Summary of potential future work

The retention of RNA in keratinized tissue opens up many possible future avenues of investigation. It may be of interest to population genetics and epidemiologists; hair and nail clippings may turn out to be a source of eQTL data that could be obtainable on a large scale. In terms of diagnostic or personalized medicine, there may contain expression data relating to disease states, exposure that could reflect the gene expression profile that leads to observed disease morphology. Even more exciting is the possibility that these changes could be present in the more sensitive RNA transcripts before a disease or condition becomes observable to the naked eye. Further work elucidating the expression profile of hair of various morphologies, or even the hardened structures of other species may identify the key genes that affect structure, size and organization of these organs. This finding of RNA retention also raises interesting questions about the biology of terminal differentiation. There may be undiscovered roles for the persistence of RNA long past the point of cell death.

References

Almasy L, Blangero J. Human QTL linkage mapping. Genetica. 2010;136(2):333-340.

Aoki Y, Niihori T, Narumi Y, Kure S, Matsubara Y. The RAS/MAPK syndromes: Novel roles of the RAS pathway in human genetic disorders. Hum Mutat. 2008 Aug;29;(8):992-1006.

Bal E, Baala L, Cluzeau C, El Kerch F, Ouldim K, Hadj-Rabia S, Bodemer C, Munnich A, Courtois G, Sefiani A, Smahi A. Autosomal dominant anhidrotic ectodermal dysplasias at the EDARADD locus. Hum Mutat. 2007 Jul;28(7):703-9.

Barth J, Telfer N, Dawber R. Nail abnormalities and autoimmunity. J Am Acad Dermatol. 1988;18(5):1062-1065.

Bauer M. RNA in forensic science. Forensic Sci Int Genet. 2007 Mar;1(1):69-74.

Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. Nucleic Acids Res. 2011; 39: D842–848

Borradori L, Blanchet-Bardon C. Skin manifestations of cardio-facio-cutaneous syndrome. J Am Acad Dermatol. 1993 May;28(5 Pt 2):815-9.

Botchkarev VA, Fessing MY. Edar signaling in the control of hair follicle development. J Investig Dermatol Symp Proc. 2005 Dec;10(3):247-51.

Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, Parker HG, Vonholdt BM, Rhue A, Boyko A, Byers A, Wong A, Mosher DS, Elkahloun AG, Spady TC, Andre C, Lark KG, Cargill M, Bustamante CD, Wayne RK, Ostrander EA. Coat variation in the domestic dog is governed by variants in three genes. Science. 2009 Oct 2;326(5949):150-3.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011 Sep 15;25(18):1915-27.

Candi E, Schmidt R, Melino G. The cornified envelope: a model of cell death in the skin. Nat Rev Mol Cell Biol. 2005;6(4):328-40.

Chang CH, Tsai RK, Yu HS. Apoptosis coordinates with proliferation and differentiation during human hair follicle morphogenesis. J Dermatol Sci. 2005;39(1):9-16.

Chang SH, Jobling S, Brennan K, Headon DJ. Enhanced Edar signalling has pleiotropic effects on craniofacial and cutaneous glands. PloS one. 2009;4(10):e7591.

Chen CYA, Shyu AB. Mechanisms of deadenylation-dependent decay. Wiley Interdiscip Rev RNA. 2011;2(2):167-183.

Cheng AS, Bayliss SJ. The genetics of hair shaft disorders. J Am Acad Dermatol. 2008;59(1):1-22

Cheung VG, Spielman RS. The genetics of variation in gene expression. Nat Genet. 2002;32 Suppl:522-5.

Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 2005 Sep 1;437(7055):69-87.

Clement JQ, Maiti S, Wilkinson MF. Localization and stability of introns spliced from the Pem homeobox gene. J Biol Chem. 2001 May 18;276(20):16919-30.

Collins FS, Green ED, Guttmacher AE, Guyer MS (2003) A vision for the future of genomics research. Nature 422: 835–847.

Dalla Valle L, Benato F, Rossi R, Alibardi L, Tschaler E, Eckhart L. Deleterious Mutations of a Claw Keratin in Multiple Taxa of Reptiles. J Mol Evol. 2011:265-273.

Davenport GC, Davenport CB. Heredity of hair-form in man. Am Nat. 1908 May;42(497):341-9.

de Viragh PA, Huber M, Hohl D. Involucrin mRNA Is More Abundant in Human Hair Follicles Than in Normal Epidermis. J Invest Dermatol. 1994;103(6):815-819.

Dean I, Siva-Jothy MT. Human fine body hair enhances ectoparasite detection. Biol Lett. 2011:rsbl.2011.0987.

Del Prete MJ, Robles MS, Guáo A, Martínez-A C, Izquierdo M, Garcia-Sanz JA. Degradation of cellular mRNA is a general early apoptosis-induced event. FASEB J. 2002;16(14):2003-5.

Denecker G, Hoste E, Gilbert B, Hochepied T, Ovaere P, Lippens S, Van den Broecke C, Van Damme P, D'Herde K, Hachem JP, Borgonie G, Presland RB, Schoonjans L, Libert C, Vandekerckhove J, Gevaert K, Vandenabeele P, Declercq W. Caspase-14 protects against epidermal UVB photodamage and water loss. Nat Cell Biol. 2007;9(6):666-74.

Dhouailly D. A new scenario for the evolutionary origin of hair, feather, and avian scales. J Anat. 2009;214(4):587-606.

Driskell RR, Giangreco A, Jensen KB, Mulder KW, Watt FM. Sox2-positive dermal papilla cells specify hair follicle type in mammalian epidermis. Development. 2009;136(16):2815-23.

Drogemuller C, Rufenacht S, Wichert B, Leeb T. Mutations within the FGF5 gene are associated with hair length in cats. Anim Genet. 2007 Jun;38(3):218-21.

Eckhart L, Valle LD, Jaeger K, Ballaun C, Szabo S, Nardi A, Buchberger M, Hermann M, Alibardi L, Tschachler E. Identification of reptilian genes encoding hair keratin-like proteins suggests a new scenario for the evolutionary origin of hair. Proc Natl Acad Sci U S A. 2008;105(47):18419-23.

Enari M, Sakahira H, Yokoyama H, Okawa K, Iwamatsu A, Nagata S. A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. Nature. 1998;391(6662):43-50.

Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473(7345):43-9.

Fischer H, Eckhart L, Mildner M, Benarafa C, Torriglia A, Tschachler E, Eckhart L. DNase1L2 degrades nuclear DNA during corneocyte formation. J Invest Dermatol. 2007;127(1):24-30. 1.

Fischer H, Szabo S, Scherz J, Jaeger K, Rossiter H, Buchberger M, Ghannadan M, Hermann M, Theussl HC, Tobin DJ, Wagner EF, Tschachler E, Eckhart L. Essential role of the keratinocyte-specific endonuclease DNase1L2 in the removal of nuclear DNA from hair and nails. J Invest Dermatol. 2011;131(6):1208-15.

Franbourg A, Hallegot P, Baltenneck F, Toutain C, Leroy F. Current research on ethnic hair. J Am Acad Dermatol. 2003 Jun;48(6 Suppl):S115-9.

Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, Batubara L, Mustofa MS, Samakkarn U, Settheetham-Ishida W, Ishida T, Morishita Y, Furusawa T, Nakazawa M, Ohtsuka R, Tokunaga K. A scan for genetic determinants of human hair morphology: EDAR is associated with asian hair thickness. Hum Mol Genet. 2008 March 15;17(6):835-43.

Fujimoto A, Nishida N, Kimura R, Miyagawa T, Yuliwulandari R, Batubara L, Mustofa MS, Samakkarn U, Settheetham-Ishida W, Ishida T, Morishita Y, Tsunoda T, Tokunaga K, Ohashi J. FGFR2 is associated with hair thickness in Asian populations. J Hum Genet, 2009 Aug; 54(8), 461-465.

Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. 2008;24(8):408-15.

Gough M. Rapid and quantitative preparation of cytoplasmic RNA from small numbers of cells. Anal Biochem. 1988;95:93-95.

Hafner M, Landgraf P, Ludwig J, Rice A, Ojo T, Lin C, Holoch D, Lim C, Tuschl T. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. Methods 2008;44(1):3-12.

Hamilton WD, Zuk M. Heritable true fitness and bright birds: a role for parasites? Science 1982; 218(4570) 384-387.

Harfe BD, Scherz PJ, Nissim S, Tian, H, McMahon, and Tabin C. Evidence for an Expansion-Based Temporal Shh Gradient in Specifying Vertebrate Digit Identities. Cell 2004;118:517-528.

He B, Lu N, Zhou Z. Cellular and nuclear degradation during apoptosis. Curr Opin Cell Biol. 2009;21(6):900-12.

Houseley J, Tollervey D. The many pathways of RNA degradation. Cell. 2009;136(4):763-76.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: Detection of large-scale variation in the human genome. Nat Genet. 2004 Sep;36(9):949-51.

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature 431: 931–945.

Janne PA, Suchy SF, Bernard D, MacDonald M, Crawley J, Grinberg A, Wynshaw-Boris A, Westphal H, Nussbaum RL. Functional overlap between murine Inpp5b and Ocrl1 may explain why deficiency of the murine ortholog for OCRL1 does not cause lowe syndrome in mice. J Clin Invest. 1998 May 15;101(10):2042-53.

Kalinowska M, Garncarz W, Pietrowska M, Garrard WT, Widlak P. Regulation of the human apoptotic DNase/RNase endonuclease G: involvement of Hsp70 and ATP. Apoptosis. 2005;10(4):821-30.

Kai ZS, Pasquinelli AE. MicroRNA assassins: factors that regulate the disappearance of miRNAs. Nat Struct Mol Biol. 2010;17(1):5-10.

Kariya N, Shimomura Y, Ito M. Size polymorphisms in the human ultrahigh sulfur hair keratin-associated protein 4, KAP4, gene family. J Invest Dermatol. 2005;124(6):1111-8.

Kasahara T, Miyazaki T, Nitta H, Ono A, Miyagishima T, Nagao T, Urushidani T. Evaluation of methods for duration of preservation of RNA quality in rat liver used for transcriptome analysis. J Toxicol Sci. 2006 Dec;31(5):509-19.

King KL, Jewell CM, Bortner CD, Cidlowski JA. 28S ribosome degradation in lymphoid cell apoptosis: evidence for caspase and Bcl-2-dependent and -independent pathways. Cell Death Differ. 2000;7(10):994-1001.

Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011;39(Database issue):D152-7.

Kroemer G, Galluzzi L, Vandenabeele P, Abrams, J, Alnemri ES, Baehrecke, EH, Blagosklonny, MV, El-Deiry, WS, Golstein, P, Green, DR, Hengartner M, Knight RA, Kumar S, Lipton SA, Malorni W, Nunez G, Peter, ME, Tschopp J, Yuan J, Piacentini M, Zhivotovsky B, and Melino G.. Classification of cell death: recommendations of the Nomenclature Committee on Cell Death 2009. Cell Death Differ. 2009;16(1):3-11.

Kudaravalli S, Veyrieras J-B, Stranger BE, Dermitzakis ET, Pritchard JK. Gene expression levels are a target of recent natural selection in the human genome. Mol Biol Evol. 2009;26(3):649-58.

Langbein L, Rogers MA, Praetzel-Wunder S, Helmke B, Schirmacher P, Scheizer J. K25 (K25irs1), K26 (K25irs2), K27 (K25irs3), and K28 (K25irs4) represent the type I inner root sheath keratins of the human hair follicle. J Invest Dermatol. 2006;126(11):2377-86.

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biol. 2009;10(3):R25.

LeBeau MA, Montgomery MA, Brewer JD. The role of variations in growth rate and sample collection on interpreting results of segmental analyses of hair. Forensic sci int. 2011;210(1-3):110-6.

Lee AT, Zane LT. Dermatologic manifestations of polycystic ovary syndrome. Am J Clin Dermatol. 2007;8(4):201-19.

Lestrade, L., and Weber, M. J. (2006). snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nucleic acids res. 34, D158-162.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

Linch CA. Degeneration of nuclei and mitochondria in human hairs. J forensic sci. 2009;54(2):346-9.

Lind LK, Stecksen-Blicks C, Lejon K, Schmitt-Egenolf M. EDAR mutation in autosomal dominant hypohidrotic ectodermal dysplasia in two swedish families. BMC Med Genet. 2006 Nov 24;7:80.

Lippens S, Hoste E, Vandenabeele P, Agostinis P, Declercq W. Cell death in the skin. Apoptosis. 2009;14(4):549-69.

Locker J. Analytical and preparative electrophoresis of RNA in agarose-urea. Anal biochem. 1979;98(2):358-67.

Magerl M, Tobin DJ, Müller-Röver S, Hagen E, Lindner G, McKay IA, and Paus R. Patterns of proliferation and apoptosis during murine hair follicle morphogenesis. Journal invest dermatol. 2001;116(6):947-955.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy, MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak, L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747-53.

Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, Gordon SD, Ferreira MA, Wright MJ, Henders AK, Campbell MJ, Duffy DL, Hansell NK, Macgregor S, Slutske WS, Heath AC, Montgomery GW, Martin NG. Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet. 2009 Nov;85(5):750-5.

Medland SE, Zhu G, Martin NG. Estimating the heritability of hair curliness in twins of european ancestry. Twin Res Hum Genet. 2009 Oct;12(5):514-8.

Mendel, Gregor. Experiments in plant hybridization. Royal Horticultural Society of London, Translator. Cambridge: Harvard University Press, 1967. 41 p.

Mesko N, Bereczkei T. Hairstyle as an adaptive means of displaying phenotypic quality. Hum Nat. 2004;15(3):251-270.

Millar SE. Molecular mechanisms regulating hair follicle development. J invest dermatol. 2002;118(2):216-25.

Moore MJ. Nuclear RNA turnover. Cell. 2002 Feb 22;108(4):431-4.

Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA. Application of

massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome res. 2008;18(4):610-21.

Morioka, K. Hair follicle: differentiation under electron microscope : an atlas. 1st ed. New York: Springer; 2005. 165p.

Mou C, Thomason HA, Willan PM, Clowes C, Harris WE, Drew CF, Dixon J, Dixon MJ, Headon DJ. Enhanced ectodysplasin-A receptor (EDAR) signaling alters multiple fiber characteristics to produce the east asian hair form. Hum Mutat. 2008 Dec;29(12):1405-11.

Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature (2002) 420: 520 -562.

Muller, H. J. 1932. Further studies on the nature and causes of gene mutations. Proceedings of the 6th International Congress of Genetics, pp. 213–255.

Naeem M, Muhammad D, Ahmad W. Novel mutations in the EDAR gene in two pakistani consanguineous families with autosomal recessive hypohidrotic ectodermal dysplasia. Br J Dermatol. 2005 Jul;153(1):46-50.

Naito T, Yokogawa T, Takatori S, et al. Role of RNase L in apoptosis induced by 1-(3-C-ethynyl-beta-D-ribo-pentofuranosyl)cytosine. Cancer chemother pharmacol. 2009;63(5):837-50.

Newman MA, Hammond SM. Lin-28: an early embryonic sentinel that blocks Let-7 biogenesis. Int J Biochem Cell Biol. 2010;42(8):1330-3.

Ogle, R, Fox, MJ. Atlas of human hair microscopic characteristics. 1st ed. New York: CRC Press; 1998. 96p.

Pace CN, Laurents DV, Thomson J a. pH dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1. Biochemistry. 1990;29(10):2564-72.

Pagel M, Bodmer W. A naked ape would have fewer parasites. Proc Biol Sci. 2003;270 Suppl (Suppl_1):S117-9.

Pennacchio L,A., Ahituv ,Nadav, Moses A,M., Prabhakar ,Shyam, Nobrega M,A., Shoukry ,Malak, Minovitsky ,Simon, Dubchak ,Inna, Holt ,Amy, Lewis K,D., Plajzer-Frick ,Ingrid, Akiyama ,Jennifer, De Val ,Sarah, Afzal ,Veena, Black B,L., Couronne ,Olivier, Eisen M,B., Visel ,Axel, Rubin E,M. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 2006;444(7118):499-502.

Perrin C. Expression of Follicular Sheath Keratins in the Normal Nail With Special Reference to the Morphological Analysis of the Distal Nail Unit. Am J Dermatopathol. 2007;29(6):543-550.

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC. Diet and the evolution of human amylase gene copy number variation. Nat Genet. 2007 Oct;39(10):1256-60.

Pillai S, Bikle DD, Hincenbergs M, Elias PM. Biochemical and morphological characterization of growth and differentiation of normal human neonatal keratinocytes in a serum-free medium. J cell physiol. 1988;134(2):229-37.

Pispa J, Pummila M, Barker PA, Thesleff I, Mikkola ML. Edar and troy signalling pathways act redundantly to regulate initiation of hair follicle development. Hum Mol Genet. 2008 Nov 1;17(21):3380-91.

Plato. The Republic. Jowett, B, Translator; [Internet]. Project Gutenberg. October 1998. [cited 2012 Jan 19] Available from: http://www.gutenberg.org/1/4/9/1497/

Plato. The Republic (Kindle Location 9724).

Porter RM. The new keratin nomenclature. J invest dermatol. 2006;126(11):2366-8.

Price VH, Odom RB, Ward WH, Jones FT. Trichothiodystrophy: Sulfur-deficient brittle hair as a marker for a neuroectodermal symptom complex. Arch Dermatol. 1980 December 1;116(12):1375-84.

Proudfoot NJ. Ending the message: poly(A) signals then and now. Genes Dev. 2011;25(17):1770-1782.

Quinlan, A. R., and Hall, I. M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MT, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Gronnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TF, Ramsey CB, Hansen TV, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Ponten T, Villems R, Nielsen R, Krogh A, Wang J, Willerslev E. Ancient human genome sequence of an extinct palaeo-eskimo. Nature. 2010 Feb 11;463(7282):757-62.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. Nature. 2006 Nov 23;444(7118):444-54.

Rendl M, Lewis L, Fuchs E. Molecular dissection of mesenchymal-epithelial interactions in the hair follicle. PLoS biol. 2005;3(11):e331.

Rice RH, Xia Y, Alvarado RJ, Phinney BS, Proteome NP. Proteomic Analysis of Human Nail Plate. J Proteome Res. 2010:6752-6758.

Rinn JL, Wang JK, Allen N, Brugmann SA, Mikels AJ, Liu H, Ridky TW, Stadler HS, Nusse R, Helms JA, Chang HY. A dermal HOX transcriptional program regulates site-specific epidermal fate. Genes dev. 2008;22(3):303-7.

Robbins CR. Chemical and physical behavior of human hair. 3rd ed. New York: Springer; 1994. 391p.

Rockman MV, Kruglyak L. Genetics of global gene expression. Nat Rev Genet. 2006;7(11):862-872.

Rogers MA, Schweizer J. Human KAP genes, only the half of it? Extensive size polymorphisms in hair keratin-associated protein genes. J Invest Dermatol. 2005;124(6):vii-ix.

Roush S, Slack FJ. The let-7 family of microRNAs. Trends cell biol. 2008;18(10):505-16.

Rubin GM, Yandell MD, Wortman JR, Gabor GL, Miklos, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, Cherry JM, Henikoff S, Skupski MP, Misra S, Ashburner M, Birney E, Boguski MS, Brody T, Brokstein P, Celniker SE, Chervitz SA, Coates D, Cravchik A, Gabrielian A, Galle RF, Gelbart WM, George RA, Goldstein LSB, Gong F, Guan P, Harris NL, Hay BA, Hoskins RA, Li J, Li Z, Hynes RO, Jones SJM, Kuehl PM, Lemaitre B, Littleton JT, Morrison DK, Mungall C, O'Farrell PH, Pickeral OK, Shue C, Vosshall LB, Zhang J, Zhao Q, Zheng XH, Zhong F, Zhong W, Gibbs R, Venter JC, Adams MD, Lewis S. Comparative genomics of the eukaryotes. Science. 2000 March 24;287(5461):2204-15.

Runkel F, Klaften M, Koch K, Bohnert V, Bussow H, Fuchs H, Franz T, Hrabe de Angelis M. Morphologic and molecular characterization of two novel Krt71

(Krt2-6g) mutations: Krt71rco12 and Krt71rco13. Mamm Genome. 2006 Dec;17(12):1172-82.

Rutberg SE, Kolpak ML, Gourley JA, Tan G, Henry JP, Shander D. Differences in expression of specific biomarkers distinguish human beard from scalp dermal papilla cells. J invest dermatol. 2006;126(12):2583-95.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES; International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Johnson TA, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM,

Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007;449(7164):913-8.

Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 2008 May 6;6(5):e107.

Schneider MR, Schmidt-Ullrich R, Paus R. The hair follicle as a dynamic miniorgan. Curr biol. 2009;19(3):R132-42.

Schwertz H, Köster S, Kahr WH, Michetti N, Kraemer BF, Weitz DA, Blaylock RC, Kraiss LW, Greinacher A, Zimmerman GA, Weyrich AS. Anucleate platelets generate progeny. Blood. 2010;115(18):3801-9.

Schuster S,C., Miller ,Webb, Ratan ,Aakrosh, Tomsho L,P., Giardine ,Belinda, Kasson L,R., Harris R,S., Petersen D,C., Zhao ,Fangqing, Qi ,Ji, Alkan ,Can, Kidd J,M., Sun ,Yazhou, Drautz D,I., Bouffard ,Pascal, Muzny D,M., Reid J,G., Nazareth L,V., Wang ,Qingyu, Burhans ,Richard, Riemer ,Cathy, Wittekindt N,E., Moorjani ,Priya, Tindall E,A., Danko C,G., Teo W,Siang, Buboltz A,M., Zhang ,Zhenhai, Ma ,Qianyi, Oosthuysen ,Arno, Steenkamp A,W., Oostuisen ,Hermann, Venter ,Philippus, Gajewski ,John, Zhang ,Yu, Pugh B,Franklin, Makova K,D., Nekrutenko ,Anton, Mardis E,R., Patterson ,Nick, Pringle T,H., Chiaromonte ,Francesca, Mullikin J,C., Eichler E,E., Hardison R,C., Gibbs R,A., Harkins T,T., Hayes V,M. Complete khoisan and bantu genomes from southern africa. Nature. 2010 Feb 18;463(7283):943-7.

Sebastiani P, Riva A, Montano M, Pham P, Torkamani A, Scherba E, Benson G, Milton JN, Baldwin CT, Andersen S, Schork NJ, Steinberg MH and Perls TT (2012) Whole genome sequences of a male and female supercentenarian, ages greater than 114 years. Front Gene. 2:90. doi: 10.3389/fgene.2011.00090

Shan X, Chang Y, Lin CL. Messenger RNA oxidation is an early event preceding cell death and causes reduced protein expression. FASEB J. 2007;21(11):2753-64.

Sharov AA, Sharova TY, Mardaryev AN, di Vignano AT, Atoyan R, Weiner L, Yang S, Brissette JL, Dotto GP, Botchkarev VA. Bone morphogenetic protein signaling regulates the size of hair follicles and modulates the expression of cell cycle-associated genes. Proc Natl Acad Sci U S A. November 28;103(48):18166-71.

Shimomura Y, Aoki N, Ito K, Ito M. Gene expression of Sh3d19, a novel adaptor protein with five Src homology 3 domains, in anagen mouse hair follicles. J Dermatol Sci. 2003;31(1):43-51.

Shimomura Y, Aoki N, Schweizer J, Langbein L, Rogers MA, Winter H, Ito M: Polymorphisms in the human high sulfur hair keratin-associated protein 1, KAP1, gene family. J Biol Chem 2002, 277(47):45493-45501.

Shimomura Y, Wajid M, Petukhova L, Kurban M, Christiano AM. Autosomal-dominant woolly hair resulting from disruption of keratin 74 (KRT74), a potential determinant of human hair texture. Am J Hum Genet. 2010 Apr 9;86(4):632-8.

Shiokawa D, Tanuma S. Characterization of human DNase I family endonucleases and activation of DNase gamma during apoptosis. Biochemistry. 2001;40(1):143-52.

Shyu A-B, Wilkinson MF, van Hoof A. Messenger RNA regulation: to translate or to degrade. EMBO J. 2008;27(3):471-81.

Skałba P, Dabkowska-Huć A, Kazimierczak W, Samojedny A, Samojedny MP, Chełmicki Z. Content of 5-alpha-reductase (type 1 and type 2) mRNA in dermal papillae from the lower abdominal region in women with hirsutism. Clin Exp Dermatol. 2006 Jul;31(4):564-70.

Stelzer G, Inger A, Olender T, Iny-Stein T, Dalah I, et al. (2009) GeneDecks: paralog hunting and gene-set distillation with GeneCards annotation. OMICS 13: 477–487.

Soligo C, Müller A. Nails and claws in primate evolution. J Hum Evol. 1999;36(1):97-114.

Sriwiriyanont P, Hachiya A, Pickens WL, Moriwaki S, Kitahara T, Visscher MO, Kitzmiller WJ, Bello A, Takema Y, Kobinger GP. Effects of IGF-binding protein 5 in dysregulating the shape of human hair. J Invest Dermatol. 2011 Feb;131(2):320-8. Epub 2010 Oct 14.

Sundberg JP, King LE,Jr. Morphology of hair in normal and mutant laboratory mice. Eur J Dermatol. 2001 Jul-Aug;11(4):357-61.

Synnott A. Shame and Glory: A Sociology of Hair. Br J Sociol. 1987 Sep; 38 (3): 381-413.

Szabo S, Jaeger K, Fischer H, Tschachler E, Parson W, Eckhart L. In situ labeling of DNA reveals interindividual variation in nuclear DNA breakdown in hair and

may be useful to predict success of forensic genotyping of hair. Int j legal med. 2011:1-8-8.

Tano K, Mizuno R, Okada T, et al. MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. FEBS lett. 2010;584(22):4575-80.

The ENCODE Project Consortium 2011 A User's Guide to the Encyclopedia of DNA Elements (ENCODE). PLoS Biol 9(4): e1001046.

The International HapMap Consortium. A haplotype map of the human genome. 2005 Nature 437, 1299-1320.

Thibaut S, Cavusoglu N, de Becker E, Zerbib F, Bednarczyk A, Schaeffer C, van Dorsselaer A, Bernard BA. Transglutaminase-3 enzyme: a putative actor in human hair shaft scaffolding? J Invest Dermatol. 2009 Feb;129(2):449-59.

Thibaut S, Candi E, Pietroni V, et al. Transglutaminase 5 expression in human hair follicle. J invest dermatol. 2005;125(3):581-5.

Thomas DJ, Rosenbloom KR, Clawson H, Hinrichs AS, Trumbower H, Raney BJ, Karolchik D, Barber GP, Harte RA, Hillman-Jackson J, Kuhn RM, Rhead BL, Smith KE, Thakkapallayil A, Zweig AS, The ENCODE Project Consortium , Haussler D, Kent WJ. The ENCODE project at UC Santa Cruz. Nucl Acids Res. 2006 December 13

Tsang M, Dawid IB, Promotion and Attenuation of FGF Signaling Through the Ras-MAPK Pathway. Sci. STKE 2004, pe17 (2004).

Vaux DL, Korsmeyer SJ. Cell death in development. Cell. 1999;96(2):245-54.

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat rev genet. 2009;10(1):57-63.

Warren Z, McPheeters ML, Sathe N, Foss-Feig JH, Glasser A, Veenstra-Vanderweele J. A systematic review of early intensive intervention for autism spectrum disorders. Pediatrics. 2011 May;127(5):e1303-11. Epub 2011 Apr 4.

Watts JA, Morley M, Burdick JT, Fiori JL, Ewens WJ, Spielman RS, Cheung VG. Gene expression phenotype in heterozygous carriers of ataxia telangiectasia. Am J Hum Genet. 2002;71(4):791-800.

Whiting DA, Dy LC (2006) Office diagnosis of hair shaft defects. Semin Cutan Med Surg 25: 24–34.

Winter H, Hofmann I, Langbein L, Rogers MA, Schweizer J. A splice site mutation in the gene of the human type I hair keratin hHa1 results in the expression of a tailless keratin isoform. J Biol Chem. 1997;272(51):32345-52.

Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, Lam WL. A comprehensive analysis of common copy-number variations in the human genome. Am J Hum Genet. 2007 Jan;80(1):91-104.

Wu DD, Irwin DM, Zhang YP. Molecular evolution of the keratin associated protein gene family in mammals, role in the evolution of mammalian hair. BMC evol biol. 2008;8(1):241.

Yamamoto S, Hirai K, Hasegawa-Oka Y, Hirai Y. Molecular elements of the regulatory control of keratin filament modulator AHF/trichohyalin in the hair follicle. Exp Dermatol. 2009 Feb;18(2):152-9.

Yell JA, Mbuagbaw J, Burge SM. Cutaneous manifestations of systemic lupus erythematosus. Br j dermatol. 1996;135(3):355-362.

Yi R, O'Carroll Donal, Pasolli H,A., Zhang ,Zhihong, Dietrich F,S., Tarakhovsky ,A, Fuchs ,Elaine. Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. Nat Genet. 2006 Mar;38(3):356-62.

Yi R, Poy MN, Stoffel M, Fuchs E (2008) A skin microRNA promotes differentiation by repressing 'stemness'. Nature 452: 225–229.

Yuan J, Shaham S, Ledoux S, Ellis HM, Horvitz HR. The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 beta-converting enzyme. Cell. 1993;75(4):641-52.

Zhang Y, Tomann P, Andl T, Gallant NM, Huelsken J, Jerchow B, Birchmeier W, Paus R, Piccolo S, Mikkola ML, Morrisey EE, Overbeek PA, Scheidereit C, Millar SE, Schmidt-Ullrich R. Reciprocal requirements for EDA/EDAR/NF-kappaB and Wnt/beta-catenin signaling pathways in hair follicle induction. Dev Cell. 2009 Jul;17(1):49-61.