Utilizing functional genomics approaches to characterize risk genes in alopecia areata

Stephanie O. Erjavec

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

Abstract

Utilizing functional genomics approaches to characterize risk genes in alopecia areata

Stephanie O. Erjavec

Understanding the genetic architecture of complex disorders is important for identifying disease mechanisms and potential molecular targets for therapeutic interventions. Genetic diseases are broadly classified as either Mendelian (monogenic) diseases or complex (polygenic) diseases. Common, polygenic disorders result from inheritance of multiple common variants with low penetrance. In contrast, monogenic, Mendelian disorders are caused by rare variants with high penetrance at a single genetic locus. However, an increasing number of studies support a role for rare variants of moderate effect size in complex diseases. As a result, genetic approaches previously utilized for discovering rare variants in Mendelian diseases, such as next generation sequencing, can now effectively be applied to complex diseases to define the contribution of both rare and common variants to the genetic burden of polygenic traits and diseases.

Alopecia Areata (AA) is a complex autoimmune disease characterized by non-scarring hair loss that is due to a combination of both environmental and genetic factors. Our previous Genome-wide Association Study (GWAS) identified at least 14 genetic regions contributing to AA disease susceptibility. Although useful in identifying disease-associated loci and surrounding linkage disequilibrium (LD) blocks, GWAS is not sufficiently granular to 1) elucidate causal (association-driving) variants; and 2) discover rare risk variants. This level of resolution can only be achieved by deep sequencing followed by functional validation of variants.

The goal of this thesis was to address these challenges in AA using two genetic approaches that have not been previously utilized in the context of this disease. In Chapter 2, I performed a *hypothesis-driven* analysis of *common* variants in a GWAS-associated locus using targeted genomic sequencing. In Chapter 3, I utilized whole exome sequencing (WES) in an

unbiased approach to assess *rare* variant contribution in AA disease risk. To conduct these analyses simultaneously, we designed a whole exome sequencing (WES) chip that also included custom capture of 24 Mb of genomic regions covering the 14 genetic loci previously identified using GWAS. I applied these two sequencing approaches in a large AA patient cohort and identified potentially causal variants in several genes. To interrogate the consequences of these variants, I performed functional analyses to determine the effects of disease-causing variants on the target organ of AA attack, the hair follicle (HF).

In Chapter 2, I report on the use of targeted genomic sequencing to interrogate the coding and non-coding regions surrounding a previously implicated GWAS locus. This approach provided fine mapping of coding and non-coding *common* variants in regions that may be contributing to disease risk. In this thesis, I focused on the effect of genetic perturbations on the end-organ HF, and consequentially prioritized my functional analysis using two criteria: 1) genes expressed in the (HF) and; 2) GWAS regions that were not previously implicated in other autoimmune diseases. One of the regions that satisfied these criteria harbored the *Syntaxin 17* (*STX17*) gene, which encodes a SNARE protein involved in autophagy and mitochondrial fission. Targeted genomic sequencing of the *STX17* region in 849 AA cases identified 35 non-coding and 0 coding variants in high LD with the GWAS SNP. Thirty-three variants were significantly enriched in cases compared to controls, and the remaining two were nominally significant. Thirty-two of the significantly associated AA variants were confirmed to be AA skin eQTLs that downregulated expression of STX17 in affected scalp skin of AA patients. Downstream analyses incorporated *in silico* and functional cell assays that uncovered a novel autophagy-independent role for *STX17* in melanocyte biology. I discovered that a reduction of *STX17* expression was associated with an accumulation of a melanocyte-specific antigen and increased immunogenicity, as seen by CD8+ T cell infiltrates in the skin of AA patients with low levels of *STX17* expression. I used a targeted sequencing approach to successfully identify candidate causal variants driving the GWAS association at the *STX17* locus, and propose a novel mechanism underlying *STX17*-dependent melanocyte perturbation and AA disease.

In the second section of this thesis, we used the WES feature of the chip to assess the genetic contribution of *rare* variation in AA, in a genome-wide and unbiased manner. WES data and gene-level burden analyses of 18,653 genes in 849 AA patients was compared to 15,640 controls to identify rare variants associated with AA. Unexpectedly, this analysis identified one gene, encoding a hair-specific keratin, *Keratin 82* (*KRT82)* that harbored significantly more rare damaging mutations in AA cases compared to controls (p=2.68E-06). Eleven rare damaging mutations were found in 51 AA patients in the heterozygous state (6.01%) compared to 2.58% controls. These variants resided in evolutionary conserved amino acid residues, and nine out of the eleven mutations were located in established disease-causing domains in keratin proteins. I determined that *KRT82* expression was absent or largely reduced in AA hair follicles, including the bulb region, the site of AA immune attack. Moreover, AA patients with damaging variants *and* reduced *KRT82* expression had increased perifollicular CD8+ T cell infiltrates in comparison to control HFs with intact *KRT82* expression remaining. I proposed that damaging mutations in the coding regions of *KRT82* resulted in loss of functional protein, thereby weakening the protective HF cuticle and predisposing the HF to immune attack.

In summary, I used two genetic approaches (targeted genomic sequencing and WES) to identify *common* (Chapter 2) and *rare* variants (Chapter 3) with novel contributions to the complex genetic architecture of AA. I focused my functional studies on genes expressed in the target HF, with the goal of defining the role of unidentified, variant-mediated end-organ disruption in the predisposition of AA patient HFs to aberrant autoimmune attack. Up to now, most efforts in AA mechanistic studies have focused on the aberrant immune response. The work in my thesis uncovered novel roles for perturbations in the HF itself as a participating factor in AA disease risk.

Table of Contents

List of Charts, Graphs, Illustrations

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations

A: adenine
AA: alopecia areata
AAP: alopecia areata patchy
ALADIN: Alopecia areata disease severity index
ALS: amyotrophic lateral sclerosis
αMSH: alpha-melanocyte stimulating hormone
AP1: adaptor protein 1
AP3: adaptor protein 3
APC: antigen presenting cells
AT: alopecia areata totalis
ATXN2: ataxin 2
AU: alopecia areata universalis
BIM: Bcl-2 interacting mediator of cell death
bp: base pair
BL: basal lamina
BLOC1/2: biogenesis of lysosome-related organelles complex 1/2
C: cytosine
cAMP: cyclic adenosine monophosphate
CD4: cluster of differentiation 4
CD8: cluster of differentiation 8
CD200: cluster of differentiation 200
ChIP-seq: chromatin immunoprecipitation sequencing
Chr: chromosome
CIITA: class II major histocompatibility complex transactivator
CLEC16A: c-type lectin domain containing 16A
CNVs: copy number variants
CTL: cytotoxic T lymphocytes
CTLA4: cytotoxic T-lymphocyte-associated protein 4
CTS: connective tissue sheath
CXCL9: C-X-C Motif Chemokine Ligand 9
CXCL9: C-X-C Motif Chemokine Ligand 10
CXCL9: C-X-C Motif Chemokine Ligand 11
DLX4: distal-less homeobox 4
DNase: deoxyribonuclease
DP: dermal papilla
DSG4: desmoglein 4
EMH: exclamation mark hairs
eQTL: expression quantitative trait loci
EM: electron microscopy
ENCODE: Encyclopedia of DNA elements
ER: endoplasmic reticulum
ExAC: Exome Aggregation Consortium
FUN-LDA: Functional- Latent Dirichlet Allocation
G: guanine
gnomAD: Genome Aggregation Database
GOF: gain-of-function
GPR143: G protein-coupled receptor 143
GWAS: Genome-wide association study
GZMB: granzyme B

HF: hair follicle
HFPU: hair follicle pigmentary unit
HLA: human leukocyte antigens
HOXC13: homeobox C13
ICOS: inducible T cell costimulator
IF: intermediate filament
IFN: interferon
IKZF4: IKAROS family zinc finger 4
IKZF1: IKAROS family zinc finger 1
IL2: interleukin-2
IL2RA: interleukin 2 receptor subunit alpha
IL4: interleukin-4
IL10: interleukin-10
IL13: interleukin-13
IL21: interleukin-21
IL23R: interleuin-23 receptor
Indels: insertions/deletions
IP: immune privilege
IRS: inner root sheath
kb: kilobase
kDa: kilodalton
KIT: KIT proto-oncogene, receptor tyrosinase kinase
K/KRT: keratin
KRT31: keratin 31
KRT32: keratin 32
KRT33B: keratin 33B
KRT35: keratin 35
KRT39: keratin 39
KRT40: keratin 40
KRT82: keratin 82
LC3: microtubule-associated protein 1A/1B- light chain 3
LD: linkage disequilibrium
LOF: loss-of-function
LRRC32: leucine rich repeat containing 32
MX: matrix
MAF: minor allele frequency
MART1: melanoma antigen recognized by T cells 1
Mb: mega base
MC1R: melanocortin 1 receptor
MHC: major histocompatibility complex
MICA: MHC class I polypeptide-related sequence A
MIF: macrophage migration inhibitory factor
MITF: microphthalmia-associated transcription factor/ melanocyte inducing transcription factor
mRNA: messenger RNA
MS: multiple sclerosis
MX1: MX Dynamin Like GTPase 1
NFJS: Naegeli-Franceschetti-Jadassohn syndrome
NGS: next generation sequencing
NK: natural killer
NKG2D: Natural killer group 2D receptor
NM: normal unaffected melanocyte

NR4A3: nuclear receptor subfamily 4 group A member 3
OCA2: oculocutaneous albinism II
OMIM: Online Mendelian Inheritance in Man
ORS: outer root sheath
PKP1: plakophilin 1
PM: degenerating pigment cell
PMEL17: premelanosome protein
PRDX5: peroxiredoxin 5
PRF1: perforin 1
PTPN22: protein tyrosine phosphatase, non-receptor type 22
Q-SNARE: glutamine (Q) -SNARE
Rab7: member RAS Oncogene Family 7
Rab27a: member RAS Oncogene Family 27a
Rab32: member RAS Oncogene Family 32
Rab38: member RAS Oncogene Family 38
ROS: reactive oxygen species
R-SNARE: Arginine (R) -SNARE
rs: reference SNP
SG: sebaceous gland
SH2B3: SH2B adaptor protein 3
SLC45A2: solute carrier family 45 member 2
SLE: systemic lupus erythematosus
SNARE: soluble N-ethylmaleimide-sensitive factor attachment protein receptor
SNAP-23: synaptosomal-associated protein 23
SNAP-29: synaptosomal-associated protein 29
SNP: single nucleotide polymorphism
SNV: single nucleotide variant
SOCS1: Suppressor of cytokine signaling 1
SOD: superoxide dismutase
SOX21: SRY-related HMG-box 21
STAT1: signal transducer and activator of transcription 1
STAT4: signal transducer and activator of transcription 4
STX3: syntaxin 3
STX13: syntaxin 13
STX17: syntaxin 17
T: thymine
TGF-β: transforming growth factor beta
TMD: transmembrane domain
TraP: Transcript-inferred Pathogenicity
TSS: Transcription start sites
TYR: tyrosinase
TYRP1: tyrosinase related protein 1
TYRP2: tyrosinase related protein 2
ULBP3/6: UL16 Binding Protein 3/6
UTR: untranslated region
UV: ultraviolet
VAMP7: vesicle-associated membrane protein 7
VAMP8: vesicle-associated membrane protein 8
VIPR: vasoactive intestinal polypeptide receptor 1
WES: whole exome sequencing
WGS: whole genome sequencing

Acknowledgments

First and foremost, I would like to thank my mentor, Dr. Angela Christiano, who gave me the incredibly rare gift of undertaking the precise thesis that I intended for when I embarked on my Ph.D. journey. In our first meeting, we spoke about my ideal thesis and my desire to focus on human disease and the characterization of a specific gene in disease development. Over the last 5 years, Angela artfully guided my thesis to achieve that goal. She provided me with the independence to explore my interests, think critically, and develop confidence in myself as a scientist; all while serving as an impressive example of a hard-working and successful woman in science. She additionally provided me the experiences and opportunities to speak at international conferences, which undoubtedly shaped not only my scientific endeavors but also my growth as an individual. Most importantly, she also fostered a lab environment that encouraged collaboration, hard work, and incredible co-workers. Colleagues became friends and made the day-to-day of lab exponentially more enjoyable. Additionally, without the pivotal guidance of key senior lab mates, Drs. Yanne Doucet, Lynn Petukhova, Eddy Wang, and Rolando Perez-Lorenzo, this thesis would not have been possible. They selflessly spent countless hours mentoring me, teaching me techniques, and developing both my hard and soft skills. Within the Christiano lab, my scientific experiments were dependent on the hard-working and kind assistance of Ming Zhang, Wangyong Zeng, and Rong Du. The Christiano lab also brought me one of the best gifts of my Ph.D. journey, Dr. Alexa Abdelaziz. Alexa was a great source of support for me in the most unknown and trying times. She gave me encouragement when I needed it and watching her succeed was inspiring and motivating. I feel infinitely lucky that we were able to accomplish our Ph.D. goals side-by-side and become life-long friends.

I am also grateful for the additional friends that I made through the Genetics & Development program, Dr. Meghan Pantalia, Dr. Tarun Nambiar, Daniel Krizay, Corey Garyn, and Laura Crowley. I could not have asked for a more supportive, fun, and amusing cohort and I cannot wait to watch everything that they will achieve throughout their careers. In the Genetics & Development Department, I'd like to especially thank Stacy Warren for all her help over the past 5 years. I would also like to express my deepest gratitude to my thesis committee: Drs. Eric Schon, Luke Berchowitz, David Owens, and Rudy

Leibel. Our discussions were lively and exciting, and I cherish the extensive knowledge and expertise they brought to my thesis work.

Although not sources of scientific support, my friends both in New York and California provided me the emotional support necessary to complete this thesis. I am eternally grateful to Rachel Duc, Rachel Nieman, Claire Paganussi, Tessa Van Bergen, Kristine Vargo, Ali Duvaras and John English for not only attempting to understand the intricate workings of biology and genetics, but also for being so actively proud of my achievements. Their genuine and constant ever-flowing pride gave me the foundation to have confidence in myself as an individual as well as a scientist. Not only would I not have been able to undertake this program without their unwavering support, but the last 5 years would not have been a fraction as fun and memorable without them by my side. My gratitude for their humor, liveliness, and love is infinite.

Finally, and most importantly, I would like to thank my family. My grandmother, Helen O'Toole, inspired me to pursue a Ph.D. in the study of human diseases because of her fearless and tenacious mentality battling multiple sclerosis. She was one the strongest woman I ever met and never let her limitations define her, which was an invaluable lesson to learn from her as a child. Equally as strong and inspiring is my mother, Staci O'Toole Erjavec, who has been my role model and best friend since the day I was born. Her ability to care for others, persevere with positivity, and provide tough love when needed has made her the rock of our family and a pivotal part of my thesis. She and my father, Tim Erjavec, provided me with the structure, encouragement, and unconditional love that was necessary for developing my character and discipline that I leaned on throughout my Ph.D. career. My dad always gave me insightful and nuanced advice that shaped how I see the world, and how I approach problems. His unique ability to identify a challenge and work through it in a methodic and stepwise fashion influenced my critical thinking skills and how I approached much of my thesis work. Finally, I'd like to thank my sister, Natalie Erjavec, who provided selfless support and love whenever I needed her. My most cherished memories revolve around my family, and I look forward to the many more to come. I am eternally grateful to them for their support, love, guidance, and the life that we have together. You have all my love.

Dedication

To my family, you are my whole world.

Chapter 1 Introduction

1.1 Genetic Disease and Patterns of Inheritance

Mendelian Diseases

The study of human genetics and disease heritability have provided unparalleled insights into pathogenic mechanisms underlying human disease. Inherited disorders are transmitted across generations by either Mendelian or non-Mendelian modes of inheritance. Mendelian inheritance follows a pattern in which genetic alleles are tightly linked to a phenotypic trait. In this model, inheriting a pathogenic allele(s) results in disease in a dominant or recessive mode of inheritance. In these disease types, penetrance is high and mutations in a single gene strongly influence disease development. In 2020, there were approximately 5,550 documented single-gene traits and diseases catalogued in in publicly available OMIM (Online Mendelian Inheritance in Man) database; annotating causal genes in disease such as Cystic Fibrosis, Sickle Cell Anemia, and Huntington's disease[3]. These diseases are usually rare in the general population due to purifying selection, which prevents the accumulation of deleterious alleles within a population. Consequentially, monogenic diseases are primarily a result of deleterious rare variants (population Minor Allele Frequency (MAF) < 0.5%) with high effect sizes in single genes (Figure 1). These pathogenic Mendelian variants

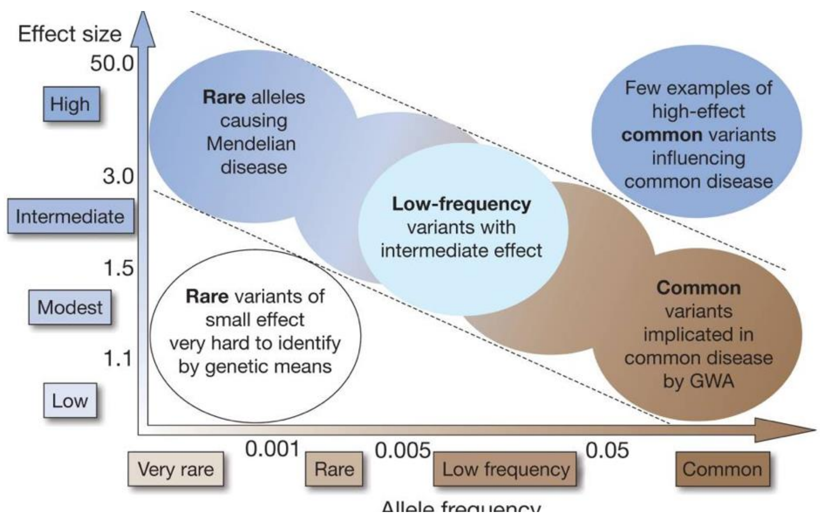


Figure 1 | Variant Frequency and Disease Associations. Rare variants (MAF<0.5%) with large effect sizes (dark blue circle) contribute to monogenic diseases while common variants (MAF>5%) with smaller effect sizes (brown circle) effect complex disease risk and are primarily identified using GWAS. Recent evidence suggests low frequency variants with moderate effect sizes (light blue circle) may also contribute to complex disease. (Adapted[2]) (mutations) consist mainly of large genomic rearrangements, copy number variants (CNVs), insertions/deletions (indels), or rare single nucleotide variations (SNVs)[4].

Complex Diseases

Unlike Mendelian diseases, complex disorders are the result of multiple genetic contributions, as well as environmental factors. These disorders are more common in the general population, and include diseases such as asthma, type I and II diabetes, depression, rheumatoid arthritis, and various autoimmune diseases. The central dogma, the ‘common disease-common variant’ model, asserts that common, complex diseases are caused by low effect, common variations (MAF > 5%) across multiple genes (Figure 1). Common variants often exhibit reduced, or incomplete penetrance, in which the presence of a disease-associated allele does not singularly determine whether the individual will develop disease. Rather, other factors, such as gene-gene interactions, epigenetic modifications, and environmental effect also contribute to disease susceptibility. In contrast to rare Mendelian mutations with high effect sizes, the majority of common disease variants only confer about a 1.1-1.5-fold increase in risk (Figure 1) and largely reside in the non-coding regions of the genome[2]. Due to the low effect and heterogenous nature of these variants, it has been more challenging to identify causal mutations in complex diseases. Consequentially, the common variants identified up to this point contribute to a small proportion of disease heritability, inviting further investigation into factors to account for this ‘missing heritability’. This thesis will center around the genomic investigation and functional studies of complex disease, Alopecia Areata (AA). To interrogate AA disease risk, I utilized two genetic approaches: 1) *hypothesis-driven* targeted sequencing of a genome-wide sequencing (GWAS) locus and 2) *unbiased* whole exome sequencing (WES) (discussed in detail in Section 1.4).

1.2 Role for rare variants in complex diseases

Several models have been proposed in an attempt to explain the missing heritability of polygenic disorders. Some authors speculate that our lack of understanding of the genetic architecture of complex diseases can be attributed to environmental factors, gene-gene interactions, chromosome structural variants, or the need for increased power of cohorts in order to detect larger numbers of small effect variants[2]. Interestingly, an increasing number of studies have reported a role for rare and low frequency variants of moderate effect size in complex diseases (Figure 1). Although rare variation (with large effects) was once thought to contribute mainly to Mendelian diseases, the contribution of rare (MAF < 0.5%) and low frequency (0.5% < MAF < 5%) variants to common disease risk is increasingly emerging. A wide variety of complex disorders such as congenital kidney malformation, amyotrophic lateral sclerosis (ALS), pulmonary fibrosis, epilepsy, coronary artery disease, myocardial infarction, and autism have all found rare variants that significantly associate with disease[5-11]. The identification of novel, low frequency variants with modest effect sizes have thus been shown to play a role in the genetic architecture of complex diseases and could feasibly explain a significant portion of the 'missing heritability'. In Chapter 3, we investigated the effect of novel, rare variants on risk susceptibility in AA.

1.3 Types of pathogenic variants

Structural Variants

Both monogenic and polygenic diseases result from DNA alterations known as mutations or variants. Disease-causing variations can affect both the non-coding (intronic, intergenic) and coding regions (exonic) of the genome. Chromosomal structural variations include insertions, deletions, duplications, translocations, and inversions that span >1 kb of the chromosomal DNA (Figure 2)[12]. Copy-number Variations (CNVs) are a subcategory of structural variants that result in genomic imbalances due to large insertions or deletions of chromosome regions[12]. CNVs span > 1 kb, often resulting in the disruption of multiple genes, and complicating the identification of a definitive disease-causing gene.

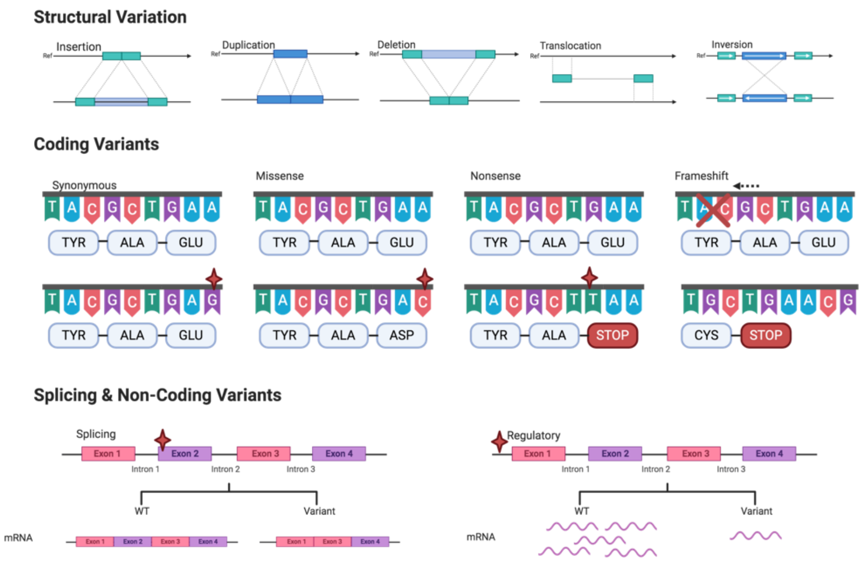


Figure 2 | Effects of genomic variants. Types of structural, coding, and non-coding variants and their corresponding consequences. Figure generated using BioRender.

Coding Single Nucleotide Variants

Single Nucleotide Variants/Polymorphisms (SNVs/SNPs) are variants that result in single nucleotide alterations, including both insertions/deletions (indels) and substitutions. SNVs that affect the coding, exonic regions of genes are classified as either synonymous or non-synonymous. Synonymous, or silent, substitutions modify the nucleotide base pair without affecting the resulting amino acid sequence (i.e. CGT, CGC, CGA, CGG = arginine) (Figure 2). These are generally considered benign, since the protein product remains unchanged.

Non-synonymous mutations are more likely to be pathogenic due to a change in the DNA sequence that results in an alteration of the protein product. Missense mutations are classified as nucleotide substitutions that change the corresponding amino acid in the protein sequence (Figure 2)[13]. Nonsense mutations incorrectly introduce a stop codon, resulting in a prematurely truncated protein[13]. Additionally, 1-2 nucleotide indels (or insertions/deletions of any number of nucleotides not divisible by 3) typically result in frameshift mutations that alter the mRNA reading frame and disrupt all downstream amino acids of the protein (Figure 2)[13].

Non-synonymous mutations can further be classified as loss-of-function (LOF) or gain-of-function (GOF). Gain-of-function mutations increase gene activity or result in a novel protein function. Conversely, loss-of-function mutations, such as nonsense and frameshift mutations, reduce the presence, function, or activity of the gene product. Heterozygotes for loss-of-function alleles (carriers) are often asymptomatic due to the ability of the wildtype allele to compensate for the reduced function of the mutated allele[14]. However, haploinsufficiency results when one LOF allele causes a reduction of the normal gene product to about 50%, which is not sufficient for adequate protein function. In comparison to LOF mutations, predicting the effects of missense variants on protein function is not as straightforward. As a result, the PolyPhen tool was developed to predict damaging effects of mutations using both sequence and structure-based predictive models[15]. PolyPhen assigns coding variants a score of likelihood to be damaging and allows prioritization of the predicted damaging mutations for further functional analyses. The effect of LOF and damaging coding mutations in AA will be discussed further in Chapter 3.

Splicing Mutations

Exon-intron boundaries are defined by splice sites that direct the splicing machinery to form a mature mRNA transcript consisting of only coding sequences. Most splice sites contain 5'-GT and 3'-AG motifs at the intron boundaries that are recognized by the splicing machinery to bind, cut, remove the intron, and ligate the ends of exons[16,17]. SNVs occurring in these tightly regulated exon/intron junctions can result in improper intron removal and aberrant transcript formation (Figure 2). Additionally, splicing mutations can occur at other coding and non-coding locations within the gene that result in novel splice sites or splicing regulatory disruption, impacting mRNA transcripts and protein translation. In an effort to identify possible causal splicing mutations associated with disease risk, *in silico* algorithms, such as Transcript-inferred Pathogenicity (TraP), incorporate predicted splicing effect disruption into their variant interpretation tools[18]. TraP provides pathogenic prediction scores for synonymous and intronic SNVs based on gene information acquisition, splice sites/regulatory features, and modeling[18]. As a result, TraP effectively predicts the probability of a synonymous or intronic variant having a damaging effect on transcript formation due to interference with the regulatory and/or splicing machinery. In Chapter 3, we incorporate TraP prediction scores to identify potential disease-causing variants in AA.

Non-coding Mutations

Pathogenic variants can also reside in non-coding regions; however, their disease-causing effects are not as well characterized (Figure 2 and 3). Non-coding DNA accounts for 98% of the human genome and harbors many important gene regulatory elements such as promoters, enhancers, and structural elements[19]. Although not fully understood, non-coding variants can exert regulatory effects on downstream gene products through both proximal and long-range interactions on surrounding genes. Interestingly, approximately 90% of susceptibility and causal variants in autoimmune diseases have been mapped to non-coding regions[20,21]. The emerging realization of the contribution of non-coding regions to disease risk underscores the need for further understanding of intronic and intergenic regions. Large consortia such as ENCODE (Encyclopedia of DNA Elements Consortium) and Roadmap Epigenomics have propelled the field forward through the compilation of datasets annotating the functional landscape of non-coding regions into publicly available databases[22,23]. By integrating a wide range of histone modifications (acetylation and methylation), DNA binding sites (ChIP-seq), and chromatin accessibility (DNase I hypersensitive sites), these databases provide an invaluable resource for understanding the non-coding architecture of the genome and identifying the location of potential regulatory elements.

Utilizing these datasets, *in silico* algorithms, such as RegulomeDB and FUN-LDA (Functional-Latent Dirichlet Allocation), were designed to predict the probability of variant regulatory effects at individual positions throughout the genome. Our collaborator, Dr. Iuliana Ionita-Laza at Columbia University, developed the FUN-LDA algorithm, which integrates histone modifications and chromatin accessibility from the ENCODE database to predict the likelihood of a regulatory effect at each position of the genome in 127 distinct cell and tissue types[24]. FUN-LDA enables functional annotation of any non-coding variant in the genome, and also provides context-dependent regulatory effects based on relevant cell and tissue types. These data help to distinguish which cell types are relevant to disease and are affected by the variant in question. Such large regulatory databases and *in silico* models have provided the field with invaluable new tools to gain insight into the roles of non-coding variants in disease pathogenesis.

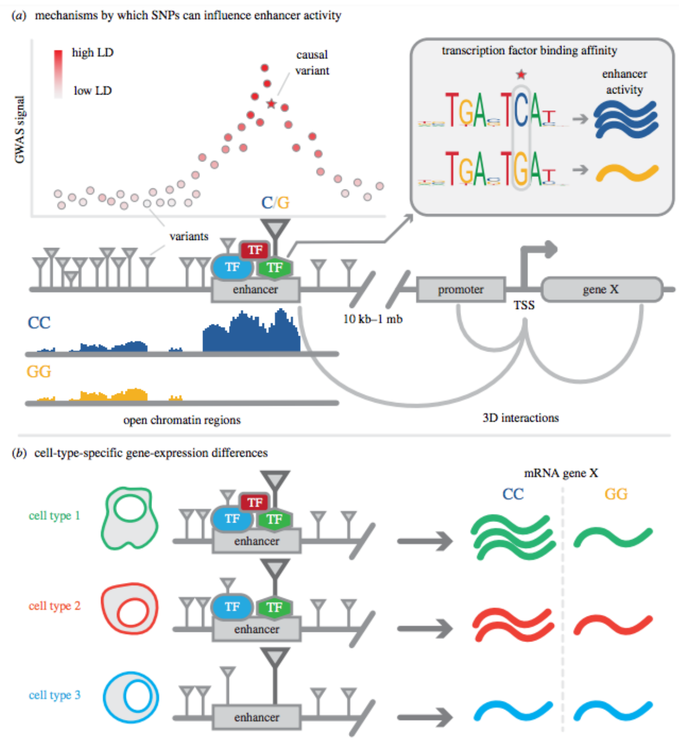


Figure 3 | Effect of non-coding variants on cell type-specific expression. Non-coding variants can fall in regulatory regions (enhancer, promoters, etc.) that have effects on gene expression in a cell-type specific manner. When a variant genotype (GG) correlates with altered expression levels of the mRNA, the variant is annotated as an eQTL for that cell type. (Adapted[25])

eQTL effects of non-coding variants

One of the most fundamental effects of causal non-coding variants is the dysregulation of gene expression. Variants falling in promoters, enhancers, transcription factor binding sites, or DNA modification regions all have the potential to disrupt transcription and alter gene expression levels (Figure 3). In order to identify non-coding variants with direct effects on gene expression, eQTL (expression quantitative trait loci) analyses are utilized[26]. This approach uses multi-dimensional patient data to integrate the presence of variant alleles with changes in gene expression levels in a targeted tissue type. eQTLs establish a direct correlation between an allele and corresponding expression levels in a disease-relevant tissue[26]. Identification of eQTLs allows pinpointing of non-coding variants that have a functional effect on the gene product, and thus have a greater likelihood of being causal in disease. Through the use of *in silico* annotations and multidimensional data, our understanding of the role of non-coding variants in complex diseases has made substantial progress. In Chapter 2, we integrate the use of *in silico* algorithms and sequence-expression data to identify potential non-coding causal variants in an AA risk locus.

1.4 Genetic approaches for identifying pathogenic variants

Historically, the genetic approaches used to identify disease-causing variants have been dependent on the type of disease inheritance. Rare, Mendelian disorders are primarily mapped using family-based linkage analysis, while complex diseases routinely utilize case-control genome-wide association studies (GWAS). However, the rapid advancements in next-generation sequencing (NGS) technology have ushered in new sequence-based approaches to identify novel variants in complex diseases. This dissertation focuses on variant identification and functional annotation in AA; therefore, we will focus on the genetic approaches available for mapping genes in common, complex diseases.

Genetic Approaches for Mapping Disease Genes

The two main approaches for gene identification in genetic diseases have been *linkage* analysis in *rare* diseases and *GWAS* in *common* disorders[27,28].

Linkage

Linkage analysis is based on the principal of co-segregation, in which a genetic region co-segregates with the disease phenotype in *a pedigree of multiple affected individuals*. Regions of the genome can be inherited together based upon recombination events, in which proximal genomic locations are inherited together with higher frequency than distant or intrachromosomal locations due to a reduced probability of chromosomal crossover events between the proximal loci[29]. Linkage disequilibrium (LD) occurs when proximal loci are disproportionately inherited together across a population, forming haplotypes, or groups of co-occurring alleles on a single chromosome[29,30]. LD is measured by r^2 and D’ metrics that take into account the statistical comparison of the frequency of an observed haplotype versus its expected frequency in a given population[31]. r^2 and D' values that approach 1 correlate with increased LD between two genomic positions.

Linkage analysis utilizes the principles of co-segregation and recombination by incorporating markers spaced across the genome to identify, in an unbiased manner, chromosomal regions that segregate with a disease or trait[32]. The co-segregating marker tags a genetic allele that is in linkage with disease, identifying nearby regions of the genome that may harbor a causal disease gene. The large candidate region can then be further investigated for the causative disease gene driving the co-segregation. Gene detection using this method is most effective for single genes of complete penetrance[32], therefore, linkage analysis has been very successful in identifying Mendelian disease genes. Linkage has also been applied with moderate success to the genetic mapping of complex disorders, including Type I and II diabetes, Crohn's Disease, Alzheimer's Disease, psoriasis, atopic dermatitis, and AA in small cohorts of families with multiple affected members[33-35]. Although linkage studies have been moderately successful in common diseases, the more fruitful approach is GWAS, a population-based association study comparing cases and controls.

GWAS

The principles of linkage and LD are also fundamental to understanding population-based GWAS association studies[30]. Association studies employ various markers dispersed across the genome to identify regions of association with a trait. The LD architecture of the human genome allows association studies to take advantage of marker genotyping without requiring sequencing of every genomic base pair[36]. Since co-occurring alleles in LD blocks are inherited together on a haplotype, the markers can successfully identify regions and their surrounding LD blocks that are associated with disease status. GWAS are population-based association studies that involve genotyping thousands of individuals at 500,000-2.5million SNPs (tag-SNPs) across the genome[37]. The tag-SNPs involved in GWAS platforms are common (MAF$\geq$ 5%) in the general population due to the basis of the ‘common variant, common disease’ dogma[2,36]. By genotyping individuals with the trait of interest at these pre-determined tag-SNPs and comparing them to unaffected controls, we can effectively identify haplotypes that are enriched at increased frequency in the trait/disease cohort. Since these SNPs are common and predicted to be of low effect size, large cohorts of patients with a given disease are required in order to identify significant associations. GWAS studies have successfully identified >40,000 SNPs significantly associated with various complex diseases and traits, conveniently compiled in a publicly-available database, GWAS Catalogue[38].

Although GWAS is an effective tool for identifying low effect, common disease variants, there are two main limitations with this approach. First, GWAS studies utilize SNP arrays that contain 500K-2.5M common SNPs, which enable identification of associated regions, but are unable to provide the base-pair (bp) level resolution necessary for distinguishing the association-driving variant(s) from the non-functional variants on the same haplotype. The tag-SNP effectively points to a large LD block associated with disease; however, it is not possible to determine whether that SNP itself is causal, or whether a proximal variant(s) in high LD with the mutation is driving the association. The accurate identification of the causal variant(s) is crucial for understanding the effect of the variant, the pathomechanism of disease, and relevant targets for potential therapeutics and interventions. As a result, deep sequencing and fine mapping of GWAS associated regions used to distinguish the surrounding genetic architecture with higher resolution in order to identify the precise causal variant(s) driving disease association at GWAS loci.

The second issue inherent to GWAS relates to the ‘missing heritability’ paradigm of complex diseases, in which GWAS findings alone are unable to account for a large portion of genetic susceptibility[2]. Due to the basis of ‘common variant, common disease’ and the use of common tag-SNPs on GWAS SNP chips, rare variants are not genotyped. Consequentially, rare disease variants are seldom identified using this method[36]. However, as discussed in section 1.2, there is increasing evidence for the role of rare variants in complex disease risk[39-43]. Since GWAS is an inadequate method for effectively identifying rare variant associations, the recent advancements in NGS have enabled the investigation of rare variants in complex disease risk (discussed in the next section). In this thesis, we addressed both of these fundamental GWAS limitations through (1) targeted sequencing of a GWAS association region to elucidate the causal variant(s) driving association in Chapter 2; and (2) investigation of the role of rare variants in the genetic architecture of AA using WES in Chapter 3.

NGS and Complex Diseases

Due to the high cost and low throughput nature of early sequencing methods, sequencing studies were primarily limited to small candidate gene regions identified by linkage analysis in large multiplex pedigrees with monogenic diseases[44]. In the mid-2000s NGS was developed as a high throughput sequencing approach that drastically lowered the cost of sequencing[45]. However, at $10,000 USD per genome, NGS approaches were still too costly to sequence large sample numbers with the power required to detect associations in complex disease studies[44]. As a result, a more cost-effective method of targeted sequencing in GWAS susceptibility regions was developed to follow-up these genetic studies and identify the functional, causal variant(s) in the associated regions of interest.

The development of targeted exon enrichment enabled the emergence of whole exome sequencing (WES), which was quickly and successfully applied to monogenic diseases on the basis that the majority of monogenic disease variants are rare coding variants with high effect sizes[46,47]. WES was extremely successful in Mendelian diseases and contributed significantly to our understanding of Mendelian disease genes. The ability of WES to effectively identify rare disease variants made it an attractive approach to address the role of rare variants in the missing heritability paradigm of complex diseases[48]. As NGS and WES technology continued to advance, cost continued to decline, and WES became feasible for large-scale studies of complex diseases. However, the successful detection of rare variants was dependent once again on the need for large sample sizes, due to the lack of power to detect variant level associations[44,48,49]. As a variant's frequency decreases, the ability to detect significant associations also decreases because of the sparse observations of the rare allele in a given population[1,44]. Subsequently, the lack of rare allele observations inhibits the ability to achieve statistically significant associations in moderate cohort sizes. To overcome the inability to identify variant level associations, burden tests were applied to WES data as a method for aggregating the variants within a given gene and comparing the frequency of genetic variants in cases and controls (Figure 4)[44,50]. In a gene-collapsing burden test, variants are annotated as "qualifying variants" based on specified criteria, such as allele frequency in the general population and variant effect[50]. Variants that meet the parameters are identified as 'qualifying' and are then combined together based on gene boundaries, thereby determining whether a gene harbors more *total rare* damaging variants in cases compared to controls (Figure 4). WES and burden tests have been successfully applied to many common diseases, including ALS, Alzheimer's disease, myocardial infarction, and schizophrenia[6,10,51,52]. In summary, the significant advances in NGS technology have allowed adaptation of these methods for complex disease studies in both *hypothesis-driven* (i.e. targeted resequencing of GWAS regions) and *unbiased* (i.e. WES, WGS) approaches.

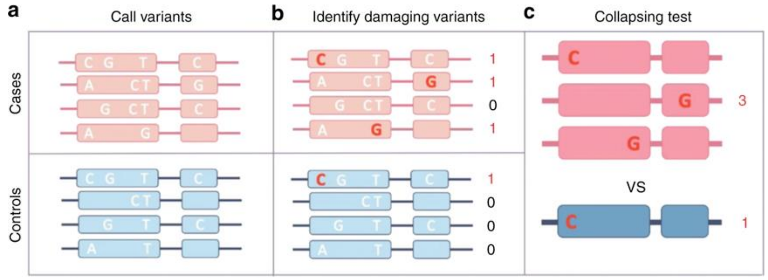


Figure 4 | Gene-collapsing burden analysis. The number of qualifying variants that meet study criteria (i.e. MAF<1% and damaging effect) in cases (pink) compared to controls (blue) in a given gene. (Adapted[1])

I utilized both of these NGS approaches in AA to (1) resolve variant annotations in non-coding regions of a previously identified AA risk locus (hypothesis-drive approach); and (2) discover rare variation in genes with novel disease association (unbiased approach).

1.5 Alopecia Areata

Clinical Characteristics and Pathology

Alopecia areata (AA) is one of the most prevalent autoimmune diseases, with a lifetime risk of 2.1%[55]. In AA, transient non-scarring hair loss is a result of immune-directed attack of the hair follicle (HF). The disease can affect any hair-bearing site and manifests in three main presentations: Alopecia Areata Patchy (AAP), Alopecia Areata Totalis (AT), and Alopecia Areata Universalis (AU). AAP is the most common form of disease and results in well-demarcated regions of hair loss on the scalp (Figure 5). In a subset of cases, disease progresses resulting in hair loss of the entire scalp (AT) or body (AU). Other rarer forms of disease include diffuse alopecia areata, which results in widespread hair thinning across the entire scalp, and ophiasis, a thick band of hair loss on the occipital scalp hair line. Pathological features of AA include yellow and black dots on the scalp, and exclamation mark hairs (EMH). EMHs are often found within lesional areas of active disease, and are distinguished as fractured, short hairs with frayed distal ends that taper towards the scalp. The presence of cuticle defects in these vulnerable/weak hairs may suggest a potential role for disruption of hair follicle integrity in AA[56].

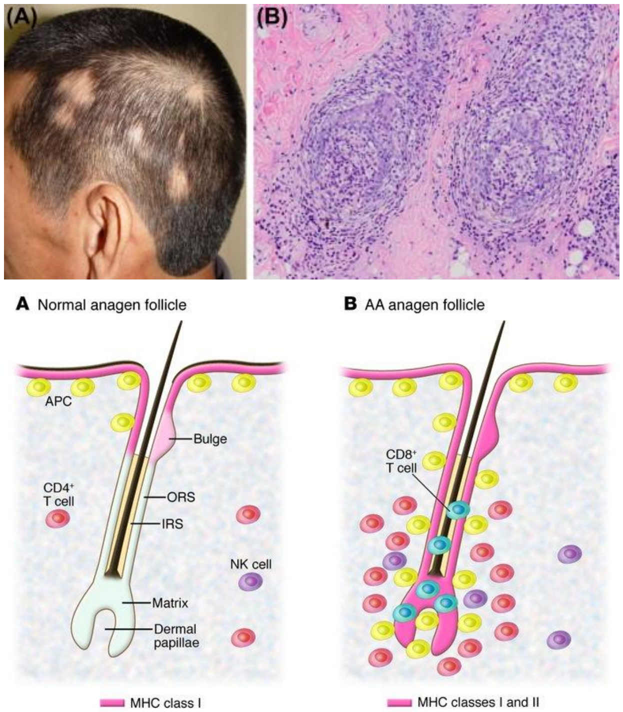


Figure 5 | Pathology of AA. Patient with AAP disease manifestation. Immunohistochemistry stain of "swarm of bees" phenotype characterized by lymphocytic infiltrate of AA hair follicles. Graphic of healthy hair follicle compared to an AA-affected hair follicle which upregulates MHC class I and II expression and is infiltrated by auto-reactive lymphocytes (APCs, NK cells, CD4+ and CD8+ T cells). APC= antigen presenting cell, NK cell= natural killer cell, MHC= major histocompatibility complex, IRS= inner root sheath, ORS= outer root sheath. (Adapted[53,54])

AA is a complex genetic disease with an incidence of positive family history reported in up to 42% of cases[57]. Onset can occur at any age, however, an early age of onset (5-10 years old) tends to correlate with disease severity(i.e. Alopecia Areata Totalis/Universalis (ATAU))[58]. Additionally, AA affects both genders and all ethnic groups equally. Common comorbidities with other disorders include atopic dermatitis (39%), autoimmune thyroid disease (8-28%), vitiligo (4%), and lupus erythematosus (0.6%)[59].

A positive diagnosis of AA is established by: 1) Clinical evaluation (hair loss pattern; AAP, AT, AU); 2) Dermoscopy (yellow/black dots and EMHs); 3) family history and comorbidities; and 4) histologic evaluation. The pathognomonic histopathology of AA is inflammatory cell infiltrate around the base of the hair follicle (Figure 5). The high density of lymphocytic cells in the infiltrate is composed largely of CD8+ T cells that exhibit a "swarm of bees" phenotype in affected hair follicles (Figure 5). This inflammatory infiltrate is responsible for the hair-follicle directed autoimmune attack and resulting hair loss.

Another measure of disease diagnosis and progression was developed in our lab using molecular gene expression signatures of skin biopsies. The Alopecia Areata Disease severity Index (ALADIN) was developed based on gene expression signatures of inflammatory genes and correlates with disease severity. AA skin samples have increased cytotoxic T lymphocytes (CTL) and interferon (IFN) gene expression signatures and a decreased hair keratin (KRT) index that are significantly distinguishable from controls[60]. The ALADIN metric is dependent on expression of the following signatures: CTL (*PRF1, CD8A, GZMB, ICOS*), IFN (*STAT1, MX1, CXCL9, CXCL10, CXCL11*), and KRT (*DSG4, PKP1, KRT31, KRT32, KRT33B, HOXC13, KRT82*). These groups of genes represent critical molecular signatures that are prominently dysregulated in AA skin, providing crucial insight into the underlying pathomechanisms of AA. Additionally, the characteristic AA immune signatures (CTL and IFN) support the autoimmune-driven pathology of hair loss in this disease.

Autoimmune-directed attack of the hair follicle

Autoimmune diseases are a result of modifications in the immune system that result in autoreactive lymphocytes, loss of self-tolerance, and subsequent attack of bodily organs and tissues[61,62]. Autoreactive lymphocytes (B and T cells) aberrantly recognize self-proteins as autoantigens, stimulating an inflammatory response and resulting in systemic or organ-specific attack. Common dysfunctional biological pathways underlie autoimmunity as evidenced by common genetic associations shared among autoimmune diseases[62]. However, differences in phenotypic manifestations between diseases are largely dependent on the target organ. For example, multiple sclerosis attacks the myelin sheath, rheumatoid arthritis targets synovial tissue in joints, and type I diabetes affects beta cells in the pancreas[61,63]. In AA, the target organ is the hair follicle. However, the precise autoantigens within the HF that are responsible for immune activation and disease precipitation remain unknown.

Sudden Whitening of the Hair

An interesting clinical phenomenon is observed in a subset of AA patients, known as *canities subita*, or Marie Antionette Syndrome. Historical accounts of Marie Antionette on the eve of her execution in 1793 reported 'sudden whitening' of her hair. Many clinicians attribute this striking observation to the acute onset of AA, in which it is postulated that AA immune cells preferentially attack the pigmented HFs and spare the white hairs, giving the appearance of 'sudden whitening'. Reports of AAP patients with residual white hairs in lesional areas of hair loss supports this hypothesis. Additionally, many AA patients who begin to recover initially regrow white hairs, before the hair shaft fully re-pigments. As a result, these observations have fueled the long-standing hypothesis that the pigment-producing cells, melanocytes, are the HF-specific targets in AA autoimmune attack. In support of this theory, histopathology of AA skin biopsies reveals dysmorphic hair bulb melanocytes (Figure 6, discussed in more detail in section 1.6)[64]. The

Figure 6 | Aberrant AA melanocytes. Melanocyte cell death (marked by PM-degenerating pigment cell) in hair follicle of AA patient. Healthy melanocytes are denoted with NM (normal healthy melanocytes). HF structural components are annotated as follows: MX=matrix, DP=dermal papilla, BL=basal lamina. (Adapted[64]) preferential attack of pigmented hair follicles in AA is an unanswered observation in the field, and the mechanism underlying this phenomenon has yet to be explained. I will address this question in Chapter 2.

Genetic Architecture of AA

As a complex disease, both environmental and genetic factors contribute to the risk of developing AA. Environmental triggers such as trauma, infections, the gut microbiome, and emotional stress are all correlated with disease onset and progression[55,65,66]. In addition to possible environmental triggers, AA also has a strong genetic component, as evidenced by 1) familial aggregation of increased disease incidence; 2) twin concordance rates of 42-55%; 3) heritability with a 10-42% incidence of AA between relatives; and 4) examples of early onset cases of AA in childhood, all of which strongly indicate a heritable component of disease. Individuals with family members diagnosed with AA are at an increased risk of developing AA themselves compared to the general population. A study of 206 patients and 3654 relatives concluded that lifetime risk of AA raises to 7.8% in first-degree relatives, compared to 2.1% in the general population[67]. Studies that investigated the occurrence of disease in monozygotic and dizygotic twins also provide insight into the heritability of AA. Due to the identical genetic makeup of monozygotic twins, a trait that is more prevalent in monozygotic compared to dizygotic twins supports a genetic contribution to disease. The first twin study in AA reported a concordance rate of 55% in monozygotic twins versus 0% in dizygotic twins[68]. The same group subsequently performed a larger twin study and found concordance rates of 42% in monozygotic twins versus 10% in dizygotic twins[69]. Due to the finding that monozygotic concordance rates were less than 100%, it is postulated that the environmental factors influence a portion of disease risk and supports a role for non-genetic components in AA development. Nonetheless, the 42-55% monozygotic concordance rate supports a strong genetic component in AA.

A linkage analysis study from our lab confirmed genetic etiology of disease in a genome-wide search for susceptibility loci in 20 AA-affected families[34]. This study found four susceptibility loci on Chr. 6, 10, and 16, with the strongest linkage on Chr. 18. To further understand the genetic architecture of AA, our lab initially performed a GWAS and meta-analysis in 3,253 AA cases and 7,543 controls. These genetic studies successfully yielded common SNP associations in at least 14 risk loci in AA[70,71]. As seen in other autoimmune diseases, most risk loci were shared among other autoimmune diseases (*HLA*, *MICA, IL21/2*, *IL2RA*, *CTLA4*, *PRDX5*, *IKZF4, LRRC32, SH2B3/ATXN2, ACOXL/BIM, CIITA/CLEC16A/SOCS1, IL13/4*). However, two associated regions (*STX17*, *ULBP3/6*) harbored genes that were specific to AA and had not been associated with other autoimmune diseases [70,71]. Interestingly, both genes were also expressed in the target organ, the HF[70]. Genes expressed in the HF and specific to AA pathogenesis were prioritized as candidates for investigating the mechanisms of organ-specific attack in AA. Genes implicated in HF biology offered potential insight for understanding the elusive AA autoantigen(s) and its role in target organ mediated immune attack. As discussed in Section 1.4, GWAS lack the resolution to identify causal functional variants in risk loci, therefore, deep sequencing studies and experimental assays are necessary for determining the effects of disease-associated variants in GWAS-identified loci.

In Chapter 2, I used targeted sequencing and functional experiments to elucidate the function of *STX17* in AA pathogenesis.

Genetic and pathologic implication of STX17

Unlike many of the other AA-associated genes discovered in the GWAS, *STX17* had not been previously implicated in any other autoimmune diseases, suggesting that its disruption may play a specific role in the AA target organ, the hair follicle. *Syntaxin17* (*STX17*) was further prioritized for functional analysis due to its expression in the HF and previous implication in hair color variation. Although not shared among other autoimmune diseases, genetic variation in

STX17 was previously associated with differences in hair pigmentation and color[72,73]. A GWAS of 27 human traits identified a SNP (rs9556) in the 3' UTR of *STX17* that was significantly associated with human hair color[72]. Moreover, a duplication in intron 6 of *STX17* was identified as a causal variant for a premature hair graying phenotype in Lipizzaner horses[73]. In this breed, horses are born with fully pigmented coats and completely lose hair pigmentation and become gray before 8 years old. Additionally, the gray horses have significantly higher incidences of melanomas. The causal mutation was mapped to a CNV resulting in a 4.6 kb duplication that harbors a melanocyte-specific transcription factor, MITF[73,74]. Melanomas biopsied from gray horses expressed higher levels of *STX17* and neighboring gene, *NR4A3*. Interestingly, the AA associated-SNP identified by our GWAS fell within the same intron as the duplication with reported melanocyte-regulatory effects in horses. The association of *STX17* with hair color variation was intriguing because of the clinical correlations between AA and the preferential attack of pigmented hair follicles. We postulated that variation in *STX17* may disrupt hair pigmentation biology and result in the aberrant inflammatory attack in AA-susceptible individuals.

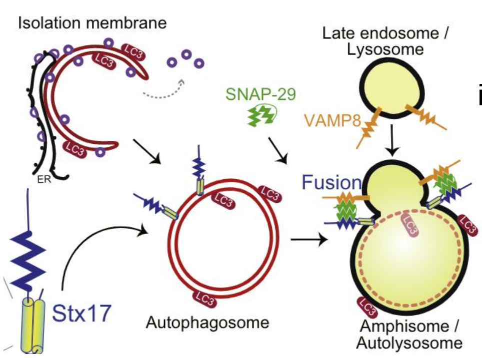


Figure 7 | Role of STX17 in autophagy. The hairpin SNARE protein, STX17, contains two transmembrane domains (TMDs) that form a hairpin structure and aids in localization of STX17 to mature autophagosomes to facilitate fusion with lysosomes through complexes with SNAP-29 and VAMP8. LC3 (microtubule-associated protein 1A/1B-light chain 3) marks isolation membranes and autophagosomes. (Adapted[75])

Although previous studies implicated *STX17* in hair color determination, the functional effect of *STX17* on pigmentation was not further explored. *STX17* encodes the STX17 protein that was originally characterized by its role in the autophagy pathway. Autophagy is a critical cellular process that regulates degradation of cellular contents and aggregated proteins through the fusion of autophagosomes with lysosomes[75]. STX17 is a member of the soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) protein family that facilitate membrane fusions through the complex of Qa-, Qb-, Qc-, and R-SNARE proteins that form helix bundles. STX17 is a Qa-

SNARE protein that traffics to mature autophagosomes to facilitate autophagosomal-lysosomal fusion with SNAP-29 (Qbc-SNARE) and VAMP8 (R-SNARE) (Figure 7)[75]. The majority of SNARE proteins are single transmembrane domain (TMD) tail-anchored proteins that are initially localized to the ER before trafficking to their target organelle membranes. However, SXT17 is unique in that it contains 2 TMDs that form a hairpin structure and are critical for selective targeting to the autophagosome membrane (Figure 7)[75]. This allows specific trafficking to mature autophagosomes and accurate regulation of autophagosome-lysosome fusion in the autophagy pathway. Since this initial finding, additional biological roles for STX17 in autophagosome formation[76], mitophagy[77] and mitochondrial fission[78] have been reported. Although a functional role for STX17 in pigmentation has not yet been investigated, multiple studies have reported a central role for autophagy in melanocyte (pigment cell) biology. Taken together, these findings suggest that multiple aspects of the autophagy and pigment production pathways overlap, and that disruption of autophagy function can perturb melanocyte function and pigmentation[79-83]. In Chapter 2, I investigated the functional role of *STX17* in pigment production and melanocyte biology.

1.6 Hair Follicle Biology

As the target organ of AA immune attack, hair follicle biology is a primary focus of this thesis. The HF is a dynamic mini organ whose intricate and accessible nature make it a unique model to study. In this thesis, I focused on how dysfunction in the HF itself primes the organ for attack, resulting in subsequent HF destruction and hair loss. The next section outlines normal HF structure and biology to understand how genetic disruption of this organ can result in autoimmunity.

Hair Follicle Structure

The hair follicle is an intricate mini-organ that produces mammalian hair, which possesses a wide range of evolutionarily conserved purposes including thermoregulation, tactile sensing, protection, and social impact[84]. During fetal skin development, epithelial, neural crest, and mesenchymal stem cells of the neuroectodermal-mesodermal system produce the developing HF, which is continually re-generated throughout its lifetime[85]. Each stem cell source generates different compartments of the HF: epithelial stem cells produce sebocytes of the sebaseous glands (SG) and keratinocytes that form the epithelial components of the HF including the matrix (MX), outer root sheath (ORS) and inner root sheath (IRS) (Figure 8). Mesenchymal stem cells give rise to fibroblasts which comprise the dermal papilla (DP) and connective tissue sheath (CTS) (Figure 8). The neural crest cells produce the melanocytes of the hair follicle pigmentary unit (HFPU) which resides atop the basement membrane of the matrix, directly above the DP[84,85].

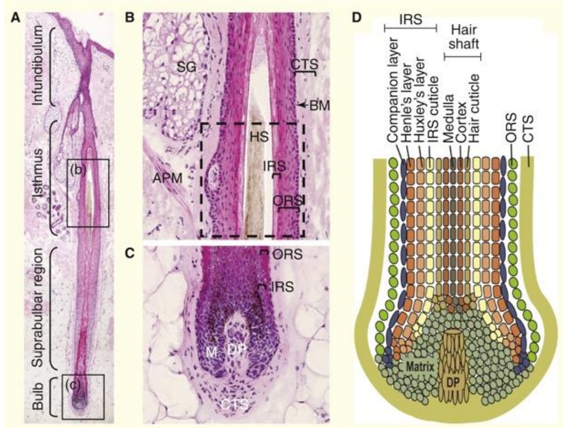


Figure 8 | Hair Follicle Structure. The HF is a mini-organ composed of many different compartments, regions, and cells types. The bulb region, suprabulbar, isthmus, and infundibulum span the HF from the base to the opening at the epidermis. Concentrically from inner to outer the layers of the hair follicle are the hair shaft (medulla, cortex, cuticle), IRS (IRS cuticle, Huxley's layer, Henle's layer, companion layer), ORS, CTS. The DP and matrix reside in the bulb. Dashed box indicates the approximate location of the HF bulge. (Adapted[84])

Beginnning at the base of the HF, the bulb region contains the DP, matrix, and HFPU (Figure 8). The DP is made up of specialized fibroblasts that regulate the proliferation and differentiation of the surrounding matrix cells. A basement membrane separates the DP from the matrix[86]. The matrix is composed of actively proliferating keratinocytes that differentiate to form the terminal keratinocyte lineages of the IRS and hair shaft (Figure 8). The HFPU also resides in the bulb, above the basement membrane and is composed of melanocytes, which are responsible for producing melanin pigment that is deposited into the surrounding matrix keratinocytes as they proliferate and differentiate up the hair shaft. Across the widest part of the bulb spans the Line of Auber, a critical separation between proliferating, undifferentiated keratinocytes below the line and diferentiated terminal keratinocytes above the line[87].

The differentiated keratinocytes of the matrix give rise to the concentric layers that constitute the hair shaft and IRS. Each of the individual layers are differentiated by their unique expression of different keratin proteins (discussed in more detail in section 1.5.2)[84]. The hair shaft is made up of 3 layers that constitute the external hair fiber that emerges from the scalp. It is composed of highly keratinized terminally differentiated cells that are responsible for its enormous tensile strength. The medulla and the cortex make up the inner layers of the shaft, which are surrounded by the cuticle, the outermost shaft layer which protects the inner layers from environmental injuries and anchors the growing shaft to the HF unit (Figure 8)[84]. The cuticle possesses a scaled texture that allows interlocking with the IRS cuticle during growth. The IRS is made up of 4 consecutive layers consisting of the IRS cuticle, Huxley's layer, Henle's layer, and the companion layer. The companion layer loosely binds to the stationary ORS, allowing the IRS and hair shaft to grow upward towards the skin epidermis[84]. The ORS is the outermost epithelial layer that harbors the bulge, a reservoir of epithelial HF stem cells (Figure 8, dashed box). Finally, the connective tissue sheath envelops the entire epithelial unit of the HF.

These layers extend concentrically from the bulb region, up the HF towards the skin epidermis (Figure 8). Ascending the hair shaft, the HF structure consists of the bulb, the suprabulbar region, the isthmus, and the infundibulum (Figure 8). The upper part of the hair follicle (the infundibulum and isthmus) is static and non-cycling while the suprabulbar and bulb region are constantly regenerating during the hair cycle (discussed in more detail in section 1.6.4)[84]. The isthmus contains the bulge which houses HF stem cells responsible for providing the HF with its regenerating capacity (Figure 8, dashed box). Damage or loss of the bulge results in irreversible hair loss due to the inability of the HF to produce the necessary epithelial keratinocyte and melanocyte stem cells required for generating a new follicle[88]. The melanocyte and keratinocyte cell properties will be discussed in further detail in the next sections.

Keratinocyte Cell Biology

Keratinocytes are epithelial cells that make up approximately 95% of the skin epidermis and are characterized by their expression of keratin proteins during differentiation[90,91]. Keratins are intermediate filaments (IF) that provide structure and stability to the cells through heterodimerization and bundling[89]. Keratin heterodimers from tetramers that are surrounded by a network of keratin associated proteins, which interact with the keratin tetramers through disulfide bridges[92]. Heterodimers are composed of one type I and one type II keratin. Type I keratins (*K9-K28*, *K31-K40*) are smaller in size (40-56.5 kDa) and acidic in charge. Type II keratins (*K1-K8*, *K71-K86*) range from 53-67 kDa and are neutral to basic in charge[89]. Interestingly, a subset of keratinocytes produces hair specific keratins that are exclusively expressed in the HF and the nail matrix/bed. The hair specific keratins are further sub-divided into "soft" keratins (type I *K25-K28*, type II *K71-74*) expressed in the epithelial root sheaths, and "hard" keratins (type I *K31-40*; type II *K81-86*) in the hair shaft and nails[89]. The hard keratins possess a higher sulfur content, allowing for increased cross-linking capabilities[89]. The increased disulfide bonds and cross-linking strengthen the stability and structure of hard keratin interactions, which are responsible for the tensile strength of the hair fiber[92,93]. Epithelial and hair keratins exert highly specific expression patterns within the HF, thereby defining the different layers and compartments (Figure 9).

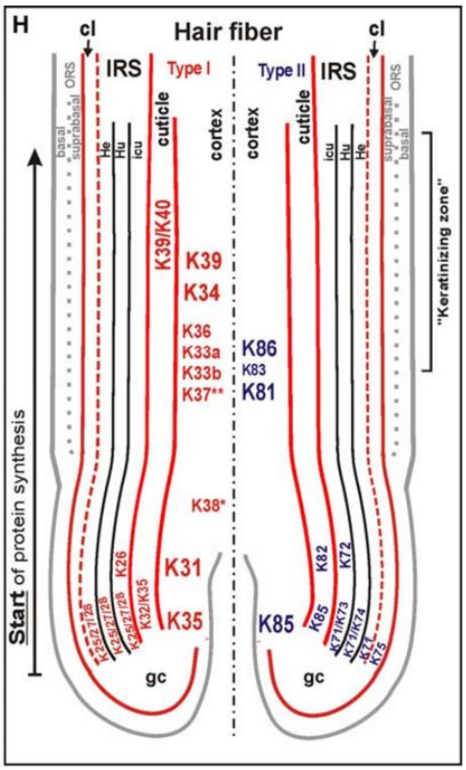


Figure 9 | Keratin expression in the HF. Both type I and type II keratins are expressed in each layer of the HF to form strong IF bundles through cross-linking and heterodimerization with one another. Type I in red, type II in blue (Adapted[89])

The hair specific keratins are intriguing antigenic candidates for AA immune attack since they could provide a potential explanation for the hair-targeted attack that is characteristic of disease. Approximately 25% of AA patients present with a “pitting” nail abnormality as a symptom, which is more common in severe forms of disease[94]. Since hair keratins are expressed in both the nail matrix and HF, a hair specific keratin autoantigen would explain the affected nail and hair sites that are observed in disease. In support of this, previous work found that AA antibodies recognized HF keratins, and specifically that AA patient serum reacted with human HF proteins of size 44/46 kDa and 60 kDa[95]. Furthermore, 44/46 kDa keratin proteins were immunoprecipitated from murine AA serum but not control. Interesting to note, 60-kDa keratins were present in both AA and control serum[95]. These findings suggest that a 44/46 kDa hair keratin(s), whose size is characteristic of type I keratins, possess antigenic qualities in AA. Although follow up work has not elucidated the precise antigen(s), type I hair keratins are attractive candidate proteins.

Melanocyte Biology

Melanocytes are neural crest-derived cells responsible for pigmenting the skin and hair. They are found in the basal layer of the epidermis as well as the HF, and possess dendrite-like projections that interact with surrounding keratinocytes and deposit pigment[96,97]. Epidermal melanocytes are smaller than HF melanocytes, with shorter dendrites and smaller melanosomes, and one melanocyte pigments approximately 30-40 surrounding keratinocytes[96-98]. The HF melanocytes are distinct from skin melanocytes in that they 1) regenerate with each new hair follicle; 2) pigment approximately 5 keratinocytes; and 3) are larger in size with larger, more pigmented melanosomes[98]. Despite these differences, the current dogma maintains that the biochemical characteristics of the melanin production pathway are shared between skin and HF melanocytes.

Melanin is a chemo- and photoprotective pigment that supports cellular health through UV light absorption and free radical scavenging[96,98,100]. However, the production of melanin itself results in toxic oxidative intermediates that must be tightly regulated in specialized organelles called melanosomes[100]. Melanosomes are lysosome-like organelles that undergo four stages of maturation during melanogenesis, the process of melanin production. First, vacuolar early endosomes form non-pigmented Stage 1 melanosomes (Figure 10). PMEL17 (premelanosome protein; also known as gp100/SILV) is a critical structural transmembrane protein that is sorted to Stage 1 and subsequently cleaved and released from the membrane into the melanosomal lumen[100,101]. Melanosome protein, MART1 (Melanoma antigen recognized by T cells 1) complexes with PMEL17 in the ER/early Golgi to regulate PMEL17 trafficking, stability, and processing[102]. After PMEL17 cleavage and processing, PMEL17 forms amyloid fibrils that striate the melanosome, a hallmark characteristic of stage 2 (Figure 10). In the Stage 2-to-3 transition, the melanin biosynthetic enzymes Tyrosinase (TYR), tyrosinase-related protein 1 (TYRP1), and tyrosinase-related protein 2 (TYRP2) are trafficked to the melanosome. TYR is the rate-limiting enzyme in melanin production that catalyzes the reaction to produce melanin pigment[99,100]. In stage 3, melanin

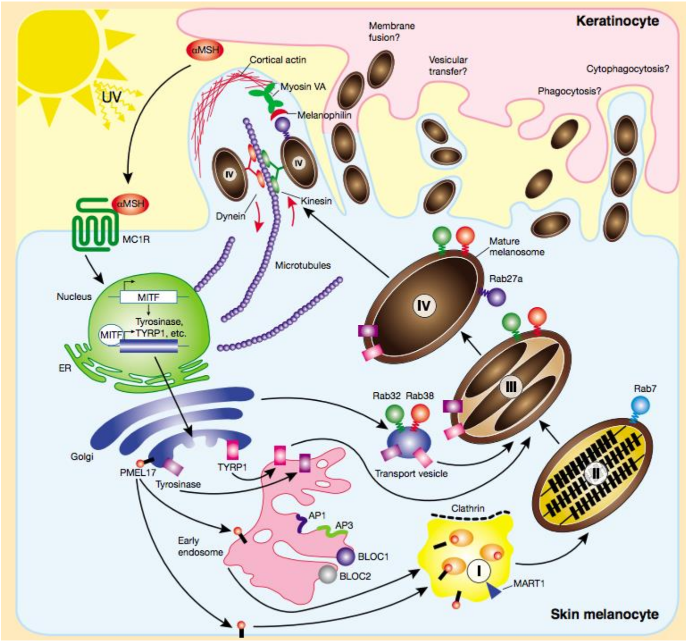


Figure 10 | Melanogenesis Pathway. Melanogenesis produces pigment by undergoing four melanosome stages of maturation within the melanocyte. PMEL17 and MART1 sort to early endosomal stage 1 melanosomes, where PMEL forms fibrils for stage 2. Enzymatic proteins TYR and TYRP1 are trafficked to maturing stage 2. In stage 3, pigment is produced and deposited on the fibrils. In stage 4, melanosomes are fully pigmented and deposited into surrounding keratinocytes. αMSH= α-melanocyte-stimulating hormone; AP= adaptor protein; BLOC= biogenesis of lysosome-related organelles complex; cAMP= cyclic adenosine monophosphate; MART= melanoma antigen recognized by T cells; MC1R= melanocortin 1 receptor; MITF= microphthalmia-associated transcription factor; PMEL17= premelanosome protein; TYRP1= tyrosinase-related protein 1. (Adapted[99]) granules are deposited on the PMEL fibrils, and in stage 4, the melanosomes reach full maturation and pigmentation (Figure 10). After stage 4 maturation, melanosomes undergo anterograde transport to the dendritic projections where they are deposited into the surrounding keratinocytes (Figure 10). The melanogenesis pathway is regulated both externally (UV light) and internally (keratinocytes-melanocyte crosstalk). Keratinocytes produce α-melanocyte stimulating hormone (αMSH) that interacts agonistically with melanocyte membrane receptor, melanocortin 1 receptor (MC1R) (Figure 10)[99]. MC1R signaling stimulates the downstream cAMP pathway which induces transcription of melanocyte-specific transcription factor, MITF (microphthalmia-associated transcription factor). MITF is a master regulator of melanocyte processes and subsequently regulates transcription of critical melanosome proteins, such as TYRP1, TYRP2, TYR, MART1, and PMEL.

As mentioned in section 1.5, melanocytes are also postulated to be potential targets of AA autoimmune attack. Clinical evidence (previously discussed) such as *canities subita* and dysmorphic follicular melanocytes in AA HFs suggest that autoreactive CD8+ T cell infiltrate targets follicular melanocytes for destruction in AA. Moreover, the autoreactive lymphocytes infiltrate the HF bulb region, where HF melanocytes reside. Melanogenesis also exclusively occurs during the growth phase (anagen) of the hair cycle, which is precisely when AA attack occurs (discussed in more detail in next section). Additional studies investigating melanocyte proteins further support the role for melanocyte antigens in AA pathogenesis. Firstly, the highly immunogenic nature of melanocyte antigens is supported by the CD8+ T cell recognition of PMEL, MART1, TYR, TYRP1, and TYRP2[103]. Specifically, PMEL, MART1, and TYR induced T cell activation and were implicated as auto-antigens in another skin autoimmune disease, vitiligo[104,105]. In AA, patient serum reacted with a wide range of epidermal and HF melanocytic proteins[97]. HF-specific melanocyte proteins that were recognized by AA serum were identified as 52-,67-, and 127 kDa in size[97]. Further elucidation of melanocyte antigens in AA revealed PMEL, MART1, and TYRP2 as capable of activating T cells in AA, supporting a potential role for these proteins as AA autoantigens. Collectively, these findings point to a role for melanocyte-specific proteins as candidate autoantigens and possible contributors to disease pathogenesis. Taken together, melanocytes are present at the right time, right place, and induce the right response to be feasible targets for inciting AA autoimmunity.

Hair cycle and Immune Privilege

The hair follicle is the only organ that continually regenerates throughout its lifetime. In order to accomplish this, the HF undergoes three stages of the hair cycle: anagen (growth), catagen (regression), and telogen (quiescence) (Figure 11). As the HF enters the last stage of telogen, a new follicle begins anagen in its place, shedding the previous follicle. Anagen, the growth stage, has the longest duration, lasting 1-8 years for human scalp hairs[59]. During this time, the shaft is continually growing, and the HF matrix is actively proliferating and undergoing melanogenesis. Next, the HF enters the catagen stage which is characterized by HF regression driven by apoptosis. Although catagen is considered to be mostly a result of apoptosis-mediated cell death, other cell death mechanisms and processes such as autophagy may also be implicated in the anagen-to-catagen transition[106]. During catagen, the lower dynamic region of the HF (bulbar and suprabulbar) recedes to the upper non-cycling HF region (Figure 11)[93]. After catagen, the HF enters telogen, a stage of quiescence, following which HF stem cells in the bulge activate to produce a new anagen HF[59,93].

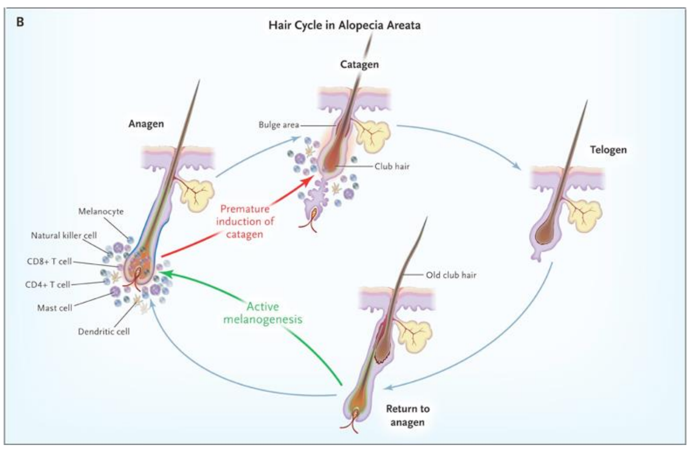


Figure 11 | Hair Cycle and AA attack. HFs undergo a hair cycle consisting of anagen (growth), catagen (regression), and telogen (rest). In AA, immune attack occurs during anagen, while melanogenesis is also active. Lymphocytic infiltrate localizes to the bulb and consists of NK cells, CD8+ T cells, CD4+ T cells, mast cells, and dendritic cells. As a result, the anagen HF is prematurely driven into apoptosis-driven catagen. (Adapted[59])

During anagen, the HF is an immune privileged (IP) site, which refers to tissue compartments that avoid immune self-recognition. Other IP sites include vital tissues that are necessary for an individual's survival and propagation, such as the testis, fetomaternal placenta, blood-brain barrier, and the anterior eye chamber[107]. Complex mechanisms are necessary for maintaining tissue IP including 1) downregulation of MHC class I (molecules for antigen presentation to CD8+ T cells); 2) local immunosuppressants such as α-MSH, TGF-β, MIF, and IL-10; 3) immunoinhibitory signals (CD200 and VIPR); and 4) repression of antigen presenting cells (APC)[58,107]. In AA, IP collapse during anagen is considered to be an initial step in disease pathogenesis. The immunogenic nature of melanocyte antigens is postulated to be a driving factor in IP collapse of AA HFs[58,107]. Melanocytes also possess the capability to process and present MHC class II antigens[96]. Following IP collapse in the AA HF, immunoinhibitory signals are downregulated and HFs express MHC class I and II molecules aberrantly[58]. As a result, immune infiltrate consisting of APCs, CD4+ T cells, and CD8+ T cells invade the bulb region of the anagen HF (Figure 11).

HF and Skin Hereditary Disorders

Dermatologic conditions affecting the skin and hair are caused by both monogenic and polygenic types of disease. Genetic mutations in keratins contribute to a wide range of monogenic skin disorders including epidermolysis bullosa (*K5, K14*), epidermolytic hyperkeratosis (*K1, K10*), epidermolytic palmoplantar keratoderma (*K9*), ichthyosis hystrix (*K1*), and white sponge nevus (*K4, K13*) (Figure 12)[108]. Mutations in both type I and type II keratins are causal in these skin diseases and a majority of pathogenicity is a result of missense mutations[108]. Monogenic keratin mutations also result in pathogenic hair disorders including, monilethrix (*K81, K83, K86*), pseudofolliculitis barbae (*K75*), and ectodermal dysplasia of hair and nail type (*K85)* (Figure 12)[109]. Interestingly, all of the monogenic hair keratin disorders have been attributed to mutations in only Type II hair keratins.

The keratin protein family contains three domains consisting of the head, the alpha-helical rod (1A, 1B, 2A, 2B subdomains), and the tail domain (Figure 12). The rod domain is made up of 310 amino acids that are critical for heterodimerization and IF formation[108]. The vast majority of monogenic keratin disorders result from mutational hotspots at the terminal ends of the 1A and 2B rod subdomains in both type I and type II keratins[108] (Figure 12, orange subdomain). In type II keratins, disease-causing mutations have also been annotated in the head and tail domains[110]. The resulting proteins generated from these mutations have functional defects in heterodimerization causing fragile keratin networks and subsequent weak tissue integrity[108,110].

Human Disorder	Type I Keratin	Type II Keratin
Epidermolysis bullosa simplex (AD)	K14	K5
Epidermolysis bullosa simplex (AR)	K14	
EBS with migratory circinate erythema		K5
EBS with mottled pigmentation	K14	K5
Dowling-Degos disease		K5
Naegeli-Franceschetti-Jadassohn	K14	
Epidermolytic hyperkeratosis (AD)	K10	K1
Epidermolytic hyperkeratosis (AR)	K10	
Ichthyosis hystrix Curth-Macklin		K1
Diffuse non-epidermolytic PPK		K1
Palmoplantar keratoderma with		K1
Ichthyosis bullosa of Siemens		K2
Epidermolytic palmoplantar	K9	
Pachyonychia congenita type I	K16	K6a
Pachyonychia congenita type II	K17	K6b
Steatocystoma multiplex	K17	
Focal non-epidermolytic PPK	K16	
Monilethrix		K81
		K83
		K86
Pseudofolliculitis barbae		K75a
Ectodermal dysplasia of hair and nail		K85
Meesmann corneal epithelial	K12	K3
White sponge nevus	K13	K4
Familial cirrhosis	K18a	K8a
Inflammatory bowel disease		K8a

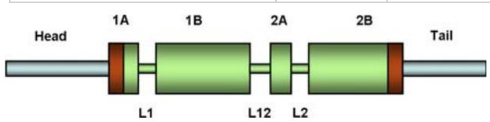


Figure 12 | Keratin disorders and domains. Table of human keratin disorders and the mutations in keratin type I and II proteins attributed to disease. Structure of keratins consisting of a head domain, a rod domain (1A, 1B, 2A, and 2B separated by linker sequences (L)) and tail domain. (Adapted[108])

In addition to a wide range of monogenic keratin disorders, mutations in melanocyte genes also give rise to Mendelian pigmentary skin disorders. As discussed in the previous section, melanogenesis is a tightly regulated process. Therefore, melanocyte dysfunction caused by single gene mutations result in both hypo- and hyper-pigmentary skin disorders. Hypopigmentary disorders include piebaldism, Waardenburg syndrome, oculocutaneous albinism, Hermansky-Pedlak syndrome, Chediak-Higashi syndrome, and Griscelli Syndrome. Although some monogenic *hypo*pigmentary diseases are caused by mutations in signaling genes non-specific to melanocytes (i.e. *KIT* in piebaldism), particularly interesting are the monogenic forms of oculocutaneous albinism that are caused by the mutations in the melanogenesis pathway genes, *TYR, OCA2, TYRP1, SLC45A2, and GPR143*. On the other end of the pigmentation disease spectrum, *hyper*pigmentary disorders such as neurofibromatosis 1, dyskeratosis congenital, Naegeli-Franceschetti-Jadassohn Syndrome (NFJS), or Dowling-Degos disease, are not caused by mutations in the melanocyte-specific genes. Interestingly, NFJS and Dowling-Degos disease are actually caused by mutations in K14 and K5, respectively[109]. Both diseases share a common mesh-like pattern of hyperpigmentation of the skin. The involvement of keratin genes in these similar pathologies suggests dysfunction in keratinocyte retention of melanosomes, rather than skin melanocytes ability to produce melanin pigment.

Complex dermatologic diseases include common disorders that affect either the skin or the HF. Other complex forms of alopecia (hair loss) include androgenetic alopecia (the most common), cicatricial alopecia, telogen effluvium, and lichen planopilaris. AA is unique from these other forms of hair loss because of its autoimmune disease mechanism. Systemic lupus erythematosus (SLE), psoriasis, atopic dermatitis, systemic sclerosis, vitiligo, pemphigus vulgaris and Sjogren's syndrome are similarly all inflammatory/autoimmune diseases that result in varying degrees of skin inflammation and can be distinguished by their unique phenotypes[111]. For example, vitiligo targets and destroys epidermal melanocytes resulting in a depigmentation phenotype in demarcated skin regions. Distinctly, pemphigus vulgaris attacks the epithelial cell adhesion proteins, resulting in painful blisters and lesions on the skin and mucus membranes[112]. Genetic studies have revealed substantial genetic overlap between autoimmune diseases, including the *CTLA4, IL2RA, IL23R, IL2/21, STAT4*, and *PTPN22* loci[113]. Genes that are not shared amongst inflammatory disorders are postulated to explain the disease-specific targets of autoimmune attack unique to each skin autoimmune disease. This thesis will focus on the functional interrogation of two HF-genes (*STX17* and *KRT82*), and how genetic disruption of these genes leads to HF dysfunction and autoimmunity.

1.7 Focus of Thesis

In this thesis, I utilized two main genetic approaches to identify variants associated with AA: targeted sequencing and WES. WES allowed us to perform an unbiased search across the entire exome to identify genes harboring rare variation associated with AA, while targeted sequencing provided a hypothesis-driven approach that allowed us to interrogate all variants, coding and non-coding, in a previously identified GWAS locus associated with AA.

First, we focused our targeted sequencing analyses on loci harboring 1) genes expressed in the hair follicle; and 2) genes specific to AA autoimmunity. This allowed us to focus on end-organ target genes whose disruption was potentially unique to AA disease pathogenesis. Using this strategy, we sequenced a 550 kb region spanning *Syntaxin17* (*STX17*) and its associated LD block in 849 AA cases. We identified 35 variants that defined a risk haplotype responsible for driving the GWAS association. Using a combination of multi-dimensional analyses and *in silico* algorithms, we annotated 32 variants that were significantly associated with AA and downregulated STX17 expression in the skin of AA patients. Interestingly, two variants resided in the *STX17* promoter region and one variant fell in a melanocyte-specific enhancer. I followed up this discovery by determining the functional role of risk gene, *STX17* in melanocyte biology and AA disease pathogenesis.

Secondly, to uncover rare unknown risk variants contributing to AA, we performed WES on the same 849 AA cases and compared these results to 15,640 controls. Gene-level burden tests and collapsing analyses identified the *Keratin 82* (*KRT82*) gene as the highest associated AA gene, achieving genome-wide significance ($p<6.7E-07$), in 3 rare damaging collapsing models. Interestingly, KRT82 is a hair-specific type II keratin that is expressed exclusively in the hair shaft cuticle during the anagen (growth) stage of the hair cycle. *KRT82* is also a HF-specific gene without any previously reported associations with other autoimmune or Mendelian diseases. In Chapter 3, we used WES methods, gene burden analyses, and functional interrogation to understand a potential new mechanism of hair shaft defects and their contribution to AA pathogenesis. Our use of WES and gene collapsing analyses for the first time in AA illustrated the ability of these approaches to identify rare disease associations in functionally relevant risk genes using a moderate sequencing cohort size.

Importantly, I defined two new disease pathomechanisms driving AA susceptibility that implicated the HF itself in disease. Up to now, most of the therapeutic approaches in AA have focused on targeting the immune response and clearance of pathogenic T cells. The findings presented in this thesis suggest that restoring the integrity of structural elements in the HF itself may represent new therapeutic targets that could be useful in combination with immunomodulation. This thesis centered around the use of sequencing strategies to functionally interrogate the role of genetic perturbation in the target organ, HF, and how its dysfunction contributes to AA pathogenesis.

Chapter 2 Functional genomics analysis of STX17 in AA revealed a novel role in melanocyte function

(Manuscript #1, in preparation)

2.1 Introduction

Historical accounts of Marie Antionette described the rapid whitening of her hair in the days prior to her execution in 1793[114]. Since then, both historical and clinical case reports have outlined the occurrence of a phenomenon known as *canities subita,* or sudden whitening of the hair[114,115]. Many dermatologists attributed this phenomenon to Alopecia Areata (AA), a highly prevalent autoimmune disease that is characterized by aberrant CD8+ T cell attack of the hair follicle (HF), resulting in patchy (AAP) or widespread hair loss of the scalp (AT) or body (AU)[114-116]. In *canities subita*, the immune cells preferentially attack the pigmented hair follicles, sparing unpigmented follicles from death and eliciting a deceptive appearance of sudden whitening of the patient's hair[64,117]. In AA, pigmented hairs with active melanocytes, pigment-producing cells, are aberrantly targeted for destruction by immune cells, while white hairs without functionally active melanocytes are spared. Consequentially, a widely accepted hypothesis in the dermatology field is that the pigment-producing melanocytes are the target of AA attack[64,117]. The precipitating AA antigen is still not fully elucidated; however, convincing evidence suggest melanocytic antigens as potential targets for AA autoimmune activation.

To further define the role of melanocyte antigens and the preferential attack of pigmented hair follicles in AA, we investigated the 14 previously-identified GWAS risk loci for pigment-related genes that may explain this phenomenon[70,71]. Of the 14 risk regions, a locus on chromosome 9 harbored a gene that was previously implicated in hair pigmentation. *Syntaxin17* (*STX17*) encodes a protein (STX17) with a well-defined role in autophagy through mediation of autophagosome-lysosome fusion and autophagosome formation at mitochondria-ER contact sites[75,76]. A possible new role for *STX17* in pigment regulation was identified in a recent genetic study that investigated 27 traits in ~415,000 subjects, and found a SNP in the 3' UTR of *STX17* that was significantly associated with human hair color[72]. Prior to identification of the *STX17* association with human hair color, a 4.6 kb duplication in intron 6 of *STX17* was identified as a causal mutation for premature graying in Lipizzaner horses[73]. However, the mechanism underlying the relationship between *STX17* variants and hair color/graying remains undefined.

The significant associations of *STX17* in both hair pigmentation and AA suggested that *STX17* function may be a potential genetic link between AA pathogenesis and the targeted attack on pigmented HFs[70-73]. In this study, we utilized a targeted sequencing fine mapping approach to identify significant variant(s) driving the GWAS association between the 550 kb *STX17* region and AA risk. Through the use of *in silico* methods, multi-dimensional datasets, and *in vitro* assays, we investigated the role of *STX17* in pigmentation and the impact of its disruption on AA susceptibility.

2.2 Targeted genomic sequencing of the *STX17* locus identified 35 non-coding risk variants

GWAS is a useful tool for elucidating part of the genetic architecture of complex disorders by identifying common variants associated with a disease or trait of interest. However, fine mapping of the associated regions is necessary to determine the genetic variants that are driving the GWAS signal[118]. GWAS test only a portion of common variants present in the human genome, resulting in the possibility of causal SNPs not being directly interrogated using this method. The surrounding SNPs residing on a given linkage disequilibrium (LD) block are disproportionately inherited together, allowing identification of associated regions but incapable of distinguishing the true risk variant(s) from variants carried along on the same risk haplotype. As a result, follow-up fine mapping of association regions is necessary for identifying the causal variant(s). Our GWAS identified an association of a SNP (rs10760706) in intron 6 of *STX17* with AA risk (p=3.6E-07)[70]. To resolve the potential variants contributing to this association, we performed targeted sequencing of a 550 kb region including the *STX17* gene and surrounding LD block (chr9:102560179-103110179; hg19) (Figure 13). We sequenced both coding and non-coding regions given that the GWAS SNP, rs10760706, is an intronic variant and 90% of disease-causing variants in autoimmune diseases are non-coding[20].

Sequencing, alignment, variant calling, and quality control were performed as previously described[119] in 849 AA cases and 62 controls of European descent. Following our analytic pipeline (Figure 13), we identified 5,841 variants in AA cases across the 550 kb region. In order to prioritize variants and define those responsible for driving the association with AA, we identified a region (chr9:102669557-103053538) linked with our GWAS SNP, whose boundaries were defined by variants with $r^2 \geq 0.4$. 4,037 AA variants fell within these boundaries and,

therefore, we further defined a LD block of interest by variants in strong linkage disequilibrium (LD) with our GWAS SNP, rs10760706 ($r^2 \geq 0.8$). This resulted in a region spanning

chr9:102669557-102734130 and contained 693 AA variants. Of the 693 variants in the region, 35 were commonly inherited with rs107060706 ($r^2 > 0.8$), thus defining our risk haplotype. We found that the 35 variants defined two common haplotypes[120], one of which contained the GWAS risk allele (C) and was present in 26.3% of the European (CEU) population (Figure 14). Interestingly, the SNP identified in the hair color GWAS (rs9556) was simultaneously carried on this haplotype[72] (Figure 14-15a). Additionally, the AA risk haplotype carrying the 35 variants spanned the same region as the *STX17* 4.6 kb duplication previously identified as causal for regulating premature hair greying in Lipizzaner horses (Figure 15a)[73].

Next, we tested all of the sequencing variants (n=465) that fell within the haplotype-defined boundaries (chr9:102669557-102734130) and compared case allele counts to quality-controlled allele data available from the gnomAD database (Genome Aggregation Database)[121]. We analyzed significant associations using Non-Finnish European (NFE) controls from gnomAD and found that 33 of the 35 variants that defined the risk haplotype were significantly associated with AA after accounting for multiple testing (α=0.05/465=1.08E-04). The other two variants (rs10124366 and rs10116142) reached nominal significance. Other variants within the region that were not in strong LD with rs10760706 and did not define the risk haplotype demonstrated weaker association with AA (Figure 15b, Table 1). This suggested that the risk haplotype carrying 35 variants was responsible for driving the GWAS association signal, and the 33 significantly associated variants were contributing to AA risk. The discovery that the AA risk haplotype included a variant previously reported to be associated with hair color supported the hypothesis that the disruption of hair pigmentation may play a causal role in AA and *STX17* could be a critical mediator of that disease mechanism.

2.3 eQTL analysis identified 32 significantly associated functional variants

Given that all of the SNPs that defined our risk haplotype were noncoding, and that most common risk variants are non-coding[20,122], we next leveraged multi-dimensional datasets made available by consortiums such as ENCODE and GTEx[22,123]. Expression Quantitative Trait Loci (eQTLs) are a particularly useful annotation for identifying meaningful regulatory effects of non-coding variants, due to their direct integration of genotype alleles and gene expression levels in a given cell or tissue type[26]. Previous expression data from our lab reported a significant decrease of *STX17* in the scalp skin of AA patients compared to controls (p=3.08E-04)[124]. However, the other four genes (*NR4A3, ERP44, INVS, TEX10*) located in this 550 kb region were not differentially expressed in AA. As a result, we postulated that *STX17* gene expression had a potential role in disease pathogenesis and investigated the functional effects of these 35 variants on *STX17* gene expression levels in skin.

We used human tissue expression and genotype data from the GTEx portal and analyzed eQTL effects of each significantly associated risk haplotype variant on the expression of *STX17* in the skin, the target organ in AA, and found that all 35 SNPs were significant eQTLs in the skin (Table 1). To test whether these regulatory effects were relevant to AA, we used a multi-dimensional data set that integrated genetic data with expression data from lesional scalp samples in 41 AA patients. We tested for association of genotypes with STX17 expression levels in the skin, and assessed significance using a pairwise t-test with FDR (false discovery rate) correction and confirmed that 33 of the 35 variants on the risk haplotype significantly downregulated the expression of *STX17* in lesional (hair loss) AA skin biopsies. Nominally associated rs10124366 and hair-color SNP rs9556, did not achieve statistical significance for genotype-dependent effects on expression. However, both were trending towards decreased expression of the risk alleles. Interestingly, rs10116142 reached eQTL significance even though it was not significantly associated with AA. Taken together, we identified 32 variants on the risk haplotype that were significantly associated with AA and had a functional effect on skin expression. While only homozygous carriers of the risk alleles expressed significantly reduced levels of *STX17* in the scalp skin, trending decreased expression levels of the heterozygotes suggested that the effect may be dominant, although a larger sample size would be needed to establish significance (Figure 16). AA GWAS SNP, rs10760706, was one of the 32 variants identified as a significant skin eQTL. Using this approach, we annotated a functional effect of 32 sequencing variants that not only significantly associated with AA, but also, significantly reduced expression of *STX17* in the scalp of AA patients.

Because these 32 SNPs are in high LD, we next sought to identify the genetic mechanism regulating expression levels of STX17. This would allow us to distinguish regulatory variant(s) that result in dysregulation from those that appear to effect regulation simply because they are being carried along a shared haplotype. *In silico* algorithms are useful in predicting variants that fall in regulatory regions of specific cell types and likely exerting regulatory consequences cell-specific expression. The Functional Latent Dirichlet Allocation model (FUN-LDA) predicts functional effects of every genomic position in 127 human cell and tissue types, including the three skin cell types: melanocytes, keratinocytes, and fibroblasts[24]. By incorporating H3K4me1, H3K4me3, H3K9ac, H3K27ac histone modifications and DNase hypersensitivity sites, the algorithm provides a reliable and accurate predictive score of regulatory effects (scale of 0-1) for each position of the genome. Integrating the FUN-LDA model with our risk haplotype variants, we identified a single SNP, rs7039716, that was predicted to be functional in all three skin cell types (Figure 17a). This SNP falls in the first intron of *STX17* in the 5' gene end. FUNLDA regulatory peaks at the 5' gene end and present in all cell datasets are indicative of promoter regions, predicting the presence of rs7039716 in the *STX17* promoter (Figure 17a). Further interrogation of this predicted regulatory SNP using data extracted from Roadmap chromatin states confirmed that rs7039716, in addition to rs2416935, fell in active transcription start sites (TSS) in primary melanocytes, keratinocytes, and fibroblasts (Figure 17b). Interestingly, melanocyte cells had additional regulatory regions (annotated as enhancers by Roadmap) located across the gene that were not present in keratinocytes or fibroblasts (blue peaks, Figure 17a; yellow bars Figure 17b). We found that three of the AA-associated SNPs (rs2416936, rs10760706, rs66799478) fell proximal to melanocyte enhancers (<400bp) and one variant resided in a melanocyte enhancer (rs7027813) (Figure 17c). The increased frequency of regulatory regions in melanocytes compared to other skin cell types, in addition to presence of one of the AA risk variants in a melanocyte enhancer allowed us to prioritize melanocyte cells for follow-up functional assays.

In addition to identifying melanocyte-specific enhancers dispersed across the STX17 locus, previous findings on preferential attack of pigmented HFs, dysmorphic follicular melanocytes in AA HFs, and melanocytic antigens eliciting T cell responses in AA models, further supported a role for melanocytes in AA pathogenesis. We found additional evidence for a disease-relevant role of melanocytes in AA from our microarray expression data in AA scalp biopsies[124]. Pathway analysis performed using DAVID (Database for Annotation, Visualization and Integrated Discovery) identified melanogenesis, a pigment-producing process performed by melanocytes, as one of the top 30 most significantly affected pathways in AA, with 30 melanogenesis genes differentially expressed in AA cases compared to controls (p=7.6E-03) (Figure 18). This provided added support for a pathomechanistic disruption of melanocyte function in AA and justified further investigation into the possible role of *STX17* in melanocyte dysfunction and AA disease.

2.4 Aberrant expression of STX17 proximal to dysmorphic melanocytes in AA HFs

We identified and functionally annotated 32 AA-associated variants that downregulate *STX17* expression in the skin of AA patients and suggested implication of melanocytes as the candidate cell type mediating the genetic association of *STX17* with AA using *in silico* methods and published literature findings. Since we previously reported that *STX17* expression was reduced in AA whole skin, we next wanted to investigate whether the HF and HF melanocytes, specifically, exhibited dysregulated *STX17* expression in AA patients. In the normal healthy HF, we observed heterogenous *STX17* expression throughout the hair follicle with a strong immunofluorescent (IF) band of expression along the basement membrane outlining the dermal papilla (DP) (Figure 19a). The follicular melanocytes, marked by premelanosome protein (PMEL), localize to the same region surrounding the DP. However, both nonlesional and lesional biopsies from the same AAP patient revealed reduced staining along the basement membrane (Figure 19b-c). Additionally, the AA hair bulb melanocytes were dysmorphic and displaced, supporting previous observations[64]. In the nonlesional, non-balding site, we observed a large reduction of *STX17* along the basement membrane proximal to the melanocytes, which displayed disrupted morphology and early detachment away from the DP and into the HF matrix (Figure 19b). In the lesional active disease site, the basement membrane *STX17* expression was almost completely lost with fewer hair bulb melanocytes (Figure 19c). The melanocytes that remained in the lesional HF appeared dysmorphic and completely detached from the basement membrane surrounding the DP. Collectively, these findings confirmed aberrant *STX17* expression in AA HFs with proximity to disrupted bulbar melanocytes *in vivo.*

2.5 STX17 reduction resulted in melanocyte dysfunction and antigen accumulation

Both *in vivo* findings of dysregulated *STX17* expression proximal to dysmorphic melanocytes in AA HFs and *in silico* annotations for melanocyte-specific regulatory regions across the *STX17* gene locus prompted us to perform functional follow-up analyses in human primary melanocytes. Melanocytes pigment the skin and hair through a process called melanogenesis, or the production of melanin, a highly oxidative reaction that must be tightly regulated in specialized organelles called melanosomes[125,126]. Melanosome biogenesis occurs in four stages and is dependent upon trafficking of specified proteins and enzymes to the melanosomal vesicles where melanin is produced, and pigmented organelles are formed. Genetic disruption of melanogenesis can lead to a wide range of disorders including oculocutaneous albinism, Hermansky-Pudlak syndrome, Charcot-Marie-Tooth disease, and vitiligo, another dermatological autoimmune disease[126,127].

Because melanocyte processes are highly regulated and known disruptions cause other diseases, we investigated whether melanogenesis was affected by *STX1*7, and sought to determine whether *STX17*-mediated disruption of melanogenesis was a contributing factor in AA risk associated with *STX17* variation. Due to our finding that AA risk haplotype variants decreased expression in the skin, we used *STX17*-targeted siRNA knockdown to deplete STX17 levels in primary human melanocytes. We found that in the absence of STX17, melanin production was impaired indicated by an overall decrease in melanin content compared to controls (Figure 20a). This suggested, for the first time, that STX17 played a critical role in the melanogenesis pathway. Due to the established role of STX17 in autophagy, we investigated whether this observed effect was a result of a newly identified role of STX17 in melanocyte function or simply a result of autophagy disruption. To test this, we incubated primary human melanocytes with 3 different autophagy inhibitors, 3-Methyladenine (3-MA), Chloroquine (CQ), and Bafilomycin A1 (BA1). Interestingly, in the presence of autophagy inhibitors, melanocytes exhibited the opposite phenotype as STX17-knockdown and had elevated levels of cellular melanin content (Figure 20b). Taken together, this supported a novel autophagy-independent role for STX17 in melanogenesis.

To further understand the novel role of STX17 in melanogenesis, we investigated melanosome biology using electron microscopy (EM). Unexpectedly, we saw equivalent numbers of total and late-stage melanosomes (Figure 20c-d). This was surprising to us since late-stage melanosomes produce melanin pigment, yet the overall melanin content for *STX17*-knockdown cells was decreased. This suggested that in the absence of STX17, the melanosomes could form and mature, however, they were unable to produce melanin pigment within the melanosomes as efficiently as controls. Interestingly, the number of early melanosomes was significantly decreased in STX17-deficient melanocytes (Figure 20c-d). This supported a potential role for STX17 in melanosome biogenesis and early melanosome trafficking. Further support of an autophagy-independent role of STX17 in melanogenesis was evident by the EM phenotype of BA1-treated melanocytes that appeared strikingly different from STX17-depletion conditions with a high concentration of autophagosomes and multivesicular bodies as is characteristic of late-stage autophagy inhibition in cells (Figure 20e)[128]. Taken together, STX17 exerts an autophagy-independent effect on early melanosome biology that results in an overall reduction in melanin production. STX17-deficient cells were able to overcome early melanosome reduction and mature to produce equivalent numbers of late stage melanosomes and low levels of melanin, suggesting the presence of a secondary redundant pathway that may act in parallel to STX17. We postulated that the reduction of early melanosomes may be a result of a trafficking defect, and since early melanosomes are responsible for structuring the organelle for melanin deposits, melanin may not be as efficiently produced in the absence of STX17 and proper cargo trafficking.

To further investigate the potential role for STX17 in early melanosome cargo trafficking and biogenesis, we interrogated the effect of STX17 loss on levels of early and late stage melanosome proteins. MART1 (Melanoma Antigen Recognized by T-Cells 1) is an early melanosome protein critical for melanosomal structural formation, and TYR (Tyrosinase) is the rate-limiting enzyme that is trafficked to late stage III melanosomes to synthesize melanin and form pigmented melanosomes[99,102,129,130]. Since we observed that STX17-depleted cells were less capable of producing melanin, we postulated that the proteins necessary for the reaction would also be negatively affected, explaining the cell's inability to efficiently perform melanogenesis. We found that late stage protein, TYR, levels were unaltered and that surprisingly, early stage protein MART1 levels accumulated in STX17-KD melanocytes (Figure 20f). Once again, we observed a STX17-KD specific effect, as there were negligible increases of MART1 in the autophagy inhibition conditions (Figure 20g). The accumulation of early melanosome MART1 protein, decreased number of early melanosomes, and reduced melanin content suggested a defect in early melanosome cargo trafficking. The lack of early melanosomes and reduced ability to produce melanin was speculatively due to aberrant accumulation of MART1 unable to traffic to early melanosomes and regulate proper melanosome structure and formation. Preliminary evidence supports this hypothesis in which we observed MART1 and STX17 co-localization at baseline conditions and after melanogenesis stimulation with αMSH (Figure 21). However, future work will be needed to pinpoint the precise mechanism of STX17 cargo trafficking in melanocytes. Collectively, these findings support a novel role for STX17 in melanocyte biology that when disrupted, results in accumulation of key melanosome protein, MART1.

The accumulation of MART1 is particularly interesting as it has previously established antigenic roles in cancer and autoimmunity. It was reported as a potential autoantigen in both AA and vitiligo, another autoimmune disease that targets epithelial melanocytes[104,105,131,132]. The cellular accumulation of MART1, a protein with auto-antigenic properties could feasibly act as a precipitating event for autoimmune attack due to an increase in the cell's antigenic load. Cellular accumulation of a reported antigen in the absence of STX17 provides a mechanism to explain how the genetic predisposition to decreased STX17 expression levels impairs melanocyte function and results in the accumulation of antigens recognized by auto-reactive T cells. We investigated this hypothesis *in vivo* and found that patients categorized with low levels of STX17 expression in the skin had large amounts of follicular CD8+ infiltrate (Figure 20h). In conclusion, we found that reduction of STX17 in melanocytes disrupted melanocyte function and caused accumulation of melanosome autoantigen MART1 *in vitro*. Furthermore, *in vivo* we observed that decreased STX17 expression correlated with high CD8+ infiltrate surrounding the HF, a pathognomonic characteristic of AA.

2.6 Discussion

This study outlines the successful integration of regional fine mapping and *in silico* prediction methods in AA to identify a functionally relevant cell type and elucidate a mechanistic role of a GWAS-identified gene, *STX17*, in cellular dysfunction and disease pathogenesis. Using targeted sequencing, we identified a risk haplotype of 35 functional variants in the STX17 region that was responsible for driving the GWAS-identified disease association. Through multidimensional datasets obtained from GTEx and 41 AA patients, we annotated that 32 of the significantly associated variants also significantly downregulated STX17 expression in AA skin. Using *in silico* methods, we prioritized three variants as likely causal due to their co-localization with regulatory regions, thus distinguishing them from the rest of the haplotype variants that were not annotated as regulatory, but rather were just carried along on the same risk haplotype. The FUN-LDA model predicted one of these variants (rs7039716) to be likely causal in the promoter region of skin cell types, and further suggested that melanocytes may be the functionally relevant skin cell type with regions of high predicted regulatory effect in melanocytes that were not predicted in fibroblasts or keratinocytes. Roadmap chromatin state data confirmed the finding of rs7039716, in addition to another variant (rs2416935) in the transcription start sites (TSS) of all skin cell types. Additionally, we annotated one variant (rs7027813) that fell in a melanocyte enhancer and three other variants proximal to melanocyte enhancers, allowing us to prioritize melanocytes as the candidate cell type for functional analyses. Interestingly, aberrant STX17 expression *in vivo* correlated with dysmorphic melanocytes in AA HFs. Functional studies in melanocytes further revealed a novel autophagy-independent role for STX17 in melanin production and melanocyte biology.

We propose that under normal conditions, STX17 aids in the transportation and delivery of early melanosome protein, MART1, to forming melanosomes (Figure 22). MART1 forms a complex with PMEL (premelanosome) protein and is necessary for regulating its expression, processing, and function in melanosome formation[102]. After PMEL fibrils form, pigment is deposited in the late-stage melanosomes (III and IV) before they are sent into surrounding keratinocytes to pigment the growing hair shaft (Figure 22). However, in susceptible individuals with the AA risk haplotype, STX17 levels are decreased (Figure 22). Consequentially, melanosome protein MART1 is not efficiently trafficked to the early melanosome targets. Aberrant MART1 trafficking would likely affect PMEL function, inhibiting the efficient structure formation and subsequent melanin production within melanosomes. We observed that melanin production is largely inhibited but not completely blocked providing evidence for redundant pathways. Melanosomes are still able to form, mature, and traffic; however, their ability to efficiently and effectively produce melanin is hindered. We propose that in the absence of STX17, MART1 is unable to be transported to the melanosomes and aberrantly accumulates in the cell. Moreover, due to the antigenic nature of MART1, the melanocyte antigenic load increases resulting in inappropriate recognition by auto-reactive CD8+ T cells and subsequent autoimmune attack (Figure 22). Although we propose compelling evidence for this model, more work is needed to elucidate the precise mechanism of STX17 function in melanocytes.

This functional approach using targeted sequencing of a GWAS region in a complex disease and application of both *in silico* and experimental assays allowed us to elucidate the role of genetic variants in AA disease pathogenesis and can be applied as a model to the genetics field as whole. Additionally, the novel role identified for *STX17* in melanocyte biology and antigenicity is pertinent to both the autoimmune and oncology fields. The study in Lipizzaner horses that identified the causal *STX17* 4.6kb-duplication in premature graying, additionally found that gray horses with the duplication had higher incidence of melanomas. The melanomas from the gray horses exhibited increased expression of both *STX17* and neighboring gene, *NR4A3*[73]. Our data suggested that in the absence of *STX17*, melanocytes accumulate antigens that result in aberrant autoimmune attack. Conversely, perhaps increased levels of *STX17* leads to an immune evasion phenotype and subsequent melanoma development. Therefore, *STX17* could be an important target for not only treating melanocyte-related autoimmune diseases, such as AA, but also melanomas.

Through the use of a multi-dimensional functional genomics approach, we successfully identified candidate causal variants in the *STX17* region that were responsible for driving the GWAS association and proposed a potential mechanism explaining the elusive preferential attack of pigmented hair follicles in AA.

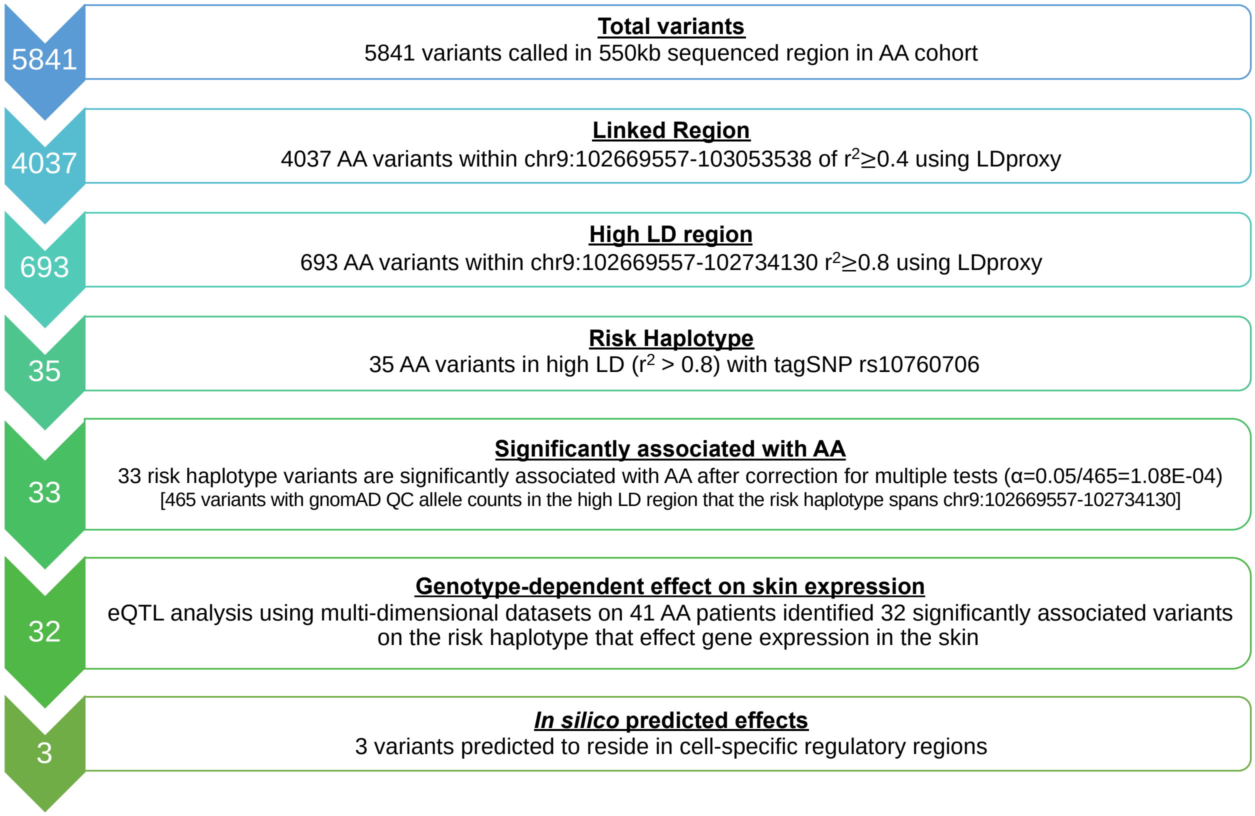


Figure 13 | Flowchart of variant identification strategy in *STX17* risk region. Targeted sequencing was performed on 849 AA individuals in the 550 kb GWAS association region spanning the *STX17* LD block and identified 5841 variants present in AA cases. 4037 AA variants fell in a loosely linked region defined as $r^2 \geq 0.4$. 693 AA variants were identified in region of tight linkage, defined by $r^2 \geq 0.8$ 35 of the variants were in high LD with GWAS tagSNP, rs107060706, with an individual $r^2 \geq 0.8$. Variant-level significance was calculated for all variants in the region spanning the high LD region (64.5 kb, chr9:102669557-102734130) using gnomAD Non-Finnish European (nfe) control allele counts and Fisher's exact test. P-value threshold was adjusted for number of tests performed (n=465). 33 of the 35 variants on the risk haplotype were significantly associated with AA (p>0.000108). 32 of the significant AA variants were identified as eQTLs in the scalp skin of AA cases (using a pairwise t-test). *In silico* algorithms (FUN-LDA) and Roadmap epigenetic annotations predicted 3 variants to likely be driving the regulatory effects of the haplotype.

		Common		Rare			
		Freq=65.7%	Freq=26.3%	Freq=4%	Freq=3%	Freq=0.5%	Freq=0.5%
	AA Risk Allele	**Hap01**	**Hap02**	**Hap03**	**Hap04**	**Hap05**	**Hap06**
rs2416935	**T**	G	T	G	T	G	T
rs7039716	**T**	A	T	A	T	A	T
rs7027813	**G**	A	G	A	G	A	G
rs2416936	**T**	A	T	A	T	A	T
rs4742777	**C**	T	C	T	C	T	C
rs2416937	**C**	A	C	A	C	A	C
rs4743370	**T**	G	T	G	T	G	T
rs10760700	**G**	A	G	A	G	A	G
rs2900224	**T**	C	T	C	T	C	T
rs10121880	**G**	A	G	A	G	A	G
rs10121601	**T**	G	T	G	T	G	T
rs10124366	**G**	A	G	A	A	A	G
rs10117828	**A**	G	A	G	A	G	A
rs4282626	**G**	A	G	A	G	A	G
rs11387575	**A**	AT	A	AT	A	AT	A
rs9299335	**G**	A	G	A	G	A	G
rs2416940	**C**	T	C	T	C	T	C
rs7022999	**T**	C	T	C	T	C	T
rs5899398	**TG**	T	TG	T	TG	T	TG
rs4742778	**T**	G	T	G	T	G	G
rs2416942	**G**	C	G	C	G	C	G
rs10760704	**G**	A	G	A	G	A	G
rs7038506	**T**	C	T	C	T	C	T
rs10217337	**G**	A	G	A	G	A	G
rs10217692	**C**	T	C	T	C	T	C
rs10217366	**G**	T	G	T	G	T	G
rs2031035	**A**	G	A	G	A	G	A
rs1852863	**A**	G	A	G	A	G	A
rs1997367	**A**	G	A	G	A	G	A
rs7027619	**T**	G	T	G	T	G	T
rs10760706	**C**	T	C	T	C	T	C
rs66799478	**TACA**	T	TACA	T	TACA	T	TACA
rs10116142	**G**	C	G	G	G	C	G
rs9556*	**T**	C	T	C	T	T	T
rs11336521	**CT**	C	CT	C	CT	CT	CT

Figure 14 | Observed Haplotypes in the European population. The risk alleles (in red) of the 35 variants that compose the risk haplotype all reside on Hap02, which is observed in 26.3% of European (CEU) individuals. Hap02 frequency is comparable to the risk allele frequency of the individual risk variants in controls. Four rare (<5%) haplotypes are observed, black boxes denote locations were genotypes differ from the common haplotypes. AA GWAS SNP in red and underlined. * denotes the SNP identified in hair color heritability GWAS.

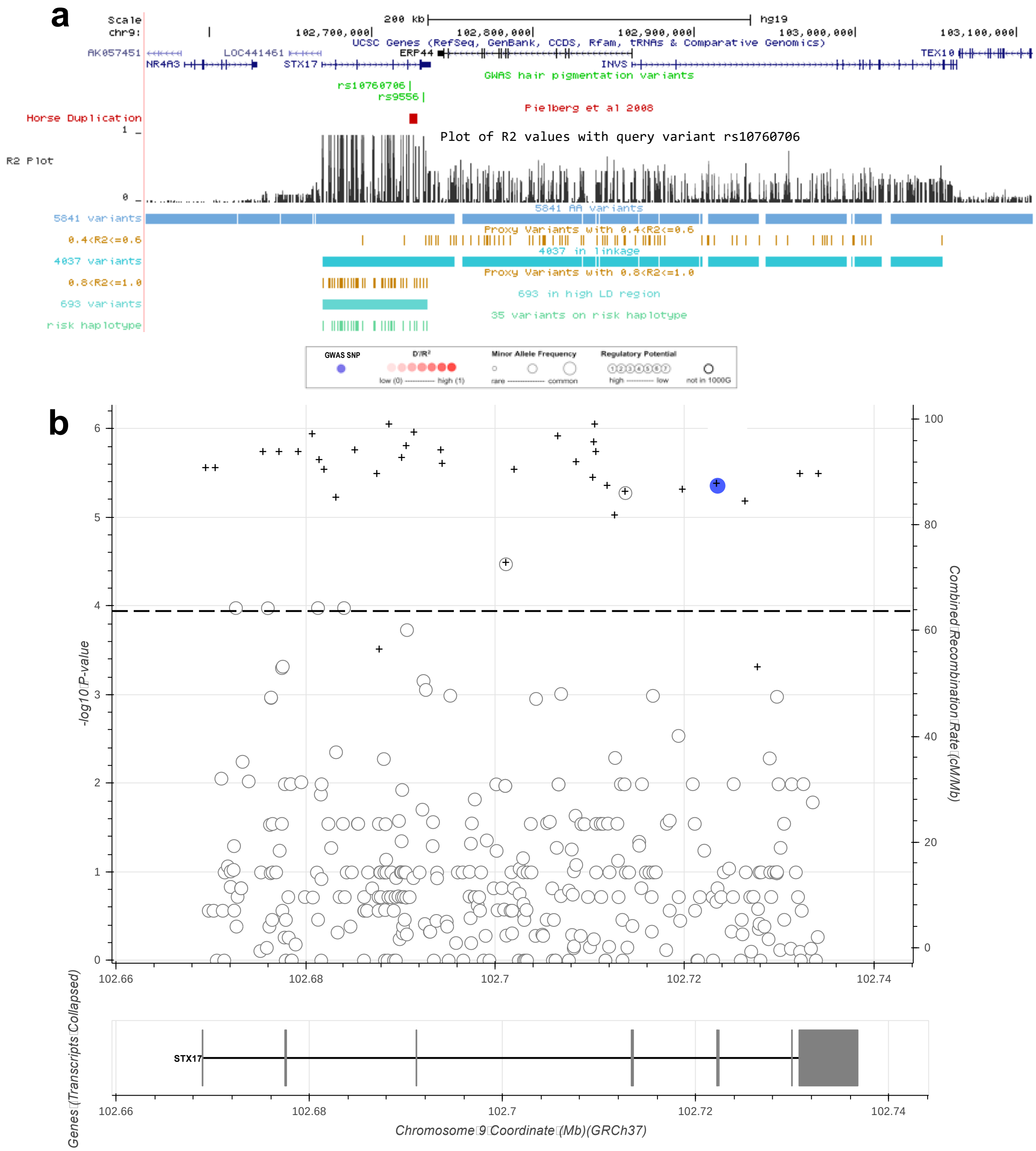


Figure 15 | 35 risk variants define the AA risk haplotype. a, Plot of 550 kb region spanning *STX17* GWAS association region. STX17 GWAS variants in green; rs10760706 associated with AA, rs9556 associated with hair color heritability. Horse duplication associated with premature hair graying (in red) resides in same 3' *STX17* gene end as the GWAS variants. r^2 plot (black peaks) depicts r^2 value of variants in relation to rs10760706. Yellow ticks denote location of variants with $r^2 \geq 0.4$ or 0.8, respectively. Dark blue, light blue, and turquoise tracks denotes the location of AA variants in the whole region, in the loosely linked region($r^2 \geq 0.4$), and in the tight LD region ($r^2 \geq 0.8$), respectively. 35 variants (green-blue ticks) were in high LD ($r^2 > 0.8$) with rs10760706 and define the risk haplotype driving association at this locus. **b,** 465 variants in the LD block with gnomAD data available. GWAS SNP in blue. 35 variants in high LD (red). 33 are significantly associated with AA ($-\log_{10}$ p-value). Dotted line signifies p-value threshold. + represent 35 variants on risk haplotype.

rsID	Protective Haplotype	Risk Haplotype	gene	AA MAF	Ctrl MAF	r2	Assoc p-value	GTEX p-value	AA eQTL p-value
rs2416935	G	T	'STX17'	36.26%	30.62%	1	2.9E-06	1.1E-30	0.044
rs7039716	A	T	'STX17'	36.26%	30.64%	1	2.92E-06	1.1E-30	0.044
rs7027813	A	G	'RP11-60I3.4'	36.29%	30.56%	1	1.86E-06	3.9E-31	0.044
rs2416936	A	T	'RP11-60I3.4'	36.35%	30.60%	1	1.87E-06	3.9E-31	0.044
rs4742777	T	C	'STX17'	36.32%	30.59%	1	1.92E-06	2.8E-31	0.044
rs2416937	A	C	'STX17'	36.49%	30.64%	1	1.22E-06	2.8E-31	0.044
rs4743370	G	T	'STX17'	36.30%	30.64%	1	2.51E-06	2.8E-31	0.044
rs10760700	A	G	'STX17'	36.25%	30.62%	1	2.88E-06	1.1E-30	0.044
rs2900224	C	T	'STX17'	36.24%	30.78%	1	6.23E-06	2.8E-31	0.044
rs10121880	A	G	'STX17'	36.30%	30.58%	1	1.9E-06	2.8E-31	0.044
rs10121601	G	T	'STX17'	36.30%	30.70%	1	3.31E-06	2.8E-31	0.044
rs10124366	A	G	'STX17'	29.67%	25.55%	0.86	0.000318	8.3E-25	0.12
rs10117828	G	A	'STX17'	36.45%	30.55%	1	9.17E-07	2.8E-31	0.044
rs4282626	A	G	'STX17'	36.30%	30.59%	1	2.08E-06	2.8E-31	0.044
rs11387575	AT	A	'STX17'	36.43%	30.64%	1	1.6E-06	2.8E-31	0.044
rs9299335	A	G	'STX17'	36.37%	30.50%	1	1.18E-06	2.8E-31	0.044
rs2416940	T	C	'STX17'	36.35%	30.60%	1	1.88E-06	2.8E-31	0.044
rs7022999	C	T	'STX17'	36.39%	30.72%	1	2.49E-06	2.8E-31	0.032
rs5899398	T	TG	'STX17'	36.20%	31.17%	1	0.000034	1.1E-30	0.044
rs4742778	G	T	'STX17'	36.29%	30.67%	0.98	2.88E-06	2.8E-31	0.044
rs2416942	C	G	'STX17'	36.45%	30.61%	1	1.23E-06	2.8E-31	0.044
rs10760704	A	G	'STX17'	36.27%	30.59%	1	2.43E-06	1.1E-30	0.048
rs7038506	C	T	'STX17'	36.19%	30.63%	1	3.75E-06	1.1E-30	0.044
rs10217337	A	G	'STX17'	36.37%	30.60%	1	1.59E-06	2.8E-31	0.044
rs10217692	T	C	'STX17'	36.39%	30.66%	1	1.88E-06	2.8E-31	0.044
rs10217366	T	G	'STX17'	36.49%	30.56%	1	9.03E-07	2.8E-31	0.044
rs2031035	G	A	'STX17'	36.15%	30.62%	1	4.55E-06	2.8E-31	0.048
rs1852863	G	A	'STX17'	35.89%	30.54%	1	9.68E-06	1.1E-30	0.044
rs1997367	G	A	'STX17'	36.04%	30.55%	1	5.35E-06	3.1E-30	0.044
rs7027619	G	T	'STX17'	36.13%	30.62%	1	4.87E-06	2.8E-31	0.044
rs10760706	T	C	'STX17'	36.08%	30.56%	1	4.28E-06	1.3E-30	0.044
rs66799478	T	TACA	'STX17'	36.30%	30.86%	1	6.6E-06	2.7E-31	0.044
rs10116142	C	G	'STX17'	40.13%	35.82%	0.83	0.000522	2.3E-15	0.02
rs9556*	C	T	'STX17'	36.30%	30.72%	0.98	3.37E-06	4.9E-30	0.078
rs11336521	C	CT	'STX17'	36.70%	31.07%	0.98	3.36E-06	1.1E-29	0.047
	Freq=65.7%	Freq=23.3%							

Table 1 | 35 AA risk variants identified as skin eQTLs. Table of 35 variants that define the risk haplotype. Risk alleles are in red. AA and control (ctrl) minor allele frequencies (MAF) are listed with r^2 values relative to GWAS SNP, rs10760706 (red underlined). Association (Assoc) p-value is the FET p-value comparing the distribution of AA allele counts compared to gnomAD non-Finnish European (nfe) allele counts. GTEx column lists p-values ascertained from GTEX eQTL data of individual variant's effect of STX17 expression in suprapubic skin (not sun exposed). AA eQTL p-value is the statistical significance measured by pairwise t-test with FDR correction of the genotype-dependent effect of variant risk alleles on the expression of STX17 in the scalp skin of 41 AA patients. * denotes the SNP identified to be associated with hair color heritability.

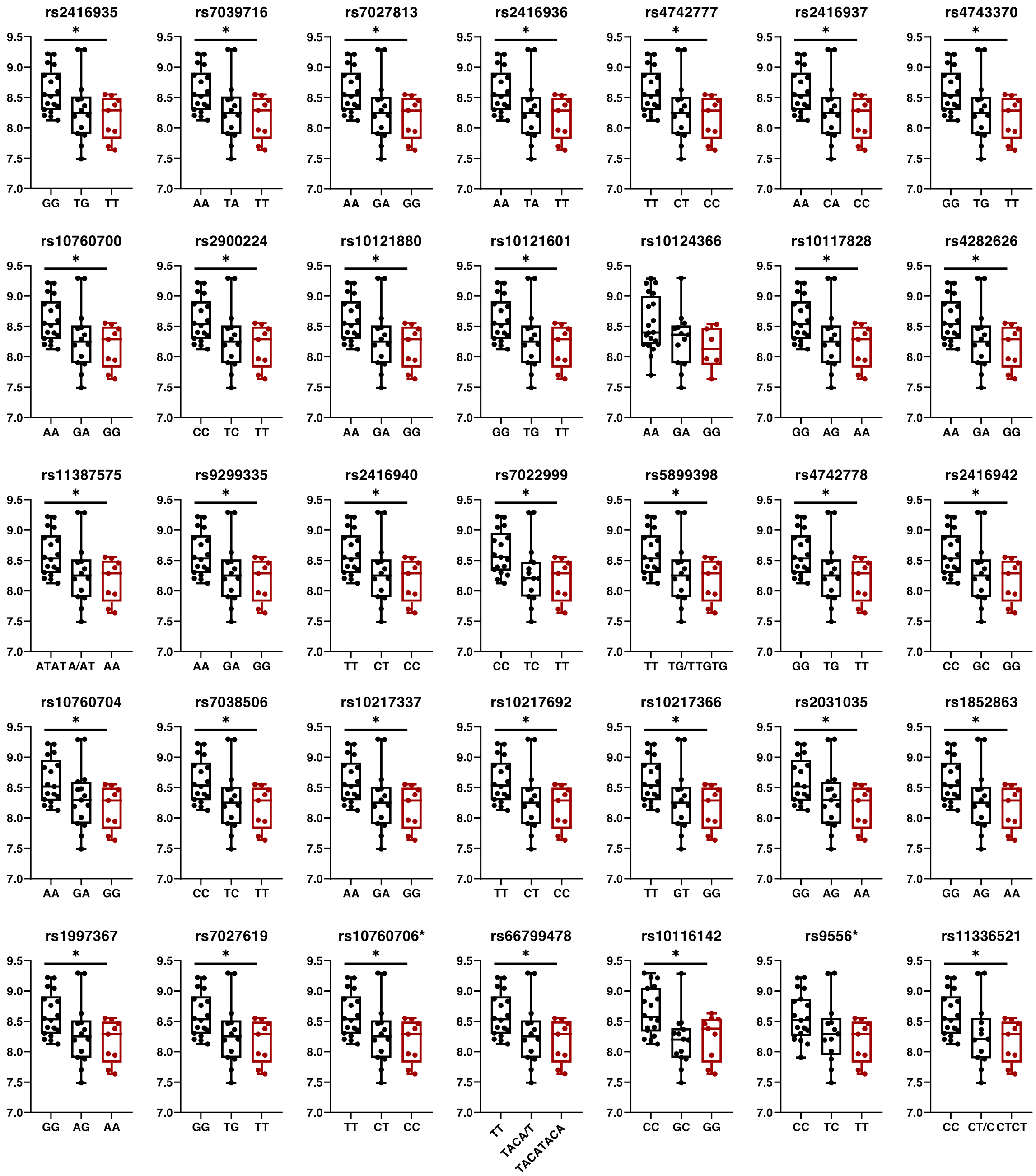


Figure 16 | 32 significant AA risk variants reduce STX17 expression in AA skin. Whole skin scalp RNA expression data and genotypes displayed as boxplots for 35 variants on the risk haplotype. 32 significant variants have an allele-specific effect on STX17 expression in AA whole skin with significantly decreased expression in homozygotes with risk allele (red). *=p<0.05. eQTL p-value in AA scalp biopsies was determined using a pairwise t-test with false discovery rate (FDR) correction. rs10124366 fails to reach both association and eQTL significance. rs10116142 is not significantly associated with AA but still has a significant genotype-dependent effect on skin expression. rs9556 is the hair color-associated SNP that was also significantly associated with AA but does not have a significant effect on skin expression; however, the allele-dependent expression reduction is trending decreased and may achieve significance with a larger cohort.

Figure 17 | *In silico* analysis predicts 3 regulatory variants in STX17 and implicates melanocytes as the disease-relevant cell type. a, FUN-LDA algorithm generated prediction scores (0-1, displayed as peaks) of tissue-specific regulatory effects for each genomic position spanning *STX17* (genomic position, x-axis) in 8 skin cell types. SNPs denoted as black vertical line. rs7039716 predicted as functional in skin cell types (falls within peaks across all 8 cell types). rs10760706 and rs9556 GWAS SNPs do not fall in predicted regulatory regions. Melanocytes (Mel03) show additional predicted regulatory regions (blue peaks) in STX17, not seen in keratinocytes or fibroblasts. **b,** Plot of Roadmap chromatin state data in fibroblasts (Fib), keratinocytes (Ker), and melanocytes (Mel) across the variant-defined risk haplotype. 35 AA-associated variants displayed as green ticks. Red represents transcription start sites (TSS) and yellow represents enhancers. Increase frequency of enhancer regions across the risk haplotype in melanocyte cells, specifically. Two variants (rs24169356 and rs7039716) fell in TSS across cell types and 1 variant (rs7027813) fell within a melanocyte-specific enhancer. **c,** Zoomed in image of region spanning intron 1 and exon 2 reveals rs7027813 residing in a melanocyte-specific enhancer and rs2416936 proximal to a melanocyte-specific enhancer (36 bp).

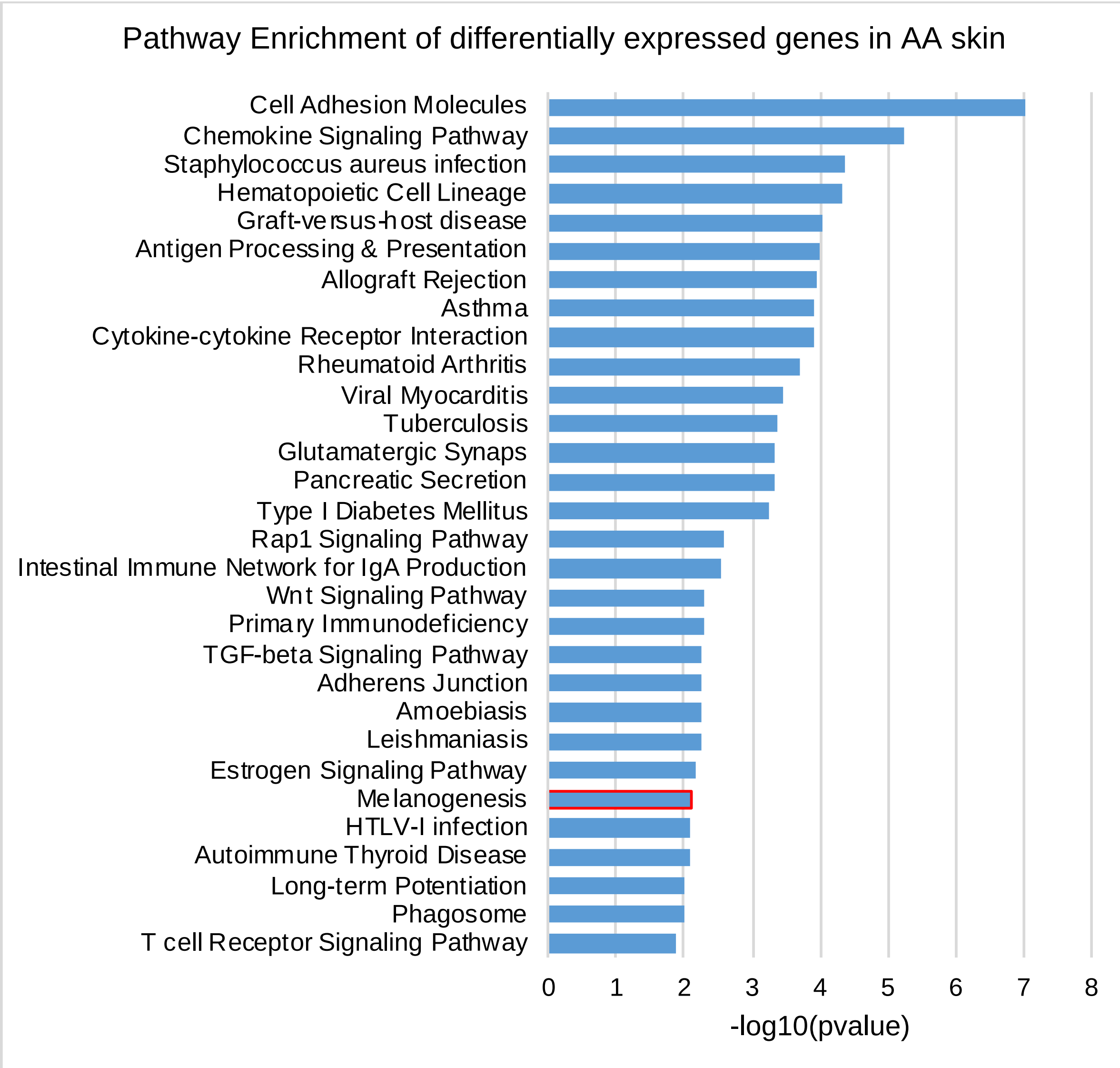


Figure 18 | DAVID Pathway analysis identifies melanogenesis pathway disrupted in AA skin. DAVID pathway analysis provided functional annotation of affected pathways based on differentially expressed genes from microarray expression data in the scalp of AA patients compared to controls. Melanogenesis is the 25th most significantly enriched pathway in the differential expression (DE) data with 30 DE genes and p-value=7.6E-03.

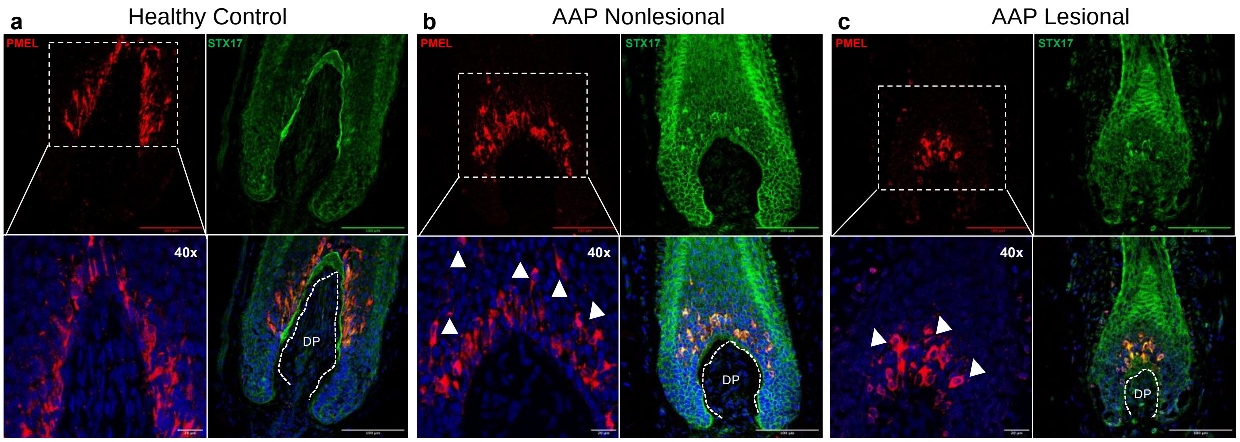


Figure 19 | STX17 localized to aberrant melanocytes in AA. Immunofluorescent staining of STX17 (green) and premelanosome protein (PMEL) to mark melanocytes in hair follicles of **a**, Healthy Control **b**, non-lesional biopsy of an Alopecia Areata Patchy (AAP) patient and **c**, lesional biopsy of same AAP patient. White box inset denotes zoomed in area of dysmorphic melanocytes at 40x magnification; white arrowheads denote detached, dysmorphic melanocytes; DP = dermal papilla. Strong STX17 staining of the control basement membrane (white dotted line) is lost with disease progression. In both lesional and non-lesional biopsies we see lack of strong STX17 basement membrane staining and increased detachment of melanocytes into the matrix.

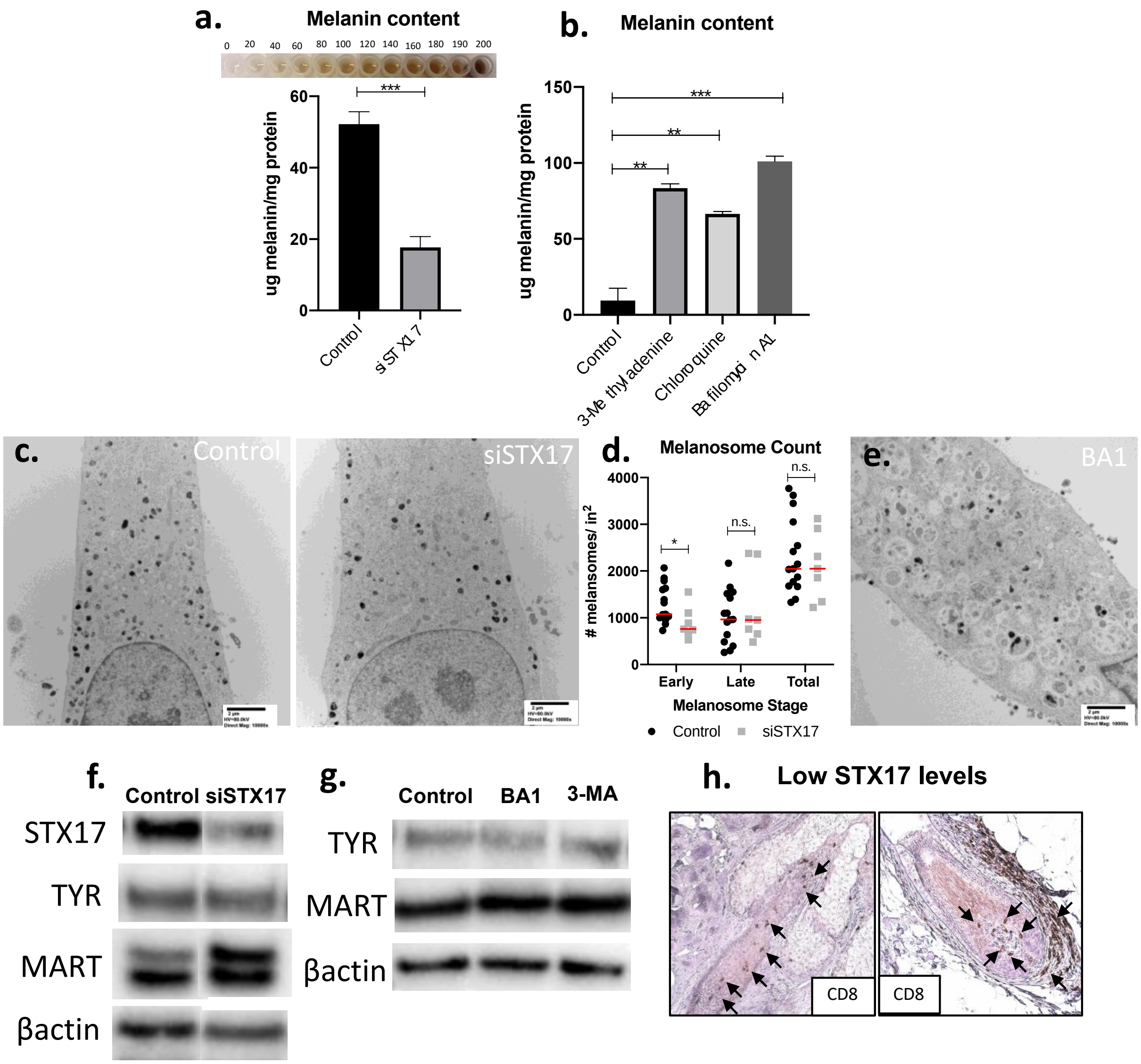


Figure 20 | STX17 has a novel autophagy-independent role in melanocyte function and antigen accumulation. a, Decreased melanin production in primary human melanocytes treated with STX17 siRNA. Melanin normalized to lysate protein levels (µg melanin/mg protein). Visual representation of µg melanin shown in top box. **b,** Melanin production increased in melanocytes treated with autophagy inhibitors 3-methyladenine (3-MA, 1 mM for 24hrs), chloroquine (CQ, 50 µM for 24hrs), and bafilomycinA1 (BA1, 10nM for 24hrs) compared to baseline. **c,** Electron microscopy (EM) of human melanocytes at baseline and after STX17 depletion reveal no striking differences in morphology and melanosome number/localization; 10,000x mag. **d,** In control and STX17 knockdown conditions, melanosomes were staged based on the amount of melanin and melanosome morphology. Quantification was performed blinded to conditions and determined by number of melanosomes per in^2 of cellular area. We observed a significant decrease in early melanosomes in the STX17 KD condition. However, late and total number of melanosomes are unchanged between conditions. P-value was determined by Welch's t-test (*=0.03) **e,** EM of BA1-treated melanocytes showed accumulation of autophagosomes and multi-vesicular bodies. **f,** Accumulation of established AA antigen MART1, an early melanosome protein in STX17-depleted melanocytes. TYR levels are unaffected. **g,** After autophagy inhibition with BA1 and 3-MA, TYR and MART1 levels are unchanged. **g,** CD8+ T cell infiltrate (arrows) in HFs of two AA patients (one AAP and one AT/AU) with low STX17 skin levels. The median level of STX17 expression in 123 scalp skin samples was 8.5. Low is defined as expression levels below the median.

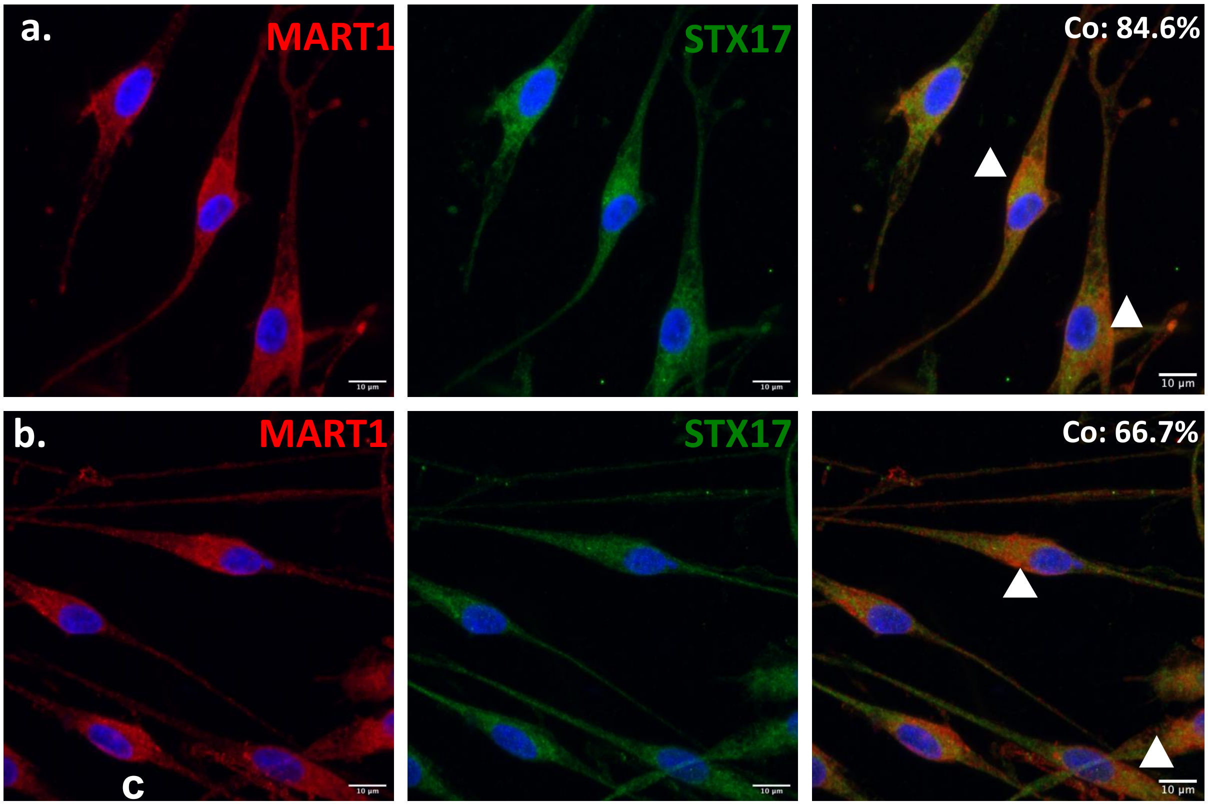


Fig. 21 | STX17 colocalizes with MART1 peri-nuclearly. a, Immunofluorescent staining of primary human melanocytes for MART1 (red) and STX17 (green) at baseline and **b,** after 1 µM αMSH treatment for 24hrs to stimulate melanogenesis. Arrowheads denote the co-localization of MART1 and STX17 in the perinuclear region. The percentage of co-localization (Co) was calculated in FIJI and represents the percentage of total MART1 that overlaps with STX17.

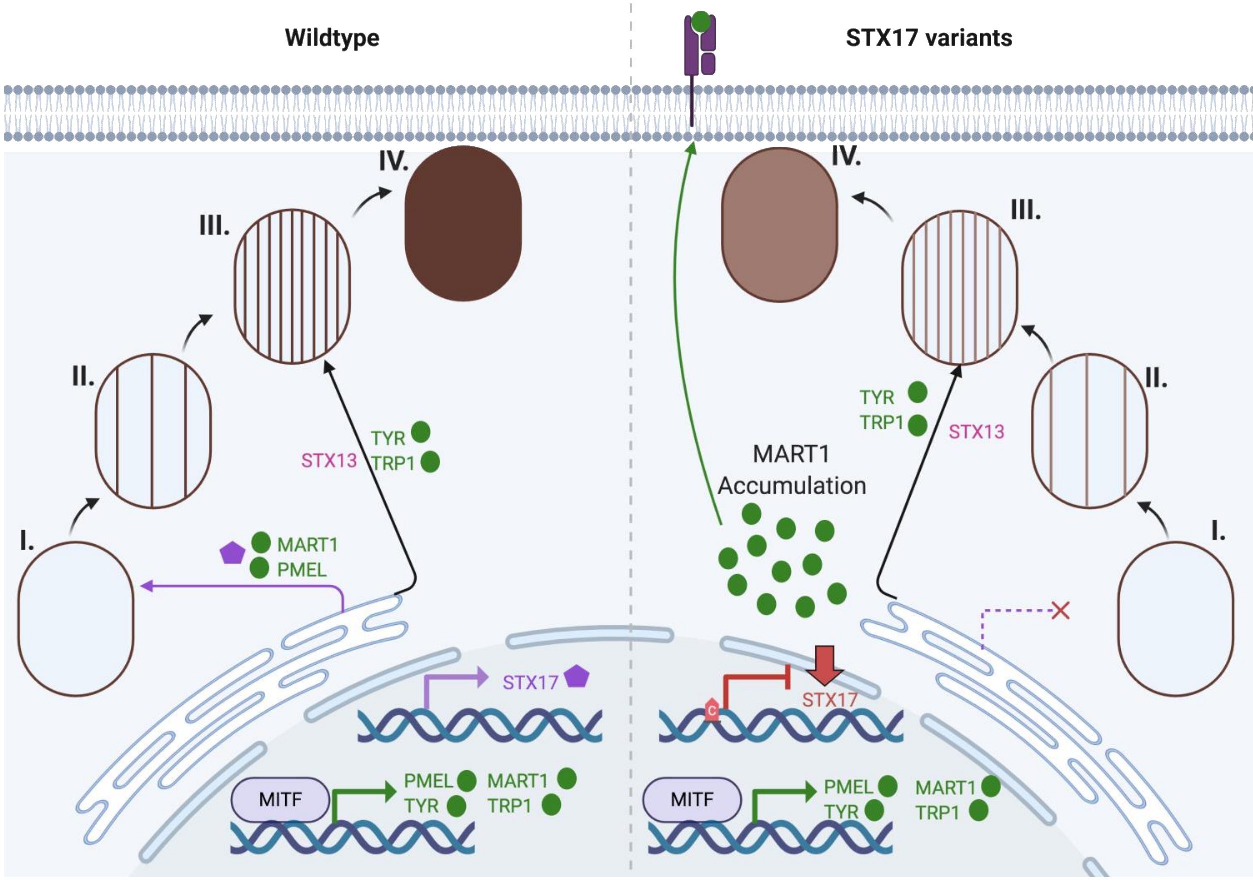


Figure 22 | STX17 mechanism in melanogenesis. Our data suggest a role for STX17 in trafficking of early protein MART1 to target melanosome organelles. In AA patients with the identified risk haplotype, variants inhibit transcription resulting in decreased STX17 levels. Reduced levels of STX17 result in inefficient MART1 enzyme transportation, consequentially hindering melanin production in melanosomes. MART1 antigens unable to properly traffic to the melanosomes accumulate in the cell and are aberrantly recognized by auto-reactive T cells, eliciting a T-cell response and subsequent autoimmune attack of the pigmented hair follicles in AA.

Chapter 3 Whole exome sequencing in alopecia areata identified mutations in KRT82

(Manuscript #2, in preparation)

3.1 Introduction

Alopecia areata (AA) is one of the most prevalent autoimmune diseases and is characterized by T-cell mediated immune attack of the hair follicle (HF) resulting in hair loss[55]. As a complex disease, both environmental and genetic factors contribute to AA risk[55,65,66]. Our genome-wide association study (GWAS) and meta-analysis identified common variation in 14 genetic risk loci, supporting the role for common variants in this complex disease. We previously performed linkage analysis in a cohort of families and found evidence for rare co-segregating mutations[34], however, the regions we identified were large and precluded the identification of causal genes. Thus, rare genetic variants with strong effect sizes remain uncharacterized in AA.

Whole exome sequencing (WES) and exome-wide association studies in large cohorts of patients with complex disorders such as congenital kidney malformations[119], amyotrophic lateral sclerosis[6], pulmonary fibrosis[7], epilepsy[8], coronary disease[9], myocardial infarction[10], and autism[11] have successfully identified rare disease variants using WES followed by gene-level collapsing approaches. As variant allele frequencies decrease, the power to detect associations simultaneously decreases due to the lack of observations to inform statistical analyses[44]. As a result, variant-level analysis in WES is limited by the need for large sample sizes to achieve power[133]. Therefore, gene-level collapsing methods have emerged as the accepted framework for identifying enrichment of variants in a given gene, or genetic burden, in disease cases compared to controls[119]. Model parameters, such as variant function and population frequency, are then used to select qualifying variants for testing. The frequency of qualifying variants within a gene is compared between cases and controls, and statistical significance is established to determine whether a gene has increased mutational burden in cases.

We performed WES and gene-based collapsing in 849 AA cases and 15,640 controls to assess the genetic burden of rare damaging mutations in unknown genes associated with AA.

3.2 KRT82 identified as an AA risk gene

In the first rare variant collapsing model, we performed a gene-based analysis by evaluating only rare, loss of function (LOF) variants with a minor allele frequency (MAF) ≤1% in the general population (gnomAD populations) and our sequencing cohorts (cases and internal controls)[121]. Stringent quality control (QC) metrics were further imposed on the variants. Using this approach across 18,653 protein-coding genes, we identified *Keratin 82* (*KRT82)* as the only genome-wide significant (at a significance level 0.05/18,653 = 2.68x10E-06) AA case-enriched gene (OR=4.04, p=2.03×10E−06), with variants detected in 19 out of 849 cases (2.24%) compared to 88 out of 15,640 controls (0.56%) (Figure 23a,d). *KRT82* is predicted to be tolerant to protein truncating variants due to the observed frequency of LOF variants in the general population (gnomAD, pLI=0; o/e=1.12)[121]. The tolerant nature of *KRT82* prompted us to perform an additional analysis of common LOF variants to ensure that rare *KRT82* LOF variants were truly enriched in our AA population, and not an artifact of genic tolerance and exclusion of common variants. The common (MAF >1%) LOF analysis did not identify additional mutations in *KRT82*, indicating that the rare model did not exclude any relevant common variants that would have dampened the significant enrichment of *KRT82* LOF mutations in our AA population. The rare LOF model also identified another hair keratin-related protein, *KRTCAP3* (*Keratinocyte Associated Protein 3*) as the second most significant AA association (p=1.28E-04) with rare LOF qualifying variants in 9/849 (1.1%) AA cases and 30/15,640 (0.19%) controls (Figure 23d). However, testing in a larger cohort may be needed to achieve a genome-wide significant association with AA.

Next, the second model expanded upon the previous model to include rare variants (MAF ≤1%) that were non-synonymous coding (missense) or canonical splice variants and predicted to be possibly or probably damaging by PolyPhen-2 HumVar[134]. The inclusion of damaging missense variants strengthened the evidence for *KRT82* association (p=9.19×10E−07, Figure 23b, d), and identified qualifying variants in 47 out of 849 cases (5.54%) compared to 376 out of 15,640 controls (2.4%; OR=2.38). This indicated that the *KRT82* association with AA was evident both when protein expression was decreased due to protein truncating variants (LOF model), as well as when protein structure was altered by a missense variant (LOF + missense model).

We postulated that if *KRT82* was enriched for both LOF and damaging missense variants, a stronger signal would be observed with the inclusion of qualifying synonymous and splice region variants predicted to be pathogenic. We used the Transcript-inferred Pathogenicity score (TraP-score, version 2.0)[18], which predicts whether a synonymous or intronic variant can damage the mRNA transcript by disrupting the splicing process, resulting in perturbed protein formation and subsequent disease risk[135-137]. We constructed an additional splicing model (Figure 23c-d) that expanded on the previous LOF and missense models to also include predicted pathogenic splice variants and non-synonymous variants with a TraP-score ≥ 0.2, which is above the 90th percentile of pathogenic scores and considered possibly damaging (http://trap-score.org/about.jsp). Notably, this model also identified *KRT82* as harboring the most significantly enriched variants in AA cases, achieving a stronger enrichment of p=2.2×10E−07, OR= 2.41 (Figure 23c-d) and surpassing genome-wide and study-wide thresholds for significance (α=2.68E−06 and 3.83E−07 respectively). The increased significance was due to an additional splice region variant (Arg314Gln) that had a high TraP-score (0.55) prediction as protein-damaging and was found in four additional cases. As a result, this splicing model identified a total of 51 of 849 AA patients (6.01%) with rare damaging variants in *KRT82*, compared to 404 out of 15,640 controls (2.58%).

We found that the genomic inflation factor (λ), which measures potential false-positive associations, varied across these models in a manner that is consistent with the mutation type that is included in each model. For example, LOF variants have a higher probability of exerting a phenotypic effect than other mutation types[138,139]. Accordingly, we observed the lowest estimate of genomic inflation for the LOF model (λ=1.03) (Figure 23a). We found greater estimates for the LOF + missense analysis (λ=1.18) and the splicing analysis (λ=1.23), models for which we expected a potentially higher proportion of clinically benign qualified variants (Figure 23c). To investigate further, we tested an additional model as a negative control, using rare synonymous variants as qualifying criteria (MAF ≤ 0.01%). As expected, we did not observe any genes with significant enrichment. The genomic inflation factor (λ) for this model was 0.91 (Figure 24). Taken together, the overall low inflation of the cohort in the synonymous analysis represented the null distribution and absence of global inflation due to genotyping discrepancies or population stratification. It provided further confidence for a true association between rare, damaging *KRT82* mutations and AA.

Qualifying variants in these gene-collapsing models included predicted damaging mutations such as stop-gained (nonsense), frameshift, splice site, and nonsynonymous damaging amino acid substitutions (missense) (Table 2). Eleven protein-altering variants (3 nonsense, 7 damaging missense, 1 splicing) accounted for the variation detected in 51 of our AA cases (Table 2). One *KRT82* mutation with the highest frequency in our AA cohort was found in 15 cases (1.77%) and resulted in a nonsense mutation at amino acid position 47 (R47X). Furthermore, 7 out of the 11 variants were shared among more than one individual. The pathogenicity of these individual variants was supported by their recurrent nature and increased frequency in our disease cohort, even though the mutations were rare (MAF ≤ 1%) in the general population. However, none of the individual variants could exceed the genome-wide significance threshold on their own, underscoring the necessity of gene-level analysis in obtaining sufficient power to identify meaningful disease risk genes.

3.3 AA KRT82 variants annotated as likely disease-causing

The identification of *KRT82* as an AA risk gene emphasized the utility of WES in uncovering functionally relevant disease loci in complex diseases. *KRT82* encodes a type II hair-specific keratin with exclusive expression in the cuticle of the hair shaft[140-142]. Keratins are intermediate filaments (IF) that provide a cellular structural network via the heterodimerization of type I with type II keratins in epithelial cells[89]. The keratin family of proteins are composed of well-defined evolutionarily conserved domains consisting of a head domain, a 310-amino acid α-helical rod domain, and a tail domain[143]. The rod domain is necessary for forming the coiled-coil α-helical dimer between type I and II keratins, and is made up of 4 subdomains, 1A, 1B, 2A, 2B[143,144]. Additionally, the keratin II head domain is crucial for tetramer stabilization and IF assembly[145]. Variants that disrupt the 1A, 2B, or type II head and tail domains in keratins have been reported as pathogenic in various other diseases, underscoring the importance of these domains in proper keratin function[110,143].

We identified 11 rare, damaging variants in 51 patients that contributed to the *KRT82* association with AA (Table 2). Strikingly, 9 out of the 11 *KRT82* variants fell within one of the keratin disease-associated subdomains, consistent with the predicted damaging effect of these variants (Figure 25a-b). Three *KRT82* variants resided in the head domain (Ser16fs, Arg47X, Ile92Val), one in coil 1A (Asn129Lys), three in coil 2B (Arg314Gln, Arg314Trp, Glu351Lys), and two in the tail domain (Gly436Arg, Gly436Trp) of *KRT82* (Figure 25a-b). Since these regions are essential for keratin function and dimerization, we postulated that the *KRT82* AA risk variants likely disrupt KRT82 function and dimerization with its respective potential binding partners (KRT32, KRT35, KRT39, or KRT40), consequentially destabilizing the hair shaft cuticle. Therefore, damaging variants in these well-defined disease-associated regions have the potential to impair normal IF assembly and disrupt the integrity of the hair shaft cuticle in AA cases.

Additionally, all identified damaging variants fell within evolutionarily conserved residues across distantly related species, further supporting the importance of these amino acids in KRT82 biology (Figure 25c). Highly conserved genes are often considered intolerant to mutations and undergo negative purifying selection against deleterious alleles[146]. Since important functional residues are often evolutionarily conserved across species, variants occurring at these positions are more likely to be pathogenic[146-148]. The enrichment of *KRT82* variants in established functional disease domains and evolutionarily conserved amino acids supported a damaging effect of these variants on KRT82 function.

3.4 KRT82 variants associated with AA pathology

Within the large family of epithelial keratins, hair keratins (type I and type II) are a subfamily responsible for the "hard" keratinization of hair and nails[89,140]. Interestingly, mutations in type II hair keratins (*KRT81-86*), but not type I (*KRT31-40)*, have reported associations with other genetic hair disorders such as Monilethrix, pseudofolliculitis barbae, and hair ectodermal dysplasia [108,110,142]. Specifically, mutations in the 1A and 2B domains of *KRT81, KRT83*, and *KRT86* resulted in Monilethrix, a rare monogenic hair loss disorder (Figure 25a)[108,143]. The causal role of type II hair keratins in genetic hair loss diseases validated their important role in healthy hair structure and further strengthened the functional association between AA and *KRT82*, a type II hair keratin. Due to the known roles of other type II hair keratin genes in other monogenic hair loss disorders, we next confirmed that our patients were accurately diagnosed with AA, and not a monogenic hair loss disorder misdiagnosed as AA in our cohort. We confirmed accurate AA diagnoses through clinical validation, biopsy immunohistochemistry (IHC), and inflammatory ALADIN gene expression signatures (Figure 26a-c). In four AA patients with a *KRT82* missense mutation, IHC of scalp samples revealed interfollicular infiltrate of CD8+ lymphocytes, a diagnostic feature of AA (Figure 26b). The ALADIN (Alopecia Areata Disease Severity Index) index utilizes expression profiles of keratin and inflammatory gene signatures to distinguish AA scalp from control[124]. Gene expression data from three patients with a *KRT82* mutation confirmed an increased inflammatory ALADIN signature, with the patients clustering in the AA cohort and away from controls (Figure 26c). Using these three methods, we were confident that the patient's diagnosis of AA was accurate, and not a form of a monogenic hair loss disorder.

3.5 AA susceptibility was specific to KRT82 variation

Due to close chromosomal localization and structural similarities in the keratin gene family, we tested whether the observed increased variation in *KRT82* was due to hypermutability in the chromosomal region or keratin family of genes. We applied the same model parameters to specifically investigate variant frequencies in 7 other keratin genes (*KRT5, KRT84, KRT85*, *KRT32, KRT35, KRT39, KRT40*) located both proximally and distally to *KRT82* (Figure 27a-b). *KRT84* and *85* are type II hair keratins located 8.3 kb and 26.4 kb upstream of *KRT82*, respectively. *KRT5* is a type II epithelial keratin that is located on chromosome 12, 120.6 kb away[149]. *KRT32, KRT35, KRT39, KRT40* are all type I hair keratins and proposed binding partners of *KRT82* due to their reported co-expression patterns in the HF cuticle[89]. Each of these keratin genes harbored markedly less variation than *KRT82*. For example, only 1 AA patient carried a LOF mutation (*KRT85*) out of the 7 of the additional keratin genes queried, whereas 19 AA cases harbored a *KRT82* LOF mutation (Figure 27b). Furthermore, none of the additional keratin genes that we investigated carried a significantly higher genetic burden of rare variants in cases compared to controls. The lack of significant AA-associated variation in any of the putative binding partners of *KRT82* or keratin family members suggested a specific role for *KRT82* in the genetic pathomechanism in AA. Additionally, these results confirmed that the AA-associated risk is driven specifically by *KRT82* variation and was not a result of broad hypermutability across the keratin gene family or chromosome block.

3.6 Haplotype analysis identified a hotspot mutation R47X in KRT82

We next asked whether the *KRT82* mutation R47X was a “hotspot” versus a founder mutation that was dispersed across the population. Some regions of the genome are “hotspots” that have a propensity to spontaneously give rise to *de novo* mutations, increasing the frequency of disease mutations in the population. For example, another dermatologic disorder, epidermolysis bullosa simplex (EBS), is associated with hotspot mutations in keratin genes *KRT5* and *KRT14* that result in ineffective cross-linking and bundle formation[150-152]. For a hotspot mutation, the same mutation will be carried on multiple different haplotypes, whereas a founder mutation will be found on one common haplotype, suggesting the presence of a common ancestor. To distinguish these possibilities, we analytically phased our sequence data to identify haplotypes carrying *KRT82* mutations[153]. We found that the most frequently occurring nonsense mutation, R47X was present on at least 6 different haplotypes, indicative of a hotspot mutation (Figure 28, Table 3).

Disease-associated mutational hotspots commonly occur at CpG dinucleotides that result in spontaneous deamination of 5-methylcytosine in C > T transitions[154-158]. Interestingly, the R47X mutation in KRT82 resulted from a C > T transition at a methylated CpG residue (Figure 28b). In this context the 5’-CGA-3’ encoded an arginine (Arg; R) that resulted in a nonsense mutation when 5’-methylcytosine was deaminated to thymine (T). As a result, R47X, the most frequent *KRT82* mutation in AA patients (15/849; 1.77%), is predicted to be a hotspot mutation arising from a spontaneous deamination at a methylated CpG site.

3.7 KRT82 is exclusively expressed in anagen phase

Previous studies reported the expression of KRT82 in the hair shaft cuticle of both human and sheep hair follicles and identified it as a hair-specific type II keratin[140-142]. However, its functional role in hair follicle biology has yet to be elucidated. Cell culture studies investigating the role of KRT82 are challenging due to its differentiated nature as a hair follicle keratin. Therefore, we focused functional analyses on *in vivo* mouse models, to characterize the spatiotemporal expression of KRT82 throughout the murine hair follicle cycle, which consists of the three stages including anagen (growth), catagen (regression), and telogen (quiescence). We found that KRT82 was expressed in the hair follicle cuticle exclusively during the growth stage of anagen (Figure 29) and was absent in catagen and telogen. Thus, KRT82 expression in the anagen HF is biologically relevant in the context of AA, since the autoimmune attack is known to occur exclusively during anagen, resulting in the premature entry of the HF into catagen.

3.8 Reduced expression of KRT82 in AA patients

We next investigated the expression of KRT82 in control human hair follicles using immunofluorescence (IF) staining and found strong expression of KRT82 in the anagen hair shaft cuticle, as previously reported (Figure 30a). In control HFs, KRT82 cuticle expression began in the bulb, just above the Line of Auber, and extended up the HF towards the epidermis. We found specific subcellular KRT82 localization in the cytoplasm of the hair shaft cuticle cells, consistent with the role of keratins as cytoskeletal intermediate filaments (IF)[143]. We compared KRT82 expression in control versus AA patient HFs and found markedly decreased expression of KRT82 in AA hair shaft cuticles (Figure 30a). In contrast to control HF, KRT82 expression in AA follicles was absent in the cuticle including the bulb, the site of AA immune attack. This finding is noteworthy in the context of AA disease onset due to the characteristic immune infiltrate surrounding the follicular bulb, where we identified loss of functional KRT82 expression in AA HF.

In support of this observation, our previous gene expression studies reported a significant -4.5 log-fold decrease in expression of KRT82 in the scalp skin of AA patients compared to controls (Benjamini-Hochberg-corrected p=3.55E-10)[124]. We found that not only was KRT82 expression significantly decreased in AA patients relative to controls, but KRT82 downregulation was also significantly correlated with disease severity (Figure 30b). Biopsies taken from nonlesional (hair-bearing) sites of AA Patchy (AAP) scalps revealed a slight reduction in KRT82 expression compared to controls ($p=0.028$) (Figure 30b). In lesional sites of hair loss in AAP patients and patients with total loss of scalp and body hair (AT/AU; alopecia areata totalis/universalis), KRT82 was significantly decreased compared to healthy controls ($p<0.0001$) (Figure 30b). The insignificant change in KRT82 levels in AAP nonlesional compared to lesional biopsy sites ($p=0.068$) suggested that decreased expression observed in patients was not simply a result of loss of the KRT82-bearing HF, but rather a predisposed reduction in functional KRT82 expression that strongly correlated with active disease sites and disease severity (Figure 30b). However, it cannot be ruled out that significant differences between nonlesional and lesional samples may be observed with a larger sample size. Our results suggested that decreased functional KRT82 in the hair cuticle may predispose genetically susceptible individuals to AA-mediated hair loss.

To further investigate the role of the identified rare, damaging variants on skin tissue expression of *KRT82*, we tested whether individuals with qualifying *KRT82* variants demonstrated altered KRT82 levels in the scalp skin independent of the presence of hair follicles (i.e. nonlesional & lesional biopsies) in 20 AAP individuals. Patients with a confirmed damaging mutation in *KRT82* trended towards lower expression levels of KRT82, independent of biopsy type (Figure 30c), although the difference was not statistically significant due to the small sample size of patients with a damaging variant and expression data. This suggested that the identified damaging mutations functionally altered/truncated KRT82 protein structure, resulting in decreased functional expression in the skin of AA patients.

We next characterized the functional effect of *KRT82* variants in the HF by assessing KRT82 and CD8 expression in lesional scalp biopsies of 3 patients with one of the damaging *KRT82* mutations. Lesional biopsies from two of the patients with a missense *KRT82* mutation (Asn129Lys, Gly436Trp) revealed reduced or absent expression of KRT82 in the HF cuticle

(Figure 30d). Both patients also had increased CD8+ T cell follicular infiltrate compared to control HFs. In the third patient with a missense *KRT82* mutation (Arg302His), we observed HFs with cuticular KRT82 expression intact, however, with less severe perifollicular CD8+ T cell infiltrate (Figure 30d) compared to the patients with absent KRT82. These findings suggested a model in which rare damaging mutations in *KRT82* resulted in reduced protein function or expression, predisposing AA patients to compromised HF cuticles, and rendering the HF vulnerable to autoimmune attack, supported by the observed negative correlation between perifollicular CD8+ T cell infiltrate and KRT82 HF cuticle expression.

3.9 Discussion

This study presents the first large-scale WES of AA patients and identification of novel rare variants contributing to disease risk. Our analysis approach utilized a gene-level collapsing approach to identify a type II hair keratin gene, *KRT82*, as the only gene harboring significantly more rare damaging variants in AA cases (6.01%) compared to controls (2.58%) (p=2.18E-07), achieving both genome-wide and study-wide significance. The predicted damaging effects of the variants were supported by their localization to evolutionary conserved residues and known disease-causing keratin domains.

In order to define the underlying mechanism for the association between *KRT82* and AA, we interrogated the role of KRT82 in hair follicle biology. We observed that KRT82 is expressed in the hair shaft cuticle exclusively in the growth phase of anagen, when AA-directed immune attack occurs. During disease, anagen-directed inflammatory infiltrate prematurely propels the anagen HF into catagen. The absence of KRT82 in AA hair shaft cuticles may cause HF dysfunction and subsequent premature entry into catagen, as seen in disease. In support of this, hair follicles in *KRT17*-null mice exhibit hair shaft fragility, aberrant apoptosis, and premature entry into catagen[159]. In the absence of *KRT17*, keratinocytes are more susceptible to TNFα-mediated apoptosis and HFs prematurely undergo the anagen-catagen transition, demonstrating a critical role of a single keratin, *KRT17,* in hair cycle regulation and suggesting a plausible role for other keratins, such as *KRT82* in regulation of apoptosis and HF cycling. Further support for the role of *KRT82* in hair loss is seen in the *Sox21* knockout (*Sox21*$^{-/}$*)* mouse model of cyclic alopecia that demonstrated altered keratin expression signatures, including a prominent reduction of KTR82 in the hair shaft cuticle[141]. The resulting phenotype led to hair loss during late first telogen and catagen stages. Although other keratin levels were affected in this animal model, the decrease in KRT82 and association with cyclical alopecia substantiates an important role for *KRT82* in hair follicle biology, underscoring the role for LOF and damaging variants in AA pathogenesis.

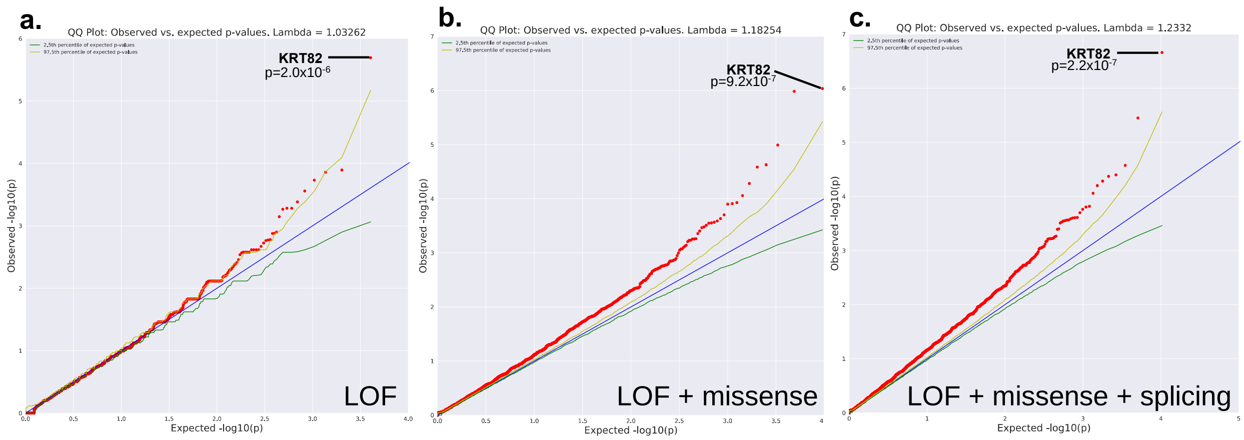
Further functional annotation of *KRT82* revealed that its expression was largely reduced in the scalp skin biopsies and HFs of AA patients. All 51 AA patients in which mutations were detected are heterozygous carriers of damaging risk variants, which correlated with lower levels of KRT82 expression supporting a haploinsufficiency model. The absence or reduced expression of KRT82 in the HF, in addition to increased perifollicular CD8+ T cells support an important relationship between KRT82 function and the organ-specific AA-mediated immune attack. We postulate that hair shaft cuticle formation is dependent on KRT82 and absence of normal expression renders the bulb vulnerable to infiltration of auto-reactive immune cells.

As the outermost layer of the hair shaft, the hair cuticle protects the internal layers of the HF against mechanical and environmental insults through a high content of crosslinking between type I and type II keratins[160]. It also plays a crucial role in anchoring the growing hair shaft to the inner root sheath (IRS) during anagen[160]. Disrupted and absent cuticles are seen in exclamation mark hairs (EMH), a characteristic diagnostic feature of AA[56]. Previous reports demonstrated that absent cuticles in AA EMHs revealed areas of high shaft disorganization[56]. Additionally, some hairs revealed structural cuticle deformities with a high density of vacuoles[56]. The authors suggested that the cuticle disruption negatively impacts AA HF integrity, causing vulnerability to mechanical insults[56]. Our findings of a genetic association between damaging

KRT82 variants with AA and the literature reports of aberrant cuticles in AA EMHs supports a critical role for the hair shaft cuticle in AA pathogenesis. Additionally, we found that *KRT82* mutations were predicted to perturb the protein product, potentially resulting in disorganized or absent hair shaft cuticle structures.

Although the target antigens in AA have not been conclusively identified, keratins are attractive candidate auto-antigens due to their abundant expression in the HF. Auto-antibodies against anagen keratinocytes in the cuticle, matrix, and cortex were previously identified in a subset of autoimmune polyendocrine syndrome type 1 patients with alopecia areata totalis (AT)[161]. Human and mouse AA auto-antibody reactivity was also reported against anagen HF-specific antigens ranging from 40-60 kDa, specifically 44/46 kDa keratins[95,161,162]. Type I keratins are 44-48 kDa and dimerize with type II keratins (55-60 kDa) to form IF bundles and structural networks in cells[89,140]. In AA, absence of the type II keratin, KRT82 in the cuticle would leave its type I binding partner (KRT32, KRT35, KRT39, or KRT40) unable to heterodimerize. We postulate that without functional KRT82, unbound type I binding partners could be aberrantly presented as autoantigens. However, further work on the potential AA keratin auto- antigens remains to be done.

Here, we used WES and gene-level mutation analyses to identify a functionally relevant disease gene, *KRT82*. We identified eleven rare damaging variants, including the most frequent, R47X, a mutational hotspot, and characterized negative consequences on KRT82 expression and function in AA. Our work uncovered a novel mechanism involving genetic susceptibility for disrupted cuticle structures rendering susceptibility to AA. This work not only elucidates a previously unknown mechanism of end organ genetic susceptibility in an autoimmune disease but also encourages potential new therapeutic approaches for treatment of AA. Treatments targeted at restoring HF integrity are advantageous due to the distinct ability of the HF to regenerate in AA, since the target organ is damaged, but not destroyed. In summary, we characterized the role of rare variants in AA susceptibility and identified a novel disease mechanism for hair shaft cuticle integrity in pathogenesis.

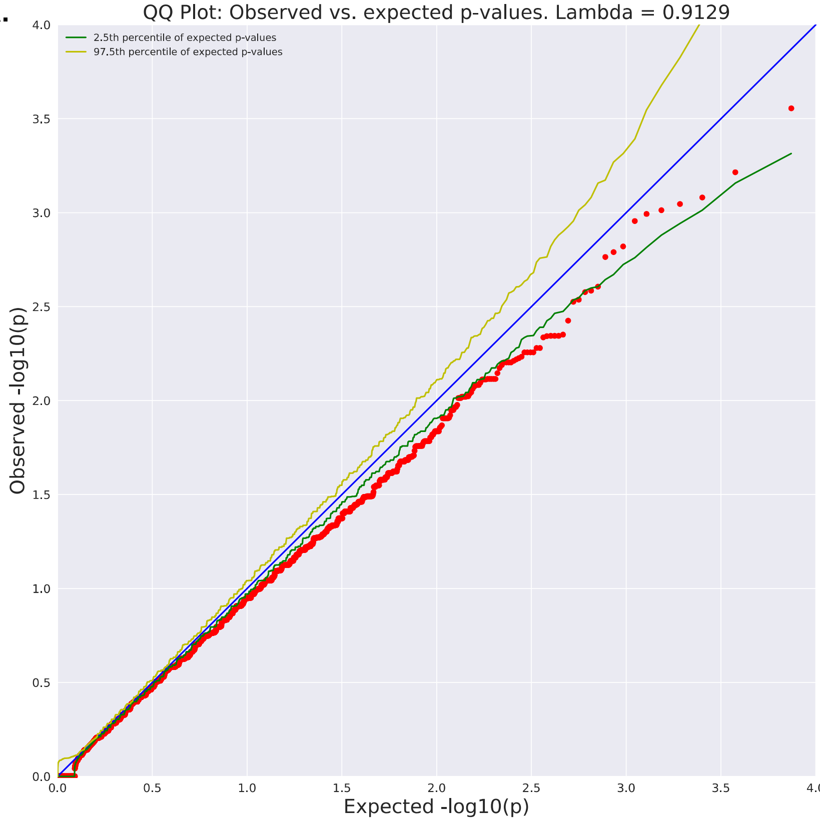


d.

Model	Gene	Cases	Controls	Case Freq	Ctrl Freq	P-value
LOF	**KRT82**	**19**	**88**	**2.2%**	**0.56%**	**2.03E-06**
	KRTCAP3	9	30	1.1%	0.19%	1.28E-04
	DECR2	6	11	0.7%	0.01%	1.39E-04
LOF + missense	**KRT82**	**47**	**376**	**5.5%**	**2.4%**	**9.19E-07**
	ZNF418	0	271	0%	1.7%	1.04E-06
	PCCB	2	321	0.2%	2.1%	1.02E-05
LOF + missense + splicing	**KRT82**	**51**	**404**	**6.0%**	**2.6%**	**2.18E-07**
	NCOA3	4	419	0.4%	2.7%	3.54E-06
	IGFBL1	0	214	0%	1.4%	2.67E-05

Figure 23 | Gene-level collapsing models identify significant rare variation in KRT82. Q-Q plots of **a,** Loss of Function (LOF) Model. **b,** LOF + damaging missense Model. **c,** LOF + missense + splicing variants predicted by high TraP scores (≥ 0.2). The genomic inflation factor (λ) is 1.03, 1.18, 1.23, respectively. The blue line represents the null hypothesis (observed p-values correspond to expected p-values), the yellow line marks the 2.5th percentile of expected p-values, and the green line marks the 97.5th percentile of expected p-values. **d,** Table of three models that identified rare variants in *KRT82* that were enriched in AA cases with genome-wide significance (p>2.68E-06). Each model lists the top three genes with highest significance (Gene Column). Table outlines the number and frequencies (Freq) of cases and controls with qualifying model variants in each gene, and their corresponding p-value, determined by two-tailed Fisher's exact test (FET).

a.



b.

Model	Gene	Cases	Controls	Case Freq	Ctrl Freq	P-value
Synonymous	AQP2	6	13	7.1%	0.08%	2.79E-04
	OTOF	0	151	0.0%	0.97%	6.09E-04
	HS1BP3	7	24	0.8%	0.15%	8.30E-04

Figure 24 | Rare but Synonymous Variant Gene Collapsing Model. Qualifying variants in this model included any rare (MAF ≤ 0.01%) and synonymous variants, that are predicted to have no pathogenic role. **a,** Low inflation (lambda (λ) = 0.91) of representative Q-Q plot. **b**, No genes reached genome-wide significance in rare synonymous variant model, serving as a negative control. Synonymous variants do not affect the protein product, and therefore should not have a pathogenic effect or be associated with disease. Under this rationale, distribution of p-values should follow the null distribution without any any significant associations, as seen here.

Type	Cases	Controls	Case Freq	Ctrl Freq	P-value
LOF	**19**	**88**	**2.24%**	**0.56%**	**2.03E-06**
Arg47X	15	72	1.77%	0.46%	3.42E-05
Ser272X	3	3	0.35%	0.02%	2.42E-03
Ser16fs	1	3	0.12%	0.02%	0.19
LOF + missense	**47**	**376**	**5.54%**	**2.40%**	**9.19E-07**
Asn129Lys	14	101	1.65%	0.65%	2.34E-03
Ile92Val	4	46	0.47%	0.29%	0.33
Glu351Lys	4	37	0.47%	0.24%	0.16
Arg314Trp	3	22	0.35%	0.14%	0.14
Arg302His	1	7	0.12%	0.04%	0.34
Gly436Arg	1	2	0.12%	0.01%	0.15
Gly436Trp	1	2	0.12%	0.01%	0.15
LOF + missense + splicing	**51**	**404**	**6.01%**	**2.58%**	**2.18E-07**
Arg314Gln	4	28	0.47%	0.18%	0.08

Table 2 | Damaging KRT82 amino acid changes in AA cases. List of the 11 qualifying variants present in AA cases, sub-divided by variant types included in each of the three models. Each model includes qualifying variants from the previous model. For example, LOF + missense Model includes the 3 LOF variants in 19 cases, plus 7 missense variants in 28 new cases, for a total of 10 variants in 47 cases. Qualifying variants exclusive to cases (n=50) were not individually listed in this table but accounted for the total number of controls with a qualifying variant (Controls column). As an example, 88 controls had a LOF variant in *KRT82* but the Arg47X, Ser27X, and Ser16fs variants account for 78 individuals. 10 controls harbored qualifying *KRT82* LOF variants that were not present in AA cases and therefore not listed in this table.

a.

Disease Domain	No. KRT82 AA Variants	AA Variant Types	Domain-associated Disease	References
Type II Head	3	Ser16fs, Arg47x, Ile92Val	EBS, DNEPPK	Smith et al. (2003)
Coil 1A	1	Asn129Lys	BCIE, DNEPPK, IBS, MECD, WSN, EBS, PC-1, FNEPPK, PC-2, SM, EPPK, M	Smith et al. (2003)
Coil 2B	3	Arg314Gln, Arg314Trp, Glu351Lys	BCIE, DNEPPK, IBS, MECD, EBS, PC-1, FNEPPK, PC-2, EPPK, M	Smith et al. (2003)
Type II Tail	2	Gly436Arg, Gly436Trp	Ichthyosis Hystrix	Smith et al. (2003)

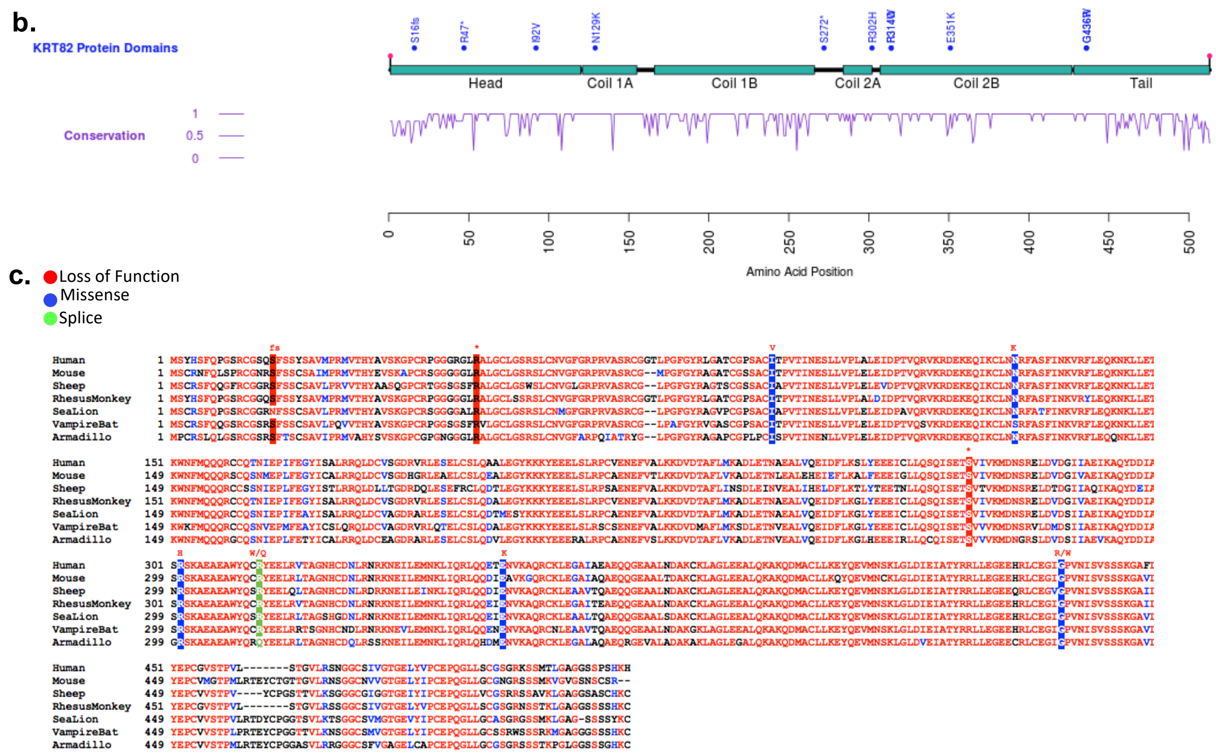


Figure 25 | KRT82 variants are localized to highly conserved disease-annotated domains. **a**, Table of 9 of the KRT82 damaging mutations that fall in a domain previously associated with pathogenic keratin mutations in other diseases. **b**, schematic representation of *KRT82* with domains (turquoise), AA variants (blue), and conservation score (purple) (1=highly conserved). **c,** Conservation of *KRT82* amino acids across 7 different species: human, mouse, sheep, rhesus monkey, sea lion, vampire bat, and armadillo. AA variants are highlighted in red (LoF), blue (Missense), and green (Splicing). Amino acid change of variant is annotated in red above. **Abbr**: EBS= epidermolysis bullosa simplex; CC= Cryptogenic Cirrhosis; DNEPPK= diffuse non-epidermolytic palmoplantar keratoderma; BCIE= bullous congenital ichthyosiform erythroderma; IBS= ichthyosis bullosa of Siemens; MECD= Meesmann's epithelial corneal dystrophy; WSN= white sponge nevus; PC=pachyonychia congenita; FNEPPK= focal non-epidermolytic palmoplantar keratoderma; SM= steatocystoma multiplex; EPPK= epidermolytic palmoplantar keratoderma; M= monilethrix.

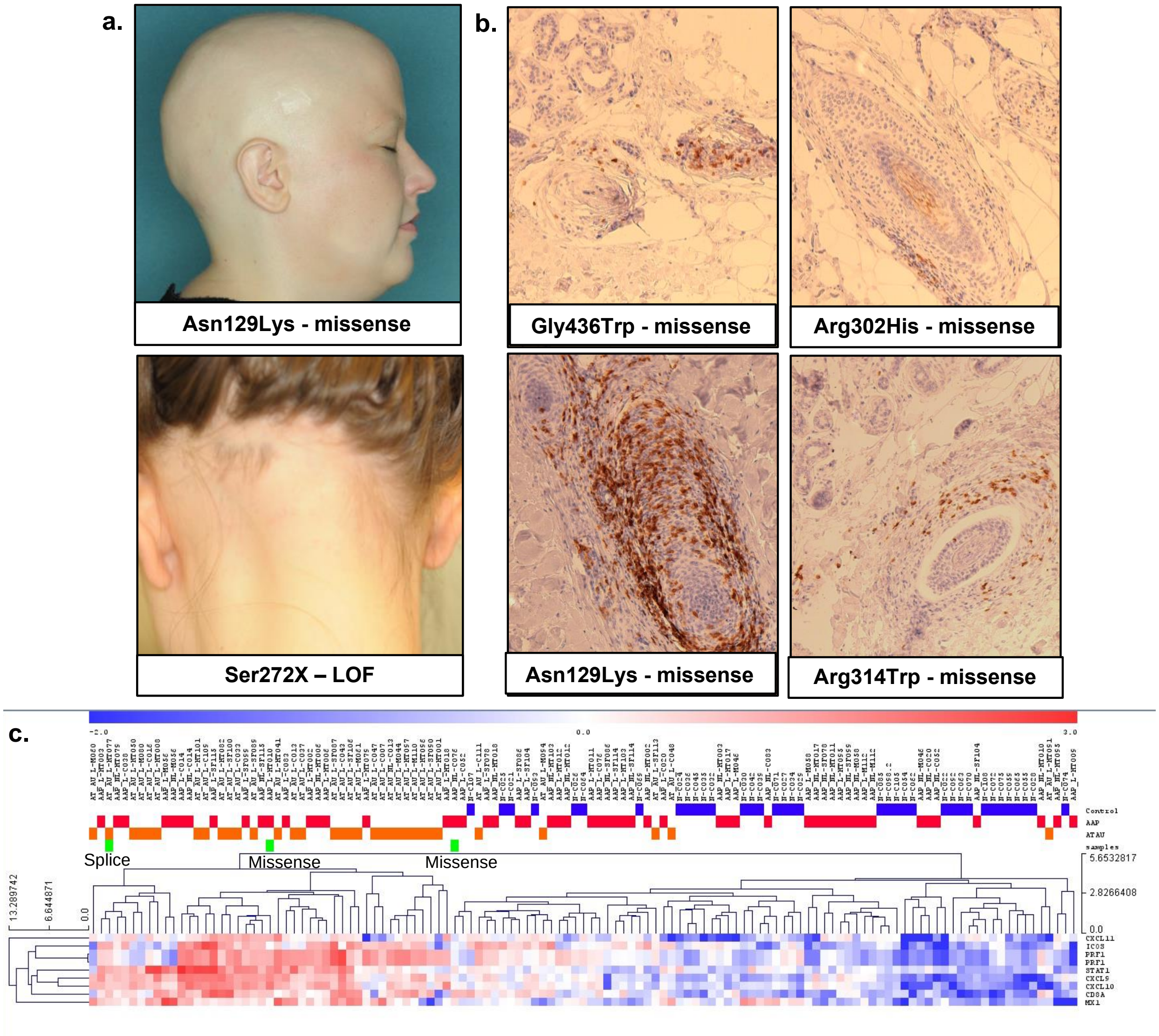
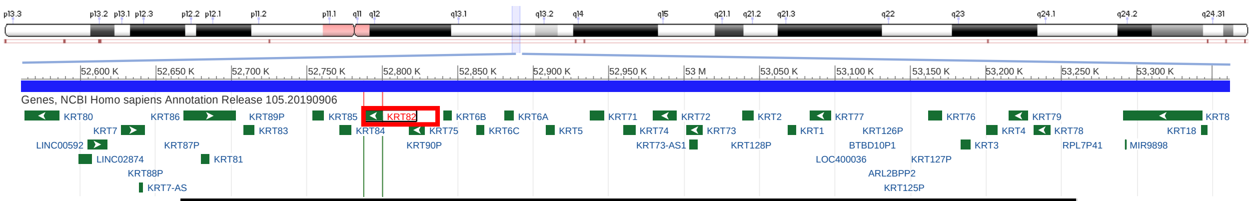


Figure 26 | Patients with KRT82 variants confirmed to have AA. a, Clinical photos of two patients with a missense and nonsense (LOF) mutation in KRT82, respectively. Both show hair loss patterns characteristic of AA. **b**, Immunohistochemistry (IHC) staining of CD8+ T cells in 4 patients with missense mutations in *KRT82*. Perifollicular CD8+ T cell infiltrate is a diagnostic characteristic of AA. **c,** RNA expression profiles of the ALADIN (Alopecia Areata Disease Severity Index) inflammatory signature confirms that three patients (green bars) with KRT82 mutations (splice and missense) have increased expression of the inflammatory signature, characteristic of AA. Additionally these three KRT82-carrying patients (green) cluster within AA patients (AAP-red, ATAU-orange), and not with controls (blue).

a. Type II Keratins Chr. 12q13



b.

Type	Case	Control	Case Freq	Ctrl Freq	P-value
LOF					
KRT82	**19**	**88**	**2.24%**	**0.56%**	**2.03E-06**
KRT84	0	13	0.00%	0.08%	1
KRT85	1	3	0.12%	0.02%	0.1906
KRT5	0	1	0.00%	0.01%	1
KRT32	0	4	0.00%	0.03%	1
KRT35	0	3	0.00%	0.02%	1
KRT39	0	16	0.00%	0.10%	1
KRT40	0	29	0.00%	0.19%	0.4014
LOF + missense					
KRT82	**47**	**376**	**5.54%**	**2.40%**	**9.19E-07**
KRT84	25	428	2.94%	2.74%	0.6668
KRT85	8	159	0.94%	1.02%	1
KRT5	3	80	0.35%	0.51%	0.8012
KRT32	6	111	0.71%	0.71%	1
KRT35	12	269	1.41%	1.72%	0.5869
KRT39	8	127	0.94%	0.81%	0.693
KRT40	17	267	2.00%	1.71%	0.4973
LOF + missense + splicing					
KRT82	**51**	**404**	**6.01%**	**2.58%**	**2.18E-07**
KRT84	25	457	2.94%	2.92%	0.9168
KRT85	25	344	2.94%	2.20%	0.999
KRT5	6	144	0.71%	0.92%	0.709
KRT32	6	142	0.71%	0.91%	0.7081
KRT35	12	284	1.41%	1.82%	0.5053
KRT39	8	127	0.94%	0.81%	0.693
KRT40	18	267	2.12%	1.71%	0.3438

Figure 27 | AA-associated variation is specific to KRT82. a, 700 kb chromosomal locus encompassing the type II keratin gene cluster (green bars) on chromosome 12. *KRT82* is boxed, in red. **b**, Table of rare damaging variants in seven other keratin genes: four type I keratins (*KRT32, KRT35, KRT39, KRT40*), two type II hair keratins (*KRT84, KRT85*), and one type II epithelial keratin (*KRT5*). None of the other keratins reach significance in any of the 3 models. Type II keratins were chosen for their similar classification and positional proximity to *KRT82*. The type I hair keratins are the putative binding partners of *KRT82.*

a.

Sample ID	12-52799341-C-T	12-52799419-T-C	12-52799464-A-G	12-52799551-C-T	12-52799675-G-T	12-52799717-C-T	12-52799788-T-C	12-52799923-G-A	12-52799963-C-T	12-52800014-ACT-A	12-52800269-A-G	12-52800286-G-T
aa00673	T T	C C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa00777	T T	C C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa01795	0 0	T C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa01849	T T	C C	A A	C C	G G	C C	T T	A G	C C	I I	0 0	0 0
aa2867	T T	C C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa4493	0 0	T C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa6943	C C	T C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa7311	C T	T C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa7407	C T	T C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa8629	T T	T C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa8878	C T	T C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa8905	C T	T C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa9123	C T	T C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aa9318	T T	C C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
aaHT005	T T	C C	A A	C C	G G	C C	T T	A G	C C	I I	A A	G G
	582 bp											

b.

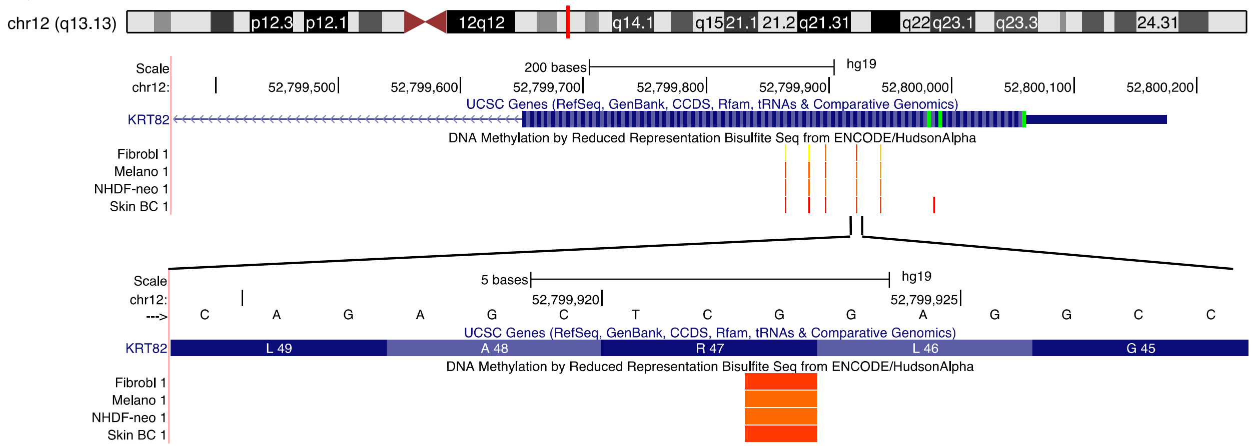


Figure 28 | Haplotype Analysis identifies Arg47X as a hotspot mutation. a, 15 AA patients heterozygous for the Arg47X mutation (12-52799923-G-A; highlighted in pink). Different genotypes (blue=alternate homozygous, yellow=heterozygous, green=reference homozygous) at variant 12-52799341-C-T (582 bp away) confirms the presence of the Arg47X A allele on different haplotypes, identifying it as a hotspot mutation. **b**, Schematic of *KRT82* exon 1 (blue) with ENCODE methylation patterns (orange) in 4 skin cell types: fibroblasts (Fibrobl), Melanocytes (Melano), Neonatal Dermal Fibroblasts (NHDF-neo), and Skin Tissue (Skin BC). This is a visualization of the methylated cytosine (C) of the sense strand that is spontaneously deaminated resulting in a mutational hotspot and premature stop codon at position 47.

Haplotype ID	12-52788928-G-A	12-52789140-T-C	12-52789198-C-A	12-52789208-T-C	12-52789509-T-C	12-52789597-A-G	12-52789605-T-A	12-52789691-A-C	12-52789693-GT-G	12-52790194-C-A	12-52790385-T-C	12-52790816-C-T	12-52790899-T-C	12-52790940-C-T	12-52794642-A-G	12-52795351-C-A	12-52797617-G-A	12-52797862-T-C	12-52799341-C-T	12-52799419-T-C	12-52799923-G-A
Hap01	A	C	A	C	C	G	A	C	G	A	C	T	C	T	G	A	A	C	T	C	G
Hap02	A	C	A	C	C	G	A	C	G	A	C	T	C	T	G	A	A	C	C	C	G
Hap03	G	T	C	T	T	A	A	A	GT	A	T	T	T	C	A	A	A	C	C	T	G
Hap04	A	C	A	C	C	G	A	C	G	C	C	T	C	T	G	A	A	C	T	C	G
Hap05	A	C	C	T	C	G	A	C	G	A	C	T	C	T	G	A	A	T	T	T	G
Hap06	A	C	A	C	C	G	A	C	G	A	T	T	T	T	A	C	A	C	T	C	G
Hap07	A	C	A	C	C	G	A	C	G	A	C	C	C	T	G	A	A	C	T	C	G
Hap08	A	C	A	C	C	G	A	C	G	A	C	T	C	T	G	A	A	C	T	C	A
Hap09	G	T	C	T	C	G	A	A	GT	A	C	T	C	C	G	A	G	C	C	T	A
Hap10	A	C	A	C	C	G	T	C	G	A	C	T	C	T	G	A	A	C	T	C	A
Hap11	G	T	C	T	C	G	A	C	G	A	C	T	C	C	G	A	A	C	C	T	A
Hap12	G	T	C	T	T	A	T	A	GT	A	C	T	C	C	G	A	A	C	C	T	A
Hap13	G	T	C	T	C	G	A	A	GT	A	C	T	C	C	G	A	A	C	C	T	A

Table 3 | Arg47X variants resides on at least 6 distinct haplotypes. Each column represents a variant and each row represents a haplotype. Arg47X alternate allele (A) that results in a premature stop codon is highlighted in red in the last column. After phasing, we report that the A allele resides on at least six distinct haplotypes. In this region, thirteen total haplotypes are observed. Black boxes denote where the haplotypes differ.

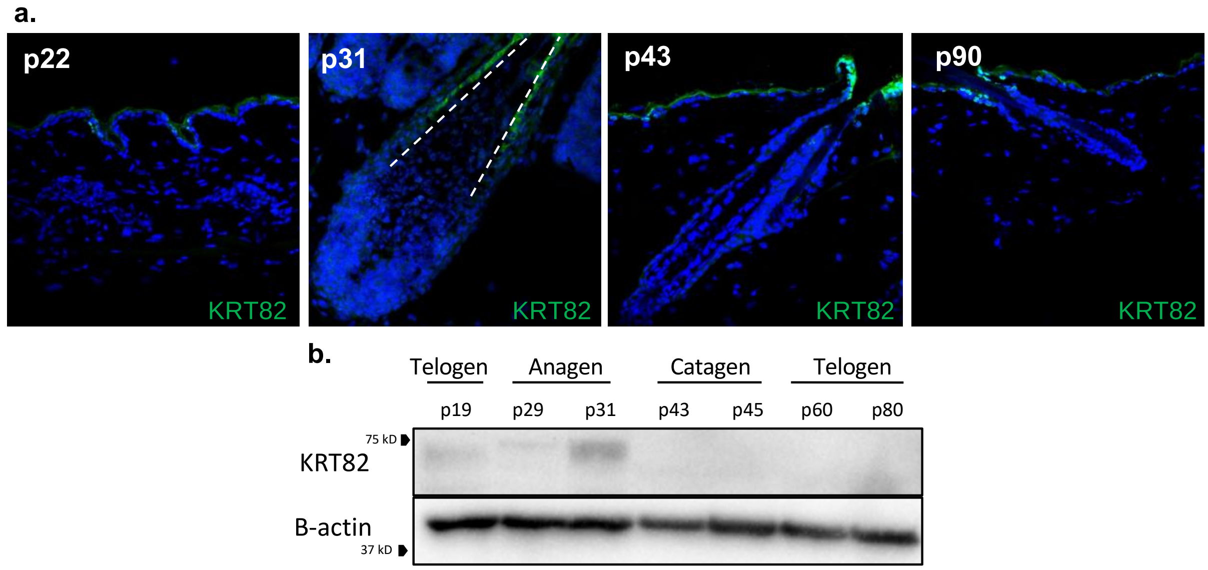


Figure 29 | KRT82 is expressed exclusively in anagen. a, IF of KRT82 (green) in postnatal (p) mouse skin. Mouse HFs are in telogen at postnatal days 22 and 90 (p22, p90), in anagen at p31, and catagen at p43. Cuticular staining of KRT82 is only observed in the p31 (anagen) HF. White dashed line outlines the hair shaft cuticle. Weak staining observed in mouse epidermis may be a result of green autofluorescent properties of skin epidermal cells. **b**, Western blot analysis of whole mouse skin at varying stages of the hair cycle shows that keratin protein is only present in the anagen phase (p31) of the mouse hair cycle.

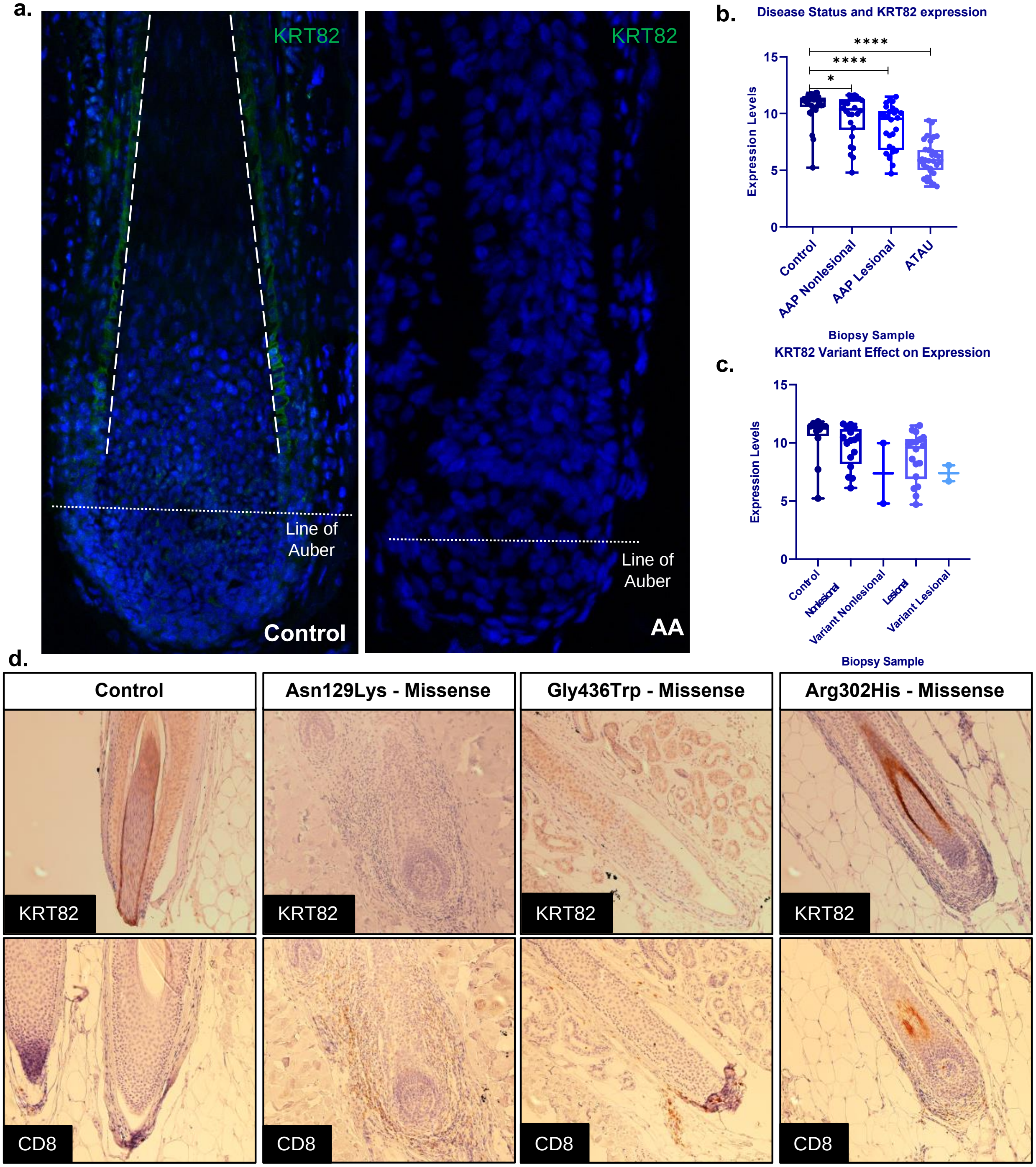


Figure 30 | Functional KRT82 expression is reduced in AA HF and skin. a, IF staining of KRT82 (green) in human healthy control and AA HF. Dashed line outlines the cuticle, dotted line represents the Line of Auber **b**, RNA expression of KRT82 from whole skin scalp biopsies of control, AAP nonlesional sites, AAP lesional sites, and AT/AU sites. Each circle represents one sample. Statistical analysis was performed using unpaired t-test with Welch's correction. *=p<0.05, ****p<0.0001. **c**, Scalp RNA expression of patients carrying *KRT82* variants at nonlesional (Variant Nonlesional) and lesional (Variant Lesional) sites compared to AA patients without *KRT82* mutations (Nonlesional and Lesional) and controls without AA or *KRT82* mutations. **d**, IHC of KRT82 and CD8 in healthy control and lesional biopsies in three AA patients with missense mutations.

Chapter 4 Conclusion

Defining the genetic architecture of autoimmune diseases is complex due to their polygenic nature and heterogenous tissue interactions between the immune system and the target organ. The common immune pathways among different autoimmune diseases explains in part the shared genetic susceptibility to general autoimmunity, however, the divergent mechanisms of tissue-specific attack (i.e. HF in AA, myelin sheath in MS, melanocytes in vitiligo, islet cells in type I diabetes, etc.) remain less well understood. Identifying genes that contribute to dysfunction in the target organ itself is crucial for fully defining the genetic architecture across the broad class of autoimmune diseases. Discovery of target-organ risk genes unique to a given disease may reveal the mechanism of action, and uncover therapeutic targets aimed at mitigating end organ dysfunction, rather than subjecting autoimmune disease patients to the substantial toxicities of broad spectrum and lifelong immunosuppression.

The overarching goal of my thesis was to elucidate the contribution of genetic variation in genes specific to the target-organ in autoimmune disease and determine how disruption of target tissue biology contributes to autoimmune-mediated attack. In this context, AA was an ideal model to study due to the accessible and regenerating nature of the targeted organ, the HF. My thesis focused on target-organ dysfunction to elucidate the mechanisms of target selection that differentiates the autoimmune diseases. In AA, genetic loci shared with other autoimmune diseases include CTLA4, IL21/2, IL2RA, PRDX5, IKZF4, ERBB3, MICA, HLA-DQA, most of which have demonstrated roles in immune signaling and function. These common pathways can help explain aberrant modifications in the immune system that result in inflammatory responses and autoreactive lymphocytes across autoimmunity. However, it remains uncertain as to why autoreactive lymphocytes react to antigens in the target organ of one disease but another. The answer to this question lies in the target tissue, and how disruption to the local biology results in improper recognition by the immune system. In order to understand whether genetic variation plays a role in target-tissue biology, we must focus on genetic associations unique to each autoimmune disease. This will provide the groundwork for novel therapeutics targeting new relevant disease pathways, instead of treating with modest effect immunosuppressants repurposed for multiple autoimmune diseases. The significant advances in NGS methods over the past decade allowed us to utilize two distinct yet powerful sequencing approaches in order to investigate genetic variation in HF-specific genes.

First, based on our *hypothesis* that GWAS associations unique to AA will play a target-specific role, we used targeted sequencing to determine the complete genomic sequence (coding and non-coding) of a GWAS-associated locus. Our previous GWAS and meta-analysis identified at least 14 risk loci associated with AA susceptibility, only two of which, *STX17* and *ULBP3*/6, were not implicated in other autoimmune diseases. To address the role of genetic variation in HF integrity, we selected *STX17*, a gene that was exclusively associated with AA (and not other autoimmune diseases), is highly expressed in the hair follicle, and had a known genetic role in hair pigmentation. The GWAS association was detected at a common variant, rs10760706, that resided in intron 6 of *STX17.* To identify the complete scope of variants in the region that may be driving this association, we used a *hypothesis-driven* targeted sequencing approach.

Secondly, we used WES to investigate the role of rare variants in AA. Due to the *unbiased* nature of this technique, we were interested in identifying any genes that had rare variant enrichments and novel associations with AA. To our surprise, only one gene, *KRT82,* a hair specific type II keratin, reached genome-wide significance with rare variants enriched in AA. Using these two methods, I investigated both common and rare variation in AA. I functionally defined the mechanisms by which genetic perturbances in these HF-expressed genes, *SXT17* and *KRT82*, contribute to disease pathogenesis. These findings elucidated the effect of common and rare genetic variants on the disruption of hair follicle function in AA and serve as framework for targeted therapies aimed at restoring HF function in AA.

Targeted Sequencing of GWAS-associated locus

As discussed in previous sections, one of the main limitations of GWAS is the lack of resolution to determine the causal disease variant(s) in a given associated region. Since GWAS is the main method for identifying pathogenic common variation in complex diseases, the true causal variant(s) and their downstream effects remain largely undiscovered in many common disorders. As a result, fine mapping and targeted sequencing of GWAS associated loci were implemented to identify causal variants within the region. In our study, we designed a custom sequencing chip that contained whole exome sequencing (WES), in addition to genomic sequencing of large regions customized to include GWAS loci. We included all 14 of our GWAS regions, in addition to other genes with potential implications in AA pathology. Using this approach, we sequenced the entire STX17 GWAS region spanning 550 kb in 849 AA patients and 62 controls and defined the risk haplotype driving AA disease risk. This approach enabled us to resolve the surrounding LD block for additional common variants contributing to disease and functionally annotate their effects. The success of this approach would not have been possible without the significant advances made in sequencing technologies in the years leading up to this study.

Novel role for STX17 in melanocyte biology

We performed fine mapping targeted sequencing of the 550 kb GWAS-associated *STX17* region in 849 AA patients to understand the correlation between *STX17* variation and hair pigmentation in the context of AA disease. We identified 35 non-coding *STX17* variants that resided on the same risk haplotype as the initial GWAS SNP (rs107060706) associated with AA susceptibility. We used this sequencing data to define a ~64.6 kb haplotype that contained 35 common non-coding variants, including the AA and human hair color GWAS SNPs (rs10760706 and rs9556, respectively). The presence of a known hair pigmentation variant on our AA risk haplotype underlined the correlation between AA and hair pigment biology.33 of these risk variants were significantly enriched in the AA population.

Using the publicly available database, GTEx, we found that all 35 SNPs on the risk haplotype had a regulatory effect on the expression of STX17 in the skin. To understand whether this functional effect was relevant to AA, we used multi-dimensional genome sequencing and skin expression data from patients and discovered that 32 significant AA variants were skin eQTLs that significantly downregulated expression of STX17 in AA skin. This integrative approach revealed the effect of these variants in total skin tissue RNA, however, due to the heterogenous nature of the skin, we were unable to distinguish the precise cell type implicated. In order to resolve this question, we used *in silico* methods that annotated regulatory regions across the *STX17* gene in the three relevant skin cell types: fibroblasts, keratinocytes, and melanocytes. These analyses revealed that two variants (rs7039716 and rs2416935) fell in the promoter region of *STX17* in all skin cell types and one (rs7027813) resided in a melanocyte specific enhancer. These three variants are likely to be the driving functional effects of the *STX17* risk haplotype, and moreover, indicated a preferential effect of STX17 regulation in melanocytes. Concordantly, immunofluorescent (IF) staining of AA hair follicles revealed dysregulated expression of STX17 proximal to dysmorphic follicular melanocytes.

Due to the increased frequency of melanocyte-specific enhancers in *STX17*, the presence of one associated SNP in a melanocyte enhancer, and clinical associations of AA with pigmentary phenotypes, we performed *in vitro* analyses of *STX17* function in primary human melanocytes. Since AA patients with risk variants have decreased expression of STX17 in the scalp, we used an siRNA knockdown model to recapitulate AA. Using this approach, we reported a novel role for *STX17* in melanocyte biology that was autophagy independent. We found that in the absence of *STX17*, melanocytes were unable to produce melanin as efficiently. The melanocytes were still able to produce a lower volume of melanin, which was supported by the presence of pigmented late melanosomes. In fact, the number of late-stage, pigmented melanosomes and total melanosomes were unchanged after *STX17* knockdown. Since the total number of melanosomes was unchanged, but the melanin content was reduced, we hypothesized that the melanosomes were unable to produce melanin as efficiently within the organelles. Interestingly, the pre-pigmented early-stage melanosomes were slightly, but significantly reduced after *STX7* depletion. The role of early stage melanosomes is to provide structure to the organelle via PMEL fibrils, onto which melanin can be deposited in stage III and IV melanosomes. We observed that STX17 co-localized with MART1 at baseline and after melanogenesis stimulation, suggesting that STX17 may play a role in MART1 trafficking. The reduction of early stage melanosomes in *STX17*-deficient cells supports this hypothesis, since cells would be unequipped to efficiently traffic the critical structural protein, MART1, to the early melanosomes. Moreover, lack of integral melanosome proteins would inhibit the organelle's ability to efficiently undergo melanogenesis, as we observed in the *STX17* knockdown conditions. Finally, we tested whether loss of STX17 would result in accumulation or degradation of melanosome cargo proteins. We found that levels of TYR, a protein that is present in late stage melanosomes, were not affected by STX17 reduction. However, early stage protein, MART1, accumulated in *STX17*-knockdown melanocytes.

This finding is interesting for two main reasons. First, the function of MART1 in the melanogenesis pathway is not well documented; however, it has been reported to play a critical role in the trafficking, processing, and functioning of PMEL[102]. MART1 and PMEL co-localize in the peri-nuclear area and form a complex in the ER, the Golgi, and stage I and II melanosomes[102]. The stability and expression levels of PMEL are dependent on MART1, and in its absence, PMEL is unable to efficiently traffic to early melanosomes. The MART1-PMEL complex is relevant to our findings, since we observed MART1-STX17 colocalization and an accumulation of MART1 in the absence of STX17. We postulate that STX17 assists in the transport of the MART1 to melanosomes which subsequently regulates PMEL expression, processing, and trafficking to early melanosomes. Furthermore, *STX17*-depletion causes aberrant MART1 and consequentially PMEL trafficking resulting in the decreased formation of early melanosomes, inefficient melanin production, and accumulation of MART1.

Second, MART1 has well-documented antigenic properties in cancer and autoimmunity literature, which is extremely relevant in the context of disease. The discovery of MART1 (Melanoma antigen recognized by T cells 1) was first identified in 1994 and aptly named because of its initial discovery as an antigen recognized by cytotoxic lymphocytes in melanoma patients[163,164]. It was found to be exclusively expressed in a subset of melanoma cell lines, melanocytes, and retinal cells[164]. In melanoma studies, MART1 was recognized by CD8+ T cells and capable of activating lymphocytes activation and inducing lysis of melanoma cells, indicating the cytotoxic activity [103,165-167]. In another autoimmune disease targeting the skin, vitiligo, MART1 was proposed as an auto-antigen due to the high frequencies of melanocyte-specific cytotoxic T lymphocytes that recognized MART1[105,168,169]. In the rare autoimmune disorder, Vogt-Koyanagi-Harada disease, MART1 was capable of inducing cytotoxic T lymphocytes to lyse melanocytes[170]. Moreover, an elegant study by Gilhar et al implicated MART1 as a potential autoantigen in AA[131]. This work showed that human MART1, in addition to PMEL, was sufficient to activate AA scalp T cells and induce hair loss in an explant/ SCID mouse model of disease. The antigenic nature of MART1 in both melanoma and autoimmune diseases supports its potent immunogenic nature and ability to elicit T cell responses. Taken together with work presented in this thesis, we showed that genetic variants in AA patients decrease STX17 in the scalp skin, and subsequently that reduced STX17 levels in melanocytes result in an accumulation of a known antigen, MART1. We propose that accumulation of MART1 has the capability to induce cellular dysfunction or aberrant antigen presentation that can result in subsequent autoimmune attack of the pigmented hair follicles.

Based on our findings, we propose the following model. Individuals carrying the AA risk haplotype are susceptible to disease through the downregulation of STX17 in the skin. In the absence of STX17, the PMEL/MART1 complex is not efficiently transported to the stage I melanosomes. As a result, there is a reduction in early melanosome numbers. We postulate the presence of a compensatory mechanism since melanosomes still form, mature, and produce reduced (but not absent) melanin pigment. However, the deficiency of critical structural protein, PMEL, in melanosomes results in the inability to produce melanin efficiently, lowering the total production of melanin in the cell. Finally, in the absence of STX17, MART, a member of the early melanosome machinery, is mis-trafficked and accumulates in the cell, and may be aberrantly presented to auto-reactive CD8 + T cells through MHC1 antigen presentation (Figure 31). Prior work supports the validity of MHC I presentation of MART1 through the ubiquitination and proteasomal degradation pathways[171]. In our model, increased MART1 presentation and identification by inflammatory T cells would elicit attack on the follicular melanocytes and surrounding HF unit, resulting in AA-mediated immune attack. We recognize that this model most likely accounts for a subset of AA cases and that disease presence is dependent on a range of other factors, such as gene-gene interactions, the presence of auto-reactive immune cells, and environmental insults. Nonetheless, our findings invite a novel mechanism linking genetic predisposition in a HF gene and consequential disruption of the melanocyte pathway in AA pathogenesis.

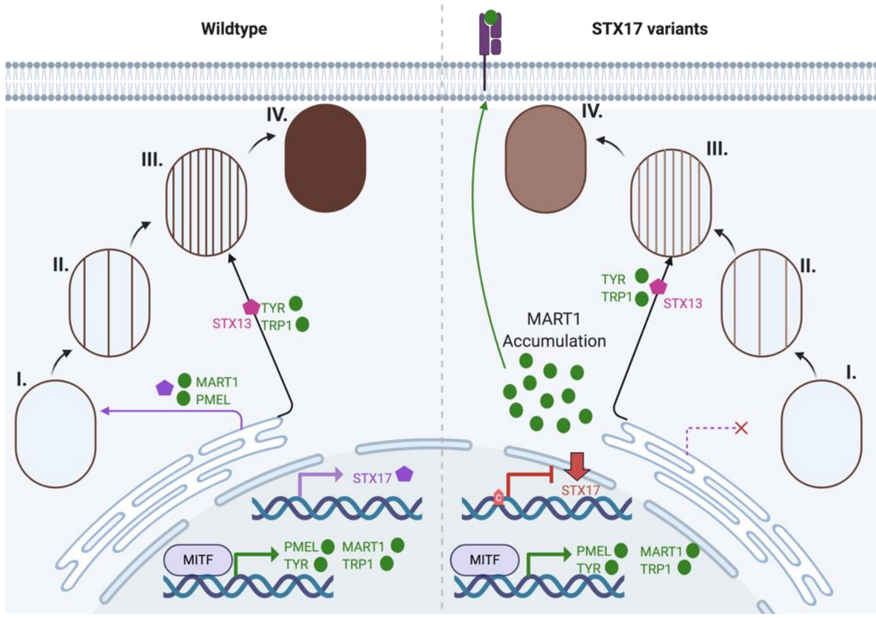


Figure 31 | Proposed model of STX17-mediated cargo trafficking. Under normal conditions, STX17 aids in trafficking MART1 to early melanosomes. STX13 has an established role TYR trafficking. In the absence of STX17, MART1 is not trafficked to forming melanosomes causing a de-pigmentation phenotype and accumulation of MART1 antigen that is aberrantly present to auto-reactive immune cells.

Although more work will be needed to further define the precise trafficking role of STX17 in melanogenesis, the role of SNARE proteins in melanocyte enzyme trafficking is not without precedent. To date, four SNARE proteins (VAMP7, STX3, STX13, and SNAP23) have well-documented roles in cargo trafficking and membrane fusion in the melanosome maturation pathway[172]. All four SNARE proteins were demonstrated to have a role in TYRP1 transport to melanosomes via two separate complexes, the VAMP7-STX13 complex and the VAMP7-STX3-SNAP23 complex (Figure 31)[172-174]. Additionally, VAMP7 and STX13 played a role in TYR cargo transport and biogenesis of PMEL+ melanosomes (Figure 31)[174]. Furthermore, in the absence of STX13, TYR is misrouted and PMEL and TYRP1 mis-localized to the lysosomes[174]. It is interesting to note that depletion of either STX3 or STX13 result in hypopigmentation[172], a similar phenotype to what we observe in STX17-depleted cells. The dependence of melanosome cargo trafficking on other SNARE proteins supports a potential role for STX17 in protein transport and membrane fusion with maturing melanosomes. Additionally, the current role of SNARES is mostly in late stage protein trafficking, and understanding the machinery for early stage cargo trafficking, such as PMEL and MART1, remains largely unknown. The results presented in this thesis support a plausible role for STX17 in the transport of the PMEL/MART1 complex to early melanosomes.

In conclusion, we utilized targeted sequencing to resolve a GWAS associated disease region and annotated the functional effects of the risk variants on regulatory expression in AA. We reported that STX17 is necessary for melanogenesis and that reduction of STX17 resulted in accumulation of melanocyte antigens and increased immunogenicity of the hair follicle. We determined that the accumulation and mis-localization of melanosome proteins in the absence of STX17 supported a strong role for enzyme trafficking to early melanosome organelles. This work addresses a long-standing, widely observed phenomenon of preferential attack of pigmented HFs and provides a mechanistic link between genetic AA risk and melanocyte dysfunction in disease, for the first time.

Whole Exome Sequencing of 849 AA patients

To identify the role of rare variants in AA risk burden, we next performed whole exome sequencing (WES) on 849 AA patients compared to 15,640 controls using an unbiased genome-wide approach. This enabled us to investigate whether rare variation in coding sequences plays a role in AA susceptibility. WES is now a cost-effective approach (compared to Whole Genome Sequencing), and its success has been well-documented in identifying rare coding variants associated with other complex diseases such as ALS and Alzheimer's Disease[6,51]. Based on these successes of WES in complex diseases and the potential role of rare variation in the 'missing heritability' paradigm, we selected WES as the appropriate approach for identifying novel rare coding mutations in AA. This strategy also implemented the use of gene-level burden analysis models, which aggregate genetic variants across a gene and compare genetic burden in cases compared to controls. Gene-collapsing analysis approaches were necessary to identify significant associations, since it provided aggregated power to test association across an entire candidate gene, whereas variant-level significance would be unattainable. As discussed previously, detection of individual variant association is not feasible in modest-sized studies because power to detect statistically significant associations decreases as MAF decreases due to the lack of observations of the given rare allele. Together, these approaches revealed a previously unknown AA candidate gene association and provided a major new insight into genetic mechanisms that predispose individuals to AA disease susceptibility.

Rare variants in *KRT82* contribute to disease risk

By utilizing WES together with gene-level burden analyses, we identified a novel association between rare variation in *KRT82* and AA disease risk. We surveyed the entire exome for genes that harbored an increased frequency of rare, damaging mutations in AA cases compared to controls. This allowed us to identify genes with damaging mutations that contributed to disease risk. Furthermore, discovery of novel genetic contributions to AA provided insight into a novel disease mechanism and how loss of functional proteins in the HF may play a role in AA pathogenesis.

In Chapter 3, I successfully identified a type II hair keratin, *KRT82,* which harbored a significantly higher frequency of rare, damaging mutations in AA patients (6.0%) compared to controls (2.6%) (p=2.18E-07). From our cohort of 849 AA patients, we identified 51 individuals who were heterozygous for one of the eleven rare damaging mutations that were predicted as loss of function, missense, and splicing mutations; suggesting a haploinsufficiency model of disease. Previous reports have identified keratin haploinsufficiency in the context of two diseases: 1) LOF mutations resulted in KRT5 (type II) haploinsufficiency in Dowling-Degos Disease[175]; and 2) KRT14 (type I) haploinsufficiency from LOF mutations in Naegeli-Franceschetti-Jadassohn syndrome[176]. The eleven variants we identified in AA patients fell in highly conserved residues and the majority resided in keratin domains established as disease-causing in other disorders. Since these mutations were all predicted to have a damaging effect on protein function, we interrogated the expression of KRT82 in AA and found that both scalp biopsies and HFs exhibited reduced levels of functional expression. We also noted that patients with one of the damaging *KRT82* mutations, and absent KRT82 expression in the HF cuticle, had increased CD8+ T cell infiltrates, the hallmark characteristic histopathology of disease.

Previous knowledge of *KRT82* was limited to a type II hair keratin with specific cuticle layer expression in the human hair follicle. Its potential association with HF dysfunction was not previously well-studied. Interestingly, three previous reports implicated *KRT82* expression in hair loss phenotypes in human and mouse studies. KRT82 is highly conserved across mammalian species, including *mus musculus* (Figure 25). First, one mouse model of cyclic alopecia is a result *Sox21* knockout[141]. In response to loss of *Sox21*, the mice expressed altered levels of several keratins and keratin-associated proteins, including a striking decrease in murine *Krt82*. Moreover, *Sox21*$^{-/-}$ knockout mice were unable to form healthy hair shaft cuticles and completely lost and regrew their hair in cyclic waves. Consequentially, *Sox21* was proposed to be a regulator of keratins and keratin-associated proteins that are necessary for hair shaft cuticle keratinization. Due to the reported repressor function of *Sox21*, it is likely that the effects on keratin expression are mediated through another unknown transcriptional repressor that is the direct target of *Sox21*[141]. The loss of KRT82 and healthy cuticle structures in a mouse model of cyclic alopecia strongly supports an important role for KRT82 in hair follicle function.

The other two studies indirectly integrate KRT82 expression levels with AA pathology. Previous work in our lab developed the ALADIN score (ALopecia Areata Disease Severity INdex) as a clinical and research tool for monitoring AA disease status using a distinct molecular signature dependent on a subset of genes differentially expressed in patients and controls[124]. It is also capable of deciphering disease severity (Control vs AAP vs ATAU)[124]. ALADIN is dependent on the expression profiles of only 18 genes divided into 3 signatures (Cytotoxic T Lymphocyte, Interferons, and Keratins) as markers of disease. The 8 signature keratin genes are decreased in patients. Interestingly, *KRT82* is one of the defining genes of the Keratin molecular signature in ALADIN, which was developed several years prior to the discovery of *KRT82* variants in this thesis[60]. Since ALADIN is a distinguishing tool in AA to measure disease status, progression, and drug response, the inclusion of *KRT82* points to an integral role for its expression in disease pathology.

Additionally, recent work published by our lab investigated the role of skin-specific master regulators in the context of autoimmunity[177]. The identification of master regulators can provide information about disease mechanisms, driver mutations, and druggable targets for therapeutic intervention. In autoimmune diseases and other heterogenous diseases, interactions between different implicated tissues, such as the skin and the immune infiltrate in AA, can convolute regulatory networks and hinder accurate identification of the driving master regulators. By using differential gene expression profiles and mapping to skin regulatory networks, our lab effectively deciphered the molecular profiles that regulate disease pathogenesis and induce immune infiltrate recruitment to skin in AA. 123 genes were annotated as skin-specific gene signatures that were differentially expressed in AA scalp and were responsible for end-organ gene dysfunction in disease, including immune infiltrate recruitment. Interestingly, one of these signature genes that informed the master regulator analysis was significant decreased expression of *KRT82.* Of the four potential type I keratin binding partners of *KRT82* (*KRT32, KRT35, KRT39, KRT40*), three including *KRT32, KRT35*, and *KRT40* were also signature genes that informed the analysis. Subsequently, two master regulators, IKZF1 and DLX4, were identified to regulate more than 60% of the AA skin gene signature. Furthermore, overexpression of IKZF1 and DLX4 *in vitro* resulted in NKG2D-dependent cytotoxicity in primary hair/skin cells. *KRT82* specifically interacted with DLX4, a homeobox transcription factor. Even though the networking algorithm is unable to infer direction, we can assume that KRT82 is the regulatory target of DLX4 because of DLX4's classification as a transcription factor and master regulator. Intriguingly, in one specific cellular setting, TGF-β suppressed DLX4 expression. This is relevant to disease since TGF-β is an immunosuppressive molecule that is responsible in part for maintaining immune privilege of the hair follicle. We postulate that when immune privilege collapses in AA, TGF-β levels decrease, DLX4 increases expression and induces regulatory changes in skin-specific gene signatures (including *KRT82)*, which subsequently recruit immune infiltrate and result in NKG2D-cytotoxicity. These findings included both *KRT82* and its potential hair shaft cuticle type I keratin binding partners as members of the molecular profile responsible for immune infiltration induction in AA skin. Taken together, these studies lend additional support for the crucial role of KRT82 in healthy hair follicle function and support our evidence for the absence of KRT82 in association with disease pathology. In summary, the work presented in Chapter 3 provides a genetic explanation to the reduced expression of KRT82 that has been observed in previous studies.

We propose four potential mechanisms for the role of KRT82 in cuticle formation and subsequent AA pathogenesis: 1) loss of structural integrity renders the HF vulnerable to infiltration; 2) inability to bind to type I keratin partners results in increased antigen availability; 3) loss of expression induces premature entry into catagen; and 4) increased ROS levels may result in a metabolically vulnerable tissue.

First, we postulate that loss of a critical cuticle keratin, such as KRT82, would result in a structurally defective cuticle during anagen. The loss of functional KRT82 would prevent accurate cross-linking between KRT82 and its type I binding partner. With a structurally compromised cuticle, the hair shaft would not be ineffectively protected from mechanical and environmental insults. Additionally, without the outer layer of the protective cuticle the hair shaft could be rendered vulnerable to inflammatory auto-reactive T cells, thus inciting the AA immune attack. In support of this mechanism, Tobin et al. previously showed absent or defective hair shaft cuticles in AA exclamation mark hairs (EMH)[56]. Furthermore, areas of the hair shaft without cuticles revealed the most disorganized shaft damage. This supports a protective role of the cuticle in shielding the hair shaft from potential insults associated with the AA immune attack. Additionally, structurally defective cuticles were observed in the *Sox21*$^{-/-}$ knockout mice with cyclic alopecia, further supporting a critical role for cuticle structural integrity in proper hair shaft growth and HF maintenance[141].

Second, the absence of KRT82 from AA cuticles may result in the aggregation of its type I binding partners (KRT32, KRT35, KRT39, KRT40). In the absence of functional KRT82, the type I cuticle keratins cannot properly dimerize and keratinize to form the hair shaft cuticle. The type I keratins would either be unable to bind to their type II partner (KRT82) and accumulate in the cells; or improperly form disrupted dimers that could be targeted for degradation and aberrantly presented to immune cells. Interestingly, previous reports support a role for 44/46 kDa keratins as auto-antigens in AA patients[95]. Since the size of type I keratins range from 40-

48 kDa, we propose that unbound type I keratins in the hair shaft cuticle may be aberrantly recognized by the immune system as autoantigens in the absence of functional KRT82.

Third, KRT82 is expressed in the hair shaft cuticle only while the hair shaft is actively proliferating and growing during anagen. One highly accepted pathomechanism of AA onset is that the AA HF enters catagen prematurely, thereby resulting in hair loss[54,178]. Premature catagen induction in AA HFs is a widely accepted mechanism of disease, and perhaps the absence of KRT82 (or a functional hair shaft cuticle) is responsible for this phenotype. In support of this, the Sox21$^{-/-}$ cyclic alopecia mouse model had increased hair shaft detachment in catagen, but not anagen, suggesting that the catagen HF is dependent on the hair shaft cuticle and IRS cuticle interactions to remain anchored[141]. We postulate that in the absence of a structurally intact cuticle, the HF may enter catagen prematurely, and become unable to anchor the hair shaft, resulting in hair loss.

Finally, the well-established roles of intermediate filaments in preventing cell damage and oxidative stress response may help elucidate the mechanism of KRT82 function in anagen HF biology[179]. The role of ROS in the growing anagen hair follicle has been previously characterized[180]. Tight regulation of oxidative stress is required to carry out necessary functions without inducing cellular stress. In the hair follicle, an oxidative environment is required for keratin cross-linking[180]. As a result, the anagen hair follicle generates high levels of ROS above the Line of Auber, with highest levels at the “ring of fire” where the hair shaft outer cortex and cuticle form[180]. Interestingly, this localization coincides with the zone of KRT82 expression in the healthy hair follicle. However, in the AA hair follicle, KRT82 is reduced or absent, and we postulate that the high levels of ROS generated proximal to the growing shaft cannot be utilized to mediate KRT82 crosslinking with its binding partner. As a consequence, ROS levels may accumulate within the hair follicle, leaving it metabolically vulnerable to ROS-induced cellular stress. In support of this hypothesis, the skin of AA patients exhibits increased oxidative stress markers, such as lipid peroxidation and superoxide dismutase (SOD) levels[181,182]. We postulate that the bulb may be rendered metabolically vulnerable due to excess ROS not utilized by keratin cross-linking reactions during cuticle formation. This vulnerability may result in unregulated cell death and/or aberrant immune infiltration of the HF, inciting AA disease.

The work reported in Chapter 3 reports the successful identification of a disease-relevant gene, *KRT82*, which harbors rare variation in 6% of AA patients. We show that genetically susceptible individuals carry rare damaging mutations in *KRT82* that are predicted to cause protein dysfunction. Concordantly, functional KRT82 levels are decreased in the skin and HF cuticles of AA patients. This novel finding presents a potential disease mechanism for genetic predisposition to disrupted hair shaft integrity and AA. Future studies may further elucidate the precise mechanism of decreased KRT82 levels, cuticle formation, and AA disease pathogenesis.

Perspectives

In my thesis work, I used two sequencing approaches to understand how both common and rare risk variants in genes expressed in the hair follicle contribute to AA autoimmune susceptibility. The implication of genetic perturbations in the HF within the context of AA pathology has yet to be explored and using modern genetic approaches we identified novel biological mechanisms underlining HF dysfunction and pathogenesis. Understanding target organ function and how genetic variants disrupt their integrity can help us elucidate the complex heterogenous interactions of autoimmune disorders. Additionally, by identifying target organ dysfunction, we may gain insight into novel therapeutic interventions to correct such aberrant function(s), instead of relying solely on lifelong immunosuppression for patients. This thesis work distinguishes two types of variation leading to HF susceptibility: a common risk haplotype in *STX17* that exerts negative regulatory effects on follicular melanocytes in about 36% of patients, and rare mutations in *KRT82* that disrupt cuticle integrity in approximately 6% of AA cases. These findings open up the possibility of genetically informed precision medicine approaches for the treatment of AA in the future.

Understanding the genetic architecture of rare and common diseases alike is crucial for the progress of the scientific and biomedical fields. Interpreting how genetic variants predispose individuals to disease can inform our efforts in tailoring treatments to patients in an efficient, effective, and genetically determined manner. The past decade has seen unparalleled advances in sequencing technologies that have amassed large amounts of data and contributed substantially to our knowledge of the human genome. Databases such as ExAC and gnomAD were created to collect sequencing data across thousands of individuals and ethnicities and provide public access to this data for the entire scientific community. Yet, with the accumulation of extensive amounts of genetic information, challenges arose in the accurate interpretation of variants. In response, PolyPhen was designed to predict the effect of coding variants, while the ENCODE project collated DNA regulatory annotations spanning the genome in order to interpret non-coding variants. Through these types of consortia, researchers were not only able to identify variants associated with disease risk but were also able to predict whether these variants would have damaging effects on protein products or gene regulation, effectively informing the likelihood of pathogenicity. Due to the high variability between individual human genomes, these efforts have been crucially important in identifying meaningful variants with actual effects on disease across patient cohorts. Without these advances in the field, the work presented herein would have been infinitely more challenging, if not impossible. As the field continues to move forward at a rapid pace, our understanding of the genome, disease, and the human condition will simultaneously and swiftly evolve. I am grateful to have contributed this work to these ongoing efforts.

References

1 Cirulli, E. T. *et al.* Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat Commun* **11**, 542, doi:10.1038/s41467-020-14288-y (2020).

2 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753, doi:10.1038/nature08494 (2009).

3 McKusick, V. A. Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet* **80**, 588-604, doi:10.1086/514346 (2007).

4 Posey, J. E. Genome sequencing and implications for rare disorders. *Orphanet J Rare Dis* **14**, 153, doi:10.1186/s13023-019-1127-0 (2019).

5 Sanna-Cherchi, S. *et al.* Exome-wide Association Study Identifies GREB1L Mutations in Congenital Kidney Malformations. *Am J Hum Genet* **101**, 1034, doi:10.1016/j.ajhg.2017.11.003 (2017).

6 Cirulli, E. T. *et al.* Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347**, 1436-1441, doi:10.1126/science.aaa3650 (2015).

7 Petrovski, S. *et al.* An Exome Sequencing Study to Assess the Role of Rare Genetic Variation in Pulmonary Fibrosis. *Am J Respir Crit Care Med* **196**, 82-93, doi:10.1164/rccm.201610-2088OC (2017).

8 Epi, K. c. & Epilepsy Phenome/Genome, P. Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. *Lancet Neurol* **16**, 135-143, doi:10.1016/S1474-4422(16)30359-3 (2017).

9 Myocardial Infarction, G. *et al.* Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N Engl J Med* **374**, 1134-1144, doi:10.1056/NEJMoa1507652 (2016).

10 Do, R. *et al.* Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**, 102-106, doi:10.1038/nature13917 (2015).

11 Ionita-Laza, I., Capanu, M., De Rubeis, S., McCallum, K. & Buxbaum, J. D. Identification of rare causal variants in sequence-based studies: methods and applications to VPS13B, a gene involved in Cohen syndrome and autism. *PLoS Genet* **10**, e1004729, doi:10.1371/journal.pgen.1004729 (2014).

12 NCBI. *Overview of Structural Variation*, 2020).

13 Lodish, H. B., A.; Zipursky, S.L.; Matsudaira, P.; Baltimore, D.; Darnell, J. in *Molecular Cell Biology* (ed W.H. Freeman) (2000).

14 Deutschbauer, A. M. *et al.* Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* **169**, 1915-1925, doi:10.1534/genetics.104.036871 (2005).

15 Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* **Chapter 7**, Unit7 20, doi:10.1002/0471142905.hg0720s76 (2013).

16 Szafranski, K. *et al.* Violating the splicing rules: TG dinucleotides function as alternative 3' splice sites in U2-dependent introns. *Genome Biol* **8**, R154, doi:10.1186/gb-2007-8-8-r154 (2007).

17 Clancy, S. RNA splicing: introns, exomes and spliceosome. *Nature Education* **1**, 31 (2008).

18 Gelfman, S. *et al.* Annotating pathogenic non-coding variants in genic regions. *Nat Commun* **8**, 236, doi:10.1038/s41467-017-00141-2 (2017).

19 Gloss, B. S. & Dinger, M. E. Realizing the significance of noncoding functionality in clinical genomics. *Exp Mol Med* **50**, 97, doi:10.1038/s12276-018-0087-0 (2018).

20 Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-343, doi:10.1038/nature13835 (2015).

21 Yousefian-Jazi, A., Jung, J., Choi, J. K. & Choi, J. Functional annotation of noncoding causal variants in autoimmune diseases. *Genomics* **112**, 1208-1213, doi:10.1016/j.ygeno.2019.07.006 (2020).

22 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

23 Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-1048, doi:10.1038/nbt1010-1045 (2010).

24 Backenroth, D. *et al.* FUN-LDA: A Latent Dirichlet Allocation Model for Predicting Tissue-Specific Functional Effects of Noncoding Variation: Methods and Applications. *Am J Hum Genet* **102**, 920-942, doi:10.1016/j.ajhg.2018.03.026 (2018).
25 Broekema, R. V., Bakker, O. B. & Jonkers, I. H. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biol* **10**, 190221, doi:10.1098/rsob.190221 (2020).
26 Nica, A. C. & Dermitzakis, E. T. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120362, doi:10.1098/rstb.2012.0362 (2013).
27 Baron, M. The search for complex disease genes: fault by linkage or fault by association? *Mol Psychiatry* **6**, 143-149, doi:10.1038/sj.mp.4000845 (2001).
28 Jin, W., Qin, P., Lou, H., Jin, L. & Xu, S. A systematic characterization of genes underlying both complex and Mendelian diseases. *Hum Mol Genet* **21**, 1611-1624, doi:10.1093/hmg/ddr599 (2012).
29 Burton, P. R., Tobin, M. D. & Hopper, J. L. Key concepts in genetic epidemiology. *Lancet* **366**, 941-951, doi:10.1016/S0140-6736(05)67322-9 (2005).
30 Wall, J. D. & Pritchard, J. K. Assessing the performance of the haplotype block model of linkage disequilibrium. *Am J Hum Genet* **73**, 502-515, doi:10.1086/378099 (2003).
31 VanLiere, J. M. & Rosenberg, N. A. Mathematical properties of the r2 measure of linkage disequilibrium. *Theor Popul Biol* **74**, 130-137, doi:10.1016/j.tpb.2008.05.006 (2008).
32 Teare, M. D. & Santibanez Koref, M. F. Linkage analysis and the study of Mendelian disease in the era of whole exome and genome sequencing. *Brief Funct Genomics* **13**, 378-383, doi:10.1093/bfgp/elu024 (2014).
33 Glazier, A. M., Nadeau, J. H. & Aitman, T. J. Finding genes that underlie complex traits. *Science* **298**, 2345-2349, doi:10.1126/science.1076641 (2002).
34 Martinez-Mir, A. *et al.* Genomewide scan for linkage reveals evidence of several susceptibility loci for alopecia areata. *Am J Hum Genet* **80**, 316-328, doi:10.1086/511442 (2007).
35 Martinez-Mir, A., Zlotogorski, A., Ott, J., Gordon, D. & Christiano, A. M. Genetic linkage studies in alopecia areata. *J Investig Dermatol Symp Proc* **8**, 199-203, doi:10.1046/j.1087-0024.2003.00809.x (2003).
36 Li, B., Liu, D. J. & Leal, S. M. Identifying rare variants associated with complex traits via sequencing. *Curr Protoc Hum Genet* **Chapter 1**, Unit 1 26, doi:10.1002/0471142905.hg0126s78 (2013).
37 Ha, N. T., Freytag, S. & Bickeboeller, H. Coverage and efficiency in current SNP chips. *Eur J Hum Genet* **22**, 1124-1130, doi:10.1038/ejhg.2013.304 (2014).
38 Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012, doi:10.1093/nar/gky1120 (2019).
39 Yu, C. *et al.* Low-frequency and rare variants may contribute to elucidate the genetics of major depressive disorder. *Transl Psychiatry* **8**, 70, doi:10.1038/s41398-018-0117-7 (2018).
40 Gorlov, I. P., Gorlova, O. Y., Frazier, M. L., Spitz, M. R. & Amos, C. I. Evolutionary evidence of the effect of rare variants on disease etiology. *Clin Genet* **79**, 199-206, doi:10.1111/j.1399-0004.2010.01535.x (2011).
41 Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387-389, doi:10.1126/science.1167728 (2009).
42 Momozawa, Y. *et al.* Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease. *Nat Genet* **43**, 43-47, doi:10.1038/ng.733 (2011).
43 Raychaudhuri, S. *et al.* A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. *Nat Genet* **43**, 1232-1236, doi:10.1038/ng.976 (2011).
44 Petersen, B. S., Fredrich, B., Hoeppner, M. P., Ellinghaus, D. & Franke, A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet* **18**, 14, doi:10.1186/s12863-017-0479-5 (2017).
45 Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333-351, doi:10.1038/nrg.2016.49 (2016).

46 Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A. & Jabado, N. What can exome sequencing do for you? *J Med Genet* **48**, 580-589, doi:10.1136/jmedgenet-2011-100223 (2011).
47 Kuhlenbaumer, G., Hullmann, J. & Appenzeller, S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Hum Mutat* **32**, 144-151, doi:10.1002/humu.21400 (2011).
48 Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K. & Mardis, E. R. The next-generation sequencing revolution and its impact on genomics. *Cell* **155**, 27-38, doi:10.1016/j.cell.2013.09.006 (2013).
49 Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* **95**, 5-23, doi:10.1016/j.ajhg.2014.06.009 (2014).
50 Povysil, G. *et al.* Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat Rev Genet* **20**, 747-759, doi:10.1038/s41576-019-0177-4 (2019).
51 Raghavan, N. S. *et al.* Whole-exome sequencing in 20,197 persons for rare variants in Alzheimer's disease. *Ann Clin Transl Neurol* **5**, 832-842, doi:10.1002/acn3.582 (2018).
52 Singh, T. *et al.* The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat Genet* **49**, 1167-1173, doi:10.1038/ng.3903 (2017).
53 Yang, C.-C., Chen, C.-C. & Chen, W.-C. in *Inflammation, Advancing Age and Nutrition* (eds Irfan Rahman & Debasis Bagchi) 231-246 (Academic Press, 2014).
54 Gilhar, A., Paus, R. & Kalish, R. S. Lymphocytes, neuropeptides, and genes involved in alopecia areata. *J Clin Invest* **117**, 2019-2027, doi:10.1172/JCI31942 (2007).
55 Villasante Fricke, A. C. & Miteva, M. Epidemiology and burden of alopecia areata: a systematic review. *Clin Cosmet Investig Dermatol* **8**, 397-403, doi:10.2147/CCID.S53985 (2015).
56 Tobin, D. J., Fenton, D. A. & Kendall, M. D. Ultrastructural study of exclamation-mark hair shafts in alopecia areata. *J Cutan Pathol* **17**, 348-354, doi:10.1111/j.1600-0560.1990.tb00111.x (1990).
57 Madani, S. & Shapiro, J. Alopecia areata update. *J Am Acad Dermatol* **42**, 549-566; quiz 567-570 (2000).
58 Pratt, C. H., King, L. E., Jr., Messenger, A. G., Christiano, A. M. & Sundberg, J. P. Alopecia areata. *Nat Rev Dis Primers* **3**, 17011, doi:10.1038/nrdp.2017.11 (2017).
59 Gilhar, A., Etzioni, A. & Paus, R. Alopecia areata. *N Engl J Med* **366**, 1515-1525, doi:10.1056/NEJMra1103442 (2012).
60 Xing, L. *et al.* Alopecia areata is driven by cytotoxic T lymphocytes and is reversed by JAK inhibition. *Nat Med* **20**, 1043-1049, doi:10.1038/nm.3645 (2014).
61 Arakelyan, A. *et al.* Autoimmunity and autoinflammation: A systems view on signaling pathway dysregulation profiles. *PLoS One* **12**, e0187572, doi:10.1371/journal.pone.0187572 (2017).
62 Zhernakova, A., Withoff, S. & Wijmenga, C. Clinical implications of shared genetics and pathogenesis in autoimmune diseases. *Nat Rev Endocrinol* **9**, 646-659, doi:10.1038/nrendo.2013.161 (2013).
63 Richard-Miceli, C. & Criswell, L. A. Emerging patterns of genetic overlap across autoimmune disorders. *Genome Med* **4**, 6, doi:10.1186/gm305 (2012).
64 Tobin, D. J., Fenton, D. A. & Kendall, M. D. Ultrastructural observations on the hair bulb melanocytes and melanosomes in acute alopecia areata. *J Invest Dermatol* **94**, 803-807, doi:10.1111/1523-1747.ep12874660 (1990).
65 Islam, N., Leung, P. S., Huntley, A. C. & Gershwin, M. E. The autoimmune basis of alopecia areata: a comprehensive review. *Autoimmun Rev* **14**, 81-89, doi:10.1016/j.autrev.2014.10.014 (2015).
66 Taheri, R., Behnam, B., Tousi, J. A., Azizzade, M. & Sheikhvatan, M. R. Triggering role of stressful life events in patients with alopecia areata. *Acta Dermatovenerol Croat* **20**, 246-250 (2012).
67 Blaumeiser, B. *et al.* Familial aggregation of alopecia areata. *J Am Acad Dermatol* **54**, 627-632, doi:10.1016/j.jaad.2005.12.007 (2006).
68 Jackow, C. *et al.* Alopecia areata and cytomegalovirus infection in twins: genes versus environment? *J Am Acad Dermatol* **38**, 418-425, doi:10.1016/s0190-9622(98)70499-2 (1998).
69 Rodriguez, T. A., Fernandes, K. E., Dresser, K. L., Duvic, M. & National Alopecia Areata, R. Concordance rate of alopecia areata in identical twins supports both genetic and environmental factors. *J Am Acad Dermatol* **62**, 525-527, doi:10.1016/j.jaad.2009.02.006 (2010).
70 Petukhova, L. *et al.* Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature* **466**, 113-117, doi:10.1038/nature09114 (2010).

71 Betz, R. C. *et al.* Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. *Nat Commun* **6**, 5966, doi:10.1038/ncomms6966 (2015).
72 Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* **104**, 65-75, doi:10.1016/j.ajhg.2018.11.008 (2019).
73 Rosengren Pielberg, G. *et al.* A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nat Genet* **40**, 1004-1009, doi:10.1038/ng.185 (2008).
74 Sundstrom, E. *et al.* Copy number expansion of the STX17 duplication in melanoma tissue from Grey horses. *BMC Genomics* **13**, 365, doi:10.1186/1471-2164-13-365 (2012).
75 Itakura, E., Kishi-Itakura, C. & Mizushima, N. The hairpin-type tail-anchored SNARE syntaxin 17 targets to autophagosomes for fusion with endosomes/lysosomes. *Cell* **151**, 1256-1269, doi:10.1016/j.cell.2012.11.001 (2012).
76 Hamasaki, M. *et al.* Autophagosomes form at ER-mitochondria contact sites. *Nature* **495**, 389-393, doi:10.1038/nature11910 (2013).
77 Xian, H., Yang, Q., Xiao, L., Shen, H. M. & Liou, Y. C. STX17 dynamically regulated by Fis1 induces mitophagy via hierarchical macroautophagic mechanism. *Nat Commun* **10**, 2059, doi:10.1038/s41467-019-10096-1 (2019).
78 Arasaki, K. *et al.* A role for the ancient SNARE syntaxin 17 in regulating mitochondrial division. *Dev Cell* **32**, 304-317, doi:10.1016/j.devcel.2014.12.011 (2015).
79 Ho, H. & Ganesan, A. K. The pleiotropic roles of autophagy regulators in melanogenesis. *Pigment Cell Melanoma Res* **24**, 595-604, doi:10.1111/j.1755-148X.2011.00889.x (2011).
80 Ganesan, A. K. *et al.* Genome-wide siRNA-based functional genomics of pigmentation identifies novel genes and pathways that impact melanogenesis in human cells. *PLoS Genet* **4**, e1000298, doi:10.1371/journal.pgen.1000298 (2008).
81 Kim, J. Y. *et al.* Autophagy induction can regulate skin pigmentation by causing melanosome degradation in keratinocytes and melanocytes. *Pigment Cell Melanoma Res* **33**, 403-415, doi:10.1111/pcmr.12838 (2020).
82 Yang, F. *et al.* Dysregulation of autophagy in melanocytes contributes to hypopigmented macules in tuberous sclerosis complex. *J Dermatol Sci* **89**, 155-164, doi:10.1016/j.jdermsci.2017.11.002 (2018).
83 Yang, Y. *et al.* Central role of autophagic UVRAG in melanogenesis and the suntan response. *Proc Natl Acad Sci U S A* **115**, E7728-E7737, doi:10.1073/pnas.1803303115 (2018).
84 Schneider, M. R., Schmidt-Ullrich, R. & Paus, R. The hair follicle as a dynamic miniorgan. *Curr Biol* **19**, R132-142, doi:10.1016/j.cub.2008.12.005 (2009).
85 Schmidt-Ullrich, R. & Paus, R. Molecular principles of hair follicle induction and morphogenesis. *Bioessays* **27**, 247-261, doi:10.1002/bies.20184 (2005).
86 Yang, C. C. & Cotsarelis, G. Review of hair follicle dermal cells. *J Dermatol Sci* **57**, 2-11, doi:10.1016/j.jdermsci.2009.11.005 (2010).
87 Hoover, E., Alhajj, M. & Flores, J. L. in *StatPearls* (2020).
88 Myung, P. & Ito, M. Dissecting the bulge in hair regeneration. *J Clin Invest* **122**, 448-454, doi:10.1172/JCI57414 (2012).
89 Moll, R., Divo, M. & Langbein, L. The human keratins: biology and pathology. *Histochem Cell Biol* **129**, 705-733, doi:10.1007/s00418-008-0435-6 (2008).
90 Johansen, C. Generation and Culturing of Primary Human Keratinocytes from Adult Skin. *J Vis Exp*, doi:10.3791/56863 (2017).
91 Eckert, R. L. & Rorke, E. A. Molecular biology of keratinocyte differentiation. *Environ Health Perspect* **80**, 109-116, doi:10.1289/ehp.8980109 (1989).
92 Deb-Choudhury, S. Crosslinking Between Trichocyte Keratins and Keratin Associated Proteins. *Adv Exp Med Biol* **1054**, 173-183, doi:10.1007/978-981-10-8195-8_12 (2018).
93 Alonso, L. & Fuchs, E. The hair cycle. *J Cell Sci* **119**, 391-393, doi:10.1242/jcs02793 (2006).
94 Kasumagic-Halilovic, E. & Prohic, A. Nail changes in alopecia areata: frequency and clinical presentation. *J Eur Acad Dermatol Venereol* **23**, 240-241, doi:10.1111/j.1468-3083.2008.02830.x (2009).
95 Tobin, D. J., Sundberg, J. P., King, L. E., Jr., Boggess, D. & Bystryn, J. C. Autoantibodies to hair follicles in C3H/HeJ mice with alopecia areata-like hair loss. *J Invest Dermatol* **109**, 329-333, doi:10.1111/1523-1747.ep12335848 (1997).

96 Colombo, S. B., I.; Delmas, V.; Larue, L. in *Melanins and Melanosomes* (ed J.; Riley Borovansky, P.A.) Ch. 2, 21-61 (2011).
97 Tobin, D. J. & Bystryn, J. C. Different populations of melanocytes are present in hair follicles and epidermis. *Pigment Cell Res* **9**, 304-310, doi:10.1111/j.1600-0749.1996.tb00122.x (1996).
98 Cichorek, M., Wachulska, M., Stasiewicz, A. & Tyminska, A. Skin melanocytes: biology and development. *Postepy Dermatol Alergol* **30**, 30-41, doi:10.5114/pdia.2013.33376 (2013).
99 Wasmeier, C., Hume, A. N., Bolasco, G. & Seabra, M. C. Melanosomes at a glance. *J Cell Sci* **121**, 3995-3999, doi:10.1242/jcs.040667 (2008).
100 Delevoye, C. G. F. M., M.S.; Raposo, G. in *Melanins and Melanosomes* (ed J.; Riley Borovansky, P.A.) Ch. 9, 247-294 (2011).
101 Kushimoto, T. *et al.* A model for melanosome biogenesis based on the purification and analysis of early melanosomes. *Proc Natl Acad Sci U S A* **98**, 10698-10703, doi:10.1073/pnas.191184798 (2001).
102 Hoashi, T. *et al.* MART-1 is required for the function of the melanosomal matrix protein PMEL17/GP100 and the maturation of melanosomes. *J Biol Chem* **280**, 14006-14016, doi:10.1074/jbc.M413692200 (2005).
103 Kawakami, Y. *et al.* T cell immune responses against melanoma and melanocytes in cancer and autoimmunity. *Pigment Cell Res* **13 Suppl 8**, 163-169, doi:10.1034/j.1600-0749.13.s8.29.x (2000).
104 Cui, T. *et al.* Identification of Novel HLA-A*0201-Restricted CTL Epitopes in Chinese Vitiligo Patients. *Sci Rep* **6**, 36360, doi:10.1038/srep36360 (2016).
105 Palermo, B. *et al.* Specific cytotoxic T lymphocyte responses against Melan-A/MART1, tyrosinase and gp100 in vitiligo by the use of major histocompatibility complex/peptide tetramers: the role of cellular immunity in the etiopathogenesis of vitiligo. *J Invest Dermatol* **117**, 326-332, doi:10.1046/j.1523-1747.2001.01408.x (2001).
106 Parodi, C. *et al.* Autophagy is essential for maintaining the growth of a human (mini-)organ: Evidence from scalp hair follicle organ culture. *PLoS Biol* **16**, e2002864, doi:10.1371/journal.pbio.2002864 (2018).
107 Paus, R., Ito, N., Takigawa, M. & Ito, T. The hair follicle and immune privilege. *J Investig Dermatol Symp Proc* **8**, 188-194, doi:10.1046/j.1087-0024.2003.00807.x (2003).
108 Arin, M. J. The molecular basis of human keratin disorders. *Hum Genet* **125**, 355-373, doi:10.1007/s00439-009-0646-5 (2009).
109 Hershkovitz, D. & Sprecher, E. Monogenic pigmentary skin disorders: genetics and pathophysiology. *Isr Med Assoc J* **10**, 713-717 (2008).
110 Langbein, L. & Schweizer, J. Keratins of the human hair follicle. *Int Rev Cytol* **243**, 1-78, doi:10.1016/S0074-7696(05)43001-6 (2005).
111 DeStefano, G. M. & Christiano, A. M. The genetics of human skin disease. *Cold Spring Harb Perspect Med* **4**, doi:10.1101/cshperspect.a015172 (2014).
112 Berkowitz, P., Chua, M., Liu, Z., Diaz, L. A. & Rubenstein, D. S. Autoantibodies in the autoimmune disease pemphigus foliaceus induce blistering via p38 mitogen-activated protein kinase-dependent signaling in the skin. *Am J Pathol* **173**, 1628-1636, doi:10.2353/ajpath.2008.080391 (2008).
113 Gregersen, P. K. & Olsson, L. M. Recent advances in the genetics of autoimmune disease. *Annu Rev Immunol* **27**, 363-391, doi:10.1146/annurev.immunol.021908.132653 (2009).
114 Navarini, A. A., Nobbe, S. & Trueb, R. M. Marie Antoinette syndrome. *Arch Dermatol* **145**, 656, doi:10.1001/archdermatol.2009.51 (2009).
115 Nahm, M., Navarini, A. A. & Kelly, E. W. Canities subita: a reappraisal of evidence based on 196 case reports published in the medical literature. *Int J Trichology* **5**, 63-68, doi:10.4103/0974-7753.122959 (2013).
116 Asz-Sigall, D. *et al.* White hair in alopecia areata: Clinical forms and proposed physiopathological mechanisms. *J Am Acad Dermatol*, doi:10.1016/j.jaad.2018.12.047 (2019).
117 Paus, R., Slominski, A. & Czarnetzki, B. M. Is alopecia areata an autoimmune-response against melanogenesis-related proteins, exposed by abnormal MHC class I expression in the anagen hair bulb? *Yale J Biol Med* **66**, 541-554 (1993).

118 Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491-504, doi:10.1038/s41576-018-0016-z (2018).
119 Gelfman, S. *et al.* A new approach for rare variation collapsing on functional protein domains implicates specific genic regions in ALS. *Genome Res* **29**, 809-818, doi:10.1101/gr.243592.118 (2019).
120 Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555-3557, doi:10.1093/bioinformatics/btv402 (2015).
121 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *bioRxiv*, 531210, doi:10.1101/531210 (2020).
122 Spielmann, M. & Mundlos, S. Looking beyond the genes: the role of non-coding variants in human disease. *Hum Mol Genet* **25**, R157-R165, doi:10.1093/hmg/ddw205 (2016).
123 Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585, doi:10.1038/ng.2653 (2013).
124 Jabbari, A. *et al.* Molecular signatures define alopecia areata subtypes and transcriptional biomarkers. *EBioMedicine* **7**, 240-247, doi:10.1016/j.ebiom.2016.03.036 (2016).
125 Denat, L., Kadekaro, A. L., Marrot, L., Leachman, S. A. & Abdel-Malek, Z. A. Melanocytes as instigators and victims of oxidative stress. *J Invest Dermatol* **134**, 1512-1518, doi:10.1038/jid.2014.65 (2014).
126 Videira, I. F., Moura, D. F. & Magina, S. Mechanisms regulating melanogenesis. *An Bras Dermatol* **88**, 76-83 (2013).
127 Yamaguchi, Y. & Hearing, V. J. Melanocytes and their diseases. *Cold Spring Harb Perspect Med* **4**, doi:10.1101/cshperspect.a017046 (2014).
128 Yamamoto, A. *et al.* Bafilomycin A1 prevents maturation of autophagic vacuoles by inhibiting fusion between autophagosomes and lysosomes in rat hepatoma cell line, H-4-II-E cells. *Cell Struct Funct* **23**, 33-42, doi:10.1247/csf.23.33 (1998).
129 Yamaguchi, Y., Brenner, M. & Hearing, V. J. The regulation of skin pigmentation. *J Biol Chem* **282**, 27557-27561, doi:10.1074/jbc.R700026200 (2007).
130 Sitaram, A. & Marks, M. S. Mechanisms of protein delivery to melanosomes in pigment cells. *Physiology (Bethesda)* **27**, 85-99, doi:10.1152/physiol.00043.2011 (2012).
131 Gilhar, A. *et al.* Melanocyte-associated T cell epitopes can function as autoantigens for transfer of alopecia areata to human scalp explants on Prkdc(scid) mice. *J Invest Dermatol* **117**, 1357-1362, doi:10.1046/j.0022-202x.2001.01583.x (2001).
132 Mandelcorn-Monson, R. L. *et al.* Cytotoxic T lymphocyte reactivity to gp100, MelanA/MART-1, and tyrosinase, in HLA-A2-positive vitiligo patients. *J Invest Dermatol* **121**, 550-556, doi:10.1046/j.1523-1747.2003.12413.x (2003).
133 Flannick, J. *et al.* Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* **570**, 71-76, doi:10.1038/s41586-019-1231-2 (2019).
134 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).
135 Shohet, A. *et al.* Variant in SCYL1 gene causes aberrant splicing in a family with cerebellar ataxia, recurrent episodes of liver failure, and growth retardation. *Eur J Hum Genet*, doi:10.1038/s41431-018-0268-2 (2018).
136 Okur, V. *et al.* Homozygous noncanonical splice variant in LSM1 in two siblings with multiple congenital anomalies and global developmental delay. *Cold Spring Harb Mol Case Stud* **5**, doi:10.1101/mcs.a004101 (2019).
137 Hausman-Kedem, M. *et al.* VPS53 gene is associated with a new phenotype of complicated hereditary spastic paraparesis. *Neurogenetics* **20**, 187-195, doi:10.1007/s10048-019-00586-1 (2019).
138 Yamaguchi-Kabata, Y. *et al.* Distribution and effects of nonsense polymorphisms in human genes. *PLoS One* **3**, e3393, doi:10.1371/journal.pone.0003393 (2008).
139 MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823-828, doi:10.1126/science.1215040 (2012).
140 Langbein, L., Rogers, M. A., Winter, H., Praetzel, S. & Schweizer, J. The catalog of human hair keratins. II. Expression of the six type II members in the hair follicle and the combined catalog of

human type I and II keratins. *J Biol Chem* **276**, 35123-35132, doi:10.1074/jbc.M103305200 (2001).
141 Kiso, M. *et al.* The disruption of Sox21-mediated hair shaft cuticle differentiation causes cyclic alopecia in mice. *Proc Natl Acad Sci U S A* **106**, 9292-9297, doi:10.1073/pnas.0808324106 (2009).
142 Yu, Z. *et al.* Annotation of sheep keratin intermediate filament genes and their patterns of expression. *Exp Dermatol* **20**, 582-588, doi:10.1111/j.1600-0625.2011.01274.x (2011).
143 Smith, F. The molecular genetics of keratin disorders. *Am J Clin Dermatol* **4**, 347-364, doi:10.2165/00128071-200304050-00005 (2003).
144 Wu, K. C. *et al.* Coiled-coil trigger motifs in the 1B and 2B rod domain segments are required for the stability of keratin intermediate filaments. *Mol Biol Cell* **11**, 3539-3558, doi:10.1091/mbc.11.10.3539 (2000).
145 Hatzfeld, M. & Burba, M. Function of type I and type II keratin head domains: their role in dimer, tetramer and filament formation. *J Cell Sci* **107 (Pt 7)**, 1959-1972 (1994).
146 Roca, I., Fernandez-Marmiesse, A., Gouveia, S., Segovia, M. & Couce, M. L. Prioritization of Variants Detected by Next Generation Sequencing According to the Mutation Tolerance and Mutational Architecture of the Corresponding Genes. *Int J Mol Sci* **19**, doi:10.3390/ijms19061584 (2018).
147 Sun, H. & Yu, G. New insights into the pathogenicity of non-synonymous variants through multi-level analysis. *Sci Rep* **9**, 1667, doi:10.1038/s41598-018-38189-9 (2019).
148 Kim, D. *et al.* Evolutionary coupling analysis identifies the impact of disease-associated variants at less-conserved sites. *Nucleic Acids Res*, doi:10.1093/nar/gkz536 (2019).
149 Cotsarelis, G. & Botchkarev, V. in *Fitzpatrick's Dermatology in General Medicine, 8e* (eds A. Lowell, L. Goldsmith, & et al.) (McGraw-Hill, 2012).
150 Hut, P. H. *et al.* Exempting homologous pseudogene sequences from polymerase chain reaction amplification allows genomic keratin 14 hotspot mutation analysis. *J Invest Dermatol* **114**, 616-619, doi:10.1046/j.1523-1747.2000.00928.x (2000).
151 Stephens, K. *et al.* Primers for exon-specific amplification of the KRT5 gene: identification of novel and recurrent mutations in epidermolysis bullosa simplex patients. *J Invest Dermatol* **108**, 349-353, doi:10.1111/1523-1747.ep12286486 (1997).
152 Ma, L., Yamada, S., Wirtz, D. & Coulombe, P. A. A 'hot-spot' mutation alters the mechanical properties of keratin filament networks. *Nat Cell Biol* **3**, 503-506, doi:10.1038/35074576 (2001).
153 Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-181, doi:10.1038/nmeth.1785 (2011).
154 Ruzicka, M. *et al.* DNA mutation motifs in the genes associated with inherited diseases. *PLoS One* **12**, e0182377, doi:10.1371/journal.pone.0182377 (2017).
155 Youssoufian, H. *et al.* Recurrent mutations in haemophilia A give evidence for CpG mutation hotspots. *Nature* **324**, 380-382, doi:10.1038/324380a0 (1986).
156 Lieb, M. Spontaneous mutation at a 5-methylcytosine hotspot is prevented by very short patch (VSP) mismatch repair. *Genetics* **128**, 23-27 (1991).
157 Baugh, E. H., Ke, H., Levine, A. J., Bonneau, R. A. & Chan, C. S. Why are there hotspot mutations in the TP53 gene in human cancers? *Cell Death Differ* **25**, 154-160, doi:10.1038/cdd.2017.180 (2018).
158 Cooper, D. N., Mort, M., Stenson, P. D., Ball, E. V. & Chuzhanova, N. A. Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Hum Genomics* **4**, 406-410, doi:10.1186/1479-7364-4-6-406 (2010).
159 Tong, X. & Coulombe, P. A. Keratin 17 modulates hair follicle cycling in a TNFalpha-dependent fashion. *Genes Dev* **20**, 1353-1364, doi:10.1101/gad.1387406 (2006).
160 Wolfram, L. J. Human hair: a unique physicochemical composite. *J Am Acad Dermatol* **48**, S106-114, doi:10.1067/mjd.2003.276 (2003).
161 Hedstrand, H. *et al.* Antibodies against hair follicles are associated with alopecia totalis in autoimmune polyendocrine syndrome type I. *J Invest Dermatol* **113**, 1054-1058, doi:10.1046/j.1523-1747.1999.00778.x (1999).
162 Tobin, D. J., Orentreich, N., Fenton, D. A. & Bystryn, J. C. Antibodies to hair follicles in alopecia areata. *J Invest Dermatol* **102**, 721-724, doi:10.1111/1523-1747.ep12375477 (1994).

163 Coulie, P. G. *et al.* A new gene coding for a differentiation antigen recognized by autologous cytolytic T lymphocytes on HLA-A2 melanomas. *J Exp Med* **180**, 35-42, doi:10.1084/jem.180.1.35 (1994).
164 Kawakami, Y. *et al.* Cloning of the gene coding for a shared human melanoma antigen recognized by autologous T cells infiltrating into tumor. *Proc Natl Acad Sci U S A* **91**, 3515-3519, doi:10.1073/pnas.91.9.3515 (1994).
165 Rivoltini, L. *et al.* Induction of tumor-reactive CTL from peripheral blood and tumor-infiltrating lymphocytes of melanoma patients by in vitro stimulation with an immunodominant peptide of the human melanoma antigen MART-1. *J Immunol* **154**, 2257-2265 (1995).
166 Butterfield, L. H. *et al.* Adenovirus MART-1-engineered autologous dendritic cell vaccine for metastatic melanoma. *J Immunother* **31**, 294-309, doi:10.1097/CJI.0b013e31816a8910 (2008).
167 Butterfield, L. H. *et al.* Generation of melanoma-specific cytotoxic T lymphocytes by dendritic cells transduced with a MART-1 adenovirus. *J Immunol* **161**, 5607-5613 (1998).
168 Ogg, G. S., Rod Dunbar, P., Romero, P., Chen, J. L. & Cerundolo, V. High frequency of skin-homing melanocyte-specific cytotoxic T lymphocytes in autoimmune vitiligo. *J Exp Med* **188**, 1203-1208, doi:10.1084/jem.188.6.1203 (1998).
169 Lang, K. S. *et al.* HLA-A2 restricted, melanocyte-specific CD8(+) T lymphocytes detected in vitiligo patients are related to disease activity and are predominantly directed against MelanA/MART1. *J Invest Dermatol* **116**, 891-897, doi:10.1046/j.1523-1747.2001.01363.x (2001).
170 Sugita, S., Sagawa, K., Mochizuki, M., Shichijo, S. & Itoh, K. Melanocyte lysis by cytotoxic T lymphocytes recognizing the MART-1 melanoma antigen in HLA-A2 patients with Vogt-Koyanagi-Harada disease. *Int Immunol* **8**, 799-803, doi:10.1093/intimm/8.5.799 (1996).
171 Setz, C. *et al.* Just one position-independent lysine residue can direct MelanA into proteasomal degradation following N-terminal fusion of ubiquitin. *PLoS One* **8**, e55567, doi:10.1371/journal.pone.0055567 (2013).
172 Ohbayashi, N. & Fukuda, M. SNARE dynamics during melanosome maturation. *Biochem Soc Trans* **46**, 911-917, doi:10.1042/BST20180130 (2018).
173 Yatsu, A., Ohbayashi, N., Tamura, K. & Fukuda, M. Syntaxin-3 is required for melanosomal localization of Tyrp1 in melanocytes. *J Invest Dermatol* **133**, 2237-2246, doi:10.1038/jid.2013.156 (2013).
174 Jani, R. A., Purushothaman, L. K., Rani, S., Bergam, P. & Setty, S. R. STX13 regulates cargo delivery from recycling endosomes during melanosome biogenesis. *J Cell Sci* **128**, 3263-3276, doi:10.1242/jcs.171165 (2015).
175 Betz, R. C. *et al.* Loss-of-function mutations in the keratin 5 gene lead to Dowling-Degos disease. *Am J Hum Genet* **78**, 510-519, doi:10.1086/500850 (2006).
176 Lugassy, J. *et al.* KRT14 haploinsufficiency results in increased susceptibility of keratinocytes to TNF-alpha-induced apoptosis and causes Naegeli-Franceschetti-Jadassohn syndrome. *J Invest Dermatol* **128**, 1517-1524, doi:10.1038/sj.jid.5701187 (2008).
177 Chen, J. C., Cerise, J. E., Jabbari, A., Clynes, R. & Christiano, A. M. Master regulators of infiltrate recruitment in autoimmune disease identified through network-based molecular deconvolution. *Cell Syst* **1**, 326-337, doi:10.1016/j.cels.2015.11.001 (2015).
178 Whiting, D. A. Histopathologic features of alopecia areata: a new look. *Arch Dermatol* **139**, 1555-1559, doi:10.1001/archderm.139.12.1555 (2003).
179 Toivola, D. M., Strnad, P., Habtezion, A. & Omary, M. B. Intermediate filaments take the heat as stress proteins. *Trends Cell Biol* **20**, 79-91, doi:10.1016/j.tcb.2009.11.004 (2010).
180 Lemasters, J. J. *et al.* Compartmentation of Mitochondrial and Oxidative Metabolism in Growing Hair Follicles: A Ring of Fire. *J Invest Dermatol* **137**, 1434-1444, doi:10.1016/j.jid.2017.02.983 (2017).
181 Prie, B. E. *et al.* Oxidative stress and alopecia areata. *J Med Life* **8 Spec Issue**, 43-46 (2015).
182 Akar, A. *et al.* Antioxidant enzymes and lipid peroxidation in the scalp of patients with alopecia areata. *J Dermatol Sci* **29**, 85-90, doi:10.1016/s0923-1811(02)00015-4 (2002).