Genetic and Environmental Determinants of Alopecia Areata

Alexa R. Abdelaziz

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

Genetic and Environmental Determinants of Alopecia Areata
Alexa R. Abdelaziz

Alopecia Areata (AA) is a highly prevalent autoimmune disease in the US with a lifetime risk of 2.1%. In AA, autoimmunity develops against the hair follicles, which leads to infiltration of immune cells around affected follicles. Among genetic risk factors in complex autoimmune diseases, variants cluster in genes regulating the immune response, as well as the target organ. AA is believed to result from both genetic and environmental factors. To identify underlying genetic drivers in AA, we analyzed AA risk genes using various sequencing techniques and analysis methods to identify causal variants and placed them in functionally relevant contexts using innovative mapping techniques.

To address the role of variants in immune function, we studied the Interleukin-2 Receptor Alpha (IL2RA), which we identified as a significant locus to study genetic factors underlying immune function from our AA GWAS studies \( (p=1.74 \times 10^{-12}) \). IL2RA plays a crucial role in regulating immune tolerance and controlling activity of regulatory T cells \( (T_{\text{reg}}) \). We identified significant causal variants in the IL2RA region associated with AA using GWAS, targeted resequencing, and custom capture exome sequencing approaches. We validated the expression of these variants in immune cell cluster tissue types \textit{in silico}, and specifically in CD4+ T cells. The variant rs3118740 increases AA susceptibility for carriers of the C allele. Such allele specific effects could lead to a perturbation of \( T_{\text{reg}} \) function, for example, one study in T1D where patients with the rs3118470 risk variant have \( T_{\text{reg}} \) with IL-2 signaling defects. These studies demonstrated that identifying causal variants may lead to an improved understating of \( T_{\text{reg}} \) function and risk of autoimmunity in AA.

Next, to study genetic susceptibility in the target organ in AA, the hair follicle (HF), the second candidate GWAS susceptibility gene we studied was peroxiredoxin 5 (PRDX5) \( (p= 8.7 \times 10^{-14}) \), which is also a GWAS gene in Crohn’s disease, sarcoidosis, and psoriasis. PRDX5
is a member of the family of antioxidant enzymes that are crucial for regulating oxidative stress. Our lab performed whole exome sequencing in 849 AA patients, together with selected custom capture regions of genomic sequencing. Using a test of variant enrichment, we identified variants in PRDX5 that were significant in both our GWAS and exome studies, and thus represented likely candidate causal variants. Using Bayesian fine mapping, we identified a GWAS and exome sequencing variant, rs574087, that was significantly enriched in both, and is predicted to be a causal variant in keratinocytes and melanocytes. To functionally validate PRDX5, we immunostained healthy human HF and AA affected HF, and found that PRDX5 is upregulated AA human HF. PRDX5 is expressed in cultured melanocytes by immunostaining, which is consistent with melanocytes exhibiting high levels of oxidative stress. We postulated that PRDX5 may be involved in protection from oxidative stress, and that its dysregulation may contribute to autoimmunity.

Finally, along with genetic predisposition, environmental triggers such as the microbiome have emerged as potential factors contributing to pathologic immune responses in autoimmune diseases. To determine the role of the microbiome in AA pathobiology, we performed 16S rRNA sequencing of skin swabs, hair follicles, and stool samples from a cohort of 34 AA patients and 12 healthy controls (HCs). Unexpectedly, we found evidence of striking gut dysbiosis, consisting of over-representation of Firmicutes and under-representation of Bacteroides in the gut microbiome of AA patients compared with healthy subjects, but no significant differences in skin or hair follicle (HF) microbiome composition. To investigate the role of the gut microbiome in AA development in vivo, we depleted the gut microbiome in C3H/HeJ mice and found that the mice were largely protected from AA developing. These data revealed a requirement for gut microbiota in the onset of murine AA. Taken together with recent reports in the literature of reversal of AA in several patients following fecal microbiota transplant (FMT)\textsuperscript{17,18}, our findings suggest that restoring homeostasis of the gut microbiome may represent an effective new treatment modality in the management of AA.
# Table of Contents

List of Figures and Illustrations ................................................................. iii
Acknowledgments .................................................................................. v
Dedication ............................................................................................... ix

Chapter 1. General Introduction .............................................................. 1

1.1 The Genetic Basis of Alopecia Areata ................................................ 1
   A. Disease Background and Immunopathology ...................................... 1
   B. GWAS ......................................................................................... 3
   C. GWAS Studies in AA ................................................................. 4
   D. Targeted Genomic Sequencing ..................................................... 6
   E. Whole Exome and Targeted Genomic Sequencing ......................... 7
   F. Analytical Methods ...................................................................... 7
   G. Functional Genomics Studies ...................................................... 8
   H. Genotype-dependent Changes in Gene expression ......................... 9
I. FUN-LDA Algorithm ......................................................................... 9
J. C3H/HeJ Mouse Model of AA .......................................................... 9
K. Scarring Alopecia ........................................................................... 10

1.2 Environmental Factors Contributing to Complex Disease ................ 11
   A. Human Microbiome ................................................................... 12
   B. Autoimmunity and the Microbiome .............................................. 13
   C. The Leaky Gut .......................................................................... 13
   D. Short Chain Fatty Acids ............................................................... 15

1.3 Work Described in this Thesis ......................................................... 16

Chapter 2. Immune System: Fine mapping of the IL2RA locus in AA ....... 18

2.1 Introduction .................................................................................. 18
2.2 Targeted Enrichment Pipeline Analysis .......................................... 19
2.3 Whole Exome and Custom Capture Sequencing .............................. 23
2.4 Genotype-dependent Expression ................................................... 24
2.5 Functional Relevance in Tissue Types ............................................ 25
2.6 IL2RA in Autoimmunity .................................................................. 28
2.7 Materials and Methods .................................................................. 29
2.8 Conclusions .................................................................................. 33

Chapter 3. Hair Follicle: Fine Mapping and Functional Studies of PRDX5 ... 35

3.1 Introduction .................................................................................. 35
3.2 Fine mapping of GWAS region PRDX5 ......................................... 37
3.3 Genotype-dependent Expression ................................................... 39
3.4 Functional Analysis of PRDX5 Locus ............................................ 40
3.5 Expression of PRDX5 in HF and Melanocytes ................................. 42
3.6 Materials and Methods .................................................................. 44
3.7 Conclusions .................................................................................. 45

Chapter 4. Environmental Factors: Gut Microbiome Dysbiosis in AA ...... 47

4.1 Introduction .................................................................................. 47
4.2 16S rRNA sequencing of AA Patients ........................................... 48
4.3 Depletion of Gut Microbiome in C3H/HeJ Mouse model of Alopecia prevents AA ......................................................... 56
4.4 Metagenomics Sequencing Revealed AA-specific microbes in patients ......................................................... 59
4.5 Short Chain Fatty Acid Analysis .................................................................63
4.6 Materials and Methods ...........................................................................64
4.7 Conclusions ...............................................................................................68

Chapter 5. Discussion .....................................................................................70
  5.1 Overview Section ....................................................................................70
  5.2 Genetic Factors in AA ...........................................................................71
  5.3 Environmental Factors in AA ...............................................................73
  5.4 Future Directions: Translational Microbiome Studies .........................75

References ......................................................................................................77
List of Figures and Illustrations

Introduction
Figure 1.1. Clinical manifestations of Alopecia
Figure 1.2. ‘Swarm of bees’ immune infiltrate and collapse of immune privilege seen in AA versus normal anagen hair follicles
Figure 1.3. Statistically significant genes identified in GWAS and meta-analysis studies
Figure 1.4. Peroxiredoxin 5 (PRDX5) is expressed in human HF
Table 1.1. Comparison of sequencing approaches.
Figure 1.5. Tissue-specific responses of host and commensals at barrier sites
Figure 1.6. Potential mechanism in Type 1 Diabetes induction, allowing microbial antigens to enter the gut and induce disease

Chapter 2
Table 2.1. Regions prioritized in our targeted resequencing studies
Figure 2.1. Analytical pipeline for exploring genomic regions with targeted resequencing data
Figure 2.2. All 458 variants in IL2RA region identified from targeted resequencing
Table 2.2. Functional annotation of rare enriched variants
Figure 2.3. All rare enriched variants in IL2RA region identified from targeted resequencing
Figure 2.4. Tissue Distribution of IL2RA Rare Enriched Variants
Table 2.3. Exome data replicates enrichment of a subset of GWAS and Targeted Resequencing Variants
Figure 2.5. Genotype dependent expression differences from whole skin biopsies in patients
Figure 2.6. In silico studies of IL2RA variants in immune cells.
Figure 2.7. IL2RA variants in the context of AA and other autoimmune diseases.
Table 2.4. AA significant risk variants implicated in other autoimmune diseases

Chapter 3
Figure 3.1. The role of Antioxidant Enzymes
Table 3.1. Exome data replicates enrichment of a subset of GWAS Variants
Figure 3.2. Genotype differences from whole skin biopsies in patients
Figure 3.3. In silico validation of PRDX5 variants in keratinocytes
Figure 3.4. In silico validation of PRDX5 variants in melanocytes
Figure 3.5. FUN-LDA analysis of the PRDX5 region
Figure 3.6. PRDX5 is overexpressed in AA human HF
Figure 3.7. PRDX5 is expressed in Human Melanocytes

Chapter 4
Table 4.1. Clinical Data for all AA patients in the microbiome cohort
Figure 4.1. Skin and hair follicle microbiome in AA show no dysbiosis between patients and healthy controls
Figure 4.2. Microbiome Analysis in 34 patients reveals distinct clustering of AA patients compared to controls
Figure 4.3. Most significant microbial genera associated with AA
Figure 4.4. Training-Validation workflow shows differences in bacterial composition in the gut microbiome between patients and healthy controls
Table 4.2. Directionality of changes in composition of Firmicutes and Bacteroides across different diseases
Figure 4.5. Scarring Alopecias show no gut dysbiosis
Figure 4.6. Depletion of gut microbiome prevents onset of AA
Figure 4.7. Reduced gut microbiome diversity in AA
Figure 4.8. Gut Microbiome Beta Diversity in AA
Figure 4.9. Metagenomic differential abundance of species
Figure 4.10. Representative microbes identified by metagenomic sequencing
Figure 4.11. Short Chain Fatty Acid Analysis in AA Patients

Chapter 5
Figure 5.1. Patients treated with FMT resulted in hair regrowth
Acknowledgments

There are so many wonderful and important people I would like to thank -- for their support, guidance, friendship, love, and encouragement throughout this whole process, both in my professional and personal development. I am so grateful to be surrounded by such wonderful family, friends, and colleagues.

I would like to first thank my mentor, Dr. Angela Christiano. I have genuinely learned so much being in her group and am grateful for her support and willingness to adapt my projects to better suit my future goals and interests. I have grown exponentially during the past five years and a huge part of that was her challenging me to step out of my comfort zone, learn, be patient, and guide me on how to think and ask questions like a true scientist. I am thankful for the many rewarding and unique opportunities that Dr. Christiano made possible, such as attending numerous national and international conferences. The opportunities to share my work, enhance my public speaking skills, and develop my scientific confidence were so essential to my training. I am truly grateful for her willingness to go above and beyond to make this as seamless of a PhD process as possible.

Another key group of wonderful people I want to thank are the members of the Christiano lab. I truly cannot say enough amazing things about this group of people. Supportive, fun, and encouraging, this was the most incredible environment to work in during these past five years. Whether it was helping with an experiment, talking about science, complaining about any and everything, or just having random conversations over lunch, interacting with the group made my time truly memorable and honestly made a very challenging PhD process so much more enjoyable. First off, I have to thank my partner in crime through this whole PhD process, Steph. I genuinely could not have done this without her. Between staying late in lab together, having a frustrating day and going to get drinks, venting about any and everything, or working together on
projects, she absolutely kept me sane and made this process so much more fun. It has been the most insane journey and there’s no one else I’d rather have by my side. All my love and gratitude.

I would like to thank Rolando, Eddy, and James for their unparalleled help and guidance along this journey. Regardless of how busy they were, they were always there to answer my questions, help explain complex concepts, and truly help me grow scientifically. I have to thank Rolando for his patience with me on my mouse work, along with all of our cooperation on projects and scientific brainstorming sessions throughout the years. Eddy, for his wonderful scientific advice and comments that always made me laugh. James, for all of the bioinformatic guidance and general advice. Eunice, for being the best sounding board and for her thoughtfulness. Lynn, for all of her support and guidance throughout this process. Gwen, Erbil, Zongyou, Rupali, Jojo, Lauren, Corey, Avina, Alberto, Isha, Jin, Zhenpeng, Ming, Jade and Emily, I genuinely cannot thank them all enough for being there with good advice and support and always a smile and a laugh. I also would like to thank some key former lab members, Etienne and Brigitte, for their direction and support; thank you for being flexible and incredible people to work with- even willing to take my blood at my desk because extra control samples were needed. Yanne was one of the most pivotal people thought my PhD, whether it was career or networking advice, life advice, or just being a good friend, she were always there for a laugh or to listen to me and I truly value her friendship.

I would also like to thank the Institute of Human Nutrition (IHN). Debra, Richard, Tony, Zach, Alex, Sara, Leslie; they took a chance on me as a student from an engineering background, and I’m so grateful I was able to be a part of the IHN and the NMB PhD program. It has been such an enjoyable process. I cannot thank them enough for all of their help, support, advice, patience, and so much more. I would like to thank my colleagues in the IHN and the deans, Mao and Art, for all of their support and encouragement throughout this whole process. I would like to thank Ron Liem for allowing me to TA the Mechanisms of Human Disease course, and the wonderful students over the years for making that such a rewarding and fulfilling experience. I would like to
thank both my qualifying and thesis committees and our lab’s wonderful collaborators for their support, advice, and patience. I would also like to thank my mentors, past and present, whether they welcomed me into their lab for undergraduate research, a rotation, or a collaboration, I am eternally grateful for all of the wonderful experiences that have helped me grow as a scientist and as a person.

I want to give a huge thanks to my friends from home, Columbia, Rutgers, and my previous career. I am blessed to have truly the most amazing support system. Whether they were there for a conversation, helping with homework, sending me coffee gift cards, travelling, working out, watching my practice talks, tailgating at Rutgers, snail-mailing each other cards, allowing me to drag them out, or just having a good time and a laugh, I am so grateful and blessed to have so many wonderful, supportive friends in my life. I thank them all from the bottom of my heart. Special thanks to Sara and Alexa for being the best 5th Abdelaziz sisters and Kumps, Ryan, Sebastian, KP, Cat, Sara P, Kate, Elise, and everyone else for their love and support.

I have to give a huge thanks to my incredible boyfriend, Nick. I truly cannot thank him enough for his support. He saw the worst of it all- the tears, the practice talks, the stress- every step of the way he was there with a smile, a funny text/call, or just something small to cheer me up and remind me this was all worth it and that I was more than capable. He was and is always there to lift my spirits and keep me positive, and I could not ask for a better support system. I love you and cannot thank you enough.

The biggest thank you has to go to my family- Mom, Dad, Sas, Maya, and Nora. From the beginning of my schooling, they have all encouraged me to work hard and follow every ambition. I am so grateful for the wonderful support, thoughtful advice, encouragement, and laughs. Mom and Dad, thank you for always being there with a supportive text or call and talking me out of stressful situations when I needed it the most, for being amazing sounding boards, and for supporting me in any pursuit I set my mind to, including pursuing my doctorate degree. I would like to give a huge thanks to all three of my younger sisters. I would like to thank Sas, my best
friend, who I always look to for advice, most of it usually reminding me to be confident and let things brush off my shoulders. I thank her for always being there to chat or just distract me from getting stressed. Maya and Nora, while the majority of our conversations consist of teasing me for being old and uncool, I know they do it to make me laugh and they have honestly always managed to do just that. I love how they keep me sane, grounded, and constantly entertained with funny memes, thoughtful gestures, and on-demand pictures of our dogs, Zeus and Thor. My three younger sisters remind me why I strive to be better each and every day to be a good role model. I am so thankful to have such a great group of sisters and the best parents. I love them all so much and thanks again to everyone who has made this process such an enjoyable and formational journey!
Dedication

To my loving family and friends who have always been the most supportive and encouraging:
Thank you for helping me pursue my dreams and realize what I am truly capable of; I love and appreciate you all more than anything.
Chapter 1. General Introduction

1.1 The Genetic Basis of Alopecia Areata

A. Disease Background and Immunopathology

Alopecia Areata (AA) is a highly prevalent organ-specific autoimmune disorder with a lifetime prevalence of 2.1%\textsuperscript{19}. In AA, autoimmunity develops against the hair follicle (HF), which leads to infiltration of immune cells around affected follicles and is believed to involve both genetic and environmental factors. There are three main clinical presentations of alopecia: alopecia areata (AA) which is patchy hair loss, alopecia totalis (AT) which is total loss of scalp hair, and alopecia universalis (AU) which is total loss of scalp and body hair (Figure 1.1).

![Figure 1.1 Clinical manifestations of Alopecia. Left to right, Alopecia Areata (AA patchy hair loss), Alopecia Totalis (AT, total loss of scalp hair), and Alopecia Universalis (AU, total loss of scalp and body hair)\textsuperscript{20,21}}

The hair follicle has innate immune privilege, suggesting that it can tolerate antigen introduction without eliciting inflammatory immune responses\textsuperscript{22}. Other body sites that exhibit immune privilege are sites like the eyes, placenta, and testes\textsuperscript{23}. The normal hair cycle is a process in which human HFs are continuously transformed in a repetitive cycle of organ construction and deconstruction. The first stage is anagen, during which a pigmented hair shaft is generated. Anagen is followed by catagen, a rapid, apoptosis-driven phase of regression that lasts several weeks, during which the hair shaft is transformed into a “club hair.” The HF then enters telogen,
a phase of relative quiescence that varies in duration, and then returns to anagen. There is a disordered, shortened hair cycle in patients with AA, in which the hair cycle is disrupted by early onset into catagen\textsuperscript{20}.

Histopathology of AA scalp reveals a characteristic inflammatory-cell infiltrate (resembling a ‘swarm of bees’, Figure 1.2) that attacks the base of pigment-producing HFs, predominantly those in anagen. This mixed inflammatory-cell infiltrate contains both CD4+ and CD8+ T cells, natural killer (NK) cells, and others, among which CD8+ T cells are typically the first inflammatory cells seen to be entering the anagen hair-bulb epithelium\textsuperscript{20}. In AA patients, it is believed that recognition of follicular autoantigens leads to the “swarm of bees” phenotype of CD4+ and CD8+ T cells, and the subsequent IP collapse, autoimmune response and the clinical phenotype of AA.

![Figure 1.2. 'Swarm of bees' immune infiltrate and collapse of immune privilege seen in AA versus normal anagen hair follicles\textsuperscript{20}](image)

AA is an ideal autoimmune disease to study, since the target organ is highly accessible and is not permanently destroyed, unlike other autoimmune diseases (i.e. Type I Diabetes (T1D) or Rheumatoid Arthritis (RA)) where the end organ is irreversibly damaged. Although the
autoimmune attack causes temporary regression of the HF, there is no permanent damage to the stem cell compartment, thus hair regrowth remains possible. Like many other organ-specific autoimmune diseases, AA is associated with abnormal infiltration of T cells to the HF; however, the environmental and physiological mediators of systemic autoimmune responses or site-specific inflammation in AA have not been identified.

This thesis is focused on defining genetic variants predisposing AA in both the immune system and HF, as well as the microbiome as an environmental trigger in AA.

B. GWAS

Genome-wide association studies (GWAS) are hypothesis free methods to discover associations between genetic regions or loci and traits (including diseases) that can identify causal genes. GWAS is regarded as the gold standard technique for detecting genetic variants enriched in cases versus controls\textsuperscript{24}. Identifying susceptibility variants and loci associated with the trait compared to the general population, allows detection of genetic risk factors. Large sample sizes (usually thousands) are needed to achieve statistical power to detect differences between controls and patients\textsuperscript{25}.

To date, large-scale GWAS studies use microarray-based chips with millions of SNP probes across the genome to identify associations with disease variants. In addition, targeted analysis such as ImmunoChip are used for autoimmune diseases and are applicable for GWAS studies\textsuperscript{26}. GWAS studies depend on accurate diagnosis of the disease being investigated, since inconsistencies in diagnosis or phenotype can dramatically decrease power. This emphasizes the need for large, well-phenotyped cohorts to detect associations between a SNP and the disease. Typically, a meta-analysis study is performed to replicate the associations in an independent cohort. These studies can expand sample size, increase power, and ideally lead to new genomic risk loci associated with disease.
After GWAS studies identify regions of genetic variation, deep sequencing studies (discussed below) are needed to further interrogate regions of interest and identify causal variants. Public databases like dbSNP, 1000genomes, gnomAD, etc., are used to draw upon general population statistics to query genetic variation in control individuals.

C. GWAS Studies in AA

The Christiano lab conducted the first AA GWAS using 1054 cases and 3278 controls, and identified eight genomic regions of association with disease\textsuperscript{11}. A follow-up meta-analysis replication study was performed that combined two previous GWAS studies supplemented with ImmunoChip data for a total of 3,253 cases and 7,543 controls\textsuperscript{27}. In the meta-analysis, several loci were identified (Figure 1.3).

![Table of gene associations](image)

**Figure 1.3. Statistically significant genes identified in AA GWAS and meta-analysis studies.**\textsuperscript{11,27}

Notably, the GWAS genes fell broadly into genes involved in the immune response and HF. Several significant immune-related genes are associated with activation and proliferation of T\textsubscript{regs} (cytotoxic T lymphocyte-associated antigen 4 (CTLA4), interleukin (IL)-2/IL-21, IL-2 receptor α (IL-2RA; CD25) and Eos (also known as Ikaros family zinc finger 4; IKZF4), as well as the human leukocyte antigen (HLA) region). As well as genes expressed in the HF itself, such as PRDX5 and STX17. PRDX5 (Peroxiredoxin 5), was biologically supported by immunostaining,
showing expression in the HF (Figure 1.4)\textsuperscript{11}. A strong association was noted in ULBP (cytomegalovirus UL16-binding protein) gene cluster, which encodes the activating ligands of the natural killer cell receptor, NKG2D.

![Image](image.png)

**Figure 1.4. Peroxiredoxin 5 (PRDX5) is expressed in human HF.** PRDX5 (green) is expressed in the hair shaft and IRS of human HF, green\textsuperscript{11}.

AA shares a significant number of risk loci with other autoimmune conditions\textsuperscript{11}. Susceptibility loci with genes associated with conditions such as rheumatoid arthritis (RA), type 1 diabetes (T1D), celiac disease (CD), systemic lupus erythematosus (SLE), multiple sclerosis (MS) and psoriasis (Ps) were also found in AA\textsuperscript{11}, including CTLA4, IL2/IL2RA, IL21, NKG2D ligands, and genes critical for T\textsubscript{reg} function.

Whereas GWAS studies can delineate the genetic architecture of complex disorders, applying functional genomics and sequencing of variants (discussed in detail below), enables integration of genome- and exome-wide datasets with gene expression and various functional information at the protein level. In the work described in Chapters 2 and 3, these approaches were used to illustrate the roles of an immune system gene (IL2RA) and an end organ specific gene (PRDX5) in AA disease pathogenesis.
D. Targeted Genomic Sequencing

Targeted genomic sequencing is used to identify all of the variants (both common and rare) present in a selected genomic region, such as those identified by GWAS studies. This can help prioritize loci that are biologically relevant and may contribute to disease pathogenesis.

Our initial targeted deep resequencing was performed using the RainDance platform. This method allows targeting of very specific regions of the genome for deep sequencing and achieves high quality sequencing and variant detection\(^\text{28}\) (See Materials and Methods 2.7.1). We prioritized five of our previously reported GWAS regions (CTLA4, IL2/IL21, ULBP3, IL2RA, and IKZF4) and sequenced 122 AA cases (Table 2.1). A RainDance primer library was designed to perform targeted resequencing across the 297kb costimulatory CTLA4 locus (Hg19:chr2:204544318 - 204841364), the 72kb ULBP locus (Hg19:chr6:150330051-150402187), the 100 kb IL2RA locus (Hg19: chr10: 6030076-6130226), the 277 kb IL2/IL21 locus (Hg19:chr4:123373133-123649223), and the 300 kb IKZF4 locus (hg19:chr12: 56313733-56613733) including all exons, introns, and intergenic regions. Sequencing was performed in 122 AA patients selected from our original GWAS cohort\(^\text{11}\). We then conducted a case-only analysis for risk haplotype (See Materials and Methods 2.7.1 and Table 1.1). Using this approach, we found novel variants in these five GWAS regions that showed rare enrichment, defined as 1) present in less than 1% of the population (in databases such as 1000genomes or gnomAD) and 2) present in 3 or more AA patients in our 122 patient cohort, discussed further in chapter 2.

In this thesis, I used this approach to study the functionally relevant Interleukin-2 Receptor Alpha (IL2RA) locus. The IL-2 receptor is crucial in regulating immune tolerance and controlling activity of regulatory T cells (T\(_\text{reg}\)), which suppress activation and proliferation of autoreactive T cells\(^\text{13}\). These findings will be discussed in Chapter 2.
E. Whole Exome and Targeted Genomic Sequencing

Whole genome and whole exome sequencing approaches are newer methods used to detect genetic variants. Whole exome sequencing is a genomic technique used for sequencing all exons in genes across the genome, referred to as the ‘exome.’ This method identifies genetic variants that alter protein sequences, splice sites, and 50-100bp intronic regions at exon borders. Variants across the genome can be identified using these approaches, rather than in a few selected regions, as in the RainDance platform (Table 1.1).

Our study utilized whole exome sequencing plus a custom capture region that included 25 Mb of genomic regions, that we customized to cover our previously implicated GWAS regions. This approach enabled comprehensive sequencing of causal coding and non-coding variants. Our custom capture strategy covered all previous AA GWAS/ meta-analysis regions, as well as key genes from functional and gene expression studies; candidate genes that may confer risk based on biological function; and regions associated with other relevant autoimmune diseases. The sequencing strategy and targeted regions can be found in Materials and Methods 2.7.2 and 3.6.1 (see Table 1.1).

Table 1.1 Comparison of sequencing approaches. Studies discussed in Targeted Genomic Sequencing and Whole exome (section 1.1D) and Targeted Genomic Sequencing (section 1.1E).

<table>
<thead>
<tr>
<th>Sequencing Method</th>
<th>Patients Sequenced</th>
<th>Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targeted Genomic Sequencing</td>
<td>122</td>
<td>5 GWAS regions</td>
</tr>
<tr>
<td>Whole exome + Custom Capture Genomic Regions</td>
<td>849</td>
<td>Whole Exome plus 25 Mb genomic sequencing</td>
</tr>
</tbody>
</table>

F. Analytical Methods

Other studies utilizing whole exome sequencing and exome wide association studies in complex diseases like amyotrophic lateral sclerosis, congenital kidney malformations, epilepsy, pulmonary fibrosis, coronary disease, autism, and myocardial infarction successfully detected significant rare disease variants using gene-level collapsing approaches. These enable
identification of enrichment of novel gene signals that may be associated with disease. (discussed further in chapters 2 and 3).

Gene-level collapsing analyses involve a gene burden approach to discover the relative contribution of rare variants in 849 AA cases and 15,640 controls. This approach consists of gene-based analysis involving only rare loss of function variants with a minor allele frequency ≤1% in cases and controls and a high minor allele frequency (≤1%) for each population in gnomAD database\textsuperscript{37}. Gene-level collapsing models can be expanded to include variants that are canonical splice variants or non-synonymous coding, and predicted as probably damaging by PolyPhen-2 HumVar\textsuperscript{38}. In chapters 2 and 3 of this thesis, I utilized whole exome and targeted genomic sequencing datasets to identify causal variants associated with AA disease pathogenesis.

G. Functional Genomics Studies

After identifying sequence variants, functional genomics studies are used to investigate the role of variants by integrating expression data and utilizing fine mapping techniques. These approaches are aimed at determining the biological involvement of disease variants and their interactions. Projects like ENCODE (Encyclopedia of DNA Elements) Project were developed to detect all of the functional regulatory elements in the human genome\textsuperscript{39}. Both ENCODE and the Roadmap Epigenomics Projects allowed for identification of various DNA regulatory elements in the context of both genetic disease and natural human variation, and are an essential resource to support our fine mapping studies\textsuperscript{40}. After GWAS studies identify loci and subsequent sequencing efforts identify variants (most of which are in regulatory regions), examination of the relationship between genetic variation and gene expression is an essential next step. The GTEx resource is a well-established database to study the relationship between genetic variation and gene expression in human tissues\textsuperscript{41}.
H. Genotype-dependent Changes in Gene expression

We also drew upon our database of microarray-based gene expression studies to generate new fundamental insights into the biology of AA. Our lab utilized skin biopsy samples from 96 patients with a range of AA phenotypes and normal control patients\textsuperscript{42}. Skin biopsy samples were interrogated using microarray-based gene expression analysis to identify AA-specific gene expression signatures\textsuperscript{42}. Changes in gene expression were correlated to different risk alleles in AA, as seen in Chapters 2 and 3. Details can be found in Methods and Materials 2.7.4 and 3.6.3.

I. FUN-LDA Algorithm

Methods such as the FUN-LDA score (See Materials and Methods 2.7.3 and 3.6.2) enable mapping of AA sequence variants to disease-relevant cell types\textsuperscript{43}. The FUN-LDA algorithm uses large genomics data sets (from the ENCODE and Roadmap Epigenomics projects) to integrate various epigenetic annotations for particular cell and tissue types to generate a ‘score’\textsuperscript{43}. FUN-LDA scores can then be correlated with our GWAS studies and various other datasets, utilizing Bayesian fine mapping, to identify the causal effect at each base pair along the gene. Fine mapping using Bayesian modeling and inference is a novel fine mapping method for genetic studies\textsuperscript{44}. This approach allows us to generate a posterior inclusion probability index (PIP), which represents the likelihood of causality of a variant in specific tissue types. This information enables prioritization of functional studies like immunostaining in the most biologically relevant tissue types (See Chapter 3.4).

In chapters 2 and 3 of this thesis, sequencing approaches were followed by functional studies that allowed for the identification of causal variants in AA, and assignment of these effects to specific tissue types.

J. C3H/HeJ Mouse Model of AA

Utilizing animal models for AA lends an additional line of evidence in support of our genetic findings in humans. The C3H/HeJ inbred mouse strain spontaneously develops AA at a \textasciitilde 20\%
rate, but the disease can be induced using a well-established skin graft model\textsuperscript{45,46}. This model enables us to study the disease in the context of both disease prevention and reversal\textsuperscript{45,46}. The C3H/HeJ mouse model replicates many of the pathological features of human AA, such as the T lymphocyte infiltration at the base of the HF. Since the HF is not destroyed in AA, they remain accessible and useable for molecular and histologic studies in AA. Using both mouse model and human AA patient samples, we can isolate affected tissues to better understand disease pathology throughout AA pathogenesis. The AA mouse model has been widely utilized in the field to test intervention methods, drug efficacy, and to define disease mechanisms. For example, one study from our lab identified that CD8+NKG2D+ cytotoxic T lymphocytes are both necessary and sufficient for AA induction in this mouse model\textsuperscript{47}. Utilization of this mouse model, along with some mouse and human gene expression studies, ultimately led to the clinical utility of JAK inhibition in human AA and a subsequent clinical trial\textsuperscript{21,47}. In Chapter 4 (Section 4.2), I utilized the C3H/HeJ mouse model to validate our human AA microbiome findings.

**K. Scarring Alopecia**

Primary cicatricial Alopecias (PCA) are a group of irreversible, scarring inflammatory hair loss disorders characterized by immune cell infiltrates attacking the HF. The most common subtypes are lichen planopilaris (LPP) and frontal fibrosing alopecia (FFA). Unlike AA, these forms of hair loss manifest with permanent destruction of the pilosebaceous unit of the HF, followed by fibrosis, and irreversible hair loss\textsuperscript{48,49}. Clinical diagnosis of LPP is characterized by patchy areas of hair loss with perifollicular erythema, whether FFA presents as a band-like loss of hair and follicular ostia across the frontal temporal scalp region\textsuperscript{48,49}. We utilized a database of PCA 16S rRNA sequencing data as a negative control for the gut microbiome dysbiosis observed in our AA cohort (Chapter 4).
1.2 Environmental Factors Contributing to Complex Disease

Gene-environment interactions lead to inflammatory responses in barrier organs like skin, gut and lung exposed to environmental stresses. Depending on genetic predisposition, stress signals may lead to homeostasis and health or dysregulation and negative outcomes like autoimmune disease. Environmental insults such as trauma, viral infection, bacteria, and other pathogens are believed to contribute to disease. In certain diseases, lifestyle factors such as unhealthy diet, physical inactivity, excess use of alcohol, tobacco, and exposure to environmental toxins may also play a significant role. One popular environmental example is the hygiene hypothesis. This hypothesis proposes a relationship between the increase in allergic disease of childhoods spent in “clean” environments as a result of decreased exposure to pathogens. This is suggested as the model for the increased prevalence of asthma and other atopic diseases in industrialized communities. The hygiene hypothesis can also be applied to the increased incidence of autoimmune diseases. It suggests that a decreased microbial load or diversity (potentially from antibiotic use) in early life may impact the programming of the immune system, particularly the gut associated lymphoid tissue. One case-control study identified a correlation between Type 1 Diabetes (T1D) and degree of social mixing in children. They discovered that children that attended day care centers and had older siblings at home, had a lower incidence of asthma and T1D than children who had no siblings and did not attend day care centers.

One of the most cited environmental triggers in AA revolves around psychological stress. However, in clinical studies, no correlations have been observed, suggesting that stress in AA may not be the only environmental trigger of disease onset. Studies have shown stress may play a role in modulating disease involving the hypothalamic-pituitary-adrenal (HPA), however whether this is sufficient to induce AA remains to be determined. In AA, it was also proposed that cytomegalovirus (CMV) could be a potential promoter of disease, however subsequent studies were not able to confirm this finding. A large-scale mouse model study in AA was unable to determine a correlation between AA and vaccinations like hepatitis B vaccine.
strong case for the potential role of stress, viruses, or vaccine, we turned to the role of the microbiome in triggering AA disease onset.

**A. Human Microbiome**

The human microbiome is comprised of commensal bacteria, viruses, bacteria, fungi and other microbes that colonize barrier sites. Commensal bacteria evolved to be beneficial to the host and serve many purposes including providing nutrients and assisting in digestion. These bacteria maintain intestinal homeostasis and inhibit the growth of pathogens. Growing interest in the human microbiome led to the initiation of projects like the Human Microbiome Project, which catalogued over 690 healthy human samples across fifteen body sites. Studies of this magnitude can help elucidate the microbial components of the 'normal' microbiome composition in healthy individuals and how they contribute to normal physiology. Each anatomic site on the body contains a different microbial composition, and site-specific differences can help identify the pathological and biological importance of particular microbes and microbial communities. In recent years, some of the most striking discoveries regarding host-commensal dependency centered on the role of the microbiome in influencing the immune system (Figure 1.5).

![Figure 1.5. Tissue-specific responses of host and commensals at barrier sites. The gastrointestinal tract has the highest abundance of microbes and influences immune function.](image)
The intestinal microbial community in humans and other mammals is highly complex, and its diversity contributes to several essential processes that maintain intestinal homeostasis and health. Microbial diversity in the gut recently emerged as a potential immunomodulatory factor with the capacity to elicit physiologic and/or pathologic responses in the host. Although a complete understanding of the underlying mechanisms is still emerging, mounting evidence suggests two predominant mechanisms: 1) dysbiosis-induced disruption of the gut epithelial barrier function; and 2) perturbation of the systemic immune response. Recently, the gut microbiome was identified as a potential trigger for the development of several autoimmune disease.

**B. Autoimmunity and the Microbiome**

The microbiome has been implicated in many autoimmune diseases in disease-relevant sites, such as the gut microbiome in celiac disease and the skin microbiome in psoriasis. There are also reports of the microbiome playing a role at distant sites like gut microbiome in uveitis (eye) and the lung microbiome in RA development (joints). Research is ongoing in these diseases to elucidate the exact mechanisms by which the microbiome is acting over great distances.

In one example, a mouse model of experimental autoimmune uveoretinitis (EAU) was either subject to germ-free conditions or broad-spectrum antibiotic treatment. Induction of EAU and T cell infiltration in the retina was reduced, as long as the reduction occurred before T cell activation was induced. In another, mice with a humanized gut microbiome were more susceptible to the development of autoimmune neuropathy (Guillan-Barre Syndrome, GBS) upon infection with *C. jejuni* from a GBS patient, supporting the hypothesis that a balanced microbiome is required to prevent/trigger autoimmunity.

**C. The Leaky Gut**

The presence of commensal and potential pathogenic microorganisms in the gastrointestinal system requires a specialized barrier function to block the entry of microbes, antigens and toxins,
while allowing the absorption of nutrients. This barrier is maintained by a single layer of specialized epithelial cells that are linked by tight junctions (TJ), in combination with other factors including cytokines and immunoglobulins. The intestinal epithelium is the key barrier layer of cells that absorb beneficial substances and provide protection against harmful agents. Loss of integrity of the intestinal epithelium plays a key pathogenic role in autoimmune diseases and leads to the “leaky gut” phenotype, since disrupted barrier function increases, intestinal permeability, and harmful materials can then enter the epithelial membrane.

The intestinal barrier is mostly formed by intestinal epithelial cells (IECs). The intestinal epithelium lies on the crypt-villus axis, with the base of the crypt region forming a niche for stem cells capable of differentiating into different types of IECs69. However, this system may be disrupted by different factors including microbiota dysbiosis, or harmful reactions to microbes in the intestinal defense mechanisms70, resulting in a leaky gut capable of promoting local and systemic immune responses. A leaky gut has been associated with multiple autoimmune diseases as inflammatory bowel disease, celiac disease, autoimmune hepatitis, type 1 diabetes, multiple sclerosis and systemic lupus erythematosus62,71.

Studies in type 1 diabetes, which is genetically similar to AA, demonstrated the capacity for microbial dysbiosis to promote pathogenesis72. These studies suggest a potential model of bacterial involvement in inducing disease by skewing bacterial composition toward microbes that promote gut permeability, allowing for microbial antigens to enter the gut and induce T1D in individuals that are genetically susceptible (Figure 1.6)72. In this thesis, the microbiome and autoimmunity in AA are discussed in Chapter 4.
Figure 1.6. Potential mechanism in Type 1 Diabetes induction, allowing microbial antigens to enter the gut and induce disease.72

D. Short Chain Fatty Acids

Short chain fatty acids (SCFA) are bacterial metabolites, specifically a subset of fatty acids that contain six or less carbon molecules, produced during the fermentation of partially and nondigestible polysaccharides by gut bacteria.73 They include propionate, butyrate, acetate, valeric (pentanoic) acid, and hexanoic (caproic) acid. SCFAs are crucial in the maintenance of immune and gut homeostasis and can play a role in disease development.73 SCFAs can influence processes such as cell differentiation, proliferation and gene expression.74 SCFAs in the GI tract are produced by the bacterial fermentation of dietary fibers and undigested complex carbohydrates and can be influenced by diet.75,76 Members of the Bacteroidetes and Firmicutes phyla are key SCFA producers, with Bacteroidetes mainly producing acetate and propionate and Firmicutes mostly producing butyrate in the gut.74 One study in IBD showed that alterations in butyrate-producing species could delineate between Ulcerative Colitis (UC) and Crohn’s Disease (CD).74 Bacterial species identified through 16S sequencing and metagenomics sequencing correlate with which SCFAs are produced and which may be deficient.
Recent studies showed that the normal microbiota may be maintained in part by reactive oxygen species (ROS)-dependent mechanisms, specifically, that enteric microbiota influence the regulatory networks of the intestinal epithelia\textsuperscript{77}. Some of these studies described the mechanisms by which oxidative stress can modulate microbiota quality and diversity in relation to chronic inflammatory diseases. The mechanisms in these studies resolve damaging signaling pathways by activating the antioxidant defense system\textsuperscript{78}. SCFAs perform anti-inflammatory functions by modulating immune cell chemotaxis, ROS release, as well as cytokine release. SCFA have been shown to induce ROS\textsuperscript{73}, suggesting the increase in SCFA observed in AA is correlated with the upregulation of oxidative stress genes like PRDX5 from our AA GWAS. ROS and oxidative stress-related gene PRDX5 are described in Chapter 3, and SCFA studies are utilized in Chapter 4.

### 1.3 Work Described in this Thesis

First, to study genetic variation in immune genes in AA, we identified key causal variants in AA risk genes using various sequencing techniques and analysis methods and placed them in functionally relevant contexts using innovative mapping techniques. From our AA GWAS studies, the Interleukin-2 Receptor Alpha (IL2RA) immune region was identified as a significant locus that plays a crucial role in regulating immune tolerance and controlling activity of regulatory T cells (T\textsubscript{reg})\textsuperscript{13}. We identified causal variants in the IL2RA region for AA using GWAS, targeted resequencing, and custom capture exome sequencing approaches. IL2RA variants identified in our sequencing and \textit{in silico} studies support that certain polymorphisms in IL2RA cause lower CD25 expression on CD4 naïve T cells, and that reduction of CD25 expression by certain protective variants could reduce the likelihood that naïve CD4 T cells are activated in a proinflammatory state.

Secondly, to study susceptibility genes affecting the HF, we analyzed the Peroxiredoxin 5 (PRDX5) gene. PRDX5 is an antioxidant enzyme crucial for regulating oxidative stress. Using
Bayesian fine mapping we identified a variant, rs574087, that is GWAS and exome significant and predicted to be a causal variant in keratinocytes and melanocytes by the posterior inclusion probability index (PIP), which predicts and scores likelihood of causality in specific tissue types. We functionally validated our in silico findings by immunostaining HFs and cultured melanocytes. Oxidative stress and excess reactive oxygen species (ROS) production occurs in many autoimmune diseases and suggests that the increased PRDX5 expression observed in AA is perhaps involved in counteracting the excess ROS. PRDX5 is crucial for protection from oxidative stress and its dysregulation may predispose to autoimmunity.

Lastly, we investigated the role of the microbiome in human and mouse AA. We found striking gut dysbiosis, consisting of over-representation of Firmicutes and under-representation of Bacteroides in the gut microbiome of AA patients (but not in skin or HF) as compared with healthy controls. We also identified several differentially abundant bacterial taxa in AA gut microbiome that could potentially modulate disease activity. To functionally test the requirement of the gut microbiome in vivo, we depleted the gut microbiome in C3H/HeJ mice and showed that they were largely protected from developing AA, confirming a requirement for the gut microbiome in the onset of murine AA. Strategies such as FMT, antibiotic therapies, and other methods for microbial perturbation could provide new methods for restoring healthy gut composition and improving disease outcomes in AA.
Chapter 2. Immune System: Fine mapping of the IL2RA locus in AA

2.1 Introduction

AA is a highly prevalent organ-specific autoimmune disorder with a lifetime prevalence of 2.1%\textsuperscript{19}. In AA, there is an aberrant interaction between the immune system and the HF which leads to non-scarring hair loss. We conducted a Genome Wide Association Study (GWAS) on patients from the AA Registry to identify variants that contribute to AA susceptibility. Our first study uncovered eight loci or regions of association containing candidate susceptibility genes\textsuperscript{11}. A replication study and meta-analysis from an independent northern European cohort identified several new regions of significance and additional loci, bringing the total number of loci of interest to 14\textsuperscript{27}.

From our AA GWAS studies, here we focused on the functionally relevant Interleukin-2 Receptor Alpha (IL2RA) locus. The IL-2 receptor is crucial in regulating immune tolerance and controlling activity of regulatory T cells (T\textsubscript{reg}), which suppress activation and proliferation of autoreactive T-cells\textsuperscript{13}. A key component of this pathway is IL-2 binding to its heterotrimeric receptor IL-2R which is made up of alpha, beta, and gamma subunits\textsuperscript{13}. IL2RA (the alpha subunit, also known as CD25) is crucially important in protecting an individual from autoimmunity\textsuperscript{79}. When IL-2 binds to its receptor, it activates both regulatory and effector T cells. Effector T cells (CD8 cells), enhance autoimmune processes and regulatory T cells (CD4 cells) attenuate and terminate inflammatory responses\textsuperscript{80}.

The IL2RA locus has been implicated by GWAS in many autoimmune diseases. Several polymorphisms surrounding the IL2RA locus have been reported to be associated with type I diabetes (T1D), rheumatoid arthritis, multiple sclerosis (MS), Crohn’s Disease (CD), Graves’ disease, generalized vitiligo, and AA\textsuperscript{81}. Several IL2RA variants are known to regulate the levels of expression of the IL-2 Receptor on CD4+T cells\textsuperscript{81}. Depending on the IL2RA haplotype in a given disease, the level of CD25 expression can confer either a higher risk or protection from
autoimmunity\textsuperscript{81}. One example is the rs11594656 variant, which is associated with protection from T1D, but confers risk for MS\textsuperscript{81}.

Our targeted resequencing was performed in two stages. First, we sequenced the IL2RA locus on the RainDance platform in 122 AA cases (Table 2.1).

<table>
<thead>
<tr>
<th>Name</th>
<th>Chromosome</th>
<th>Position (hg19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTLA4</td>
<td>2</td>
<td>204544427-204840948</td>
</tr>
<tr>
<td>IL2/IL21</td>
<td>4</td>
<td>123351431-123662586</td>
</tr>
<tr>
<td>ULBP3</td>
<td>6</td>
<td>150402185-150402187</td>
</tr>
<tr>
<td>IL2RA</td>
<td>10</td>
<td>6030076-6130226</td>
</tr>
<tr>
<td>IKZF4</td>
<td>12</td>
<td>56313733-56613733</td>
</tr>
</tbody>
</table>

We conducted a case-only analysis of risk haplotype for enrichment and identified novel rare variants that showed enrichment. In the second stage, we conducted whole exome sequencing plus custom capture targeted sequencing of our previously identified GWAS regions (including IL2RA), as well as linkage peaks, confirmed CNVs, and pathways identified using gene expression analysis (see Table 1.1). These studies were completed in 849 patients, a cohort size with robust power to reveal causal variants. We compared rare variants identified by the targeted RainDance resequencing in 122 patients to the whole exome/custom capture rare variants in 849 patients, are observed that many were replicated in both analyses.

### 2.2 Targeted Enrichment Pipeline Analysis

Typically, GWAS regions are complex and challenging to analyze. Here, we developed a pipeline to streamline this process. To begin, we chose the loci of interest and analyzed the
targeted resequencing variants in that region. We then narrowed down to the rare enriched variants as seen in Figure 2.1.

![Figure 1](image1.png)

**Figure 2.1. Analytical pipeline for exploring genomic regions with targeted resequencing data.** From our 14 GWAS regions, we focused on one region of interest (IL2RA). With our targeted resequencing studies in 122 AA patients, we identified 458 variants. Of these, 65 variants were rare enriched (defined as present in less than 1% of the population in population databases; and present in 3 or more patients in our cohort).

We defined a variant as ‘rare enriched’ on the basis of two criteria: 1) if it is present in less than 1% of the population (in databases such as 1000genomes or gnomAD); and 2) is present in 3 or more AA patients in our 122 patient cohort. We identified a total of 458 variants in the IL2RA region by our targeted resequencing, shown in Figure 2.2. When we filtered the data for ‘rare

![Figure 2](image2.png)

**Figure 2.2. All 458 variants in IL2RA region identified from targeted resequencing.** Each of the 458 enriched as tick marks along the IL2RA gene. Variants fall mostly in intergenic and intronic regions.
enriched’ variants, we identified 65 variants, which fell mainly in introns or intergenically, as seen in Table 2.2 and in Figure 2.3.

**Table 2.2. Functional annotation of rare enriched variants.** Rare enriched variants from Targeted Resequencing studies performed on the RainDance platform for the IL2RA region. Rare enriched variants are present in 3 or more of our patient cohort and less than 1% of the population. These 65 rare enriched variants are located intergenically, intronically, UTR3, and downstream of the IL2RA locus.

<table>
<thead>
<tr>
<th>Number of Variants</th>
<th>Functional Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>Intergenic</td>
</tr>
<tr>
<td>35</td>
<td>Intronic</td>
</tr>
<tr>
<td>2</td>
<td>UTR3</td>
</tr>
<tr>
<td>1</td>
<td>Downstream</td>
</tr>
<tr>
<td><strong>Total Rare Enriched Variants</strong></td>
<td><strong>65</strong></td>
</tr>
</tbody>
</table>

**Figure 2.3. All rare enriched variants in IL2RA region identified from targeted resequencing.** Each of the 65 rare enriched variants as tick marks along the IL2RA gene. As seen also in Table 1, the variants fall mostly intergenically and intronically, with two UTR3 and one downstream.

Once the rare enriched variants were identified, a crucial step in the analysis pipeline is functional interrogation. By exploring the tissue types in which these rare enriched variants were expressed, we integrated GWAS-identified variants with sequence-identified variants in target loci to specific cell types. The FUN-LDA score, developed by Dr. Iuliana Ionita-Laza, is used to predict the functional effects of certain non-coding genetic variants in particular tissue types. The FUN-LDA algorithm uses large genomics data sets like ENCODE and the Roadmap Epigenomics projects to integrate various epigenetic annotations for particular cell and tissue types. This unsupervised approach can predict functional effects in a tissue-specific manner. The integrated
functional scores also predict the most likely tissue/cell-type in which the variant will exert a functional effect in a complex disease.

For each IL2RA variant in our AA GWAS study, we summed the FUN-LDA score for cell types across a designated tissue cluster (Immune Cells, Stem Cells, Mesodermal Cells, Connective Mesenchymal, and Neural Crest) and divided this by the overall total score of the variant across all cell types. This yields the proportion of the variants predicted to have a regulatory effect in each tissue cluster, while considering the strength of scores. This analytical method allows identification of variants that are predicted to have a functional effect in specific cell types. FUN-LDA implicates these cell types in AA and the algorithm serves as a preliminary model for determining the appropriate cell types that should be used for investigating and validating the functional effect of these variants in relation to AA disease pathogenesis. As shown

![Figure 2.4. Tissue Distribution of IL2RA Rare Enriched Variants.](image)

IL2RA variants identified in our targeted resequencing displayed across the IL2RA gene and their highest tissue type expression. Red shading indicates variants that are highly likely to have a functional effect in a specific tissue type as identified by the FUN-LDA score. This figure shows that IL2RA variants identified tend to cluster in the immune cell populations (81% in the immune cell cluster relative to the 19% in the other four clusters).
in Figure 2.4, variants are shown as tick marks across the gene, with the tissue expression pattern of each variant shown vertically above across each cell cluster. The red clusters indicate higher scores, suggesting that the variant is more likely to have an effect in a given cell type. This data shows that most of our variants fall within the promoter region of intron 1 of IL2RA (Figure 2.4, red box) and are expressed primarily in the immune cell cluster, with a few others interspersed.

2.3 Whole Exome and Custom Capture Sequencing

To replicate and validate some of rare enriched variants for functional testing, we next performed deep sequencing in an independent cohort of 849 patients. Deep sequencing analysis was conducted across the IL2RA region to identify causal variants, together with whole exome sequencing using the Agilent SureSelect Human All Exon 50Mb plus 6.8-24Mb custom option for targeted enrichment of the genome. Genes were selected for targeted enrichment using GWAS regions, linkage peaks, confirmed CNVs, and pathways identified using gene expression analysis.

We designed the custom capture regions to cover all AA GWAS regions and linkage disequilibrium (LD) blocks implicated in our two previous GWAS in AA. Additionally, we sequenced key genes identified in our gene expression studies and functional experiments, as well as candidate genes selected using biological knowledge of their function and potential to contribute to the pathogenesis of AA. Lastly, we included a few LD blocks implicated by GWAS in type 1 diabetes (T1D) and rheumatoid arthritis, and harbor genes that are differentially regulated in peripheral blood mononuclear cells of patients compared to controls. In total, our custom capture regions cover approximately 17 MB.

Analyzing this data allowed us to link some of the IL2RA causal variants to those found in our targeted resequencing analysis or GWAS study as seen in Table 2.3, with custom capture exome p-values assessed by chi-squared test of variant enrichment (Table 2.3).
We analyzed whole skin biopsies from AA patients from our previous microarray studies (see Materials and Methods 2.7.4), to search for gene expression changes that correlated with genotype. We found rs3118740 decreased IL2RA gene expression for carriers of the T allele.
Such allele specific effects could be due to a perturbation of T\textsubscript{reg} function, i.e. one study in T1D where patients with the rs3118470 risk variant have T\textsubscript{reg} with IL-2 signaling defects\textsuperscript{12}. This variant also has a PIP score of 0.2899, suggesting it is highly likely to be a causal variant in immune cells, most likely due to the large number of T cells infiltrating the skin. This experiment validates some of our targeted resequencing variants and enables prioritization of variants for future work.

![Genotype dependent expression differences from whole skin biopsies in patients.](image)

**Figure 2.5.** Genotype dependent expression differences from whole skin biopsies in patients. Using microarray data to study patient’s whole skin biopsies, we found that targeted resequencing, GWAS, and exome significant variant rs3118470 decreases IL2RA gene expression for carriers of the T allele. This variant also has a PIP score of 0.2899, suggesting it is highly likely to be a causal variant.

### 2.5 Functional Relevance in Tissue Types

Identifying causal variants, while important, does not allow us to place these findings into a disease-relevant context. Using a fine-mapping algorithm, we placed some of our significant variants into relevant tissue types and predict the likelihood of them being causal in that tissue type. This is done using the posterior inclusion probability (PIP) score which is a result of Bayesian
fine-mapping and provides the weight of evidence that a SNP should be included as potential causal variant. PIP scores above 0.1 and 0.2 are predicted to be causal in the tissue type.

*In silico* analysis suggests that IL2RA is strongly implicated in immune cells (specifically CD4+ T cells), rather than HF cells (Figure 2.6). This shows high FUN-LDA peaks (yellow) overlapping with significant SNPs with high PIP scores (red and blue X’s), suggesting there is high likelihood of causality of these snps in this particular tissue type (CD4+ T cells). Comparing this to another cell type, for example, keratinocytes (Figure 2.6), we observed no peaks. The high PIP SNPs were overlapping, suggesting a specific role for IL2RA variants in immune cell types (see red box).
Figure 2.6. *In silico* studies of IL2RA variants in immune cells.
Map of the IL2RA gene shows the FUN-LDA score, defined as the likelihood of each base pair in the gene to be functional in a specific tissue type (here CD4+ T cells, or E038 Primary T helper naïve cells from peripheral blood as in the GTEx database). The blue dots represent significant SNPs in this region. The green peaks represent the FUN-LDA score of likelihood of being causal in selected tissue types at a specific base pair position. The red and blue X’s are significant SNPs predicted by posterior inclusion probabilities or PIP scores (red PIP>0.2, blue PIP>0.1). A high green peak together with an overlapping red or blue X SNP, indicates a high likelihood that SNP is causal in that specific tissue or cell type.
2.6 IL2RA in Autoimmunity

IL2RA is known to play different roles in relation to autoimmune diseases, and variants can cause either susceptibility or risk to disease. For example, there are IL2RA variants such as rs2104286, where variants confer the same risk (AA, MS, and RA), and other examples, such as rs7072793 and rs4147359, in which one allele confers susceptibility to SLE, but protection from AA and T1D. Figure 2.7 represents a comparison of IL2RA variants genotyped in various autoimmune diseases.

![Figure 2.7. IL2RA variants in the context of AA and other autoimmune diseases.](image)

Allelic differences in IL2RA locus between diseases indicates the need for fine mapping and functional studies. The variant rs2104286 (far left) clusters together in all of the diseases. Most likely this is due to the reduction of CD25 expression by the protective phenotype of the rs2104286 allele which may reduce the likelihood of naïve CD4+ T cells being activated under proinflammatory conditions. Other variants are predicted to confer risk in some diseases but not others. An odds ratio value <1 indicates protection from disease, whereas an odds ratio value >1 indicates increased risk for disease. Data adapted from Maier et al.4.

The common autoimmune variant rs2104286 clusters together in all of the diseases. Since CD25-positive naïve CD4+ T cells expressing the high affinity IL-2 receptor can simultaneously be activated by IL-2 and engagement of the T cell receptor, the reduction of CD25 expression by the protective phenotype of the rs2104286 allele may reduce the likelihood of naïve CD4+ T cells being activated under proinflammatory conditions. Since we have established AA contains many
shared variants with other autoimmune diseases, we contextualized our exome and targeted resequencing genotype data to see whether these autoimmune diseases contain similar risk alleles. Table 2.4 shows a representative set of variants and the relative risk alleles.

Table 2.4. AA significant risk variants implicated in other autoimmune diseases. AA variants implicated in our studies and studies in MS, T1D, and RA, which represent autoimmune diseases that commonly harbor variants in the IL2RA locus. We also note increased statistical significance and power in almost all of these variants when expanding our cohort from the 122 patients in AA Targeted Resequencing study to the 849 patients in the AA Whole Exome study.

<table>
<thead>
<tr>
<th>rsID</th>
<th>Hg19 position</th>
<th>dbSNP Predicted</th>
<th>Published Studies</th>
<th>Studies in Alopecia Areata (AA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ref</td>
<td>Alt</td>
<td>Risk</td>
<td>Risk</td>
</tr>
<tr>
<td>rs12722489</td>
<td>6102012</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>rs2104286</td>
<td>6099045</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>rs12722561</td>
<td>6069893</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>rs61839660</td>
<td>6094697</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
</tbody>
</table>

2.7 Materials and Methods

2.7.1 Targeted Resequencing

Our targeted resequencing was performed on the RainDance platform and was performed on five of our previously reported GWAS regions and LD block regions in 122 AA cases (Table 2.1). A RainDance primer library was designed to perform targeted resequencing across the 297kb costimulatory CTLA4 locus (Hg19:chr2: 204544318 - 204841364), the 72kb ULBP locus (Hg19:chr6:150330051-150402187), the 100 kb IL2RA locus (Hg19: chr10: 6030076-6130226), the 277 kb IL2/IL21 locus (Hg19:chr4:123373133-123649223), and the 300 kb IKZF4 locus (h19:chr12: 56313733-56613733) including all exons and introns, and intergenic regions.
Sequencing was performed in 122 AA patients from our GWAS cohort\textsuperscript{11}. Briefly, genomic DNA was sheared into 3-5kb fragments using Covaris S2. RainDance 1000 was used to merge primer libraries with fragmented genomic DNA prior to PCR amplification. PCR products were pooled together for each sample and indexed during library preparation. Samples were then sequenced on the Roche 454 FLX.

Mapping and calling of variants followed the Genome Analysis Toolkit (GATK) Best Practices workflow\textsuperscript{82}. Raw paired-end reads (read size: 101bp) in FASTQ format were mapped to the human reference genome (NCBI GRCh37) using a Burrow-Wheeler Alignment Tool (BWA-version 0.5.9)\textsuperscript{83}. Our region of interest was specified in a BED file. The resultant BAM files were refined with GATK (version 1.6-13) by performing indel realignment and base quality score recalibration. Variants were called on the refined BAM files using GATK Unified Genotyper. In addition to default variant filters in GATK, filters were applied to ensure a low false-discovery rate. The filters included: (1) a genotype quality score less than 30, (2) a quality-to-depth ratio less than 5.0, (3) a strand bias score greater than -0.1, and (4) more than 10% of reads at a locus with zero-mapping quality \textsuperscript{84}. Stringent quality control (QC) measures were used to remove samples and markers that did not exceed predefined thresholds. These included excessive failure rates (missing greater than 5% of data) and Hardy-Weinberg equilibrium (p<1\times 10^{-2}).

Variants that passed QC measures were phased, and missing data was imputed using Beagle\textsuperscript{85}. Phased chromosomes were computed and stratified on the basis of containing the AA GWAS risk haplotype or the AA GWAS protective haplotype.

We then conducted a case-only analysis for risk haplotype versus all other for enrichment. To identify variants enriched on chromosomes carrying the AA risk haplotype, a two-sided Fisher’s exact test was performed. A Bonferroni correction was used to determine a cut-off for significance (p=0.05/variant count). Variants with alleles significantly enriched on the risk haplotype were further prioritized for investigation based on a regulatory function annotation in the Ensembl Variation database in Biomart\textsuperscript{86}.
To facilitate analysis at sites observed in the sequencing study, 934 cases and 3278 controls were imputed to the panel of sequenced SNPs. Briefly, 122 sequenced AA cases were merged with the 1000 Genomes Phase 1 version 3 global reference haplotypes. The 1000 Genomes samples were added to both increase the reference information added in imputation and to reduce the direct correlation between the control dataset and a subset of its paired cases. These combined samples were then used to impute the remaining samples using MACH version 1.0.18. Association was performed and a Bonferroni correction was used to determine a cut-off for significance \((p=.05/636=7.9\times10^{-5})\). From this dataset, we found novel variants that show rare enrichment, defined as 1) present in less than 1% of the population (in databases such as 1000genomes or gnomAD) and 2) present in 3 or more AA patients in our 122 patient cohort.

### 2.7.2 Whole Exome/Genome Sequencing

Deep sequence analysis was conducted in several candidate regions to identify causal variants, together with whole exome sequencing using the Agilent SureSelect Human All Exon 50Mb plus 6.8-24Mb custom option for targeted enrichment of the genome for indexed sequencing on Illumina HiSeq sequencers. Genes were selected for targeted enrichment using GWAS regions, linkage peaks, confirmed CNVs, and pathways identified using gene expression analysis. We designed our custom capture regions to cover all AA GWAS regions. We also included in our custom capture regions linkage disequilibrium (LD) blocks implicated in our two previous GWAS in AA. Additionally, we sequenced key genes identified in our gene expression studies and functional experiments, as well as candidate genes guided by our biological knowledge of their function and potential to contribute to the pathogenesis of AA. Lastly, we included a few LD blocks implicated by GWAS in type 1 diabetes (T1D) that were associated with other autoimmune diseases and have expression or functional relevance to AA. In total, our custom capture regions covered approximately 17 MB.

### 2.7.3 Mapping AA Sequence Variants to Disease-Relevant Cell Types
The FUN-LDA score predicts the functional effects of non-coding genetic variants in particular tissue types. The FUN-LDA algorithm uses large genomics data sets like the ENCODE and Roadmap Epigenomics projects to integrate various epigenetic annotations for particular cell and tissue types. This unsupervised approach predicts the functional effects in a tissue-specific manner for every position in the human genome and for different functional variant classes and locations (i.e. promoters, enhancers, etc.). These integrated functional scores may elucidate the most likely tissue/cell-type for complex diseases using GWAS summary statistics.

For each variant in our AA GWAS study, we summed the FUN-LDA score for each cell type across a designated tissue cluster (A-Immune Cells, B-Stem Cells, C-Mesodermal Cells, D-Connective Mesenchymal, and E-Neural Crest) and divided this by the overall total score of the variant across all cell types. This generated the proportion of variants having a regulatory effect in each tissue cluster, while considering the strength of scores. We then used a threshold of the proportion > 0.5 to assign specific tissue clusters to variants. This analytical method allows identification of variants that are predicted to have a functional effect in specific cell types. FUN-LDA implicates these cell types in AA disease and the algorithm serves as a crucial preliminary model for determining the appropriate cell types that should be used for investigating and validating the functional effect of these variants in relation to AA disease pathogenesis.

2.7.4 Gene Expression Analysis of AA Patient Tissue

Quality control of microarrays was performed using the affyAnalysisQC package from http://arrayanalysis.org/. Microarray preprocessing was performed using BioConductor in R. Differential expression in these studies was defined by an absolute fold change threshold of 1.5 with a Benjamini–Hochberg-corrected significance threshold of 0.05. The dataset was normalized using GCRMA and MAS5. The Affymetrix HGU-133Plus2 array contains 54675 probe sets (PSIDs). Filtering was performed so that PSIDs that were on the X or Y chromosome, that were Affymetrix control probe sets, or that did not have Gene Symbol annotation were removed from all arrays for further downstream analysis. Correction for batch effects was performed using the
implementation of the function ComBat available in the sva package with gender and AA group (AT/AU, AAP, and normal) used as covariates.

Differential analysis was performed on the batch corrected discovery data set using linear models as implemented in the limma package in Bioconductor\(^8\). Two-sample comparisons were performed separately to identify PSIDs differentially expressed in AA patients versus normal controls, in AAP patients versus normal controls, and in AT/AU patients versus normal controls, treating gender as a fixed factor. This dataset was previously used in our lab in other studies\(^4^2\).

2.8 Conclusions

The IL-2 receptor is critical in regulating immune tolerance and controlling activity of regulatory T cells (T\(_{\text{reg}}\)). IL2RA (the alpha subunit, also known as CD25) is crucially important in protecting an individual from autoimmunity\(^7^9\). We identified significant causal variants in the IL2RA region for AA using GWAS, targeted resequencing, and custom capture exome sequencing approaches. Here, we identified causal variants identified in both our targeted resequencing studies and our exome studies, with high PIP scores. Using the FUN-LDA algorithm, we validated the expression of these variants in immune cell cluster tissue types \textit{in silico} as compared to other tissue types. Utilizing Bayesian fine mapping, we discovered high likelihood of causality in subtypes of immune cells, such as CD4+ T cells, representing T\(_{\text{regs}}\).

We found genotype dependent alterations in gene expression and showed that the variant rs3118740 decreases IL2RA gene expression for carriers of the T allele in our patient cohort. This could be due to a perturbation of T\(_{\text{reg}}\) function, for example, one study in T1D where patients with the rs3118470 risk variant have T\(_{\text{reg}}\) with IL-2 signaling defects\(^1^4\). Placing these findings in the context of autoimmunity, we identified some variants in IL2RA (such as rs2104286) that confer the same risk for autoimmune disease. Likewise, we identified variants (such as rs7072793 and rs4147359), in which one allele confers susceptibility to one disease, but protection from another. We observed the same protective effect of IL2RA variant rs2104286 in AA as reported across all
autoimmune diseases, which may reduce the likelihood of naïve CD4+ T cells being activated T\(_{\text{regs}}\) under proinflammatory conditions. These studies demonstrated causal variants in AA, may correlate with reduced T\(_{\text{reg}}\) function and risk of autoimmunity, and leave open further therapeutic options beyond low dose IL-2 therapy (see Chapter 5).
Chapter 3. Hair Follicle: Fine Mapping and Functional Studies of PRDX5

3.1 Introduction

AA is a highly prevalent organ-specific autoimmune disorder with a lifetime occurrence of 2.1%\textsuperscript{19}. In AA, there is an aberrant interaction between the immune system and the hair follicle, which leads to non-scarring hair loss. Our lab conducted an AA GWAS to identify variants that contribute to AA susceptibility. The first study uncovered eight loci or regions of association containing candidate susceptibility genes\textsuperscript{11}. Additional replication study and meta-analysis from an independent northern European cohort identified several new loci of significance, bringing the total loci of interest to fourteen\textsuperscript{27}. One of the HF specific genes, Peroxiredoxin 5 (PRDX5), is also a known susceptibility gene in other autoimmune diseases including Crohn’s Disease, Psoriasis, and Sarcoidosis, among others\textsuperscript{15,16}. PRDX5 is a member of the peroxiredoxin family of antioxidant enzymes crucial for regulating oxidative stress, which causes disruption in redox potentials that extend to the endoplasmic reticulum (ER).

Since AA has multifactorial pathology, several non-genetic factors must be considered to elucidate the pathophysiology of disease. One topic of interest is oxidative stress, in light of the significance of the PRDX5 gene in AA GWAS and meta-analyses replications. The dysregulation of antioxidant enzymes, such as PRDX5, has been implicated in rheumatoid arthritis (RA) and has been linked to inflammatory processes. ROS represent important intracellular signaling molecules, and when present in excess, they can enhance inflammatory responses and require an antioxidant defense system to prevent ROS toxicity.

It is well-established that many dermatological disorders including vitiligo, lichen planus, atopic dermatitis, and psoriasis, are characterized by increased levels of ROS\textsuperscript{89}. Notably, in AA, there are reports of increased ROS production from perifollicular inflammatory cells\textsuperscript{89}. These ROS are formed by incomplete reduction of atmospheric oxygen generating potent oxidizing agents and their toxic byproducts\textsuperscript{90}. Under normal physiological conditions, antioxidant enzymes play an
important role in protecting tissues from the damaging effects of ROS. Antioxidants are substances that scavenge oxygen free radicals or inhibit the cellular oxidation process\textsuperscript{91}. Under homeostasis, cells and the extracellular environment contain antioxidant defenses to oppose the toxic effects of lipid peroxides and maintain normal and stable physiology. Many of the damaging reactions in the cell are due to lipid peroxides and oxygen radicals, which can damage cell membranes, and eventually cause cell death\textsuperscript{92}.

Hydrogen peroxide (H\textsubscript{2}O\textsubscript{2}) is a covalent liquid that readily crosses cell membranes. Elevated H\textsubscript{2}O\textsubscript{2} levels are cytotoxic, and in order to keep these levels at a minimum, antioxidant enzymes like peroxidases (especially GPx), catalases, and thioredoxin-linked antioxidant enzymes like PRDX5 are deployed\textsuperscript{93}. H\textsubscript{2}O\textsubscript{2} itself is not highly reactive, but the danger comes from its ability to break down into highly reactive hydroxyl radicals (free radical \textsuperscript{\cdot}OH). Normally, antioxidant enzymes are the first line of defense and free radicals undergo a decomposition reaction to form water and oxygen using the general mechanism seen in Figure 3.1\textsuperscript{12,92}.

Recent efforts have been focused on the understudied peroxiredoxin family of antioxidant enzymes. Six isoforms of the PRDX family have been characterized in mammals, all of which are directly implicated in helping to eliminate H\textsubscript{2}O\textsubscript{2} and play a role in neutralizing various other oxidizing molecules\textsuperscript{94}. The crystal structure of PRDX5 suggests that it may have a broader activity against ROS relative to the other PRDX isoforms and other antioxidant enzymes\textsuperscript{95}. PRDX5 overexpression has been shown to have a protective role against oxidative damage in human tendon degeneration diseases\textsuperscript{94}.

The body relies on these antioxidant enzymes to ensure minimal damage from free radicals, however, when these free radicals are overproduced, a state of oxidative stress ensues.
The cause of this imbalance is usually a stressor such as an environmental exposure inciting an immune reaction, for example, it was previously shown that there is a dysfunction of SOD in AA patients\cite{96}. ROS genes such as PRDX5 have been implicated in our GWAS, however, their role in disease pathogenesis has not been fully elucidated.

### 3.2 Fine mapping of GWAS region PRDX5

To fine map the PRDX5 region and further our understanding of variants, deep sequencing was conducted in several candidate regions, including PRDX5, to identify causal variants, together with whole exome sequencing using the Agilent SureSelect Human All Exon 50Mb plus 6.8-24Mb custom option for targeted enrichment of the genome. These studies were performed on 849 patients, which gives us robust power to reveal causal variants.

We designed our custom capture regions to cover all AA GWAS regions and included linkage disequilibrium (LD) blocks implicated in our two previous GWAS in AA. Additionally, we sequenced key genes identified in our gene expression studies and functional experiments, as well as on the basis of our biological knowledge of their function and potential contribution to the pathogenesis of AA. Lastly, we included representative LD blocks implicated by GWAS in type 1 diabetes (T1D) or rheumatoid arthritis, and harbor genes that are differentially regulated in peripheral blood mononuclear cells compared to controls. In total, our custom capture regions cover approximately 17 MB. To fine map the PRDX5 region, we performed whole exome sequencing with a custom capture region of genomic sequencing from 849 AA patients. In our custom capture exome data, we identified representative variants in PRDX5 that were enriched in both exome and GWAS studies as compared to the gnomAD database, with custom capture exome p-values assessed by chi-squared test of variant enrichment (Table 3.1). This exercise
validates our previous GWAS variants and shows that these variants should be prioritized for future functional work.

**Table 3.1. Exome data replicates enrichment of a subset of GWAS Variants.** Shown here are the top 20 significant variants from our exome custom capture sequencing, with their predicted locations. A subset of variants identified in our GWAS as significant (bolded), were also seen in our exome custom capture studies. Case allele frequency (AF) refers to the allele frequency in cases, gnomAD AF is the population frequency, and p-value was assessed by chi-squared test of variant enrichment.

<table>
<thead>
<tr>
<th>RS Number</th>
<th>Case AF</th>
<th>gnomAD AF</th>
<th>Location</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7112596</td>
<td>0.0171</td>
<td>0.0016</td>
<td>Intergenic region</td>
<td>2.2E-24</td>
</tr>
<tr>
<td>rs10501395</td>
<td>0.0083</td>
<td>0.0002</td>
<td>Downstream gene variant</td>
<td>9.08E-21</td>
</tr>
<tr>
<td>rs497544</td>
<td>0.9864</td>
<td>0.9986</td>
<td>Intergenic region</td>
<td>1.73E-18</td>
</tr>
<tr>
<td>rs1783519</td>
<td>0.9841</td>
<td>0.9978</td>
<td>Intergenic region</td>
<td>2.14E-17</td>
</tr>
<tr>
<td>rs11231696</td>
<td>0.0130</td>
<td>0.0014</td>
<td>5 prime UTR premature start codon gain variant</td>
<td>3.22E-17</td>
</tr>
<tr>
<td>rs539810</td>
<td>0.9864</td>
<td>0.9983</td>
<td>Intergenic region</td>
<td>2.11E-16</td>
</tr>
<tr>
<td>rs57779243</td>
<td>0.0124</td>
<td>0.0648</td>
<td>Downstream gene variant</td>
<td>2.46E-16</td>
</tr>
<tr>
<td>rs56241103</td>
<td>0.0124</td>
<td>0.0648</td>
<td>Downstream gene variant</td>
<td>2.46E-16</td>
</tr>
<tr>
<td>rs12271952</td>
<td>0.0193</td>
<td>0.0034</td>
<td>Intron variant</td>
<td>1.67E-15</td>
</tr>
<tr>
<td>rs7118199</td>
<td>0.9811</td>
<td>0.9963</td>
<td>5 prime UTR variant</td>
<td>6.62E-15</td>
</tr>
<tr>
<td>rs527465</td>
<td>0.3632</td>
<td>0.4557</td>
<td>Downstream gene variant</td>
<td>5.41E-12</td>
</tr>
<tr>
<td>rs618160</td>
<td>0.2783</td>
<td>0.3658</td>
<td>Downstream gene variant</td>
<td>1.4E-11</td>
</tr>
<tr>
<td>rs602534</td>
<td>0.2783</td>
<td>0.3654</td>
<td>Downstream gene variant</td>
<td>1.62E-11</td>
</tr>
<tr>
<td>rs12421615</td>
<td>0.2730</td>
<td>0.3584</td>
<td>Intron variant</td>
<td>3.39E-11</td>
</tr>
<tr>
<td>rs541207</td>
<td>0.3703</td>
<td>0.4581</td>
<td>Downstream gene variant</td>
<td>5.95E-11</td>
</tr>
<tr>
<td>rs499425</td>
<td>0.2916</td>
<td>0.3764</td>
<td>Intergenic region</td>
<td>7.72E-11</td>
</tr>
<tr>
<td>rs574087</td>
<td>0.2930</td>
<td>0.3770</td>
<td>Intergenic region</td>
<td>1.22E-10</td>
</tr>
<tr>
<td>rs694739</td>
<td>0.2960</td>
<td>0.3796</td>
<td>Intergenic region</td>
<td>1.55E-10</td>
</tr>
<tr>
<td>rs627425</td>
<td>0.2942</td>
<td>0.3750</td>
<td>Downstream gene variant</td>
<td>4.53E-10</td>
</tr>
<tr>
<td>rs2510066</td>
<td>0.2948</td>
<td>0.3752</td>
<td>5 prime UTR variant</td>
<td>6.59E-10</td>
</tr>
</tbody>
</table>
3.3 Genotype-dependent Expression

Studying patients' whole skin biopsies, we found that the variant rs694739 decreases the expression of PRDX5 for carriers of the A allele (Figure 3.2), as shown previously in other autoimmune diseases such as Crohn's Disease (p=6E-10), psoriasis (p=3.71E-09), and MS (p=2E-09)\textsuperscript{1,2,97}.

![Figure 3.2. Genotype differences from whole skin biopsies in patients.](image)

Using microarray data to study patient's whole skin biopsies, we found GWAS and exome significant variant rs694739 decreases the expression of PRDX5 for carriers of the A allele. This is a shared risk allele with Crohn's Disease and MS\textsuperscript{1,2}. 

39
3.4 Functional Analysis of PRDX5 Locus

Our experimental approaches included fine mapping of the PRDX5 locus, Bayesian fine mapping, and using a functional prediction score (FUN-LDA) to predict which SNPs are causal in different tissue types. Using a fine-mapping algorithm, we placed some of the significant variants into relevant tissue types and predict the likelihood of them being causal in that tissue type. This is done using the posterior inclusion probability (PIP) score which is a result of Bayesian fine-mapping and provides weight of evidence that a SNP should be included as potential causal variant. PIP scores above 0.1 and 0.2 are predicted to be causal in the chosen tissue type. Using Bayesian fine mapping, we found a GWAS and exome significant variant, rs574087, that is predicted to be a causal variant in keratinocytes (Figure 3.3) and melanocytes (Figure 3.4) (with

![Figure 3.3. In silico validation of PRDX5 variants in Keratinocytes.](image)

The blue dots represent significant SNPs in this region. The green peaks represent the FUN-LDA score of likelihood of being causal in selected tissue type at specific gene base pair position. The red and blue X’s are significant SNPs predicted by posterior inclusion probabilities or PIP scores (red PIP>0.2, blue PIP>0.1). A high green peak together with an overlapping red or blue X SNP indicates high likelihood that SNP is causal in that specific tissue or cell type.
a posterior inclusion probability index or PIP score greater than 0.1, indicating high likelihood of causality).

This analysis shows high FUN-LDA peaks (yellow) overlapping with significant SNPs with high PIP scores (red and blue X's), suggestive of a high likelihood of causality of these SNPs in particular tissue types. Compared to another cell type, for example, neurospheres (Figure 3.5), we observed no overlap of peaks and high PIP SNPs, suggesting there is a specific role for PRDX5 variants in keratinocytes and melanocytes, but not in other cell types.

![Image](image-url)

**Figure 3.4. In silico validation of PRDX5 variants in melanocytes.** Map along the PRDX5 gene shows the FUN-LDA score, defined as the likelihood of each base pair on the gene to be functional in a specific tissue type (here Melanocytes, or E059 Foreskin Melanocyte Primary Cells skin01 as in the GTEx database). The blue dots represent significant SNPs in this region. The green peaks represent the FUN-LDA score of likelihood of being causal in selected tissue type at specific gene base pair position. The red and blue X's are significant SNPs predicted by posterior inclusion probabilities or PIP scores (red PIP>0.2, blue PIP>0.1). A high green peak and an overlapping red or blue X SNP indicates high likelihood that SNP is causal in that specific tissue or cell type.
Figure 3.5. FUN-LDA analysis of the PRDX5 region. We compared the PRDX5 region in a cohort of AA patients, using the FUN-LDA score in various tissue types from the GTEx database. The tissue types are keratinocytes (E058 as in the GTEx database), melanocytes (E059) and neurospheres (E053). These findings support that PRDX5 is likely to harbor causal variants in AA-relevant keratinocytes and melanocytes, and not in other tissue types such as brain stem cell related neurospheres.

3.5 Expression of PRDX5 in HF and Melanocytes

To functionally validate our in silico studies, we immunostained healthy human HF and AA-affected HF and found PRDX5 is upregulated in AA human HF (Figure 3.6).
PRDX5 is expressed in cultured melanocytes, which was noteworthy since melanocytes are known to have high levels of oxidative stress (Figure 3.7). We postulate that PRDX5 plays a crucial role in oxidative stress and its dysregulation can contribute to autoimmunity. PRDX5 may be enabling the survival of aberrant cells which may lead to presentation of damaged self-antigens to the immune system. Our findings establish a connection between PRDX5 causal variants and

**Figure 3.6. PRDX5 is overexpressed in AA human HF.** Immunostaining of PRDX5 (green) in unaffected HF and AA HF. PRDX5 is expressed throughout the HF in the inner and outer root sheath and the hair shaft cortex. In the AA HF, PRDX5 is upregulated compared to control. DAPI, blue.

**Figure 3.7. PRDX5 is expressed in Human Melanocytes.** Immunostaining of PRDX5 (green) in human cultured melanocytes. PRDX5 localizes with melanocytes (marked in red by melanocyte marker MART-1), which are known to have high levels of oxidative stress. DAPI, blue.
relevant tissue types which provides a framework for additional studies on the role of PRDX5 in AA disease pathogenesis.

3.6 Materials and Methods

3.6.1. Whole Exome/Genome Sequencing

Deep sequence analysis was conducted in several candidate regions to identify causal variants, together with whole exome sequencing using the Agilent SureSelect Human All Exon 50Mb plus 6.8-24Mb custom option for targeted enrichment of the genome for indexed sequencing on Illumina HiSeq sequencers. Genes were selected for targeted enrichment using GWAS regions, linkage peaks, confirmed CNVs, and pathways identified using gene expression analysis. We designed our custom capture regions to cover all AA GWAS regions. We also included in our custom capture regions linkage disequilibrium (LD) blocks implicated in our two previous GWAS in AA. Additionally, we sequenced key genes identified in our gene expression studies and functional experiments, as well as candidate genes guided by our biological knowledge of their function and potential to contribute to the pathogenesis of AA. Lastly, we included a few LD blocks implicated by GWAS in type 1 diabetes (T1D) that were associated with other autoimmune diseases and have expression or functional relevance to AA. In total, our custom capture regions covered approximately 17 MB.

3.6.2. Mapping AA Sequence Variants to Disease-Relevant Cell Types

The FUN-LDA score\textsuperscript{43} predicts the functional effects of non-coding genetic variants in particular tissue types. The FUN-LDA algorithm uses large genomics data sets like the ENCODE and Roadmap Epigenomics projects to integrate various epigenetic annotations for particular cell and tissue types. This unsupervised approach predicts the functional effects in a tissue-specific manner for every position in the human genome and for different functional variant classes and locations (i.e. promoters, enhancers, etc.). These integrated functional scores may elucidate the most likely tissue/ cell-type for complex diseases using GWAS summary statistics.
3.6.3. Gene Expression Analysis of AA Scalp Skin

Quality control of microarrays was performed using the affyAnalysisQC package from http://arrayanalysis.org/. Microarray preprocessing was performed using BioConductor in R. Differential expression in these studies was defined by an absolute fold change threshold of 1.5 with a Benjamini–Hochberg-corrected significance threshold of 0.05. The dataset was normalized using GCRMA and MAS5. The Affymetrix HGU-133Plus2 array contains 54675 probe sets (PSIDs). Filtering was performed to removed PSIDs that were on the X or Y chromosome, that represented Affymetrix control probe sets, or that did not have Gene Symbol annotations were removed from further downstream analysis. Correction for batch effects was performed using the implementation of the function ComBat available in the sva package with gender and AA group (AT/AU, AAP, and control) used as covariates.

Differential analysis was performed on the batch corrected discovery data set using linear models as implemented in the limma package in Bioconductor88. Two-sample comparisons were performed separately to identify PSIDs differentially expressed in AA patients versus normal controls, in AAP patients versus normal controls, and in AT/AU patients versus normal controls, treating gender as a fixed factor. This dataset was previously used in our lab in other studies42.

3.7 Conclusions

To study genetic susceptibility in the HF in AA, we analyzed peroxiredoxin 5 (PRDX5) (p= of 8.7*10^{-14}), which is also a GWAS gene in Crohn’s disease, sarcoidosis, and psoriasis15,16. PRDX5 is a member of the family of antioxidant enzymes that are crucial for regulating oxidative stress. The dysregulation of antioxidant enzymes has been linked to inflammatory processes since ROS can serve as important intracellular signaling molecule. An excess of ROS can enhance the inflammatory response, and thus require an antioxidant defense system to prevent ROS toxicity.
We identified variants in PRDX5 that were significantly enriched in both our GWAS and deep sequencing studies, and thus likely represent candidate causal variants in AA pathogenesis. Using Bayesian fine mapping, we discovered a significant variant, rs574087, in GWAS and exome sequencing, that is predicted to be causal in keratinocytes and melanocytes, with a posterior inclusion probability index or PIP score greater than 0.1, predicting a high likelihood of causality. This increased causality was not seen in non-AA relevant tissue types such as neurospheres, further supporting effects in keratinocytes and melanocytes. This variant, rs574087, was also identified in our meta-analysis studies (p=8.7*10^{-14}) and is located within the intergenic region.

Functional validation was performed by immunostaining healthy human HF and AA affected HF. Upregulation of PRDX5 in AA human HF was seen, likely due to PRDX5 enabling the survival of aberrant cells, which may lead to the presentation of damaged self-antigens to the immune system and the clinical manifestation of AA. PRDX5 was also expressed in cultured melanocytes by immunostaining, consistent with melanocytes having high levels of oxidative stress. We identified a connection between causal variants in PRDX5 and affected tissue types, which provides a mechanism for linking PRDX5 and AA disease pathogenesis.
Chapter 4. Environmental Factors: Gut Microbiome Dysbiosis in AA

4.1 Introduction

The mechanisms by which autoimmunity against the HF develops in AA are not fully elucidated. Our lab and others have demonstrated a strong genetic component in the development of AA\(^{47,98}\). However, there is only a 50% concordance rate of AA in identical twins, suggesting there exist yet unidentified environment factors in AA pathogenesis\(^{98}\). The microbiome has emerged as an environmental factor in many autoimmune diseases with roles in both disease onset and progression.

The mammalian microbial niche is highly complex and its diverse composition has immunomodulatory properties capable of eliciting both physiologic and pathologic responses in the host\(^{61}\). The microbiome has been linked to increased susceptibility for the development of a variety of autoimmune diseases such as rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, systemic lupus erythematosus and type I diabetes among others\(^{99}\). Interestingly, microbiome perturbations are evident both at the site of disease (i.e. atopic dermatitis and skin microbiome\(^{100,101}\)), as well as at distant sites than disease (i.e. uveitis and gut microbiome\(^{65}\)). Growing evidence suggests that gut dysbiosis induces disruption of epithelial barrier function and leads to gut-initiated systemic immune responses\(^{102}\). The mechanisms by which this occurs draws upon models of molecular mimicry, epitope spreading, and/or bystander effects\(^{102,103}\).

Studies in other skin autoimmune and inflammatory dermatological conditions such as atopic dermatitis (AD) and vitiligo have shown significant dysbiosis in the composition of the skin microbiome between patients and control individuals\(^{100,101}\). Other diseases such as uveitis and rheumatoid arthritis have shown gut dysbiosis and identified certain causal species contributing to disease pathogenesis\(^{65,104}\).
4.2 16S rRNA sequencing of AA Patients

To comprehensively investigate the role of the microbiome in AA, we first performed 16S rRNA sequencing in skin swabs from four distinct anatomical skin sites: lesional scalp (L), non-lesional scalp (NL), antecubital fossa (AC, inner elbow) and posterior auricular (PA, behind the ear), as well as plucked hairs (when present). We also collected nares samples for a mucus membrane (data not shown). We studied a cohort of 34 AA patients (14 new-onset cases, with AA presenting for the first time within the last year and 20 AA cases with longstanding disease, average ~7-8 years, see Materials and Methods Section 4.6.1). Analysis of the composition of the skin and HF microbiome showed no significant differences between AA patients and healthy controls (HCs) (Figure 4.1). From these analyses, we concluded that skin and HF microbial composition was not dysregulated in AA and is therefore not likely to be a trigger of disease onset.
Figure 4.1. Skin and hair follicle microbiome in AA show no significant dysbiosis between patients and healthy controls. No significant clustering or differences between patients and healthy controls in skin- PA and AC sites (A), skin- Scalp Non-Lesional (NL) (B) and Lesional (C) and hair follicle (D) microbiomes. Principal component analysis (PCA) plots are used to visualize relatedness between populations. PCA plot shows patients (pink) and HCs (blue) with no significant clustering. The overlap of samples and HCs is consistent with no significant dysbiosis.
Next, we investigated the role of the gut microbiome in AA pathogenesis, and performed 16S rRNA sequencing on the fecal samples from our cohort. Unexpectedly, unsupervised analysis of the bacterial taxa revealed significant differences in the gut microbiome composition of the AA group compared to HCs, with cluster fidelity assessed by Fisher’s exact test (p<0.00001) (Figure 4.2). Using 16S rRNA sequencing and bioinformatic analyses including linear discriminant analysis effect size (LefSe) (data not shown) and unsupervised hierarchical clustering, we detected a pronounced bidirectional dysbiosis characterized by an increased representation of members of the Firmicutes phylum and under-representation of the Bacteroides phylum in the gut microbiome of AA patients as compared with HCs.
Figure 4.2. Microbiome Analysis in 34 patients reveals distinct clustering of AA patients compared to controls. Differential expression of the most significantly associated taxa shows distinct differentially abundant OTUs. Unsupervised hierarchical clustering shows robust clustering between AA and HCs. Heatmap shows 34 patients (red bars) and 12 HCs (blue bars) across the top clustered by unsupervised hierarchical clustering using Pearson’s correlation. OTU IDs Phylum are listed on the right of each row. Highly abundant OTUs are in red and lower abundance OTUs are in blue. (p < 0.00001)
Using unsupervised hierarchical clustering on the most statistically significant Operational Taxonomic Units (OTUs), we found robust cluster fidelity between AA and HCs, with *Bacteroides* and *Firmicutes* bacterial genera representing the top differentially represented genera (Figure 4.3A). For example, certain OTUs in *Bacteroides* phyla, like OTU4339144 *B. Butyricimonas*, were significantly (p=0.00024) underrepresented in AA patients compared to HCs (Figure 4.3B). In contrast, OTU4381430 and OTU3444430 *Firmicutes* (family *Ruminococcaceae*) were statistically significantly (p=0.007 and p=0.0289) overrepresented in AA patients compared to HCs (Figure 4.3C). Notably, studies of the gut microbiome in other autoimmune diseases as well as cancer implicated a crucial role for the same predominant microbial phyla, *Bacteroides* and *Firmicutes* that were perturbed in AA patients (Table 4.2).

To test the robustness of these findings, we utilized a comprehensive machine-learning analytic pipeline novel to the field of microbiome studies. We first identified statistically significant microbiome components from an initial cohort of patients (n=13, p<0.00001) (Figure 4.3A and Figure 4.4). We validated the association in an independent cohort of patients, demonstrating the reproducibility of this microbial signature and its strong association with disease (n=21, p=0.000028) (Figure 4.4C).
Figure 4.3. Most significant microbial genera associated with AA. A. Most significantly different OTUs between AA and HCs. Red OTUs are highly abundant in AA and blue being lower abundance in AA compared to HCs. These 17 OTUs make up the training set (p<0.00001). B. OTU4339144 *Butyricimonas* is significantly (p=0.00024) underrepresented in AA, displayed as filtered abundance counts after rarefaction to minimum library size and Cumulative Sum Scaling (CSS) normalization. C. OTU344430 *Faecalibacterium* is significantly (p=0.0289) overrepresented in AA patients compared to HCs, displayed as filtered abundance counts after rarefaction to minimum library size and CSS normalization.
Figure 4.4. Training-Validation workflow shows differences in bacterial composition in the gut microbiome between patients and healthy controls. A. AA patients and HCs show distinct gut signatures. B. Identification of Discovery set. Differential expression of the most significantly associated taxa shows distinct differentially abundant OTUs in the initially sampled 13 patients. Unsupervised hierarchical clustering shows robust clustering between AA and HCs. Heatmaps both show AA patients (Red) and HCs (Blue) across the top clustered by unsupervised hierarchical clustering using Pearson’s correlation, and OTU IDs on the right. High abundance OTUs are shown in red and lower abundance OTUs are shown in blue (p<0.00001). C. Validation set. Differential abundance of the most significantly associated taxa from our training cohort (same 17 OTUs, just displayed in a different order) shows significant differential abundance and statistically significant clustering between patients and HCs in other 21 patients (p=0.000028).
Table 4.2. Directionality of changes in composition of *Firmicutes* and *Bacteroides* across different diseases

<table>
<thead>
<tr>
<th>Disease</th>
<th><em>Bacteroides</em></th>
<th><em>Firmicutes</em></th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitiligo</td>
<td>Increase</td>
<td>Increase</td>
<td>105</td>
</tr>
<tr>
<td>Colorectal Cancer</td>
<td>Decrease</td>
<td>Decrease</td>
<td>106</td>
</tr>
<tr>
<td>Inflammatory Bowel Disease</td>
<td>Decrease</td>
<td>Decrease</td>
<td>107</td>
</tr>
<tr>
<td>Systemic Lupus Erythematosus</td>
<td>Decrease</td>
<td>Decrease</td>
<td>99</td>
</tr>
<tr>
<td>Type 1 Diabetes</td>
<td>Increase</td>
<td>Decrease</td>
<td>107</td>
</tr>
<tr>
<td>Autism Spectrum Disorder</td>
<td>Increase</td>
<td>Decrease</td>
<td>107</td>
</tr>
<tr>
<td>Celiac Disease</td>
<td>Increase</td>
<td>Decrease</td>
<td>108</td>
</tr>
<tr>
<td>Scleroderma</td>
<td>Decrease</td>
<td>Increase</td>
<td>99</td>
</tr>
<tr>
<td>Multiple Sclerosis</td>
<td>Decrease</td>
<td>Increase</td>
<td>109</td>
</tr>
<tr>
<td>Grave’s Disease</td>
<td>Decrease</td>
<td>Increase</td>
<td>99,110</td>
</tr>
<tr>
<td>Alopecia Areata</td>
<td>Decrease</td>
<td>Increase</td>
<td>This study</td>
</tr>
</tbody>
</table>

To determine whether this pattern of dysbiosis was a unique characteristic of the gut microbiome in AA, we compared these results to a different cohort of patients with inflammatory (scarring) alopecias. We analyzed the gut microbiome of 52 scarring alopecia patients and found no clustering or distinct differential OTU representation in the skin (*data not shown*), HF (*data not shown*), or gut microbiome (p=0.2932) (Figure 4.5). Together, these results support a unique role for the gut microbiome in AA autoimmunity that is not shared by other forms of inflammatory alopecia.
4.3 Depletion of Gut Microbiome in C3H/HeJ Mouse model of Alopecia prevents AA

To determine the requirement for the gut microbiome in the initiation and onset of AA in vivo, we utilized the well-established C3H/HeJ mouse model\(^\text{15}\). We depleted the mouse gut microbiome using broad-spectrum antibiotic regimen (Ampicillin, Neomycin and Vancomycin) in drinking water ad libitum, that was given at three different timepoints: before, after, or at the time

Figure 4.5. Scarring Alopecias show no gut dysbiosis. A. PCA plot of gut microbiome shows no significant clustering (p=0.2932) between scarring alopecia patients (blue) and HCs (pink). Unlike the distinct clustering seen in our AA autoimmune gut samples, PCA samples show no distinct separation or gut dysbiosis.
of AA induction by skin grafting. When the microbiome was depleted 5 weeks before grafting or at the time of grafting (8 weeks), the recipient animals were protected against hair loss within the observation period of 10 weeks (Figure 4.6A&C). In contrast, when microbiome was depleted 2 weeks after grafting and induction of disease, no protection from AA was observed (Figure 4.6A). These data suggest that the gut microbiome is required for AA induction and disease onset in the C3H/HeJ mouse model.

In order to determine whether changes in the skin or gut microbiome composition played a role in the development of disease in the murine AA model, we conducted 16S rRNA sequencing of dorsal skin swabs and fecal samples from the antibiotic treated and untreated mice. We found no significant changes between antibiotic treated and untreated groups in the skin microbiome (Figure 4.6B). In contrast, 16S rRNA sequencing in fecal samples (Figure 4.6D) revealed statistically significant differences between antibiotic-treated and untreated groups 10 weeks after baseline (no treatment), with the most statistically significant OTUs between groups being of phyla Firmicutes and Bacteroides, consistent with our findings in AA human gut microbiota. For example, Bacteroidaceae showed a statistically significant under-representation in both mice with hair loss and AA patients. Together, dysbiosis in the mouse model of AA reflected our findings of bidirectional dysbiosis in human AA.
Figure 4.6. Depletion of gut microbiome prevents onset of AA. A. Experimental design of antibiotic treated C3H/HeJ mice. Mice were given antibiotics (ANV) in drinking water at different time points before grafting (Experiments 1), at the time of grafting (Experiment 2), or after grafting (Experiment 3). We found that the animals were partially protected from hair loss giving antibiotics before grafting (Experiment 1) and not protected from hair loss giving antibiotics after grafting (Experiment 3). Optimal protection from hair loss was achieved by giving antibiotics at time of grafting (Experiment 2). B. No differences in microbiome diversity in skin is shown by non-clustering or random distribution of samples (Experiment 2) C. Oral Antibiotic treatment delays onset of AA in C3H/HeJ mice (Experiment 2) D. Principal Component Analysis (PCA) reveals differences in gut microbiome composition at baseline, 10 weeks untreated and 10 weeks antibiotic-treated mice. Mice at baseline (upon arrival in our facility, green), mice after 10 weeks either untreated (adapted to Columbia University mouse facility microbiome, blue) or antibiotic treated (red).
4.4 Metagenomics Sequencing Revealed AA-specific microbes in patients

We next performed metagenomics on human stool samples to further identify the microbes potentially contributing to disease. Metagenomics is a deep sequencing technique that has higher resolution than 16S rRNA sequencing and can identify individual causal microbes. The first finding from our metagenomics study examined alpha diversity in our patients vs. controls. Alpha diversity is a metric that predicts (within-group) diversity and it accounts for evenness and distribution in each sample in each group. We noted an overall Shannon diversity decrease in AA compared to controls (Figure 4.7). Shannon diversity is a metric used to measure alpha diversity, typically values are between 1.5 and 4\(^1\). This result indicates there is a normal distribution of gut microbes, but overall, there are less diverse taxa in the gut of AA patients, consistent with the gut dysbiosis observed in AA.

\[\text{Figure 4.7. Reduced gut microbiome diversity in AA.}\] Plot shows alpha diversity, which is a metric showing within-group diversity. It accounts for evenness and distribution in each sample in each group. Overall, our data shows normal distributions of microbes, but less diversity in taxa of AA gut.
We next examined beta diversity using Bray Curtis dissimilarly (which measures compositional dissimilarity between groups). This measurement is based on the abundance of the microbial species present in each sample in the diseased group, and is then compared to the control group. From this deeper resolution metagenomics sequencing, we observed a clear separation of AA vs. controls, further supporting the distinct separation between AA and HCs demonstrated in our 16S PCA plots (Figure 4.8).

**Figure 4.8. Gut Microbiome Beta Diversity in AA.** Plot shows beta diversity, using Bray Curtis dissimilarity metric, which measures compositional dissimilarity between groups. Overall, our data shows that between the cases and control groups, there are distinct and differentially abundant taxa.
Using unsupervised hierarchical clustering of our metagenomics data, we also observed robust clustering between cases and controls. We identified several bacterial species that are over- and under-represented in AA versus healthy controls (Figure 4.9).

![Heatmap showing metagenomic differential abundance of species.](image)

**Figure 4.9. Metagenomic differential abundance of species.** Heatmap shows distinct clustering between cases and controls with most significantly differentially abundant genera expressed. High abundance species are shown in red and low abundance species are shown in blue.

We next examined our metagenomic data for particular species that were overabundant in AA and could be potentially pathogenic or under-represented in AA. Figure 4.10 shows some of these specific microbes.
Figure 4.10. Representative microbes identified by metagenomic sequencing. Utilizing metagenomic sequencing, we identified representative species underrepresented in AA (top row) and overabundant in AA (bottom row). A. Dialister was statistically significantly (p=0.0048) reduced in AA versus controls. B. Oscillibacter was statistically significantly (p=0.0381) reduced in AA versus controls. C. Lactobacillales was statistically significantly (p=0.0048) reduced in AA versus controls. D. Ruminococcus sp. 5_1_39BFAA was statistically significantly (p=0.0048) increased in AA versus controls. E. Alistipes Shahii was statistically significantly (p=0.0333) increased in AA versus controls. F. Blautia sp. Marseille-P2398 was statistically significantly (p=0.0095) increased in AA versus controls. (See Methods).
4.5 Short Chain Fatty Acid Analysis

Various metabolites and nutrients from the host and commensal bacteria regulate $T_{reg}$ generation, trafficking and function. Among these metabolites, short chain fatty acids (SCFA) are products of the bacterial fermentation of dietary fiber, such as butyrate, acetate, and propionate, which can influence $T_{reg}$ differentiation and function$^{112,113}$. In order to investigate the relationship between the gut microbiome and its products to the development of AA, we quantitated the levels of SCFA in fecal samples from 13 AA patients to 13 controls. While no specific acid was statistically significant, we identified changes in each of the SCFAs studied (Figure 4.11).

![Figure 4.11: Short Chain Fatty Acid Analysis in AA Patients.](image)

We observed a trend toward an increase in SCFA production in AA patients versus controls. Analysis of variance was used to test statistical significance between AA and controls. A. Acetic Acid (p=0.202), B. Butyric Acid (p=0.393), C. Hexanoic Acid (p=0.076), D. Propionic Acid (p=0.35), E. Isobutyric Acid (p=0.217), F. Valeric Acid (p=0.152), and G. Isovaleric Acid (p=0.324).
4.6 Materials and Methods

4.6.1. Subject selection

Patients with AA at all stages from new onset disease to long standing stable diseases in the subtypes patchy, alopecia totalis (AT), and alopecia universalis (AU) were enrolled from a tertiary care dermatology clinic; from the public at large as interest accumulated from public advertisement; and from AA support group recruitment. Consistent with the known female-to-male distribution for AA (2.3:1)\textsuperscript{114}, this cohort included 23 females and 11 males. In the cohort, 14 patients were diagnosed with alopecia universalis (AU)/ alopecia totalis (AT) whereas 20 patients had patchy AA. We also studied twelve healthy age matched controls for comparison. We did not find any disparities based on age, gender, or race. All patients stopped AA treatment for at least 30 days before participation, however some continued concurrent medications for co-morbid conditions including cardiovascular disease, hypothyroidism, and asthma (Table 4.1). Each patient enrolled contributed scalp samples, through biopsy, peripheral blood samples, surface skin microbiome swabs, HF microbiome samples and GI samples as available. We collected lesional and non-lesional scalp biopsies from patients with patchy alopecia and only lesional biopsies from patients with AU and AT. Surface skin swabs were collected from AA lesional and non-lesional scalp whenever possible as well as matching posterior auricular, antecubital fossa and nares swabs. Scalp lesional and non-lesional swabs, as well as posterior auricular and antecubital fossa swabs were first dipped in 20 mM Tris pH 8, 2 mM EDTA, 1.2% Triton X-100 buffer then rubbed on the skin surface about 50 times\textsuperscript{115}. Nares swabs were not dipped in buffer and were swabbed 10 times in one nare. Each patchy AA patient contributed HF microbiome samples through anagen hair pluck, and HF microbiome sample was collected from AT and AU patients, whenever possible (patients with SALT score of >95%). GI stool samples were collected from each patient within a few days the office visit, using OMNIgene-GUT (OMR-200) kits.
Table 4.1. Clinical Data for all AA patients in the microbiome cohort.

<table>
<thead>
<tr>
<th>Number</th>
<th>Race</th>
<th>Gender</th>
<th>Age</th>
<th>Disease Subset</th>
<th>Time since Onset</th>
<th>Other Diseases</th>
<th>Previous Medications</th>
<th>Special Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>942</td>
<td>Caucasian</td>
<td>F</td>
<td>23</td>
<td>AT</td>
<td>8 years</td>
<td>None</td>
<td>None</td>
<td>patchy onset 15 years ago</td>
</tr>
<tr>
<td>944</td>
<td>Caucasian</td>
<td>F</td>
<td>22</td>
<td>AU</td>
<td>9 years</td>
<td>Ilihaisisemia, Raynaud's disease</td>
<td>OCP</td>
<td></td>
</tr>
<tr>
<td>946</td>
<td>African</td>
<td>F</td>
<td>24</td>
<td>AT</td>
<td>4 years</td>
<td>None</td>
<td>biotin</td>
<td></td>
</tr>
<tr>
<td>947</td>
<td>Caucasian</td>
<td>F</td>
<td>56</td>
<td>AT</td>
<td>8 months</td>
<td>Charcot Marie Tooth, peripheral neuropathy</td>
<td>metformin</td>
<td></td>
</tr>
<tr>
<td>948</td>
<td>Caucasian</td>
<td>M</td>
<td>18</td>
<td>AAP</td>
<td>4 months</td>
<td>Sleep apnea</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>949</td>
<td>Caucasian</td>
<td>M</td>
<td>65</td>
<td>AAP</td>
<td>9 months</td>
<td>Acne</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>950</td>
<td>Caucasian</td>
<td>F</td>
<td>18</td>
<td>AT</td>
<td>4 months</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>954</td>
<td>Caucasian</td>
<td>M</td>
<td>26</td>
<td>AT</td>
<td>10 months</td>
<td>Asthma</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>955</td>
<td>African</td>
<td>F</td>
<td>46</td>
<td>AT</td>
<td>20 years</td>
<td>Hypothyroidism</td>
<td>levothyroxine</td>
<td></td>
</tr>
<tr>
<td>956</td>
<td>Caucasian</td>
<td>F</td>
<td>72</td>
<td>AT</td>
<td>5 years</td>
<td>Crow's disease, seriate bruising, low back spinal surgery, osteoporosis, eye</td>
<td>cefuroxime, topical retinoid</td>
<td></td>
</tr>
<tr>
<td>957</td>
<td>Caucasian</td>
<td>F</td>
<td>60</td>
<td>AAP</td>
<td>4 months</td>
<td>History of cervical cancer, OA</td>
<td>metformin</td>
<td></td>
</tr>
<tr>
<td>958</td>
<td>Caucasian</td>
<td>F</td>
<td>37</td>
<td>AU</td>
<td>24 months on</td>
<td>Abnormal pap smear (unspecified)</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>959</td>
<td>Caucasian</td>
<td>M</td>
<td>30</td>
<td>AAP</td>
<td>15 years</td>
<td>Seasonal allergies</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>961</td>
<td>Caucasian</td>
<td>F</td>
<td>27</td>
<td>AAP</td>
<td>8 years</td>
<td>Bleeding disorder</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>963</td>
<td>Asian</td>
<td>M</td>
<td>44</td>
<td>AAP</td>
<td>28 years</td>
<td>Hypertension</td>
<td>Amlodipine, fish oil, biotin; Vitamin D</td>
<td></td>
</tr>
<tr>
<td>968</td>
<td>Hispanic</td>
<td>F</td>
<td>32</td>
<td>AAP</td>
<td>3 years</td>
<td>Ulitmate polyph, GERD</td>
<td>Prenatal vitamin</td>
<td></td>
</tr>
<tr>
<td>968</td>
<td>Hispanic</td>
<td>M</td>
<td>48</td>
<td>AAP</td>
<td>8 months</td>
<td>Eczeema, Allergies, Acne vulgaris, Linea corporis, Seborehe dematitis</td>
<td>Xeljanz, Tacrolimus, cyclophosphamide</td>
<td></td>
</tr>
<tr>
<td>969</td>
<td>African</td>
<td>F</td>
<td>19</td>
<td>AAP</td>
<td>1.5 years</td>
<td>Seasonal allergies, Asthma</td>
<td>None</td>
<td>was at 9 years old, now patchy for the last 11 months</td>
</tr>
<tr>
<td>974</td>
<td>Caucasian</td>
<td>M</td>
<td>31</td>
<td>AAP</td>
<td>2 years</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>975</td>
<td>Caucasian</td>
<td>M</td>
<td>30</td>
<td>AAP</td>
<td>3 months</td>
<td>Seborrheic Dermatitis, Pyrilarion rosea, hx of appendicitis</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>976</td>
<td>Caucasian</td>
<td>F</td>
<td>54</td>
<td>AAP</td>
<td>6 months</td>
<td>Hx of endometriosis, Hashimoto thyroiditis, Mononaplasis</td>
<td>Escitalopram, Levothyroxine, estrace, vitamin D, fish oil, mycetes</td>
<td></td>
</tr>
<tr>
<td>979</td>
<td>Caucasian</td>
<td>M</td>
<td>37</td>
<td>AU</td>
<td>1 year</td>
<td>Hx of gallstones</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>983</td>
<td>Caucasian</td>
<td>F</td>
<td>52</td>
<td>AAP</td>
<td>11 months</td>
<td>Contact dermatitis; Malaria, Pailapitans</td>
<td>Zolit, cionaapram, Propanol, hidin, tin, vitamin D, probiotic</td>
<td></td>
</tr>
<tr>
<td>985</td>
<td>Asian</td>
<td>M</td>
<td>30</td>
<td>AT</td>
<td>10 years</td>
<td>Scaly temnails, acne, hydropigmentation</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>986</td>
<td>Caucasian</td>
<td>F</td>
<td>31</td>
<td>AU</td>
<td>2 years</td>
<td>Hyperhidrosis</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>987</td>
<td>Caucasian</td>
<td>F</td>
<td>44</td>
<td>AAP</td>
<td>4 years</td>
<td>Lactose intolerance</td>
<td>Xeljanz, patchy of the time of study of been at 2 other times in her life, 3 years rapidly progressive to at 2/2 weight gain</td>
<td></td>
</tr>
<tr>
<td>990</td>
<td>African</td>
<td>F</td>
<td>24</td>
<td>AU</td>
<td>1 year</td>
<td>None</td>
<td>None</td>
<td>Has used Xeljanz in the past</td>
</tr>
<tr>
<td>991</td>
<td>Caucasian</td>
<td>F</td>
<td>60</td>
<td>AAP</td>
<td>1.5 years</td>
<td>None</td>
<td>OCP (birth control), spiroclolaterone</td>
<td></td>
</tr>
<tr>
<td>994</td>
<td>Caucasian</td>
<td>F</td>
<td>33</td>
<td>AAP</td>
<td>5 years</td>
<td>Asthma</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>995</td>
<td>Caucasian</td>
<td>F</td>
<td>15</td>
<td>AAP</td>
<td>1.5 years</td>
<td>Anemia</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>996</td>
<td>Caucasian</td>
<td>F</td>
<td>55</td>
<td>AU</td>
<td>2.5 years</td>
<td>Hyperlipidemia, Hypertension, History of herpes zoster</td>
<td>Amlodipine, Clofibrate, Ibesartan-HCTZ, Metoprolol, Valacyclovir, Cefuroxime, Aspirin</td>
<td></td>
</tr>
<tr>
<td>997</td>
<td>Asian</td>
<td>M</td>
<td>28</td>
<td>AAP</td>
<td>3 months</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>998</td>
<td>Hispanic</td>
<td>F</td>
<td>41</td>
<td>AAP</td>
<td>1 month</td>
<td>Asthma</td>
<td>Inhaler (LABA, albuterol), singular</td>
<td></td>
</tr>
<tr>
<td>999</td>
<td>Hispanic</td>
<td>F</td>
<td>22</td>
<td>AAP</td>
<td>7 years</td>
<td>Recurrent sinus infections, Asthma, major depression</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>
The microbiome of skin samples from the posterior auricular (PA), antecubital fossa (AC), Non-lesional scalp (NL), and Lesional scalp (L) sites showed a similar distribution of Operational Taxonomic Units (OTU) in AA patients and HCs, with no obvious clustering (Figure 4.1A-C). In addition, when we compared the OTUs represented in the HF, we found a similar microbiome composition between AA patients and HCs (p=0.5692) (Figure 4.1D). The profile of the microbiome at the skin sites was comparable to previously published studies showing abundances of *Firmicutes, Bacteroides, Actinobacteria, and Proteobacteria*[^118]. One study from the literature showed different microbial composition of different areas surrounding the HF (from the epidermis to hypodermis) in AA versus HCs[^117], but other studies have shown that irrespective of disease, the different skin compartments contain a diverse and partially distinct microbial community[^118], supporting our finding of minimal, if any, detectable differences in the skin and HF microbiome composition.

### 4.6.2. 16S v3-v4 rRNA sequencing.

Fecal DNA was extracted using the MagAttract PowerSoil DNA kit (Qiagen), and DNA from hair and swab samples was extracted using the AllPrep PowerViral RNA/DNA kit (Qiagen). Extracted DNA was quantified using the Quant-iT dsDNA High Sensitivity assay kit (Invitrogen). We amplified the V3-V4 16S rRNA region and barcoded samples using the Nextera XT Index Kit (Illumina). Libraries were sequenced on an Illumina MiSeq platform using the 2 X 300bp v3 kit with a 10% PhiX spike (Columbia University Irving Medical Center Microbiome Core).

### 4.6.3. 16S rRNA data analysis.

Raw sequences were filtered and closed-reference operational taxonomic unit (OTU) calling was performed against Greengenes database (97% cut-off) using QIIME. The phyloseq package in R was used to filter low-abundance taxa (average relative abundance <0.005%) and calculate both alpha-diversity (Shannon, Chao indices) and beta-diversity (unweighted and weighted UniFrac). Based on α-diversity rarefaction curves, minimum count cutoffs were established for each sample type to accurately calculate diversity metrics (stool: 7500; hair: 4000; swabs: 4000)
4.6.4. 16S Statistical analyses

OTU mapping was handled by the Columbia University Microbiome Core using standard, published alignment protocols. OTU tables were then internally normalized and corrected for batch effects using Bioconductor and QIIME software packages available from R\textsuperscript{119,120}. All subsequent analyses were conducted both with normalized OTU depth, i.e. considering only OTUs that map to the same phylogenetic depth. This step was taken to prevent analytic bias towards the more generally conserved sequences contained in higher-order depths, and to provide a standardized depth that provided sufficient OTUs for clustering analysis. OTU differential distribution across samples was assessed by non-parametric distribution statistics, e.g. both supervised and unsupervised machine learning algorithms using leave-in-out cross-validation for initial cohorts using MeV: MultipleExperiment viewer software\textsuperscript{121}. Subsequent cohorts were analyzed independently to be leveraged as validation sets in machine learning analyses. Associations between OTU clusters and clinical features was assessed using concordance statistics such as Fisher’s exact test and programs such as the DESeq2 package in Bioconductor which can estimate variance-mean dependence in count data from high-throughput sequencing assays and test for differential expression based on a model using the negative binomial distribution. Unless explicitly specified, an alpha of 0.05 was used as the threshold for consensus-based statistical tests.

4.6.5. Metagenomics

Stool was collected in the OMNIgene-GUT kit, as previously described. This study utilized the same patient cohort as previously described. Six AA patients and four controls had DNA extracted and sequenced at an external vendor (Genewiz). OTU mapping and preliminary bioinformatic analysis were also performed at Genewiz. Associations between OTU clusters and clinical features was assessed using concordance statistics such as Fisher’s exact test and Mann-Whitney U Test. and programs such as the DESeq2 package in Bioconductor which can estimate variance-mean dependence in count data from high-throughput sequencing assays and test for
differential expression based on a model using the negative binomial distribution. Unless explicitly specified, an alpha of 0.05 was used as the threshold for consensus-based statistical tests.

4.6.6. Short Chain Fatty Acid Analysis

The SCFA extraction procedure utilizing fecal samples is similar to that reported by Zhao et al (2006)\textsuperscript{122}. Material was resuspended in MilliQ-grade H2O, and homogenized using MP Bio FastPrep, for 1 minute at 4.0 m/s. 5M HCl was added to acidify fecal suspensions to a final pH of 2.0. Acidified fecal suspensions were incubated and centrifuged at 10,000 RPM to separate the supernatant. Fecal supernatants were spiked with 2-Ethylbutyric acid for a final concentration of 1 mM. Extracted SCFA supernatants were stored in 2-ml GC vials with glass inserts. SCFA were detected using gas chromatography (Thermo Trace 1310), coupled to a flame ionization detector (Thermo). Our SCFA column is 'Thermo TG-WAXMS A GC Column, 30 m, 0.32 mm, 0.25 um, which is very similar to the instrument and method adapted from the literature\textsuperscript{122}. This was performed by external vendor (Microbiome Insights).

Short chain fatty acid (SCFA) data was analyzed on a gas chromatograph (GC) coupled with a flame ionization detector (FID). Concentrations were normalized to the amount of input material (mmol SCFA/ kg stool aliquoted from OMNIgene-GUT kit). Analysis of variance was used to test statistical significance among the variables, along with, if applicable, a Tukey’s honest significance test (Tukey-HSD) to report pairwise significance.

4.7 Conclusions

To determine the role of various body site microbiomes in AA pathobiology, we performed 16S rRNA sequencing of skin swabs, HFs, and stool samples from a cohort of 34 AA patients and 12 healthy controls (HCs). We found no significant differences in skin and HF microbiome composition between AA patients and HCs. Unexpectedly, we found striking gut dysbiosis, consisting of over-representation of \textit{Firmicutes} and under-representation of \textit{Bacteroides} in the gut microbiome of AA patients as compared with HCs. We also identified several differentially
abundant taxa in AA gut microbiome that potentially associated with disease. Using a computational machine-learning analytic pipeline, we first identified dysbiotic microbiome components from an initial cohort of 11 AA patients. We then validated the discovery association in an additional independent patient cohort. The strong association of the dysbiotic microbial signature to AA demonstrated reproducibility.

To functionally test the requirement of the gut microbiome in AA development *in vivo*, we depleted the gut microbiome in C3H/HeJ mice, and found that the mice were largely protected from AA induction, confirming a requirement for the gut microbiome in the onset of murine AA. We found that the skin and HF microbiome is unchanged with disease; however, gut microbiota dysbiosis is evident in AA patients and required for AA onset in the murine model of disease. Our findings revealed strong evidence for the role of gut microbiome in AA, and how perturbations of the gut microbiota can delay disease onset in mice. Whether dysbiosis in humans is causal or a consequence of AA is still under investigation, our study with the pre-clinical mouse model suggests that depletion of gut microbiota (or potentially depletion of certain phyla) has a protective role. We identified robust dysbiosis and key species involved in AA in our metagenomics deep sequencing. SCFA analysis revealed changes in SCFA production in AA patients. We identified trends towards overabundance of SCFAs in AA patients versus controls, which supports the increase in microbial species known to produce SCFA we observed in our 16S and metagenomics sequencing (increases in *Firmicutes* phyla, such as *Clostridia*). Taken together, this work uncovered a novel role of the gut microbiome in AA autoimmunity.

Our findings suggest restoring the homeostasis of the gut microbiome may be useful for the successful clinical management of patients with AA. This provides rationale for further functional analyses into the mechanism of gut dysbiosis and AA and potential therapeutic targets. Strategies such as FMT, antibiotic therapy, and other methods for microbial perturbation could provide safe methods for restoring healthy gut composition and improving disease outcomes.
Chapter 5. Discussion

5.1 Overview Section

AA is believed to result from both genetic and environmental factors, both of which were investigated in this thesis. Among genetic risk factors in AA, we found that risk variants cluster in genes regulating the immune response, as well as the target organ. To identify underlying genetic drivers in AA, we analyzed AA risk genes using various sequencing techniques and analysis methods to identify causal variants, and placed them in functionally relevant contexts using innovative mapping techniques. First, we studied the Interleukin-2 Receptor Alpha (IL2RA) region, which was identified as a significant locus in our AA GWAS\textsuperscript{11}, and plays a crucial role in regulating immune tolerance and controlling activity of T\textsubscript{reg}\textsuperscript{13}. IL2RA variants identified in our sequencing and \textit{in silico} studies support that certain polymorphisms in Il2RA result in lower CD25 expression on CD4 naïve T cells. In turn, that reduction of CD25 expression by variants could reduce the likelihood that naïve CD4 T cells are activated to an inflammatory state.

Secondly, to study susceptibility genes in the hair follicle (end organ), we examined Peroxiredoxin 5 (PRDX5). PRDX5 is an antioxidant enzyme crucial for regulating oxidative stress. Oxidative stress and excess reactive oxygen species (ROS) production occurs in many autoimmune diseases and suggests that the increased PRDX5 expression observed in AA, may be involved in combatting the excess ROS. Lastly, we investigated the role of the microbiome in human and mouse AA. We found striking gut dysbiosis, consisting of over-representation of \textit{Firmicutes} and under-representation of \textit{Bacteroides} in the gut microbiome of AA patients as compared with healthy controls. We also identified several differentially abundant bacterial taxa in AA gut microbiome that could potentially modulate disease activity in some of our metagenomics studies. Strategies such as FMT, antibiotic therapies, and other approaches could provide new methods for restoring healthy gut composition and improving disease outcomes in AA.
5.2 Genetic Factors in AA

**IL2RA**

IL2RA, identified in our GWAS\(^1\), plays a critical role in regulating immune tolerance and controlling activity of T\(_{\text{regs}}\)^\(^1\). We identified significant causal variants in the IL2RA region using GWAS, targeted resequencing, and custom capture exome sequencing approaches. We then validated *in silico* the expression of these variants in immune cell types and specifically in CD4+ T cells versus end organ-specific cells like keratinocytes. The variant rs3118740 decreased IL2RA gene expression for carriers of the T allele. Allele specific effects such as this may lead to a perturbation of T\(_{\text{reg}}\) function as seen in one study in T1D where patients with the rs3118470 risk variant have T\(_{\text{regs}}\) with IL-2 signaling defects\(^4\). Since the IL2RA locus is a widely studied autoimmune locus, there are extensive opportunities for functional analysis. Testing the function of certain rare variants in T cell differentiation might be informative for protective or pathogenic roles they may be playing in AA pathogenesis.

Since CD25-positive naïve CD4+ T cells expressing the high affinity IL-2 receptor can simultaneously be activated by IL-2 and engagement of the T cell receptor, the reduction of CD25 expression by the protective phenotype of the rs2104286 allele may reduce the likelihood of naïve CD4+ T cells being activated under proinflammatory conditions. Further functional validation can be done to study the role of IL2RA over the progression of disease and whether it correlates to disease severity.

Low dose recombinant IL-2 therapy has been used in many autoimmune diseases (such as T1D, SLE, among others) to promote T\(_{\text{reg}}\) development and expansion, and thus suppress unwanted immune responses to treat these diseases\(^{123}\). Studies in SLE specifically have shown success in human clinical trials, with another 14 ongoing trials\(^{123}\). In AA, low dose IL-2 therapy has been shown to increase numbers of T\(_{\text{reg}}\) cells, while decreasing the number of infiltrating CD8+ T cells in scalp biopsies of AA patients during and after treatment\(^{124}\). Our *in silico* validation
confirms the role of IL2RA in AA consistent with this mechanism supporting IL-2 therapy\textsuperscript{124,125} to improve T\textsubscript{reg} function.

\textbf{PRDX5}

To uncover genetic susceptibility in the target organ in AA, we studied a HF specific candidate gene from our GWAS, PRDX5\textsuperscript{11}. This gene is also significant in GWAS of other diseases such as psoriasis, Crohn’s disease, and sarcoidosis\textsuperscript{15,16}. Using a test of enrichment, we identified variants that are significant in both our GWAS and exome studies, and thus represent likely candidate causal variants in our cohort of 849 patients. Using Bayesian fine mapping, we found a GWAS and exome significant variant in PRDX5, rs574087, that is predicted to be a causal variant in keratinocytes and melanocytes, with a posterior inclusion probability index or PIP score greater than 0.1, lending to high likelihood of causality. We functionally validated these findings by immunostaining HFs and cultured melanocytes. AA HFs showed increased expression of antioxidant enzyme PRDX5 in HF epithelial cells. We postulate that PRDX5 plays a crucial role in oxidative stress and its dysregulation can predispose to autoimmunity.

PRDX5 has a known role in oxidative stress and we observed an upregulation of PRDX5 in the AA HFS. PRDX5 may enable the survival of aberrant cells, which may lead to the presentation of damaged self-antigens to the immune system\textsuperscript{11}. PRDX5 is also expressed in human melanocytes, which exhibit high levels of oxidative stress. We also detected PRDX5 expression in the colonic mucosa, consistent with the association of PRDX5 in Crohn’s disease\textsuperscript{15}. One study in RA, identified a single microbe that restored gut dysbiosis and protected from disease in a rat model, which was hypothesized to occur by modulating redox balance\textsuperscript{126}. This mechanism was proposed due to an increase in key genes involved in redox balance being more abundant in the rats after transplanting the gut microbiome with this single microbe, which promoted anti-arthritis effects by maintaining the redox balance of oxidative stress\textsuperscript{126}. There is an established bidirectional crosstalk between the microbiota and the mitochondria in redox balance (key in maintaining the gut barrier integrity and immune response) which are crucial functions in
the microbiome\textsuperscript{127}. Our findings of high expression of PRDX5 in tissues known to have high levels of ROS (i.e. AA HF or colonic mucosa), enable us to correlate the increase in PRDX5 to redox balance and maintenance of gut barrier integrity.

5.3 Environmental Factors in AA

To determine the role of various body site microbiomes in AA pathobiology, we performed 16S rRNA sequencing of skin swabs, HFs, and stool samples from a cohort of 34 AA patients and 12 healthy controls (HCs). Our results revealed an under-representation of the \textit{Bacteroides} phyla and over-representation of the \textit{Firmicutes} phyla in AA patient gut microbiomes, but no significant differences in skin or HF microbiome composition. To study the role of the gut microbiome in AA development \textit{in vivo}, we depleted the gut microbiome in C3H/HeJ mice, and found that the mice were largely protected from AA induction, revealing a requirement for gut microbiota in mouse AA.

Notably, there is precedent for relative abundance changes in \textit{Firmicutes} and \textit{Bacteroides} in both autoimmune and non-autoimmune processes. \textit{Bacteroides} and \textit{Firmicutes} have also been implicated in differential outcomes of cancer immunotherapy. For example, \textit{B. fragilis} has been reported to enhance efficacy of anti-CTLA-4 cancer immunotherapy through involvement of Th1 cell activation and cross-reactivity of bacterial neoantigens with tumor neoantigens\textsuperscript{128}. Although the underlying mechanisms are still incompletely defined, there is growing evidence suggesting that the gut microbiome composition can play either a immunosuppressive or immunostimulatory role in anti-tumor immunity\textsuperscript{128}. Studies in other autoimmune diseases, such as T1D, have shown that gut microbiome perturbations can accelerate disease onset and incidence\textsuperscript{129,130}. Thus, changes in the composition of the gut microbiota may also induce immunoregulatory mechanisms triggering and/or preventing the onset of autoimmunity. The composition and variability of the gut microbiome may contribute to autoimmunity through molecular mimicry, direct microbial stimuli,
loss of the beneficial anti-inflammatory metabolites, and/or pathogen translocation that stimulate immune mediators\textsuperscript{102}.

The microbiome holds great therapeutic potential for the management of AA. Clinical trials and longitudinal studies in FMT will elucidate potentially pathogenic or protective microbes or communities of microbes that could be used in treatment of AA. Deeper resolution sequencing approaches, like metagenomics and metabolomics, in larger cohorts of AA patients can help further identify bacterial species and metabolic pathways involved in disease. Our metagenomic studies identified a decrease in \textit{Firmicutes} phyla dialister. Not much is known about this particular species, but there is also a decrease in this microbe in T1D and lupus, suggesting a connection to autoimmunity\textsuperscript{131,132}. Our metagenomic studies also revealed an increased expression of \textit{Ruminococcus} sp. 5\_1\_39BFAA in AA. This species has also been shown to be increased in Type 1 Diabetes and Psoriasis\textsuperscript{133,134}. While the function of this microbe is currently unknown, in the literature other members of the \textit{Ruminococcus} genera have been shown to produce inflammatory polysaccharides\textsuperscript{135}. SCFA butyrate production is limited to few genera and \textit{Ruminococcus} is one of the key producers\textsuperscript{136} consistent with the trend towards increases in SCFA observed in AA patients.

Our SCFA studies revealed trends toward overabundance of SCFAs in AA patients versus controls, which reflects similar trends in other disease states. An interesting study showed that nonalcoholic fatty liver disease (NAFLD) patients have higher SCFA concentrations and higher prevalence of SCFA-producing bacteria. The authors hypothesized this could contribute to disease progression by maintaining inflammatory processes and influencing circulating immune cells, this having an impact on peripheral target organs, in this case, the liver\textsuperscript{137}. Other showed that SCFAs and secondary bile acids produced by \textit{Clostridia} can be sensed by epithelial cells and may be involved in the initiation of immunological signaling\textsuperscript{138}. Our SCFA data in AA patients, together with the increases in certain SCFA-producing microbial species we observed in our AA studies (increases in \textit{Firmicutes} phyla, such as \textit{Clostridia}) support this potential mechanism.
Additional SCFA studies may define certain metabolites involved in AA versus controls and provide a rationale to restore these SCFAs through diet or other interventions. These types of analyses and approaches can help provide novel alternatives for the treatment of AA and further elucidate the exact mechanism of gut microbes in disease pathogenesis.

5.4 Future Directions: Translational Microbiome Studies

Despite the many commonalities between AA and other autoimmune diseases, there are no studies to date on the effects of the gut microbiome and its interactions with the immune system in the development of AA. However, there is accumulating evidence in the literature that modification of the gut microbiome in AA patients undergoing fecal microbiota transplantation resolved their AA and resulted in durable hair regrowth (Figure 5.1)\(^{17,18}\). The first study reported two AA patients undergoing fecal microbiota transplantation (FMT) for the treatment of *Clostridium difficile* infection (CDI). In these subjects, the FMT not only resolved their intestinal disease, but also elicited surprising and durable hair regrowth in their pre-existing AA (Figure 5.1, top)\(^{17}\). Another recent study reported treatment of an AA patient with noninfectious diarrhea using FMT, and reported new hair growth and repigmentation on affected regions of the scalp (Figure 5.1, bottom)\(^{18}\).

![Figure 5.1. Patients treated with FMT resulted in hair regrowth\(^{17,18}\)](image)
The efficacy of FMT has been studied in other autoimmune diseases, including in clinical trials for inflammatory bowel disease (IBD) (NCT02391012, NCT02460705). In addition, some case reports have suggested improvements in Parkinson’s disease, multiple sclerosis, and idiopathic thrombocytopenic purpura after FMT. FMT can correct dysbiosis successfully by improving the ratio of certain bacterial phyla in the gut, and skews the recipient’s gut bacterial community toward that of the healthy donor. In cancer, studies have shown that resistance to immune checkpoint inhibitors (ICI) can be attributed to atypical gut microbiome composition. By utilizing FMT, the resistance to ICI can be improved by restoration of the healthy gut composition. FMT has also been shown to reconstitute the gut in ICI-associated colitis, and even increase the proportion of regulatory T cells in the colonic mucosa. Another recent study involving FMT in children with *C. difficile* infections showed distinct differences between pre-FMT and post-FMT microbiomes. After FMT treatment, gut bacterial communities similar to the healthy donors were seen, consistent with the successful establishment of a healthy microbiome and resolution of disease. To test the translational potential of these findings of microbial dysbiosis in AA patients, we initiated a clinical trial of FMT in AA (IRB AAAS4183) to repopulate the gut of AA patients with a healthy microbiome as a therapeutic alternative for AA. These studies will expand our understanding of the mechanisms underlying the development of AA and dysbiosis in the gut, and may offer new therapeutic options and improved patient care.
References


86 Biomart, E. http://.ensembl.org/biomart.


