UNIVERSITY OF CALIFORNIA, SAN DIEGO

Identification and mapping of a CNV associated with Psoriasis

A Thesis submitted in partial satisfaction of the requirements for the degree

Master of Science

in

Biology

By

Alvin Chen

Committee in charge:

Professor Benjamin Yu, Chair
Professor Colin Jamora, Co-chair
Professor Andrew Chisholm

2011

The Thesis of Alvin Chen is approved and it is accepted in quality and form for publication on microfilm and electronically:

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2011

Dedication

This is dedicated to my family.

Helen, Eric, Justin, and Brian for always providing the support and encouragement needed for my success.

Table of Contents

List of Figures and Tables

Acknowledgements

There are several people I would like to acknowledge in my research including my PI Benjamin Yu for allowing me the opportunity to join his lab and contribute to his work. I have spent the last 2 years working in his lab and have thoroughly enjoyed his enthusiasm and kindness in teaching me about science.

I would also like to than Anan Mukhopadhy for working with me on a daily basis and teaching me the techniques required for completing my research. Without Anan I would not have been able to accomplish what I have in the past 2 years.

Finally, I would like to acknowledge all the other lab members of Yu lab including, Suguna Krishnaswami, Gloria Kuo Lefkowitz, Ivy Hung, Shantanu Kumar, and Chris Cowing-Zitron, for not only helping me in my research but also making the lab an enjoyable and fun place to work.

ABSTRACT OF THE THESIS

Identification and mapping of a CNV associated with Psoriasis

By

Alvin Chen
Master of Science in Biology

University of California, San Diego, 2011
Professor Benjamin Yu, Chair
Professor Colin Jamora, Co-Chair

My project involves looking at sequence variation among individuals. Recent studies have shown that CNVs encompass a larger source of variation than SNPs, with 12% of the human genome thought to contain CNVs. CNVs are a form of structural variation, greater than 1kb, that is found in a variable number of copies in the human genome. The goal of my project is to accurately characterize and define a CNV that is implicated with hair type and disease susceptibility. We have thus far

not seen an association between CNV with hair type among the hair morphogen genes tested so far including BMP6 and WNT3 in a small sample study. Psoriasis has a known genetic component that has been linked through Genome Wide Association Studies to a region on chromosome 6p21.3. Through previous studies, this region is composed of many genes with SNP markers that have been associated with psoriasis patients. Of these genes, we have thus far analyzed BDEF4, PSORS1C1, and CDSN. Buccal swabs from 28 psoriasis and 8 control patients were collected. Utilizing qPCR analysis we have observed no significant change in copy number from BDEF4 but a reproducible 2-fold reduction of an intronic region of PSORS1C1 in psoriasis patients. A closer look at this region revealed that it also contained an exon of CDSN on the reverse strand. Furthermore, analysis on the remaining portion of exon 2 of CDSN revealed a consistent decrease among afflicted patients with this copy number variant in 19 out of 28 psoriasis patients.

I.

Introduction

Human Genetic Variation

Genetic variation is the basis for human evolution and individual biological differences among our species. Genetic variation among us can be caused by a variety of structural rearrangements mechanisms not limited to insertions, deletions, inversions, and large tandem repeats. Gene variations can have a major impact on every level of cellular processes and play a role in how our species evolves. Sequencing of the entire human genome has shown that there is approximately 5% sequence variation among us affecting over 800 independent genes (Sharp et al. 2006). Recent advancements and development of high-throughput sequencing techniques has allowed us to look at these differences in human DNA sequences ranging from single nucleotide resolutions to multiple kilobase scale comparisons. Looking at the variations among individuals will give us insight into how these genetic differences influence our unique biological differences. The effect of DNA variation can range from beneficial, neutral or deleterious. As scientists, we are interested in finding out what these structural variations in the genome lead to phenotypically. Populations and individuals vary in their susceptibility to diseases, sizes, life spans, and external characteristics. Natural variations in the skin and hair such as, hair type, skin color, numbers of moles, baldness, as well as disease susceptibility are just a few phenotypes that can be traced through our genes as a result of structural variation in our genome. By analyzing and understanding the genetic variations that occur in our genome we can better understand how we have evolved as humans and the mechanism by which our genes function.

Copy Number Variation

Structural DNA variation exists as a variety of different forms including but not limited to single nucleotide polymorphisms (SNPs), small insertions and deletions (indels), and also larger copy number variations (CNVs). These genetic variations contribute and affect the physical and biological differences among us. For instance, recent studies have identified a non-synonymous SNP in the ectodysplasin receptor in Asians that represents a derived or new allele. This polymorphism has been proposed to contribute to differences between Asian, European and African hair, which are distinguishable by diameter, cross-section and mechanical properties (Fujimoto et al. 2008). Genome wide association studies using SNP markers have successfully identified susceptibility genes for many common diseases, however this represents only a fraction of the total diseases to be uncovered. A more exciting new area of research has come in the form of larger deletions and duplications commonly referred to as copy number variants. These variants have not been studied to the extent of SNPs and many believe that they play a large role in disease formation. Recent genome wide analysis has shown that CNVs represent a larger source of variation than SNPs, 12% of the human genome was found to contain CNVs, which suggests that they could play a more important role in human differences than previously thought (Franciulli et al. 2010). A few association studies have already demonstrated the importance of CNVs as disease susceptibility variants, with specific CNVs found to confer differential risk to HIV infections, autoimmune disease, and asthma (Iuliana et al. 2008). In a famous instance of HIV susceptibility, a low copy number of CCL3L1 is associated with risk

of contracting HIV-1 (Gonzalez et al. 2005). CNVs may include deletions, insertions, and duplications typically larger than 1 kb and up to several Mb with nearly 3000 genes associated with CNVs (Kehrer-Sawatzki et al. 2007). CNVs are important to study because they often contain coding regions of expressed genes. Several published studies implicate that the coding regions of some 800-1800 known genes that are subject to variation in copy number through deletion, duplication, or tandem rearrangements (Sharp et al. 2006). Any such deletion or duplication of an exon could lead to a different mRNA transcript and subsequent aberrant protein being produced. Additionally, CNVs lying in noncoding regions may have an impact on gene expression due to the variability, gain or loss, of regulatory regions, enhancers or suppressors, which may lie within these introns. The effect that CNVs play on gene expression has be studied by analyzing the expression levels of 14,000 genes in lymphoblastoid cell lines of 210 unrelated individuals and finding an association of 17.7% of the genes with CNVs (Stranger et al. 2007). Understanding the impact that CNVs have on human phenotypes will help us to better comprehend the complexity and intricacy of genetic variation in the human genome. Thus far, only a fraction of CNV implicated genes have been analyzed and it seems that this type of variation may play a large role in disease susceptibility. We are interested in finding out the cellular processes where CNVs play a role to create biological differences among us.

Detecting CNVs

DNA typically exists as two copies in the genome, one from the mother and one from the father. Many genes however get duplicated and over time are inherited as multiple copy numbers. For instance, B-defensin genes show a common copy number between two and seven copies (Wain et al. 2009). While next-gen sequencing techniques have allowed easy detection of SNPs by single nucleotide comparison templates, reliable mapping of all CNVs has proved to be much more difficult. Since the majority of CNVs are merely duplications or deletions of existing sequence, mapping the exact locations, size, and boundaries of these CNVs on a genome wide scale is much harder. Many previous studies done on CNVs have been inexact and at low resolutions, only able to detect large CNVs greater than 50kb. The development of FISH, fluorescent in situ hybridization, allowed copy number changes to be viewed under a microscope however, the inexact boundaries and the large amount of DNA required for this experiment makes it less than ideal. Recent advancements in sequencing technologies have allowed a shift from locus-specific studies to a large scale genome-wide assessment of genetic variation using genomic clone and oligonucleotide based arrays in parallel with a reference human genome sequence providing a more comprehensive look at the prevalence of CNVs in the genome (Feuk et al. 2006). In my experiment, the primary technique to detect copy number will be utilizing Quantitative PCR, qPCR, to look at amplification curves and cycle numbers. qPCR has been shown to be a perfectly suited method for the detection of copy number variations in targeted regions due to the low screening costs and fast turn around time (D'haene et al. 2009). qPCR has also been shown to be able to distinguish a quantitative resolution as low as 1.25-fold differences in

copy number with the appropriate number of runs (Weaver et al. 2010). Since we are only looking at a cluster of target genes and have relatively low amounts of patient DNA to work with, qPCR is an ideal quantitative experimental method of choice for us.

Using Hair phenotypes to study CNVs

My research is both focused on finding the impact of CNVs on common phenotypic variations specifically hair type and curliness, and well as its impact on disease disposition to psoriasis. Several natural occurring CNVs lie in coding regions of hair shaft maintenance genes. These genes including, BMPs, WNTs, FGFs, and EDARs, play a role in differentiating stem cells in the hair follicle. For instance, EDAR has been implicated as a major genetic determinant for Asian hair thickness in East Asian populations (Fujimoto et al. 2008). The same study also implicated CUTL1, EGFR, and TGM3 as strong candidates for hair morphology-determining genes. We believe that a CNV in one more of these loci could potentially have an affect on various different hair types. We had previously constructed a hair shaft expression library and located exons of the hair shaft maintenance genes that exhibited peaks in gene expression. No research has been done implicating copy number variation as a mechanism for maintaining hair morphology.

Psoriasis and CNVs

Psoriasis is a common genetic inflammatory skin condition that results in red scaling and elevated plaques commonly on the elbows, knees, and trunk. It affects over 2% of the Caucasian population and 1-2% of the general population. Psoriasis

is primarily characterized by angiogenesis, increase of T cells in the skin, hyperproliferation of keratinocytes, and altered epidermal differentiation (Barker, Pathology of psoriasis). Studies looking at twins and familial concordance show a high incidence of genetic factors playing a role in susceptibility to psoriasis. Since psoriasis is a multifaceted disease there will be no clear-cut gene that leads to the development of psoriasis. Instead we must look at a combination of factors that could lead to increased susceptibility to psoriasis perhaps through external environmental factors. We are interested to see whether a CNV located in a particular gene has an effect on expression and protein production that can lead to an increased susceptibility to inflammatory disease. An initial study on psoriasis and copy number described an association between risk of psoriasis and increased copy number of a 300-kb genomic region that contains seven genes belonging to the B-defensin family (Hollox et al. 2007). Other studies analyzing SNP markers found, a distinct region in the HLA-C haplotype block corresponding to the gene PSORS1 was found to be associated with psoriasis in Northern European families (Helms et al. 2005). Additionally a separate study found several coding polymorphisms to be associated with psoriatic patients of Caucasian descent (Ameem et al. 2005). Cataloging of CNVs in the Database of Genomic Variants has provided a map of 1237 CNVs covering an estimated 143 Mb of genomic sequence allowing us to easily find and design primers against CNVs in target regions (Freeman et al. 2006). Studying the function and correlation of each CNV to a particular phenotype is still being explored. Currently, only basic genetic association tests have been done identifying diseases associated with CNVs. No studies have been done directly implicating CNVs

and their functional role with psoriasis. We are hoping to discover the mechanism by which CNVs act in the cell and to find the link between CNVs and its role in psoriasis onset.

We hypothesize that the CNV genetic variant, depending on the location within the genome, alters the regulation of gene expression or splicing and also, in some cases, modifies protein function. In addition, we believe that some of these variants will be linked to phenotypic characteristics in regards to hair type and psoriasis susceptibility. The approach to this experiment is first to determine the frequency and structure of CNVs in targeted genes. The use of Hapmap and Database of Genomic Variants will allow us to map the CNVs to particular affected genes and to determine whether there are dominant patterns using qPCR analysis of published CNV regions. Collection of patient hair samples and buccal swabs will allow us to collect phenotypic data on hair and psoriasis deposition to compare with their genomic data. We also hope to characterize the impact of CNVs on gene regulation and activity. Once the targeted CNVs are successfully mapped, the ones located on coding regions will be examined by cloning transcripts from this region to see if the coding region becomes altered or extended and the impact of these "duplicated" exons on protein function will be determined. And finally once the CNVs are mapped and expression is determined, we can then determine the molecular and developmental consequence of CNV on hair morphogenesis and psoriasis susceptibility. By answering these questions about the function of CNVs and their impact on the cell we can better understand the role that CNVs play in creating the individual differences among us.

II.

Results

The initial project involved finding the frequency and structure of CNVs in several hair expressed genes. The genes were initially targeted using the database for genomic variants, which showed locations of reported CNVs in the genome. Several genes were targeted because they were thought to play a role in the formation and structure of various different hair types. We wanted to characterize possible CNVs that are correlated with different ethnic phenotypic hair variations. We initially looked at 5 cell lines from 3 different populations, Asian, Caucasian, and African, to delineate a difference in copy number between the hair expressed genes. Using qPCR we wanted to more accurately map the types and locations of these CNVs within our target genes. What we found was differential copy numbers of several genes including a 2-fold copy number increase of WNT3 and BMP6, to go along with our positive control 2-fold increase in copy number of EDA and EDA2R, which are known duplications in females (Figure 2A). After obtaining the initial copy number differences among WNT3 and BMP6 across the different ethnic cell lines, we were interested in seeing if there was any correlation between hair phenotype and CNVs in these regions. Hair was collected at random from researchers around the lab and grouped into 3 different hair phenotypes; curly, straight, and wavy. The ethnicities and ages of the patients were noted to see if there was any correlation. DNA was analyzed for copy number and we were able to see an increase between 1.5 to 2-fold in BMP6 in 7 out of 15 samples (Figure 2B). The copy number increase was evenly distributed across the three hair types and no correlation was found among ethnicities. We also examined WNT3 copy number and saw a greater than 1.5-fold increase in only two samples (Figure 2B). EDARADD

was also screened as a control with little variation among the samples (Figure 2B). Extraction of hair RNA was difficult and gave extremely low yields combined with poor quality. Using several different protocols, hair RNA was unable to be purified reliably for amplification via qPCR with large variation in triplicate results, as well as failed amplification of several primers. A_{260}/A_{280} numbers confirm the poor quality of DNA obtained from hair purification. The lack of RNA for hair samples prohibited us from measuring gene expression as it correlated to copy number in the hair expressed genes.

The other aspect of my research project we decided to pursue involved looking at the CNV impact of disease susceptibility to psoriasis. Working with Caroline Piggot, MD from UCSD School of Medicine, buccal swabs were obtained from various clinically approved psoriasis patients with varying ages of onset. Clinical controls were also obtained for comparison. DNA was purified from buccal swabs resulting in low yield however relatively pure (A260/280 1.8-2.1) genomic DNA samples (Table 1). Due to the low yield of DNA purified, limited numbers of qPCR runs could be conducted. Based on previous studies, certain regions of the genome associated with psoriasis would be examined for psoriasis-linked genes. Of these, the first two to be tested were DEFB4 and PSORS1C1 (Figure 2C). While no correlation was found in DEFB4, we found a significant link between the primers designed against an intronic region of PSORS1C1 with psoriasis patients. In 10 out of 12 patients we were able to see a reproducible 2-fold copy number reduction in PSORS1C1 relative to a control sample. None of the control patients exhibited this same copy number reduction (Figure 2C).

In further analyzing the intronic region of PSORS1C1 that represented a low copy number in psoriasis patients (Chromosome 6: 31192162-31192342), we observed that this region also consisted of an exon of another gene on the reverse strand, CDSN (Figure 1). Analysis of CDSN shows that it contains 2 exons with the longer exon 2 containing the amplified region. CDSN was targeted as our gene of interest due to various linkage studies of CDSN (Sources) as well as the fact that the potential CNV was in a coding region of CDSN making it more likely to play a functional role in the cell. Additional primers were then designed for qPCR analysis flanking the original amplified region of exon 2. The regions flanking the original region demonstrated a very similar decrease in copy number in affected psoriasis patients (Figure 2D). We found that 19 out of the total 28 psoriasis patients examined were seen to have at least a 25% decrease in copy number. Additionally, 12 of the 20 psoriasis patients with this copy number decrease exhibited a greater than 2-fold decrease (50%) in copy number (Figure 2D). The qPCR runs were successfully duplicated and triplicate data showed minimal deviation providing verification of these results.

The qPCR amplification regions were run through BLAST to see if there was nonspecific binding to other areas of the genome. The regions amplified span nearly the entire portion of the coding region of exon 2 of CDSN implicating perhaps that entire copies of exon 2 of CDSN are being deleted in psoriasis patients.

To study the functional role of CDSN ex2 and to further map the specific boundaries of this CNV, we wanted to clone and sequence the region of interest to

obtain sufficient DNA quantities to work with. Primers were designed around the previous qPCR amplification regions to span the entire coding region of CDSN exon 2. To encompass the entire CNV of interest, 1.5kb and 2.4kb amplification regions were designed across exon 2 of CDSN. The idea was to observe an overall shift in the size of the band corresponding to a CNV in affected psoriasis patients with a copy number decrease. In the case that the individual is heterozygous for this gene, we would expect to see two bands. One of the same length as the wild type and an additional truncated band caused by a CNV. The segmented band would then be gel purified and used for cloning and sequencing to identify the exact boundaries of the deleted segment and to see further functional roles it may have. There was incredible difficulty in amplifying the targeted segments (1.5kb and 2.6kb) using DNA purified from buccal swabs. We increased the cycle numbers and DNA concentration using Onetaq DNA polymerase and observed 3 bands among psoriasis samples and control (Figure 3A). Two bands were ruled nonspecific due to their presence in all samples and likely caused by the high cycle numbers. In the psoriasis samples 100819A and 101129A, which showed a copy number reduction via QPCR, we observed either a faint band or complete lack of band amplification. To confirm that the lack of band was due to a deleted segment rather than failed annealing of the primers, we designed forward and reverse confirmation primers against the original amplified primers. In the two psoriasis samples 100819A and 101129A, we were able to clearly see the reverse confirmation band. However the forward confirmation bands were either absent or very faint (Figure 3B). This indicates perhaps that there is faulty annealing of the forward end of the amplified region in

these samples. This could be due to a CNV deletion in this region so further primers were designed to test the boundaries.

We attempted to amplify a larger 2.6kb fragment that encompassed all of exon 2 of CDSN. After several failed attempts using multiple polymerases at varying conditions, we conducted a nested PCR. The advantages of nested PCR are the ability to amplify low yield DNA and the specificity of the amplified product. One of our psoriasis samples with a reduced copy number, 110105A, exhibited a reduced band after nested PCR (Figure 3C). The band was subsequently extracted and cloned using pGEM-T vector plasmid. Sequencing of the suspected band revealed no alignment to the CDSN gene and the band was likely nonspecific caused by the large cycle numbers of nested PCR (1^{st} round: 30 cycles 2^{nd} round: 20 cycles). A 10-cycle reduction to the first round of the PCR amplification eliminated the nonspecific band from showing. In total, 7 psoriasis patients DNA exhibiting a copy number reduction were amplified using a 2.6kb probe and a nested 2.4kb PCR probe. None of the other samples revealed a reduced band indicating the presence of a CNV within the region (Figure 3C).

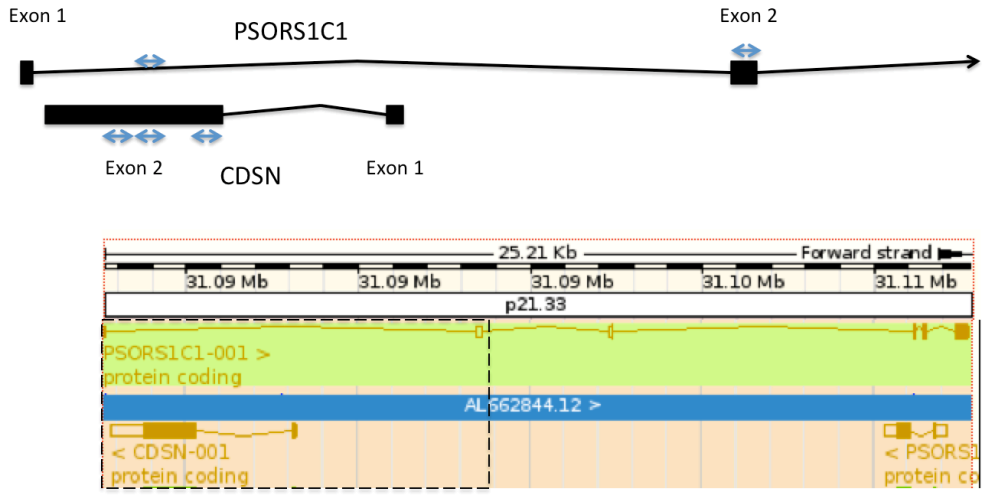


Figure 1. Gene map with qPCR probes, shown as arrows, of the region of CDSN/PSORS1C1 on Chromosome 6.

Table 1. Samples of buccal swab purified DNA from Radys Childrens Hospital showing age, sex, age of onset, psoriasis condition, and family history of psoriasis patients and control.

Sample (10ng/ul)	DOB	Age at Collection	Sex	Age of Onset	Psor	FHX
100709-A (standard)	1964	46	M		N	N
100709-B	1963	47	F		N	Y
100709-C	1988	22	F		N	N
100709-D	1985	25	F		N	N
100709-E	1983	27	F		N	N
100709-F	1969	41	F		N	N
100709-G	1980	30	F		N	N
100707-A	1992	18	M	9	P	Y
100708-A	1992	18	F	9	P	N
100726-A	1949	61	M	50	P	N
100726-B	1940	70	F	69	P	N
100803-A	2003	7	M	5	P	N
100803-B	1995	15	F	14	P	N
100723-A	1977	33	M	22	P	Y
100721-A	1978	32	F	0.3	P	N
100903-A	2006	4	F	3	P	?
100819-A	2003	7	M	4	P	Y
100601-P	1952	58	F	15	P	Y
100929-A	1997	13	F	11	P	Y
100929-B	1962	48	F	32	P	N
101022-A	1994	16	M		N	N
101027-A	1988	22	M	14	P	N
101029-A	1993	17	F	17	P	N
101108-A	1993	17	M	15	P	N
101117-A	1993	17	M	16	P	Y
101117-B	1994	16	M	15	P	Y
101119-A	1999	11	F	5	P	N
101129-A	1997	13	M	9	P	Y
101210-A	1995	15	F	10	P	N
101217-A	1993	17	M	5	P	Y
101220-A	1997	13	M	8	P	N
101223-A	1994	16	F	12	P	Y
110105-A	2000	10	M	9	P	N
110119-A	2002	8	M	2	P	UNK
110128-A	1992	18	F	12	P	N
110131-A	1996	14	F	6	P	N
110131-B	1994	16	M	0.2	P	Y
110318-A (standard)	1978	32	M		N	N

(A)

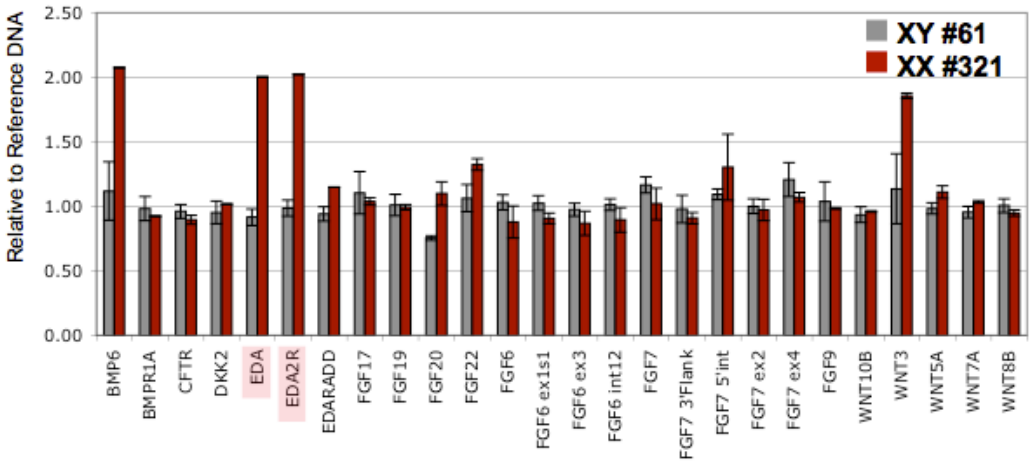
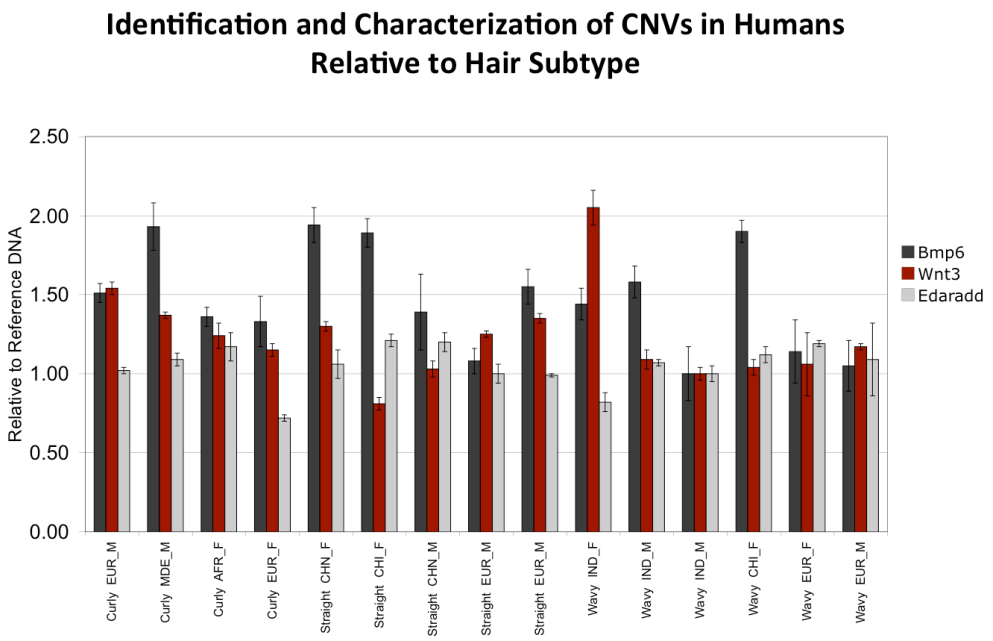


Figure 2. Copy number data acquired from ROCHE lightcycler qPCR using cycle number methods of (A) hair implicated genes in two cell lines, (B) BMP6 and WNT3 in hair purified DNA samples, (C) PSORS1C1 and DEFB4 in case controlled psoriasis patient buccal swabs, (D) CDSN exon 2 in case controlled psoriasis patient buccal swabs.

(B)



(C)

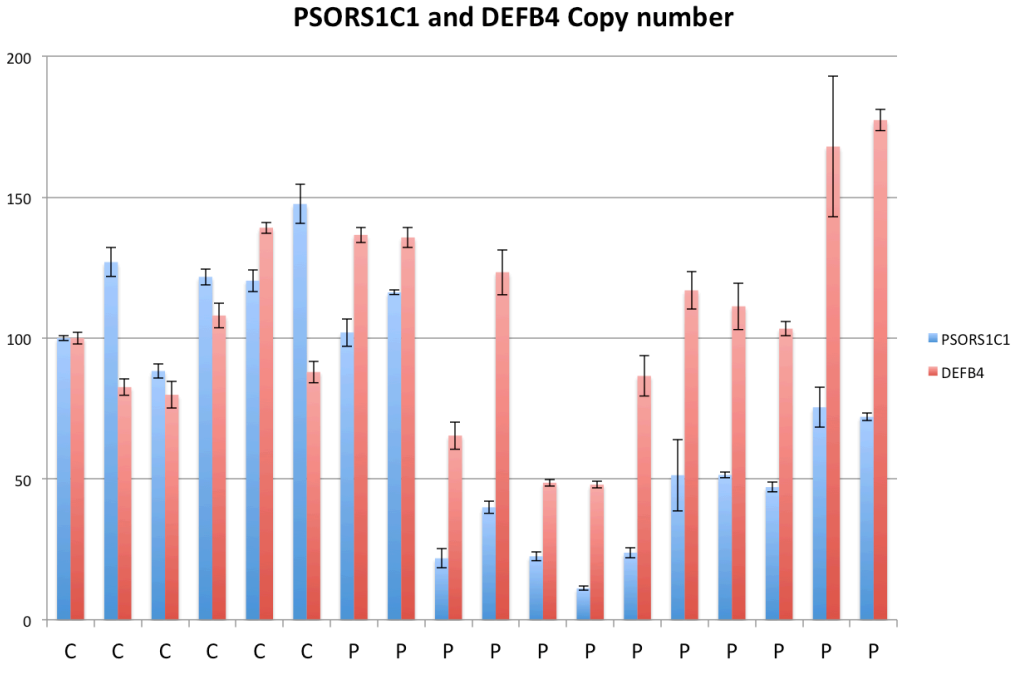


Figure 2, continued

(D)

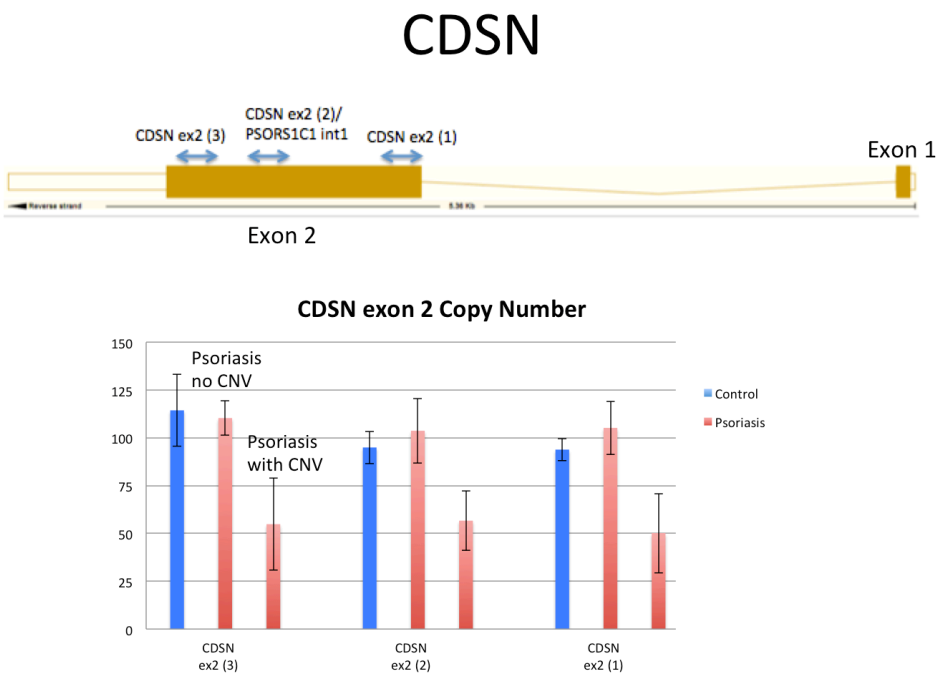


Figure 2, continued

(A)

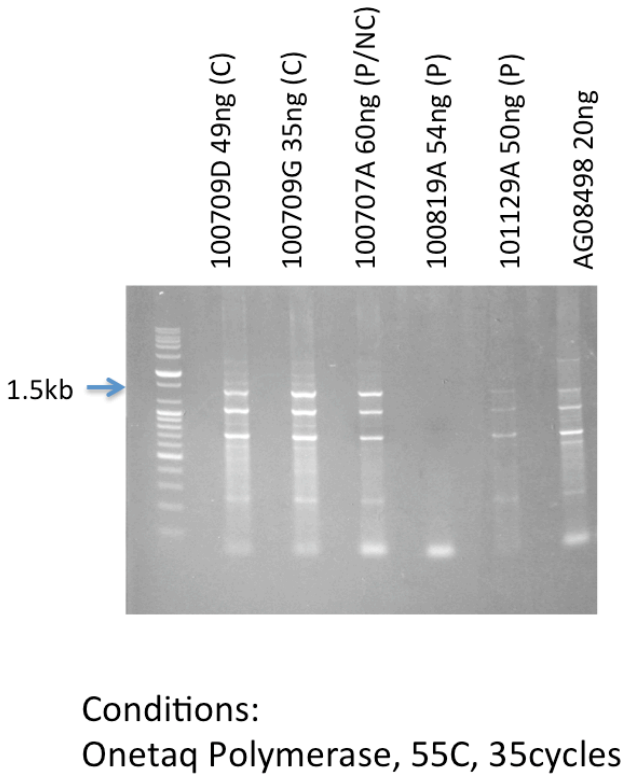


Figure 3. PCR Amplification on agarose 1.0% gels visualized under UV lamp of various probes: (A) 1.5kb CDSN exon2 Amplification, (B) Forward and Reverse confirmation sets for 1.5kb primers, (C) PCR amplification of CDSN exon2 2.6kb and nested 2.4kb probes of psoriasis buccal swab DNA and controls.

(B)

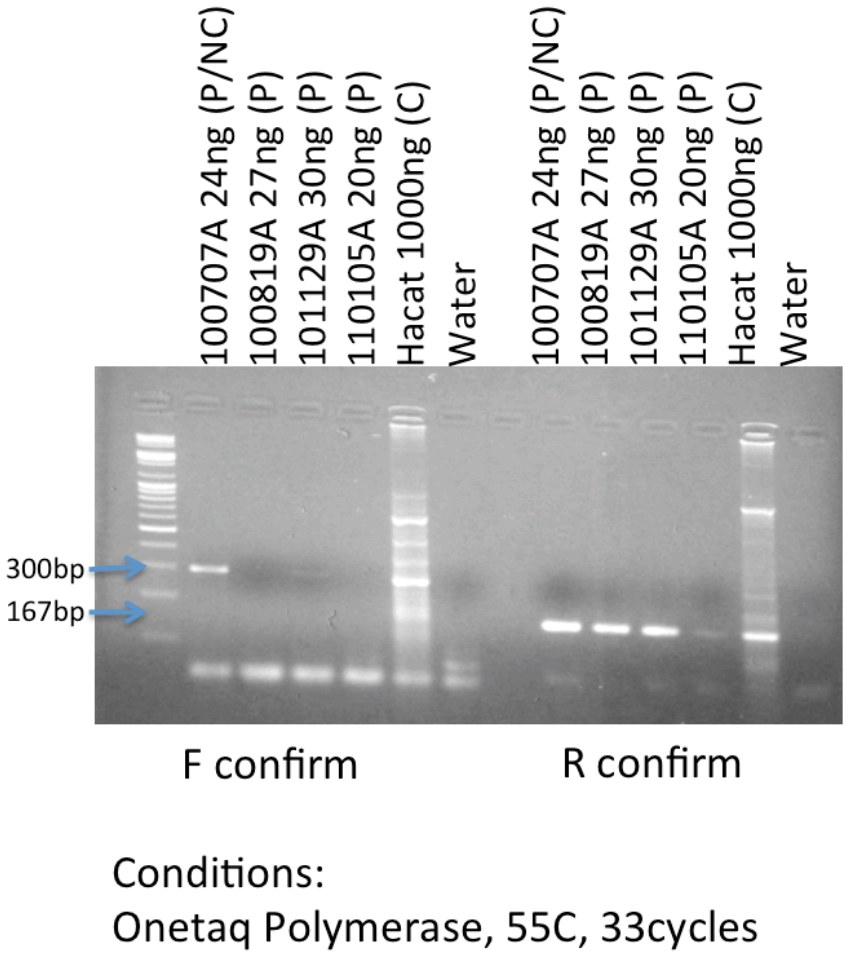


Figure 3, continued

(C)

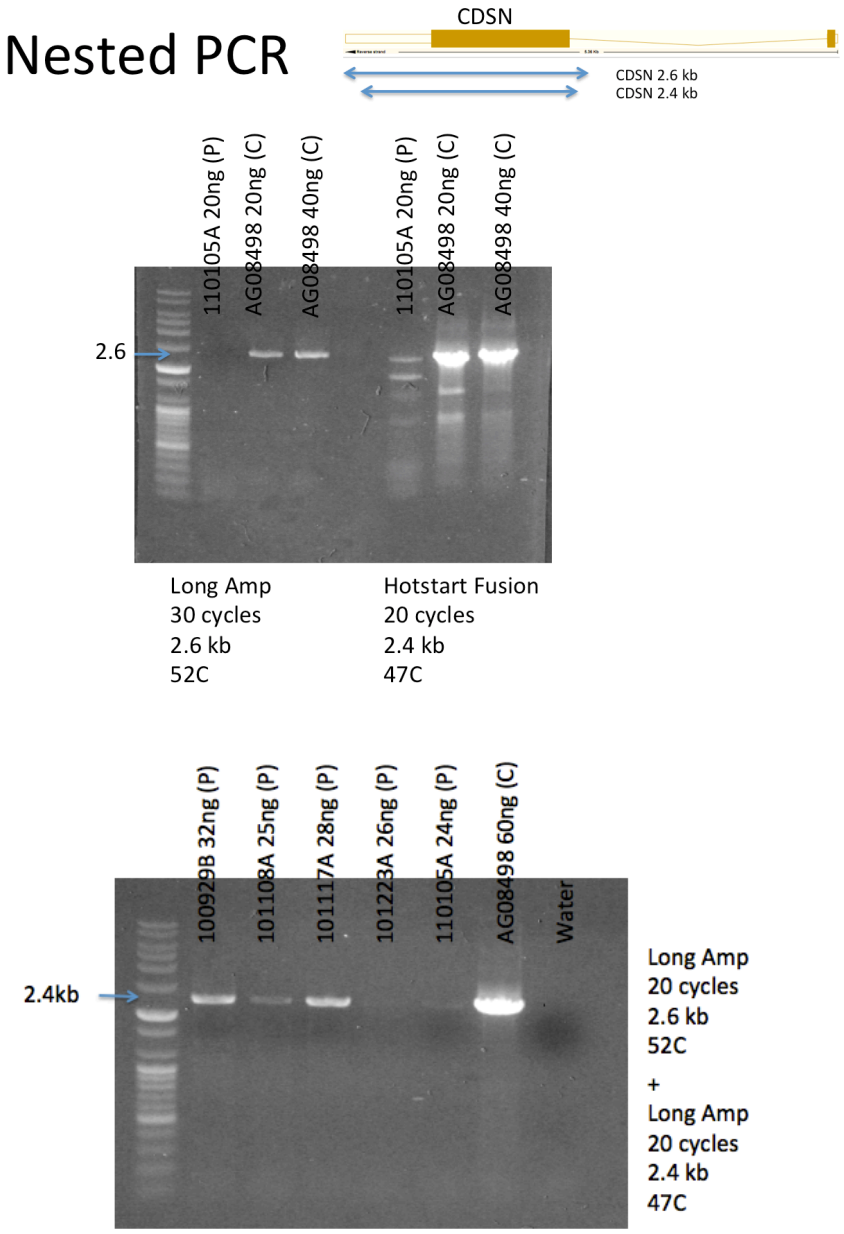


Figure 3, continued

CDSN enhancer regions

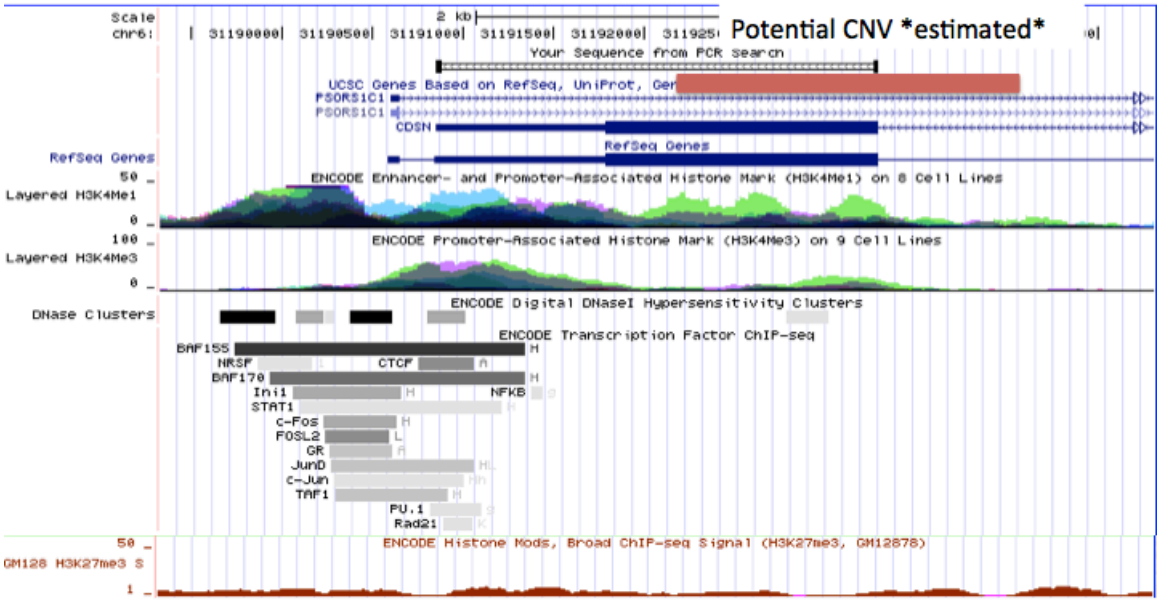


Figure 4. CNV enhancer regions of CDSN/PSORS1C1

III.

Discussion

We were initially interested in characterizing a CNV that played a role in a phenotypic outcome by analyzing both natural variations, in the form of hair type, and disease susceptibility, in the form of psoriasis. Looking at two genes, WNT3 and BMP6, that showed differential copy number and were implicated in hair shaft structure and maintenance, we were unable to find a correlation between hair type and CNVs. This study would be improved by increasing the sample size as well as overcoming technical issues with purifying RNA resulting in gene expression data.

Psoriasis provided us with a genetic disease component to identify a potential CNV. Previous studies had implicated a region of DEFB4 copy number as being associated with CNVs (Hollox et al. 2007). We were unable to reproduce the data they had seen in the one probe we had tested. Other association studies (Helms et al. 2005) had shown SNP linkage analysis of various regions likely to be associated with psoriasis. Based on this information, we identified one region in PSORS1C1 as another candidate for CNV study in psoriasis patients. Upon finding an initial correlation with psoriasis and a region of PSORS1C1, we examined the area further and found the coding region of the gene CDSN on the reverse strand of this region. To determine the boundaries of this CNV we attempted to amplify this region of DNA in hopes of seeing a reduced band size indicating the presence of a deletion CNV. Since we did not know how large the CNV would be, we attempted several amplifications with probes of varying sizes 1.5kb, and 2.6kb. We did not observe a legitimate reduced band of our CNV in any of the psoriasis patients or controls using these probes. A likely possibility for these results would be that the CNV we are interested in is larger than the 2.6kb region we tested for. A smaller

band would only be visible if the two amplification primers fully encompassed the boundaries of the CNV. Another possibility of not seeing a reduced band is that the CNV deletion occupies a segment of one of the amplification primers preventing annealing of one of the primers. We believe this to be a possibility as well due to failed binding of a subset of confirmation primers we had designed in two psoriasis patients with the CNV.

We have extracted DNA from 28 psoriasis patients and have observed a reproducible copy number reduction in a region spanning CDSN exon 2 in 19 out of 28 of the total samples. Additionally, we have not seen the same decrease in copy number in the 8 control patients as well as 4 separate fibroblast cell lines. Based on multiple qPCR probes spanning the entire exonic coding region as well as amplification data designed across the entire exon, we believe the boundaries of this CNV to be larger than 2.6kb. The functional effect that this CNV has on gene expression is not yet known due to the inability of extracting RNA from buccal swabs. The CNV we have identified, due the location in the genome, likely plays a role in susceptibility to psoriasis. We believe that understanding the functional role of CDSN is important in the susceptibility of psoriasis. Cornedesmosin has been shown to play a role in reinforcing cell-cell cohesion within the upper epidermis and stratum cornuem. A peeling skin disease has been characterized by the loss of CDSN and is implicated with a defect in epidermal barrier integrity and strong predisposition to atopic diseases (Oji et al. 2010). A separate study looked at adult CDSN knockout mice and observed a very similar persistent barrier defect without the inflammatory response seen in psoriasis. While several studies have shown the

increase of CDSN expression in multiple layers of stratum corneum in psoriasis lesions, this may be caused by psoriasis associated SNPs participating in development of hyperproliferation as a compensatory mechanism to barrier impairment (Leclerc et al. 2009). It is possible that the observed reduction in CDSN is one of many factors that play a role in susceptibility to psoriasis.

Mechanistically, CNVs can have an effect on several different regions contributing to an alteration of gene expression. We propose 3 general possibilities by which this CNV may play a role in susceptibility to psoriasis.

A coding region CNV

A CNV lying in a coding region may play a role affecting gene dosage where increased copies of a dosage sensitive gene would lead to increased transcription and ultimately lead to increased protein levels and vice versa. It may be that decreased copies of CDSN ex2 leads to a dosage dependent reduction of CDSN expression which has been shown to play a role in skin barrier disruption and epidermis integrity in mice (Leclerc et al. 2009). CDSN is an extracellular protein that integrates into desmosomes before cornification and forms an intercellular structure responsible for corneocyte cohesion and transformation of desmosome into corneodesmosome (Orru et al. 2005). Furthermore, a peeling skin disorder in humans has been found to be associated with loss of the CDSN gene caused by a premature stop codon SNP (Oji et al. 2010). The functional role that CDSN plays in psoriasis has not yet been studied and we will not know the exact affect that CDSN reduction has until more studies have been done. While CDSN presents a very

attractive biological candidate, there are several caveats to proposing that CDSN reduction causes psoriasis. Firstly, in the case of the peeling skin syndrome characterized before, heterozygotes carrying the mutation do not show any symptoms. Homozygotes with the affected mutation have a skin disease that is distinct from psoriasis. Finally, studies have shown that psoriatic lesions have increased CDSN expression due to a reduction in mRNA degradation of corneodesmosomal proteins (Simon et al. 2008). It was shown that the peeling skin phenotype is due to a complete loss of CDSN and we believe that psoriasis acts on an alternative pathway to this.

A regulatory region CNV

A CNV lying in a regulatory region may play role in transcription of the gene by altering binding sites of promoters, enhancers, transcription factors, and chromatin structure often times leading to reduced or altered expression. Using data from UCSC genome browser, we see that the proposed deletion consists of many enhancer and transcription factor binding sites in the local area. An alteration to one of these locations could lead to inability for binding by one or more of these factors. It has been shown that CNVs which lie in regulatory regions can play a role in expression of genes due to deletion of enhancer regions in particular with SOX9 in regards to disorders of sex development (White et al. 2011). It has also been shown that a CNV localized to a specific area has a proportional affect on expression of the surrounding genes. We have shown that the CNV consists minimally of the genes CDSN and PSORS1C1, therefore we hypothesize that this region may play a

regulatory role in the expression of various downstream genes impacting susceptibility to psoriasis.

An alteration to slice sites causing aberrant protein

It is alternatively possible that the CNV deletion on exon 2 of CDSN could cause an alteration to the slice sites of the intron-exon junction creating the formation of a slice variant. This splice variant may in turn create an alternate mRNA transcript leading to the production of an aberrant protein. An aberrant protein may create deleterious effects on the cell leading to susceptibility of psoriasis. Functional studies on the CNV using model cell lines as well as assays for their mRNA transcripts would have to be done to confirm these hypotheses.

Future Work

Thus far we have characterized a potential CNV that is associated with just over two thirds of the tested psoriasis patients. We are further interested in the functional impact of this CNV and its prevalence in psoriasis patients. We would greatly benefit from gene expression data as well as an increased sample size. Future psoriasis samples would ideally ask for blood collection to provide adequate DNA and RNA collection for analysis. Hopes to amplify and clone the entire CNV would also benefit from the increased DNA yield. To fully map the boundaries of the CNV, additional qPCR and amplification probes would need to be created and tested. While we were able to obtain one psoriasis cell line, it did not exhibit the same copy number reduction preventing its use as a model cell line. Locating a model cell line would help to characterize the CNV further by providing renewable resources for

genetic analysis. We are also further interested in the mRNA transcripts produced by the CNV and seeing if there is any abnormalities created by splice variants. Research on CNVs is an increasingly popular field of study. Every year, more and more papers are showing the impact that CNVs have on diseases and susceptibility to diseases. My hopes in this research is to gain a better understanding into how our genetic differences influence human variations and ultimately how CNVs play a role in making us different as individuals.

IV.

Materials and Methods

DNA and RNA extraction from fibroblast cell lines

DNA and RNA were extracted from 5 fibroblast cell lines. The 5 cell lines were obtained from my PI, Ben Yu, and included an Asian male (AG08498), two Africans (GM05757, GM03529), and two Caucasians (AG012261C, GM0321B). DNA extraction technique involved suspending the cells in 1mL TE buffer, 25 µl 20% SDS, and 10 µl of 10mg/mL Proteinase K. The cells were hybridized overnight at 45° C. Subsequent washes were performed with Phenol:Chlorophorm (25:1) and Chlorophorm:IAA (24:1) to remove impurities in the sample. DNA was precipitated with ethanol (100%), 3M NaOAC pH5.2, and 3 µl glycogen overnight at -20°C. DNA concentrations were fairly good ranging from 120ng/µl to 400 ng/µl with adequate A_{260}/A_{280} between the 1.6-2.1 range.

RNA was extracted using standard protocol from Zymo Research RNA mini-prep kit.

Primer Sequences

Primers were designed using sequence data from UCSC genome browser and Ensembl genome browser. Primers were created by both primer 3 and IDT primerquest software. The following are the primer sequences used in my research:

BMP6Fw	CCCTCTTCATGCTGGATCTG
BMP6Rev	AGAAGGCTCTTGCGGTTGA
BMPR1AFw	GACAGAATCTGGATAGTATGCTTCA
BMPR1ARev	TAATAGCATCATCTGGACAGTGC
DKK2Fw	GCCTACCCTTGTAGCAGTGA
DKK2Rev	ATTATTGCAGCGGGTACTGG
EDAFw	GTGGAGCGCAGGGAACTC

EDARev	AGGTAGCAGCACAACGTCAG
EDA2RFw	TATGGAGAGGGTGGAGATGC
EDA2RRev	AAACAGTCCCCACAGACAGC
EDARADDFw	CTTGTTCCTCCTGCTTGCTC
EDARADDRev	GGAGCAAGAACTCCAAGGTG
FGF17Fw	ACCAGTACGTGAGGGACCAG
FGF17Rev	CAAACTTGTTGCCGTCCTC
FGF19Fw	GTGGTGGTCCACGTATGGAT
FGF19Rev	GCAGGAAGCAGCTGGAGAG
FGF20Fw	GTGGGTTCGCATTTCCTGT
FGF20Rev	AGGATCTGCAGGTGGAAGC
FGF22Fw	CGACTCTACACCGTGGACTG
FGF22Rev	GAAGTGGGCGGACAGGTG
FGF6Fw	AACAACACGCTGCTGGACTC
FGF6Rev	AGCACCTGGAGGTGAAAGC
FGF7Fw	TTCTGGAAAACCATTACAACACA
FGF7Rev	TTGCCATAGGAAGAAAGTGG
FGF9Fw	TGGGAACTATTTCGGTGTGC
FGF9Rev	CGCCTGAGAATCCCCTTTA
WNT10BFw	AGATTCTGGGCCTGAAGTTG
WNT10BRev	AGCTGGTGCTGACACTCGT
WNT3Fw	GGGCCAGCAGTACACATCTC
WNT3Rev	AACTGGTGCTGGCACTCC
WNT5AFw	CCAACTGGCAGGACTTTCTC
WNT5ARev	CCTGCCAAAAACAGAGGTGT
WNT5BFw	GTGCAGAGACCCGAGATGTT
WNT5BRev	GCATGACTCTCCCAAAGACAG
WNT7AFw	GCGCAAGCATCATCTGTAAC
WNT7ARev	CTCTCCCAGTGCAGAGCAGT
WNT8BFw	GAAGTACCACGCAGCACTCA
WNT8BRev	CAGCCCTAGCGTTTTGTTCT
BMP6int1(1)-F	CTGAAGTGGGTGAGCAGACA
BMP6int1(1)-R	CCCATGTTCTCTCCCTCAAA
BMP6int1(2)-F	CTCAGAGTCGTGGAGCCTTC
BMP6int1(2)-R	CTCTCCTCCCGGCTACTCTT
BMP6ex2-F	GTCAGCGACACCACAAAGAG
BMP6ex2-R	TGTGCTGATGCTCCTGTAAGA
KRT33A/B -F	CACCAGGGCCTCATACTGAT
KRT33A/B -R	AGCTGCAATGCAATCCTCTT
RAB3C -F	TCAGCTCTTGTGGGAGGTCT
RAB3C -R	CTCCATGGGCTCTGACAAAT
EDARADD -F	CCTCCTGGTTGTCAGGATGT
EDARADD -R	GGTCTTGAATGGCCTAACCA

MYO9A -F	GAGGTGGGAGGATCACTTCA
MYO9A -R	TCCTCCTCCTCCTCCTCTTC
SMOC2 -F	T C G G A G G A G G T G T C A T T T T C
SMOC2 -R	T G C A G T G A G A G C C T T G T G T C
FSTL1 -F	GACCAGGGCAAAGTGGTCTA
FSTL1 -R	GCTACAAGGATGTGCAGCAA
KRTAP9-2 -F	CTGGTTCATGCCATTCTCCT
KRTAP9-2 -R	AATAATAGCCGGGTGCAGTG
TYR -F	TTGTTTTTCCAGGGAGTTGG
TYR -R	CTTGCTAGGGGCTAATGCAG
COL23 -F	AGGACCCGAAGGCTTTATCTA
COL23 -R	GGTCAGGCCTGTGTTCCTAA
KRT33A/B -F	TGCAGTGGGCCATTGTATTA
KRT33A/B -R	CTTTGCCTTCCACCATGATT
COL4A2 -F	C A C A C T G C T C T G T C C T T C C A
COL4A2 -R	A C T G T G T G T C T C C A G C A A C G
CUX1 -F	GCTTTGGCCACTGATATGGT
CUX1 -R	GGTGAAGGAGAAGCAAGCAC
KRTAP9-2 -F	ACATTCTCGTCCTGGAATCG
KRTAP9-2 -R	CACCTGTGCCTTTCCGTTAT
COL23 -F	GGACCCGAAGGCTTTATCTA
COL23 -R	GGTCAGGCCTGTGTTCCTAA
Psors1c1-3 -F	CTAGAACTGCTGGGGACTCG
Psors1c1-3 -R	CCCATCTCTGAGGGCAAATA
DefB4 -F	TTTGGTGGTATAGGCGATCC
DefB4 -R	GAGACCACAGGTGCCAATTT
psors1c1 ex2 -F	AAGGGTGGAATCTACAGTCCGTGA
psors1c1 ex2 -R	TGTGCTTTGCCCACTGGCAT
Psors1c1 ex1-ex2 -F	TCCCAGAAACCCAGGAAATCGAGA
Psors1c1 ex1-ex2 -R	CTGTTGCATTTCAAAGCCACTGGG
Psors1c1 ex1-int1 -F	TCCCAGAAACCCAGGAAATCGAGA
Psors1c1 ex1-int1 -R	TTTCCCTCCTCTCTCATTGCTCCA
Psors1c1 int1 -F	AATGCTCTTAGCCAAGGTCCCTGT
Psors1c1 int1 -R	TTTGCCCACCCTTCCCAAACAAAG
Psors1c1 int1-ex2 -F	TCCTTTGTTCCTCTCTGCTTGGCT
Psors1c1 int1-ex2 -R	ACCAGGTGAAGAGTGTGGATGACT
Psors1c1 ex3 dna -F	TCAGCCTGGGAAGAATTGGTTTGC
Psors1c1 ex3 dna -R	TCACAGACGAGGAACCAACACTGA
Psors1c1ex1-3 rna-F	TCAAATAGACTCTGCAGCCAGCCA
Psors1c1ex1-3rna-R	GCAAACCAATTCTTCCCAGGCTGT
cdsn ex2 -F	CTTGGCTAAGAGCATTGGCACCTT
cdsn ex2 -R	AGAGCTTCTGGCACTGGAAATGGA
Psors1c1 int1-2 -F	GAAGTATTTGCCCTCAGAGATGGG
Psors1c1 int1-2 -R	ACAAATCCTATGGTGGCTACGAGG

Psors1c1 ex3-5 -F	TGGGAAGAATTGGTTTGCAGCCAG
Psors1c1 ex3-5 -R	AGTTTCCTTGGAGCTGGGATCTGT
Psors1c1 ex2-3 -F	TCAAATAGACTCTGCAGCCAGCCA
Psors1c1 ex2-3 -R	CACAAGTGCCAAAGTCAAGACTGG
Psors1c1 ex2 -F	AGACGATCAAGGGTGGAATCTACAG
Psors1c1 ex2 -R	TGTGCTTTGCCCACTGGCAT
Psors1c1 ex1-2 -F	CCAGAAACCCAGGAAATCGAGACTCA
Psors1c1 ex1-2 -R	CTGTTGCATTTCAAAGCCACTGGG
CDSN 1.4kb -F	A A G A G C A T T G G C A C C T T C T C
CDSN 1.4kb -R	G T A A C T C T C C T T G G G G T A G G A A A
CDSN 2.4kb -F	A C T G C T T T C C T C C A C A G G G
CDSN 2.4kb -R	T C C T C C T C T G T G G G A G C A
CDSN 1kb -F	C C T T G G C T A A G A G C A T T G G
CDSN 1kb -R	C A A A G G A A G G G A C C C C T G
CDSN 2.4(2) -F	G G A C C T T G G C T A A G A G C A T T G
CDSN 2.4(2) -R	TCAACAAATATTTATTGTCTTCCTCCT
CDSN ex2 (1) -F	TGGCTAAGAGCATTGGCACCTTCT
CDSN ex2 (1) -R	ACCAGAGCTTCTGGCACTGGAAAT
CDSN ex2 (3) -F	TCTTCTGGTCACCCTTGCATGTCT
CDSN ex2 (3) -R	AGAGGCTTCACTTGGGCTAGGATA
CDSN 1.4 (2) -F	ATTGGCACCTTCTCAGACCCTTGT
CDSN 1.4 (2) -R	AGAGGCTTCACTTGGGCTAGGATA
CDSN 1.3 nest -F	AAGGGTGACTCCAGCGGTTTCAGTA
CDSN 1.3 nest -R	AGAGGCTTCACTTGGGCTAGGATA
CDSN 1.4 F confor -R	AACCGGATGCACCTTGTAGACTAGAG
CDSN 2.6 -F	ATCCTTTAGAGACAATAGCCAGTTCTTCC
CDSN 2.6 -R	ACAAATATTTATTGTCTTCCTCCTCTGTGG

qPCR

Quantitative PCR done by ROCHE 480 light cycler was used to measure copy number of DNA sequences. Master mix was created using Sybr Green and appropriate primer. 10µl reactions were done in triplicates. Several housekeeping genes were used including ACTIN, CFTR, GAPDH, and YWHZ. Typically two housekeeping genes were used as control in each run to ensure there was no large variability in cycle numbers. Input DNA concentration was 50ng/µl for fibroblast

DNA and ranges between 4ng/µl and 40ng/µl for buccal swab DNA due to low DNA quantity. Verification of successful amplification was done by viewing melting peaks and standard amplification curves of the samples. Water controls were conducted to ensure no contamination of water as well as to rule out nonspecific binding of the primers.

Hair RNA and DNA extraction

Hair DNA and RNA was purified using techniques found in other papers and through collaboration with Gloria Leftkowiz, a graduate student in the same lab. Between 10 and 15 strands of hair cut from the base of the head were used in each sample. Hair was initially washed twice with ethanol (100%) and water. Zirconia beads were added for breakdown of hair. Hair was treated with Gough Buffer, 1M DTT and heated for 1 hour at 55° C with persistent vortexing. DNA purification required the addition of Proteinase K. Gough buffer consisted of 7M Urea, 1% SDS, 0.35M NaCl, 10mM EDTA, and 10mM Tris HCl. After sufficient digestion of hair, standard phenol chlorophorm washes followed by precipitation in ethanol (100%), 7M NaOAC, and glycerol were done. DNA was treated with RNase while RNA was treated with DNase I for removal of cross contaminants.

Buccal swab DNA purification from Psoriasis patients and control

Buccal swab samples were obtained from Caroline Piggot, a UCSD medical student, who collected samples from 30-40 patients so far. The psoriasis patients were recruited from clinics with any type of psoriasis as diagnosed by an attending dermatologist. Control samples were also collected from patients with no personal

or family history of psoriasis. DNA was purified using Qiagen QIamp DNA Mini and DNA Blood kit. Spin protocol was followed for DNA purification from Buccal swabs. Data from buccal swabs is shown in Table 1.

PCR Amplification

The primers CDSN 1, CDSN 1.5, CDSN 2.4, and CDSN 2.6 were used to attempt to amplify the segment of interest. Typically between 20-50 ng of DNA was used as template. Various different polymerase enzymes were used including: Hot start fusion, Phusion DNA polymerase, One taq DNA polymerase, Long amp DNA polymerase. Successful conditions including template DNA concentration, cycle number, annealing temperature, and polymerase are shown in the captions of figure 3.

Cloning and Sequencing of Amplified Fragment

Cloning of DNA was done using the amplification products of Long amp DNA polymerase. 1.8kb and 2.4kb target bands were gel extracted using Bioneer gel extraction kit. After purification, DNA insert was ligated to pGEM-t easy vector overnight. DH5α competent cells were used for transformation of ligated vector. Transformed product was grown overnight in 37°C on ampicillin/X-gal/IPTG plates. Blue/white colony selection followed by inoculation and Bioneer plasmid mini prep yielded sufficient DNA for sequencing. Eton Bio sequencing was analyzed using Sequencher 4.10.1 software. Sequenced inserts were aligned to known CDSN exon2 gene sequence showing 80-90% match for the positive control.

References

Sharp, AJ., Cheng, Z., Eichler, EE. (2006). Structural variation of the human genome. *Annu Rev Genomics Hum Genet*. 7:407-442.

Freeman, JL., Perry, GH., Feuk, L., Redon R et al. (2006). Copy number variation: new insights in genome diversity. *Genome Res.* 16(8):949-61.

Stranger, BE., Forrest, MS., Dunning, M., Ingle, CE. at al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 315(5813):848-53.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J. et al. (2004). Large-scale copy number polymorphism in the human genome. *Science*. 305(5683):525-8.

Franbough, A., Hallegot, P., Baltenneck, F., Toutain, C et al. (2003). Current research on ethnic hair. *J Am. Acad. Dermatology*. 48, S115-S119.

Fujimoto, A., Kimura, R., Ohashi, J., Omi, K. et al. (2008). A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum Mol Genet*. 17(6):835-43. Epub.

Hollox, EJ., Huffmeier, U., Zeeuwen, PL., et al. (2007). Psoriasis is associated with increased beta-defensin genomic copy number. *Nat Genet*. Jan;40(1):23-5.

Iuliana, I., Angela, R., Christoph, L., Benjamin, R., Charles, L. (2008). Genetic association analysis of copy number variation in Human Disease Pathogenesis. Genomics. 93(1), 22-26.

Kehrer-Sawatzki, H. (2007). What a difference copy number variation makes. *Bioessays*. 4,311-313.

Speleman, F., Kumps, C., Buysse, K., Poppe, B., Menten, B., De Preter, K. (2008). Copy Number alterations and copy number variation in cancer: close encounters of the bad kind. *Cytogenet Genome Res*. 123, 176-182.

Weaver, S., Dube, S., Mir, A., Qin, J, Gang, S., Ramakrishnan, R., Jones, R., Livak, K. (2010). Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. *Methods* 50(4),271-276.

Leszek, W. (2003). Human Hair: A unique physiochemical composite. *Journal of the American Academy of Dermatology*. 48(6), S106-S114.

Fanciulli, M., Petretto, E., Aitman, TJ. (2010). Gene copy number variation and common human disease. *Clin Genet*. 77:201-213.

Feuk, L., Marshall, C., Wintle, R., Scherer, S. (2006). Structural variants: changing the landscape of chromosomes and design of disease studies. *Hu Mol Genet*. 15:R57-R66.

Orru, S., Giuressi, E., Carcassi, C., Mirella, C., Licinio, C. (2005). Mapping of the Major Psoriasis-Susceptibility Locus (PSORS1) in a 70-Kb Interval around the Corneodesmosin Gene (CDSN). *Am. J. Hum. Genet*. 76:164-171.

Capon, S., Toal, I., Evans, J., Allen, M., Patel, S., Tillman, et al. (2003). Haplotype analysis of distantly related populations implicates corneodesmosin in psoriasis susceptibility. *J Med Genet*. 40:447-452.

Simon, M.,Tazi-Ahnimi, R., Jonca, N., Caubet, C., Cork, M.J., Serre, G. (2008). Alterations in the desquamation-related proteolytic cleavage of corneodesmosin and other corneodesmosomal proteins in psoriatic lesional epidermis. *British Journal of Dermatology*. 159:77-85.

Israeli, S., at al. Inflammatory Peeling Skin Syndrome Caused by a Mutation in CDSN Encoding Corneodesmosin. (2011). *Journal of Investigative Dermatology*. 131:779-781.

Matsumoto, M., Zhou, Y., Matsuo, S., Nakanishi, H., et al. (2008). Targeted deletion of the murine corneodesmosin gene delineates its essential role in skin and hair physiology. *PNAS*. 105:6720-6724.

Wain, L., Armour, J., Tobin, M. (2009). Genomic copy number variation, human health, and disease. *Lancet*. 374:340-350.

Weaver, S., Dube, S., Mir, A., Qin, J., Sun, G., et al. (2010). Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. *Methods*. 50:271-276.

D'haene, B., Vandosompele, J., Hellemans, J. (2009). Accurate and objective copy number profiling using real-time quantitative PCR. *Methods.* 50:262-270.

Gonzalez, E. et al. (2005). The Influence of *CCL3L1* Gene-Containing Segmental Duplications on HIV/AIDS Susceptibility. Science. 307:1434-1440.

Mills, R. et al. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature*. 470:59-65.

Henrichsen, C., Chaignat, E., Reymond, A. (2009). Copy number variants, diseases and gene expression. *Human Molecular Genetics*. 18:R1-R8.

Henrichsen, C., Vinkenbosch, N., Zollner, S., et al. (2009). Segmental copy number variation shapes tissue transcriptomes. *Nature genetics*. 41:424-429.

Gallinaro, H., Jonca, N., Langbein, L., Vincent, C., Simon, M., Serre, G., Guerrin, M. (2004). A 4.2 kb Upstream Region of the Human Corneodesmosin Gene Directs Site-Specific Expression in Hair Follicles and Hyperkeratotic Epidermis of Transgenic Mice. *J Invest Derma*. 122:730-738.

Oji, V., Ecki, M., Aufenvenne, K., et al. (2010). Loss of Corneodesmosin Leads to Severe Skin Barrier Defect, Pruritus, and Atopy: Unraveling the Peeling Skin Disease. American Society of Human Genetics. 87:274-281.

Leclerc, E., Huchenq, A., Mattluzzo, N., Metzger, D., et al. (2009). Corneodesmosin gene ablation induces lethal skinbarrier disruption and hair-follicle degeneration related to desmosome dysfunction. *Journal of Cell Science*. 122:2699-2709.

Helms, C., Saccone, N., Cao, L., Daw, J., et al. (2005). Localization of PSORS1 to a haplotype block harboring HLA-C and distinct from corneodesmosin and HCR. *Hum Genet*. 118:466-476.

Ameen, M., Allen, M., Fisher, S., Lewis, C., Cuthbert, A., et al. (2005). Corneodesmosin (CDSN) gene association with psoriasis vulgaris in Caucasian but not in Japanese populations. *Clinical and Experimental Dermatology*. 30:414-418.