Published in final edited form as:
Hum Genet. 2014 December ; 133(12): 1513–1523. doi:10.1007/s00439-014-1490-9.

Admixture mapping identifies a locus at 15q21.2-22.3 associated with keloid formation in African Americans

Digna R Velez Edwards[1,2,3,4,*], **Krystal Tsosie**[3,5,*], **Scott M. Williams**[6], **Todd L. Edwards**[1,2,3,5], and **Shirley B. Russell**[3,7]

[1]Vanderbilt Epidemiology Center, Vanderbilt University, Nashville, Tennessee

[2]Institute for Medicine and Public Health, Vanderbilt University, Nashville, Tennessee

[3]Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee

[4]Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN

[5]Division of Epidemiology, Department of Medicine, Vanderbilt University, Nashville, TN

[6]Department of Genetics, Geisel School of Medicine, Dartmouth University, Hanover, NH

[7]Division of Dermatology, Department of Medicine, Vanderbilt University, Nashville, TN

Abstract

Keloids are benign dermal tumors that occur ~20-times more often in African versus Caucasian descent individuals. While most keloids occur sporadically, a genetic predisposition is supported by both familial aggregation of some keloids and the large differences in risk among populations. Yet, no well-established genetic risk factors for keloids have been identified. In this study we conducted admixture mapping and whole exome association using 478 African Americans (AAs) samples (122 cases, 356 controls) with exome genotyping data to identify regions where local ancestry associated with keloid risk. Logistic regression was used to evaluate associations under admixture peaks. A significant mapping peak was observed on chr15q21.2-22.3. This peak included *NEDD4*, a gene previously implicated in a keloid genome-wide association study (GWAS) of Japanese individuals later validated in a Chinese cohort. While we observed modest evidence for association with *NEDD4*, a more significant association was observed at (myosin 1E) *MYO1E*. A genome-scan not including the 15q21-22 region also identified associations at *MYO7A* (p rs35641839, odds ratio [OR]=4.71, 95% confidence interval [CI] 2.38–9.32, $p=8.34x10^{-6}$) at 11q13.5. The identification of SNPs in two myosin genes strongly associated with keloid formation suggests that an altered cytoskeleton contributes to the enhanced migratory and invasive properties of keloid fibroblasts. Our findings support the admixture mapping approach for the

CORRESPONDING AUTHORS: Digna R. Velez Edwards, Ph.D., M.S., Vanderbilt Epidemiology Center, 2525 West End Ave., Suite 600 6th Floor, Nashville, TN 37203, Tel: 615-322-1288, Fax: 615-322-8291, digna.r.velez.edwards@vanderbilt.edu. OR Todd L. Edwards, Ph.D., M.S., Vanderbilt Epidemiology Center, 2525 West End Ave., Suite 600 6th Floor, Nashville, TN 37203, Tel: 615-322-3652, Fax: 615-322-8291, todd.l.edwards@vanderbilt.edu.
*Co-first author equal contribution

Its contents are solely the responsibility of the authors and do not necessarily represent official views of the National Center for Advancing Translational Sciences or the NIH.

CONFLICT OF INTEREST STATEMENT
None Declared

study of keloid risk, and indicate potentially common genetic elements on chr15q21.2-22.3 in causation of keloids in AAs, Japanese, and Chinese populations.

Keywords

keloids; admixture mapping; fibrosis; genetic epidemiology; ancestry

INTRODUCTION

Keloids are benign fibrotic tumors of the dermis that form during a prolonged wound healing process. (Marneros *et al.*, 2004; Niessen *et al.*, 1999) They are characterized by expansion beyond the bounds of the original wound, invading surrounding normal skin. (references cited in(Bond *et al.*, 2011)) Keloids have been estimated to occur in ~1/30 African Americans (AAs) and ~1/625 of the overall US population. (Barrett, 1973) Keloid formation is one of a group of fibroproliferative diseases characterized by an exaggerated response to injury that occur at higher frequency or with more severe manifestations in people of African descent. (Smith *et al.*, 2008) These diseases include hypertension, nephrosclerosis, scleroderma, sarcoidosis, allergic disease, and uterine fibroma. We and others have suggested that a common etiopathology may operate in these diseases, and that common genetic factors may account for their unusual racial distribution. (references cited in(Smith *et al.*, 2008)) The key alteration(s) responsible for the pathological processes resulting in keloid formation has not been identified, and there is no satisfactory treatment for this disorder.

While most cases of keloid formation occur sporadically, familial forms have been observed. The hypothesis of a genetic component for keloid formation is supported by the occurrence of keloids at different frequencies in different ancestral populations, regardless of current location. Although some cases of keloid formation may be due to somatic mutation,(Saed *et al.*, 1998) multiple keloids in the same individual and evidence for a multicellular origin of keloids argue against somatic mutation as the primary event in most cases. (Chevray and Manson, 2004; Moulton-Levy *et al.*, 1984; Trupin *et al.*, 1977) There have been several studies attempting to identify the genetic basis of keloids. Linkage studies have been performed in both AA and Japanese. The AA studies identified loci at 14q22-q23, (Davis *et al.*, 2000) and 7p11 (1), whereas linkage to 2q23 was found in Japanese. (Marneros *et al.*, 2004) None of these studies, however, was able to localize the linkage signal. A recent genome-wide association study (GWAS) in a Japanese population identified four single nucleotide polymorphism (SNP) loci in three chromosomal regions (1q41, 3q22.3-23 and 15q21.3) that showed significant association with keloid formation. (Nakashima *et al.*, 2010) The results from this study were independently replicated in a Chinese population. (Zhu *et al.*, 2013) From the limited data available in the above studies and in pedigrees supporting autosomal dominant inheritance with reduced penetrance, genetic complexity is indicated for keloids, consistent with contributions of multiple susceptibility loci, each of which that may provide a different genetic model of disease risk.

The approximate 20-fold increase observed for keloid formation in AA relative to European American (EA) in the US indicates that an alternative approach to identifying keloid loci, admixture mapping, is likely to yield informative results. While admixture mapping has been used in studies of several disorders that occur more frequently in individuals of African ancestry, including asthma,(Flores *et al.*, 2012) hypertension,(Zhu *et al.*, 2005) and nondiabetic end-stage kidney disease (ESKD),(Winkler *et al.*, 2010) it is potentially even more powerful for mapping keloid loci, given the greater prevalence differential between AAs and EAs. Therefore, in the present study we used admixture mapping together with whole exome association analysis of keloids in an AA population. The goals of this study are to identify the genetic determinants of keloid formation in an AA population.

MATERIALS AND METHODS

Study populations

This study used AA keloid cases and controls currently available from the BioVU DNA Repository (or BioVU), as well as samples from previously collected keloid fibroblasts and lymphocytes for both admixture mapping and whole exome association analyses,. (Russell *et al.*, 2010; Smith *et al.*, 2008) The study included DNA from 81 AA keloid cases and 399 controls from the BioVU DNA Repository (BioVU, 2007 – present), Vanderbilt University, Nashville, TN. BioVU was designed to link clinical data available from de-identified electronic medical records (EMR) to DNA specimens. The BioVU DNA Repository consists of de-identified blood samples obtained from patients at the Vanderbilt University Medical Center Hospital, including all clinics that are part of the hospital system. (Pulley *et al.*, 2010) De-identified data from multiple sources are available within BioVU, including diagnostic and procedure codes (International Classification of Diseases 9th edition [ICD-9] and Current Procedural Terminology [CPT]), basic demographics, discharge summaries, nursing notes, progress notes, health history, multi-disciplinary assessments, laboratory values, echocardiogram diagnoses, imaging reports, electronically derived data, and inpatient medication orders. BioVU keloid samples were defined as AAs 18 years or older who were diagnosed with keloids in the EMR and who have at least two mentions of a keloid diagnosis in their record, (either two diagnostic codes [ICD 9 = 701.4] or a code and mention of a keloid within their record). Controls were AA subjects 18 years and older who have had surgical procedures performed at the Vanderbilt University Medical Center Hospital that involved an open wound, such as breast surgery, cesarean section, open heart surgery, etc, and have two years of follow-up in the EMR with no evidence of keloid formation. Excluded from controls are individuals with either a diagnosis of keloids (ICD9 = 701.4) or other fibroproliferative disorders such as asthma (ICD9 = 493.*), nephrosclerosis (ICD9 = 403.90), or fibroids (ICD9 = 68.2, 68.29, 218.*, 654.1*; CPT = 58140). Also excluded from controls are individuals with the key words “excessive scarring” in their clinical record. While individuals with keloids are not more susceptible to these conditions than the general population, common susceptibility variants may be present. Initial validation analyses via manual review of records by DRVE and SBR demonstrated that the phenotyping algorithms were able to accurately capture cases and controls. A detailed description of the human subjects protection applied to BioVU is described by Pulley et al (Pulley *et al.*, 2010).

Keloid cases (n=36) were also obtained from a keloid biospecimen repository (15–53 years) that includes collection of cultured fibroblasts from normal and keloid scar tissue. (5,17) The diagnosis was made both by the surgeon or dermatologist removing the tissue and by the pathologist who examined the tissue. The principal criterion used to differentiate keloid from other hypertrophic scars was the extent to which the scar exceeded the boundary of the initiating wound. The majority of specimens met three additional criteria that help to differentiate keloids from other hypertrophic scars: (1) the patient has multiple keloids; (2) tumors have recurred following surgical removal or other treatment; (3) they do not spontaneously regress over long periods of time. DNA was also obtained from another repository of blood samples of unrelated individuals (n=15) who were part of multiplex families. These studies were approved by the Vanderbilt University Institutional Review Board, and qualified as exempt non-human subjects research.

DNA Extraction and Genotyping

All DNA samples were isolated from whole blood or from fibroblasts using the Autopure LS system (QIAGEN Inc., Valencia, CA).

Genomic DNA was quantitated via an ND-8000 spectrophotometer and DNA quality was evaluated via gel electrophoresis. We genotyped DNA from the 494 participants using the custom Affymetrix Axiom Exome Genotyping Array (Affymetrix Inc., Santa Clara, CA). The genomic DNA samples were processed according to standard Affymetrix procedures for processing of the assay. The data were processed for genotype calling using the Affymetrix Power Tools software (APT, Affymetrix Inc., Santa Clara, CA).

Microarray Data

We used previously published microarray data to evaluate associations located under the admixture mapping peak. (Smith *et al.*, 2008) Methods describing isolation and propagation of fibroblasts from keloids and normal scars and microarray analysis approaches have been previously described. (Russell *et al.*, 1978; Smith *et al.*, 2008)

Whole Exome Genotyping Study Quality Control

Data on 319,283 SNPs and 494 individuals were available prior to implementation of quality control. Six individuals were removed after excluding individuals with low genotyping efficiency (<95%). Three subjects were removed due to inconsistencies between reported sex and genetically estimated sex. Five individuals were removed after kinship estimates identified related individuals using identity-by-descent sharing from a random selection of 100,000 autosomal SNPs. When a pair of related individuals was identified, only one member (parent or sibling) of the family was included, with priority given to parent over offspring. 478 unrelated women remained in the final dataset.

All SNPs were tested for deviation from Hardy-Weinberg equilibrium (HWE) using PLINK software (Purcell et al., 2007a). We excluded SNPs with HWE $P \leq 1.0\text{x}10^{-6}$, dropping 498 SNPs. Markers with low genotyping efficiency (<95%) were also excluded, 5,216 SNPs. Twenty duplicate SNPs, 146 non-autosomal SNPs and 149,790 monomorphic SNPs were removed. Quality control procedures are outlined in Supplemental Figure 1. After removing

subjects and SNPs for quality control, 478 subjects (122 cases and 356 controls) and 163,613 SNPs remained for analyses. Due to our sample size we limited association analyses to SNPs with an MAF ≥ 0.05 (47,599 SNPs).

In order to assess population stratification among AAs samples EIGENSTRAT software(Price *et al.*, 2006) was used to estimate continuous axes of ancestry. The top ten principle components (PCs) were extracted and used as covariates in regression models to test for ancestry and genotype associations (Supplemental Figure 2).

Statistical Analyses

Descriptive statistics of demographic data were expressed as means and standard deviations for continuous covariates and as frequencies and proportions for categorical data, and compared between cases and controls with logistic regression using STATA 11.0 statistical software (College Station, TX).

Admixture mapping was performed using whole exome genotyping data. Admixture mapping has been shown to be valid approach to identify novel gene-regions of interest in studies of several complex diseases. (Anum *et al.*, 2009; Cooper *et al.*, 2008; Rosenberg *et al.*, 2010; Zhu *et al.*, 2005) Admixture mapping were performed using Local Ancestry in Admixed Populations – Ancestry (LAMP-ANC), which uses the ancestral frequencies to infer the ancestral origins of each observed allele for every SNP. For our analyses we used African ancestry and European ancestry allele frequency estimates from the 1000 Genomes reference populations (September 2013 release). LAMP-ANC was used to identify unlinked markers using an r^2=0.10 regardless of allele frequency, resulting in 36,344 SNPs available for analyses. LAMP computes the structure of ancestry using overlapping windows of contiguous SNPs and combines the results with estimates of ancestry for a region. (Sankararaman *et al.*, 2008) Admixture mapping association analyses were performed in PLINK using logistic regression with variables coding for 0, 1, or 2 European ancestry chromosomes inferred from SNP genotypes by LAMP-ANC, adjusting for the top 10 PCs and sex. The significance threshold of for the admixture mapping analyses was calculated from n=1000 permutations. Permutation testing was conducted by permuting the phenotype to obtain an empirical p-value.

Association analyses were conducted on genotyped SNPs for the global whole exome analysis and for genotyped and imputed SNPs within the genomic region with the strongest evidence from admixture mapping. Common variants, minor allele frequency (MAF) ≥ 0.05, were analyzed for single SNP association with keloid risk in AAs using logistic regression adjusted for sex and 10 PCs using PLINK software. (Purcell et al., 2007b) Analyses were performed assuming an additive genotypic model. We performed a Bonferroni corrections adjusting for the total number of markers that passed QC (0.05/47,599 SNPs = 1.05×10^{-6}).

We used IMPUTE software to impute ungenotyped SNPs in the chromosomal region where we observed the strongest evidence from admixture mapping (chr15q21.2-22.3), using the entire panel from 1000 genomes reference population (build 37, 2013). (Howie *et al.*, 2009) Studies have shown that using the entire reference panel increases imputation accuracy. (Howie *et al.*, 2011) SNPTEST statistical software was used to test imputed SNPs for

association with keloids adjusting for sex and 10 PCs. (Marchini *et al.*, 2007) Analyses were performed assuming an additive genotypic model.

We also evaluated the strongest associated SNPs through conditional analyses, where we tested for single SNP association at each SNP again adjusting for top PCs, sex, and the top associated index SNP within the region. We also performed secondary analyses of SNP association, accounting for admixture mapping evidence using the approach described by Zhu et al. (Zhu et al., 2011) We further analyzed the Spearman rho correlation coefficients between the SNPs evaluated through conditional analyses, and then report r^2 between SNPs evaluated.

RESULTS

478 AA subjects were evaluated by admixture mapping (122 cases and 356 controls) by Local Ancestry in Admixed Populations – Ancestry (LAMP-ANC), which uses ancestral frequencies to infer ancestral origins of each observed allele for every SNP. The majority of subjects were female (70% cases and 62% controls)(Table 1). The mean age of study participants was 43(standard deviation [SD] = 17) for cases and 52 (SD = 18) controls. AA keloid cases on average had a significantly lower proportion European ancestry (19%) than controls (21%) (p= 0.011).

Admixture Mapping and Single SNP Associations Under Identified Regions

The most significant admixture mapping peak was observed on chromosome 15 (q21.2 to 22.3); therefore we focused our analyses of SNP associations to this region (Figure 1). The most significant p values for SNPs within this region from the admixture mapping analyses (association of SNPs with differences in percentage European ancestry across keloid cases and controls) were for myosin 1E (*MYO1E*, rs140447165), cyclin B2 (*CCNB2*, rs28383563), glucosaminyl (N-acetyl) transferase 3, mucin type (*GCNT3*, rs139614046), and BCL2/adenovirus E1B 19kDa interacting protein 2 (*BNIP2*, rs141152343) all of which had a p value of 1.13×10^{-4} (Supplemental Table 1). Also of note, a SNP in neural precursor cell expressed, developmentally down-regulated 4, E2 unbiquitin protein ligase (*NEDD4*) a gene previously implicated in the prior keloid GWAS within a Japanese population,(Nakashima *et al.*, 2010) had an admixture mapping analysis p value of 2.73×10^{-4}.

We further evaluated SNPs within one $-\log_{10}$ (p) of the smallest p value from admixture mapping analyses (p-value limit = 1.13×10^{-3}) using single SNP association with keloid status. To more comprehensively evaluate the region we also imputed ungenotyped SNPs within chr15q21.2-22.3 using IMPUTE2 software. (Howie *et al.*, 2009) Genes located under the admixture mapping peak along with the $-\log_{10}$ (P) from the single SNP association analyses within this region are shown in Figure 2. Although we did see evidence of association in *NEDD4* (smallest p = 3×10^{-5}), along with other genes associated with fibroproliferative disorders in prior studies, including aldehyde dehydrogenase 1 family, member A2 (*ALDH1A2*, smallest p = 9×10^{-3}), aquaporin 9 (*AQP9*, smallest p = 3×10^{-3}), ring finger protein 111 (*RNF111*, smallest p = 8×10^{-3}), *CCNB2* (smallest p = 0.015), RAR-related orphan receptor A (*RORA*, smallest p = 8×10^{-3}) and A disintegrin and metalloproteinase 10 (*ADAM10*, smallest p = 8×10^{-3}) (Figure 2, Supplemental Table 2),

they were not among our most significant associations within the region. The most significant single SNP associations were seen, as in the admixture analysis, at several SNPs in *MYO1E* (strongest associated SNP rs747722, odds ratio [OR] = 4.41, 95% confidence interval [CI] 2.29 to 8.50, $p = 9.07x10^{-6}$, Table 2). As seen in Figure 3, there is good agreement on the involvement of genes at 15q21.2-22.3 from both admixture mapping and single SNP association within the region.

Global Whole Exome Association

In addition to admixture mapping and tests of single SNP associations with imputation under the 15q peak, we conducted agnostic tests of association of all genotyped SNPs. Strong associations were shown at multiple SNPs in the myosin VIIA (*MYO7A)* gene located on chromosome 11q13.5, most significant association at rs35641839 (OR = 4.71, 95% CI 2.38 to 9.32, $p = 8.34x10^{-6}$) and one at chromosome 4q31.21 SNP in the gene inositol polyphosphate-4-phosphatase, type II, 105kDa (*INPP4B*, rs2636675, OR = 3.23, 95% CI 1.89 to 5.54, $p = 1.97x10^{-5}$) (Table 3). Two of the top associated SNPs in *MYO7A* are missense mutations (rs35641839 and rs111033183) and one is a synonymous mutation (rs78871677).

Conditional Analyses

Conditional analyses were performed conditioning on the index SNP (most statistically significant association) within a region (Table 4). We performed conditional analyses for top associated SNPs within chr15q21.2-22.3 with the *MYO1E* SNP rs747722 as the index SNP and for the global scan using the *MYO7A* SNP rs35641839 as the index SNP. These analyses show that the associations within *MYO1E* are due to rs747722; however, there are independent association signals for *NEDD4* and the other intergenic SNPs within chr15q21.2-22.3. Conditional analyses of *MYO7A* SNPs showed that the signal for *MYO7A* SNPs was due to rs35641839. A summary of the r^2 between the SNPs evaluated in conditional analyses is provided in Supplemental Table 3.

In a secondary analyses conditioning admixture mapping analyses on the strongest associated SNP from association analyses (rs747722 none of the SNPs in Supplemental Table 1 were statistically significant ($P<0.05$), suggesting that the association between rs747722 and keloids accounts for the evidence observed in admixture mapping analyses.

DISCUSSION

We observed strong evidence of local ancestry associating with keloid risk on chr15q21.2 to 22.3, a region that includes a gene previously associated in a keloid GWAS of Japanese. (Nakashima *et al.*, 2010) Other GWAS have implicated this chromosomal locus with three other fibroproliferative diseases, hypertension in AAs,(Adeyemo *et al.*, 2009) asthma in a European population,(Moffatt *et al.*, 2010) and atherosclerosis in Finland. (Inouye *et al.*, 2012) A summary of SNPs implicated with fibroproliferative disorders within this genomic region is provided in Table 5. Significant associations were seen in *NEDD4* in keloids, *ALDH2* in hypertension, *RORA* in asthma, and *AQP9, CCNB2, RNF11*, and *MYO1E* in atherosclerosis.

Examination of data from an earlier microarray study of fibroblasts from keloids and normal scars that included some of the subjects evaluated in the admixture mapping and exome association analyses, revealed altered expression of several genes under the admixture mapping peak on chr15q21.2 to 22.3 (Table 6), including increased expression for *MYO1E* and *ADAM10* and decreased expression for *TCF12 a*nd *CCNB2* in keloid versus normal scar fibroblasts. (Russell *et al.*, 1978; Smith *et al.*, 2008) Of interest was the observation of increased expression of *ADAM10* only in cultures grown with hydrocortisone, a condition we have shown to downregulate a subset of fibrosis related genes in normal but not keloid fibroblasts.

The most significant association with keloid formation at 15q22.2 was within the *MYO1E* gene. *MYO1E* is a widely expressed nonmuscle membrane-associated class I myosin with a motor-head domain that binds ATP and F-actin, a calmodulin-binding neck domain, and a long tail consisting of a plasma membrane binding domain, a proline rich TH2 domain and a C-terminal SH3 domain. (Mele *et al.*, 2011) Mutations in this gene have been identified as the cause of some cases of focal segmental glomerulosclerosis, a genetically heterogeneous fibroproliferative disease. (Butler *et al.*, 2008) It is involved in multiple actin-dependent processes at the cell membrane including cell migration, cell adhesion and clathrin-mediated endocytosis. (Bond *et al.*, 2011; Cheng *et al.*, 2012) Recently *MYO1E* has been shown to be a crucial component of the invadosome, a specialized structure involved in matrix degradation and invasion that may have implications for wound healing. These findings suggest that mutations in the *MYO1E* gene are excellent candidates for a role in the increased migration and invasive properties of keloid fibroblasts. (Bond *et al.*, 2011; Smith *et al.*, 2008) Our previous observation of increased expression of *MYO1E* in keloid fibroblasts, further supports a functional role for this gene. (Smith *et al.*, 2008) Our findings that several genes, including *MYO1E*, *CCNB2* and *ADAM10* were significantly associated with keloid formation in three different analyses (admixture, exome association, and gene expression) strengthen the possibility that one or more of these are keloid susceptibility loci. We note that *CCNB2* and *MYO1E* are within 11 kb from each other.

The most significant associations we observed across the genome, excluding 15q21.2-22.3, were from SNPs in another unconventional myosin gene, *MYO7A at 11q13.5*, two of which were missense mutations. The gene region where the associated SNPs are located is highly conserved across vertebrate species. The SNPs that associated with keloids in *MYO7A* are only variable in African ancestry individuals according to HapMap reference populations (YRI and ASW) and are located in a regulatory region of the *MYO7A* gene. The presence of a risk increasing allele only existing in the African populations is consistent with variants at this locus contributing to the observed ancestral prevalence disparity and provides further reason to argue that this is not a false positive finding, especially since this gene was not found in the admixture analyses. Defects in the *MY07A* gene have associated with the mouse shaker-1 phenotype and human skin pigmentation. (Fernandez *et al.*, 2009) Furthermore, the top associated SNPs (rs35641839) and nearby SNPs have been associated with Usher Syndrome, hearing loss, and retinitis pigmentosa. (Liu et al., 1998; Roux et al., 2006) How these conditions may be related to keloids is unclear but the associations with several biologically related traits support the possibility of common biological mechanisms.

The GWAS of keloids in a Japanese population(Nakashima *et al.*, 2010) found significant associations of four loci in three chromosomal regions, 1q41, 3q22.3-23 and 15q21.3 with keloid formation. The association observed at 15q21.3 was in *NEDD4*, rs8032158 (OR 1.51, 95% CI 1.35 to 1.69, p $5.96x10^{-13}$). It is of note that within the Japanese population of the HapMap *NEDD4* rs8032158 is in a large LD block (99kb) and is in very strong LD with multiple SNPs within the gene, in contrast to African Americans who have much weaker LD in this region. The results from this study were independently replicated in a Chinese population. (Zhu *et al.*, 2013) Although we did observe some evidence of association ($p < 0.05$) within this gene, they were not among our most statistically significant associations (Supplemental Table 1). It may be that the association signal observed at *NEDD4* in Asians is due to SNPs in the gene being in linkage disequilibrium (LD) with the causal variant(s) elsewhere in the region and not actually being in *NEDD4.* Alternatively, NEDD4 causative variants may not be captured well by the exome chip platform. It may also be that the effect of *NEDD4* is strongest among the Japanese population based on other interacting factors.

Although our study identified several SNPs associated with keloid risk we were limited by our sample size; we were not powered to identify associations with smaller effect sizes or associations with SNPs with low allele frequencies. As a result we were unable to use several of the SNPs included on the exome chip. Use of exome chip data also resulted in fewer markers being included in admixture mapping analyses compared to admixture mapping with GWAS data. Furthermore, although we were able to adjust for sex as a covariate we were unable to adjust for a larger set of potential confounders for the statistical analyses due to poor capture of those variables across all samples. We also note that although our EMR control algorithm was validated after chart review, it is possible that physicians may underreport keloids. However, if keloids were underreported this would only serve to decrease our power and result in weaker risk estimates for our associations, implying that the reported associations are likely stronger than presented. Finally, although we detected statistically significant associations at multiple chromosomal regions (Tables 2 and Table 3), none of the associations passed the Bonferroni multiple testing correction ($p<1.05x10^{-6}$). We note, there is no established threshold for multiple testing for whole exome analyses. We expect this threshold was conservative due to the lack of independence of tests due to LD between multiple SNPs. Therefore, it is possible that some of our associations would meet genome-wide significance if we accounted for the actual number of independent tests performed and not the conventional threshold. However, we emphasize the need to replicate our study findings.

In summary, we evaluated whole exome genotyping data for evidence of local ancestry associating with risk of keloids and observed strong evidence of association in several genes, including from gene expression data. The finding that African ancestry at 15q21.2-22.3 was strongly associated with keloids is the first evidence of this type in AAs and points toward common elements in the genetics of keloids in African and Asian populations. GWAS and other studies implicating chromosomal region 15q21.2-22.3 in multiple fibroproliferative diseases with identification of different associating genes accentuates the importance of the region but also the need for additional studies to validate findings and to identify causal variants. Lastly, the identification of SNPs in two different

myosin genes that are strongly associated with keloid formation suggests that an altered cytoskeleton contributes to the enhanced migratory and invasive properties of keloid fibroblasts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grants F33AR052241 (SBR), P50DE10595 (SBR), RO3AR048938 (SMW), P30AR041943 (SMW), P20GM103534 (SMW), pilot grant from the Center for Human Genetics Research at VU (SBR, DRVE and TLE), 1UL1RR024975 (SBR), VICTR for TLE, 5K12HD04383-12 (DRVE) and by resources of Vanderbilt University School of Medicine and University of North Carolina Functional Genomics Core. The project described was also supported by CTSA award No. UL1TR000445 from the National Center for Advancing Translational Sciences.

Reference List

Adeyemo A, Gerry N, Chen G, Herbert A, Doumatey A, Huang H, Zhou J, Lashley K, Chen Y, Christman M, Rotimi C. A genome-wide association study of hypertension and blood pressure in African Americans. PLoS Genet. 2009; 5:e1000564. [PubMed: 19609347]

Anum EA, Springel EH, Shriver MD, Strauss JF III. Genetic contributions to disparities in preterm birth. Pediatr Res. 2009; 65:1–9. [PubMed: 18787421]

Barrett. Keloid. In: Bergsma, D., editor. Birth Defext Compendium. Williams and Wilkens Company; Baltimore: 1973.

Bond JE, Bergeron A, Thurlow P, Selim MA, Bowers EV, Kuang A, Levinson H. Angiotensin-II mediates nonmuscle myosin II activation and expression and contributes to human keloid disease progression. Mol Med. 2011; 17:1196–1203. [PubMed: 21792479]

Butler PD, Longaker MT, Yang GP. Current progress in keloid research and treatment. J Am Coll Surg. 2008; 206:731–741. [PubMed: 18387480]

Cheng J, Grassart A, Drubin DG. Myosin 1E coordinates actin assembly and cargo trafficking during clathrin-mediated endocytosis. Mol Biol Cell. 2012; 23:2891–2904. [PubMed: 22675027]

Chevray PM, Manson PN. Keloid scars are formed by polyclonal fibroblasts. Ann Plast Surg. 2004; 52:605–608. [PubMed: 15166997]

Cooper RS, Tayo B, Zhu X. Genome-wide association studies: implications for multiethnic samples. Hum Mol Genet. 2008; 17:R151–R155. [PubMed: 18852204]

Davis, Garcia M.; Phillips, J., III; Hedges, L.; Haines, J.; Carneal, J., et al. Detection of a critical interval for a familial keloid locus on chromosome 14q22-q23 in an African-American pedigree. 67. 2000. p. A21

Fernandez LP, Milne RL, Pita G, Floristan U, Sendagorta E, Feito M, Aviles JA, Martin-Gonzalez M, Lazaro P, Benitez J, Ribas G. Pigmentation-related genes and their implication in malignant melanoma susceptibility. Exp Dermatol. 2009; 18:634–642. [PubMed: 19320733]

Flores C, Ma SF, Pino-Yanes M, Wade MS, Perez-Mendez L, Kittles RA, Wang D, Papaiahgari S, Ford JG, Kumar R, Garcia JG. African ancestry is associated with asthma risk in African Americans. PLoS One. 2012; 7:e26807. [PubMed: 22235241]

Howie, Marchini J.; Stephens, M. Genotype Imputation with Thousands of Genomes. 1. 2011. p. 457-470.

Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5:e1000529. [PubMed: 19543373]

Inouye M, Ripatti S, Kettunen J, Lyytikainen LP, Oksala N, Laurila PP, Kangas AJ, Soininen P, Savolainen MJ, Viikari J, Kahonen M, Perola M, Salomaa V, Raitakari O, Lehtimaki T, Taskinen MR, Jarvelin MR, Ala-Korpela M, Palotie A, de Bakker PI. Novel Loci for metabolic networks

and multi-tissue expression studies reveal genes for atherosclerosis. PLoS Genet. 2012; 8:e1002907. [PubMed: 22916037]

Liu XZ, Hope C, Walsh J, Newton V, Ke XM, Liang CY, Xu LR, Zhou JM, Trump D, Steel KP, Bundey S, Brown SD. Mutations in the myosin VIIA gene cause a wide phenotypic spectrum, including atypical Usher syndrome. Am J Hum Genet. 1998; 63:909–912. [PubMed: 9718356]

Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007; 39:906–913. [PubMed: 17572673]

Marneros AG, Norris JE, Watanabe S, Reichenberger E, Olsen BR. Genome scans provide evidence for keloid susceptibility loci on chromosomes 2q23 and 7p11. J Invest Dermatol. 2004; 122:1126–1132. [PubMed: 15140214]

Mele C, Iatropoulos P, Donadelli R, Calabria A, Maranta R, Cassis P, Buelli S, Tomasoni S, Piras R, Krendel M, Bettoni S, Morigi M, Delledonne M, Pecoraro C, Abbate I, Capobianchi MR, Hildebrandt F, Otto E, Schaefer F, Macciardi F, Ozaltin F, Emre S, Ibsirlioglu T, Benigni A, Remuzzi G, Noris M. MYO1E mutations and childhood familial focal segmental glomerulosclerosis. N Engl J Med. 2011; 365:295–306. [PubMed: 21756023]

Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von ME, Farrall M, Lathrop M, Cookson WO. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. 2010; 363:1211–1221. [PubMed: 20860503]

Moulton-Levy P, Jackson CE, Levy HG, Fialkow PJ. Multiple cell origin of traumatically induced keloids. J Am Acad Dermatol. 1984; 10:986–988. [PubMed: 6736343]

Nakashima M, Chung S, Takahashi A, Kamatani N, Kawaguchi T, Tsunoda T, Hosono N, Kubo M, Nakamura Y, Zembutsu H. A genome-wide association study identifies four susceptibility loci for keloid in the Japanese population. Nat Genet. 2010; 42:768–771. [PubMed: 20711176]

Niessen FB, Spauwen PH, Schalkwijk J, Kon M. On the nature of hypertrophic scars and keloids: a review. Plast Reconstr Surg. 1999; 104:1435–1458. [PubMed: 10513931]

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]

Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR. Principles of human subjects protections applied in an opt-out, de-identified biobank. Clin Transl Sci. 2010; 3:42–48. [PubMed: 20443953]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007a; 81:559–575. [PubMed: 17701901]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007b; 81:559–575. [PubMed: 17701901]

Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. Nat Rev Genet. 2010; 11:356–366. [PubMed: 20395969]

Roux AF, Faugere V, Le GS, Pallares-Ruiz N, Vielle A, Chambert S, Marlin S, Hamel C, Gilbert B, Malcolm S, Claustres M. Survey of the frequency of USH1 gene mutations in a cohort of Usher patients shows the importance of cadherin 23 and protocadherin 15 genes and establishes a detection rate of above 90%. J Med Genet. 2006; 43:763–768. [PubMed: 16679490]

Russell JD, Russell SB, Trupin KM. Differential effects of hydrocortisone on both growth and collagen metabolism of human fibroblasts from normal and keloid tissue. J Cell Physiol. 1978; 97:221–229. [PubMed: 701387]

Russell SB, Russell JD, Trupin KM, Gayden AE, Opalenik SR, Nanney LB, Broquist AH, Raju L, Williams SM. Epigenetically altered wound healing in keloid fibroblasts. J Invest Dermatol. 2010; 130:2489–2496. [PubMed: 20555348]

Saed GM, Ladin D, Olson J, Han X, Hou Z, Fivenson D. Analysis of p53 gene mutations in keloids using polymerase chain reaction-based single-strand conformational polymorphism and DNA sequencing. Arch Dermatol. 1998; 134:963–967. [PubMed: 9722726]

Sankararaman S, Sridhar S, Kimmel G, Halperin E. Estimating local ancestry in admixed populations. Am J Hum Genet. 2008; 82:290–303. [PubMed: 18252211]

Smith JC, Boone BE, Opalenik SR, Williams SM, Russell SB. Gene profiling of keloid fibroblasts shows altered expression in multiple fibrosis-associated pathways. J Invest Dermatol. 2008; 128:1298–1310. [PubMed: 17989729]

Trupin, Williams J.; Hammons, J.; Russell, J. Multicellular Origin of Keloids. Birth Defects. Fifth International Conference; Montreal: Elsevier; 1977. p. 121

Winkler CA, Nelson GW, Smith MW. Admixture mapping comes of age. Annu Rev Genomics Hum Genet. 2010; 11:65–89. [PubMed: 20594047]

Zhu F, Wu B, Li P, Wang J, Tang H, Liu Y, Zuo X, Cheng H, Ding Y, Wang W, Zhai Y, Qian F, Wang W, Yuan X, Wang J, Ha W, Hou J, Zhou F, Wang Y, Gao J, Sheng Y, Sun L, Liu J, Yang S, Zhang X. Association study confirmed susceptibility loci with keloid in the Chinese Han population. PLoS One. 2013; 8:e62377. [PubMed: 23667473]

Zhu X, Luke A, Cooper RS, Quertermous T, Hanis C, Mosley T, Gu CC, Tang H, Rao DC, Risch N, Weder A. Admixture mapping for hypertension loci with genome-scan markers. Nat Genet. 2005; 37:177–181. [PubMed: 15665825]

Zhu X, Young JH, Fox E, Keating BJ, Franceschini N, Kang S, Tayo B, Adeyemo A, Sun YV, Li Y, Morrison A, Newton-Cheh C, Liu K, Ganesh SK, Kutlar A, Vasan RS, Dreisbach A, Wyatt S, Polak J, Palmas W, Musani S, Taylor H, Fabsitz R, Townsend RR, Dries D, Glessner J, Chiang CW, Mosley T, Kardia S, Curb D, Hirschhorn JN, Rotimi C, Reiner A, Eaton C, Rotter JI, Cooper RS, Redline S, Chakravarti A, Levy D. Combined admixture mapping and association analysis identifies a novel blood pressure genetic locus on 5p13: contributions from the CARe consortium. Hum Mol Genet. 2011; 20:2285–2295. [PubMed: 21422096]

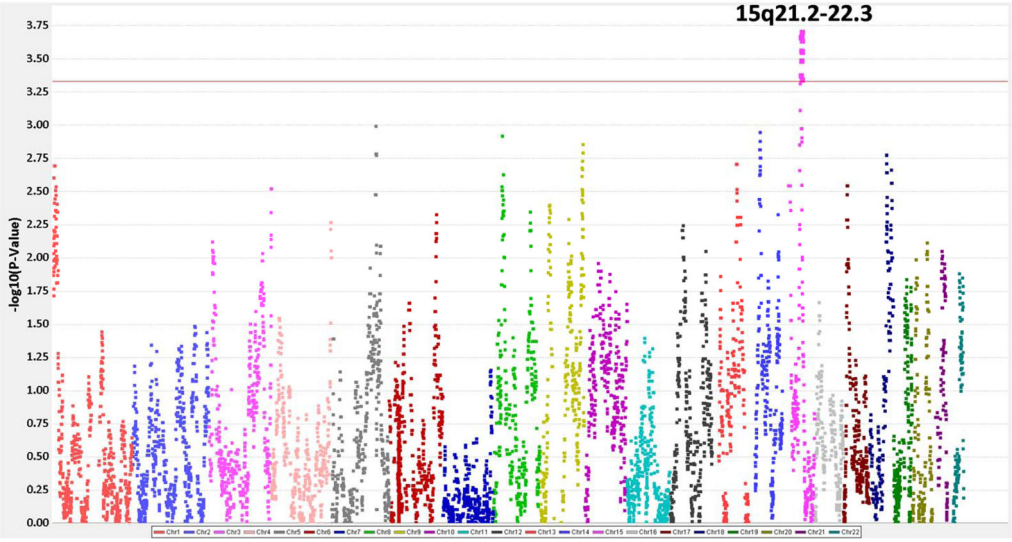

Figure 1. Summary of admixture mapping analyses

X-axis indicates the genomic position along the chromosome beginning from chromosome 1 to chromosome 22. Y-axis indicates the −log10(P-Value) for the admixture mapping analyses. The red line indicates one $\log_{10}$ down from the family-wise error rate (FWER) calculated from 1000 permutations.

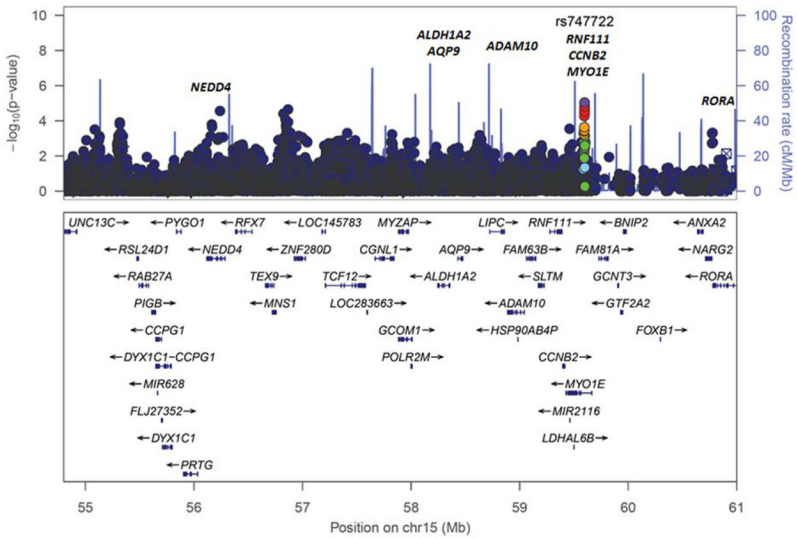

Figure 2. Focused LocusZoom plot of SNP association analyses of genotyped and imputed SNPs within the strongest association interval on chr15q21.2-22.3

Labeled are candidate genes associated within the regions based on prior studies. We focus on the chr15q21.2-22.3, as it is the region one $-\log_{10}$ down from the $-\log_{10}$ value of the FWER, calculated from n=1000 permutations. The index SNP (SNP with most statistically significant p value) is color coded in purple. The LD between the index SNP and nearby SNPs is indicated by the color coding for the SNP associations, where red indicates high LD (r^2 <0.8), orange r^2 0.6–0.8, green r^2 0.4–0.6, light blue r^2 0.2–0.4 and dark blue indicates r^2 0.20-0.

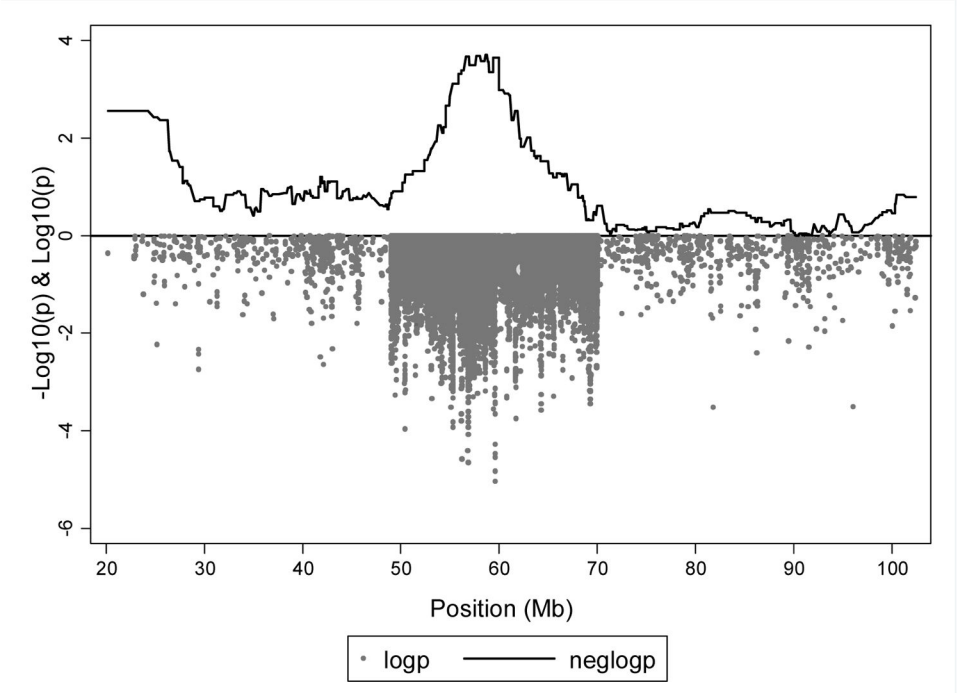

Figure 3. Admixture mapping peak on chromosome 15 with overlapping keloid single SNP association analysis results

X-axis indicates genomic position along the chromosome. The top of the Y-axis indicates $-\log_{10}$ (p-value) from logistic regression of admixture mapping values generated from LAMP-ANC (solid black line). The bottom portion of the Y-axis (grey circles) indicates $\log_{10}$ (p-value) for SNP association analyses with keloid outcome of genotyped and imputed SNPs (imputed only within admixture mapping peak region).

Table 1

Summary of demographic characteristics and ancestry estimates

	Keloid Cases (N = 122)	Controls (N=356)	p[1]
Age mean (standard deviation), years	43 (17)	52 (18)	$2.1x10^{-5}$
Sex			0.119
Male (%)	30	38	
Female (%)	70	62	
European Ancestry (%)	19	21	0.011

* Age and BMI were only available for samples obtained from BioVU

[1] Logistic regression analysis P value of demographic variables compared between cases and controls adjusted for ancestry principle components, except the European Ancestry variable which was not adjusted for principle components

Table 2

Summary of genotyped and imputed SNPs associated under chr15q21.2-22.3

Gene/Nearby Gene[1]	SNP rs#	Position (bp)	MA	MAF Case/Control	OR	95% CI		P-Value
						Lower	Upper	
MYO1E	rs747722	59599222	T	0.34/0.24	4.41	2.29	8.50	$9.07x10^{-6}$
MYO1E	rs28394564	59599606	G	0.34/0.24	4.41	2.29	8.50	$9.07x10^{-6}$
MYO1E	rs28673219	59599996	A	0.34/0.24	4.24	2.21	8.16	$1.47x10^{-5}$
MNS1||ZNF280D	rs55780277	56866974	T	0.43/0.51	0.30	0.17	0.53	$2.23x10^{-5}$
NEDD4	rs138585173	56239500	A	0.14/0.09	8.00	3.03	21.10	$2.67x10^{-5}$
MYO1E	rs8034553	59598008	T	0.33/0.24	4.12	2.12	7.99	$2.83x10^{-5}$
MYO1E	rs747721	59598381	G	0.33/0.24	4.12	2.12	7.99	$2.84x10^{-5}$
MYO1E	rs16941266	59596619	C	0.33/0.23	4.04	2.09	7.81	$3.41x10^{-5}$
MNS1||ZNF280D	rs3858886	56824818	A	0.25/0.32	0.34	0.20	0.56	$3.88x10^{-5}$
MYO1E	rs28583561	59597454	G	0.33/0.23	3.93	2.02	7.62	$5.23x10^{-5}$
MNS1||ZNF280D	rs57287820	56844596	A	0.29/0.22	3.69	1.93	7.08	$8.23x10^{-5}$

BP-base pair (GRCH37.p13/hg19); MA-minor allele; MAF-minor allele frequency; OR-odds ratio; CI-confidence interval

[1]Nearby genes, noted with//, are the closest genes to the associated SNP regardless of distance

Table 3

Summary of associated SNPs from genome-wide analyses (47,599 SNPs) excluding 15q.21-22.3

Chr	Gene/Nearby Genes[1]	SNP rs#	Position (bp)	MA	MAF Case/Control	OR	95% CI		P-Value
							Lower	Upper	
11	*MYO7A*	rs35641839[1]	76885901	A	0.12/0.04	4.71	2.38	9.32	$8.34x10^{-6}$
4	*INPP4B*	rs2636675	143224472	G	0.18/0.13	3.23	1.89	5.54	$1.97x10^{-5}$
4	*ASSP8||LOC100130396*	rs10016143	149615561	A	0.45/0.38	2.37	1.56	3.59	$5.00x10^{-5}$
11	*MYO7A*	rs111033183[2]	76910600	T	0.12/0.06	3.79	1.98	7.28	$6.20x10^{-5}$
13	*LOC101928697*	rs77809485	19761359	G	0.11/0.05	3.96	2.02	7.78	$6.55x10^{-5}$
11	*MYO7A*	rs78871677[2]	76901819	A	0.12/0.06	3.55	1.89	6.69	$8.66x10^{-5}$

BP-base pair (GRCH37.p13/hg19); MA-minor allele; MAF-minor allele frequency; OR-odds ratio; CI-confidence interval

[1]Nearby genes, noted with//, are the closest genes to the associated SNP regardless of distance

[2]rs35641839 missense mutation, rs111033183 missense mutation, rs78871677 synonymous mutation

Table 4

Summary of conditional association analyses for SNPs within chr15q21.2-22.3 conditioning on rs747722 and for *MYO7A* (strongest global scan association) conditioning on rs35641839

Gene/Nearby Gene[1]	SNP rs#	P-Value
Conditioning on *MYO1E* rs747722		
NEDD4	rs138585173	7.20×10^{-6}
MNS1||ZNF280D	rs3858886	1.13×10^{-4}
	rs57287820	4.38×10^{-4}
	rs55780277	1.57×10^{-5}
MYO1E	rs16941266	0.904
	rs28583561	0.658
	rs8034553	0.924
	rs747721	0.923
	rs28394564	1.000
	rs28673219	0.910
Conditioning on *MYO7A* rs35641839		
MYO7A	rs78871677	0.791
	rs111033183	0.820

[1] Nearby genes, noted with//, are the closest genes to the associated SNP regardless of distance

Hum Genet. Author manuscript; available in PMC 2015 February 19.

Table 5

GWAS studies of fibroproliferative disorders with associated SNPs within chr15q21.2-22.3

Study Population Trait	PUBMED ID	Region	Position (bp)	SNPs	Reported Gene(s)	Risk Allele	OR or beta	P-Value[2]
Adeyemo A et al (2009)(Adeyemo *et al.*, 2009) AAs Hypertension	19609347	15q21.3	58213414	rs1550576	*ALDH1A2*	NR	1.92	$3x10^{-6}$
Nakashima M et al (2010)(Nakashima *et al.*, 2010) Japanese Keloid	20711176	15q21.3	56194877	rs8032158	*NEDD4*	C	1.51	$6x10^{-13}$
Moffatt MF et al (2010)(Moffatt *et al.*, 2010) European Ancestry Asthma	20860503	15q22.2	61069988	rs11071559	*RORA*	C	1.18	$1x10^{-7}$
Inouye M et al (2012) European Ancestry Metabolite(Inouye *et al.*, 2012) levels/Atherosclerosis	22916037	15q22.2	59487930	rs2306786	*MYO1E, CCNB2, RNF111*	NR	NR	$1x10^{-10}$
Inouye M et al (2012) European Ancestry Metabolite(Inouye *et al.*, 2012) levels/Atherosclerosis	22916037	15q22.2	58471979	rs16939881	*AQP9*	NR	NR	$3x10^{-27}$

*Data abstracted from the NHGRI GWAS catalog of published GWAS studies (https://www.genome.gov/26525384)

[1]NR-not reported

[2]P-Value reported in published study

Table 6

Differential expression in keloid versus normal fibroblasts in genes at chr15q21.2-22.3

Chromosome Location	Gene	Keloid/Normal (HC)	Keloid/Normal (no HC)
15q21.3	*ADAM10*	2.19	1.04 (ns)
15q21	*TCF12*	0.56	0.58
15q21.3	*CCNB2*	0.34	0.43
15q21-q22	*MYO1E*	2.01	1.93 (ns)

ns-not significant; HC-hydrocortisone