Qualifying Exam Preparation Materials

Students will be required to demonstrate expertise in their chosen areas of concentration, which may include deeper knowledge than covered in the required coursework. In addition to mastery in their specific area, students will be required to demonstrate broad knowledge of the field of Human Genetics as a whole. Students will be responsible for showcasing their ability to utilize knowledge and skills obtained through coursework and research experiences to inform their interpretation, synthesis, and evaluation of the contemporary peer-reviewed literature.

The following list of knowledge requirements that students are expected to master may be used to assist in preparing for the qualifying exam. Likewise, course objectives for the core Human Genetics curriculum are provided below and may serve as an outline of topics to assist in students’ preparations. Material covered in additional courses relevant to an individual student’s academic experience and trajectory may also be germane. Lastly, example qualifying exam questions pertaining to specific publications from the primary literature are provided. Please be aware that some questions raised during the oral examination may not have clear or objective “right” answers, but may instead reflect current uncertainties in the field of human genetics. In these cases, students may be expected to provide thoughtful discussion regarding prevailing hypotheses or themes, demonstrate knowledge of ambiguities in the scientific evidence or limitations of our current knowledge, and propose new experiments that could yield insight into the question at hand.

Knowledge requirements:

- **DNA, Genes, Chromosomes, and Genomes**
  - definition, structure, function, and organization of genes, chromosomes, and genomes
  - mitosis, meiosis, and recombination
  - central dogma and mechanisms of DNA replication, transcription, and translation
  - gene expression and regulation during development, normal, and abnormal cellular functions
  - genetic variation at the nucleotide, gene, chromosome, and genome levels

- **Genetic Control of Phenotypes**
  - modes of inheritance (autosomal/sex-linked, dominant/additive/recessive, mitochondrial, etc.) and corresponding patterns of phenotypic variation in human pedigrees
  - basic principles and examples of inborn errors of metabolism and classic genetic syndromes
  - the multifactorial nature of complex traits and the principles of multifactorial inheritance
  - molecular techniques for detecting genetic mutations and polymorphisms
  - use of model organisms to study human genetic diseases
  - principals of cancer genetics, the two-hit hypothesis, and somatic mutations

- **Disease Gene Discovery and Gene Mapping**
  - genetic epidemiology: modeling genotype-phenotype relationships, and identifying the genetic contributors of a phenotype
  - how multifactorial inheritance impacts disease gene mapping in humans
  - rationale, advantages, and limitations of linkage analysis study designs
- rationale, advantages, and limitations of genetic association study designs
- interpretation of classical statistical tests commonly used in genetic studies
- approaches for studying the role of copy number variation (CNV) in complex disease
- applications of next-generation sequencing (NGS) in the studying human disease
- the role of rare variants in common complex disease and how rare variants are tested for association
- approaches for finding functional variations
- examples of complex disease genes that have been successfully identified
- definition of multiple testing and common methods for correction
- use of animal models to define the pathophysiology of disease

- Population Genetics
  - impact of fundamental principles of population genetics including Hardy-Weinberg Equilibrium, its assumptions, and the effects of violations of these assumptions
  - how evolutionary principles (like bottlenecks, founder effects, and population isolates) can affect or be exploited for human genetic studies
  - how allele frequencies differ in populations and the impact on genetic studies
  - linkage disequilibrium, how it is described and measured, how it relates to haplotypes, and how it impacts association studies

- Genetics in Society
  - how novel scientific discoveries are evaluated in a clinical context and applied appropriately to the care of patients, the function of genetic counseling
  - issues in genetic testing
  - how legal and ethical issues affect research, including informed consent

- Molecular Techniques
  - definition, implementation, advantages, and limitations of PCR, genotyping methods, gene expression methods, sequencing methods

Course objectives:

HUGEN 2022: Population Genetics

- apply the Law of Hardy-Weinberg Equilibrium and its assumptions to calculate allele and genotype frequencies
- predict the consequences of genetic inheritance and recombination in populations including the concepts of linkage and linkage disequilibrium
- interpret the qualitative effects of violations of Hardy-Weinberg Equilibrium and solve simple quantitative problems demonstrating these effects
- express the fundamental goals and principles of genetic epidemiology by modeling genotype-phenotype relationships, quantitative traits, and heritability
- interpret results from gene discovery methods such as linkage analysis and large-scale genetic association studies, and critically evaluate the strengths, limitations, and appropriate applications of these methods.

HUGEN2031: Chromosomes and Human Disease

- compare and contrast the strengths and limitations of the various cytogenetic assays and the appropriate tests to apply in a particular situation
- describe chromosomal alterations and chromosome breakage syndromes and their clinical implications
• describe the role of chromosomal alterations in development of cancer and their use in cancer diagnostics
• interpret cytogenetic nomenclature
• describe the cellular basis of chromosome segregation, chromosome structure, meiosis and mitosis, numerical and structural chromosome abnormalities in clinical disorders, including chromosomal syndromes, chromosome breakage syndromes, and cancer, sex determination and sex chromosome abnormalities, genomic imprinting, trinucleotide repeat disorders, classical and molecular cytogenetic methods including copy number arrays and nomenclature, the role of cytogenetic techniques in diagnosis of disorders, the history of cytogenetic methods, and ethical issues related to cytogenetic testing and results.

HUGEN 2034: Biochemical and Molecular Genetics of Complex Disease

• Describe the complex interactions between genetic variation and environmental exposures contributing to the following selected examples of complex human diseases: cardiovascular disease, neurodegenerative and psychiatric diseases, diabetes, lupus, age-related macular degeneration, cancer, osteoporosis, infectious disease, and asthma
• Identify the general tenets regarding the genetic susceptibility to common, complex diseases
• Describe the public health impact of common, complex diseases

HUGEN 2040: Molecular Genetics of Mendelian Disease

• The objectives of the course are to present a variety of Mendelian diseases from the perspective of the complexity of contemporary genetic thought. There are substantial readings assigned. The core of understanding expected of each student for each disorder will be:
  • the protein or transcription factor causal for each disease
  • the clinical signs and symptoms characterizing the phenotype
  • any unique features which exemplify non-mendelian characteristics of the disorder
  • the treatment strategies for the disorder
  • any unique human populations which harbor alleles for the disorder. Other features may be emphasized on a case by case basis
1. Why would investigators perform a GWAS of endometriosis when many previous candidate gene studies have been reported?

2. Surgically-confirmed cases of endometriosis were compared to controls. Given that endometriosis is a common gynecological condition and that diagnosis is typically delayed 7-10 years from onset, do you expect that some control participants may actually be undiagnosed cases? How would the presence of cryptic endometriosis cases in the control sample effect this study? Comment on how this scenario would impact the reported odds ratios.

3. Previous twin studies have estimated that the narrow-sense heritability of endometriosis is 51%, 75%, and 87%. Do you interpret these three estimates as consistent with each other? If so, explain why. If not, explain possible reasons for this inconsistency. What additional pieces of information could help you jointly interpret these results?

4. Genomic data were used to test for unknown relatedness among participants, and if detected, related samples were excluded from analysis. Why did the investigators do this?

5. Genomic data were used to determine genetic ancestry, and only samples of 95% or more European ancestry were included in statistical analysis. Why did the investigators do this?

6. Numerous genotyped and imputed SNPs in the chromosomal 1 region containing WNT4, CDC42, and HSPC157 showed evidence of genetic association (Figure 1). Based on these associated SNPs, is it possible to determine how many causal alleles are in this region? Are variants in all three genes causal? What additional information could help answers these questions?
# Genome-Wide Association Study Link Novel Loci to Endometriosis

Hans M. Albertsen, Rakesh Chettier, Pamela Farrington, Kenneth Ward

Juneau Biosciences, LLC, Salt Lake City, Utah, United States of America

## Abstract

Endometriosis is a common gynecological condition with complex etiology defined by the presence of endometrial glands and stroma outside the womb. Endometriosis is a common cause of both cyclic and chronic pelvic pain, reduced fertility, and reduced quality-of-life. Diagnosis and treatment of endometriosis is, on average, delayed by 7–10 years from the onset of symptoms. Absence of a timely and non-invasive diagnostic tool is presently the greatest barrier to the identification and treatment of endometriosis. Twin and family studies have documented an increased relative risk in families. To identify genetic factors that contribute to endometriosis we conducted a two-stage genome-wide association study (GWAS) of a European cohort including 2,019 surgically confirmed endometriosis cases and 14,471 controls. Three of the SNPs we identify associated at \( P < 5 \times 10^{-8} \) in our combined analysis belong to two loci: LINC00339-WNT4 on 1p36.12 (rs2235529; \( P = 8.65 \times 10^{-9} \), \( OR = 1.29 \), CI = 1.18–1.40) and RND3-RBM43 on 2q23.3 (rs1519761; \( P = 4.70 \times 10^{-8} \), \( OR = 1.20 \), CI = 1.13–1.29), and rs6757804; \( P = 4.05 \times 10^{-8} \), \( OR = 1.20 \), CI = 1.13–1.29). Using an adjusted Bonferroni significance threshold of 4.51 \( 10^{-6} \) we identify two additional loci in our meta-analysis that associate with endometriosis; RNF144B-ID4 on 6p22.3 (rs6907340; \( P = 2.19 \times 10^{-7} \), \( OR = 1.20 \), CI = 1.12–1.28), and HNRNPAP3-P1-LOC100135039 on 10q11.21 (rs10508881; \( P = 4.08 \times 10^{-7} \), \( OR = 1.19 \), CI = 1.11–1.27). Consistent with previously suggested associations to WNT4 our study implicate a 150 kb region around WNT4 that also include LINC00339 and CDC42. A univariate analysis of documented infertility, age at menarche, and family history did not show allelic association with these SNP markers. Clinical data from patients in our study reveal an average delay in diagnosis of 8.4 years and confirm a strong correlation between endometriosis severity and infertility (\( n = 1182 \), \( P < 0.001 \), \( OR = 2.18 \)). This GWAS of endometriosis was conducted with high diagnostic certainty in cases, and with stringent handling of population substructure. Our findings broaden the understanding of the genetic factors that play a role in endometriosis.

---


**Editor:** Li-Min Huang, National Taiwan University Hospital, Taiwan

**Received** October 7, 2012; **Accepted** February 1, 2013; **Published** March 5, 2013

**Copyright:** © 2013 Albertsen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** No current external funding sources for this study.

**Competing Interests:** The authors have the following interests. Hans M. Albertsen, Rakesh Chettier, Pamela Farrington and Kenneth Ward are employed by Juneau Biosciences LLC, the funder of this study. All authors have direct financial interest in Juneau Biosciences. A US provisional patent application has been filed by Juneau Biosciences that include the results and inventions reported in the manuscript. Title: Genetic Markers Associated with Endometriosis and Use Thereof Serial No.: 61/707,730 Filing date: September 28, 2012. There are no further patents, products in development or marketed products to declare. This does not alter the authors’ adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

* E-mail: kenw@juneaubiosciences.com

---

**Introduction**

Endometriosis affects 5–10% of women in their reproductive years with symptoms including pelvic pain, dyspareunia, dysmenorrhea and infertility [1]. Although ectopic endometrium has been observed in female fetuses [2], symptoms of endometriosis usually don’t manifest until adolescence, and some patients with severe endometriosis remain asymptomatic. Definitive diagnosis is often delayed 7–10 years after the onset of symptoms severely impacting quality of life [3–5]. Family history of endometriosis has been reported in multiple studies to increased relative risk about a 5-fold [6,7]. A large twin-study based on the Australian Twin Registry has shown that the ratio of mono-zygotic to fraternal twin pair correlations was in excess of 2 fold, suggesting that 51% of the variance of the liability to endometriosis may be attributable to additive genetic influences with minimal influence from environmental factors [8], and two smaller twin-studies report the concordance rate of endometriosis between monozygotic twins to range between 75% and 87% [9,10]. A large number of candidate genes have been investigated for their role in endometriosis as summarized by Montgomery et al. [11] and Rahmioglu et al. [12], but the first strong evidence to date for genetic association are reported in two large Genome-Wide Association Studies (GWAS). In the first study Uno et al. [13] identified rs10965235 located in an intron of CDKN2BAS on chromosome 9q21 to be associated in a Japanese cohort, and in the second study Painter et al. [14] identified the intergenic SNP rs12700667 on chromosome 7p15.2 to be associated in a European cohort. A meta-analysis of the two studies extend this evidence and identify a total of seven loci associated with endometriosis [15]. To replicate and extend our understanding of the genetic factors that contribute to endometriosis we have undertaken a large two-stage GWAS in a European cohort.

**Results and Discussion**

**GWAS and Replication**

We conducted a discovery GWAS on surgically confirmed endometriosis patients and population controls using the Illumina OmniExpress BeadChip. SNPs were limited to the autosomes and
SNPs with an Illumina Genotyp score ≥0.65. We further eliminated SNPs with callrate <0.98, Hardy-Weinberg Equilibrium (hwe) <0.001 and minor allele frequencies (MAF) <0.01. After filtering 580,699 SNPs remained. Next, samples with identity-by-state (IBS), and samples closer than 3rd-degree were removed. We used ADMIXTURE (ver. 1.22) to estimate individual ancestry proportions based on a subset of SNPs on the Illumina OmniExpress chip (see Materials and Methods) and restricted our analysis to samples with ≥95% European ancestry. The calculated ancestral distribution of samples within Europe is shown in Figure S1. After applying quality, relatedness and ethnicity filters 1,514 case and 12,660 control samples were used for the discovery phase of the association analysis. The genomic inflation factor lambda (λ) was determined to be 1.18, indicating measurable population stratification across the samples. To account for the elevated λ, we performed a PCA adjusted association analysis that resulted in a λ-value of 1.05 shown in QQ-plots in Figure S2. We selected the top 100 SNPs with the lowest PCA-adjusted P-values (ranging between 8.20×10⁻⁵ and 1.36×10⁻⁷) for further association analysis in the replication stage (Table S1).

The replication samples included 505 cases and 1811 controls selected for the same criteria as the discovery set. The λ-value for the replication cohort was determined to be 1.01 suggesting no measurable population stratification. After applying the same SNP filters as above we analyzed the top 100 SNPs from the discovery GWAS in the replication set. A significance threshold for the study, allowing for multiple correction, was chosen at 4.51×10⁻⁷ (0.05/108,699; 108,699 being the number of independent SNPs in the panel of 580,699 filtered SNPs with r²<0.20). A meta-analysis of the discovery and replication results was performed using Cochran-Mantel-Hanzel test and revealed 8 SNPs from 4 genomic regions that passed our genome-wide significance threshold including: LINC00339-WNT4 on 1p36.12 (rs2235529; PMeta = 3.05×10⁻⁹, OR = 1.30); RNDR3-RBM43 on 2q23.3 (rs6757804; PMeta = 6.45×10⁻⁸, OR = 1.20); RNFL14-B-Id4 on 6p22.3 (rs6007340; PMeta = 2.19×10⁻⁷, OR = 1.20); and HRNRNA3P1-LOC100130539 on 10q11.21 (rs10508881; PMeta = 4.08×10⁻⁵, OR = 1.19) and shown in detail in Table 1. Table 1 also show that three SNPs (rs2235529, rs1519761 and rs6757804) pass a conventional genome wide significance threshold (P < 5.0×10⁻⁸, Hardy-Weinberg Equilibrium (HWE) passed) and show that the risk for endometriosis is confined to a single haplotype anchored in rs10917151, rs4654783 and rs2235529 together with rs16826658, and rs7210902 using the imputed data from our population. The haplotype-results are summarized in Table S3 and show that the risk for endometriosis is confined to a single haplotype anchored in rs10917151, rs4654783 and rs2235529 (PHAP-1 = 7.26E-09, ORHAP-1 = 1.28), and that this haplotype starts to deteriorate with the addition of rs16826658 and rs7210902 (PHAP-1,2 = 8.33E-07, ORHAP-1,2 = 1.25). This analysis suggests that the risk allele observed in the present study and the two previously published reports is located on the same ancestral haplotype. Imputation was also performed on the three other significant regions on chromosome 2, 6, and 10, and on the region surrounding IL33 on chromosome 9 (Table S2a–c, Figure S3).

**Overlap with Other Reported Loci**

Uno et al. [13] reported association between endometriosis and rs19065235, located in intron 19 of CDKN2BAS, in a Japanese population. The protective minor allele A, that has a minor allele frequency of 0.20 in the Japanese population, is not observed in the European population preventing a direct comparison between the two ethnic groups. In lieu of a direct comparison between the two ethnic groups we scanned 66 SNPs from a region of 200 kb surrounding rs19065235 in our European population. After correcting for multiple-testing (P-value ≤0.05/66 = 0.00076) we didn’t find any evidence that CDKN2BAS is associated with endometriosis in the European population (Table S4a). In contrast, a second SNP (rs13271465) located on 8p22 between MTM7 and SLC7A2, that Uno et al. identified as being associated with endometriosis (Pcombined = 9.84×10⁻⁶, OR = 1.18), is present in both studies, show weak association (P = 0.0057, OR = 1.14) in our study, while none of the other 88 SNPs from the 200 kb region surrounding rs13271465 showed GWAS Link Novel Loci to Endometriosis...
any evidence for association with endometriosis (Table S4b). Uno et al. provided a supplementary list of 100 additional candidate SNPs from their study of which 60 are present in our analysis. Among the 60 SNPs only rs2473277 has a P-value <0.001 in our study (P = 5.97 × 10⁻⁶, OR = 1.17). SNP rs2473277 is located between LINC00339 and CDC42 at the left-most boundary of the WNT4 LD-block discussed above (Figure 1), and the risk allele of rs2473277 tag perfectly with the risk haplotype HAP-1 shown in Table S3 (data not shown).

Painter et al. [14] reported significant association of moderate and severe endometriosis with rs12700667 on 7p15.2 (P_all = 2.6 × 10⁻⁷, OR = 1.22), with flanking support from
rs7798431 located 41 kb away. The two SNPs were reported to be in strong linkage disequilibrium (LD), but unfortunately neither SNP is included in our study. A review of the Hapmap3 data from the region show that three SNPs in our study (rs12535837, rs10282436, rs10232819) are in moderate to strong LD with rs12700667 and rs7798431, but we find no evidence for association between endometriosis and any of these markers in our analyses of all endometriosis cases together nor do we find any evidence for association to the moderate and severe subset (Table S4c). A broader scan of the 200 kb region surrounding rs12700667, suggests weak association with rs4722551 (P = 0.000867) about 90 kb downstream of rs12700667 (Table S4c). Painter et al. provided a supplementary list of 73 candidate SNPs, but none of the thirty-nine SNPs present in our study reach a P-value threshold.

A recent meta-analysis published by Nyholt et al. [15] extend the findings by Uno et al. and Painter et al. and report a total of seven SNPs that pass a genome-wide significance (P<5 x 10^{-8}). A comparison of our results to each of the seven loci show evidence for association to endometriosis for three of the seven loci as detailed in Table S7.

Clinical Stratification and Diagnostic Delay

Clinical features commonly used to characterize and stratify endometriosis include infertility, pelvic pain, severity, age-at-menarche and familiality. To determine the diagnostic delay in our patient-population we identified a group of women (n = 874) that reported both age at onset-of-symptoms (mean-age-onset = 19.04 years) and age at diagnosis (mean-age-diagnosis = 27.49), and observed an average diagnostic delay of 8.44 years, similar to previous studies. We then went on to examine if our samples showed any clinical correlations using logistic regression. The analysis revealed strong correlations between severity and infertility (P<0.001, OR = 2.19), and between severity and diagnostic delay (P<0.001, OR = 1.04) as shown.
in Table 2. To identify loci associated with the progression of endometriosis, we compared patients with mild endometriosis to patients with moderate or severe endometriosis in a two-stage GWAS. Stage one included 657 patients with mild endometriosis vs. 523 patients with moderate-or-severe endometriosis and a stage two replication set of 318 mild vs. 519 moderate-or-severe patients. A meta-analysis using the CMH test in this limited sample set, found no loci that pass the genome-wide significant threshold which suggest the SNPs identified in the primary study contribute to the general endometriosis risk rather than endometriosis progression. A separate analysis of the top five SNPs using logistic regression also showed no correlation with severity as shown in Table S5a, and Table S5b shows no noticeable increase in effect size when comparing moderate and severe disease against all controls.

Risk Analysis of Endometriosis

After removing markers with $r^2$ > 0.8 among the top 5 associated regions (incl. the IL33 locus), we conducted multivariate logistic regression using the combined set of 2,019 cases and 14,471 controls. All of the 5 SNPs rs101917151, rs6757904, rs6907340, rs10975519 and rs10508881 analyzed remained significant with OR of 1.3, 1.2, 1.18, 1.17 and 1.17. Each marker appear to be an independent risk factors for endometriosis. Comparison of the OR between the discovery and replication datasets, shown in Table 1, does not suggest any significant inflation of effect size in the discovery dataset (winner’s curse), but this conclusion remain tentative due to significant inflation of effect size in the replication datasets, shown in Table 1, does not suggest any.

Conclusion

A two-stage GWAS and a replication study involving 2,019 cases and 14,471 controls was performed which identified four novel loci strongly associated with endometriosis and confirmed the involvement of a region around WNT4 which previously have been suggested as being associated to endometriosis. Nine other regions identified in the study also hold promise as candidate loci for endometriosis. Utmost care was taken in the clinical classification of patients and only surgically-confirmed cases with >95% European ancestry were considered in this large GWAS of endometriosis. The study is well powered (>90%) to identify a marker at or above 10% minor allele frequency (MAF) with odds-ratio (OR) >1.20, but we estimate the top 5 loci only explain about 1.5% of the phenotypic variance of endometriosis. Since the few risk loci we detected all have odds ratios <1.30 it must be assumed that any new endometriosis loci that contribute to the “missing heritability” must be rare, recent, or show minimal effect. GWAS, by design, detects only very old founder effects. When a phenotype includes infertility, like endometriosis, a high mutation rate would be required to replenish the disease-causing alleles lost from the gene-pool due to infertility. One suitable avenue to investigate under that scenario is to use whole genome sequencing of high-risk families rather than SNP-based GWAS. Little is presently known about the pathophysiology of endometriosis, but we hope that a more detailed investigation of the loci presented in this paper will help elucidate the pathogenesis of endometriosis and clarify its genetic underpinnings.

Materials and Methods

Ethics Statement

All subjects and controls provided written informed consent in accordance with study protocols approved by Quorum Review IRB (Seattle, WA 98101).

Participant Recruitment

Patients included in the present study were invited to participate via an outreach program at www.endoendo.com, where our research initiative is described in more detail. Briefly, the “End to Endo” website provides general information regarding endometriosis and our research project, and invites women diagnosed with endometriosis to participate in our study.

Medical Review

The inclusion criteria in the endometriosis case population in the present study is surgically confirmed diagnosis of endometriosis with laparoscopy being the preferred method. Trained OB/GYN clinicians performed the medical record review and clinical assessment of each individual patient. Patients were considered to be affected if they had biopsy-proven lesions or if operative reports revealed unambiguous gross lesions. Patients were further categorized by severity, clinical history of pelvic pain, infertility, dyspareunia or dysmenorrhea and family history. Patients were grouped into one of three classes of severity; mild, moderate or severe, following the general guidelines set forth by ASRM [25]. Exclusion of endometriosis also requires surgical intervention and we made no attempt to exclude endometriosis in the population controls. Thus, in this analysis we are comparing cases with 100% prevalence of endometriosis to controls with the population prevalence of endometriosis (5–10%), which leads to a systematical underestimation of the true odds ratios and a decrease in statistical power to detect associations.

DNA Extraction

Saliva samples were collected using the Oragene 300 saliva collection kit (DNA Genotek; Ottawa, Ontario, Canada) and

| Table 2. Endometriosis severity correlate with infertility and diagnostic delay. |
|-----------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Clinical Feature            | Moderate or Severe endometriosis (n = 842) | Mild endometriosis (n = 1177) | Category         | OR        | Beta     | SE      | P         |
| infertility (1182)          | 525            | 657            | Yes or No       | 2.19      | 0.78     | 0.12    | 8.52E-11  |
| Family History (8811)       | 790            | 1091           | Yes or No       | 0.89      | -0.11    | 0.09    | 0.23      |
| Age at Menarche (921)       | 405            | 516            | <12 or >12 yrs  | 1.20      | 0.18     | 0.13    | 0.18      |
| Diagnostic Delay (874)      | 383            | 491            | 0 to 35         | 1.04      | 0.04     | 0.01    | 2.18E-05  |

Clinical features were correlated to severity. Only patients that could be unambiguously categorized were included in the analysis with total counts provided in parenthesis next to the clinical feature. P-values (P) are calculated using Wald test. Beta is the regression coefficients and SE the standard error from logistic regression. doi:10.1371/journal.pone.0058257.t002
DNA was extracted using an automated extraction instrument, AutoPure LS (Qiagen; Valencia, CA), and manufacturer’s reagents and protocols. DNA quality was evaluated by calculation absorbance ratio OD$_{260}$/OD$_{280}$, and DNA quantification was measured using PicoGreen® (Life Technologies; Grand Island, NY).

**Microarray Genotyping**

The discovery set of 1514 cases and 12660 controls and replication set of 505 cases and 1011 controls were genotyped using the Illumina Human OmniExpress Chip (Illumina; San Diego, CA) according to protocols provided by the manufacture. Figure S4 show the genotype clusters for the top eight SNPs in our study. All SNPs reported in the present study passed visual inspection for cluster quality. It is our experience that technical replication does not affect genotype calls of SNPs with high quality clusters and due to cost we could not justify independent technical replication.

**Taqman Genotyping**

A Taqman® 7900 instrument (Life Technologies; Grand Island, NY) and manufacturer’s protocols were used to genotype rs61764370. Genotypes were determined using Taqman genotyping software SDS (v2.3) and the genotype cluster was visually inspected. Genotyping QC for rs61764370 passed standard criteria of call rate >95% and no deviation from HWE ($p<0.001$) was observed.

**Sample Quality Control**

Samples were excluded from the analysis if they missed any of the following quality thresholds:

a) Evidence of familial relationship closer that 3rd-degree ($\pi>0.2$) using genome-wide Identity-By-State (IBS) estimation implemented in PLINK

b) Samples with missing genotypes >0.02

c) Samples with non-European admixture >0.05 as determined by ADMIXTURE

**SNP Quality Control**

SNPs were excluded from the analysis if they missed any of the following quality thresholds:

a) SNPs with Illumina GenTrain Score <0.65

b) SNPs from copy number variant regions or regions with adjacent SNPs

c) SNPs failing Hardy-Weinberg Equilibrium (HWE) $P \leq 10^{-3}$

d) SNPs with minor allele frequency (MAF) $\leq 0.01$ in the control population

e) SNP call rate $\leq 98\%$

**Admixture**

ADMIXTURE (ver. 1.22) was used to estimate the individual ancestry proportion [16]. The software estimates the relative admixture proportions of a given number of a priori defined ancestral groups contributing to the genome of each individual. We used the POPRES dataset [26] as a reference group to create a supervised set of 9 ancestral clusters. Seven of them belong to the European subgroups along with African and Asian groups. Since POPRES dataset utilized Affymetrix 5.0 chip, we used 105,079 autosomal SNPs that overlapped with the Illumina OmniExpress dataset. Among the 105,079 SNPs we selected a subset of 33,067 SNPs that showed greater genetic variation (absolute difference in frequency) among the 9 reference groups. The pair-wise autosomal genetic distance determined by Fixation Index ($F_{ST}$) using 33,067 SNPs was calculated for the 9 reference groups and show in Table S6 [27]. Subsequently, a conditional test was used to estimate the admixture proportions in the unknown samples as described by Alexander et al. (2009).

**Principal Component Analysis (PCA)**

PCA was applied to account for population stratification among the European subgroups. We selected the previously identified 33,067 SNPs to infer the axes of variation using EIGENSTRAT [28]. Only the top 10 eigenvectors were analyzed. Most of the variance among the European populations was observed in the first and second eigenvector. The first eigenvector accounts for the east-west European geographical variation while the second accounts for the north-south component. Only the top 10 eigenvectors showed population differences using Anova statistics ($p<0.01$). We then calculated the PCA adjusted Armitrage trend P-values using the top 10 eigenvectors as covariates.

**Power Analysis**

Power calculations were performed using QUANTO (ver. 1.2), using a log-additive model. The analysis included 2019 cases and 14471 controls with the following assumptions: Type I error $= 0.05$, a minor allele frequency $= 0.10$ and the odds-ratio $= 1.2$.

**Association Analysis**

After the quality of all data was confirmed for accuracy, genetic association was determined using the whole-genome association analysis toolset, PLINK (ver. 1.07) [29].

Differences in allele frequencies between endometriosis patients and population controls were tested for each SNP by a 1-degree-of-freedom Cochran-Armitrage Trend test.

The allelic odds ratios were calculated with a confidence interval of 95%. SNPs that passed the quality control parameters were used to calculate the genomic inflation factor ($\lambda$) as well as to generate Quantile-Quantile (QQ) plots (Figure S2), which were generated by ranking a set of $-\log_{10} p$-values and plotting them against their expected values. PCA adjusted Cochran-Armitrage test $p$-values were also determined. The combined/meta-analysis of discovery and replication dataset was performed using Cochran-Mantel-Hanszel method. Breslow Day test was used to determine between-cluster heterogeneity in the odds ratio for the disease/SNP association. Multivariate Logistic regression was used to test for independence of SNP effects. Univariate Logistic regression was used to test for correlation of clinical factors to the severity of the disease.

Control samples include both male and female samples in approximately equal proportions. The allele frequencies for the 8 strongly associated SNPs and the 15 SNPs with suggested associations did not show any significant gender bias.

Haplotype-based association tests were calculated by 1-degree of freedom $\chi^2$-test, along with their respective odds ratios using PLINK.

The variance explained by logistic regression model is calculated using the Cox Snell and Nagelkerke pseudo $R^2$ method which is similar to the $R^2$ concept of linear regression [30].

**Imputation Analysis**

IMPUTE2 (ver. 2.2.2) was used for imputing SNPs against the 1000-Genome (version 3 of the Phase 1 integrated data). Samples
were pre-phased with IMPUTE2 using actual genotypes and then imputed for SNPs included in the 1000-Genome reference panel to form imputed haplotypes. Imputation was carried out within +/− 250 kb of the main marker of interest. Only SNPs that pass the confidence score of > 0.9 from imputation, call rate of 0.95 and with MAF>0.01 are reported. The imputation was performed on the total dataset of 2,019 cases and 14,471 control subjects.

Software Used
PLINK (version 1.07; http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml), QUANTO (version 1.2; http://hydra.usc.edu/gxec), Impute2 (version 2.15.0; http://www.r-project.org/).
LocusZoom (version 1.1; http://csg.sph.umich.edu/locuszoom/) was used for regional association plots.
EIGENSTRAT (version 3.0; http://genepath.med.harvard.edu/reich/Software.htm).

Supporting Information
Figure S1 PCA classification of Case and Control samples. A reference set of samples previously identified as European are shown in Panel A (reference). Samples selected for being 95% European are projected onto the European map and shown in Panel B. The Figure show that our Case and Control populations are geographically identical. A preponderance of the participants have ancestral roots in the north-western part of Europe with a Southern trend towards Italy.

Figure S2 Quantile-quantile plots for the Discovery set of 1,514 endometriosis cases and 12,660 population controls before and after PCA-based adjustment. The unadjusted QQ plot in Panel A is showing a λ = 1.18. The adjusted QQ plot in Panel B show λ = 1.05. The association analysis include 580,699 SNPs and included only samples that passed our Ethnicity, SNP and Sample quality filters.

Figure S3 Regional association plots for the five top regions (Panel A–E).

Figure S4 Genotype clusters for the 8 most strongly associated SNPs.

References
1. Authors state that “it is still unclear whether these GWAS loci [that were implicated in GWASs of European-ancestry populations] can be generalized to other ethnic groups, such as African Americans.” Do you agree or disagree with this assertion? Explain why or why not.

2. Studies have shown that obesity is highly genetic, with heritability estimates ranging from 40-70%. Furthermore, the prevalence of worldwide obesity has increased sharply over recent decades. Are these two observations conflicting? Does the recent increase in obesity indicate recent changes in the genetic composition of the human population, worldwide?

3. Investigators performed genetic imputation for the WHI SHARE cohort. In brief, 1,962 women serving as a reference panel were genotyped for both (1) the Metabochip custom panel of SNPs at putative BMI loci, and (2) the (less-dense) Affymetrix 6.0 platform. Another 6,326 women were genotyped only for the Affymetrix platform. Imputation of the Metabochip SNPs was performed for the 6,326 women. Explain the logical basis of imputation, what the investigators stood to gain from this analysis, and whether you think this was a good approach? What could the investigators have done to assess the accuracy of the imputed genotypes?

4. Authors used the p-value threshold of $5.8 \times 10^{-5}$ for statistical significance. This is a considerably smaller p-value threshold than 0.05. Explain why investigators used the stricter p-value threshold.

5. Authors were careful to model population structure (i.e., principal components of ancestry). Given that their samples were all African American, do you think this approach was necessary?

6. The investigators claim that differences in the linkage disequilibrium patterns between European ancestry and African ancestry population can aid in narrowing the association signals. Explain the logic behind this approach. What is the purpose of narrowing the association signal? Were investigators successful in this effort?

7. The investigators repeated their analysis in sex-stratified samples. Why did they do this?
Fine Mapping and Identification of BMI Loci in African Americans

Jian Gong,¹,∗ Fredrick Schumacher,⁹ Unhee Lim,⁸ Lucia A. Hindorff,¹⁰ Jeff Haessler,¹ Steven Buyske,⁴,⁵ Christopher S. Carlson,¹ Stephanie Rosse,¹ Petra Bůžková,¹⁹ Myriam Fornage,²³ Myron Gross,²⁴ Nathan Pankratz,²⁴ James S. Pankow,²⁵ Pamela J. Schreiner,²⁵ Richard Cooper,¹³ Georg Ehret,¹¹,¹² C. Charles Gu,¹⁷ Denise Houston,¹⁸ Marguerite R. Irvin,¹⁶ Rebecca Jackson,⁶ Lew Kuller,⁷ Brian Henderson,⁹ Iona Cheng,²⁷ Lynne Wilkens,⁸ Mark Leppt,¹⁵ Cora E. Lewis,¹⁴ Rongling Li,¹⁰ Khanh-Dung H. Nguyen,¹¹ Robert Goodloe,²² Eric Farber-Eger,²² Jonathan Boston,²² Holli H. Dilks,²² Marylyn D. Ritchie,²⁰ Jay Fowke,²¹,²² Loreall Pooler,⁹ Misa Graff,³ Lindsay Fernandez-Rhodes,³ Barbara Cochrane,²⁶ Eric Boerwinkle,²³ Charles Kooperberg,¹ Tara C. Matise,⁴ Loic Le Marchand,⁸ Dana C. Crawford,²² Christopher A. Haiman,⁹ Kari E. North,²,³ and Ulrike Peters¹,∗

Genome-wide association studies (GWASs) primarily performed in European-ancestry (EA) populations have identified numerous loci associated with body mass index (BMI). However, it is still unclear whether these GWAS loci can be generalized to other ethnic groups, such as African Americans (AAs). Furthermore, the putative functional variant or variants in these loci mostly remain under investigation. The overall lower linkage disequilibrium in AA compared to EA populations provides the opportunity to narrow in or fine-map these BMI-related loci. Therefore, we used the Metabochip to densely genotype and evaluate 21 BMI GWAS loci identified in EA studies in 29,151 AAs from the Population Architecture using Genomics and Epidemiology (PAGE) study. Eight of the 21 loci (SEC16B, TMEM18, DHX34, GNPDA2, TEAP2B, BDNF, FTO, and MC4R) were found to be associated with BMI in AAs at 5.8 × 10⁻⁵. Within seven out of these eight loci, we found that, on average, a substantially smaller number of variants was correlated (r² > 0.5) with the most significant SNP in AA than in EA populations (16 versus 55). Conditional analyses revealed GNPDA2 harboring a potential additional independent signal. Moreover, Metabochip-wide discovery analyses revealed two BMI-related loci, BRE (rs116612809, p = 3.6 × 10⁻⁸) and DHX34 (rs4802349, p = 1.2 × 10⁻⁷), which were significant when adjustment was made for the total number of SNPs tested across the chip. These results demonstrate that mapping in AAs is a powerful approach for both narrowing in on the underlying causal variants in known loci and discovering BMI-related loci.

Introduction

Obesity (MIM 601665) is a major risk factor for a number of chronic diseases, such as type 2 diabetes (MIM 125853), hyperlipidemia (MIM 144250), cardiovascular diseases, and several cancer types.¹,² Worldwide obesity prevalence has nearly doubled since 1980, and in 2008 more than 1.4 billion adults worldwide were obese. In the United States, more than one-third of adults (35.7%) were obese in 2010.

Studies have shown that obesity is highly heritable; heritability is estimated to fall in the range of 40%–70%.³,⁴ Genome-wide association studies (GWAS) have identified numerous loci associated with body mass index (BMI),¹⁰⁻¹⁷ a common measure of obesity. However, most of these studies were performed among European-ancestry (EA) populations. It is still unclear whether previously identified GWAS loci are population specific or whether they can be generalized to other ethnic groups, such as African Americans (AAs). Furthermore, the overall lower linkage disequilibrium (LD) patterns in AA compared to EA populations can offer opportunities to narrow in or fine-map BMI-related loci.¹⁸ This will help to reduce the number of variants for functional follow-up studies, which tend to

¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; ²Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC 27514, USA; ³Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA; ⁴Department of Genetics, Rutgers University, Piscataway, NJ 08854, USA; ⁵Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854, USA; ⁶Department of Internal Medicine, Ohio State Medical Center, Columbus, OH 43210, USA; ⁷Department of Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA; ⁸Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI 96813, USA; ⁹Department of Preventive Medicine, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA; ¹⁰Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; ¹¹Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; ¹²Division of Cardiology, Geneva University Hospital, Geneva, OH 44041, Switzerland; ¹³Preventive Medicine and Epidemiology, Loyola University, Chicago, IL 60613, USA; ¹⁴Department of Medicine, University of Alabama, Birmingham, AL 35294, USA; ¹⁵Department of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA; ¹⁶Department of Biostatistics, University of Alabama, Birmingham, AL 35294, USA; ¹⁷Department of Biostatistics, Washington University, St. Louis, MO 63110, USA; ¹⁸Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA; ¹⁹Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; ²⁰Pennsylvania State University, State College, PA 16801, USA; ²¹Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, TN 37203, USA; ²²Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN, 37203, USA; ²³The Human Genetics Center and Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; ²⁴Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA; ²⁵Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA; ²⁶Baylor College of Medicine, Houston, TX 77030, USA; ²⁷Cancer Prevention Institute of California, Fremont, CA 94538, USA

*Correspondence: jgong@fhcrc.org (J.G.), upeters@fhcrc.org (U.P.)
http://dx.doi.org/10.1016/j.ajhg.2013.08.012. ©2013 by The American Society of Human Genetics. All rights reserved.
be time and labor intensive. In addition, dense genotyping of the GWAS loci could aid the discovery of additional independent signals within the GWAS loci.

In this study, we densely genotyped 21 BMI loci identified in EA studies in 29,151 AAs from the Population Architecture using Genomics and Epidemiology (PAGE) consortium by using the Metabochip.7 We aimed to fine-map the 21 known BMI loci in the AA population and search for additional independent signals associated with BMI. For validated loci in AAs, we evaluated whether weaker LD patterns in AAs help narrow in on the underlying potential causal variants. In addition, because Metabochip was developed to test putative association signals for BMI and many obesity-related metabolic and cardiovascular traits and to fine-map established loci,9 we also conducted a Metabochip-wide discovery-oriented analysis to search for potential BMI-associated loci.

Subjects and Methods

Study Population

The National Human Genome Research Institute funds the PAGE consortium to investigate the epidemiologic architecture of well-replicated genetic variants associated with human diseases or traits.10 PAGE consists of a coordinating center and four consortia, Epidemiologic Architecture for Genes Linked to Environment (EAGLE), which is used data from Vanderbilt University Medical Center's biorepository and links it to deidentified electronic medical records (BioVU); the Multiethnic Cohort Study (MEC); the Women's Health Initiative (WHI); and Causal Variants Across the Life Course (CALIco), itself a consortium of five cohort studies—the Atherosclerosis Risk in Communities (ARIC) study, Coronary Artery Risk Development in Young Adults (CARDIA), the Cardiovascular Health Study (CHS), the Hispanic Community Health Study/Study of Latinos, and the Strong Heart Study.10

This PAGE Metabochip study included AA participants from the ARIC, BioVU, CHS, CARDIA, MEC, and WHI studies and from extended collaborations to two additional studies – GenNet and the Hypertension Genetic Epidemiology Network (HyperGEN) (Table S1, in the Supplemental Data available with this article online). The detailed description of each study can be found in the Supplemental Data. We excluded underweight (BMI < 18.5 kg/m²) and extremely overweight (BMI > 70 kg/m²) individuals under the assumption that these extremes could be attributable to data-coding errors or an underlying rare condition outside the scope of this investigation. We also limited analysis to adults (defined as having an age > 20 years). The CARDIA participants are young, and the BMI < 18.5 exclusion criterion was not applied in this cohort. All studies were approved by institutional review boards at their respective sites, and all study participants provided informed consent.

Anthropometric Measurements

For individuals from the ARIC, CHS, CARDIA, HyperGEN, GenNet, and WHI studies, BMI was calculated from height and weight measured at the time of study enrollment. For individuals from BioVU, the median height and weight across all visit years were used in BMI calculations. For individuals from MEC, self-reported height and weight were used for calculations of baseline BMI. A validation study within MEC has shown high validity of self-reported height and weight. Specifically, this study showed that BMI was underestimated on the basis of self-reports versus measured weight, but the difference was small (< 1 BMI unit) and was comparable to the findings from national surveys.11

Genotyping and Quality Control

Genotyping was performed with the Metabochip, whose design has been described elsewhere.7 In brief, the Metabochip, a custom Illumina Select genotyping array of nearly 200,000 SNP markers, is designed to cost-effectively analyze putative association signals identified through GWAS meta-analyses of many obesity-related metabolic and cardiovascular traits and to fine-map established loci.7 Metabochip SNPs were selected from the catalogs developed by the International HapMap and 1000 Genomes projects.9 More than 122,000 SNPs were included for fine mapping of 257 GWAS loci of 23 traits (including 21 BMI loci).7 For determination of the boundaries around each GWAS index SNP, all SNPs with $r^2 \geq 0.5$ with the index SNP were identified, and then initial boundaries were expanded by 0.02 cm in either direction through use of the HapMap-based genetic map. SNPs were excluded if (1) the Illumina design score was < 0.5 or (2) SNPs within 15 bp in both directions of the SNP of interest could be found with an allele frequency of > 0.02 among Europeans (CEU). SNPs annotated as nonsynonymous, essential splice site, or stop codon were included regardless of allele frequency, design score, or nearby SNPs in the primer.9 Twenty-one BMI GWAS loci identified at the time at which the Metabochip was designed were represented for signal fine mapping (Table S2).

Samples were genotyped at the Human Genetics Center of the University of Texas, Houston (ARIC, CHS, CARDIA, GenNet, and HyperGEN), the Vanderbilt DNA Resources Core in Nashville (BioVU), the University of Southern California Epigenome Center (MEC), and the Translational Genomics Research Institute (WHI). Each center genotyped the same 90 HapMap YRI (Tunuba in Ibadan, Nigeria) samples to facilitate cross-study quality control (QC), as well as 2%–3% study-specific blinded replicates to assess genotyping quality. Genotypes were called separately for each study via GenomeStudio with the GenCall 2.0 algorithm. Study-specific cluster definitions (based on samples with call rate > 95%; ARIC, BioVU, CHS, CARDIA, MEC, and WHI) or cluster definitions provided by Illumina (GenNet and HyperGEN) were used for sample calling, and samples were kept in the analysis if the call rate was > 95%. We excluded SNPs with a GenTrain score < 0.6 (ARIC, BioVU, CHS, CARDIA, MEC, and WHI) or < 0.7 (GenNet and HyperGEN), a cluster separation score < 0.4, a call rate < 0.95, and a Hardy-Weinberg equilibrium p $< 1 \times 10^{-6}$. We utilized the common 90 YRI samples and excluded any SNP that had more than 1 Mendelian error (in 30 YRI trios), any SNP that had more than two replication errors with discordant calls when comparisons were made across studies in 90 YRI samples, and any SNP that had more than three discordant calls for 90 YRI genotyped in PAGE versus the HapMap database. SNPs were excluded from the meta-analyses if they were present in less than three studies.

For ARIC, BioVU, CHS, CARDIA, MEC, and WHI combined we identified related individuals by using PLINK to estimate identical-by-descent (IBD) statistics for all pairs. When apparent pairs of first-degree relatives were identified, we excluded from each pair the member with the lower call rate. We excluded from further analysis samples with an inbreeding coefficient (F) above...
of the WHI women genotyped on the Metabochip, 1,962 women were part of the group of 8,288 WHI subjects genotyped for the WHI SNP Health Association Resource (SHARE) GWAS via the Affymetrix 6.0 platform. To improve statistical power, we imputed the Metabochip SNPs in the remaining 6,326 SHARE subjects with Affymetrix 6.0 data. Details can be found elsewhere. In brief, we first phased genotypes for the 1,962 subjects genotyped on both the Affymetrix 6.0 platform and the Metabochip and constructed haplotypes (study-specific reference panel). We then phased the haplotypes for samples genotyped on the Affymetrix 6.0 platform only and performed a haplotype->haplotype imputation on Metabochip SNPs for the 6,326 target individuals to estimate genotypes (as allele dosages). We used MACH for phasing and Minimac for final imputation. To evaluate the quality of each imputed SNP, we calculated the dosage $r^2$. We excluded imputed SNPs with $r^2 < 0.5$ for SNPs with allele frequency $< 1\%$ and with $r^2 < 0.3$ for SNPs with allele frequency $> 1\%$. Given the large reference panel and strict QC criteria, this resulted in high imputation quality. 

**WHI SHARE Imputation**

LD in the AA sample was calculated in 500 kb sliding windows via PLINK. Likewise, the Malmo Diet and Cancer Study on 2,143 controls from a Swedish population provided Metabochip LD and frequency information in Europeans to facilitate the LD pattern comparisons between AA and EA populations. We used LocusZoom plots to graphically display the fine-mapping results. SNP positions from NCBI build 37 were used, and recombination rates were estimated from 1000 Genomes Project data.

**Functional Annotation**

To inform the discussion about the underlying potential functional variants, we made functional hypotheses for each of our most significantly BMI-associated variants by compiling a list of correlated SNPs ($r^2 > 0.5$) genotyped in our AA study populations and annotating each list for potential regulatory evidence consistent with enhancers, promoters, insulators, silencers and other effects related to gene expression. Because our lead SNPs and the SNPs in strong LD with our lead SNPs were in noncoding regions, we hypothesized that the underlying biology behind the signal was likely to impact gene expression through some unknown regulatory mechanism. For each list we aligned correlated SNPs with a combined browser view of all currently available ENCODE tracks in the UCSC Genome Browser and compared each allelic region for altered transcription factor binding site (TFBS) motifs by using JASPAR and ConSite. Given that methylation patterns are highly variable, it is useful to look for, in addition to BMI-relevant tissues, histone modifications in many cell lines to identify regions that are actively regulated, meaning that histone marks are present in some but not all cell lines. Thus, in addition to adipocytes, hepatocytes, and neurons, we used ENCODE’s histone modification tracks to query a wide variety of cell lines for all available histone marks to identify SNPs falling in various regulatory regions. The DNase hypersensitivity track provided a more precise demarcation of open chromatin loci, and the ChIP-Seq TFBS track provided evidence for the binding of specific proteins. Although less specific than ChIP-Seq, JASPAR, ConSite, and HaploReg databases were used for querying a larger number of conserved TFBS and predicting alterations in predicted motifs between reference and alternate alleles. A 46-way PhastCons track in the UCSC Genome Browser was used as secondary evidence for a regulatory region, but lack of conservation did not rule out a functional candidate. The most likely functional-candidate SNPs for each locus were evaluated for statistical significance in association with BMI. The detailed list of functional annotation data sets we used is shown in Table S3.

**Results**

This study consisted of 29,151 AAs from eight studies. Study participants had an average age of 51.2 years (Table S1). Approximately 80% of study participants were women. Within and across studies, men tended to have a lower mean BMI than women. The obesity rate (BMI $\geq 30$ kg/m$^2$) ranged from 16%-46% in men and 26%-64% in women. After quality control, we tested 18,187 genetic variants across 21 BMI loci and 177,663 variants across the Metabochip.
Table 1. Association Results for the Most Significant SNP and All Previously Identified GWAS SNPs for Eight BMI-Related Loci with Significant Results in African Americans

<table>
<thead>
<tr>
<th>Region</th>
<th>Lead AA SNP on Metabochip</th>
<th>GWAS Index SNP</th>
<th>Position</th>
<th>Number of SNPs</th>
<th>Candidate Gene</th>
<th>R² Lead SNP with GWAS SNP in AAs</th>
<th>R² Lead SNP with GWAS SNP in EA Populations</th>
<th>CA in AAs</th>
<th>CAF in EA Populations</th>
<th>Effect</th>
<th>p Value</th>
<th>Rsq</th>
<th>Het P</th>
<th>I²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q25.2</td>
<td>rs543874</td>
<td>rs543874</td>
<td>177889480</td>
<td>765</td>
<td>SEC16B</td>
<td>T</td>
<td>0.75 0.81</td>
<td>−0.0110</td>
<td>2.4 × 10⁻⁹</td>
<td>NA</td>
<td>0.04</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2p25.3</td>
<td>rs6548240</td>
<td></td>
<td>636929</td>
<td>1,123</td>
<td>TMEM18</td>
<td>A</td>
<td>0.87 0.83</td>
<td>0.0130</td>
<td>1.1 × 10⁻⁷</td>
<td>0.98</td>
<td>0.96</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3q27.2</td>
<td>rs7647305</td>
<td>rs7647305</td>
<td>185834290</td>
<td>369</td>
<td>ETYS</td>
<td>A</td>
<td>0.41 0.23</td>
<td>−0.0069</td>
<td>3.2 × 10⁻⁵</td>
<td>0.95</td>
<td>0.29</td>
<td>0.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6p12.3</td>
<td>rs2744475</td>
<td></td>
<td>50784880</td>
<td>1,685</td>
<td>TFAP2B</td>
<td>C</td>
<td>0.67 0.71</td>
<td>−0.0082</td>
<td>2.8 × 10⁻⁶</td>
<td>0.98</td>
<td>0.39</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11p14.1</td>
<td>rs1519480</td>
<td></td>
<td>27675712</td>
<td>688</td>
<td>BDNF</td>
<td>A</td>
<td>0.25 0.68</td>
<td>−0.0095</td>
<td>7.8 × 10⁻⁷</td>
<td>1.00</td>
<td>0.90</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16q12.2</td>
<td>rs62048402</td>
<td></td>
<td>53803223</td>
<td>1,814</td>
<td>FTO</td>
<td>A</td>
<td>0.11 0.41</td>
<td>0.0120</td>
<td>5.1 × 10⁻⁶</td>
<td>1.00</td>
<td>0.14</td>
<td>0.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18q21.32</td>
<td>rs6567160</td>
<td></td>
<td>57829135</td>
<td>1275</td>
<td>MC4R</td>
<td>A</td>
<td>0.81 0.75</td>
<td>−0.0096</td>
<td>4.7 × 10⁻⁶</td>
<td>0.98</td>
<td>0.03</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued on next page)
Fine Mapping BMI Loci

Among the 21 BMI GWAS loci identified in EA studies, eight loci (SEC16B [MIM 612855], TMEM18 [MIM 613220], ETV5 [MIM 601600], GNPDA2 [MIM 613222], TFAP2B [MIM 601601], BDNF [MIM 113505], FTO [MIM 610966], and MC4R [MIM 155541]) displayed SNPs with significant evidence of association (Table 1; see also Table S2). The lead SNP (the most significant SNP in AAs) in each of these eight loci had a minor-allele frequency >0.05 and showed little evidence of heterogeneity. In these eight loci most GWAS index SNPs previously identified in EA GWASs (20 out of 23) had a consistent direction of the effects reported in original EA studies. Because the results of the FTO locus have been described previously, we focused on the other seven regions. Among those seven loci, the lead SNPs, rs543874 in SEC16B (p = 1.5 × 10⁻⁴), rs7647305 in ETV5 (p = 3.2 × 10⁻⁵), and rs10938397 in GNPDA2 (p = 1.7 × 10⁻⁷), were consistent with the previously identified GWAS SNP (the most significant SNP highlighted in the previous GWAS) in EA populations. For rs543874 and rs10938397, the observed effects on BMI were slightly stronger in AAs than in EA individuals (change in BMI per coded allele: 1.1% in AAs and 0.9% in EA individuals for rs543874; 1.0% in AAs and 0.8% in EAs for rs10938397), whereas they were slightly weaker in AAs than in EA individuals for rs7647305 (change in BMI per coded allele: 0.7% in AAs and 0.9% in EA individuals). The minor-allele frequency (MAF) was higher for rs543874 and rs7647305 and lower for rs10938397 in AAs than in EA individuals (Table 1). In the other four loci, the lead SNPs in AAs differed from the GWAS SNPs from EA populations. The lead SNPs in all four loci (rs6548240 in TMEM18, rs2744475 in TFAP2B, rs1519480 in BDNF, and rs6567160 in MC4R) were modestly to strongly correlated (r² ranged from 0.5–1.0) with at least one of the GWAS SNPs on the basis of LD in EA populations. However, when LD was based on AA populations, the correlation was weaker, in several cases substantially weaker (Table 1). For the 13 BMI loci that did not replicate in our AA analysis, most GWAS index SNPs (13 out of 17) from previous GWASs involving EA individuals showed effects in the same direction when results from our AA samples were compared with results from the previous EA studies.

We investigated the question of whether LD patterns in AA studies can narrow previous association signals from EA studies and found that at seven out of the eight significant loci, AA LD patterns assisted with narrowing association signals (Table 2; see also Figure S1). One of the most extreme examples was for MC4R. Among EA individuals, 107 and 119 SNPs were correlated (r² > 0.5) with the lead SNP in our analysis (rs6567160) and with the GWAS index SNPs, respectively; these SNPs represent a region spanning 184 kb and 230 kb, respectively. However, among AAs only five SNPs were correlated with rs6567160 at r² > 0.5, and these SNPs represented a region spanning 71 kb. In SEC16B, ETV5, TFAP2B, BDNF, and FTO, the number of
Table 2. Comparison of Correlation between African American and European Populations for Eight BMI-Related Loci that had Significant\(^a\) Lead SNPs

<table>
<thead>
<tr>
<th>Region</th>
<th>Gene</th>
<th>Lead SNP in AAs of PAGE</th>
<th>Region Size</th>
<th>Number of SNPs</th>
<th>Number of SNPs with (r^2 &gt; 0.5) with Lead SNP</th>
<th>Region size for SNPs with (r^2 &gt; 0.5) with Lead SNP (bp)</th>
<th>Number of SNPs with (r^2 &gt; 0.5) with GWAS Index SNPs</th>
<th>Region Size for SNPs with (r^2 &gt; 0.5) with GWAS Index SNPs (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q25.2</td>
<td>SEC16B</td>
<td>rs543874</td>
<td>180 kb</td>
<td>765</td>
<td>2</td>
<td>4,242</td>
<td>39</td>
<td>95,660</td>
</tr>
<tr>
<td>2p25.3</td>
<td>TIMM18</td>
<td>rs6548240</td>
<td>250 kb</td>
<td>1,123</td>
<td>81</td>
<td>40,820</td>
<td>104</td>
<td>53,290</td>
</tr>
<tr>
<td>3q27.2</td>
<td>ETV5</td>
<td>rs7647305</td>
<td>110 kb</td>
<td>369</td>
<td>5</td>
<td>10,348</td>
<td>28</td>
<td>49,410</td>
</tr>
<tr>
<td>4p12</td>
<td>GNPDA2</td>
<td>rs10938397</td>
<td>90 kb</td>
<td>342</td>
<td>6</td>
<td>20,482</td>
<td>5</td>
<td>20,482</td>
</tr>
<tr>
<td>6p12.3</td>
<td>TFAP2B</td>
<td>rs2744475</td>
<td>560 kb</td>
<td>1,685</td>
<td>1</td>
<td>1,128</td>
<td>44</td>
<td>15,320</td>
</tr>
<tr>
<td>11p14.1</td>
<td>BDNF</td>
<td>rs1519480</td>
<td>300 kb</td>
<td>688</td>
<td>7</td>
<td>21,942</td>
<td>42</td>
<td>22,420</td>
</tr>
<tr>
<td>16q12.2</td>
<td>FTO</td>
<td>rs6204840</td>
<td>650 kb</td>
<td>1,814</td>
<td>19</td>
<td>45,529</td>
<td>74</td>
<td>47,575</td>
</tr>
<tr>
<td>18q21.32</td>
<td>MC4R</td>
<td>rs6567160</td>
<td>360 kb</td>
<td>1,275</td>
<td>5</td>
<td>71,160</td>
<td>107</td>
<td>183,730</td>
</tr>
</tbody>
</table>

Average for number of SNPs or region size

16

26,831

55

86,208

56

115,178

\(^a\)Significance level: 0.05 divided by 866 (average number of SNPs across 21 BMI loci).

\(^b\)Only included the GWAS index SNP rs925946 because the other two GWAS SNPs, rs10767664 and rs6265, have \(r^2 < 0.15\) with the lead SNP and rs925946 in AA and EA populations.

\(^c\)GWAS index SNP rs6499640 was not included because it has \(r^2 < 0.1\) with the lead SNP and the other GWAS index SNPs in AA and EA populations.

In the Metabochip-wide analysis, we identified five SNPs in two loci (1q25.2 and 2p25.3) as being significant loci by including the lead SNP in a locus as a covariate to search for additional independent signals. In the conditional analysis, the number of correlated SNPs and the spanning region size were also limited to significant loci by the lead SNP and the spanning region size not reduce, but the number of correlated SNPs decreased (\(r^2 \sim 0.5\)). In the conditional analysis, the number of correlated SNPs and the spanning region size were also limited to significant loci by the lead SNP and the spanning region size not reduce, but the number of correlated SNPs decreased (\(r^2 \sim 0.5\)). In the conditional analysis, the number of correlated SNPs and the spanning region size were also limited to significant loci by the lead SNP and the spanning region size not reduce, but the number of correlated SNPs decreased (\(r^2 \sim 0.5\)).
Table 3. Conditional Analysis Showing a Locus with Evidence of Second Independent BMI Association Signals in African Americans

<table>
<thead>
<tr>
<th>Region</th>
<th>Gene</th>
<th>SNP Position</th>
<th>Number of Tested SNPs</th>
<th>Coded Allele</th>
<th>CAF</th>
<th>Marginal Results&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Conditional Results&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Effect</td>
<td>p Value</td>
</tr>
<tr>
<td>4p12</td>
<td>GNPDA2</td>
<td>rs186117327</td>
<td>45,101,187</td>
<td>285</td>
<td>A</td>
<td>-0.0095</td>
<td>3.7 × 10&lt;sup&gt;-7&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs10938397&lt;sup&gt;c&lt;/sup&gt;</td>
<td>45,182,527</td>
<td>-</td>
<td>A</td>
<td>-0.0099</td>
<td>1.7 × 10&lt;sup&gt;-7&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Marginal results represent results when only the single variant was in the model.

<sup>b</sup>Conditional analysis represents the result for the SNP when adjustment was made for the most significant lead SNP and vice versa.

<sup>c</sup>The most significant SNP in a locus.

<sup>d</sup>Het p: heterogeneity-test p value.

Table 4. Two BMI Loci Identified in African Americans on the Basis of Metabochip-wide Analysis

<table>
<thead>
<tr>
<th>Region</th>
<th>SNP</th>
<th>Position</th>
<th>Candidate Gene</th>
<th>Coded Allele</th>
<th>CAF in AAs</th>
<th>CAF in EA Individuals</th>
<th>Effect</th>
<th>p value</th>
<th>Rsq&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Het p</th>
<th>r&lt;sup&gt;2&lt;/sup&gt; with Lead SNP in AAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>2p23.2</td>
<td>rs116612809</td>
<td>28,301,171</td>
<td>BRE</td>
<td>A</td>
<td>0.90</td>
<td>0.001</td>
<td>-0.0151</td>
<td>3.6 × 10&lt;sup&gt;-8&lt;/sup&gt;</td>
<td>0.99</td>
<td>0.48</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>rs114584581</td>
<td>28,304,380</td>
<td>BRE</td>
<td>A</td>
<td>0.10</td>
<td>0.001</td>
<td>0.0148</td>
<td>5.9 × 10&lt;sup&gt;-8&lt;/sup&gt;</td>
<td>0.99</td>
<td>0.43</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>rs74941130</td>
<td>28,306,293</td>
<td>BRE</td>
<td>A</td>
<td>0.10</td>
<td>0.001</td>
<td>0.0148</td>
<td>6.9 × 10&lt;sup&gt;-8&lt;/sup&gt;</td>
<td>0.99</td>
<td>0.44</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>rs79329695</td>
<td>28,319,874</td>
<td>BRE</td>
<td>A</td>
<td>0.10</td>
<td>0.001</td>
<td>0.0151</td>
<td>3.7 × 10&lt;sup&gt;-8&lt;/sup&gt;</td>
<td>0.99</td>
<td>0.56</td>
<td>0.99</td>
</tr>
<tr>
<td>19q13.32</td>
<td>rs4802349</td>
<td>47,874,510</td>
<td>DHX34</td>
<td>A</td>
<td>0.48</td>
<td>0.12</td>
<td>-0.0087</td>
<td>1.2 × 10&lt;sup&gt;-7&lt;/sup&gt;</td>
<td>1.00</td>
<td>0.46</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup>Measurement of imputation accuracy, ranging from 0 (low) to 1 (high).
association (p = 0.71) with blood TG concentrations, which did not change after adjustment for BMI. Furthermore, the association between rs116612809 and BMI did not change after adjustment for TG (Table S4).

In 19q13.32/DHX34, rs4802349 is located in an intron of DHX34 and has a MAF of 0.48 in AAs and 0.12 in EA individuals. Only ten SNPs surrounding rs4802349 (±200 kb) were genotyped; none was correlated with rs4802349 in Aas, and none was associated with BMI. Previously, rs4802349 was reported to be a putative association signal for high-density lipoprotein (HDL) cholesterol at a moderate p value of 0.005.26 We examined the association between rs4802349 and blood HDL concentrations with and without BMI adjustment in ARIC and WHI (n = 11,680). We observed that the p value changed from 0.25 to 0.04 after BMI adjustment. Furthermore, the association between rs4802349 and BMI did not change after adjustment for HDL (Table S4).

Sex-Stratification Analyses
We did not observe any additional loci when we stratified the analysis by sex (Tables S5 and S6 and Figures S2 and S3). Also, the results for the SNPs that were significant in the fine-mapping analyses and the two BMI loci results were consistent across sex (p-heterogeneity ≥ 0.29, Table S7). However, as a result of the relatively small sample of men in this study, we have limited power to detect a difference-of-sex effect.

Functional Annotation
In silico analysis assessed whether each BMI-associated locus was in an “open,” transcriptionally permissive conformation, as would be expected of a functional locus. Using ENCODE data sets, we found each signal to be in a region consistent with regulatory evidence, such as active histone marks, open chromatin structure (DNase hypersensitivity), or regions experimentally shown to bind one or more transcription factors (Table S8). Bioinformatics analyses revealed that the lead SNP, rs6548240, is in a region of open chromatin that binds multiple transcription factors, and elevated levels of active histone marks associated with promoters were detected in several cell lines (Figure S4). The functional properties of TMEM18 are obscure, although a recent study indicated that TMEM18 plays a regulatory role in adipocyte differentiation and biology.27 For GNPDA2, the potential additional independent signal rs186117327 could be tagging rs7659184 (r² = 0.6 with rs186117327 in AAs), which falls in a region of open chromatin, and relative to the reference allele, the alternate allele reduces the binding affinity of GATA2 and EN1 transcription factors. In 2p23.2/BRE, bioinformatics analyses indicated that the lead SNP, rs116612809, tags another intronic SNP, rs78003529 (r² = 0.65 in AAs), which falls in a region of open chromatin with elevated enhancer histone marks and is in a region that binds both Pol2 and c-Jun. Furthermore, evidence from scans of positional weight matrices (PWMs) suggests that the alternate allele of rs78003527 has a much higher binding affinity for Nkx3 and Nkx2. In 19q13.32/DHX34, the DHX34 signal rs4802349 and two SNPs highly correlated with it are located within putative regulatory regions. rs4802349 falls within histone-modification marks associated with enhancer activity and alters the Mtf1 transcription-factor binding motif. In addition, two other functional candidates, rs2547369 and rs2341878 (r² = 0.7 and 0.6, respectively in African populations) were highly correlated with this signal and fall in regions of open chromatin having methylation patterns associated with promoter or enhancer activity. Furthermore, rs25476369 (r² = 0.65 with rs4802349 in African populations) falls in a region that binds six transcription factors, whereas rs2341878 is of particular interest because it falls in a strong promoter region in a number of relevant tissues. The function of GNPDA2, BRE, and DHX34 is unknown, which makes it difficult to link the underlying biological mechanisms of these genetic variants to BMI.

Discussion
In this study, encompassing close to 30,000 AAs, we used the Metabochip to systematically evaluate 21 BMI loci...
discovered among European descent populations in previous GWASs and to search for potential BMI loci. Eight of the 21 loci (SEC16B, TMEM18, ETV5, GNPDA2, TFAP2B, BDNF, FTO, and MC4R) were found to be associated with BMI in our AA study population. Further conditional analyses indicated that GNPDA2 contained an additional independent signal. Moreover, Metabochip-wide analyses revealed two BMI-associated loci: the BRE locus at genome-wide significance (p < 5 × 10⁻⁸) and the DHX34 locus at Metabochip-wide significance (p < 2.5 × 10⁻⁷).

Among the eight BMI loci that were significant in the AA population, the lead SNPs in SEC16B, ETV5, and GNPDA2 were the same as the GWAS SNPs identified in previous EA studies, which provided further support for the idea that the three SNPs (rs543874, rs7647305, and rs10938397) are proxies of causal variants influencing BMI. The most significant (lead) SNPs in the other five loci differed from the previously reported GWAS SNPs, although, as expected, all were moderately to highly correlated with the GWAS SNPs in European populations (r² ranged from 0.5 to 1.0). In TFAP2B and MC4R, the GWAS SNPs were not significantly associated with BMI in AAAs, suggesting that the lead SNPs in these two loci are better proxies for the underlying functional variants. We showed that the weaker LD patterns in AA populations than in EA populations substantially reduced the number of functional-variant proxies in six of the eight BMI loci that were significant in AAs (and to a limited extent in TMEM18). Our results illustrate the important contribution of AAs to systematic fine-mapping of GWAS loci originally reported in EA populations. In addition, the lead SNPs that were either consistent or correlated with the GWAS SNPs at these replicated loci in EA populations might indicate that EA populations and AA populations share the underlying causal variants at these loci.

Bioinformatics analyses revealed that the lead SNP, rs6548240 in the TMEM18 locus, was the strongest functional candidate. Although rs6548240 was located in an intergenic region 31 kb downstream of TMEM18, ChIP-seq evidence indicates that CTCF looping might anchor this distant putative enhancer to the promoter of TMEM18. Taken together, these pieces of evidence make rs6548240 an interesting functional candidate for future laboratory follow-up. This example shows that combining fine mapping with bioinformatics analysis can help to narrow in on the putative functional variants for further follow-up studies.

There are multiple reasons that 13 BMI loci originally identified in EA populations were not significantly associated with BMI in this sample of AAs. Limited statistical power could be an important reason for this observation. Statistical power is impacted by the variance of BMI, MAF, effect size, and sample size. Compared to populations in some large European-focused studies, our AA population had larger variance in BMI (standard deviation: 6.2 kg/m² versus 4.2 kg/m²), which reduces the statistical power. Also compared with primarily very large European-focused studies with sample sizes from ~90,000 to ~250,000, our study was relatively small. Nonreplication might arise because of different causal variants between EA individuals and AAs, a weak AA LD pattern, which leads to weak correlation between causal variants and marker SNPs on the Metabochip, and limited statistical power. All of this might explain that we did not observe all loci significantly associated with BMI in AAs, and it emphasizes the need for a larger sample size. We used a uniform p value threshold of 5.8 × 10⁻⁵ (0.05/average number of SNPs per locus) as an approximate correction for the average of 866 SNPs across these loci. However, if we use Bonferroni correction (i.e., 0.05/the number of SNPs at a given locus) or correction by the effective number of SNPs at each locus after accounting for LD patterns, one more locus, SH2B1 (MIM 608937), would be indicated to be significant in AAs (Table S2).

Conditional analysis indicated that the GNPDA2 locus contained a potential additional independent signal. The potential additional independent signal, rs186117327, is physically relatively close (80 kb) to the lead SNP,
rs10938397, but is only weakly correlated ($r^2 = 0.11$) with it.

We identified two BMI-related loci, 2p23.2/BRE and 19p13.32/DHX34, in the Metabochip-wide analysis. We evaluated whether African ancestry plays role in these two BMI-related loci among 8,310 AAs from a GWAS in WHI. We didn’t observe a significant effect of ancestry at these two loci ($p > 0.5$; Table S9). At 2p23.3/BRE, the most significant SNP, rs116612809, reached the conventional genome-wide significance level ($5 \times 10^{-8}$) and was surrounded (3–18 kb) by three highly correlated SNPs ($r^2 = 0.98–1.00$) showing very similar results. All four SNPs are common (MAF = 0.1), but they all are rare variants in EA populations (MAF = 0.001) and hence it is unlikely that these SNPs would be identified in EA GWASs unless they have much stronger effects than the moderate effects observed in our study or are tested in very large sample sets. The lead SNP in 2p23.3/BRE (rs116612809) is located within an intron of BRE, which is stress responsive and highly expressed in brain and reproductive organs. No previous study reported an association with BMI or any other trait; however, in support of our finding, the GIANT consortium reported that 82 of 200 SNPs within BRE had a $p$ value < 0.05 (min. $p = 2.5 \times 10^{-4}$) for a BMI association in EA populations. Accordingly, our results showed that varying allele frequency in different ancestral groups significantly contributes to the statistical power and that studying different ancestral groups helps to identify potentially functional loci.

We found another potential BMI-related locus at 19q13.32/DHX34. The SNP rs4802349, located in an intron of DHX34, was marginally significantly associated with BMI at a Metabochip-wide significance level ($p = 1.2 \times 10^{-7}$), but because the finding did not reach the conventional genome-wide significance level of $5 \times 10^{-8}$, additional replication studies in AAs are warranted. DHX34 is a putative RNA helicase and has not been reported to be associated with any diseases or traits in humans before, except for a suggestive association with HDL cholesterol ($p = 0.005$). A recent GWAS in EA populations showed an association with BMI in a neighboring region, 19q13.32/ZC3H4; rs3810291 was the most significant SNP. However, rs4802349 is physically 300 kb away from rs3810291, and a recombination hotspot lies between ZC3H4 and DHX34 in AAs and Europeans (HapMap phase II YRI and CEU), resulting in a low correlation between both SNPs in African populations ($r^2 = 0.04$) and Europeans ($r^2 < 0.001$). This observation suggests that these two associations are independent from each other. Our results along with the evidence from functional annotation warrant additional validation studies for this locus.

In conclusion, we observed that eight BMI-associated GWAS loci identified from EA populations were significantly associated with BMI in AAs and identified a potential additional independent signal in one locus. In addition, we discovered two potential BMI-related loci through Metabochip-wide analysis; one of these loci which reached the conventional genome-wide significance level. Importantly, our study demonstrated that fine mapping in AA populations in combination with bioinformatics analyses is a valuable and effective way to narrow in on the underlying causal variants in GWAS loci discovered in EA populations and that studying minority populations can contribute to loci discovery.

Supplemental Data
Supplemental data include four figures, nine tables, and supplement text, including acknowledgments and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments
See Supplemental Data.

Web Resources
The URLs for data presented herein are as follows:
Center for Disease Control and Prevention (CDC), Adult Obesity Facts, http://www.cdc.gov/obesity/data/adult.html
MACH, http://www.sph.umich.edu/csg/abecasis/mach
Minimac, http://genome.sph.umich.edu/wiki/Minimac
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org

References


Example qualifying exam (Ph.D.) and comprehensive exam (M.S.) questions pertaining to Jordan et al. (2012) *American Journal of Human Genetics.*

1. The burden test and variable threshold test (both of which collapse information on rare variants across a genomic region) were used to test rare variants in CARD14 exons for association with psoriasis. What is the rationale for such tests? Why did investigators not test each rare variant individually?

2. The common variant rs11652075 (see Figure 2A) was tested for genetic association with psoriasis in six individual cohorts, as well as across all six cohorts simultaneously using meta-analysis. The odds ratios in the individual cohorts ranged from moderately protective (OR=0.66) to weakly harmful (OR=1.09) and associations were statistically significant in some, but not all, cohorts. However, the odds ratio of the meta-analysis was protective (OR=0.87) and very significant (p=0.000002). Interpret the meta-analysis results in light of the findings from individual cohorts. Is the meta-analysis inconsistent with the individual cohorts? Why or why not?

3. Consider the “common disease, common variant” hypothesis. What is the premise of this hypothesis, and do the results from this paper support or refute this hypothesis an explanation for genetic nature of psoriasis?

4. Consider Table 2, which shows for each CARD14 coding variant (1) the bioinformatically-predicted effect on protein function, (2) the effect on NF-kB activation, and (3) the allele frequencies in cases and controls. Suppose that the bioinformatics prediction of variant’s function was “benign” and that this prediction was accurate. What would you expect for the variant’s effect on NF-kB activation? What would you expect for the relative allele frequencies in cases and controls?

Suppose instead that the bioinformatics prediction of a variant’s function was “damaging” and that this prediction was accurate. What would you expect for the variant’s effect on NF-kB activation? What would you expect for the relative allele frequencies in cases and controls?

Scanning Table 2, are there any coding variants for which functional prediction is inconsistent with effect of NF-kB activation, or relative allele frequencies in cases vs. controls, or both?

5. The effect of CARD14 variants on NF-kB activity was measured both with and without TNF-alpha stimulation. Explain the rationale for this.
Psoriasis is a common inflammatory disorder of the skin and other organs. We have determined that mutations in CARD14, encoding a nuclear factor of kappa light chain enhancer in B cells (NF-kB) activator within skin epidermis, account for PSORS2. Here, we describe fifteen additional rare missense variants in CARD14, their distribution in seven psoriasis cohorts (>6,000 cases and >4,000 controls), and their effects on NF-kB activation and the transcriptome of keratinocytes. There were more CARD14 rare variants in cases than in controls (burden test p value = 0.0015). Some variants were only seen in a single case, and these included putative pathogenic mutations (c.424G>A [p.Glu142Lys] and c.425A>G [p.Glu142Gly]) and the generalized-pustular-psoriasis mutation, c.413A>C (p.Glu138Ala); these three mutations lie within the coiled-coil domain of CARD14. The c.349G>A (p.Gly117Ser) familial-psoriasis mutation was present at a frequency of 0.0005 in cases of European ancestry. CARD14 variants led to a range of NF-kB activities; in particular, putative pathogenic mutations from these three mutations lie within the coiled-coil domain of CARD14. The c.349G>A (p.Gly117Ser) familial-psoriasis mutation was present at a frequency of 0.0005 in cases of European ancestry. CARD14 variants led to a range of NF-kB activities; in particular, putative pathogenic variants led to levels >2.5x higher than did wild-type CARD14. Two variants (c.511C>T [p.His171Asn] and c.536G>A [p.Asp179His]) stimulated tumor necrosis factor alpha (TNF-a) to achieve significant increases in NF-kB levels. Transcriptome profiling of wild-type and variant CARD14 transfectants in keratinocytes differentiated probably pathogenic mutations from neutral variants such as polymorphisms. Over 20 CARD14 polymorphisms were also genotyped, and meta-analysis revealed an association between psoriasis and rs11652075 (c.2458C>T [p.Arg820Trp]), p value = 2.1 x 10^-5. In the two largest psoriasis cohorts, evidence for association increased when rs11652075 was conditioned on HLA-Cw*0602 (PSORS1). These studies contribute to our understanding of the genetic basis of psoriasis and illustrate the challenges faced in identifying pathogenic variants in common disease.

**Introduction**

Psoriasis is a chronic, inflammatory disease of the skin and other organs. It affects approximately 2% of individuals of European descent, and in up to 30% of cases, it is associated with chronic inflammatory psoriatic arthritis. Genome-wide association studies (GWASs) have identified over 20 susceptibility loci for psoriasis. However, with the exception of psoriasis susceptibility locus 1 (PSORS1 [MIM 177900]), for which the odds ratio (OR) is approximately 3.0, risk conferred by these loci is generally small (ORs < 1.5). Moreover, less than 20% of disease variance has been explained. This implies that additional low-risk loci, genetic interactions, or rare variants of large effect account for the remaining variance.

In our accompanying paper, we identified rare, gain-of-function mutations in caspase recruitment domain family, member 14 (CARD14 [MIM 607211]) in two large multiplex families affected by Mendelian forms of psoriasis and psoriatic arthritis (see the accompanying paper in this issue of AJHG). We also identified a de novo mutation in CARD14 in a child with early-onset, severe pustular psoriasis (PSORP [MIM 614204]). These mutations are responsible for the elusive psoriasis susceptibility locus 2 (PSORS2 [MIM 602723]) in chromosomal region 17q25. These results led us to hypothesize that additional rare and common variants in CARD14 might contribute to psoriasis and/or psoriatic arthritis in the general population.

Here, we identify and characterize 15 additional rare missense variants within CARD14 and determine their frequencies in a large cohort of approximately 6,000 psoriasis cases and 4,000 controls. Statistical analyses revealed an excess of rare variants in psoriasis cases relative to controls. The potential pathogenicity of variants was demonstrated by their ability to increase transcriptional activation by nuclear factor of kappa light chain enhancer in B cells (NF-kB) and to enhance production of a subset of psoriasis-associated transcripts. A common missense
variant within CARD14 was also associated with psoriasis, and that evidence for association increased when this locus was conditioned on the presence of PSORS1. Our findings indicate that a range of NF-κB responses in the skin are mediated by CARD14 and that a subset of rare CARD14 variants leads to psoriasis and psoriatic arthritis.

Subjects and Methods

Subjects

Cases and controls for sequencing and genotyping were recruited from multiple institutions. Samples were organized into cohorts as shown in Table 1. There were six cohorts of European ancestry and one of Asian ancestry. Referring to Table 1, European cases in cohort A were recruited from either Washington University in St. Louis or the Department of Dermatology at the University of California, San Francisco (UCSF). Controls in cohort A were unaffected individuals who were over 20 years of age and who had no family history of psoriasis; they were recruited from the Texas Scottish Rite Hospital for Children or from the Cardiovascular Research Institute and Center for Human Genetics at the University of California, San Francisco or they were CEU (Utah residents with Northern and Western European ancestry from the CEPH collection) grandparents. Cohort B samples were from the National Psoriasis Foundation Victor Henschel Tissue Repository (NPF). Cases and controls in cohort C were recruited from the Department of Dermatology at the University of Utah. For cohorts A–C, psoriasis was diagnosed by a dermatologist.

Cohort D samples were recruited from the Department of Dermatology at the University of Michigan. Cases had at least two psoriatic plaques or a single plaque occupying at least 1% of the total body surface outside the scalp. Individuals presenting with only palmar/plantar psoriasis, inverse psoriasis, or seboporiasis were excluded. Controls were at least 18 years of age and had no personal or family history of psoriasis.

Cohort E samples were gathered from the University of Toronto and Toronto Western Hospital. Cohort F samples were gathered from the Department of Medicine, Division of Rheumatology, Memorial University of Newfoundland. Psoriasis was diagnosed by a dermatologist. When psoriatic arthritis was suspected, cases were evaluated according to clinical history and rheumatologic and radiologic evaluation. Control individuals showed no evidence of psoriasis, psoriatic arthritis, or any other autoimmune disease.

Asian samples in Cohort G were recruited from the Cardiovascular Research Institute and Center for Human Genetics at the University of California, San Francisco, from the University of Toronto and Toronto Western Hospital and the Department of Medicine, or from the NPF.

DNA was isolated from whole blood or saliva by standard methods. Protocols were approved by local institutional review boards. All subjects or their parents (if the subjects were minors) provided informed consent.

Sanger Resequencing and Genotyping

As a first pass, all coding exons of CARD14 (full-length, CARD14fl) were resequenced in 192 psoriasis cases and 96 controls of European ancestry. Exons in which rare missense mutations were identified were resequenced in 95 more controls of European ancestry.

Expression Plasmids

Full-length CARD14sh (GenBank BC018142, coding for 740 amino acids) and CARD14cl (RefSeq NM_052819, coding for 434 amino acids) were cloned into pReceiver-M11 (Capital Biosciences). The CARD14sh construct was subjected to in vitro mutagenesis with the QuikChange Site-Directed Mutagenesis Kit (Stratagene). The numbering of all CARD14 mutations in this manuscript is based on RefSeq NM_024110.3. For rare, nonannotated missense variants, constructs were generated with the mutant allele. For polymorphisms (rare and common), constructs were generated with the allele and amino acid used are listed in parentheses: rs11558260 (c.185A [p.Gln185]), p.Ser200Asn (c.599A [p.Asn200]), rs146214639 (c.449G [p.Arg150]), rs144475004 (c.526C [p.His176]), rs2066964 (c.930C [p.Ser547]), rs34367357 (c.1042A [p.Ile348]), rs117918077 (c.2044T [p.Arg682]), and rs151159061 (c.2140A [p.Ser714]). Full-length CARD14fl was not available for subcloning. As a result, constructs could not be created for rs144285237 (c.2919C>G [p.Asp973Gln]) or rs11652075 (c.2458C>T [p.Arg820Trp]).

NF-κB Luciferase Reporter Assay

The NF-κB luciferase reporter assay was performed with the pNFkB-luc system (Clontech) as described in our accompanying manuscript. The CARD14cl clone was used as a negative control in this assay because it lacks the sequence encoding the CARD domain, which is necessary for CARD14-induced NF-κB activation.
Expression Profiling and qRT-PCR
HEK 001 cells (human-papillomavirus-16-transformed keratinocytes) were transfected with wild-type CARD14sh or expression constructs encoding CARD14sh substitutions. Cells were cultured for 24 hr, and then RNA was isolated with the miRNEasy kit (QIAGEN). Global expression profiling of RNA from cells was performed with the HumanHT-12 v4 Expression BeadChip (Illumina). Experiments were conducted in compliance with MIAME (minimum information about a microarray experiment) guidelines. Raw and normalized expression data are deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) as series GSE36381. 100 ng to 1 μg of randomly primed total RNA was used for quantitative RT-PCR (qRT-PCR) according to standard procedures. Expression levels were normalized to 18S rRNA. Relative expression levels were calculated as follows: 2^(-ΔΔCt −ΔCt CARD14). We also normalized expression levels in transfected cells to levels of FLAG to correct for differences in transfection efficiency.

Clustering of Variants on the Basis of Expression Levels in Keratinocytes
On the basis of the NF-κB reporter assays, a subset of CARD14 variants were classified as (1) leading to enhanced basal NF-κB activation when they were compared to the effects of wild-type CARD14sh (these are c.349G>A [p.Gly117Ser], c.413A>C [p.Glu138Ala], c.424G>A [p.Glu142Lys], and c.425A>G [p.Glu142Gly]), (2) leading to downregulation of NF-κB activation (c.112C>T [p.Arg38Cys]), or (3) having no effect on NF-κB activation (these are c.185G>A [p.Arg62Gln] [rs115582620], c.930G>C [p.Arg547Ser] [rs2066964], c.449T>G [p.Leu150Arg] [rs146214639], c.854A>G [p.Asp285Gly], c.1778T>A [p.Leu150Arg] [rs115582620], and c.2140G>A [p.Gly714Ser] [rs151150961]). After performing mRNA transcriptome analysis on cells that were transfected with either wild-type CARD14sh or CARD14 variants, we found that 1,531 transcripts had at least a 2-fold change (up or down) in expression as well as a significant p value in transfected cells with the CARD14 variants causing enhanced NF-κB activation relative to transfected with wild-type CARD14sh. The variants were clustered in the use of the reduced probe set and the R (v2.10.1) randomForest package in R (R v2.13.1; Biobase v2.12.2; BeadArray v2.2.0; and limma v3.8.3). The contrast was defined as “pathogenic versus nonpathogenic,” and the t tests, fold changes, and false-discovery-rate-corrected p values were calculated with LimFit and eBayes. Given the overexpression of the constructs, larger sample sizes would be required for the detection of significant group-wise effects. However, by taking the genes with a nominal group-wise p value of 0.05 and ranking them by fold change, we generated a list of the top 200 upregulated and top 200 downregulated genes. This list was analyzed with Ingenuity pathway analysis (IPA).

Statistical Analysis

Results

Rare-Variant Screening
In our accompanying manuscript, we identified rare gain-of-function CARD14 mutations that lead to psoriasis.17 These included the familial c.349G>A [p.Gly117Ser] mutation in family PS1, affected by multiple cases of psoriasis and psoriatic arthritis,21 and the de novo c.413A>C [p.Glu138Ala] germline mutation in a case of childhood generalized pustular psoriasis. To determine whether there were additional rare variants predisposing to psoriasis in CARD14, we resequenced all coding exons of CARD14 (full-length, CARD14fl) in over 192 psoriasis cases and 96 controls (see Subjects and Methods). This revealed ten rare missense mutations in CARD14 (Figure 1, Table 2, and Table S1): c.112C>T (p.Arg38Cys), c.185G>A [p.Arg62Gln] (rs115582620), c.425A>G [p.Glu142Gly], c.449T>G [p.Leu150Arg] (rs146214639), c.599G>A [p.Ser200Asn], c.854A>G [p.Asp285Gly], c.1778T>A [p.Leu593Asn], c.2044C>T [p.Arg682Trp] (rs117918077), c.2140G>A [p.Gly714Ser] (rs1151150961), and c.2919C>G [p.Asp973Gln] (rs144285237). On the basis of the observation that several rare mutations were clustered in exon 4, which encodes part of the critical coiled-coil domain, we reasoned that it might be a mutation hotspot and resequenced an additional 1,800 cases and 900 controls for that exon. This revealed five additional rare variants within CARD14: c.424G>A [p.Glu142Lys], c.511C>A [p.His171Asn], c.526G>C

Pathway Analysis of the “CARD14 Pathogenic Keratinocyte Signature”
A list was generated of genes differentially expressed as a consequence of the introduction of CARD14 psoriasis-specific alterations into HEK 001 cells as described above. We obtained this list, termed the “CARD14 pathogenic keratinocyte signature,” by comparing the global transcriptomes of keratinocyte transfecteds with wild-type CARD14sh substitutions (p.Gly117Ser, p.Glu138Ala, p.Glu142Lys, and p.Glu142Gly) to those with nonpathogenic substitutions (p.Leu150Arg [rs146214639], p.Val191Leu, p.Asp285Gly, and wild-type CARD14sh). Expression data from each sample were quantile normalized, log2 transformed, and fitted to linear models in R (R v2.13.1; Biobase v2.12.2; BeadArray v2.2.0; and limma v3.8.3). The contrast was
(p.Asp176His) (rs144475004), c.536G>A (p.Arg179His), and c.571G>T (p.Val191Leu) (Figure 1, Table 2, and Table S1). A search in dbSNP135 revealed that 8 of the 15 identified rare variants have not been previously annotated; these eight are c.112C>T (p.Arg38Cys), c.424G>A (p.Glu142Lys), c.425A>G (p.Glu142Gly), c.511C>A (p.His171Asn), c.536G>A (p.Arg179His), c.571G>T (p.Val191Leu), c.599G>A (p.Ser200Asn), c.854A>G (p.Asp285Gly), and c.1778T>A (p.Ile593Asn). Furthermore, this search revealed that the c.349G>A (p.Gly117Ser) and c.413A>C (p.Glu138Ala) mutations have also not been previously annotated.

Population-Based Frequency Estimates of Rare Variants
We determined the frequencies of all rare CARD14 variants, including those encoding the familial p.Gly117Ser alteration (causing psoriasis and/or psoriatic arthritis) and the p.Glu138Ala alteration (causing generalized pustular psoriasis), by high-throughput genotyping seven independent case/control cohorts (>10,000 individuals; Tables 1 and 2 and Tables S1 and S2). This revealed two other unrelated psoriasis cases with the CARD14 c.349G>A (p.Gly117Ser) mutation; one woman was from Utah and had a history of psoriasis, which was diagnosed at 65 years old. This latter woman transmitted the mutation to her daughter, who developed psoriasis at age 32. This mutation was also detected in a male NPF control for whom additional data (e.g., ethnicity, age, and family history of psoriasis) were not available. None of these individuals harbored the SLC26A11 mutation (solute carrier family 26, member 11, SLC26A11 (MIM 610117)) c.365A>G (p.Tyr122Cys) mutation, which cosegregated with c.349G>A (p.Gly117Ser) in family PS1,17 providing evidence that these variants can arise independently. Numbering of the SLC26A11 mutation is based on RefSeq NM_001166347.1. In total, the frequency of the CARD14 c.349G>A (p.Gly117Ser) mutation in cases of European ancestry was 0.0005. The CARD14 c.349+5G>A mutation, which segregated with psoriasis in a large multiply affected Taiwanese family,17 was not detected in any other cases or controls. However, given the relatively small number of Asian samples screened, we were not well powered to detect variants at low frequencies in that population. The mutation encoding p.Glu138Ala seen in the single pustular psoriasis case was also not seen in any other cases or controls.

Five rare CARD14 variants (encoding p.Arg38Cys, p.Glu142Gly, p.Glu142Lys, p.Val191Leu, and p.Asp285Gly; Figure 1 and Tables S1 and S2) were seen in only one case and...
Table 2. Characteristics and Frequencies of CARD14 Coding Variants

<table>
<thead>
<tr>
<th>CARD14 Exon</th>
<th>cDNA Mutation and Corresponding Protein Change</th>
<th>Protein Domain</th>
<th>PolyPhen2 27-Predicted Effect on Protein Function</th>
<th>Effect on NF-κB Activation (FC versus Wild-Type CARD14sh)</th>
<th>Allele Frequency in Cases (Number Sampled)</th>
<th>Allele Frequency in Controls (Number Sampled)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>c.112C&gt;T (p.Arg37Cys)</td>
<td>CARD</td>
<td>probably damaging</td>
<td>0.11</td>
<td>0.00019 (2,691)</td>
<td>0 (1,271)</td>
</tr>
<tr>
<td>2</td>
<td>c.185G&gt;A (p.Arg62Gln) (rs115582620)</td>
<td>CARD</td>
<td>benign</td>
<td>1.06</td>
<td>0.0014 (3,284)</td>
<td>0.00084 (1,797)</td>
</tr>
<tr>
<td>3</td>
<td>c.349G&gt;A (p.Gly117Ser)</td>
<td>none</td>
<td>possibly damaging</td>
<td>3.71</td>
<td>0.00023 (6,630)</td>
<td>0 (4,731)</td>
</tr>
<tr>
<td>3</td>
<td>c.349+5G&gt;A</td>
<td>none</td>
<td>NA</td>
<td>ND</td>
<td>0 (2,871)</td>
<td>0 (1,339)</td>
</tr>
<tr>
<td>4</td>
<td>c.413A&gt;C (p.Glu138Ala)</td>
<td>coiled-coil</td>
<td>probably damaging</td>
<td>8.95</td>
<td>0.00015 (3,488)</td>
<td>0 (1,902)</td>
</tr>
<tr>
<td>4</td>
<td>c.424G&gt;A (p.Glu142Lys)</td>
<td>coiled-coil</td>
<td>probably damaging</td>
<td>4.03</td>
<td>0.000012 (4,107)</td>
<td>0 (1,874)</td>
</tr>
<tr>
<td>4</td>
<td>c.425A&gt;G (p.Glu142Gly)</td>
<td>coiled-coil</td>
<td>probably damaging</td>
<td>5.00</td>
<td>0.00019 (2,848)</td>
<td>0 (1,451)</td>
</tr>
<tr>
<td>4</td>
<td>c.449T&gt;G (p.Leu150Arg) (rs146214639)</td>
<td>coiled-coil</td>
<td>probably damaging</td>
<td>1.79</td>
<td>0.0025 (6,140)</td>
<td>0.0016 (4,614)</td>
</tr>
<tr>
<td>4</td>
<td>c.511C&gt;A (p.His171Asn)</td>
<td>coiled-coil</td>
<td>benign</td>
<td>0.68 (5.95 with TNF-α stimulation)</td>
<td>0.00025 (4,077)</td>
<td>0 (1,858)</td>
</tr>
<tr>
<td>4</td>
<td>c.526G&gt;C (p.Asp176His) (rs144475004)</td>
<td>coiled-coil</td>
<td>probably damaging</td>
<td>2.78</td>
<td>0.00056 (3,575)</td>
<td>0.00062 (1,609)</td>
</tr>
<tr>
<td>4</td>
<td>c.536G&gt;A (p.Arg179His)</td>
<td>coiled-coil</td>
<td>probably damaging</td>
<td>1.38 (2.19 with TNF-α stimulation)</td>
<td>0.00025 (4,061)</td>
<td>0.00027 (1,848)</td>
</tr>
<tr>
<td>4</td>
<td>c.571G&gt;T (p.Val191Leu)</td>
<td>coiled-coil</td>
<td>benign</td>
<td>1.02</td>
<td>0.00014 (3,575)</td>
<td>0 (1,613)</td>
</tr>
<tr>
<td>4</td>
<td>c.599G&gt;A (p.Ser200Asn)</td>
<td>coiled-coil</td>
<td>benign</td>
<td>0.67</td>
<td>0.011 (6,163)</td>
<td>0.0084 (4,624)</td>
</tr>
<tr>
<td>6</td>
<td>c.854A&gt;G (p.Asp285Gly)</td>
<td>none</td>
<td>possibly damaging</td>
<td>1.14</td>
<td>0.00019 (2,673)</td>
<td>0 (1,467)</td>
</tr>
<tr>
<td>13</td>
<td>c.1778T&gt;A (p.Ile593Asn)</td>
<td>PDZ</td>
<td>probably damaging</td>
<td>1.30</td>
<td>0.00024 (2,049)</td>
<td>0.00048 (1,039)</td>
</tr>
<tr>
<td>15</td>
<td>c.2044C&gt;T (p.Arg682Trp) (rs117918077)</td>
<td>SH3</td>
<td>probably damaging</td>
<td>0.95</td>
<td>0.013 (2,169)</td>
<td>0.012 (1,042)</td>
</tr>
<tr>
<td>15</td>
<td>c.2140G&gt;A (p.Gly714Ser) (rs151150961)</td>
<td>SH3</td>
<td>benign</td>
<td>1.02</td>
<td>0.0021 (2,105)</td>
<td>0.0014 (1,038)</td>
</tr>
<tr>
<td>21</td>
<td>c.2919C&gt;G (p.Asp973Glu) (rs144285237)</td>
<td>GUK</td>
<td>benign</td>
<td>NDb</td>
<td>0.0024 (5,177)</td>
<td>0.0015 (4,099)</td>
</tr>
</tbody>
</table>

CARD14 missense variants are listed with details on their locations in critical CARD14 protein domains, their predicted effect on protein function from PolyPhen2,27 their effect on NF-κB activation (fold change compared to unstimulated wild-type CARD14sh; see also Figure 3), and frequencies in unrelated cases and controls of European ancestry. The number of individuals screened is in parenthesis. The following abbreviations are used: FC, fold change; and ND, not done.

1The p.Ser200 residue lies within a 6 bp sequence separating two predicted coiled-coil regions and thus could be considered part of an overarching coiled-coil domain. Also, variant rs114688446 was identified at this location in dbSNP, but the amino acid change (p.Ser200Ile) was different.

2The impact of p.Asp973Glu on NF-κB activation could not be tested because it is exclusive to CARD14fl, for which a full-length cDNA clone was unavailable.

in no controls. The variant encoding p.His171Asn was only seen in two psoriasis- and psoriatic-arthritis-affected cases from Newfoundland and in no controls. We performed a simple burden test and a variable threshold test19 to compare the distribution of rare variants in cases and controls. These tests provided evidence of an excess of rare CARD14 variants in cases versus controls (burden test p value = 0.0015; variable threshold test p value = 0.0053).

Common-Variant Association Tests
Resequencing validated several common missense polymorphisms described in dbSNP: rs2066964, rs34367357, and rs11652075. We genotyped these and 20 other previously described SNPs in our seven psoriasis case/control cohorts (Table S2) and looked for association with psoriasis. Most variants were present at similar frequencies in cases and controls. However, three missense SNPs (rs2066964, rs34367357, and rs11652075) within CARD14 and rare variant c.599G>A (p.Ser200Asn) provided evidence of association with psoriasis in some of the cohorts (Table S2). A meta-analysis of these four missense variants was performed for the six cohorts of European ancestry (Figure 2 and Figure S1). This revealed further evidence of association between psoriasis and rs11652075 (c.2458C>T [p.Arg820Trp]) (fixed effects p value = 2.1 × 10⁻⁶, OR = 0.87 [0.83–0.92]; random effects p value = 0.031, OR = 0.86 [0.75–0.99]); c.2458C was the risk allele (Figures 2A and 2B). This SNP, as well as the same risk allele, was also associated with psoriasis in Asians (p value = 0.0029, OR = 0.64 [0.48–0.86]). The meta-analysis also revealed evidence of association with p.Ser200Asn (fixed and random effects p values = 0.05, OR = 1.00).
the rare c.599A allele increased psoriasis risk (Figure 2C). However, this would not be significant if it were adjusted for multiple testing.

Because of the large effect of HLA-Cw*0602 from the major histocompatibility complex (MHC) class I region (PSORS1),12,13 we investigated its connection with the four CARD14 variants described above. In the Michigan and Utah psoriasis cohorts, rs11652075 was found to have a higher association with psoriasis when it was conditioned on HLA-Cw*0602 (p value of rs11652075 alone = 0.023 (Michigan), 0.017 (Utah); stratified on HLA-Cw*0602, p value = 0.0021 (Michigan), 0.0086 (Utah); Table S3). No such evidence was observed with the other variants.

**Effect of Variants on CARD14 Function In Vitro**

CARD14 encodes a 1,004 amino acid protein that activates NF-kB.16 In our companion paper, we observed that compared to wild-type CARD14, the familial p.Gly117Ser and de novo p.Glu138Ala substitutions lead to enhanced NF-kB activation (3.71- and 8.95-fold enhancement, respectively). To test the effect of rare variants described here on this activity, we again used an NF-kB luciferase reporter assay. Several rare variants (p.Glu142Lys, p.Glu142Gly, and p.Asp176His [rs144475004]) in the coiled-coil domain increased NF-kB activation two to five times more than did wild-type CARD14sh (Table 2, Figure 3A). The p.Arg38Cys, p.His171Asn, and p.Ser200Asn substitutions led to less NF-kB activation than did CARD14sh, but compared to wild-type CARD14sh, other rare variants and the CARD14 missense polymorphisms (p.Arg62Gln [rs115582620], p.Arg547Ser [rs2066964], and p.Arg682Trp [rs117918077]) did not significantly alter NF-kB activation levels (Table 2, Figure 3A). As discussed below, variants that increased NF-kB activation at least 2.5× more than that seen with CARD14sh also induced greater upregulation of psoriasis-associated genes. Relative to wild-type CARD14sh, variants that resulted in a more
modest increase in NF-kB activation, those that reduced NF-kB activation, and those that did not change it did not induce upregulation of those genes to the same degree. Therefore, when compared with the level of NF-kB activation caused by wild-type CARD14, a 2.5x or greater increase in NF-kB activation is predictive of putative pathogenic CARD14 amino acid substitutions.

Onset of psoriatic lesions is thought to be triggered by an inflammatory stimulus. We therefore examined the effects of wild-type and variant CARD14 on NF-kB activation after stimulation with tumor necrosis factor alpha (TNF-α).

Compared with unstimulated CARD14sh, TNF-α-stimulated p.His171Asn and p.Arg179His resulted in a 5.95- and a 2.95-fold increase in NF-kB activation, respectively; compared with TNF-α-stimulated CARD14sh, TNF-α-stimulated p.His171Asn and p.Arg179His resulted in a 2.87- and a 1.06-fold increase in NF-kB activation, respectively (Table 2 and Figure 3B, discussed further below).

Effect of CARD14 Substitutions on Keratinocyte Gene Expression

We have shown that CARD14 is localized in keratinocytes and that the familial and pustular-psoriasis variants (p.Gly117Ser and p.Glu138Ala, respectively) lead to enhanced production of some chemokines and other transcripts that are upregulated in psoriatic skin.17 To evaluate the effect of the additional rare variants on transcription in keratinocytes, we transfected all altered constructs into the keratinocyte cell line HEK 001. The transcriptome of each transfectant was then evaluated after 24 hr by interrogation with Illumina bead arrays. A heat map with 30 probes (see Subjects and Methods, random forest classification) revealed clustering of the p.Glu142Lys and p.Glu142Gly transfectants with those with the p.Gly117Ser and p.Glu138Ala alterations (Figure 4). The p.Glu138Ala substitution clustered at the extreme end of other pathogenic variants and on its own branch of the tree. This might be expected given the severity of the disease in the child (who has generalized pustular psoriasis) in whom it was found.

Within this heat map, the following 13 genes were upregulated in keratinocytes transfected with pathogenic substitutions: superoxide dismutase 2, mitochondrial (SOD2 [MIM 147460]); interleukin 6 (interferon, beta 2 [IL6 [MIM 147620]]); colony stimulating factor 2 (granulocyte-macrophage [CSF2 [MIM 138960]]); interleukin 8 ([IL8 [MIM 146930]]); matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase [MMP9 [MIM 120361]]); BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like ([BRF2 [MIM 607013]]); chemokine (C-C motif) ligand 20 ([CCL20 [MIM 601960]]); solute carrier family 7 (cationic amino acid transporter, y+ system), member 2 ([SLC7A2 [MIM 601872]]); oxidized low density lipoprotein (lectin-like) receptor 1 ([OLR1 [MIM 602601]]); interleukin 36, gamma ([IL36G [MIM 605542]]); guanylate binding protein 2, interferon-inducible ([GBP2 [MIM 600412]]); tumor necrosis factor, alpha-induced protein 2 ([TNFAIP2 [MIM 603300]]); and tumor necrosis factor ([TNF [MIM 191160]]). We used g:Profiler23 to look for specific functions of these genes. The most significant Gene Ontology24 terms associated with this group of genes included: immune-system process/development, hematopoietic- or lymphoid-organ...
development, and response to lipopolysaccharide/bacterium/molecule of bacterial origin. **IL6, MMP9**, and **BRF2** have also been implicated in cell migration and could mediate immune cell infiltration into the skin. **CSF2, IL8,** and **SLC7A2** have been specifically implicated in myeloid leukocyte activation. Psoriasis is a disease of keratinocyte and immune cell proliferation, and genes involved in cell proliferation include **SOD2, **IL6, **CSF2**,** and **IL8**.

Other terms included response to wounding (**SOD2, **IL6, **IL8, **CCL20, **SLC7A2, and **OLR1**), nitric oxide biosynthetic process (**SOD2, **IL6, **IL8, **CLCF1, **SLC7A2, and **OLR1**), and antiapoptosis (**SOD2, **IL6, and **CSF2**). **IL6** and **CSF2** are implicated in the regulation of the JAK-STAT pathway, and **IL6, **IL8, **IL8,** and **MMP9** are implicated in angiogenesis. Genes involved in the response to biotic stimuli include **SOD2, **IL6, **IL8, **MMP9,** and **CCL20**. Psoriasis increases risk of cardiovascular disease,^25^ ^26^ and five of these genes have been implicated in the development of the cardiovascular system (**SOD2, **IL6, **IL6, **MMP9,** and **TNFAIP2**). Uptregulation of these thirteen genes thus constitutes a pathogenic psoriatic signature.

Two variants clustered together: c.526G>C (p.Asp176His) (rs144475004) and c.536G>A (p.Arg179His). Both amino acid substitutions were predicted to be damaging by PolyPhen2,^27^ and both lead to upregulation of NF-kB activation (2.78- and 1.38-fold increases, respectively). In the case of p.Arg179His, TNF-α was required for NF-kB activation to achieve pathogenic levels; compared with unstimulated CARD14sh and TNF-α-stimulated CARD14sh, TNF-α-stimulated p.Arg179His produced a 2.95- and a 1.06-fold increase, respectively. Both variants were seen in a small number of cases and controls (4:2 and 2:1, respectively).

Three other constructs clustered in the same branch with p.Asp176His (rs144475004) and p.Arg179His. These included wild-type CARD14sh, c.2140G>A (p.Gly714Ser) (rs151150961), and c.599G>A (p.Ser200Asn). However, the latter three did not induce expression of the pathogenic psoriatic signature described above to the same
degree as the overtly pathogenic variants. One noteworthy example of this is that the IL8 expression produced by these three constructs was much lower than that seen with the pathogenic variants. Moreover, clustering of these constructs was due to other transcripts that are not considered part of the pathogenic signature. Other variants exhibited reduced levels of genes in this psoriasis signature, even when compared to wild-type CARD14sh.

We confirmed altered expression of all thirteen transcripts (SOD2, IL6, CSF2, BRF2, MMP9, IL8, CCL20, SLC7A2, OLR1, IL36G, GBP2, TNFAIP1, and TNF) by qRT-PCR (Figure S2).

We also performed a group-wise comparison of the global expression profiles of the overtly pathogenic substitutions (p.Gly117Ser, p.Glu138Ala, p.Glu142Lys, and p.Glu142Gly) and several nonpathogenic variants (p.Leu150Arg [rs146214639], p.Val191Leu, p.Asp285Gly, and wild-type CARD14sh). The top 200 upregulated and top 200 downregulated genes were identified (see Subjects and Methods), and pathway analysis was performed with IPA (Table S4). A number of cytokine signaling pathways were significant (including IL-17 signaling, IL-6 signaling, and TNFR2 signaling). Also significant were communication between innate and adaptive immune cells, dendritic cell maturation, mTOR signaling, notch signaling, and atherosclerosis signaling pathways. This latter pathway is interesting given the association between psoriasis and other systemic comorbidities, including cardiovascular disease.25,26

A comparison of these results with a published psoriasis transcriptome28 revealed that a number of these pathways are significantly represented in both groups. For example, the atherosclerosis signaling pathway, the NF-kB signaling pathway, and many of the cytokine signaling pathways were significant in both the published psoriasis transcriptome and the CARD14 pathogenic keratinocyte signature (Table S4). These results indicate that some of the pathways upregulated in keratinocytes in which CARD14 harbors a pathogenic substitution are also upregulated in classic psoriatic skin. This suggests that altered keratinocyte activation might significantly contribute to the transcriptome signatures in classic psoriasis lesions.

Discussion

Here, we describe a spectrum of rare and common variation within CARD14, an activator of NF-kB16 in skin epidermis, and we demonstrate enrichment of rare variants in cases by using two independent statistical tests. The burden test, which performs a straightforward comparison of the number, or “burden,” of rare variants in cases and controls, provided a p value of 0.0015. The variable threshold test,19 which compares rare variants subject to a variable allele-frequency threshold in cases and controls, gave a p value of 0.0053. We also demonstrate that pathogenic alterations were enriched in the coiled-coil domain of CARD14. This domain is predicted to be involved in the oligomerization of CARD14 with other proteins and the formation of its active conformation.29,30 Interestingly, although some of the rare variants we identified have been annotated in dbSNP135, none of the putative pathogenic alterations are annotated. A coding polymorphism in CARD14, rs11652075 (p.Arg820Trp), and c.599G>A (p.Ser200Asn) were also associated with psoriasis in several large cohorts. In the two largest psoriasis cohorts, evidence of association between psoriasis and rs11652075 increased when rs11652075 was conditioned on PSORS1.

Two rare variants, c.424G>A (p.Glu142Lys) and c.425A>G (p.Glu142Gly), were identified in cases but not in controls and manifested as overtly causing of disease. Compared with wild-type CARD14sh, they significantly enhanced NF-kB activation (4.03- and 5.00-fold enhancement, respectively), and they clustered with p.G117Ser and p.Glu138Ala after transfection into the keratinocyte line HEK 001 and global expression profiling. The c.424G>A (p.Glu142Lys) variant was identified in a Caucasian male who was diagnosed with psoriasis at 42 years of age and who responded well to treatment with UV light and a topical mixture of corticosteroid and a vitamin D analog. The c.425A>G (p.Glu142Gly) variant was found in a Caucasian male who was diagnosed with psoriasis in infancy and whose father also had psoriasis. He experienced a partial remission of psoriasis with methotrexate treatment. It is noteworthy that after these variants were stimulated with TNF-α, levels of NF-kB activation induced by these variants and the pustular-psoriasis substitution, p.Glu138Ala, decreased at the 24 hr mark. This suggests that at this time, downregulation of the NF-kB response might have been initiated in our cell-culture system, which merits further study. Both of these variants lie in the coiled-coil domain of CARD14, as does the de novo pustular-psoriasis substitution, p.Glu138Ala.

Compared with wild-type CARD14sh, a third variant, p.Asp176His (rs144475004), leads to enhanced NF-kB activation. However, its frequency was similar in cases and controls, and it didn’t increase NF-kB activity in vitro as much as other variants did. Hence, it might lie below the NF-kB-activation threshold required for disease. Alternatively, disease might require a specific stimulus or interaction with a second genetic factor. Other variants such as p.Arg38Cys and p.Ser200Asn exhibited significantly less NF-kB activation than did wild-type CARD14sh. Previous studies have shown that decreased activation of NF-kB, much like increased activation, can induce inflammation and epidermal hyperplasia.31,32 It might be interesting to examine clinical features, such as inflammation after skin wounding, of individuals with these variants. However, it should be noted that the p.Arg38Cys and p.Ser200Asn substitutions did not induce expression of the pathogenic psoriasis signature when transfected into keratinocytes. Thus, although we cannot completely rule out a role for these variants in some aspects of disease, they are neither...
Two variants, p.His171Asn and p.Arg179His, required stimulation with TNF-α to achieve maximal levels of NF-kB activation. The p.Arg179His substitution was observed in two unrelated cases from Toronto and one control. The cases included a female who was diagnosed with psoriasis at 40 years of age and a male who was diagnosed with psoriasis at 64 years of age and who had a family history of psoriasis. The female responded well to oral and topical steroids, but the male was not treated.

The p.His171Asn alteration was seen in two unrelated psoriasis- and psoriatic-arthritis-affected individuals from Newfoundland and was not seen in controls. One individual was diagnosed with psoriasis at 40 years of age and psoriatic arthritis at 41 years of age and had a family history of psoriasis. The second individual was diagnosed with psoriasis at 55 years of age after being diagnosed with psoriatic arthritis at 53 years of age. The identification of the c.511C>A (p.His171Asn) variant in only the Newfoundland population suggests that it arose as a result of a founder effect in this population, but confirming this will require further studies. Inspection of the genotypes of 13 polymorphisms in a 75 kb region harboring CARD14 revealed a shared haplotype that is common in cases and controls from the Newfoundland cohort and our other case/control cohorts from the United States (e.g., St. Louis/Dallas/UCSF, Michigan, and Utah). Therefore, further evidence of a founder effect for the p.His171Asn substitution seen in these two cases was not possible.

The altered transcriptome signature with pathologic substitutions included upregulation of psoriasis-specific transcripts SOD2, IL6, CSF2, IL8, MMP9, BRF2, CCL20, SLC7A2, OLR1, IL36G, GBP2, TNAIP2, and TNF. Expression of these molecules is expected to be an early event in the pathogenesis of psoriasis. Many of these genes have been implicated in immune-system development. However, BRF2 is implicated in the development of squamous cell carcinoma of the lung. This suggests that it might have global effects on the transcriptional profile of squamous cell epithelia in general and might help elicit a wound-healing or regenerative response in psoriatic keratinocytes.

Despite the dramatic effects of some CARD14 variants as keratinocyte transfectants, there was a wide range of phenotypes, even among individuals who carried the same substitution. This suggests that in many instances, the variable phenotypes are likely to be due to genetic background and/or environmental factors. For example, affected members of family PS1 all harbor the c.349G>A (p.Gly117Ser) mutation, but they have variable ages of onset (ranging from infancy to 83 years of age) and variable levels of disease severity, including the presence of psoriatic arthritis. Similarly, age of onset and response to treatment differed among individuals with the putative pathogenic variants from the coiled-coil domain (p.Glu142lys, p.Glu142Gly, and p.Glu138Ala).

However, there might be some genotype-phenotype correlations because the pustular-psoriasis substitution, p.Glu138Ala, led to the most severe phenotype (in terms of both clinical presentation and increased NF-kB activation) relative to that produced by wild-type CARD14sh. The child with this alteration presented with a spectrum of plaque-type lesions, but she mostly presented with pustular lesions. This implies that some forms of plaque psoriasis might be pathogenetically linked to pustular psoriasis at the severe end of the disease spectrum. The child’s lesions also exhibited a pronounced infiltration of neutrophils. Interestingly, in keratinocyte transfectants with this CARD14 substitution, there was a higher level of IL8 than there was with other variants, which could lead to higher levels of neutrophil infiltration. The observation that the p.Glu138Ala alteration led to the most severe clinical phenotype and induced the greatest increase in NF-kB activation and upregulation of psoriasis-associated transcripts suggests that the phenotype of psoriasis could, in some cases, be predicted by the detection of pathogenically increased levels of NF-kB activation and signaling. How the CARD14 substitutions translate to variable levels of chemokine activation and how genetic background and environment trigger variable phenotypes are important areas for further study.

CARD14 missense variants rs11652075 (c.2458C>T [p.Arg820Trp]) and c.599G>A (p.Ser200Asn) were associated with psoriasis in cohorts of European ancestry. In the Asian cohort, the c.2458T>C polymorphism was also associated with psoriasis, but c.599G was monomorphic. CARD14 was not shown to be associated with psoriasis in a previous GWAS. However, none of the polymorphisms with evidence of association with psoriasis in this study were included in that GWAS and would not have been likely to exceed genome-wide significance if they had been. From the PSORS2 region of linkage, the SNP with the most significant p value in the GWAS was rs7216577 (p value = 0.00261). That SNP is located within an intron of SLC26A11 and might regulate levels of CARD14 mRNAs in the skin. However, rs11652075, although not on the Perlegen microarray in the CASP-GWAS, was part of the HapMap-based imputed dataset that was used for the association analyses in that study. Its association p value was of nominal significance (p = 0.039, OR = 1.10 for the C risk allele reported here) and would not have been reported because it was considerably below the threshold for genome-wide significant association.

PSORS2 was originally identified by linkage, and our studies with a common missense SNP (rs11652075) also indicate that it can be associated with psoriasis. Linkage and association have also been seen in some genes for other common diseases, e.g., NOD2 (nucleotide-binding oligomerization domain containing 2, [MIM 605956]) in inflammatory bowel disease (IBD1 [MIM 266600]) and CFH (complement factor H, [MIM 134370]) in inflammatory eye disease.
Supplemental Data

Supplemental Data include two figures and three tables and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

This work was supported by grants from the National Institutes of Health (NIH): AR050266 and 5RC1AR058681 (A.M.B.), T32AR007279 (E.D.O.R.), T32HL083822 and T32 GM07200 (C.E.J.), AR060222 (M.A.L.), T32HG000045 (C.E.J.), and K08AR057763 (W.L.). J.T.E., R.N., and P.S. were supported by NIH grants R01 AR042742, R01 AR050511, and AR054966 and by the Babcock Memorial Trust. J.T.E. is supported by the Ann Arbor Veterans Affairs Hospital. R.G.M. is supported by the Intramural Research Programs of the National Institute of Arthritis and Musculoskeletal and Skin Diseases. Additional funding came from the Arthritis Society of Canada, the Atlantic Innovation Fund, the Canadian Institute of Health Research, the Arthritis Society, the Dana Foundation, and the Academia Sinica and National Science Council (National Clinical Core, National Genotyping Core) of Taiwan. The authors thank the many individuals with psoriasis and the controls who participated in this study. We thank Linus Schwantes-An, Weimin Duan, and Nancy Saccone for advice on statistical analyses. We also thank Mayte Suarez-Farinaz for biostatistical assistance, Haoyan Chen, Olivia Lai, Fawnda Pellet, and M.J. Malloy for assembling the control population, Pui-Yan Kwok for sample procurement, Trilokraj Tejasvi for clinical evaluation of study subjects, Nikki Plass for patient scheduling and study logistics, Damaris Garcia for the management of samples, and Debbie Stone and Dawn Chapelle for patient care. Mike Lovett provided helpful comments on the manuscript. The authors are indebted to the National Psoriasis Foundation for continuing support during the course of this study.

Received: December 27, 2011
Revised: February 29, 2012
Accepted: March 13, 2012
Published online: April 19, 2012

Web Resources

The URLs for data presented herein are as follows:

1,000 Genomes Project, http://www.1000genomes.org
gProfiler, http://biit.cs.ut.ee/gprofiler/index.cgi
Ingenuity Pathway Analysis (IPA), http://www.ingenuity.com/
Microarray Gene Expression Data Society (MIAME), http://www.mged.org/Workgroups/MIAME/miame_checklist.html
Online Mendelian Inheritance in Man (OMIM), http://omim.org
PLINK, http://pngu.mgh.harvard.edu/purcell/plink/
PLINK/SEQ version 0.05, http://atgu.mgh.harvard.edu/plinkseq/
PolyPhen-2.0, http://genetics.bwh.harvard.edu/pph2/
rmeta, http://cran.r-project.org/web/packages/rmeta/index.html
SIFT, http://sift.jcvi.org/

References


In two cohorts of European ancestry, evidence of association between psoriasis and rs11652075 increased when the PSORS1 variant HLA-Cw*0602 was included as a covariate, suggesting a genetic connection between PSORS1 and PSORS2. This variant resides in the CARD14fl C-terminal GUK domain, which is predicted to relay external signals to the cellular milieu.16 This could explain the possible genetic connection because antigen stimulation via PSORS1 might increase the risk of psoriasis by upregulating signaling through the CARD14 pathway. This is consistent with the fact that CARD14 affects signaling downstream of antigen stimulation. However, it is important to note that the PSORS1 risk variant, HLA-Cw*060240, was not present in the affected members of the 17q-linked multiplex families, the pustular-psoriasis case described previously, or the cases with the p.Glu142Lys or p.Glu142Gly variants, indicating that rare CARD14 variants can be sufficient to lead to disease.

Our study illustrates some of the difficulties in searching for rare variants associated with common disease, even with an established gene. For example, it is not always easy to differentiate pathogenic rare variants from others when their numbers are very small. This also has an impact on the detection of gene–gene or gene–environment interactions. Moreover, it is sometimes necessary to recreate the cellular milieu (e.g., an inflammatory stimulus in the case of CARD14) when one attempts to differentiate disease-causing variants from neutral variants through functional studies. Nevertheless, our findings provide evidence that some rare CARD14 variants predispose to psoriasis and, possibly, to psoriatic arthritis, and they suggest that common variants in this region might also predispose to disease. They illustrate the challenges faced in identifying truly pathogenic rare variants in common disease and contribute to our understanding of the genetic basis of psoriasis.


