HHS Public Access

Author manuscript

J Invest Dermatol. Author manuscript; available in PMC 2018 February 13.

Published in final edited form as:

J Invest Dermatol. 2017 May ; 137(5): e113–e118. doi:10.1016/j.jid.2016.03.047.

The Molecular Revolution in Cutaneous Biology: The Era of Genome-Wide Association Studies and Statistical, Big Data, and Computational Topics

Hima Anbunathan[1] and **Anne M. Bowcock**[1]

[1]National Heart and Lung Institute, Imperial College, London, UK

Abstract

The investigation of biological systems involving all organs of the body including the skin is in an era of big data. This requires heavy-duty computational tools, and novel statistical methods. Microarrays have allowed the interrogation of thousands of common genetic markers in thousands of individuals from the same population (termed genome wide association studies or GWAS) to reveal common variation associated with disease or phenotype. These markers are usually single nucleotide polymorphisms (SNPs) that are relatively common in the population. In the case of dermatological diseases such as alopecia areata, vitiligo, psoriasis and atopic dermatitis, common variants have been identified that are associated with disease, and these provide insights into biological pathways and reveal possible novel drug targets. Other skin phenotypes such as acne, color and skin cancers are also being investigated with GWAS. Analyses of such large GWAS datasets require a consideration of a number of statistical issues including the testing of multiple markers, population substructure, and ultimately a requirement for replication. There are also issues regarding the missing heritability of disease that cannot be entirely explained with current GWAS approaches. Next generation sequencing technologies such as exome and genome sequencing of similar patient cohorts will reveal additional variants contributing to disease susceptibility. However, the data generated with these approaches will be orders of magnitude greater than that those generated with arrays, with concomitant challenges in the identification of disease causing variants.

THE GENOME-WIDE ASSOCIATION STUDY ERA

Studies on Mendelian diseases in the 1980s and 1990s were performed with traditional family and linkage studies and showed genetic regions segregating with disease. Positional cloning then led to the identification of the mutated genes (Collins, 1995). However, this approach offered low power and poor resolution when dealing with polygenic disorders, in which it is less likely for all the genetic factors to aggregate in every family under study and the penetrance of risk alleles can be low. In the case of common diseases, there has been a shift to genetic association studies, where parameters such as mode of inheritance and

Correspondence: Anne M. Bowcock, National Heart and Lung Institute, Guy Scadding Building, Dovehouse Street, Imperial College, London, SW3 6LY, UK. a.bowcock@imperial.ac.uk.

CONFLICT OF INTEREST

The authors state no conflict of interest.

penetrance do not need to be known. Although association studies between common traits and human phenotypes had been used for many years, it was only with the advent of genomic technologies that a global approach was possible. This is usually performed with arrays through which single nucleotide polymorphisms (SNPs) are interrogated by hundreds of individual DNA samples. Termed *genome-wide association studies* (GWASs), this approach has provided important insights into many common skin diseases, phenotypes, and skin cancer by providing genetic risk factors for disease (Bowcock, 2007).

GWASs are justified on the basis of the "common disease, common variant" hypothesis, where allelic variants present in more than 1e5% of the population can contribute to susceptibility to common disease. Although SNPs have usually been the polymorphic marker of choice because they are frequent in the genome and alleles are common and easy to type, there have also been association tests for other types of variation in the genome such as copy number variation (Craddock et al., 2010). The most popular study design used in GWAS is the case-control study in which the frequencies of markers are tested for association between the genotyped groups with and without disease. One of the most critical aspects of a GWAS is to use a sample size with sufficient statistical power to detect significant association with markers of low effect. The size depends in part on the study design (case-control vs. case-parents), effect sizes of genetic variants, linkage disequilibrium between markers and number of markers used. Although early GWASs were performed on hundreds of patients versus control subjects, cohort sizes in later studies increased by over 10-fold. For most diseases and phenotypes, this has led to the identification of ±10–100 genetic risk factors of low effect.

GWAS HIGHLIGHTS IN DERMATOLOGY

Alopecia areata, vitiligo, psoriasis, and atopic dermatitis are common skin diseases with complex etiologies involving genetic and environmental risk factors. In many cases, GWASs have shown the involvement of unexpected pathways or cell types. Alopecia areata is an autoimmune disease resulting from damage of the hair follicle by T cells. Associations with the major histocompatibility complex (MHC), including HLA-DR, have been described. A region of strong association also resides within the *ULBP* gene cluster on chromosome band 6q25.1, which encodes activating ligands of the natural killer cell receptor NKG2D. This had not previously been implicated in an autoimmune disease. *ULBP3* is up-regulated in lesional scalp from patients with alopecia areata in the hair follicle dermal sheath during active disease (Petukhova et al., 2010). The latter observation is consistent with the observation that cytotoxic $CD8^+NKG2D^+$ T cells are both necessary and sufficient for the induction of alopecia areata in mouse models of disease. Other pathways implicated from GWAS include autophagy/apoptosis, transforming growth factor-β/regulatory T cells, Jak signaling, and activation and proliferation of regulatory T cells (Betz et al., 2015; Petukhova et al., 2010).

Vitiligo GWASs have shown associations with genes encoding immunoregulatory components such as *TYR, CD80, CLNK, BACH2, SLA, CASP7, CD44, IKZF4*, and *SH2B3*, as well as genes encoding melanocyte components (*OCA2-HERC2* and *TYR*) (Jin et al., 2012).

GWASs for psoriasis have shown over 50 associations to date. This includes an early association with the *IL12B* 3′-untranslated-region SNP (Cargill et al., 2007), confirming the results of a small Japanese study (Tsunemi et al., 2002). Subsequently, many GWASs for psoriasis have been performed with overwhelming evidence for association with the class I region of the MHC being obtained in all studies. Other implicated pathways include T-cell development and T-cell polarization, innate immunity, and negative regulators of immune responses (Jordan et al., 2012; Nair et al., 2009; Strange et al., 2010; Stuart et al., 2010; Sun et al., 2010; Tsoi et al., 2012; Yin et al., 2014). There are also associations with a select set of genes involved in barrier formation (e.g. deletion of a region of the late cornified envelope gene cluster) (de Cid et al., 2009).

Heterozygotes with loss-of-function mutations in the *FLG* gene are at increased risk of AD (Palmer et al., 2006). This important finding implicated altered epidermal differentiation in the genesis of this disease. Subsequent GWASs of large European and Asian cohorts have shown over 20 associated loci that also implicate the innate-adaptive immune response, IL-1 family signaling, regulatory T cells, the vitamin D pathway, and the nerve growth factor pathway. Some loci are shared by both populations (e.g., that harboring the gene for *OVOL1*), and others are unique (e.g., the MHC and *CARD11* in Asia) (Ellinghaus et al., 2013; Esparza-Gordillo et al., 2009; Hirota et al., 2012; Sun et al., 2011; Tamari and Hirota, 2014; Weidinger et al., 2013).

Systemic lupus erythematosus is a heterogeneous disease with a diverse spectrum of clinical symptoms, ranging from skin rash to end-organ damage. GWASs for systemic lupus erythematosus have confirmed earlier known HLA associations and have shown additional associations at *IRF5* and genes encoding immune system components (Gateva et al., 2009; Graham et al., 2008; Han et al., 2009; International Consortium for Systemic Lupus Erythematosus et al., 2008).

Individuals with Behçet's disease suffer from recurrent inflammatory attacks leading to recurrent ocular symptoms, oral and genital ulcers, and skin lesions. Behçet's disease is associated with HLA-B*51, and GWAS of a Turkish cohort has also shown association of a variant within *IL*-10 conferring decreased mRNA expression (Remmers et al., 2010) and *IL23R-IL12RB2*. There are also associations with genes encoding other immune system components. Three risk loci are shared with ankylosing spondylitis and psoriasis and include the MHC class I region, *ERAP1* and *IL-23R*, and the MHC class I-*ERAP1* interaction), as well as two loci encoding genes shared with inflammatory bowel disease (*IL-23R* and *IL-10*). This implicates shared pathogenic pathways in the spondyloarthritides and Behçet's disease (Kirino et al., 2013).

Systemic sclerosis is an autoimmune disease characterized by fibrosis of the skin and internal organs that leads to profound disability and premature death. Associated loci include *CD247*, the MHC, *IRF5*, and the gene encoding *STAT4* (Radstake et al., 2010). Acute urticaria and angioedema (nonsteroidal anti-inflammatory drug-induced) have also shown associations through a small GWASs (Cornejo-Garcia et al., 2013).

Risk of acne vulgaris, a common inflammatory disorder of the cutaneous pilosebaceous unit is associated with a locus at chromosome band 11q13.1 in Asians (candidates are *DDB2* and *SELL*) (He et al., 2014). These are both genes encoding proteins that are involved in androgen metabolism, inflammation processes, and scar formation. In Europeans, three associated loci contain genes linked to the TGF-βcell signaling pathway (*OVOL1*, *FST*, and *TGFB2*) (Navarini et al., 2014). Although there is little overlap between the studies, these studies illustrate the power of GWAS in helping understand the causes of some common, poorly understood skin diseases.

Other skin phenotypes that have also been examined through GWAS include skin color. Nine SNPs can predict skin color in Europeans and neighboring populations and could be relevant in future forensic and anthropological investigations (Han et al., 2008; Liu et al., 2015). Several SNPs located in or adjacent to pigmentation genes play a role in tanning response after exposure to sunlight (Nan et al., 2009). There is also a locus on chromosome 3, harboring skin-expressed genes *STXBP5L* and *FBXO40*, that is associated with facial photoaging (Le Clerc et al., 2013).

GWASs have also shown associations with basal cell and melanoma skin cancer risk (Barrett et al., 2011; Bishop et al., 2009; Brown et al., 2008; Falchi et al., 2009; Gudbjartsson et al., 2008; Law et al., 2015; Macgregor et al., 2011; Stacey et al., 2008, 2009, 2011, 2015). In the case of melanoma these include genes involved in melanocyte development and telomere maintenance. In the case of basal cell carcinoma, these include *MYCN*, *CASP8-ALS2CR12*, *ZFHX4*, and *GATA3* in European populations.

STATISTICAL CONSIDERATIONS IN GWAS

Analysis methods

For binary traits such as presence or absence of disease or a particular phenotype, contingency tables are used to test deviation from independence expected under the null hypothesis of no association. Commonly used statistical tests include Fisher exact test, genotypic tests, and the Cochran-Armitage trend test, which tests if a given genetic marker is associated with the disease, in which case the allele frequencies or genotype frequencies will be different between the two groups. For complex traits for which the models need to include additional covariates such as epidemiological or clinical risk factors, logistic regression models are used. These predict the probability of having the disease for a given genotype. Analysis of variance is used for quantitative traits that are similar to linear regression with a categorical outcome variable. This tests the null hypothesis of no difference between the phenotypic means of any genotype group and can incorporate covariates with main effects and interactions.

For family-based association studies, the transmission disequilibrium test (TDT) (Spielman et al., 1994) can be used to evaluate transmission of one allele more often by chance among families with at least one affected child. PLINK is a popular tool that is used to perform these standard genome-wide association analyses and to provide a framework for modeling quantitative data, allowing incorporation of covariates and family-based genome-wide association analyses like the transmission disequilibrium, DFAM (family-based association

for disease traits), and QFAM (family-based association tests for quantitative traits) tests (Purcell et al., 2007).

Multiple testing

The enormous number of statistical tests of association performed for each SNP against the phenotype is an essential part of the GWAS. However, because many tests are done (e.g., by using thousands of SNPs), *P*-values must be adjusted for multiple testing. In general, this has been somewhat arbitrarily set at 5×10^{-8} (on the basis of the number of markers tested and a *P*-value of association of .025 per marker). However, this restrictive value means that many bona fide associations cannot be identified because they lie within a region harboring multiple associations that are due to statistical noise. There are also issues of the population being screened. Although 500,000 SNPs is sufficient for detecting a reasonable number of associated loci in European populations, older populations such as those from Africa, where regions of linkage disequilibrium are smaller (for example ±5–30 kilobase pairs for African genomes vs. 30–100 kilo base pairs for European genomes, on average), a GWAS in African populations could require a larger number of markers (probably > 1.5 million) and a concomitantly more significant *P*-value for detecting and reporting associations.

Population substructure

One of the important considerations in the GWAS design is to control for population structures between the groups. SNP allele frequencies often vary in different populations and can cause spurious associations. Population stratification can be detected with algorithms such as EIGENSTRAT, which uses a principal components analysis approach (Li and Yu, 2008). Data with subpopulations can then be used in genome-wide association analysis by excluding samples with allele frequencies dissimilar to the target population or by adding this as a covariate in the statistical model. Both principal component analysis and multidimensional scaling have been applied to correct population stratifications to reduce spurious associations (Price et al., 2010).

Replication

After detection of association, it is recommended that the association be replicated in another cohort of samples to distinguish true signals from spurious ones and to make a reliable estimate of the effect sizes. It has been a challenge to replicate many GWAS findings, and some guidelines have been outlined by National Human Genome Research Institute working group (Chanock et al., 2007). Some of these include use of sufficient sample sizes, independent datasets from the population in which the initial GWAS finding was made, and identical phenotype criteria in both the GWAS and replication studies.

Meta-analyses

To increase statistical power, multiple studies can be combined to perform a meta-analysis. This helps in the discovery of new variants or in replication of known significant variant, but this approach requires many stages of organization (Evangelou and Ioannidis, 2013). Imputation is used to predict the genotypes of SNPs that have not been typed in the study samples or are missing. A number of algorithms have been developed for this and include

IMPUTE, version 1 (hidden Markov model) (Marchini et al., 2007), fastPHASE (Scheet and Stephens, 2006), BIMBAM (expectatione—maximization algorithm) (Servin and Stephens, 2007), MACH (imputes unobserved genotypes fusing reference panel of haplotypes) (Li et al., 2010), and BEAGLE (uses graphical model) (Browning and Browning, 2011). Other imputing methods include an SNP-tagging approach (Johnson et al., 2001), which is used by PLINK (Purcell et al., 2007), SNPMSTAT (Lin et al., 2008), and TUNA (Nicolae, 2006). Approaches for imputation are rapidly evolving, and the next-generation sequencing methods can complement the array-based methods for SNP detection, to help increase the phenotypic prediction accuracy.

Conditional analysis

There are a number of instances in which the top GWA SNP does not account for all the variation at that locus. To detect this additional variation, conditional analyses are performed where secondary association signals are detected after conditioning on the primary associated SNP. Examples of this in skin diseases are the MHC in psoriasis, where there are at least two other secondary signals and secondary hits at IFIH1 and IL12B (Cargill et al., 2007; Yin et al., 2015), and Behçet's disease, where there is an MHC class II association and a second, independent association within the MHC class I region (Remmers et al., 2010).

Missing heritability

One major issue with GWASs is the missing heritability. Even with increased sample sizes it is not possible to account for a considerable fraction of the genetic contribution of common disease. Detection of missing heritability can include meta-analyses to increase the power of GWASs, resequencing to search for rare variants that cannot be detected with GWAS, and a search for epistasis (gene-gene interactions) and heritable epigenetic effects. An analysis of rare variants identified by resequencing over 1,000 genomes (Abecasis et al., 2012) and typed on an "exome chip" has shown additional variants associated with psoriasis that account for an additional 1.9% of psoriasis heritability (Zuo et al., 2015).

Modeling gene-gene interactions presents a challenge because of the small sample size and large number of SNPs interrogated. Although most computers can handle statistical analysis of few hundred SNPs, higher-order interactions, for which the search space grows with the number of SNPs, require much greater computational resources, along with robust statistical methods. A number of methods have been used for modeling epistatic interactions, including PLINK using standard regression, Random forest (trademark of Leo Breiman and Adele Cutler), and bayesian epistasis association mapping using Markov chain Monte Carlo sampling (Zhang and Liu 2007). In the case of psoriasis, interaction between *HLA-C* and *ERAP1* (Strange et al., 2010) and between *HLA-C* and a locus within the epidermal differentiation complex (the *LCE* deletion described earlier) (Bergboer et al., 2012) have been described. Interaction between HLA-B*51 and ERAP1 is described in Behçet's disease (Remmers et al., 2010). However, there are few examples of gene-gene interactions in any common human disease. This could be due to lack of power to detect such epistatic interactions, even with large cohorts. The interactions with the MHC that have been described might have achieved statistical significance because of the strong initial associations of this locus with the diseases in question.

Pathway analysis

Once a GWAS has shown its top hits and provided a preliminary framework for identifying the genetic component of a disease or phenotype of interest, additional analyses are required to accumulate evidence for association (particularly in the case of less significant findings). Taking a pathway-based approach can help prioritize GWAS results to identify disease-specific pathways, classify clinical outcomes, and find drug targets. These analyses can be loosely classified into three categories (Kraft and Raychaudhuri, 2009): user-defined gene set analysis (over-representational analysis and gene set enrichment analysis), data mining methods (Chen et al., 2010; Wu et al., 2010), and network-based approaches. The last include GRAIL (Raychaudhuri et al., 2009), DAPPLE (Rossin et al., 2011), and GeneMANIA (Warde-Farley et al., 2010). Network-based approaches interrogate a variety of sources including the literature, gene expression, known protein-protein interactions, protein domain similarity, and protein co-localization. Besides pinpointing plausible genes in regions of association, they can help identify additional susceptibility that might be missed because of the genome-wide significance threshold. This was used for osteoporosis, and additional novel genes were identified using connectivity to known bone genes (Farber, 2013).

Transcriptome analyses of psoriatic versus normal skin have been performed with both array-based and next-generation sequencing methods (Li et al., 2014; Suarez-Farinas et al., 2010), and integration of the altered transcriptome with genes identified through GWAS can also be useful for creating altered disease networks. Although currently challenging because of limited knowledge, such approaches have shown transcription factor/DNA-binding proteins that also connect transcriptomics to drug development (Swindell et al., 2015b). Recently, proteomic studies have also been performed in psoriasis (Swindell et al., 2015a), reflecting a future direction of “-omic” studies.

Computational challenges with big data

With the technological advances in the capture of biological data (including next-generation sequencing), the volume of data for common diseases and phenotypes is rapidly growing. Studies involving subjects from large populations with phenotypic and health information require a pipeline that uses advanced informatics methods to convert raw data into useful information. However, these approaches impose a heavy computational burden, and tools used for these computations need to be developed and optimized for the large amounts of data and iterative processes.

PERSONALIZED MEDICINE

With the advances in next-generation sequencing it is possible to identify highly penetrant mutations that could have a direct impact on a patient. However, unlike Mendelian traits, for which genetic risk factors (mutations) are rare and of high penetrance, risk factors identified through GWAS are common and of low penetrance. Hence, although factors identified through GWAS might help assess risk susceptibility for a disease or drug response, the genetic prediction for a complex disease using information from SNPs with alleles with small effect sizes may add little value in clinical decision making until it is possible to

explain most of the genetic heritability of any common disease. In spite of this limitation, multiple risk factors from GWASs often cluster into specific biological pathways, and in this way provide novel drug targets (Price et al., 2010). One excellent example of this is the association of psoriasis with the IL-23 receptor gene (*IL23R*) (Cargill et al., 2007) and the efficacy of biologics targeting components of the IL-23 heterodimer (IL-12B/p40 by ustekinumab) and its downstream pathway (e.g., the anti-IL-17 receptor A by brodalumab) (Lebwohl et al., 2015). In the case of common diseases of the skin, the integration of many layers of biological data (clinical, genetics/genomics, proteomics, epigenomics, metabolomics etc.) will ultimately show the basis for their complex and heterogeneous natures.

Acknowledgments

This study was supported in part by NIH grant AR050266 to AMB.

Abbreviations

GWAS	genome-wide association study
MHC	major histocompatibility complex
SNP	single nucleotide polymorphism

References

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]

Barrett JH, Iles MM, Harland M, Taylor JC, Aitken JF, Andresen PA, et al. Genome-wide association study identifies three new melanoma susceptibility loci. Nat Genet. 2011; 43:1108–13. [PubMed: 21983787]

Bergboer JG, Zeeuwen PL, Schalkwijk J. Genetics of psoriasis: evidence for epistatic interaction between skin barrier abnormalities and immune deviation. J Invest Dermatol. 2012; 132:2320–31. [PubMed: 22622420]

Betz RC, Petukhova L, Ripke S, Huang H, Menelaou A, Redler S, et al. Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. Nat Commun. 2015; 6:5966. [PubMed: 25608926]

Bishop DT, Demenais F, Iles MM, Harland M, Taylor JC, Corda E, et al. Genome-wide association study identifies three loci associated with melanoma risk. Nat Genet. 2009; 41:920–5. [PubMed: 19578364]

Bowcock AM. Genomics: guilt by association. Nature. 2007; 447:645–6. [PubMed: 17554292]

Brown KM, Macgregor S, Montgomery GW, Craig DW, Zhao ZZ, Iyadurai K, et al. Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat Genet. 2008; 40:838–40. [PubMed: 18488026]

Browning BL, Browning SR. A fast, powerful method for detecting identity by descent. Am J Hum Genet. 2011; 88:173–82. [PubMed: 21310274]

Cargill M, Schrodi SJ, Chang M, Garcia VE, Brandon R, Callis KP, et al. A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. Am J Hum Genet. 2007; 80:273–90. [PubMed: 17236132]

Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al. Replicating genotype-phenotype associations. Nature. 2007; 447:655–60. [PubMed: 17554299]

Chen X, Wang L, Hu B, Guo M, Barnard J, Zhu X. Pathway-based analysis for genome-wide association studies using supervised principal components. Genet Epidemiol. 2010; 34:716–24. [PubMed: 20842628]

Collins FS. Positional cloning moves from perditional to traditional. Nat Genet. 1995; 9:347–50. [PubMed: 7795639]

Cornejo-Garcia JA, Liou LB, Blanca-Lopez N, Dona I, Chen CH, Chou YC, et al. Genome-wide association study in NSAID-induced acute urticaria/angioedema in Spanish and Han Chinese populations. Pharmacogenomics. 2013; 14:1857–69. [PubMed: 24236485]

Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature. 2010; 464:713–20. [PubMed: 20360734]

de Cid R, Riveira-Munoz E, Zeeuwen PL, Robarge J, Liao W, Dannhauser EN, et al. Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. Nat Genet. 2009; 41:211–5. [PubMed: 19169253]

Ellinghaus D, Baurecht H, Esparza-Gordillo J, Rodriguez E, Matanovic A, Marenholz I, et al. High-density genotyping study identifies four new susceptibility loci for atopic dermatitis. Nat Genet. 2013; 45:808–12. [PubMed: 23727859]

Esparza-Gordillo J, Weidinger S, Folster-Holst R, Bauerfeind A, Ruschendorf F, Patone G, et al. A common variant on chromosome 11q13 is associated with atopic dermatitis. Nat Genet. 2009; 41:596–601. [PubMed: 19349984]

Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet. 2013; 14:379–89. [PubMed: 23657481]

Falchi M, Bataille V, Hayward NK, Duffy DL, Bishop JA, Pastinen T, et al. Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. Nat Genet. 2009; 41:915–9. [PubMed: 19578365]

Farber CR. Systems-level analysis of genome-wide association data. G3 (Bethesda). 2013; 3:119–29. [PubMed: 23316444]

Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genet. 2009; 41:1228–33. [PubMed: 19838195]

Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, et al. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nat Genet. 2008; 40:1059–61. [PubMed: 19165918]

Gudbjartsson DF, Sulem P, Stacey SN, Goldstein AM, Rafnar T, Sigurgeirsson B, et al. ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. Nat Genet. 2008; 40:886–91. [PubMed: 18488027]

Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, et al. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet. 2008; 4:e1000074. [PubMed: 18483556]

Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, Hu Z, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet. 2009; 41:1234–7. [PubMed: 19838193]

He L, Wu WJ, Yang JK, Cheng H, Zuo XB, Lai W, et al. Two new susceptibility loci 1q24.2 and 11p11.2 confer risk to severe acne. Nat Commun. 2014; 5:2870. [PubMed: 24399259]

Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Sakashita M, et al. Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. Nat Genet. 2012; 44:1222–6. [PubMed: 23042114]

International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN). Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet. 2008; 40:204–10. [PubMed: 18204446]

Jin Y, Birlea SA, Fain PR, Ferrara TM, Ben S, Riccardi SL, et al. Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. Nat Genet. 2012; 44:676–80. [PubMed: 22561518]

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, et al. Haplotype tagging for the identification of common disease genes. Nat Genet. 2001; 29:233–7. [PubMed: 11586306]

Jordan CT, Cao L, Roberson ED, Duan S, Helms CA, Nair RP, et al. Rare and common variants in CARD14, encoding an epidermal regulator of NF-kappaB, in psoriasis. Am J Hum Genet. 2012; 90:796–808. [PubMed: 22521419]

Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, Seyahi E, et al. Genome-wide association analysis identifies new susceptibility loci for Behcet's disease and epistasis between HLA-B*51 and ERAP1. Nat Genet. 2013; 45:202–7. [PubMed: 23291587]

Kraft P, Raychaudhuri S. Complex diseases, complex genes: keeping pathways on the right track. Epidemiology. 2009; 20:508–11. [PubMed: 19525687]

Law MH, Bishop DT, Lee JE, Brossard M, Martin NG, Moses EK, et al. Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. Nat Genet. 2015; 47:987–95. [PubMed: 26237428]

Le Clerc S, Taing L, Ezzedine K, Latreille J, Delaneau O, Labib T, et al. A genome-wide association study in Caucasian women points out a putative role of the STXBP5L gene in facial photoaging. J Invest Dermatol. 2013; 133:929–35. [PubMed: 23223146]

Lebwohl M, Strober B, Menter A, Gordon K, Weglowska J, Puig L, et al. Phase 3 studies comparing brodalumab with ustekinumab in psoriasis. N Engl J Med. 2015; 373:1318–28. [PubMed: 26422722]

Li B, Tsoi LC, Swindell WR, Gudjonsson JE, Tejasvi T, Johnston A, et al. Transcriptome analysis of psoriasis in a large case-control sample: RNA-seq provides insights into disease mechanisms. J Invest Dermatol. 2014; 134:1828–38. [PubMed: 24441097]

Li Q, Yu K. Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. Genet Epidemiol. 2008; 32:215–26. [PubMed: 18161052]

Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010; 34:816–34. [PubMed: 21058334]

Lin DY, Hu Y, Huang BE. Simple and efficient analysis of disease association with missing genotype data. Am J Hum Genet. 2008; 82:444–52. [PubMed: 18252224]

Liu F, Visser M, Duffy DL, Hysi PG, Jacobs LC, Lao O, et al. Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. Hum Genet. 2015; 134:823–35. [PubMed: 25963972]

Macgregor S, Montgomery GW, Liu JZ, Zhao ZZ, Henders AK, Stark M, et al. Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. Nat Genet. 2011; 43:1114–8. [PubMed: 21983785]

Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007; 39:906–13. [PubMed: 17572673]

Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, et al. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. Nat Genet. 2009; 41:199–204. [PubMed: 19169254]

Nan H, Kraft P, Qureshi AA, Guo Q, Chen C, Hankinson SE, et al. Genome-wide association study of tanning phenotype in a population of European ancestry. J Invest Dermatol. 2009; 129:2250–7. [PubMed: 19340012]

Navarini AA, Simpson MA, Weale M, Knight J, Carlavan I, Reiniche P, et al. Genome-wide association study identifies three novel susceptibility loci for severe acne vulgaris. Nat Commun. 2014; 5:4020. [PubMed: 24927181]

Nicolae DL. Testing untyped alleles (TUNA)-applications to genome-wide association studies. Genet Epidemiol. 2006; 30:718–27. [PubMed: 16986160]

Palmer CN, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. Nat Genet. 2006; 38:441–6. [PubMed: 16550169]

Petukhova L, Duvic M, Hordinsky M, Norris D, Price V, Shimomura Y, et al. Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. Nature. 2010; 466:113–7. [PubMed: 20596022]

Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet. 2010; 11:459–63. [PubMed: 20548291]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. [PubMed: 17701901]

Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet. 2010; 42:426–9. [PubMed: 20383147]

Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, International Schizophrenia C, Purcell SM, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet. 2009; 5:e1000534. [PubMed: 19557189]

Remmers EF, Cosan F, Kirino Y, Ombrello MJ, Abaci N, Satorius C, et al. Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behcet's disease. Nat Genet. 2010; 42:698–702. [PubMed: 20622878]

Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 2011; 7:e1001273. [PubMed: 21249183]

Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet. 2006; 78:629–44. [PubMed: 16532393]

Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. PLoS Genet. 2007; 3:e114. [PubMed: 17676998]

Spielman RS, McGinnis RE, Ewens WJ. The transmission/disequilibrium test detects cosegregation and linkage. Am J Hum Genet. 1994; 54:559–60. [PubMed: 8116627]

Stacey SN, Gudbjartsson DF, Sulem P, Bergthorsson JT, Kumar R, Thorleifsson G, et al. Common variants on 1p36 and 1q42 are associated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. Nat Genet. 2008; 40:1313–8. [PubMed: 18849993]

Stacey SN, Helgason H, Gudjonsson SA, Thorleifsson G, Zink F, Sigurdsson A, et al. New basal cell carcinoma susceptibility loci. Nat Commun. 2015; 6:6825. [PubMed: 25855136]

Stacey SN, Sulem P, Jonasdottir A, Masson G, Gudmundsson J, Gudbjartsson DF, et al. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. Nat Genet. 2011; 43:1098–103. [PubMed: 21946351]

Stacey SN, Sulem P, Masson G, Gudjonsson SA, Thorleifsson G, Jakobsdottir M, et al. New common variants affecting susceptibility to basal cell carcinoma. Nat Genet. 2009; 41:909–14. [PubMed: 19578363]

Strange A, Capon F, Spencer CC, Knight J, Weale ME, Allen MH, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat Genet. 2010; 42:985–90. [PubMed: 20953190]

Stuart PE, Nair RP, Ellinghaus E, Ding J, Tejasvi T, Gudjonsson JE, et al. Genome-wide association analysis identifies three psoriasis susceptibility loci. Nat Genet. 2010; 42:1000–4. [PubMed: 20953189]

Suarez-Farinas M, Lowes MA, Zaba LC, Krueger JG. Evaluation of the psoriasis transcriptome across different studies by gene set enrichment analysis (GSEA). PLoS One. 2010; 5:e10247. [PubMed: 20422035]

Sun LD, Cheng H, Wang ZX, Zhang AP, Wang PG, Xu JH, et al. Association analyses identify six new psoriasis susceptibility loci in the Chinese population. Nat Genet. 2010; 42:1005–9. [PubMed: 20953187]

Sun LD, Xiao FL, Li Y, Zhou WM, Tang HY, Tang XF, et al. Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. Nat Genet. 2011; 43:690–4. [PubMed: 21666691]

Swindell WR, Remmer HA, Sarkar MK, Xing X, Barnes DH, Wolterink L, et al. Proteogenomic analysis of psoriasis reveals discordant and concordant changes in mRNA and protein abundance. Genome Med. 2015a; 7:86. [PubMed: 26251673]

Swindell WR, Sarkar MK, Stuart PE, Voorhees JJ, Elder JT, Johnston A, et al. Psoriasis drug development and GWAS interpretation through in silico analysis of transcription factor binding sites. Clin Transl Med. 2015b; 4:13. [PubMed: 25883770]

Tamari M, Hirota T. Genome-wide association studies of atopic dermatitis. J Dermatol. 2014; 41:213–20. [PubMed: 24628071]

Tsoi LC, Spain SL, Knight J, Ellinghaus E, Stuart PE, Capon F, et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat Genet. 2012; 44:1341–8. [PubMed: 23143594]

Tsunemi Y, Saeki H, Nakamura K, Sekiya T, Hirai K, Fujita H, et al. Interleukin-12 p40 gene (IL12B) 3′-untranslated region polymorphism is associated with susceptibility to atopic dermatitis and psoriasis vulgaris. J Dermatol Sci. 2002; 30:161–6. [PubMed: 12413772]

Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010; 38:W214–20. [PubMed: 20576703]

Weidinger S, Willis-Owen SA, Kamatani Y, Baurecht H, Morar N, Liang L, et al. A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis. Hum Mol Genet. 2013; 22:4841–56. [PubMed: 23886662]

Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, et al. Powerful SNP-set analysis for case-control genome-wide association studies. Am J Hum Genet. 2010; 86:929–42. [PubMed: 20560208]

Yin X, Cheng H, Lin Y, Fan X, Cui Y, Zhou F, et al. Five regulatory genes detected by matching signatures of eQTL and GWAS in psoriasis. J Dermatol Sci. 2014; 76:139–42. [PubMed: 25205356]

Yin X, Low HQ, Wang L, Li Y, Ellinghaus E, Han J, et al. Genome-wide meta-analysis identifies multiple novel associations and ethnic heterogeneity of psoriasis susceptibility. Nat Commun. 2015; 6:6916. [PubMed: 25903422]

Zhang Y, Liu JS. Bayesian inference of epistatic interactions in case-control studies. Nat Genet. 2007; 39:1167–73. [PubMed: 17721534]

Zuo X, Sun L, Yin X, Gao J, Sheng Y, Xu J, et al. Whole-exome SNP array identifies 15 new susceptibility loci for psoriasis. Nat Commun. 2015; 6:6793. [PubMed: 25854761]