Research Techniques Made Simple: Using Genome-Wide Association Studies to Understand Complex Cutaneous Disorders

Lam C. Tsoi[1,2,3], Matthew T. Patrick[1] and James T. Elder[1,4]

Complex cutaneous disorders result from the combined effect of many different genes and environmental factors, with individual genetic variants often having only a modest effect on disease risk. The ability to examine large numbers of samples is required for correlating genetic variants with diseases/traits. Technological advances in high-throughput genotyping, along with mapping of the human genome and its associated inter-individual variation, have allowed genetic variants to be analyzed at high density in large case-control cohorts for many diseases, including several major skin diseases. These genome-wide association studies focus on showing differences in the frequencies of variants between case and control groups, rather than co-transmission of a variant and disease through a family, as is done in linkage studies. In this review, we provide overall guidance for genome-wide association study analysis and interpreting the results. Additionally, we discuss challenges and future directions for genome-wide association studies, focusing on translation of findings to provide biological and clinical implications for dermatology.

Journal of Investigative Dermatology (2018) **138,** e23–e29; doi:10.1016/j.jid.2018.01.004

CME Activity Dates: 21 February 2018
Expiration Date: 20 February 2019
Estimated Time to Complete: 1 hour

Planning Committee/Speaker Disclosure: James T. Elder, MD is a consultant/advisor for Johnson & Johnson and Novartis AG. All other authors, planning committee members, CME committee members and staff involved with this activity as content validation reviewers have no financial relationships with commercial interests to disclose relative to the content of this CME activity.

Commercial Support Acknowledgment: This CME activity is supported by an educational grant from Lilly USA, LLC.

Description: This article, designed for dermatologists, residents, fellows, and related healthcare providers, seeks to reduce the growing divide between dermatology clinical practice and the basic science/current research methodologies on which many diagnostic and therapeutic advances are built.

Objectives: At the conclusion of this activity, learners should be better able to:

- Recognize the newest techniques in biomedical research.
- Describe how these techniques can be utilized and their limitations.
- Describe the potential impact of these techniques.

CME Accreditation and Credit Designation: This activity has been planned and implemented in accordance with the accreditation requirements and policies of the Accreditation Council for Continuing Medical Education through the joint providership of Beaumont Health and the Society for Investigative Dermatology. Beaumont Health is accredited by the ACCME to provide continuing medical education for physicians. Beaumont Health designates this enduring material for a maximum of 1.0 *AMA PRA Category 1 Credit(s)*™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Method of Physician Participation in Learning Process: The content can be read from the Journal of Investigative Dermatology website: http://www.jidonline.org/current. Tests for CME credits may only be submitted online at https://beaumont.cloud-cme.com/RTMS-Mar18 – click 'CME on Demand' and locate the article to complete the test. Fax or other copies will not be accepted. To receive credits, learners must review the CME accreditation information; view the entire article, complete the post-test with a minimum performance level of 60%; and complete the online evaluation form in order to claim CME credit. The CME credit code for this activity is: 21310. For questions about CME credit email cme@beaumont.edu.

[1]*Department of Dermatology, University of Michigan Medical School, Ann Arbor, Michigan, USA;* [2]*Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan, USA;* [3]*Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA; and* [4]*Ann Arbor Veterans Affairs Hospital, Ann Arbor, Michigan, USA*

Correspondence: Lam C. Tsoi or James T. Elder, Department of Dermatology, University of Michigan, 7421 Medical Sciences I, 1301 East Catherine, Ann Arbor, MI 48109-5609, USA. E-mail: alextsoi@umich.edu or jelder@med.umich.edu

Abbreviation: GWAS, genome-wide association study

ADVANTAGES AND LIMITATIONS OF GWAS

Advantages

- GWASs can identify new susceptibility regions without the need to know which variants may be relevant in advance ("hypothesis-free" approach).
- Knowledge obtained from GWASs can be used to guide other types of experiments.
- GWAS is a well-developed approach with many tools available for data analysis and interpretation of results.
- GWAS is suitable for complex polygenic diseases, with many genes contributing only modestly to disease risk.
- GWAS has the potential to guide development of precision (personalized) medicine and health care, especially when combined with other biomarkers.

Limitations

- GWAS needs a large sample size to achieve sufficient power (i.e., the multiple testing problem).
- It is often not trivial to identify how variants affect biology.

INTRODUCTION

Large-scale efforts, such as the Human Genome Project and the 1000 Genomes Project (1000 Genomes Project et al., 2015), have allowed common genetic variations (e.g., genetic differences between individuals in which the rare variant is present in >5% of individuals) to be mapped across multiple populations. This has facilitated development of new techniques to study the genetics and genomics of human diseases, including statistical tools for correlating genetic variants with diseases/traits of interest in genome-wide association studies (GWASs). GWASs have significantly advanced the identification of susceptibility regions (i.e., disease-associated regions in the human genome) for cutaneous disorders in different populations, including psoriasis (Tsoi et al., 2017; Yin et al., 2015), atopic dermatitis (Hirota et al., 2012; Paternoster et al., 2015), alopecia areata (Betz et al., 2015), acne vulgaris (He et al., 2014; Navarini et al., 2014), vitiligo (Jin et al., 2016), and lupus (Morris et al., 2016) (see Supplementary Table S1 online). These discoveries have led to the uncovering of disease pathways and thus have potential to facilitate novel drug development, including the notable example of PCSK9 as a therapeutic target to reduce low-density lipoprotein (LDL) cholesterol levels in hypercholesterolemia (Price et al., 2015).

Results from GWASs are also shaping our understanding of biological effects. Far from the early expectations that GWAS would uncover "nonsynonymous" disease-associated mutations (i.e., genetic changes that alter protein structure), interpretation of recent GWAS results has led to an appreciation that disease-associated genetic differences commonly affect the efficiency of regulatory elements in a cell type-specific manner (Farh et al., 2015), rather than altering proteins. Coupled with the sheer numbers of variants correlating with disease (for instance, more than 60 distinct loci in psoriasis alone [Tsoi et al., 2017]), it becomes apparent why most variants, when considered individually, have only modest effect on disease risk. It is important to understand that this modest risk does not mean that these variants are unimportant, only that further experiments are needed to (i) identify which genes are actually affected by these variations and (ii) understand how the affected genes participate in the disease process.

This review aims to provide an overview of GWAS and its associated techniques. Specifically, we illustrate how GWAS data, methods, and results can be interpreted, and we discuss the benefits and limitations of GWAS. Although we focus on genotyping arrays, some topics discussed can also be applied to genetic data generated from DNA sequencing experiments.

STRATEGIES FOR GWAS

Genotyping

To understand the GWAS strategy, it is important to understand the concept of *linkage disequilibrium*. Figure 1a shows that by crossover during meiosis recombination over many generations, our ancestors' chromosomes formed small "chunks" of genetic materials (i.e., haplotypes) in which their underlying variations have been preserved (Ott, 1999). GWAS takes advantage of linkage disequilibrium structure to genotype only one or a few of the correlated variants in the haplotypes and offers clues about causal disease-associated variants.

Genotyping is the most commonly used approach to profile genetic data for GWAS (Bush and Moore, 2012). Genotyping arrays exploit DNA hybridization and fluorescence technologies (Figure 1b). To detect a single-nucleotide polymorphism, several probes are placed on the array in such a way that for any given probe, the hybridization efficiency of one single-nucleotide polymorphism allele is substantially different from the other allele(s).

Various genotyping arrays have been developed for association studies (Table 1). *Traditional* GWAS arrays cover the entire genome and focus on genotyping common variants. *Custom* arrays, such as Metabochip or Immunochip (Illumina, San Diego, CA), provide high density genotyping in specific regions of interest identified by earlier GWAS studies (Cortes and Brown, 2011). For example, the exome array (Exomechip; Illumina) focuses on the approximately 2% of the genome transcribed and translated into proteins. There is general agreement that even if most disease-associated variation relates to gene regulation, finding associations that influence protein structure is of high importance, even if this is uncommon (Rivas et al., 2011; Tang et al., 2014). Large-scale genotyping and sequencing projects (1000 Genomes Project et al., 2015) have advanced the development of genotyping platforms and efficient strategies in tagging common variants in GWAS arrays. These arrays can be used to study small insertions/deletions in addition to single-nucleotide polymorphisms. Genetic data from genotyping arrays can have many different formats, but the file format used by the PLINK software (a publicly available whole-

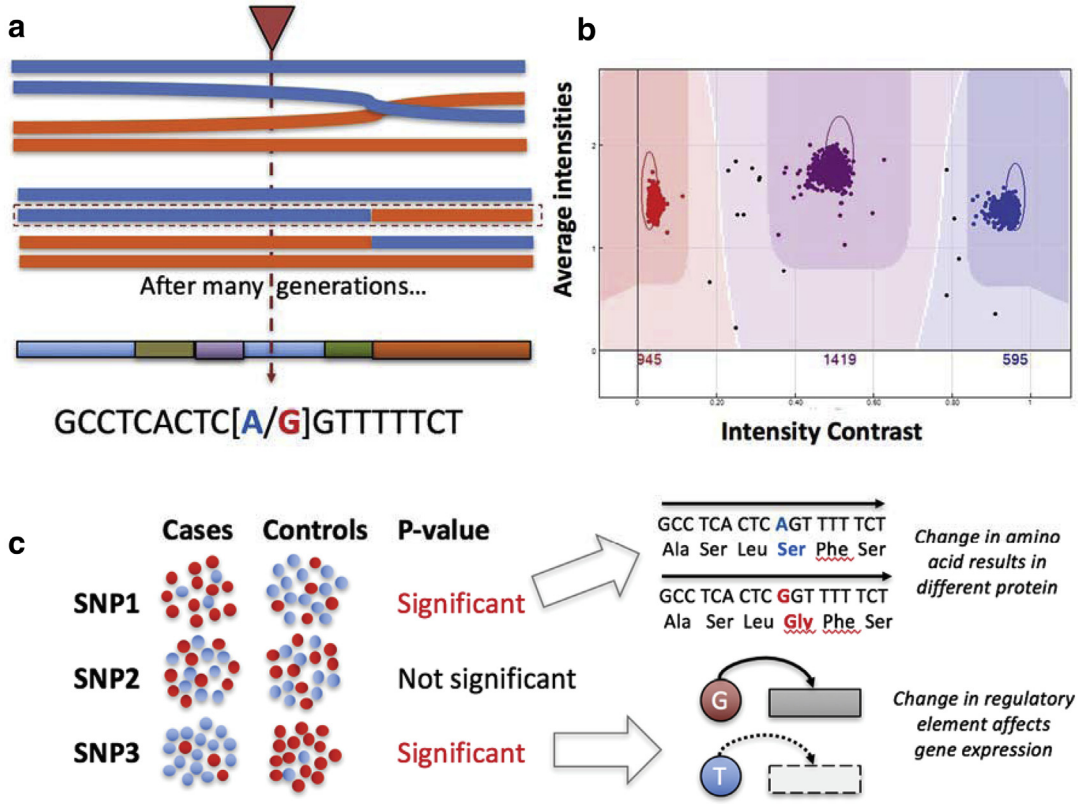

Figure 1. Basic illustrations for GWAS. (**a**) Chromosomes are "sliced and diced" by meiosis over thousands of generations, such that only small chunks of the ancestral chromosomes persist intact in the present-day chromosome (linkage disequilibrium). Each haplotype is represented by a different color, with the crossing point of the blue and orange haplotypes indicating crossover (i.e., the exchange of haplotypes during meiosis). The square brackets indicate a mutation from A to G, which occurs within a small chunk of the blue haplotype. (**b**) Hybridization and fluorescence technologies define genotypes for each marker across different samples (here represented as dots). The x-axis corresponds to the contrast between the two fluorescent intensities for the two alleles; the y-axis is the average intensity. Samples with homozygous genotypes are colored in red or blue, and the heterozygous genotype is colored purple. Classification of genotypes does not work well if the intensities of the samples do not fall in any one of the three clusters (black). (**c**) GWASs identify genetic signals (i.e., where there is a statistically-significant difference in allele frequencies) for a particular trait using a statistical model. These markers may lead to a specific phenotype through changes to proteins or regulatory mechanisms. GWAS, genome-wide association study; SNP, single-nucleotide polymorphism.

genome data analysis toolset [Purcell et al., 2007], with PLINK 2.0 being the latest version) is commonly supported.

Association

Single-variant association is performed to associate the alleles/genotypes of each variant with the trait of interest, typically through a generalized linear model (Bush and Moore, 2012). To determine which variants are associated with the trait, a genome-wide significance threshold ($P < 5 \times 10^{-8}$) is normally used (Fadista et al., 2016). This value was chosen to account for *P*-values being significant by random chance (a frequent problem in multiple testing) by controlling the family-wise error rate, under the assumption of one million independent haplotypes ($0.05/10^6 = 5 \times 10^{-8}$). The criterion is sufficiently robust for common variants in European populations; however, more stringent criteria might be needed for less common variants or association studies in other populations (Fadista et al., 2016). In case/control studies, odds ratios are often reported as effect sizes for the associated variants identified. Findings from earlier GWAS studies tend to have higher odds ratios (Tsoi et al., 2012) because of the limitations in showing modest signals with a small sample size, whereas the newer loci revealed by later association studies (revealed by larger sample size) tend to have smaller odds ratios.

Table 1. Commonly used array platforms for GWAS

Properties	GWAS Array	Exome Array	Targeted Array
Number of markers	500,000 to 5,000,000	~200,000 (can add GWAS content)	Vary (e.g., Immunochip and Metabochip, ~200,000)
Regions of interest	Whole genome	Exome	Targeted
Allow imputation	Yes	No, if without GWAS content	Yes, but well-imputed markers are limited
Requires prior knowledge	For tagging	Exonic regions	Candidate regions
Variants to study	Common	Rare	Common/low allele frequency variants

Abbreviations: GWAS, genome-wide association study.

Quality control

Quality control is a critical data processing procedure for ensuring robustness of any downstream analysis for genetic data (Winkler et al., 2014), as with other omics data. In

Table 2. Common procedures in performing genome-wide association analysis and result interpretation

Step	Description	Example Software Programs
Quality control	Array	PLINK GTOOL
	Check genomic build	
	Sample genotyping rate	
	Check sex inconsistencies	
	Marker	
	Marker genotyping rate	
	Mapping probe to genome (to ensure unique mapping)	
	Remove monomorphic markers	
	Hardy-Weinberg equilibrium	
	Genotype clustering	Z-call optiCall
	Principal component analysis	EIGENSTRAT LASER
	Relationship inference	KING
Imputation	Phasing	MaCH ShapeIT Beagle
	Imputation	Minimac IMPUTE2 Beagle
	HLA imputation	SNP2HLA
Association	Single variant association/burden test for rare variants	PLINK-1.9 EPACTS
	Meta-analysis	METAL rareMETAL
Annotation	Annotate variants	ANNOVAR
Pathway analysis	Identify enriched functions	INRICH ALIGATOR MAGENTA MEAGA
Candidate gene prioritization	Provide inference for the best candidate genes from associated loci	GRAIL OntoFing DEPICT

Table 2, we lay out some typical quality control procedures (and commonly used software) for genetic studies, and here we use the quality control metrics used by a genetic study on psoriasis for illustration (Tsoi et al., 2012). The quality of genotyped data may be evaluated using different metrics, and only samples/markers with high quality are considered (e.g., only including samples and markers with ≥95% of genotyping rate; using Hardy-Weinberg equilibrium, $P < 1 \times 10^{-6}$ as a cutoff to filter out markers with observed genotype frequencies deviating from expected) (Bush and Moore, 2012). Most often, raw intensity files are provided, which provide valuable quantitative information regarding hybridization intensity. Cluster plots (Figure 1b) can be produced from intensity files to validate the genotypes that have been generated from the genotype calling algorithm. Such plots are particularly important for rare genetic variants, for which genotype calling is often not trivial (Goldstein et al., 2012).

False positive results can creep in through dependencies between related individuals (including cryptic relatedness) and differences in the underlying genetic structures of the case and control populations (population stratification). Under the latter scenario, the difference in allele frequencies will reflect only the systematic ancestry differences between the two groups (Price et al., 2010). This can be particularly challenging if either the cases or the control groups are enriched in outliers for the population being studied and can be problematic even for studies that use shared controls. Kinship coefficient (e.g., KING) (Manichaikul et al., 2010) can be used together with a mixed model (Kang et al., 2010) to address these issues effectively. Principal component analysis or multidimensional scaling, dimension reduction techniques that project genetic data to lower-dimension space, can also be used to generate covariates for association (Price et al., 2010) to address population stratification. By performing principal component analysis/multidimensional scaling analysis together with different populations (e.g., 1000 Genomes [1000 Genomes Project Consortium, 2012]), outliers can be shown by comparing their principal component analysis/multidimensional scaling coordinates with those from the population of interest. Although these are promising approaches for common variants, more advanced techniques may be needed to control for population stratification in the context of rare variant analysis (Lee et al., 2014) and targeted/exome platforms (Wang et al., 2015). Finally, genomic control (λ_{GC}) is a metric of population stratification (Devlin and Roeder, 1999) that may be applied to evaluate association results after principal component analysis adjustment or mixed model correction (Devlin and Roeder, 1999). Under the null hypothesis that genetic variants are not associated with the trait of interest and the population stratification is adequately corrected, the λ_{GC} value would be equal to 1.

Genotype imputation

Genotype imputation is a powerful statistical genetic technique (Marchini and Howie, 2010) that allows combining multiple cohorts (through meta-analysis) by providing a common framework to analyze genotypes derived from different platforms. Meta-analysis can significantly enhance power to show more subtle signals associated with the traits of interest. Variants that are not genotyped in a cohort can be imputed (Das et al., 2016) using reference haplotypes from panels with high variant density (e.g., 1000 Genomes, Haplotype Reference Consortium [McCarthy et al., 2016]). First, the haplotypes of genotyped variants in the cohort are inferred (i.e., phased), with alleles assigned to either the maternal or paternal chromosomes (Delaneau et al., 2012). Then, the haplotype structure and frequencies (as well as the markers present in both the cohort and reference panel) are used to impute genotypes for the missing variants. In addition to single-nucleotide polymorphisms and insertions/deletions, one can also impute HLA alleles and their amino acid sequences for different classical alleles in the major histocompatibility complex region (Jia et al., 2013). This is particularly useful in fine-mapping (i.e., high-resolution mapping for disease-associated variants) major histocompatibility complex associations for immune-mediated diseases, such as psoriasis (Okada et al., 2014).

Because a statistical model is used to infer genotypes for the unobserved markers, they must be represented using a continuous "dosage" value (Howie et al., 2011). Typically, this value is set to between 0 and 2, indicating the expected number of times the alternative allele occurs. So that dosage

values may be used in regression models (i.e., as part of association analysis), genotyped markers are represented in the same way (with 0 indicating that both copies have the reference allele, 2 indicating that both copies have the alternate allele, and 1 being the heterozygous case). By comparing the allele frequency of the marker in the reference samples with that inferred from the cohort, imputation quality metrics measure the accuracy of imputed markers (e.g., r^2 for MiniMac [Das et al., 2016], "info score" for IMPUTE2 [Howie et al., 2011]). Markers with low imputation quality (e.g., $r^2 < 0.7$ [Tsoi et al., 2017]) are removed from the downstream analysis. Imputation works very well for common variants (e.g., a recent study used imputation to evaluate more than 6 times the number of genotyped markers and thus identified five novel disease susceptibility regions for psoriasis [Tsoi et al., 2015b]) but is not as effective for variants with low allele frequencies. Large reference haplotypes (e.g., 1000G via public access [1000 Genomes Project et al., 2015]) or Haplotype Reference Consortium [HRC] via Imputation Server [Das et al., 2016]) can help enhance imputation quality for less common variants.

Applications to interpret association results

The interpretation of GWAS findings is of critical importance to understand how these genetic signals relate to biological events (Foulkes et al., 2017). One of the first steps in downstream analysis is to perform functional annotations for the identified markers (Figure 1c). These annotations can be used to classify the potential role(s) of the implicated variants (e.g., coding or noncoding regions) and to identify nearby genes of interest (Wang et al., 2010). As noted, recent large-scale GWASs have found that disease-associated genetic variants (or *signals*) tend to play regulatory roles. By integrating information from recent large-scale epigenomics projects, such as ENCODE (ENCODE Project Consortium, 2012) and the National Institutes of Health (NIH) Roadmap (Romanoski et al., 2015), we can provide inference for the chromatin states and corresponding cell types of the associated regions.

Once markers have been annotated, pathway analysis can be used to identify biological functions for the genes among the disease loci. Bioinformatics approaches identify the pathways/functions that are enriched among the genes in associated loci (Lee et al., 2012) compared with genes from the (nonsignificant) background regions. Identifying the best candidate genes from disease regions can also be important, especially when designing replication experiments (e.g., resequencing selected candidate genes or regulatory sequences). Various approaches (based on text mining [Raychaudhuri et al., 2009], gene expression [Pers et al., 2015], or ontology [Tsoi et al., 2009]) have been proposed to integrate independent information with traits/tissue types of interest to enhance the prioritization of candidate genes in each locus. Pathway analysis can also prioritize genes that are mapped to the enriched functions (Tsoi et al., 2015a), and network-based approaches capturing gene-gene interactions can be used to identify gene clusters with significant connectivity (Rossin et al., 2011) or shortest distances (Tsoi et al., 2015a). Statistical genetics techniques have also been developed to provide robust estimation of heritability using GWAS data. For example, genome-wide complex trait analysis (i.e., GCTA) uses variance component estimation to estimate the heritability of genetic variants captured by the genotyping platform (Yang et al., 2011).

MULTIPLE CHOICE QUESTIONS

1. Which of the following is NOT a type of array used for genotyping?
   A. Exomechip
   B. Immunochip
   C. Metabochip
   D. Compuchip
2. What is the typical range of values for imputed genotypes?
   A. 0 to 1
   B. 0 to 2
   C. −1 to 1
   D. 0 to 100
3. Which of the following can be used to address population stratification?
   A. Annotation
   B. Genomic control
   C. Multiple testing
   D. Phasing
4. What *P*-value threshold is commonly used for genome-wide significance?
   A. 5×10^{-4}
   B. 5×10^{-6}
   C. 5×10^{-8}
   D. 5×10^{-10}
5. Which of the following is not a priority for GWAS research in skin disease?
   A. Increased sample size and integration across ethnicities
   B. Inferring the biological function of the disease loci identified
   C. Integrating information from clinical data for precision medicine
   D. Identifying differences in gene expression

CHALLENGES AND FUTURE DIRECTIONS

GWASs have facilitated both the generation and evaluation of new hypotheses in basic science and clinical research over the last decade (Claussnitzer et al., 2015; Price et al., 2015; Turner et al., 2012). Most GWASs have been conducted in European populations, with relatively less comprehensive genetic information for other underrepresented populations (e.g., Arabic, Indian). Although GWASs with increased sample size and transethnic components are ongoing (Morris et al., 2016; Paternoster et al., 2015), the current challenges are to provide biological inference for each of the disease loci

identified. Specifically, functional assays need to be in place to test hypotheses developed from GWAS results, providing in silico/in vitro experimental evaluations on the biological effects for the susceptibility loci. In addition, it is important that follow-up studies involve appropriate cell types, because disease-associated regulatory events are usually cell type specific (Farh et al., 2015). Epigenetic and expression data (ENCODE Project Consortium, 2012; Lonsdale et al., 2013) can be used to investigate whether disease-associated genetic variants alter the chromatin accessibility of specific genes and thereby gene expression. Complex cutaneous disorders are unique in that the affected tissues are readily available and relatively easy to obtain, thus making the design and implementation of downstream analysis more efficient, as illustrated in the large-scale transcriptomic studies conducted on skin tissues (Johnston et al., 2017). There is potential for a higher rate of ascertainment bias in self-reported or health record data related to skin conditions for genetic studies, but methods are being developed to address this potential challenge (Tsoi et al., 2017).

The vast amount of data and information obtained from GWAS studies may inform precision/personalized medicine for patients with cutaneous disorders. The next wave of GWASs should aim to integrate information from clinical data by associating genetic data with health records (i.e., Phenome-wide association studies [PheWAS]) (Denny et al., 2013), or drug responses (i.e., pharmacogenetics) (Whirl-Carrillo et al., 2012). The challenge, however, is that GWAS loci alone cannot yet provide clinically relevant risk assessment for disease (such as the risk of development of psoriatic arthritis in a psoriasis patient [Stuart et al., 2015]). Moving forward, efforts should focus on integrating information from GWASs with a variety of other clinical biomarkers and omics data (i.e., proteomics, metabolomics, transcriptomics, etc.) to produce useful tests to allow clinical decision making for individualized health care.

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

This work was supported by the Arthritis National Research Foundation and the National Psoriasis Foundation (LCT and MTP) and by awards from the National Institutes of Health (K01AR072129 to LCT and R01AR042742, R01AR050511, R01AR054966, R01AR063611, and R01AR065183 to JTE). LCT was also supported by the Dermatology Foundation. LCT and JTE were supported by the Dawn and Dudley Holmes Foundation and the Babcock Memorial Trust. JTE is supported by the Ann Arbor Veterans Affairs Hospital.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to this paper. Teaching slides are available as supplementary material.

REFERENCES

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.

Bush WS, Moore JH. Chapter 11: genome-wide association studies. PLoS Comput Biol 2012;8(12):e1002822.

Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. N Engl J Med 2015;373:895–907.

Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther 2011;13:101.

Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284–7.

Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods 2012;9:179–81.

Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol 2013;31:1102–10.

Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997–1004.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489(7414):57–74.

Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. Eur J Hum Genet 2016;24:1202–5.

Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 2015;518(7539):337–43.

Foulkes AC, Watson DS, Griffiths CEM, Warren RB, Huber W, Barnes MR. Research techniques made simple: bioinformatics for genome-scale biology. J Invest Dermatol 2017;137:e163–8.

Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M, et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. Bioinformatics 2012;28:2543–5.

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.

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.

Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda) 2011;1:457–70.

Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One 2013;8(6):e64683.

Jin Y, Andersen G, Yorgov D, Ferrara TM, Ben S, Brownson KM, et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. Nat Genet 2016;48:1418–24.

Johnston A, Sarkar MK, Vrana A, Tsoi LC, Gudjonsson JE. The molecular revolution in cutaneous biology: the era of global transcriptional analysis. J Invest Dermatol 2017;137:e87–91.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet 2010;42:348–54.

Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. Bioinformatics 2012;28:1797–9.

Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet 2014;95:5–23.

Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. Nat Genet 2013;45:580–5.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics 2010;26:2867–73.

Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet 2010;11:499–511.

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48:1279–83.

Morris DL, Sheng Y, Zhang Y, Wang YF, Zhu Z, Tombleson P, et al. Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. Nat Genet 2016;48:940–6.

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.

Okada Y, Han B, Tsoi LC, Stuart PE, Ellinghaus E, Tejasvi T, et al. Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. Am J Hum Genet 2014;95:162–72.

1000 Genomes Project Consotium. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491(7422):56–65.

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature 2015;526(7571):68–74.

Ott J. Analysis of human genetic linkage. 3rd ed. Baltimore, MD: Johns Hopkins University Press; 1999.

Paternoster L, Standl M, Waage J, Baurecht H, Hotze M, Strachan DP, et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. Nat Genet 2015;47:1449–56.

Pers TH, Karjalainen JM, Chan Y, Westra HJ, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun 2015;6:5890.

Price AL, Spencer CC, Donnelly P. Progress and promise in understanding the genetic basis of common diseases. Proc Biol Sci 2015;282(1821): 20151684.

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.

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.

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(6):e1000534.

Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 2011;43(11): 1066–73.

Romanoski CE, Glass CK, Stunnenberg HG, Wilson L, Almouzni G. Epigenomics: Roadmap for regulation. Nature 2015;518(7539):314–6.

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(1):e1001273.

Stuart PE, Nair RP, Tsoi LC, Tejasvi T, Das S, Kang HM, et al. Genome-wide association analysis of psoriatic arthritis and cutaneous psoriasis reveals differences in their genetic architecture. Am J Hum Genet 2015;97: 816–36.

Tang H, Jin X, Li Y, Jiang H, Tang X, Yang X, et al. A large-scale screen for coding variants predisposing to psoriasis. Nat Genet 2014;46: 45–50.

Tsoi LC, Boehnke M, Klein RL, Zheng WJ. Evaluation of genome-wide association study results through development of ontology fingerprints. Bioinformatics 2009;25:1314–20.

Tsoi LC, Elder JT, Abecasis GR. Graphical algorithm for integration of genetic and biological data: proof of principle using psoriasis as a model. Bioinformatics 2015a;31:1243–9.

Tsoi LC, Spain SL, Ellinghaus E, Stuart PE, Capon F, Knight J, et al. Enhanced meta-analysis and replication studies identify five new psoriasis susceptibility loci. Nat Commun 2015b;6:7001.

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.

Tsoi LC, Stuart PE, Tian C, Gudjonsson JE, Das S, Zawistowski M, et al. Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. Nat Commun 2017;8:15382.

Turner AR, Kader AK, Xu J. Utility of genome-wide association study findings: prostate cancer as a translational research paradigm. J Intern Med 2012;271:344–52.

Wang C, Zhan X, Liang L, Abecasis GR, Lin X. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. Am J Hum Genet 2015;96:926–37.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38(16):e164.

Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther 2012;92:414–7.

Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Magi R, et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc 2014;9:1192–212.

Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88:76–82.

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.