Genetics of atopic dermatitis: from DNA sequence to clinical relevance

Mari Løset\textsuperscript{1,2}, Sara J. Brown\textsuperscript{3,4}, Marit Saunes\textsuperscript{2,5}, Kristian Hveem\textsuperscript{1,6}

\textsuperscript{1}K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway
\textsuperscript{2}Department of Dermatology, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway
\textsuperscript{3}Skin Research Group, School of Medicine, University of Dundee, Dundee, United Kingdom
\textsuperscript{4}Department of Dermatology, Ninewells Hospital and Medical School, Dundee, United Kingdom
\textsuperscript{5}Department of Cancer Research and Molecular Medicine, NTNU, Norwegian University of Science and Technology, Trondheim, Norway
\textsuperscript{6}HUNT Research Centre, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

Short Title: Genetics of atopic dermatitis

Corresponding Author:
Mari Løset, MD, PhD
K.G. Jebsen Center for Genetic Epidemiology
Department of Public Health and Nursing
NTNU, Norwegian University of Science and Technology
Postboks 8905
N-7491 Trondheim, Norway
Tel: +47 90799117
E-mail: mari.loset@ntnu.no

Key Message: Genetic discoveries have added significantly to the biological understanding of atopic dermatitis. In this review, we describe some of the methods used for this purpose, their findings, and future directions in the field.


This is the peer-reviewed but unedited manuscript version of the following article: Løset, M., Brown, S., Saunes, M., & Hveem, K. "Genetics of Atopic Dermatitis: From DNA Sequence to Clinical Relevance", Dermatology 2019 (DOI: 10.1159/000500402). The final, published version is available at https://doi.org/10.1159/000500402.
1. Abstract

Atopic dermatitis (AD) is a complex disease that is thought to be triggered by environmental factors in genetically susceptible individuals. Twin studies have estimated the heritability of AD to be approximately 75%, with the null (loss-of-function) mutations of the gene encoding filaggrin (FLG) (chromosome 1q21.3) as the strongest known genetic risk factor. The discovery of the filaggrin gene was important in the emerging model for AD pathogenesis, combining skin barrier function with adaptive and innate immunity. Assisted by recent development of large-scale high-throughput genomics, more than 30 genetic loci have been linked to AD across different populations. Identification of these loci, together with functional studies, has already provided new insights into disease biology and identified novel drug targets. Further, these susceptibility loci are laying the groundwork for phenome-wide association studies to test their multiple phenotype relationships and application of Mendelian Randomization to investigate causal relationships. Despite many known genes, a majority of the genetic risk for AD is yet unexplored. Therefore, studies investigating refined phenotype groups, low-frequency and rare genetic variation, gene-gene and/or gene-environment interactions, epigenetic mechanisms and data from multi-omics technologies are warranted. In this review we describe genetic discoveries for AD, including results from candidate gene studies, studies of AD-like genetic diseases, genome-wide association studies and genetic sequencing studies. We explain how some of these genetic discoveries have unraveled new mechanistic insights into the pathogenesis of AD and exemplifies how personal genetic data could be used for preventive strategies and tailored treatment regimen (i.e. precision medicine).

2. Introduction

Atopic dermatitis (AD, synonymous with 'atopic eczema') is a chronic inflammatory skin disease with a lifetime prevalence of 15-20% in developed countries [1, 2]. AD usually begins in early childhood, and in 70% of cases the disease manifests before five years of age [3]. The disease is characterized by dry skin, pruritus, and eczematous lesions, which can lead to intense itch, sleep deprivation, and diminished self-esteem [2]. The disease shows large clinical heterogeneity related to age, ethnicity and disease severity [4, 5]. Consequently, topical and systemic treatments are tailored to clinical need with the aim to manage symptoms and obtain long-term disease control [6, 7].
AD is associated with atopic (asthma, allergic rhinitis, food allergies) and non-atopic (e.g., inflammatory diseases, phycological disturbances) comorbidities [8, 9]. For the atopic comorbidities, the association is strong, and family and genetic studies indicates a shared genetic susceptibility [10-12]. For the non-atopic comorbidities, the association is less clear, and underlying mechanisms are likely to be complex and involve common genetic susceptibilities, systemic inflammation, environmental exposures, medication and lifestyle factors [8].

A family history of atopic disease is the strongest known risk factor for AD [13]. This observation formed the rationale for genetic investigations to identify novel mechanisms leading to AD and related comorbidities. The search for new genetic risk factors is motivated by already highly successful discoveries in the field, and in particular by the identification of the filaggrin gene (FLG) in 2006 [14]. This completely changed the understanding of atopic diseases; from a purely immunological view, the initial pathogenesis is now known to include defects of the epidermal barrier, immunological dysregulation and triggering factors [2] and this have had implications for treatment strategies [15, 16].

Substantial advances in the field of genetic epidemiology over the past two decades are laying the framework for new gene discoveries for complex diseases. One of the groundbreaking events in the field was when the Human Genome Project published the full human genome in 2004 [17, 18]. Together with large population-based sample sizes, the innovation of cost-effective microarray technology and new analytical tools, opened for large-scale high-throughput genomics. The field has moved from candidate gene and family-based studies to large population-scale whole-genome genotyping including >1 million people and > 20 million genetic variants [19].

In this review, we summarize some of the known genetic associations of AD in relation to the pathogenesis of the disease and illustrate how these associations can be used for phenome-wide association studies (PheWAS) to test their multiple phenotype relationships and application of Mendelian Randomization (MR) to investigate causal relationships. An overview of some of the investigative strategies included in this paper is given in Figure 1. We focus primarily on research involving genetic studies and encourage the readers to study results from additional omics-data, i.e. epigenomics, transcriptomics, proteomics and
metabolomics in other publications [20-25]. Finally, we discuss future directions in the field including the potential avenues of precision medicine.

3. Establishing a genetic component and gene discovery

The heritability (the proportion of phenotypic variance due to genetic variance) of a disease or trait is usually estimated by observing patterns of inheritance among individuals either in family and twin studies or in population-based studies [26]. The first descriptions of familial aggregation of atopic skin disease goes 2,000 years back, when the Roman historian Suetonius described that Emperor Augustus (63 BC - 14 AD) had dry, itchy patches on his skin and suffered from a seasonal respiratory disorder, and both his grandson and his great-grandnephew suffered from symptoms of atopy [27]. More recent family studies have indicated a three- to fivefold increased risk for a person to develop AD if one or both parents have a history of AD [13, 28]. Twin studies of AD show that monozygotic twins have approximately three times higher concordance rate than dizygotic twins [29]. A systematic review which included information from 35,155 twin pairs estimated the heritability of AD to be approximately 75% [29], which is considered to be a high heritability for a complex trait.

3.1 Candidate gene studies. FLG. By far the strongest known genetic risk factor for AD is null (loss-of-function) mutations of the gene encoding filaggrin (FLG), a gene located in the epidermal differentiation complex on chromosome 1q21.3 [14, 30, 31]. Filaggrin has multiple interrelated functions contributing to skin barrier development and maintenance [30], and null mutations of FLG lead to truncation of profilaggrin and loss of filaggrin expression. Individuals with filaggrin deficiency experience an increased skin permeability, facilitating the effect of environmental allergens, irritants and microbes, and initiation of an inflammatory cascade. The link between FLG and AD was identified in 2006 when studying ichthyosis vulgaris [32], a genetic disease that displays AD-like characteristics. FLG was further associated with AD [14], a discovery that changed the view of the pathogenesis of AD. The focus shifted from an immunological view including imbalance of T-cells to inclusion of a functional disruption of the epidermal barrier as the primary pathogenic process [33]. About 10% of the European and Japanese ancestry population carry a null mutation within FLG exon 3 and they have mild ichthyosis and a threefold increased risk of AD compared with the general population [2, 30]. This demonstrates a strong effect for a single gene in the context of a complex disease [34].
Other genes. Based on their theoretical role in disease pathogenesis, a large number of candidate gene studies have been performed to elucidate the genetic background of AD [20, 35, 36]. This includes genes involved in epidermal differentiation, skin immunity, or systemic immunity [36]. Highlighted genes identified by candidate gene studies comprise the genes encoding interleukin (IL)-4, the IL-4 receptor, and IL-13 [37, 38], all lying in the Th2 cytokine cluster on chromosome 5q31.1. This locus is also robustly associated with AD in genome-wide association studies (GWAS) (see below).

3.2 Studies of AD-like genetic diseases. Detailed studies of genetic diseases that display AD-like characteristics have provided insight into monogenic drivers of disease pathogenesis [39]. These studies have focused on diseases characterized by skin barrier dysfunction (e.g. ichthyosis vulgaris, generalized peeling skin, Netherton syndrome), multi-system atopic inflammation (e.g. Severe dermatitis, multiple Allergies and Metabolic wasting, SAM) and immunodeficiency with skin manifestations resembling AD (e.g. Hyper-IgE, Omenn syndrome) [39]. To illustrate; the syndrome of SAM is a rare (prevalence <1/1 million) genetic disease that shows AD-like manifestations including congenital erythroderma, superficial skin erosions, fine scales and palmoplantar keratoderma, food allergies and increased IgE levels. Whole-exome sequencing revealed homozygous loss-of-function mutations in DSG1 (encoding desmoglein) [40] and heterozygous mutation in DSP (encoding desmoplakin) [40, 41] that segregated with disease. The genetic mutations of SAM lead to compromised barrier function and DSG1 deficiency was associated with increased expression of genes encoding allergy-related cytokines. This informed the knowledge of epidermal barrier dysfunctions leading to local and systemic atopic inflammation [40, 41].

3.3 Genome-wide association studies (GWAS). GWAS is an experimental design used to test the association between thousands, and even millions, of genetic variants with a disease outcome or trait [42]. The primary aim of these studies is to increase the understanding of genetic risk of disease and provide a foundation for functional and mechanistic follow-up studies that will enable more effective detection, prevention and treatment. GWAS have proven successful at mapping areas on the human chromosomes (loci) that influence AD [43-50], and since the first published study on AD in 2009 [43], >30 genetic loci have been identified, where the majority show a slight to moderate risk for disease development [47]. An overview of these loci, their likely candidate genes and known or proposed functions have recently been published [2, 4], including a review focusing on associations in diverse, non-
European ethnic groups [4]. The first GWAS on AD was performed on a German population of 939 individuals with AD and 975 controls as well as on 270 complete nuclear families with two affected siblings, and their replication population included 2,637 cases and 3,957 controls [43]. They detected the FLG locus and identified a novel susceptibility locus on chromosome 11q13.5 (top hit rs7927894), an intergenic region of unknown function. The study found that approximately 13% of individuals with European ancestry are homozygous for rs7927894, and that their risk to develop AD is 1.47 times that of non-carriers [43]. Subsequent GWAS replicated the FLG and chromosome 11q13.5 loci across different populations, and added several new loci associated with AD risk. The majority of GWAS loci identified are involved in skin barrier development and immunological dysfunctions, in particular in innate immune signaling and T cell activation and differentiation [2]. Even though most genetic variants that confer AD susceptibility have relatively small effect sizes, they have provided important insights into the biology of the disease and subsequently supported the genetic basis for established drugs [47] and identified potential drug targets for allergic diseases [10].

Since the first GWAS in 2009, large scale collaborations have made progressively larger sample sizes available, and the process of predicting genotypes through imputation, a method to estimate the genotype at genetic markers that not are directly genotyped, is commonly used. In 2015, the EAGLE consortium meta-analyzed >15 million genetic variants in 21,399 cases and 95,464 controls [47]. Their replication phase included 32,059 cases and 228,628 controls. The study included populations of European, African, Japanese and Latino ancestry, and identified ten new risk loci, increasing the total number of known AD risk loci to 31 [47]. Of the identified loci, several are common for other atopic conditions (asthma, allergic sensitization, self-reported allergy), supporting common atopic mechanisms [47]. This was further investigated by Ferreira et al., where genetic risk factors were identified between asthma, hay fever and eczema [10]. They identified 136 independent risk variants, of which 73 were novel; disease-specific effects were detected for only six variants, confirming shared genetic risk across atopic phenotypes. The risk variants were shown to dysregulate the expression of immune-related genes [10]. These results were further explored using gene-based tests and additional 11 risk loci for allergic diseases were discovered [11].

### 3.4 Phenome-wide association studies (PheWAS)

While GWAS analyze thousands to millions of genetic variants per phenotypic trait, PheWAS inverts this study design to analyze many phenotypic traits for association with a single genetic variant. The goal of this study...
design is to elucidate genetic variants with multiple phenotype relationships (e.g. comorbidities and pleiotropy) [51]. The method has also successfully been used for drug discovery and repositioning, prediction of adverse drug events and selection of indications for clinical trials [52, 53]. Figure 2 shows the results of a PheGWAS analysis on the UK Biobank PheWeb (http://pheweb.sph.umich.edu:5003) based on GWAS of 28 million variants across 1,403 ICD-derived traits identified in 408,961 individuals [54] for the lead variant (rs61816761) within the FLG gene. As expected and due to known atopic comorbidities, we find that this genetic variant is associated with asthma, dermatitis, and allergic reaction to food (the latter only nominally significant). This variant is also associated with diseases of sebaceous glands, a result that is suggestive of pleiotropy or shared disease mechanisms.

3.5 Genetic sequencing studies. To date, GWAS has mainly been conducted using genotype arrays that are designed to measure genetic variation at common variants. Imputation allows for estimation of genotypes of rare variants not included on genotyping arrays, but this is dependent on the size and quality of the imputation reference panel. By their nature, rare variants are so infrequently seen that those detected often do not reach statistical significance in GWAS, even in large sample sizes. Due to this, rare genetic variants (population minor allele frequency typically <1%) have not yet been investigated to the same extent as common variants, even though they may have from moderate to high effects on complex traits [55, 56]. Facilitated by the steady drop in sequencing costs, large-scale studies of rare variants are now possible through both whole-exome sequencing (WES) and whole-genome sequencing (WGS). WES targets all of the protein-coding regions of the genome (known as the exome) and has proven to be a useful tool to investigate rare diseases and in the search for rare variants causing polygenic diseases [55]. For AD, the WES studies performed to date have had relatively small sample sizes [57-62]. The first published study included 22 Ethiopian individuals with ichthyosis vulgaris and AD, and identified several rare variants suggesting a heterogenous disease pathogenesis [57]. Recently, 43 probands of 42 Bangladeshi families with severe AD were investigated using WES in combination with rare variant enrichment analysis [60]. This identified a rare putative loss-of-function allele in FLG as the major component of disease susceptibility, in addition to potential novel risk genes within chromosome 1 (TCHHL1, ADCY10, MTF1, MAST2), 6 (SCAND3), 9 (ORM2), 10 (MCM10), 11 (PHLDB1, PANX3) and 12 (CUX2) [60]. Further, WES showed that dominant negative loss-of-function mutations in CARD14 were associated with severe AD [62]. Upregulation of the same gene leads to psoriasis [63, 64], in line with previously identified shared loci with
opposing mechanisms for psoriasis and AD [48]. Larger and population-based WES and WGS to systematically investigate less frequent genetic variations and association to disease are ongoing in several cohorts [65]. In October 2018, it was announced that the United Kingdom’s National Health Services (NHS) are currently planning to sequence the genomes of one million people, and in the next five years, they are planning to sequence five million genomes [66].

4. Investigating causal relationships using Mendelian Randomization (MR)

Causal relationships can be investigated with MR, a method that uses genetic variants to make causal inference of non-genetic environmental exposures [67]. These studies leverage Mendel’s second law that the inheritance of one trait is independent of other traits. When satisfied, genetic variants can be used as instrumental variables to estimate causal effects. Additional, germline genetic variants are fixed at conception and this happens prior to the onset of disease, therefore these analyses are less susceptible to confounding and reverse causality [67, 68]. Through GWAS, the increasing identification of genetic determinants of modifiable exposures and phenotypes has provided a valuable source of data to use for MR analyses. Examples of MR used in dermatology include investigations of direction(s) of causality for psoriasis with higher CRP levels (supports a non-causal relationship) [69] and higher body mass index (BMI) (supports a causal relationship) [70, 71]. For AD, the relationship to lower vitamin D levels and AD has been tested and evidence indicates a non-causal relationship [72]. Observational epidemiological studies have reported an association of high BMI with risk of AD [73]. If high BMI is a causal risk factor for AD this may have implications for treatment and prevention strategies because BMI is potentially modifiable. An ongoing MR study is examining this potential causal relationship (unpublished data). The study population includes >400,000 individuals from the UK Biobank and the Nord-Trøndelag Health Study (The HUNT Study) and genetic variants identified by GWAS to be associated with BMI and AD as instrumental variables.

5. Translation into clinical relevance

The results of MR studies as described above may have implications for treatment and prevention strategies for AD. Further, genetic information leading to focus on skin barrier dysfunction as the key precursor to increased skin permeability, inflammation and percutaneous allergic sensitization [5] have given rise to the development of novel treatment strategies that target the skin barrier or cutaneous inflammation [15, 16]. However, despite an
increasing knowledge about disease mechanisms, there is still a very limited number of systemic immunomodulating biologics that have been approved for the treatment of moderate-to-severe AD [74]. One exception is Dupilumab, a fully human monoclonal antibody targeting the IL-4 and IL-13 pathway [75]. Dupilumab was the first biologic to receive US Food and Drug Administration (FDA) and European Medicines Agency (EMA) approval for the treatment of adults with moderate-to-severe AD. Clinical trials have shown that Dupilumab significantly improved clinical outcomes and quality of life in adults with moderate-to-severe AD [76]. Many other new treatments are under development with several biologics and small molecule antagonists in phase II and phase III trials [2].

AD shows large clinical heterogeneity, and a recent study divided AD into six sub-phenotypes, which showed different risk profiles and diverse prognosis [77]. With the increasing knowledge of genetic associations with AD in combination with the steady drop in genotyping costs, one can foresee a future where health care providers could have access to the patient’s genetic data [78]. Personal genetics could help stratifying individuals into AD disease subsets, and direct medical interventions by making predictions about disease risk, tailoring the right therapeutic strategy and prevention programs for the right person at the right time ('the promise of precision medicine'). This could again reduce the risk of adverse events and the costs related to treatments. For AD, genetic risk could be ascertained from birth, and thereby identify individuals at risk for atopic diseases. Such knowledge could allow early interventions to prevent disease, e.g. using daily emollients to enhance a defective skin barrier [79, 80]. It has been suggested that early preventive treatments of AD could reduce the development of associated comorbidities, such as asthma [15]. To calculate a polygenic risk score per individual for AD could be a method of identifying individuals for preventive measures. A polygenic risk score adds together the contributions of all the small genetic effects as estimated through GWAS and provides a continuous and quantitative measure of genetic risk [81]. Polygenic risk scores could create one of the most powerful genetic diagnostics to date, such as demonstrated for e.g. coronary artery disease and atrial fibrillation [82]. To date, there is no routine in integration of genetic testing in the clinics for inflammatory skin disease. However, work to predict individual treatment responses in dermatology is ongoing through e.g. the Psoriasis Stratification to Optimize Relevant Therapy (PSORT) consortium [83].
Despite the success of association analyses among complex diseases [42], there is still a significant proportion of genetic risk that has not yet been explained for most traits. This phenomenon has been called the 'missing heritability' of complex diseases [84, 85]. For AD, the identified GWAS loci explains approximately 15% of the variance in liability [47], which is in line with other highly heritable complex traits. There are many factors that could explain this still 'missing heritability' ten years after the first large scale GWAS. This includes effects due to marked heterogeneity of AD, the cumulative effects of multiple genetic variations, rare genetic variations, structural variations of the genome such as copy number variants (CNV), the existence of gene-gene and/or gene-environment interactions, and heritable epigenetic mechanisms [26]. The present times are highly promising for investigating more of these effects.

Technological advances have opened new possibilities for analyzing multi-omics data [86]. The likelihood of identifying a true gene or pathway increases if data from multiple levels of biological data support the same association [86]. By multi-staged and meta-dimensional analyses of multi-omics one could identify effective models that predict disease status, discover biomarkers and lead to an increased understanding of the role of genetics and genomics in complex traits [86] and in AD in specific [25, 39]. Multiple levels of biological data are illustrated in Figure 3 (adapted from [86]), and includes data from the genome, epigenome, transcriptome, proteome and metabolome.

7. Conclusions

Over the last decade there has been a paradigm shift in our understanding of the mechanisms causing AD. Moving away from a purely immunological view, genetic studies have played a major role in understanding defects of the epidermal barrier. In the future we hope for similar mechanistic breakthroughs lead by gene discoveries, that these discoveries will continue to inform drug development efforts and safety measures, and that personal genetic information may be included in routine clinical care to aid prevention and treatment of AD. The groundwork for these implementations is the growing large-scale high-throughput genomics, including both WES and WGS, made possible by population-based biobanks in combination with information from electronic health records, and methods developed to handle these complex health data with the necessary computational and statistical expertise [65]. Together such activities will increase the interpretation and functional validation, and future studies
will most likely take gene-gene and gene-environmental effect into larger account and include multi-omics technologies.

8. Ethics
The authors have no ethical conflicts to disclose.

9. Disclosure Statement
M.L., S.J.B., M.S. and K.H. declare no relevant conflicts of interest.

10. Funding Sources
M.L. and K.H. work in a research unit funded by Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, NTNU; The Liaison Committee for education, research and innovation in Central Norway; and the Joint Research Committee between St. Olavs hospital and the Faculty of Medicine and Health Sciences, NTNU. ML was supported by a research grant from the Liaison Committee for education, research and innovation in Central Norway. S.J.B. holds a Wellcome Trust Senior Research Fellowship in Clinical Science (106865/Z/15/Z).

11. Author Contributions
This manuscript was planned and written in collaboration by the authors.

12. Acknowledgments
The authors would like to thank Janne Tellefsen for help with design of Figure 3, Ben M. Brumpton for commenting on the manuscript and revising the section on Mendelian Randomization, and Brooke N. Wolford for insights on genetic sequencing studies and constructive criticism of the manuscript. We acknowledge researchers whose work we could not cite because of space constraints.

13. References


Fig. 1. Some of the investigative strategies that can be used to increase the knowledge of molecular mechanisms involved in a phenotypic trait. Candidate gene studies are performed on genes that are suggested to be involved in the genesis of a phenotypic trait because of their theoretically role in the trait being investigated. Genome-wide studies include genome-wide association studies (GWAS) and genetic sequencing studies. GWAS are mainly conducted in large population-based samples using genotype arrays that are designed to measure genetic variation at common variants. Genetic sequencing studies include both whole-exome sequencing (WES) and whole-genome sequencing (WGS). WES targets all of the protein-coding regions of the genome, and WGS targets the complete genome sequence. The results of candidate gene and genome-wide studies can be used in phenome-wide association studies (PheWAS) that analyze many phenotypic traits for association with a single genetic variant to test multiple phenotype relationships, and further, they could be used as instrumental variables in Mendelian Randomization (MR) studies for the investigation of causal relationships.
Fig. 2. PheGWAS plot of the lead variant (rs61816761) in FLG in UK Biobank. The variant is associated with dermatitis (P value $3.2 \times 10^{-10}$) in 2,110 cases and 40,4817 controls. As expected, the variant is also associated with asthma (P value $5 \times 10^{-12}$) and nominally associated with allergic reaction to food (P value $5.8 \times 10^{-5}$). The genetic variant is further associated to diseases of sebaceous glands (P value $2.1 \times 10^{-7}$) and sebaceous cyst (P value $4.6 \times 10^{-7}$).

Fig. 3. Technological advances have made efficient integration of multi-omics possible. This includes combined studies of the genome, epigenome, transcriptome, proteome and metabolome and the observed phenome. Adapted from [86].