

# Intragenic Copy Number Variation within *Filaggrin* Contributes to the Risk of Atopic Dermatitis with a Dose-Dependent Effect

Sara J. Brown<sup>1,2,3</sup>, Karin Kroboth<sup>1</sup>, Aileen Sandilands<sup>1</sup>, Linda E. Campbell<sup>1</sup>, Elizabeth Pohler<sup>1</sup>, Sanja Kezic<sup>4</sup>, Heather J. Cordell<sup>5</sup>, W.H. Irwin McLean<sup>1</sup> and Alan D. Irvine<sup>3,6</sup>

Loss-of-function variants within the *filaggrin* gene (*FLG*) increase the risk of atopic dermatitis. *FLG* also demonstrates intragenic copy number variation (CNV), with alleles encoding 10, 11, or 12 filaggrin monomers; hence, CNV may affect the amount of filaggrin expressed in the epidermis. A total of 876 Irish pediatric atopic dermatitis cases were compared with 928 population controls to test the hypothesis that CNV within *FLG* affects the risk of atopic dermatitis independently of *FLG*-null mutations. Cases and controls were screened for CNV and common *FLG*-null mutations. In this population the 11-repeat allele was most prevalent (allele frequency 51.5%); the 10-repeat allele frequency was 33.9% and the 12-repeat allele frequency was 14.6%. Having excluded *FLG* mutation carriers, the control group had a significantly higher number of repeats than cases ( $\chi^2 P = 0.043$ ), and the odds ratio of disease was reduced by a factor of 0.88 (95% confidence interval 0.78–0.98,  $P = 0.025$ ) for each additional unit of copy number. Breakdown products of filaggrin were quantified in tape-stripped stratum corneum from 31 atopic dermatitis patients and urocanic acid showed a positive correlation with total copy number. CNV within *FLG* makes a significant, dose-dependent contribution to atopic dermatitis risk, and therefore treatments to increase filaggrin expression may have therapeutic utility.

*Journal of Investigative Dermatology* advance online publication, 10 November 2011; doi:10.1038/jid.2011.342

## INTRODUCTION

Atopic dermatitis (eczema) is an itchy inflammatory skin disease that affects ~10.7% of children in the United States (Shaw *et al.*, 2011) and up to 25% in the United Kingdom (Kay *et al.*, 1994; Shamssain, 2007). Atopic dermatitis can cause significant morbidity (Lewis-Jones, 2006) and the associated family stress is a substantial social and financial burden (Williams, 2005).

Loss-of-function mutations in the filaggrin gene (*FLG*; OMIM \*135940) cause the dry scaly skin condition ichthyosis vulgaris (Smith *et al.*, 2006) (OMIM #146700) and are strongly and significantly associated with atopic dermatitis (Palmer *et al.*, 2006; Rodríguez *et al.*, 2009; van den Oord

and Sheikh, 2009). *FLG* is expressed in terminally differentiating keratinocytes in the outermost layers of the human epidermis as a large, insoluble polyprotein, profilaggrin. Each inactive profilaggrin molecule is dephosphorylated and cleaved into 10, 11, or 12 functional filaggrin monomers (Gan *et al.*, 1990; Sandilands *et al.*, 2007b). Filaggrin plays a key role in epithelial barrier function: it aggregates and aligns keratin bundles in the cornified cell envelope and is further degraded to release hygroscopic amino acids, constituents of the “natural moisturizing factor” in human skin (Rawlings and Harding, 2004; Sandilands *et al.*, 2009).

*FLG* is located within the epidermal differentiation complex on chromosome 1q21, a dense cluster of genes involved in keratinocyte differentiation (Mischke *et al.*, 1996). *FLG* belongs to the “fused” S100 gene family, characterized by a repetitive sequence in a large third exon (Sandilands *et al.*, 2009). The human *FLG* gene is polymorphic, with common allelic variants showing 10, 11, or 12 nearly identical tandem repeats in exon 3, each ~972 bp in length (Gan *et al.*, 1990; Sandilands *et al.*, 2007b). The *FLG* locus therefore demonstrates intragenic copy number variation, using the definition of copy number variation (CNV) proposed by Redon *et al.* (2006) and Alkan *et al.* (2011). Each repeat at the DNA level encodes a complete copy of the 324 amino-acid filaggrin polypeptide, the functional protein product of the gene. Alleles encode 10, 11, or 12 copies of the filaggrin monomer, and consequently CNV may affect the quantity of filaggrin expressed in the epidermis.

<sup>1</sup>Dermatology and Genetic Medicine, Division of Molecular Medicine, University of Dundee, Dundee, UK; <sup>2</sup>Department of Dermatology, Ninewells Hospital and Medical School, Dundee, UK; <sup>3</sup>Department of Clinical Medicine, Trinity College Dublin, Dublin, Ireland; <sup>4</sup>Academic Medical Center, Coronel Institute, Amsterdam, The Netherlands; <sup>5</sup>Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK and <sup>6</sup>The National Children's Research Centre, Our Lady's Children's Hospital, Dublin, Ireland

Correspondence: Sara J. Brown, Dermatology and Genetic Medicine, Division of Molecular Medicine, College of Life Sciences and College of Medicine, Dentistry & Nursing, Medical Sciences Institute, University of Dundee, Dundee, DD1 5EH, UK. E-mail: s.j.brown@dundee.ac.uk

Abbreviations: CNV, copy number variation; *FLG*, filaggrin gene; PCA, pyrrolidone-5-carboxylic acid; SNP, single-nucleotide polymorphism; UCA, urocanic acid

Received 20 June 2011; revised 2 August 2011; accepted 27 August 2011

Early CNV studies, based on BAC (Bacterial Artificial Chromosome) arrays or low-resolution oligonucleotide platforms, were able to identify CNVs >100 kb (Iafate *et al.*, 2004; Sebat *et al.*, 2004; Redon *et al.*, 2006; Alkan *et al.*, 2011). More recent studies have achieved genome-wide assessment of CNVs above 450–500 bp in length using comparative genomic hybridization (Conrad *et al.*, 2010; Craddock *et al.*, 2010), but this technique currently lacks the power to study rare alleles or to detect an odds ratio of <1.4 (Craddock *et al.*, 2010). We therefore used detailed PCR-based genotyping strategies to define size variants of exon 3 of *FLG* to investigate whether CNV contributes to the risk of atopic dermatitis.

## RESULTS

A case-control study using a total of 925 Irish pediatric atopic dermatitis cases and 998 Irish population controls was used to test the hypothesis that CNV within *FLG* affects atopic dermatitis risk. Of these, 876 cases (94.7%) and 928 controls (93.0%) were successfully genotyped; the demographic and clinical data are summarized in Table 1.

### FLG copy number variants in the Irish population

A total of 100 consecutive atopic dermatitis cases were screened by long-range PCR and there was no evidence of CNV within the 5' portion of *FLG* exon 3, representing repeats numbered 1 to 7 (Figures 1 and 2). All 925 cases and 998 controls were screened using a separate long-range PCR to identify copy number variation within the 3' portion of *FLG* exon 3, representing repeats 7 to the end of the repetitive region (Figures 1 and 3) and a TaqMan allelic discrimination assay for single-nucleotide polymorphism (SNP) rs12730241. In the Irish control population, the most prevalent allele size was 11 repeats, representing 51.5% of alleles; the 10-repeat allele frequency was 33.9% and the 12-repeat allele frequency was 14.6%. Allele counts in cases and controls are summarized in Table 2.

### FLG-null mutation screening

All four prevalent *FLG*-null mutations were successfully typed in 855 out of the 876 (97.6%) cases genotyped for copy number and 926 out of 928 (99.8%) controls genotyped for copy number. Of the fully genotyped individuals, 327 (38.2%)

of cases and 102 (11.0%) of controls carried one or two *FLG*-null mutations; 528 cases and 824 controls were of wild-type genotype for the four *FLG*-null mutations (Table 1).

### FLG copy number affects the risk of atopic dermatitis

*FLG*-null mutation carriers were excluded from further analysis to allow independent assessment of the CNV effect. After exclusion of the *FLG*-null mutation carriers,  $\chi^2$  testing showed a significantly higher total copy number in the controls than in the atopic dermatitis cases ( $P=0.043$ ). Furthermore, logistic regression analysis of the total copy number as continuous data (considering 20, 21, 22, 23, and 24 copies) showed that the addition of each subsequent filaggrin-encoding unit reduces the odds ratio of disease by a factor of 0.88 (95% confidence interval 0.78–0.98,  $P=0.025$ ). Therefore, an individual having two 12-repeat alleles, compared with an individual with two 10-repeat alleles, has a risk reduction of 0.88 to the power of 4, meaning a 0.60 times reduced risk of developing atopic dermatitis.

### The concentrations of filaggrin breakdown products correlate with FLG copy number

Linear regression analysis comparing filaggrin breakdown products quantified by HPLC of tape-stripped stratum corneum from 31 atopic dermatitis cases (Kezic *et al.*, 2011) showed a small but statistically significant correlation between urocanic acid concentration (UCA) and *FLG* total copy number ( $r=0.42$ ,  $P=0.018$ ; Supplementary Figure S1a online). The value  $r^2=0.18$  indicates that 18% of the variation of UCA in the stratum corneum samples may be attributed to *FLG* CNV. The concentrations of two other breakdown products of filaggrin, histidine and pyrrolidone-5-carboxylic acid (PCA), each also showed a small positive correlation with *FLG* total copy number (histidine  $r=0.20$ , PCA  $r=0.28$ ) but these were not statistically significant (Supplementary Figure S1b and c online).

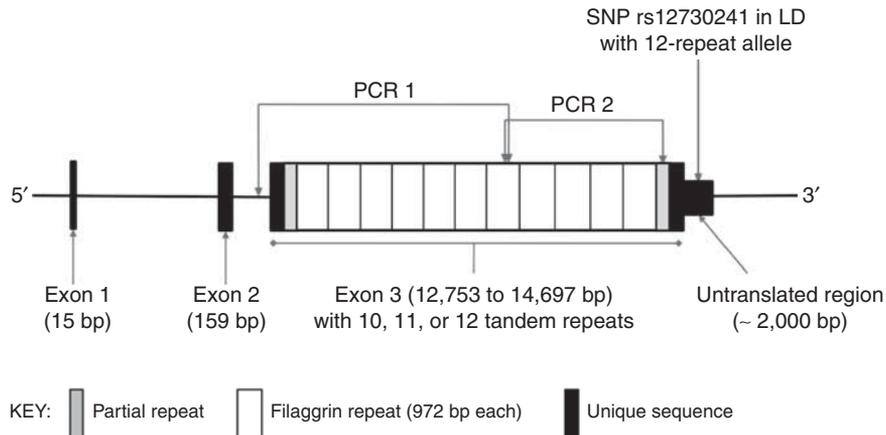
## DISCUSSION

Assessment of the entire range of genetic and genomic variation (including SNPs, structural variation, noncoding variants, long-range *trans* effects, and so on) will be required in order to fully understand the genetic predisposition to

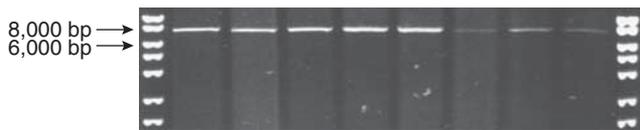
**Table 1. Demographic and clinical data relating to atopic dermatitis cases and unselected population controls**

	Irish pediatric atopic dermatitis cases		Irish population controls	
	Genotyped collection (n=876)	Filaggrin ( <i>FLG</i> ) mutation carriers excluded (n=528)	Genotyped collection (n=928)	<i>FLG</i> mutation carriers excluded (n=824)
Male sex, number (%)	545 (62.2)	332 (62.9)	278 (30.0)	251 (30.5)
Mean age, years (SD)	3.3 (3.5)	3.0 (3.3)	36.2 (12.4)	36.2 (12.5)
Nottingham eczema severity score, mean (SD)	11.2 (2.6)	11.0 (2.5)	Not known	Not known

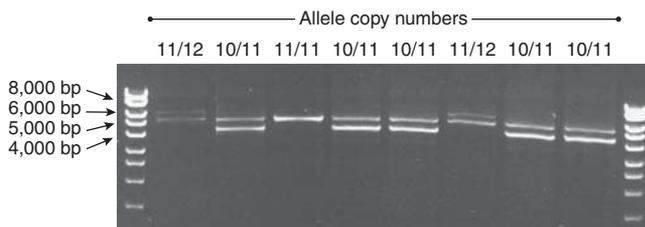
The diagnosis of atopic dermatitis was made by experienced dermatologists according to the United Kingdom diagnostic criteria (Williams *et al.*, 1994) with disease severity scored according to the Nottingham Eczema Severity Score, in which 1 to 8 denotes mild atopic dermatitis, 9 to 11 is moderate, and 12 to 15 defines severe disease (Emerson *et al.*, 2000). The control population represents consecutive successfully genotyped Trinity Biobank Control samples, derived from Irish adult blood donors.



**Figure 1.** *Filaggrin (FLG)* structure and genotyping strategies to identify copy number variants. LD, linkage disequilibrium; SNP, single-nucleotide polymorphism. Diagram adapted from Smith *et al.* (2006).



**Figure 2.** Visualization of PCR 1 products showing no evidence of copy number variation (CNV) within the 5' portion of *filaggrin (FLG)* exon 3. PCR reactions separated on 0.8% agarose gel. Bioline hyperladder I marker (Bioline, London, UK).



**Figure 3.** Visualization of PCR 2 products showing product sizes representing 10-, 11-, and 12-repeat alleles from the 3' portion of *filaggrin (FLG)* exon 3. PCR reactions separated on 0.8% agarose gel. Bioline hyperladder I marker (Bioline).

complex traits such as atopic dermatitis. Loss-of-function mutations in *FLG* represent the strongest and most significant genetic factor in the etiology of atopic dermatitis characterized to date (Baurecht *et al.*, 2007; Rodríguez *et al.*, 2009; van den Oord and Sheikh, 2009), with an odds ratio estimated by meta-analysis to be 3.12 (Baurecht *et al.*, 2007). This study contributes to the understanding of haploinsufficiency in atopic dermatitis by showing that CNV within *FLG* also significantly affects atopic dermatitis risk. The CNV effect is independent of *FLG*-null mutations but the overall effect size is smaller, as the comparable odds ratio is  $1/0.88 = 1.14$ . However, logistic regression analysis shows a dose-dependent effect, such that the risk reduces incrementally as the total number of filaggrin monomers encoded on a pair of alleles increases. Therefore, an individual with 20

**Table 2.** *FLG* genotypes, null mutations, and copy number variants in 876 Irish atopic dermatitis cases and 928 Irish population controls

Allelic variant	Whole collection		<i>FLG</i> mutation carriers excluded	
	Atopic dermatitis cases, n (%)	Population controls, n (%)	Atopic dermatitis cases, n (%)	Population controls, n (%)
10 Repeats	696 (39.7)	629 (33.9)	410 (38.8)	550 (33.4)
11 Repeats	814 (46.5)	956 (51.5)	497 (47.1)	856 (51.9)
12 Repeats	242 (13.8)	271 (14.6)	149 (14.1)	242 (14.7)
<i>Total copy number</i>				
20	NA	NA	85 (16.1)	87 (10.6)
21	NA	NA	186 (35.2)	300 (36.4)
22	NA	NA	172 (32.6)	294 (35.7)
23	NA	NA	75 (14.2)	120 (14.6)
24	NA	NA	10 (1.9)	23 (2.8)
$\chi^2$ test	$P=0.043$			
Logistic regression	$P=0.025$			

Abbreviations: *FLG*, *filaggrin*; NA, not applicable. Total copy number denotes the sum of the number of repeats on each allele in an individual showing *FLG* wild-type genotype at the four prevalent null mutations. The *FLG* and copy number allelic variants each did not show a significant deviation from Hardy-Weinberg equilibrium ( $P>0.05$ ) in the control population.

*filaggrin* repeats has an odds ratio of  $1/0.60 = 1.67$  compared with an individual having a total of 24 repeats. As all of the size variant alleles are common within the population (33.9%, 10 repeats; 51.5%, 11 repeats; and 14.6%, 12 repeats), this form of genetic variation represents a significant additional contribution to AD risk.

The polymorphic *FLG* alleles segregate in families according to Mendelian inheritance (Gan *et al.*, 1990) and therefore may contribute to the observed familial clustering of atopic disease (Brown and Reynolds, 2006). Filaggrin is a histidine-rich protein (Lynley and Dale, 1983) and *FLG*-null mutations have been shown to affect the quantitative expression of histidine and subsequent filaggrin breakdown products in the stratum corneum of atopic dermatitis patients and control individuals (Kezic *et al.*, 2008, 2011). This study has shown that CNV within *FLG* may have a significant functional effect on the amount of UCA in the stratum corneum, with a small but significant correlation accounting for ~18% of the variation in UCA concentration. UCA is a component of the “natural moisturizing factor” in the epidermis (Rawlings and Harding, 2004) and plays a role in the maintenance of stratum corneum hydration as well as being a putative UV photoprotector (Eckhart *et al.*, 2008). The quantitative data linking *FLG* CNV with stratum corneum UCA concentration therefore provide support for the observed genetic epidemiological association. The other filaggrin breakdown products, histidine and PCA, also showed a small positive correlation with filaggrin copy number; however, a larger study is required in order to confirm or refute these correlations.

Two dermatological disorders have so far been shown to exhibit CNV as a mechanism contributing to their pathogenesis: microdeletions on chromosome 17q24.2–q24.3 in congenital generalized hypertrichosis terminalis with or without gingival hyperplasia (Sun *et al.*, 2009); and deletions of the late cornified cell envelope genes *LCE3B* and *LCE3C* in psoriasis (de Cid *et al.*, 2009; Huffmeier *et al.*, 2010). The extent to which CNV contributes to the “missing heritability” of common complex traits, including psoriasis and atopic dermatitis, remains to be defined. The three common *FLG* allele sizes (10, 11, and 12) have previously been reported in American and Irish populations (Gan *et al.*, 1990; Ginger *et al.*, 2005; Sandilands *et al.*, 2007b). Gan *et al.* (1990) screened over 300 individuals by Southern blot and showed that there were approximately equal proportions of 10-, 11-, and 12-repeat alleles. In our much larger study of the Irish population, roughly one-third of alleles have 10 repeats, half have 11 repeats, and the remaining 15% are 12-repeat alleles. An investigation of 113 individuals from Winnipeg Canada suggested that the 12-repeat allele may be a protective factor against self-perceived frequent dry skin (Ginger *et al.*, 2005), but this study was performed before knowledge of *FLG*-null mutations as a cause of common ichthyosis (Palmer *et al.*, 2006; Brown *et al.*, 2008), and hence the data are difficult to interpret. In retrospect, it appears likely that the observed association is attributable to a risk effect of the two common, ancestral *FLG*-null mutations (R501X and 2282del4) that usually do not occur on the 12-repeat allele (A Sandilands, unpublished data).

It is theoretically possible that the observed association of copy number with atopic dermatitis risk may be caused by linkage of the 10-repeat allele to an unrecognized *FLG* loss-of-function mutation. However, this is unlikely to account for the statistically significant association for three reasons: the

prevalence and distribution of *FLG*-null mutations in the Irish population is well established (Sandilands *et al.*, 2007b) and mutations other than the four that were assayed in this study are rare; furthermore, the statistical analysis of the continuous data shows an incremental reduction in the risk of atopic dermatitis with each additional unit of copy number; finally, the correlation of total copy number with stratum corneum UCA concentration provides additional evidence of a stepwise functional effect.

Carriers of the four *FLG*-null mutations that are prevalent in the population were excluded from statistical analysis in order to study the independent effect of CNV on atopic dermatitis risk. This approach is necessary because *FLG*-null mutations have such a profound effect on filaggrin expression, which is reduced to near zero in homozygotes or compound heterozygotes (Sandilands *et al.*, 2007b), that the more subtle effect of CNV could be obscured. It would be interesting to study the effect of CNV *in trans* with *FLG*-null mutations to investigate the possibility that larger copy number variants may provide some protection by partially ameliorating the effect of a *FLG*-null allele, or conversely the smaller copy number variants may serve to worsen the phenotype of a *FLG*-null heterozygote. However, even the total sample size of nearly 2,000 individuals in this study was not large enough to have sufficient power to permit such haplotype analysis. It would also be interesting to investigate the effect of *FLG* CNV on atopic dermatitis severity, with the hypothesis that a greater number of repeats protects against severe disease whereas a smaller number of repeats increases the risk of severe disease. However, this study, having predominantly moderate-severe atopic dermatitis cases, did not have sufficient statistical power to draw any firm conclusions regarding the effect of copy number on disease severity.

Previously published whole-genome association studies of CNVs focused on eight common complex traits of importance to human health (Craddock *et al.*, 2010); a separate study characterized CNVs of >443 bp in 450 individuals as reference genotypes from European, East Asian, and African individuals (Conrad *et al.*, 2010). The careful and rather laborious genotyping strategy utilized for this study has given detailed information on size variation within exon 3 of *FLG* in nearly 2,000 Irish individuals and the same three size variants—encoding 10, 11, and 12 copies of the filaggrin monomer—have been identified. Rarer size variants including duplications, a 3.7-kb deletion (Matsuzaki *et al.*, 2009), and two large inversions involving the *FLG* locus (Kidd *et al.*, 2008; Pang *et al.*, 2010) are recorded as single cases in the UCSC (University of California, Santa Cruz) database. Large inversions and other structural variants may not have been detected using our PCR-based genotyping assays and these cases may have contributed to the 5.3% of cases and 7.0% of controls in whom CNV genotyping failed. In contrast to human *FLG*, mouse strains show a wide range of CNV within the orthologous murine gene (*flg*), where characterized strains show between 10 and 20 filaggrin subunits (Zhang *et al.*, 2002). Although this study focused on CNV of 972 bp sequences within *FLG* exon 3, it is clearly possible that larger duplications of the whole *FLG* third exon or indeed the entire

FLG gene may occur. In light of the dose-dependent effect of intragenic CNV, these larger CNVs would be expected to have an even greater effect on filaggrin “dose” in the epidermis and warrant further investigation.

The discovery that *FLG*-null mutations are a strong and significant risk factor for atopic dermatitis has already led to a greater understanding of the importance of skin barrier function in atopic disease (Sandilands *et al.*, 2009). The fact that CNV within *FLG* significantly affects atopic dermatitis risk raises the attractive possibility that treatments aimed at increasing filaggrin expression may have a therapeutic effect on the prevention and/or treatment of atopic dermatitis. Encouragingly, our data show that even a 5–10% increase in filaggrin “dose” results in a significant reduction in the incidence of atopic dermatitis. Furthermore, as this analysis excluded *FLG*-null mutation carriers, the benefit is not likely to be restricted to the ~10% of individuals in the population who carry a *FLG*-null mutation (Smith *et al.*, 2006; Sandilands *et al.*, 2007a). It is perhaps surprising to observe that a 5–10% increase in filaggrin copy number is sufficient to result in a reduced risk of atopic dermatitis. There is likely to be considerable inter- and intra-individual variation in the rates of biosynthesis, processing, and degradation of filaggrin, but it may be speculated that the life-long effect of a genetically determined 5–10% increase in filaggrin “dose” may provide sufficient enhancement to skin barrier function to result in the observed risk reduction. Feasibility studies are currently underway in the United Kingdom and United States (Feasibility Study of Barrier Enhancement for Eczema Prevention (BEEP); The BEEP Study—Feasibility study of Barrier Enhancement for Eczema Prevention) in preparation for a prospective clinical trial to investigate whether intensive emollient application in early life, with the aim of improving skin barrier function, may reduce the risk of atopic dermatitis in infancy. In the future, it may also be possible to target pathways that control filaggrin expression with small molecules and thereby develop therapeutics to boost filaggrin expression in the epidermis. It is hoped that these combined investigations will demonstrate the translation of functional genomics to better care for atopic dermatitis patients in the foreseeable future.

## MATERIALS AND METHODS

### Atopic dermatitis cases

A collection of 925 unrelated Irish pediatric atopic dermatitis cases was recruited through a tertiary referral dermatology clinic at one center, Our Lady’s Children’s Hospital, Crumlin, Dublin. Atopic dermatitis was defined according to the UK diagnostic criteria

(Williams *et al.*, 1994) and all cases were confirmed on examination by an experienced dermatologist. In all, 51.2% of cases had severe disease; 32.6% had moderate, and 16.2% had mild atopic dermatitis, defined according to the Nottingham Eczema Severity Score (Emerson *et al.*, 2000). Demographic and clinical data are summarized in Table 1. Carriers of the four *FLG*-null mutations that are prevalent in the Irish population (R501X, 2282del4, S3247X, and R2447X) were excluded from subsequent statistical analysis in order to study the independent effect of CNV on atopic dermatitis risk. The genetic architecture of *FLG* loss-of-function mutations has been extensively characterized in the Irish population (Sandilands *et al.*, 2007b; Chen *et al.*, 2011), and together these four mutations account for >95% of *FLG* mutations; the remaining <5% are composed of rare and family-specific mutations.

### Control population

A total of 998 control DNA samples were obtained from the Trinity College Dublin Biobank, representing an unselected Irish population, as described previously (O’Donovan *et al.*, 2008). Demographic data relating to the control population are summarized in Table 1. As with the case collection, carriers of the four prevalent *FLG*-null mutations in the control population were excluded from subsequent statistical analysis.

### Ethical considerations

This study was conducted with the approval of the local research ethics committee (Our Lady’s Children’s Hospital, Crumlin). A parent or guardian for each case and all control subjects gave written informed consent. The collection of DNA samples and clinical data were conducted in accordance with the Good Clinical Practice and the Declaration of Helsinki Principles.

### Genotyping strategy

The four *FLG* loss-of-function mutations that are most prevalent in the Irish population (R501X, 2282del4, S3247X, and R2447X) were screened in the cases and controls as previously described (Sandilands *et al.*, 2007b; O’Regan *et al.*, 2010). Copy number variants were distinguished by their different sized products following PCR amplification. Amplification of the whole of exon 3 results in rather large products, ranging from 12,413 to 14,357 base pairs, within which size variants are difficult to distinguish reliably with the necessary resolution of 1,000 base pairs. Long-range PCR reactions were therefore designed to amplify the region of interest in two partially overlapping portions (Figure 1) as follows. In the first PCR reaction (PCR 1), the 5’ portion of *FLG* exon 3, from intron 2 (at a point near exon 3) to repeat 7, was amplified by long-range PCR using primers listed in Table 3. PCR reactions (50 µl) were performed using

**Table 3. PCR primers for amplification of *FLG* exon 3**

PCR reaction	Forward primer sequence	Reverse primer sequence	Product sizes (bp)
Amplification of beginning of exon 3 to repeat 7 (PCR 1)	5’-GCTGATAATGTGATTCTGTCTG-3’	5’-CTGGCTAAAACCTGGATCCCCA-3’	7,624
Amplification of repeat 7 to 3’ partial repeat (PCR 2)	5’-CCCAGGACAAGCAGGAACT-3’	5’-CTGCACTACCATAGCTGCC-3’	4,277, 5,249, and 6,224

Abbreviations: bp, base pairs; *FLG*, filaggrin.

the TaKaRa LA Taq (Takara Bio, Shiga, Japan) with 200 ng genomic DNA, 1 unit of TaKaRa LA Taq, 8  $\mu$ l dNTPs (2.5 mM each), 25 ng each of the forward and reverse primers, 5  $\mu$ l 10 $\times$  Buffer (TaKaRa), and 5  $\mu$ l MgCl (25 mM). Cycle conditions for PCR 1 were as follows: 94 °C initial denaturation for 1 minute, followed by 33 cycles of denaturation (94 °C for 30 seconds) and elongation (68 °C for 10 minutes), followed by a final elongation at 72 °C for 10 minutes. Electrophoresis on a 0.8% w/v agarose gel, at 120V for 3 hours, was used to identify product size. In a second PCR reaction (PCR 2), the 3' portion of *FLG* exon 3, encompassing repeats 7 to 10 plus the 3' partial repeat, was amplified by long-range PCR using the primers shown in Table 3. Reactions (20  $\mu$ l) were performed using the Expand High Fidelity System (Roche Applied Science, Mannheim, Germany) with  $\sim$ 100 ng genomic DNA, 1.5 units of Expand High Fidelity enzyme, 2  $\mu$ l dNTPs (2.5 mM each), 10 ng each of the forward and reverse primers, 2  $\mu$ l Buffer 2 (Roche Expand High Fidelity System), and 4% v/v DMSO. Cycle conditions for PCR 2 were as follows: 95 °C initial denaturation for 5 minutes, followed by 35 cycles of denaturation (94 °C for 30 seconds), annealing (63.5 °C for 30 seconds) and elongation (72 °C for 5 minutes 30 seconds), followed by a final elongation at 72 °C for 7 minutes. Products were separated as for PCR 1, on a 0.8% w/v agarose gel.

The largest product from PCR 2 is most at risk of suboptimal amplification, resulting in a weaker band when visualized on a gel under UV light. In order to minimize the risk of failing to detect the largest size allele, an additional SNP genotyping strategy was used. SNP rs12730241 is in the 3' untranslated region of *FLG* (near to the duplicated repeats) and was shown, in 89 out of 90 (98.9%) fully sequenced individuals, to be associated with the 12-repeat allele. Rs12730241 was typed using the available predesigned TaqMan allelic discrimination assays (Applied Biosystems, Foster City, CA) according to the manufacturer's recommended protocols. Therefore, 12-repeat alleles shown by long-range PCR were confirmed by genotyping for SNP rs12730241.

### Quantification of filaggrin breakdown products

Data relating to the most abundant filaggrin breakdown products—histidine, UCA, and PCA—in 31 *FLG* wild-type atopic dermatitis cases were available from a previous study (Kezic *et al.*, 2011). Stratum corneum samples removed from the nonlesional skin of pediatric atopic dermatitis cases by a tape-stripping technique had undergone HPLC to quantify histidine, UCA, and PCA as previously described (Kezic *et al.*, 2011).

### Statistical analysis

In individuals with a wild-type genotype at the four screened *FLG*-null mutations, total copy number was defined as the sum of the number of tandem repeats on each allele. Total copy number genotype frequencies were compared between the case and control groups using  $\chi^2$  test and logistic regression analysis, performed with the statistical analysis package Stata (StataCorp LP, College Station, TX). Simple linear regression was used to test for correlation between total copy number and *FLG* breakdown products using SPSS v 15.0.1 (IBM Corporation, Somers, NY).

### CONFLICT OF INTEREST

WHI McLean has filed patents on genetic testing and therapy development aimed at the filaggrin gene. The other authors state no conflict of interest.

### ACKNOWLEDGMENTS

This work was supported by a Wellcome Trust Intermediate Clinical Fellowship (ref 086398/Z/08/Z) awarded to SJB; filaggrin research in the McLean laboratory is supported by grants from the British Skin Foundation, National Eczema Society, Medical Research Council (ref G0700314), the Wellcome Trust (references 090066/B/09/Z and 092530/Z/10/Z), and donations from anonymous families affected by eczema in the Tayside Region of Scotland; SJB, SK, WHIM, and ADI are members of COST Action BM 0903 (Skin Barrier and Atopic Disease (SkinBAD)); HJC is supported by a Wellcome Trust Senior Fellowship (ref 087436); ADI has received funding from the National Children's Research Centre, Dublin, and the Wellcome Trust (references 090066/B/09/Z and 092530/Z/10/Z). We are very grateful to our patients who participated in this research and for the expert help of Mrs Nuala Aylward. We also thank Dr Grainne O'Regan for allowing us to utilize previously published data on filaggrin breakdown products quantified in tape-stripped stratum corneum.

### SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/jid>

### REFERENCES

- Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet* 12:363–76
- Baurecht H, Irvine AD, Novak N *et al.* (2007) Toward a major risk factor for atopic eczema: meta-analysis of filaggrin polymorphism data. *J Allergy Clin Immunol* 120:1406–12
- Brown S, Reynolds NJ (2006) Atopic and non-atopic eczema. *BMJ* 332: 584–8
- Brown SJ, Relton CL, Liao H *et al.* (2008) Filaggrin null mutations and childhood atopic eczema: a population-based case-control study. *J Allergy Clin Immunol* 121:940–6, e943
- Chen H, Common JE, Haines RL *et al.* (2011) Wide spectrum of filaggrin-null mutations in atopic dermatitis highlights differences between Singaporean Chinese and European populations. *Br J Dermatol* 165:106–14
- Conrad DF, Pinto D, Redon R *et al.* (2010) Origins and functional impact of copy number variation in the human genome. *Nature* 464:704–12
- Craddock N, Hurler ME, Cardin N *et al.* (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464:713–20
- de Cid R, Riveira-Munoz E, Zeeuwen PL *et al.* (2009) Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet* 41:211–5
- Eckhart L, Schmidt M, Mildner M *et al.* (2008) Histidase expression in human epidermal keratinocytes: regulation by differentiation status and all-trans retinoic acid. *J Dermatol Sci* 50:209–15
- Emerson R, Charman C, Williams H (2000) The Nottingham Eczema Severity Score: preliminary refinement of the Rajka and Langeland grading. *Br J Dermatol* 142:288–97
- Feasibility Study of Barrier Enhancement for Eczema Prevention (BEEP). <http://clinicaltrials.gov/ct2/show/NCT01142999>
- Gan SQ, McBride OW, Idler WW *et al.* (1990) Organization, structure, and polymorphisms of the human profilaggrin gene. *Biochemistry* 29: 9432–40
- Ginger RS, Blachford S, Rowland J *et al.* (2005) Filaggrin repeat number polymorphism is associated with a dry skin phenotype. *Arch Dermatol Res* 297:235–41
- Huffmeier U, Bergboer JG, Becker T *et al.* (2010) Replication of LCE3C-LCE3B CNV as a risk factor for psoriasis and analysis of interaction with other genetic risk factors. *J Invest Dermatol* 130:979–84
- Iafraite AJ, Feuk L, Rivera MN *et al.* (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36:949–51
- Kay J, Gawkrödger DJ, Mortimer MJ *et al.* (1994) The prevalence of childhood atopic eczema in a general population. *J Am Acad Dermatol* 30: 35–9

- Kezic S, Kemperman PM, Koster ES *et al.* (2008) Loss-of-function mutations in the filaggrin gene lead to reduced level of natural moisturizing factor in the stratum corneum. *J Invest Dermatol* 128:2117-9
- Kezic S, O'Regan GM, Yau N *et al.* (2011) Levels of filaggrin degradation products are influenced by both filaggrin genotype and atopic dermatitis severity. *Allergy* 66:934-40
- Kidd JM, Cooper GM, Donahue WF *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature* 453:56-64
- Lewis-Jones S (2006) Quality of life and childhood atopic dermatitis: the misery of living with childhood eczema. *Int J Clin Pract* 60:984-92
- Lynley AM, Dale BA (1983) The characterization of human epidermal filaggrin. A histidine-rich, keratin filament-aggregating protein. *Biochim Biophys Acta* 744:28-35
- Matsuzaki H, Wang PH, Hu J *et al.* (2009) High resolution discovery and confirmation of copy number variants in 90 Yoruba Nigerians. *Genome Biol* 10:R125
- Mischke D, Korge B, Marenholz I *et al.* (1996) Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex ("epidermal differentiation complex") on human chromosome 1q21. *J Invest Dermatol* 106:989-92
- O'Donovan M, Craddock N, Norton N *et al.* (2008) Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 40:1053-5
- O'Regan GM, Campbell LE, Cordell HJ *et al.* (2010) Chromosome 11q13.5 variant associated with childhood eczema: an effect supplementary to filaggrin mutations. *J Allergy Clin Immunol* 125:170-4, e171-172
- Palmer CN, Irvine AD, Terron-Kwiatkowski A *et al.* (2006) Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* 38:441-6
- Pang AW, MacDonald JR, Pinto D *et al.* (2010) Towards a comprehensive structural variation map of an individual human genome. *Genome Biol* 11:R52
- Rawlings AV, Harding CR (2004) Moisturization and skin barrier function. *Dermatol Ther* 17 Suppl 1:43-8
- Redon R, Ishikawa S, Fitch KR *et al.* (2006) Global variation in copy number in the human genome. *Nature* 444:444-54
- Rodríguez E, Baurecht H, Herberich E *et al.* (2009) Meta-analysis of filaggrin polymorphisms in eczema and asthma: robust risk factors in atopic disease. *J Allergy Clin Immunol* 123:1361-70. e1367
- Sandilands A, Smith FJ, Irvine AD *et al.* (2007a) Filaggrin's fuller figure: a glimpse into the genetic architecture of atopic dermatitis. *J Invest Dermatol* 127:1282-4
- Sandilands A, Sutherland C, Irvine A *et al.* (2009) Filaggrin in the frontline: role in skin barrier function and disease. *J Cell Sci* 122: 1285-94
- Sandilands A, Terron-Kwiatkowski A, Hull P *et al.* (2007b) Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema. *Nat Genet* 39: 650-4
- Sebat J, Lakshmi B, Troge J *et al.* (2004) Large-scale copy number polymorphism in the human genome. *Science* 305:525-8
- Shamssain M (2007) Trends in the prevalence and severity of asthma, rhinitis and atopic eczema in 6- to 7- and 13- to 14-yr-old children from the north-east of England. *Pediatr Allergy Immunol* 18:149-53
- Shaw TE, Currie GP, Koudelka CW *et al.* (2011) Eczema prevalence in the United States: data from the 2003 National Survey of Children's Health. *J Invest Dermatol* 131:67-73
- Smith F, Irvine A, Terron-Kwiatkowski A *et al.* (2006) Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat Genet* 38:337-42
- Sun M, Li N, Dong W *et al.* (2009) Copy-number mutations on chromosome 17q24.2-q24.3 in congenital generalized hypertrichosis terminalis with or without gingival hyperplasia. *Am J Hum Genet* 84: 807-13
- The BEEP Study—Feasibility Study of Barrier Enhancement for Eczema Prevention. (<http://www.beepstudy.org/>)
- van den Oord R, Sheikh A (2009) Filaggrin gene defects and risk of developing allergic sensitisation and allergic disorders: systematic review and meta-analysis. *BMJ* 339:b2433
- Williams HC (2005) Clinical practice. Atopic dermatitis. *N Engl J Med* 352:2314-24
- Williams HC, Burney PG, Pembroke AC *et al.* (1994) The UK Working Party's Diagnostic Criteria for Atopic Dermatitis. III. Independent hospital validation. *Br J Dermatol* 131:406-16
- Zhang D, Karunaratne S, Kessler M *et al.* (2002) Characterization of mouse profilaggrin: evidence for nuclear engulfment and translocation of the profilaggrin B-domain during epidermal differentiation. *J Invest Dermatol* 119:905-12



**This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>**