Accepted Manuscript

Exome sequencing and rare variant analysis reveals multiple filaggrin mutations in Bangladeshi atopic eczema families and additional risk genes

Manuela Pigors, John E.A. Common, Xuan Fei Colin C. Wong, Sajid Malik, Claire A. Scott, Niloofar Tabarra, Herty Liany, Jianjun Liu, Vachiranee Limviphuvadh, Sebastian Maurer-Stroh, Mark BY. Tang, Nicholas Lench, David J. Margolis, David A. van Heel, Charles A. Mein, Natalija Novak, Hansjörg Baurecht, Stephan Weidinger, W.H. Irwin McLean, Alan D. Irvine, Edel A. O'Toole, Michael A. Simpson, David P. Kelsell

PII: S0022-202X(18)31990-0

DOI: 10.1016/j.jid.2018.05.013

Reference: JID 1443

To appear in: *The Journal of Investigative Dermatology*

Received Date: 9 January 2018

Revised Date: 1 May 2018

Accepted Date: 8 May 2018

Please cite this article as: Pigors M, Common JEA, Wong XFCC, Malik S, Scott CA, Tabarra N, Liany H, Liu J, Limviphuvadh V, Maurer-Stroh S, Tang MB, Lench N, Margolis DJ, van Heel DA, Mein CA, Novak N, Baurecht H, Weidinger S, McLean WHI, Irvine AD, O'Toole EA, Simpson MA, Kelsell DP, Exome sequencing and rare variant analysis reveals multiple filaggrin mutations in Bangladeshi atopic eczema families and additional risk genes, *The Journal of Investigative Dermatology* (2018), doi: 10.1016/j.jid.2018.05.013.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Exome sequencing and rare variant analysis reveals multiple filaggrin mutations in Bangladeshi atopic eczema families and additional risk genes

Manuela Pigors[1], John E. A. Common[2], Xuan Fei Colin C. Wong[2], Sajid Malik[3], Claire A. Scott[1], Niloofar Tabarra[1], Herty Liany[4], Jianjun Liu[4], Vachiranee Limviphuvadh[5], Sebastian Maurer-Stroh[5], Mark BY Tang[6], Nicholas Lench[7], David J. Margolis[8], David A. van Heel[1], Charles A. Mein[9], Natalija Novak[10], Hansjörg Baurecht[11], Stephan Weidinger[11], W.H. Irwin McLean[12], Alan D. Irvine[13], Edel A. O'Toole[1], Michael A. Simpson[14,15], David P. Kelsell[1,15]

AFFILIATIONS:

[1]Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK

[2]Institute of Medical Biology, ASTAR, Singapore, Singapore

[3]Department of Animal Sciences, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan

[4]Genome Institute of Singapore, A*STAR, Singapore, Singapore

[5]Bioinformatics Institute, A*STAR, Singapore, Singapore

[6]National Skin Centre, Singapore, Singapore

[7]Congenica Limited, Genome Campus, Hinxton, UK

[8]Department of Biostatistics and Epidemiology and Dermatology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[9]The Genome Center, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London , UK

[10]Department of Dermatology and Allergy, University Hospital Bonn, Bonn, Germany

ACCEPTED MANUSCRIPT

[11]Department of Dermatology, Allergology and Venereology, University Hospital Schleswig-Holstein, Kiel, Germany

[12]Dermatology and Genetic Medicine, University of Dundee, Dundee, UK

[13]National Children's Research Centre, Our Lady's Children's Hospital Crumlin, Dublin, Ireland; Department of Paediatric Dermatology, Our Lady's Children's Hospital Crumlin, Dublin, Ireland; Department of Clinical Medicine, Trinity College Dublin, Dublin, Ireland.

[14]Division of Genetics and Molecular Medicine, King's College London, Guy's Hospital, Great Maze Pond, London, UK

[15]These authors contributed equally to this work and are joint senior authors

CORRESPONDING AUTHOR:

David P. Kelsell, PhD

Centre for Cell Biology & Cutaneous Research

Blizard Institute

Barts and the London School of Medicine and Dentistry

Queen Mary University of London

4 Newark Street

London E1 2AT

United Kingdom

d.p.kelsell@qmul.ac.uk

@davidkelsell

ACCEPTED MANUSCRIPT

Letter

Atopic eczema (AE) is a heterogeneous chronic inflammatory skin condition that affects approximately 15-20% of children worldwide (Nutten, 2015). The prevalence varies widely among different countries, between rural and urban areas within a single country and particular Asian skin types are more sensitive to urban environments (Krutmann et al., 2014, Ben-Gashir et al., 2004, Nutten, 2015, Odhiambo et al., 2009). In this study, we analyzed the genetic architecture of AE patients from the South Asian Bangladeshi community in East London (UK) using Whole Exome Sequencing (WES) combined with rare variant enrichment analysis. A total of 70 Bangladeshi Sylheti families each with at least two affected siblings presenting with severe AE (determined by National Institute for Health and Care Excellence (NICE) guidelines) were recruited via the pediatric dermatology clinic at the Royal London Hospital, London (UK). The majority of the affected individuals were born in the UK and also presented with very high IgE levels and other atopic phenotypes, such as food allergy, asthma and hay fever.

Genomic DNA was extracted from EDTA-peripheral blood samples from family members after written informed consent and in adherence with the Declaration of Helsinki Principles and approval from the East London and City Health Authority. WES was performed in 43 probands of 42 Bangladeshi families and an enrichment analysis was performed towards identifying possible AE associated rare coding gene variants. Exome capture and enrichment was performed using various versions of the Agilent capture platform. The subsequent DNA library was sequenced with 100 bp paired-end reads on the HiSeq 2000 platform (Illumina). Resulting sequence data were aligned to the hg19 human reference genome using the Novoalign alignment tool (Novocraft Technologies Sdn Bhd). Sequence variants were called with SAMtools and annotated with ANNOVAR (Wang et al., 2010). Variants were filtered for novelty by comparison to ExAC, in-house exomes, dbSNP137 and

1000 genomes. The association analysis is based on the hypothesis, that multiple rare variants can contribute to the genetic susceptibility to common diseases (Schork et al., 2009). We sought to establish the presence of an excess of rare dominant (<1%) and low frequency recessive (<5%) missense, frameshift, and nonsense variants between 43 probands and 232 ethnically matched unrelated control samples, association was assessed using a simple Fisher's exact test.

This analysis revealed that filaggrin *(FLG)* harbored the highest number of enriched dominant (Odds ratio = 12.1, $P < 0.0001$) and recessive loss-of-function, i.e. nonsense and frameshift mutations (Odds ratio = 43.4, $P < 0.0001$) (**Table 1**). *FLG* missense variants were not enriched in the Bangladeshi probands. In total, 13 loss-of-function *FLG* variants, of which five were unreported variants, were identified in 21 of the 43 Bangladeshi probands using WES (**Figure 1a, Supplementary Table S1**). Moreover, the presence of numerous *FLG* loss-of-function mutations among the Bangladeshi probands, prompted us to sequence further affected family members (n=42) and probands (n=34) from 28 additional Bangladeshi families using a targeted array-based resequencing assay for *FLG* (Wong et al., 2017). All *FLG* loss-of-function variants were validated using Sanger sequencing and tested for segregation in all available family members. This revealed five additional variants in 33 probands, of which three were previously unreported mutations: c.2767insT (p.S923Ffs*2), c.4630delA (p.T1545Qfs*163), and c.7055C>G (p.S2352*) (**Figure 1a, Supplementary Table S1**). Furthermore, these genetic data revealed intrafamilial heterogeneity with multiple *FLG* variants often segregating within the Bangladeshi AE families (**Figure 1b, Supplementary Table S1**).

In addition, this rare variant association testing led to the identification of a potential burden of risk variants in as yet unreported AE associated gene loci, including *ADCY10*, *CUX2*, *MAST2*, *MCM10*, *MTF1*, *ORM2*, *PANX3*, *PHLDB1*, *SCAND3*, *TCHHL1* (**Table 1,**

Supplementary Table S2). The rare sequence variation of top 10 candidate genes in exome data from other AE cohorts with European, African-American, and Singapore Chinese ancestry was also analysed using allele frequencies of population-specific control exome data from the ExAC genome browser (accessed September 13-17, 2017; http://exac.broadinstitute.org/) (**Figure 1c, Supplementary Tables S3 and S4**). Allele frequencies were analyzed descriptively.

The comparison of the rare risk variants in AE compared to the ExAC data showed that the Bangladeshi AE-associated risk loci were, in part, also identified in the additional populations, with some AE risk variants shared between populations, e.g. *ADCY10, CUX2, MAST2, ORM2*, while others maybe ethnicity-specific, e.g. *MTF1, PANX3, TCHHL1* (**Figure 1c, Supplementary Tables S3 and S4**). Preliminary data showed that *MTF1*, *ORM2* and *TCHHL1* also appear to be replicated in Irish AE exomes (**Supplementary Tables S3 and S4**).

This rare variant enrichment strategy demonstrated that *FLG* represents the major AE risk gene in the Bangladeshi population with approximately 50% of AE affected individuals carrying one or more *FLG* loss-of-function variant. In contrast to *FLG*, whose role in the epidermis has been characterized in much detail, the precise function of the other AE gene variants is, at present, not defined (**Supplementary Table S2)**. We postulate that rare coding variants particularly *MTF1 (*a metal-regulatory transcription factor that regulates expression of, for example, cadmium and zinc) (Gunther et al., 2012, Kimura and Kambe, 2016), and *ORM2 (*a regulator of sphingolipid homeostasis) (Gururaj et al., 2013) could play a significant role in AE disease expression, regulation of immune responses, and/or environmental sensing.

This study has revealed insights into the genetic landscape of AE and adds to our multiethnic understanding of AE. Rare putative loss of function alleles in *FLG* was identified as a major component of disease susceptibility in the Bangladeshi AE cohort, however, the

burden analysis also identified a series of putative AE risk loci. Whilst the scope of replication was limited, additional evidence for the role of some of these genes in AE was also observed in three cohorts of individuals with AE from distinct ethnic ancestries.

CONFLICT OF INTEREST:

NL: Was an employee of Oxagen Limited. This work was funded in part with support from Oxagen Limited, Milton Park, Abingdon, Oxfordshire, UK.

SW: Has received honoraria for invited lectures and advisory boards from Galderma, Novartis, Pfizer, Regeneron and Sanofi, and has received research funding from La Roche-Posay, Novartis, Pfizer, and Sanofi.

All others: None.

ACKNOWLEDGEMENTS

We would like to thank the families for their participation in this study. We acknowledge the contributions of Jean Robinson and Dr David Paige for their assistance in identifying the families. M.P was supported by a Fellowship from the German Research Foundation (DFG). This work received infrastructure support through the DFG Cluster of Excellence "Inflammation at Interfaces" (grants EXC306 and EXC306/2), and was supported by grants (WE2678/6-1, WE2678/6-2, WE2678/9) from the DFG and the e:Med sysINFLAME grant no. 01ZX1306A from the German Federal Ministry of Education and Research (BMBF). J.E.A.C. and X.F.C.C.W. are funded by A*STAR SPF funding for translational skin research and genetic orphan diseases.

ACCEPTED MANUSCRIPT

REFERENCES

Ben-Gashir MA, Seed PT, Hay RJ. Predictors of atopic dermatitis severity over time. J Am Acad Dermatol 2004;50(3):349-56.

Gunther V, Lindert U, Schaffner W. The taste of heavy metals: gene regulation by MTF-1. Biochim Biophys Acta 2012;1823(9):1416-25.

Gururaj C, Federman RS, Chang A. Orm proteins integrate multiple signals to maintain sphingolipid homeostasis. J Biol Chem 2013;288(28):20453-63.

Kimura T, Kambe T. The Functions of Metallothionein and ZIP and ZnT Transporters: An Overview and Perspective. Int J Mol Sci 2016;17(3):336.

Krutmann J, Liu W, Li L, Pan X, Crawford M, Sore G, et al. Pollution and skin: from epidemiological and mechanistic studies to clinical implications. J Dermatol Sci 2014;76(3):163-8.

Margolis DJ, Gupta J, Apter AJ, Hoffstad O, Papadopoulos M, Rebbeck TR, et al. Exome sequencing of filaggrin and related genes in African-American children with atopic dermatitis. J Invest Dermatol 2014;134(8):2272-4.

Nutten S. Atopic dermatitis: global epidemiology and risk factors. Annals of nutrition & metabolism 2015;66 Suppl 1:8-16.

Odhiambo JA, Williams HC, Clayton TO, Robertson CF, Asher MI, Group IPTS. Global variations in prevalence of eczema symptoms in children from ISAAC Phase Three. J Allergy Clin Immunol 2009;124(6):1251-8 e23.

Sandilands A, Terron-Kwiatkowski A, Hull PR, O'Regan GM, Clayton TH, Watson RM, et al. Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema. Nat Genet 2007;39(5):650-4.

Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 2009;19(3):212-9.

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

Wong XFCC, Denil SLIJ, Foo JN, Chen H, Tay ASL, Haines RL, et al. Array-based sequencing of filaggrin gene for comprehensive detection of disease-associated variants. bioRxiv 2017.

ACCEPTED MANUSCRIPT

Figure 1: Filaggrin (*FLG*) risk variants identified in the Bangladeshi atopic eczema (AE) cohort and putative risk genes.

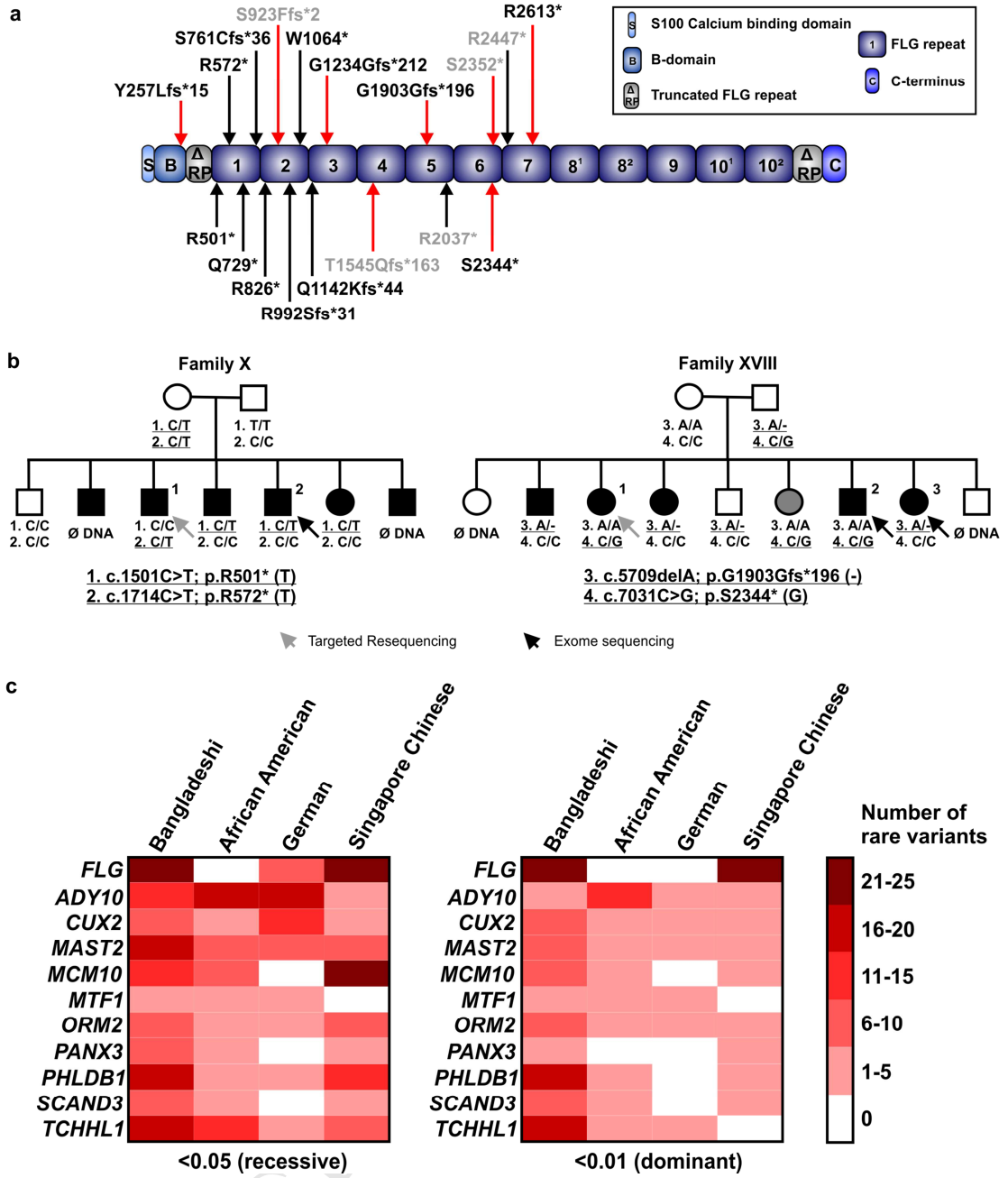
(a) Schematic representation of *FLG* showing all loss-of-function variants identified in this study. (Red arrows indicate previously unreported variants, variants shown black and gray were identified using whole exome sequencing or targeted re-sequencing, respectively) **(b)** Exome and targeted resequencing demonstrated an intrafamilial heterogeneity of *FLG* risk variants in the Bangladeshi AE cohort, segregating, in part, within families. Of note, the parents and the oldest sibling (all unaffected) of both families were born in Bangladesh. All other siblings were born in the UK. (Black symbols indicate affected individuals, gray circle indicates a previously affected daughter, white symbols represent unaffected individuals) **(c)** Heat Maps showing the number of rare variants identified in the different AE cohorts from this study. Data are based on allele frequencies from ethnically-matching subpopulations obtained from the genomic database ExAC. Recessive and dominant variants were considered rare if they had an allele frequency <0.05 and <0.01, respectively. Filaggrin (*FLG*) loss-of-function variants were excluded in the African-American cohort as described previously (Margolis et al., 2014).

ACCEPTED MANUSCRIPT

GENE/ GENE NAME	MUTATION TYPE	CHROMO-SOME	ODDS RATIO	*P*-VALUE
FLG/ Filaggrin	recessive frameshift or nonsense	**1** (152276443-152286591)	**43.4**	**< 0.0001**
	dominant frameshift or nonsense		**12.1**	
***SCAND3* (ZBED9)/** SCAN Domain Containing Protein 3 (Zinc Finger BED-Type Containing 9)	dominant missense	**6** (28542927-28554107)	**30.8**	**0.0002**
TCHHL1/ Trichohyalin Like 1	recessive missense	**1** (152057583-152060564)	**25.0**	**0.0008**
ADCY10/ Adenylate Cyclase 10	dominant frameshift or nonsense	**1** (167778964-167871251)	**19.5**	**0.0041**
MTF1/ Metal Regulatory Transcription Factor 1	dominant missense	**1** (38280953-38323065)	**7.6**	**0.0030**
MCM10/ Minichromosome Maintenance 10 Replication Initiation Factor	dominant missense	**10** (13213067-13251266)	**6.2**	**0.0004**
ORM2/ Orosomucoid 2	dominant missense	**9** (117092297-117094168)	**6.1**	**0.0015**
CUX2/ Cut Like Homeobox 2	dominant missense	**12** (111729278-111786109)	**5.3**	**0.0026**
MAST2/ Microtubule Associated Serine/Threonine Kinase 2	dominant missense	**1** (46463419-46501578)	**4.9**	**0.0021**
PHLDB1/ Pleckstrin Homology Like Domain Family B Member 1	recessive missense	**11** (118485376-118520801)	**4.5**	**0.0082**
PANX3/ Pannexin 3	dominant missense	**11** (124487339-124489689)	**infinity**	**0.0002**

Table 1: Exome sequencing and subsequent enrichment of rare variants led to the identification of risk genes linked to atopic eczema

ACCEPTED MANUSCRIPT


a
S100 Calcium binding domain
B-domain
Truncated FLG repeat
FLG repeat
C-terminus
S923Ffs*2
S761Cfs*36
W1064*
R572*
G1234Gfs*212
Y257Lfs*15
G1903Gfs*196
R2447*
S2352*
R2613*
R501*
Q729*
R826*
Q1142Kfs*44
R992Sfs*31
T1545Qfs*163
R2037*
S2344*
b
Family X
1. C/T
2. C/T
1. T/T
2. C/C
1. C/C
2. C/C
Ø DNA
1. C/C
2. C/T
1. C/T
2. C/C
1. C/T
2. C/C
1. C/T
2. C/C
Ø DNA
1. c.1501C>T; p.R501* (T)
2. c.1714C>T; p.R572* (T)
Family XVIII
3. A/A
4. C/C
3. A/-
4. C/G
Ø DNA
3. A/-
4. C/C
3. A/A
4. C/G
3. A/-
4. C/C
3. A/-
4. C/C
3. A/A
4. C/G
3. A/A
4. C/G
3. A/-
4. C/C
Ø DNA
3. c.5709delA; p.G1903Gfs*196 (-)
4. c.7031C>G; p.S2344* (G)
Targeted Resequencing
Exome sequencing
c
Bangladeshi
African American
German
Singapore Chinese
FLG
ADY10
CUX2
MAST2
MCM10
MTF1
ORM2
PANX3
PHLDB1
SCAND3
TCHHL1
<0.05 (recessive)
<0.01 (dominant)
Number of rare variants
21-25
16-20
11-15
6-10
1-5
0