

Journal Pre-proof

Improving the diagnostic yield for filaggrin; concealed mutations in the Dutch population

Frank S. van Leersum, Ivo F. Nagtzaam, Carolien N. van Oosterhoud, Stijn A.I. Ghesquiere, Roel R.H.F.J. Brandts, Antoni Gostyrński, Peter M. Steijlen, Michel van Geel

PII: S0091-6749(20)30122-6

DOI: <https://doi.org/10.1016/j.jaci.2020.01.033>

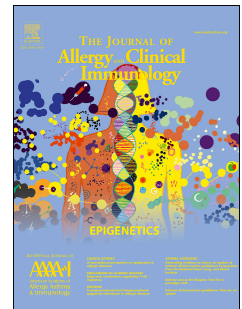
Reference: YMAI 14400

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 30 September 2019

Revised Date: 21 January 2020

Accepted Date: 24 January 2020



Please cite this article as: van Leersum FS, Nagtzaam IF, van Oosterhoud CN, Ghesquiere SAI, Brandts RRHFJ, Gostyrński A, Steijlen PM, van Geel M, Improving the diagnostic yield for filaggrin; concealed mutations in the Dutch population, *Journal of Allergy and Clinical Immunology* (2020), doi: <https://doi.org/10.1016/j.jaci.2020.01.033>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. 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.

© 2020 Published by Elsevier Inc. on behalf of the American Academy of Allergy, Asthma & Immunology.

Improving the diagnostic yield for filaggrin; concealed mutations in the Dutch population

Frank S. van Leersum^{1,2}, Ivo F. Nagtzaam^{1,2}, Carolien N. van Oosterhoud³, Stijn A.I. Ghesquiere³, Roel R.H.F.J. Brandts³, Antoni Gostyński^{1,2}, Peter M. Steijlen^{1,2}, Michel van Geel^{1,2,3}

¹ Department of Dermatology, Maastricht University Medical Center, Maastricht, the Netherlands

² GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, the Netherlands

³ Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, the Netherlands

Authors:

Frank S. van Leersum, MD, frank.van.leersum@mumc.nl

Ivo F. Nagtzaam, MD, i.nagtzaam@mumc.nl

Carolien N. van Oosterhoud, MSc, carolien.van.oosterhoud@mumc.nl

Stijn A.I. Ghesquiere, PhD, s.ghesquiere@mumc.nl

Roel R.H.F.J. Brandts, BSc, roel.brandts@mumc.nl

Antoni Gostyński, MD, PhD, antoni.gostynski@mumc.nl

Peter M. Steijlen, MD, PhD, peter.steijlen@mumc.nl

Michel van Geel, PhD, m.van.geel@mumc.nl

Keywords: filaggrin, diagnostics, small-molecule molecular inversion probes, next generation sequencing

Corresponding author:

F.S. van Leersum, MD

Maastricht University Medical Center

Department of Dermatology

P. Debyelaan 25

6229 HX Maastricht (The Netherlands)

Email: frank.van.leersum@mumc.nl

Phone: 0031 433875292

Fax: 0031 433877293

Conflicts of interest: none of the authors have competing interests to declare

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Word count: 1018

40 **Capsule summary (37 words)**

41 This study emphasizes screening of the entire *FLG* gene beyond population-specific
42 mutations. Screening was simplified by implementing single-molecule molecular-inversion-
43 probes and NextGeneration Sequencing. The diagnostic yield of *FLG* mutations increased
44 identifying novel mutations and enhancing further research opportunities.

To the editor, filaggrin (filament aggregating protein) has an important role in the terminal differentiation of the epidermis. The keratohyalin granules in the stratum granulosum contain profilaggrin which consists of 10-12 almost identical filaggrin repeats. In the uppermost layer of the granular cells - where transition to the corneocytes of the stratum corneum takes place – profilaggrin is processed into its active component, filaggrin. Filaggrin fragments are cross-linked to epidermal keratins by transglutaminase in order to form the cornified cell envelope. Remnants of filaggrin are further processed into natural moisturizing factors (NMF) (1).

Filaggrin (*FLG*) mutations show a semi-dominant inheritance: homozygous or compound heterozygous loss-of-function mutations are causally linked to ichthyosis vulgaris (2). This in contrast to heterozygous carriers who tend to have a milder phenotype or have a predisposition to develop atopic dermatitis, allergic rhinitis and asthma (3).

FLG is a remarkable gene within the epidermal differentiation complex on chromosome 1q21. The third and last exon, more than 12.7kb in size and encoding almost the entire filaggrin protein - is one of the largest exons known in the human genome (1). These filaggrin repeat units have a highly homologous and repetitive character that has hampered molecular diagnostics with conventional Sanger sequencing. Different approaches have been described in literature including Next Generation Sequencing (4, 5). By implementation of single molecule molecular inversion probes (smMIPS) and barcoded next generation sequencing (NGS) in a single tube, the entire coding region of *FLG* can be easily screened for rare family-specific mutations or mutations in ethnic groups not previously studied. Pooling barcoded samples will make this method very cost effective and superior to other established *FLG* sequencing techniques (see Table E1 in this article's Online Repository at www.jacionline.org).

MIPs are single-stranded DNA molecules that are designed to complement a target DNA. The probe contains an extension and a ligation arm flanking this target. The gap created over

the target region can be approximately 100-200 bases. Following hybridization of MIPs onto the target, gap-filling and ligation results in circularized DNA molecules containing the sequence of the target. Exonuclease removes the linear DNA and non-hybridized linear probes. Circularized MIPs are amplified by PCR and sequenced with NGS. The smMIPs were designed in accordance with the instructions published elsewhere (6). For input in the smMIP *FLG* analysis, 100 ng of genomic DNA was used (6). Since exclusion of NGS-duplicates-reads is possible with the single molecule (sm)-tag (5-N for *FLG* smMIPs), mutation percentages can be related and assigned to polymorphic duplicated filaggrin-repeat-unit 8 and 10 if a long-distance specific PCR-fragment sizing experiment is included. Sequencing was performed on a NextSeq500 instrument ((Illumina, San Diego, CA), calling of relevant variants was done with a commercial analysis software tool (Sequence Pilot version 4.4.0 Build 508; JSI medical systems, Ettenheim, Germany) and sequence reads displayed with Integrative Genomics Viewer V2.3.60.

A cohort of 202 patients residing in the Netherlands (mostly Dutch ethnicity) with a clinical manifestation ranging from atopic dermatitis to ichthyosis vulgaris was previously screened for five population-specific *FLG* mutations by Sanger sequencing. A *FLG* mutation was detected in 35% of the cohort. Retrospectively the entire *FLG* was analyzed again with the smMIP-NGS strategy. Mutations predicted to result in premature protein termination were scrutinized such that only nonsense and frameshift (small insertions and deletions) mutations were identified and selected. In several (17/202; 8%) of the screened patients, additional novel truncating mutations were identified (figure 1 and 2). These mutations were either new findings elucidating their previously unexplained phenotype, or a second previously unidentified mutation, confirming the clinical marked ichthyosis. Establishing a confirmed molecular diagnosis will exclude the often clinically challenging to distinguish X-linked or autosomal recessive congenital ichthyosis (see percentages in figure 1) as well as confirm

FLG deficiency as a potential disease modifier for severity in the other types of ichthyosis. Therefore, both patients and clinicians would benefit from a confirmed diagnosis. To date, approximately 113 unique mutations have been reported. Within this cohort and aside from 10 previously reported mutations, 17 additional new mutations have been identified. All were confirmed with Sanger sequencing consequently also verifying which of the duplicated repeat units 8 and 10 were involved. Mutations such as c.1501C>T (p.Arg501*) and c.2282_2285del comprise about 80% of the mutations found in the European population (7). Although this percentage is comparable to the Dutch cohort (69%), the distribution is significantly skewed towards the c.2282_2285del mutation (45%, figure 1B). Hence, one should consider mutation distribution variation in every European country. Currently, the most frequent four to five Western European mutations are tested with conventional Sanger sequencing or TaqMan PCR assays. However, a clear distinction can be made between population-specific (Asian and European variants) and family-specific variants (8). In order to detect all relevant variants, the focus of detection should be diverted from known population specific mutations to the entire coding region of the gene. By sequencing the entire *FLG* gene we were able to identify 23% additional mutations although this was only 4% in the Irish (figure 1B) population (8). Beyond mutations found in the large exon 3, mutations in exon 2, c.63dup and c.94G>T (p.(Glu32*)) coding for the S100 calcium binding domain, were for the first time identified (figure 2). These results emphasize that truncating mutations may be detected anywhere in the gene by using the smMIP-NGS strategy. In addition, more than two mutations could be present in the gene (figure 2B) thus underscoring that variants may be present on the same allele (in cis). Therefore, one should consider different familial inheritance patterns for counseling.

In conclusion, this and other studies (4, 5) emphasize the need to screen the entire *FLG* gene for mutations. We show a novel utility for smMIPs to establish this. Being able to screen the

entire gene improves the diagnostic yield in atopic dermatitis and ichthyosis vulgaris (in this study 8%) and makes it possible to identify novel findings or hidden variants in the homologous repeated units of the gene. Herein, the smMIP-NGS method has proven to be cost-effective and reliable. Besides its value in diagnostics, this strategy can also be applied as a tool to facilitate patient stratification in large cohort studies.

Acknowledgements

We thank Renske Janssen for her excellent technical assistance and all physicians contributing patients to this Dutch cohort.

129 Frank S. van Leersum, MD, Department of Dermatology and GROW School for Oncology
130 and Developmental Biology, Maastricht University Medical Center, Maastricht, the
131 Netherlands

132 Ivo, F. Nagtzaam, MD, Department of Dermatology and GROW School for Oncology and
133 Developmental Biology, Maastricht University Medical Center, Maastricht, the Netherlands

134 Carolien N. van Oosterhoud, MSc, Department of Clinical Genetics, Maastricht University
135 Medical Center, Maastricht, the Netherlands

136 Stijn A.I. Ghesquiere, PhD, Department of Clinical Genetics, Maastricht University Medical
137 Center, Maastricht, the Netherlands

138 Roel R.H.F.J. Brandts, BSc, Department of Clinical Genetics, Maastricht University Medical
139 Center, Maastricht, the Netherlands

140 Antoni Gostyński, MD, PhD, Department of Dermatology and GROW School for Oncology
141 and Developmental Biology, Maastricht University Medical Center, Maastricht, the
142 Netherlands

143 Peter M. Steijlen, MD, PhD, Department of Dermatology and GROW School for Oncology
144 and Developmental Biology, Maastricht University Medical Center, Maastricht, the
145 Netherlands

146 Michel van Geel, PhD, Departments of Dermatology, Clinical Genetics and GROW School
147 for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht,
148 the Netherlands

References

1. Irvine AD, McLean WH, Leung DY. Filaggrin mutations associated with skin and allergic diseases. *N Engl J Med*. 2011;365(14):1315-27.
2. Smith FJ, Irvine AD, Terron-Kwiatkowski A, Sandilands A, Campbell LE, Zhao Y, et al. Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat Genet*. 2006;38(3):337-42.
3. Palmer CN, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet*. 2006;38(4):441-6.
4. Wong X, Denil S, Foo JN, Chen H, Tay ASL, Haines RL, et al. Array-based sequencing of filaggrin gene for comprehensive detection of disease-associated variants. *J Allergy Clin Immunol*. 2018;141(2):814-6.
5. Pigors M, Common JEA, Wong X, Malik S, Scott CA, Tabarra N, et al. Exome sequencing and rare variant analysis reveals multiple filaggrin mutations in Bangladeshi atopic eczema families and additional risk genes. *J Invest Dermatol*. 2018.
6. Eijkelenboom A, Kamping EJ, Kastner-van Raaij AW, Hendriks-Cornelissen SJ, Neveling K, Kuiper RP, et al. Reliable Next-Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Tissue Using Single Molecule Tags. *J Mol Diagn*. 2016;18(6):851-63.
7. Gruber R, Janecke AR, Fauth C, Utermann G, Fritsch PO, Schmuth M. Filaggrin mutations p.R501X and c.2282del4 in ichthyosis vulgaris. *Eur J Hum Genet*. 2007;15(2):179-84.
8. Chen H, Common JE, Haines RL, Balakrishnan A, Brown SJ, Goh CS, et al. Wide spectrum of filaggrin-null mutations in atopic dermatitis highlights differences between Singaporean Chinese and European populations. *Br J Dermatol*. 2011;165(1):106-14.

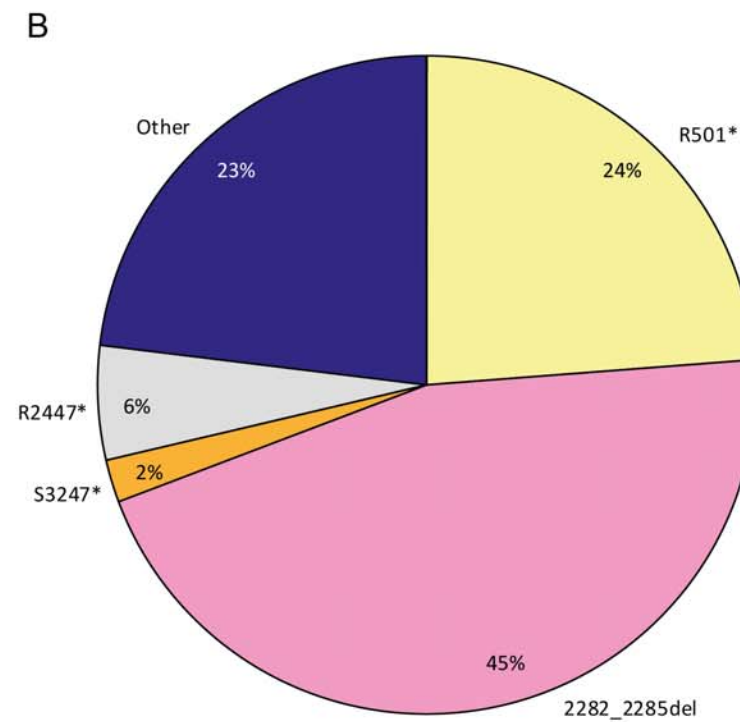
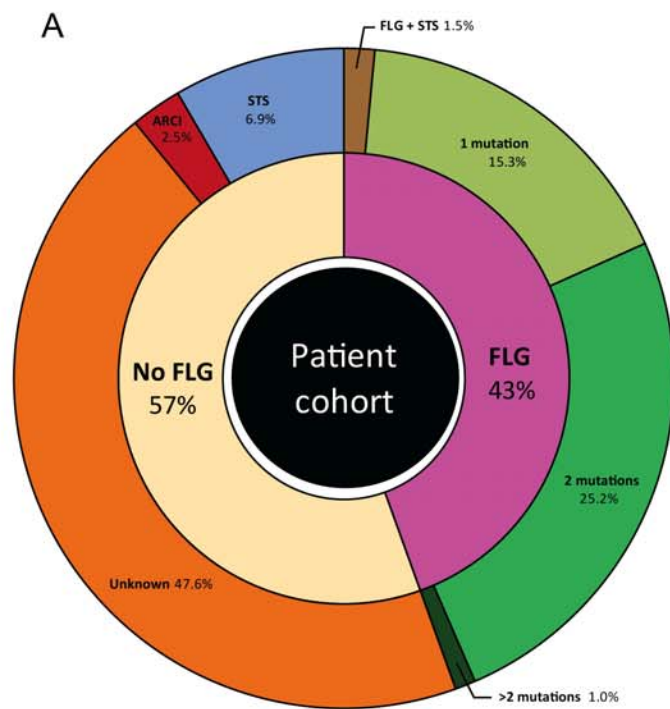
Legends

Figure 1: Overview of the *FLG* mutation results from the Dutch patient cohort. A;

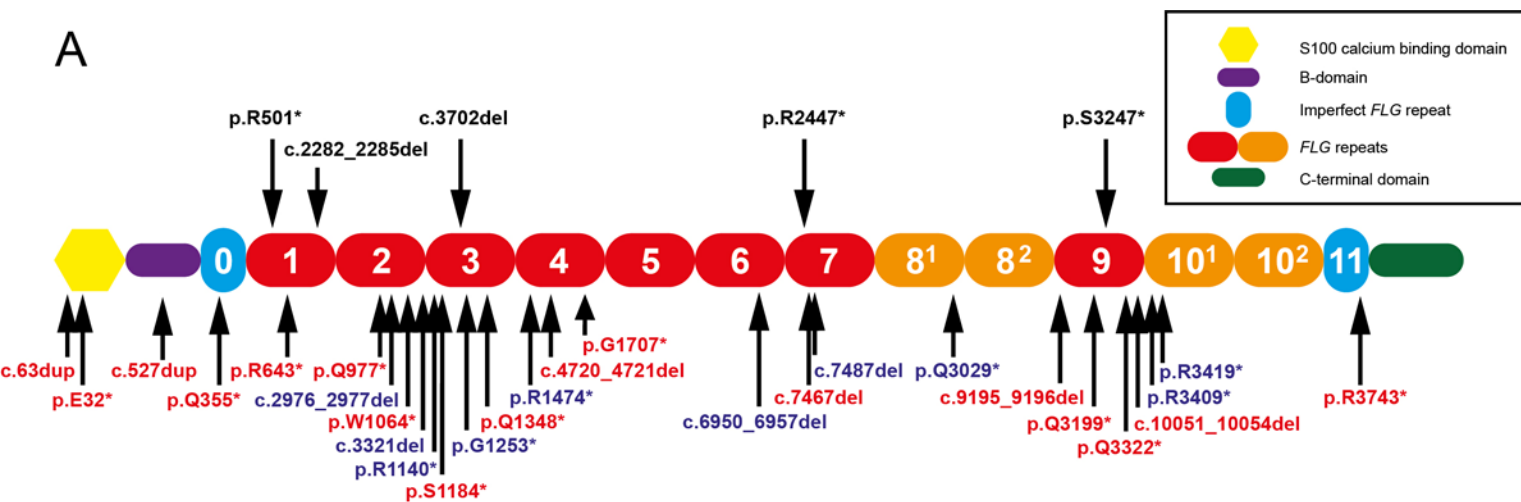
Proportion of the cohort carrying *FLG*-mutations. No mutation was detected in 57% of patients whereas the remaining 43% carried one or more mutations in *FLG* or had a coexisting *STS* mutation. *STS* mutations represent X-linked ichthyosis and ARCI autosomal recessive congenital ichthyosis patients. **B;** Spectrum of mutations in the Dutch population (N=143). The c.2282_2285del mutation is the most frequent followed by p.R501* representing 45% and 24 % of all mutations, respectively. 23 % of the mutations are different from the frequent West-European variants. Figure adapted from (8).

Figure 2: Overview of the mutations found, located outside the population specific *FLG* sites. A;

Schematic protein illustration of profilaggrin with the location of the mutations in each specific filaggrin repeat unit. Above, the 5 most frequent Western European mutations tested with conventional Sanger sequencing. Below, the additional mutations found with the smMIP-NGS strategy. The mutations in red are novel (n=17) and not described in literature before. Figure adapted from (4). **B;** Display of additional findings with the smMIP-NGS strategy in *FLG* mutation analysis. Mutations are shown according to their accompanied allele and colour coded for new (red, 17x), known (blue, 10x) and frequent West European (black) mutation. Of note, in patient 66 and 97 more than two *FLG* mutations were identified on the different alleles (3 and 4, resp.). †: Unknown if mutations are present on the different alleles (compound heterozygous), although the patients' clinical features do indeed suggest this.



A



B

patient ID	FLG allele 1	protein change	repeat	FLG allele 2	protein change	repeat
7	c.1063C>T	p.(Gln355*)	0			
16	c.9595C>T	p.(Gln3199*)	9			
17	c.2282_2285del	p.Ser761Cysfs*36	1	c.4420C>T	p.(Arg1474*)	4
18	c.7487del	p.(Thr2496Asnfs*104)	7			
37	c.2282_2285del	p.Ser761Cysfs*36	1	c.4042C>T	p.(Gln1348*)	3
43	c.6950_6957del	p.(Ser2317*)	6	c.7487del	p.(Thr2496Asnfs*104)	7
54	c.63dup	p.(Asp22Argfs*2)	S-100	c.3551C>A	p.(Ser1184*)	3
66	c.[1501C>T;4720_4721del]	p.[(R501*);(Ser1574Thrfs*21)]	1,4	c.1501C>T	p.(R501*)	1
68	c.94G>T	p.(Glu32*)	S-100	c.9740C>A	p.(Ser3247*)	9
75	c.2282_2285del	p.Ser761Cysfs*36	1	c.9964C>T	p.(Gln3322*)	9
78	c.2976_2977del	p.(Arg992Serfs*31)	2	c.3191G>A	p.(Trp1064*)	2
81	c.2282_2285del	p.Ser761Cysfs*36	1	c.10255C>T	p.(Arg3419*)	10 ¹
86	c.1501C>T	p.(Arg501*)	1	c.10225C>T	p.(Arg3409*)	10 ¹
93	c.11227C>T	p.(Arg3743*)	11			
97	c.[2976_2977del;9195_9196del]	p.[(Arg992Serfs*31);(Gln3065Hisfs*5)]	2,9	c.[9195_9196del;10225C>T]	p.[(Gln3065Hisfs*5);(Arg3409*)]	9, 10 ¹
99	c.3757G>T	p.(Gly1253*)	3			
108	c.3321del	p.(Gly1109Glufs*13)	2			
113	c.1501C>T	p.(Arg501*)	1	c.3757G>T	p.(Gly1253*)	3
117	c.527dup	p.(Asn176Lysfs*4)	0	c.2282_2285del	p.Ser761Cysfs*36	1
121	c.1927A>T	p.(Arg643*)	1			
126	c.2282_2285del	p.Ser761Cysfs*36	1	c.7467del	p.(Ser2490Leufs*110)	7
150	c.2282_2285del	p.Ser761Cysfs*36	1	c.9085C>T	p.(Gln3029*)	8 ¹
158	c.1501C>T	p.(Arg501*)	1	c.2929C>T	p.(Gln977*)	2
162	c.1501C>T	p.(Arg501*)	1	c.10051_10054del	p.(Asp3351Hisfs*39)	9
167	c.3418C>T	p.(Arg1140*)	3			
190	c.4420C>T	p.(Arg1474*)	4			
195	c.5119G>T	p.(Gly1707*)	4			

Table E1: Comparison of the smMIP-based NGS technology to the already established *FLG* sequencing tools.

Method	specificity	sensitivity	Labour intensity	cost
Sanger ^a	++	++	++++	++++
TaqMan PCR ^b	+++	+++	+++	+
WES ^c	++	++	+	+++
Array-based NGS ^d	+++	+++	++	++
Capture-based NGS ^e	++	++	+	++
smMIP-based NGS	+++	+++	+	+

+ to ++++ range illustrates increasing values (low to high) **a**: Sanger sequencing is highly dependent on primer-specific product amplification and PCR conditions are extremely delicate to prevent a-specific products (1) **b**: TaqMan PCR assay, considered as the gold standard (2), is only applicable to specific mutations and does not analyse the whole *FLG* gene (3). **c**: Whole Exome Sequencing (WES) does not select for only the *FLG* gene (more costly, using more NGS capacity) and sequencing coverage is usually <100x (4). Because alignment problems may map the homologous repeated sequence wrong, specificity and sensitivity is reduced (5) **d**: Labour and costs range higher, since no single-tube protocol is used (6). **e**: Coverage of the targeted gene was 88%, some variants could have been missed (7).

References 2, 6 and 7 also performed an in depth comparison to beadchip technology, Sanger sequencing, and WES and TaqMan analyses, respectively.

References:

- 1) 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. *Nature Genet* 2007; 39:650–4.
- 2) Margolis DJ, Apter AJ, Mitra N, Gupta J, Hoffstad O, Papadopoulos M, et al. Reliability and validity of genotyping filaggrin null mutations. *J Dermatol Sci* 2013; 70:67–8.
- 3) Margolis DJ, Apter AJ, Gupta J, Bowser M, Sharma H, Duffy E, et al. The persistence of atopic dermatitis and Filaggrin mutations in a US longitudinal cohort. *J Allergy Clin Immunol* 2012; 130:912–7.
- 4) Margolis DJ, Gupta J, Apter A, 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; 134:2272–4.
- 5) Margolis DJ, Mitra N, Wubbenhorst B, Nathanson KL. Filaggrin sequencing and bioinformatics tools. *Arch Dermatol Res.* 2019 Aug 1. doi: 10.1007/s00403-019-01956-3 [Epub ahead of print] PMID: 31372728.
- 6) 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. *J Allergy Clin Immunol* 2018;141(2):814-816.
- 7) Margolis DJ, Mitra N, Gochnauer H, Wubbenhorst B, D'Andrea K, Kraya A, et al. Uncommon filaggrin variants are associated with persistent atopic dermatitis in African Americans. *J Invest Dermatol* 2018; 138(7): 1501–1506.