Supplementary Information for

Structural properties of target binding by profilaggrin A and B domains and other S100 fused-type calcium-binding proteins

Alexander J. Hinbest[1], Sa Rang Kim[1], Sherif A. Eldirany[1], Ivan B. Lomakin[2], Joseph Watson[2], Minh Ho[1], and Christopher G. Bunick[1,2*]

[1] Department of Dermatology, Yale University, New Haven, Connecticut, 06520, USA.
[2] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, 06520, USA.

* To whom correspondence should be addressed:
E-mail: christopher.bunick@yale.edu (C.G.B.)

This file includes:

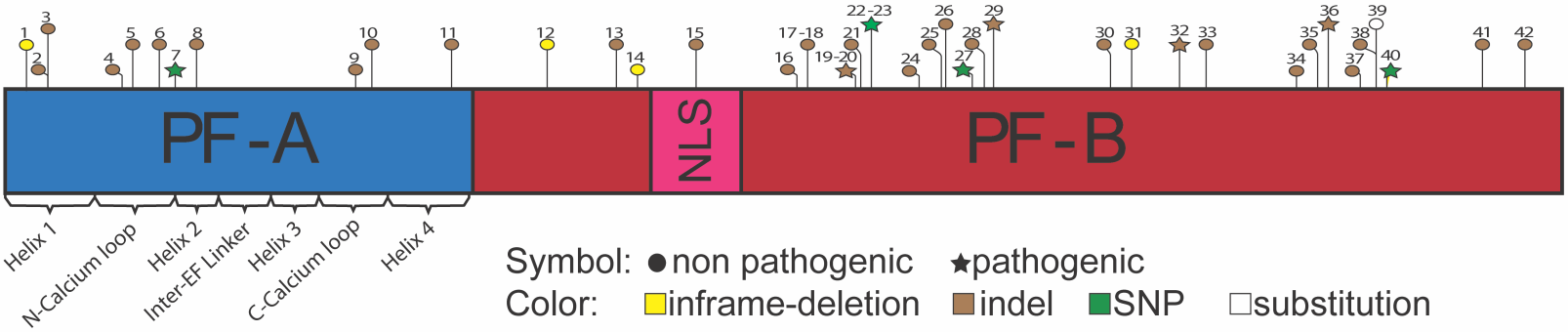
I. **Supplementary Figure S1**: Select mutations localized in the human profilaggrin A and B domains.
II. **Supplementary Figure S2**: Predicted antimicrobial peptide regions in S100 fused type proteins.
III. **Supplementary Figure S3**: Distinguishing characteristics of S100 fused type proteins and their clinical relevance.
IV. **Supplementary Figure S4**: PF-AB interactions with K1/K10 are reduced in the presence of elevated NaCl concentration.
V. **Supplementary Figure S5**: The B domain is predicted to be highly disordered with little rigid secondary structure.

a

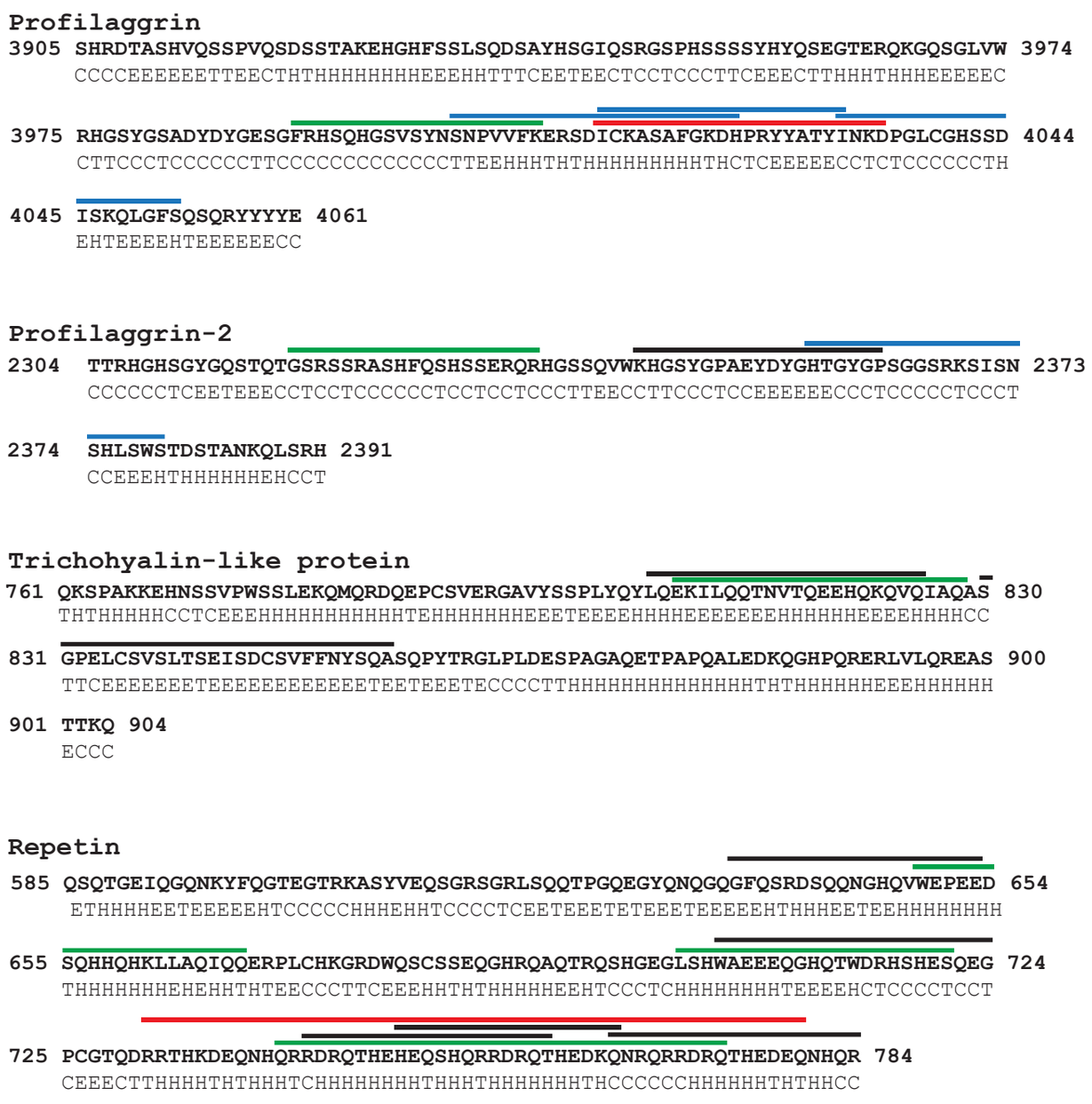
List of Profilaggrin Mutations

	Mutation ID		Transcript ID	ClinVar Accession	DNA mutation	Protein mutation	Protein domain	Mechanism	Condition
1	rs1399852786	Inframe-deletion	ENST00000368799.2		c.9_14del	Leu4_Leu5del	PFA N-terminus	Non synonymous variant	None
2	rs1395978363	indel	ENST00000368799.2		c.21_22insTTTT	p.Ile8PhefsTer7	PFA N-terminus	Extreme truncation 7 amino acids later (loss of PFAB and repeats)	None
3	rs1168885071	indel	ENST00000368799.2		c.22_23insC	p.Ile8ThrfsTer6	PFA N-terminus	Extreme truncation 6 amino acids later (loss of PFAB and repeats)	None
4	rs756605558	indel	ENST00000368799.2		c.63dupA	p.Asp22ArgfsTer2	PFA N-terminus	Extreme truncation 2 amino acids later (loss of PFAB and repeats)	None
5	rs929085219	indel	ENST00000368799.2		c.71delA	p.Asn24ThrfsTer5	PFA N-terminus	Extreme truncation 5 amino acids later (loss of PFAB and repeats)	None
6	rs1193300583	indel	ENST00000368799.2		c.83_84dup	Ser29Ter	PFA N-terminus	Truncation at mutation site (loss half end of PFB and all repeats)	None
7	rs114733570	SNP	ENST00000368799.2	RCV000709724.1	c.94G>T	p.Glu32Ter	PFA N-terminus	Extreme truncation at mutation site (loss of PFAB and repeats)	Ichthyosis vulgaris
8	rs752827914	indel	ENST00000368799.2		c.120_121insCTTCTGGAAAAGGAA	Leu36_Glu40dup	PFA N-terminus	Non synonymous variant	None
9	rs774514736	indel	ENST00000368799.2		c.198delC	p.Asn66LysfsTer11	PFA C-terminus	Extreme truncation 11 amino acids later (loss of PFAB and repeats)	None
10	rs769046914	indel	ENST00000368799.2		c.207delT	p.Ile69MetfsTer8	PFA C-terminus	Extreme truncation 8 amino acids later (loss of PFAB and repeats)	None
11	rs1195478941	indel	ENST00000368799.2		c.250_254del	p.Ala84LeufsTer2	PFA/B boundary	Extreme truncation 2 amino acids later (loss of PFB and repeats)	None
12	rs766604084	Inframe-deletion	ENST00000368799.2		c.306_314del	p.Arg102_His104del	PFB N-terminus	Non synonymous variant	None
13	rs762698116	indel	ENST00000368799.2		c.344delA	p.Gln115ArgfsTer79	PFB N-terminus	Extreme truncation 79 amino acids later (loss half end of PFB and repeats)	None
14	rs773056944	Inframe-deletion	ENST00000368799.2		c.356_364del	p.Lys119_Asn121del	PFB N-terminus	Non synonymous variant	None
15	rs769677999	indel	ENST00000368799.2		c.388_391del	p.Arg130GlufsTer63	PFB N-terminus	Extreme truncation 63 amino acids later (loss half end of PFB and repeats)	None
16	rs776289194	indel	ENST00000368799.2		c.441_442del	p.Gly149GlufsTer4	PFB N-terminus	Extreme truncation 4 amino acids later (loss half end of PFB and repeats)	None
17	rs761789405	indel	ENST00000368799.2		c.441delA	p.Arg151GlyfsTer43	PFB N-terminus	Extreme truncation 43 amino acids later (loss half end of PFB and repeats)	None
18	rs768243733	indel	ENST00000368799.2		c.451delA	p.Arg151GlyfsTer43	PFB N-terminus	Extreme truncation 43 amino acids later (loss half end of PFB and repeats)	None
19	rs746683647	indel	ENST00000368799.2	RCV000598732.1	c.477dupA	p.Glu160ArgfsTer10	PFB N-terminus	Truncation 10 amino acids later (loss half end of PFB and repeats)	Not Provided
20	rs746683647	indel	ENST00000368799.2	RCV000598732.1	c.477delA	p.Glu160LysfsTer34	PFB N-terminus	Truncation 34 amino acids later (loss half end of PFB and repeats)	Not Provided
21	rs1168180592	indel	ENST00000368799.2		c.482_485del	p.Arg161LysfsTer32	PFB N-terminus	Truncation 32 amino acids later (loss half end of PFB and repeats)	None
22	rs1214424848	SNP	ENST00000368799.2	RCV000763247.1	c.487G>T	p.Gly163Ter	PFB N-terminus	Truncation at the mutation site (loss half end of PFB and repeats)	Atopic Dermatitis, Eczema, Ichthyosis vulgaris
23	rs1214424848	SNP	ENST00000368799.2		c.487G>A	p.Gly163Arg	PFB N-terminus	Missense Variant	None
24	rs1557882660	indel	ENST00000368799.2		c.515_516del	p.Glu172ValfsTer7	PFB N-terminus	Truncation 7 amino acids later (loss half end of PFB and repeats)	None
25	rs1166843216	indel	ENST00000368799.2		c.527_528insA	p.Asn176LysfsTer4	PFB N-terminus	Truncation 4 amino acids later (loss half end of PFB and repeats)	None
26	rs1191273869	indel	ENST00000368799.2		c.529_530insTGTATTTTCATA	p.His177delinsLeuTyrPheHisAsn	PFB N-terminus	Altering Variant	None
27	rs1218912272	SNP	ENST00000368799.2	RCV000626817.1	c.544A>T	p.Lys182Ter	PFB N-terminus	Truncation at mutation site (loss half end of PFB and all repeats)	Abnormality of salivation and skin, Dry skin, Ichthyosis, Congenital cerebellar hypoplasia, Microcephaly, Feeding difficulties, Paroxysmal dystonia
28	rs1335400957	indel	ENST00000368799.2		c.549del	p.Glu184ArgfsTer10	PFB N-terminus	Truncation 10 amino acids later (loss half end of PFB and repeats)	None
29	rs771721862	indel	ENST00000368799.2	RCV000482898.1	c.557_558insA	p.Asn186LysfsTer4	PFB N-terminus	Truncation 4 amino acids later (loss half end of PFB and repeats)	Not Provided
30	rs749678415	indel	ENST00000368799.2		c.617_620del^621_624del	p.Glu208ThrfsTer15	PFB C-terminus	Truncation 15 amino acids later (loss end of PFB and all repeats)	None
31	rs778181683	Inframe-deletion	ENST00000368799.2		c.635_637del	p.Glu212del	PFB C-terminus	Non synonymous variant	None
32	rs756626190	indel	ENST00000368799.2	RCV000599043.1	c.660delA	p.Gly221GlufsTer3	PFB C-terminus	Truncation 3 amino acids later (loss end of PFB and all repeats)	Not Provided
33	rs1222586539	indel	ENST00000368799.2		c.678delA	p.Lys226AsnfsTer26	PFB C-terminus	Truncation 26 amino acids later (loss the end of PFB and all repeats)	None
34	rs753134537	indel	ENST00000368799.2		c.727delG	p.Ala243ProfsTer9	PFB C-terminus	Truncation 9 amino acids later (loss end of PFB and all repeats)	None
35	rs1381346375	Inframe-deletion	ENST00000368799.2		c.739_741del	p.Thr247del	PFB C-terminus	Non synonymous variant	None
36	rs781280495	indel	ENST00000368799.2	RCV000599537.1	c.745dupA	p.Ser249LysfsTer10	PFB C-terminus	Truncation 10 amino acids later (loss end of PFB and all repeats)	Not Provided
37	rs755000793	indel	ENST00000368799.2		c.762_766del	p.Lys255IlefsTer2	PFB C-terminus	Truncation 2 amino acids later (loss end of PFB and all repeats)	None
38	rs1557882406	indel	ENST00000368799.2		c.770delA	p.Tyr257LeufsTer15	PFB C-terminus	Truncation 15 amino acids later (loss end of PFB and all repeats)	None
39	rs386635460	substitution	ENST00000368799.2		c.769_770delinsGT	p.Tyr257Val	PFB C-terminus	Missense variant	None
40	rs199885226	SNP	ENST00000368799.2	RCV000358958.1	c.779C>R	p.Ser260Ter	PFB C-terminus	Truncation at mutation site (loss end of PFB and all repeats)	Not Provided
41	rs751629874	indel	ENST00000368799.2		c.834delT	p.Asn278LysfsTer168	PFB C-terminus	Truncation 168 amino acids later (loss start from half end of the first repeat)	None
42	rs1400304361	indel	ENST00000368799.2		c.856delG	p.Glu286SerfsTer160	PFB C-terminus	Truncation 160 amino acids later (loss start from half end of the first repeat)	None

b



Supplementary Figure S1.
Select mutations localized in the human profilaggin A and B domains.
(a) List of select mutations in profilaggrin A and B domains. The databases at the National Center for Biotechnology Information and Ensembl GRCh37 were used to identify mutations in the A and B domains that satisfied the following criteria: were deletions, insertions, or had reported clinical relevance. We excluded gain of premature stop codons unless clinical relevance had been established. These criteria identified 42 mutations, 11 in PF-A (blue, top) and 31 in PF-B (red, bottom). Under "Condition": None means no clinical effect documented to date; Not Provided means clinical relevance reported but exact condition not specified. Mutations with reported disease association have those diseases listed in red. **(b)** The 42 mutations listed in (a) are mapped to the PF-A and PF-B domains, showing their relative location. PF-A is further subdivided into structural domains, where six mutations (4-7, 9-10) lie within the calcium binding loops. One mutation (15) lies within the nuclear localization signal (NLS) of PF-B.

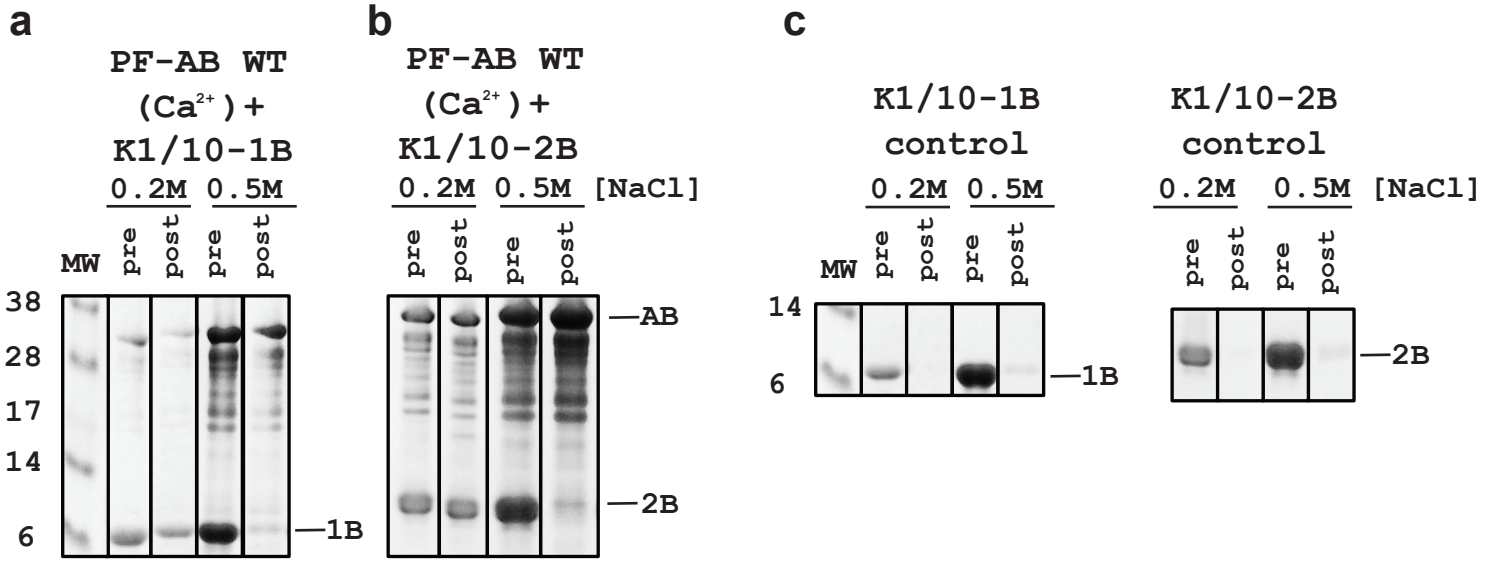


Supplementary Figure S2. **Predicted antimicrobial peptide regions in S100 fused type proteins.**
a) Multiple sequence alignment of carboxy-terminal domains (CTD) of the SFTP family shows poor conservation.
b) Predicted antimicrobial regions in the CTD for human profilaggrin, profilaggrin-2, trichohyalin-like 1 protein, and repetin. Lines above the bolded sequence represent computationally predicted antimicrobial peptide regions with the different colors indicating different algorithms used to make the prediction (black, SVM; red, Random Forest; green, Discriminant Analysis; blue, Artificial Neural Network). Letters below each bolded sequence represent respective predicted secondary structure. C, coil; T, turn; E, sheet; H, helix.

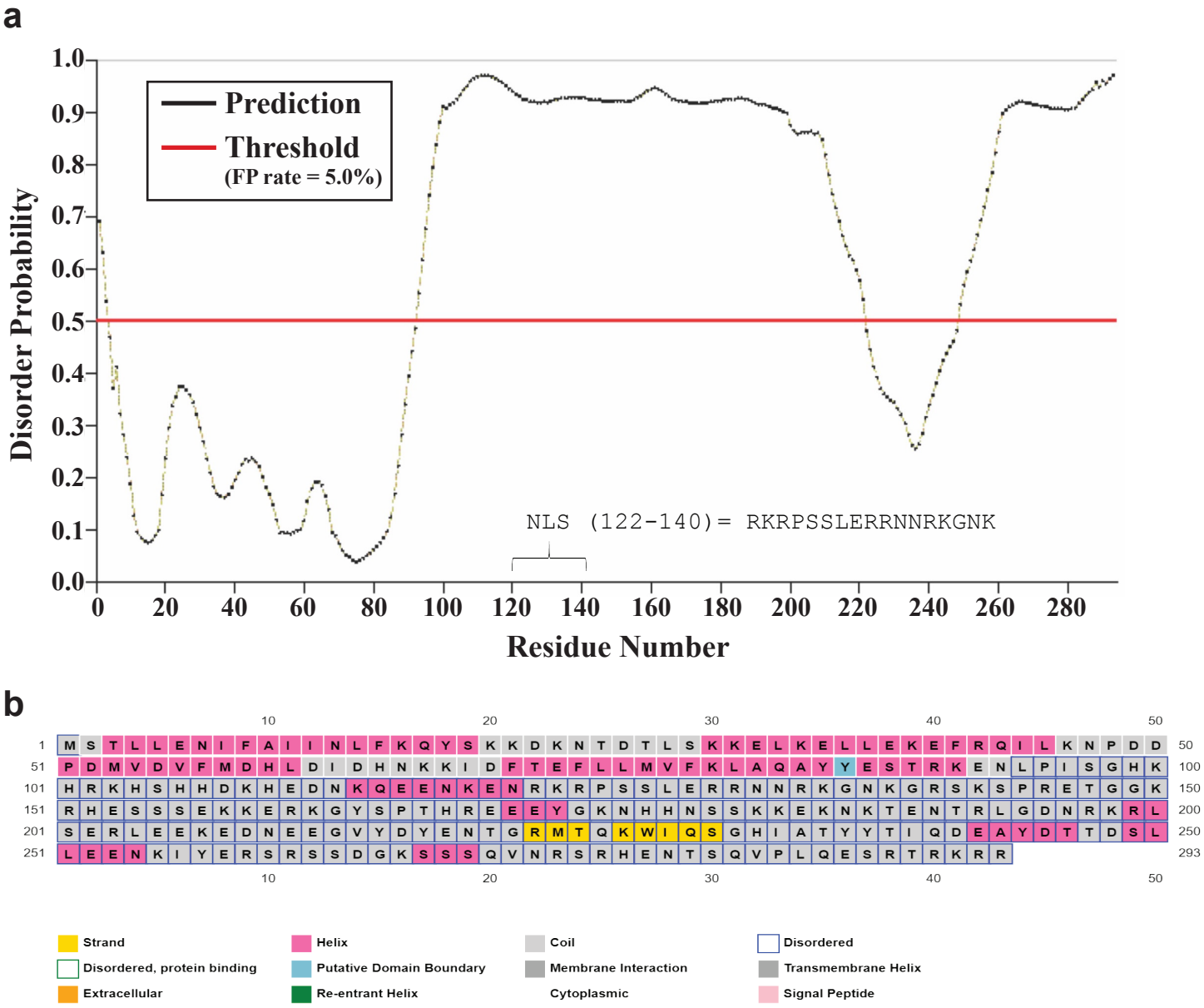
SFTP	PROFILAGGRIN (FLG)	FILAGGRIN-2 (FLG2)	HORNERIN (HRNR)	REPETIN (RPTN)	CORNULIN (CRNN)	TRICHOHYALIN (TCHH)	TRICHOHYALIN-LIKE PROTEIN-1 (TCHHL1)
CENTRAL REPEATS	10-12 monomers. Further degradation to component amino acids (NMF). Highly charged. Mostly random coil, almost no helix formation.	Two types: A-type (50-77% homology with hornerin); B-type (28-39% homology with filaggrin). Rich in histidine and glutamine.	5 repeats of three segments (A, B, and C), divided into four repeats. Glycine loops suggestive of elastic and adhesive properties.	28 repeats of 12 amino acids with positional conservation of glutamine, glycine, serine, and histidine.	2 repeat sequences, rich in glutamine and threonine. Homology to a bacterial GPI anchored protein.	Domains 2-8, each containing varying repeats. Mostly α-helical. Rich in glutamine.	Rich in glutamine and lysine. Contains NLS.
C-TERMINAL DOMAIN	Function unknown. Essential for processing.	Antimicrobial activity.	Function unknown.	Function unknown.	Function unknown. Likely random-coil.	Function unknown. Complete conservation of 13 residues.	Function unknown. Contains TM domain.
KERATIN-BINDING	Yes; filaggrin monomers.	Yes; demonstrated *in vitro*.	Yes; domain 6 and 8, CTD.	-	-	Yes; with IRS keratin intermediate filaments.	-
AMP ACTIVITY	CTD may be potential AMP.	C-terminal domain	Tandem A repeat	CTD may be potential AMP.	-	-	CTD may be potential AMP.
NUCLEUS TARGETING ACTIVITY	Yes; NLS in B domain.	Unknown	Yes; bactericidal via ribosome-targeting.	Unknown	Unknown	8 predicted monopartite NLS	Unknown
EXPRESSION	Keratohyalin granules.	Keratohyalin granules.	Keratohyalin granules.	Inter-follicular epidermis, IRS, acrosyringium.	Keratinocytes; scalp skin, foreskin.	IRS	Distal IRS, basal layer and keratinocytes.
DISEASE IMPLICATION	Ichthyosis vulgaris, Atopic dermatitis, asthma, hay-fever, peanut allergy.	Mutations associated with Atopic dermatitis, peeling skin syndromes 3 and 6; Downregulated in psoriatic lesions.	Overexpressed in: hepatocellular carcinoma, psoriatic and wound-healing skin.	Low protein level associated with schizophrenia and bipolar disorder; Mutations associated with Atopic dermatitis.	Downregulation associated with oral esophageal and squamous carcinoma.	Mutations cause uncombable hair syndrome 3; associated with Alopecia areata.	Strongly expressed in basal- and squamous-cell carcinoma; overexpression in Psoriasis vulgaris and Lichen planus.

AMP: anti-microbial peptide; CTD: C-terminal domain; IRS: inner root sheath; NMF: natural moisturizing factor; TM: transmembrane domain; NLS: nuclear localization sequence.

Supplementary Figure S3. Distinguishing characteristics of S100 fused-type proteins (SFTPs) and their clinical relevance.



Supplementary Figure S4.
PF-AB interactions with K1/K10 are reduced in the presence of elevated NaCl concentration. Ni^{2+} pulldown assays in the presence of Ca^{2+} using His_6-tagged PF-AB as bait protein for keratin 1/10-1B (a) and keratin 1/10-2B (b) heterocomplex at two different NaCl concentrations (200 mM, 500 mM). Reduced PF-AB binding of K1/K10 is observed for both the 1B and 2B domains in the 500 mM NaCl condition. Lanes are designated either "pre" or "post" the washing away of unbound proteins. (c) K1/K10-1B and K1/K10-2B do not bind the Ni^{2+} resin in the absence of PF-AB. The PF-AB (res. 1-257) construct was used in this experiment.



Supplementary Figure S5. The B domain is predicted to be highly disordered with little rigid secondary structure.

a. Disorder prediction (DISOPRED, University College London) for the human profilaggrin N-terminus (PF-AB). All data points above the red threshold line represent elevated disorder prediction. The profilaggrin S100 (A) domain (residues ~1-88) is predicted to be highly ordered compared to the B domain (~89-293) with the exception of one region (residues ~220-250) that is C-terminal to the nuclear localization signal (NLS).

b. Secondary structure (SS) prediction of the human profilaggrin N-terminus (PF-AB) (PSIPRED, University College London) demonstrates over 70% disorder prediction for the B domain. The predominant SS type is (random) coil with a few small alpha-helices and beta-sheets scattered throughout the sequence.