The American Journal of Human Genetics, Volume *100*

Supplemental Data

Duplicated Enhancer Region Increases Expression of *CTSB* and Segregates with Keratolytic Winter Erythema in South African and Norwegian Families

Thandiswa Ngcungcu, Martin Oti, Jan C. Sitek, Bjørn I. Haukanes, Bolan Linghu, Robert Bruccoleri, Tomasz Stokowy, Edward J. Oakeley, Fan Yang, Jiang Zhu, Marc Sultan, Joost Schalkwijk, Ivonne M.J.J. van Vlijmen-Willems, Charlotte von der Lippe, Han G. Brunner, Kari M. Ersland, Wayne Grayson, Stine Buechmann-Moller, Olav Sundnes, Nanguneri Nirmala, Thomas M. Morgan, Hans van Bokhoven, Vidar M. Steen, Peter R. Hull, Joseph Szustakowski, Frank Staedtler, Huiqing Zhou, Torunn Fiskerstrand, and Michele Ramsay

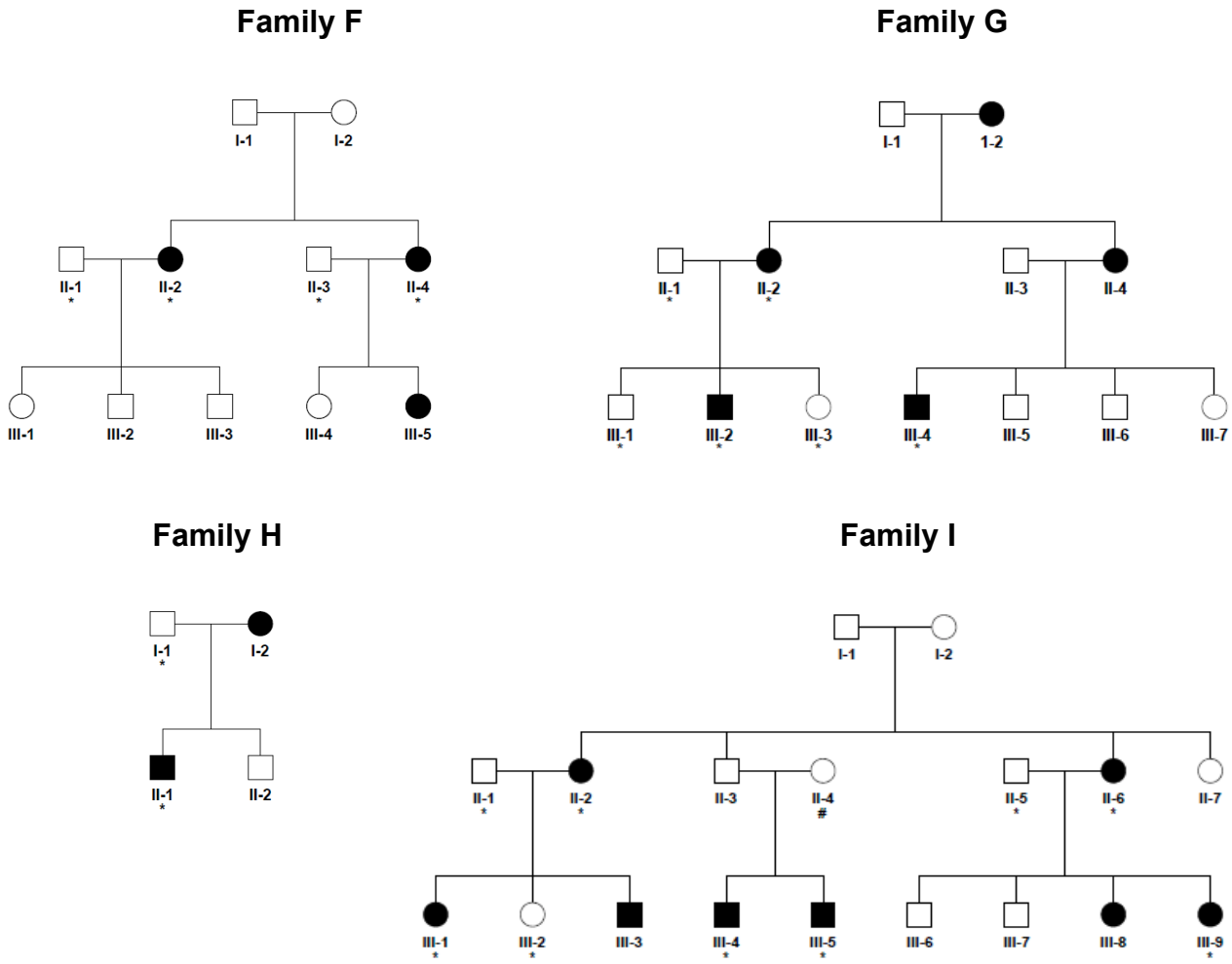

Figure S1. Validation of the tandem duplication in South African families with KWE

The tandem duplication at chr8:11729286-11736956 was analysed by PCR and Sanger sequencing (Table S1, Figure S2). Eleven affected and twelve non-affected individuals (*) from families F-I (pedigrees shown here) and individual I-1 from family C (Figure 1) were analysed. The duplication segregated completely with the disease. The pedigrees were drawn using HaploPainter V.1.043.[1]

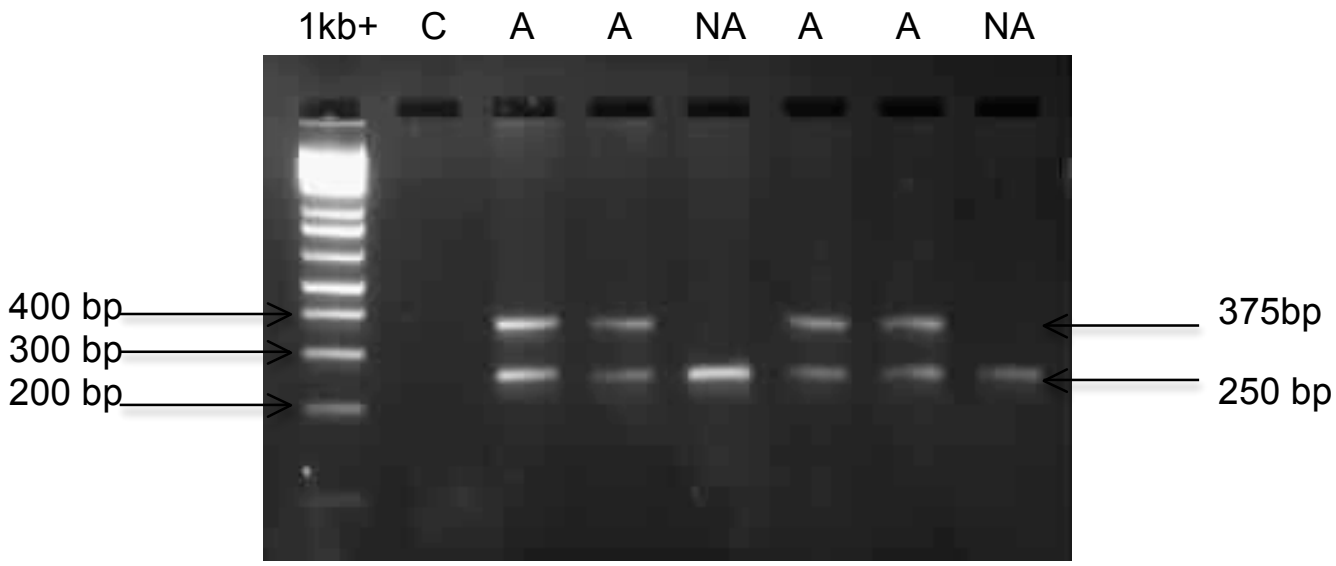

Figure S2. Mapping of duplication breakpoints in South African families. The breakpoints for the tandem duplication were determined by Pindel and visualized using the Integrative Genomics Viewer.[2] The junction sequence of the duplicated region is unique and was only present in individuals with the duplication. The picture shows gel electrophoresis of the PCR output using primer pairs for the duplication junction region (SA_Junction) and a control amplicon from the *LEP* gene (SA_Control). Primer sequences are shown in Table S1. The control amplicon (250 bp) is present in all samples and the duplication junction amplicon (375 bp) is present only in affected individuals (A), and not in unaffected individuals (NA).

.

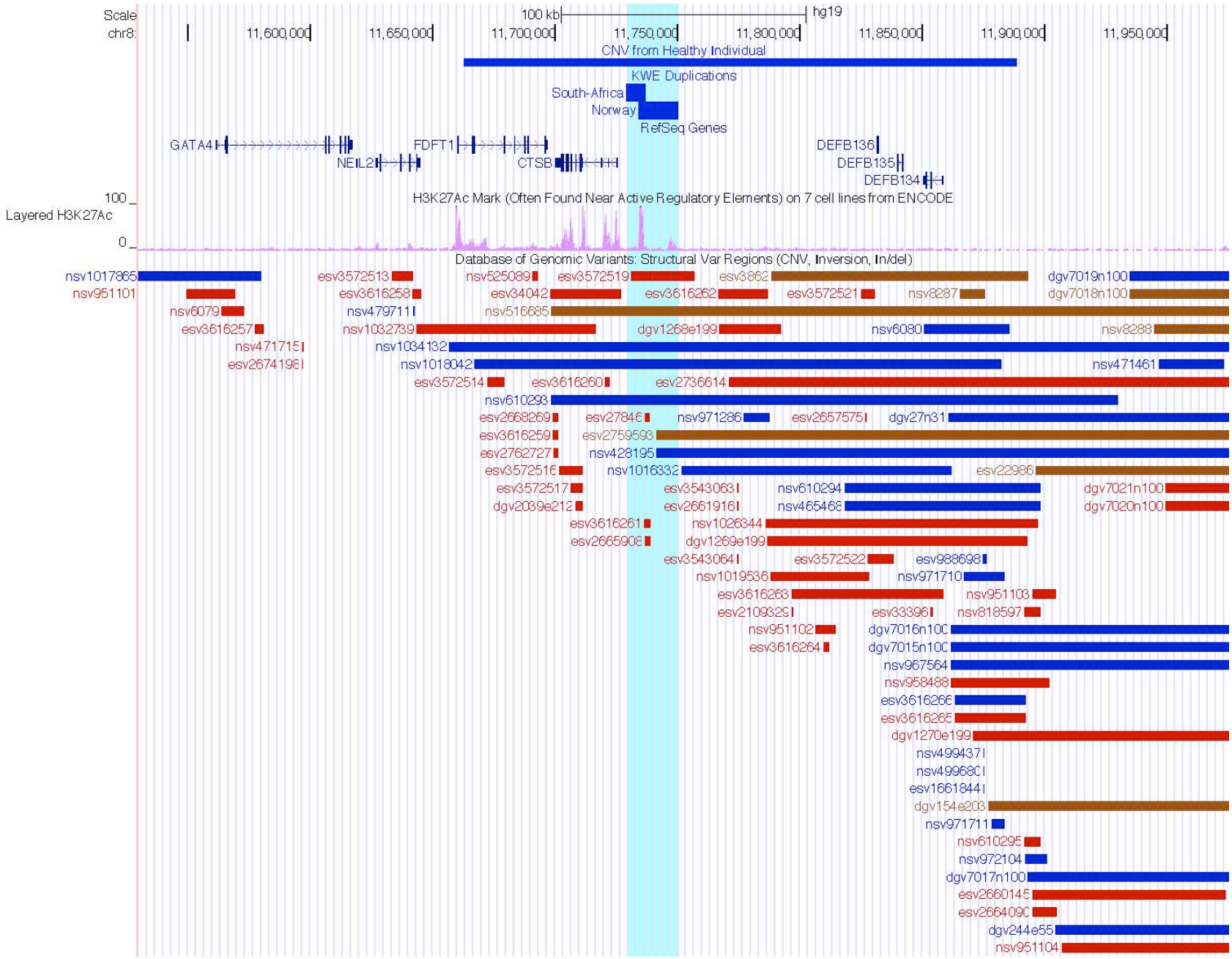
Scale
chr8:
100 kb
hg19
11,600,000
11,650,000
11,700,000
11,750,000
11,800,000
11,850,000
11,900,000
11,950,000
CNV from Healthy Individual
KWE Duplications
South-Africa
Norway
RefSeq Genes
GATA4
NEIL2
FDFT1
CTSB
DEFB136
DEFB135
DEFB134
100
Layered H3K27Ac
0
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Database of Genomic Variants: Structural Var Regions (CNV, Inversion, In/del)
nsv1017865
nsv951101
nsv6079
esv3616257
nsv471715
esv2674198
esv3572513
esv3616258
nsv479711
nsv1032739
nsv1034132
nsv1018042
esv3572514
esv3616260
nsv525089
esv3572519
esv34042
nsv516685
dgv1268e199
esv3572516
esv3572517
dgv2089e212
esv3616261
esv2665908
esv3543064
nsv1019536
esv3616263
esv2109329
nsv951102
esv3616264
esv3862
esv3616262
esv3572521
nsv8287
nsv6080
esv2736614
nsv610293
esv2668269
esv3616259
esv2762727
esv27846
esv2759593
nsv428195
nsv1016332
esv3543063
esv2661916
nsv1026344
dgv1269e199
esv3572522
nsv971286
esv2657575
nsv610294
nsv465468
esv988698
nsv971710
nsv951103
esv33396
nsv818597
dgv7016n100
dgv7015n100
nsv967564
nsv958488
esv3616266
esv3616265
dgv1270e199
nsv499437
nsv499680
esv1661844
dgv154e203
nsv971711
nsv610295
nsv972104
dgv7017n100
esv2660145
esv2664090
dgv244e55
nsv951104
dgv7019n100
dgv7018n100
nsv8288
nsv471461
esv22986
dgv7021n100
dgv7020n100

Figure S3. CNVs in the KWE critical region reported in normal individuals. Schematic overview of the KWE critical region on chr8:11525326-11987757, with scale shown in the upper panel. Blue horizontal bars indicate duplications, red bars are deletions, and brown bars are both duplicated and deleted within the normal population. At the top, a large duplication identified in a Dutch cohort of 1416 healthy students[3] and the tandem duplications identified in South African (7.67 kb) and Norwegian (15.93 kb) KWE patients are displayed as blue horizontal bars. The bottom track shows deletions and duplications in the healthy population from the Database of Genomic Variants, with accession numbers starting with "esv" or "nsv" depending on whether their source was the EBI or the NCBI. No duplications (blue bars) of similar size to those described in South African and Norwegian patients (turquoise shade across all panels) have been reported within the region. Several larger duplications (including the one in the Dutch cohort) are shown encompassing the region of the enhancer at chr8: 11734333-11736956 (large pink peak within the turquoise shade), but all of them include the *CTSB* gene and even the *FDFT1* gene, or the duplication does not extend to include the enhancer (nsv428195).

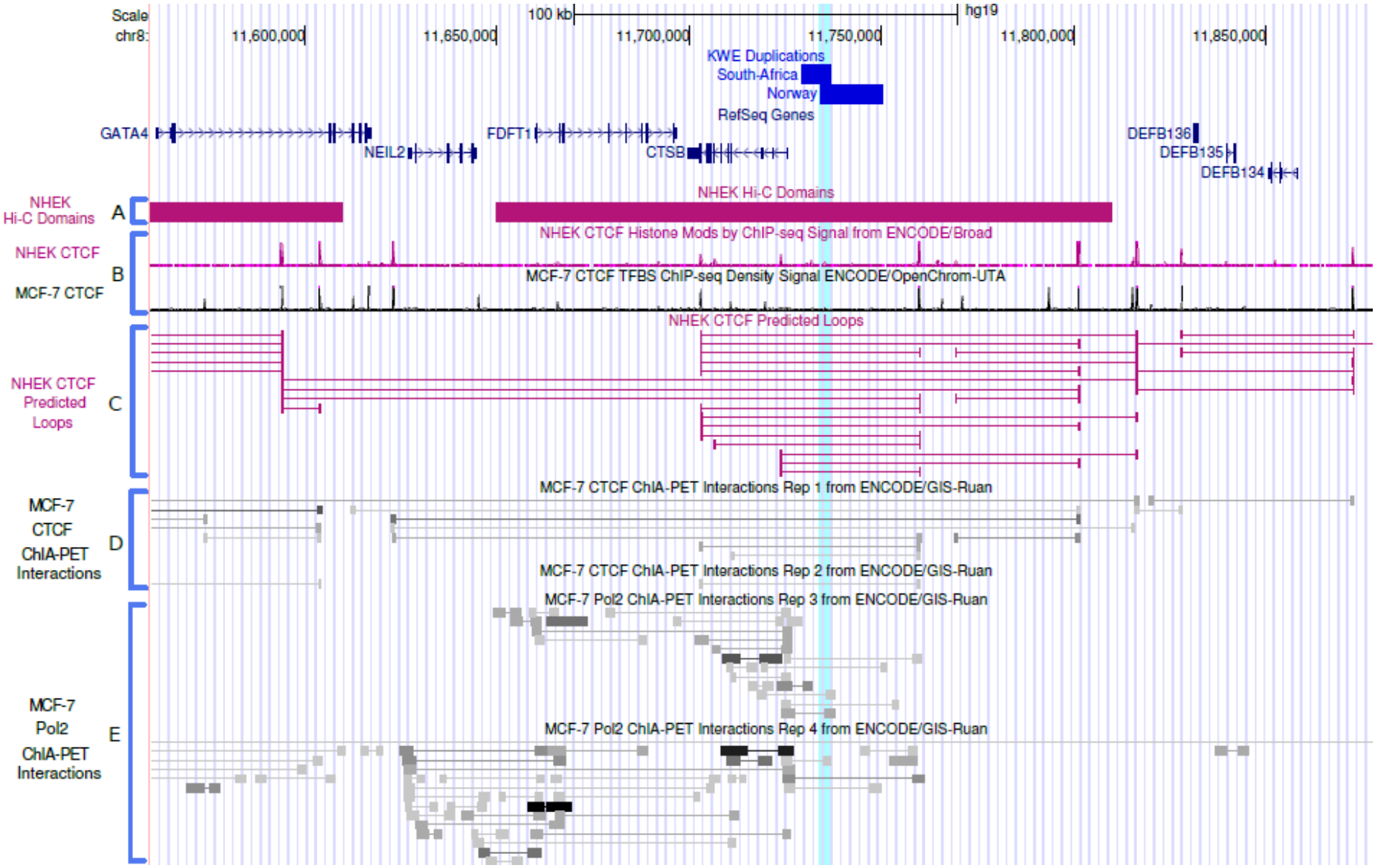
Scale
chr8:
100 kb
hg19
11,600,000
11,650,000
11,700,000
11,750,000
11,800,000
11,850,000
KWE Duplications
South-Africa
Norway
RefSeq Genes
GATA4
NEIL2
FDFT1
CTSB
DEFB136
DEFB135
DEFB134
NHEK Hi-C Domains
NHEK Hi-C Domains
A
NHEK CTCF Histone Mods by ChIP-seq Signal from ENCODE/Broad
NHEK CTCF
B
MCF-7 CTCF TFBS ChIP-seq Density Signal ENCODE/OpenChrom-UTA
MCF-7 CTCF
NHEK CTCF Predicted Loops
NHEK CTCF Predicted Loops
C
MCF-7 CTCF ChIA-PET Interactions Rep 1 from ENCODE/GIS-Ruan
MCF-7 CTCF ChIA-PET Interactions
D
MCF-7 CTCF ChIA-PET Interactions Rep 2 from ENCODE/GIS-Ruan
MCF-7 Pol2 ChIA-PET Interactions Rep 3 from ENCODE/GIS-Ruan
MCF-7 Pol2 ChIA-PET Interactions
E
MCF-7 Pol2 ChIA-PET Interactions Rep 4 from ENCODE/GIS-Ruan

Figure S4. Topological subdomains, CTCF binding sites and chromatin interactions involving the enhancer and nearby genomic regions

Schematic overview of the KWE critical region on chr8:11560000-11880000, with scale shown in the upper panel. The South African (7.67 kb) and Norwegian (15.93 kb) tandem duplications are displayed as blue horizontal bars, and the 2.62 kb overlap (chr8:11734333-11736955) is marked across all horizontal panels (turquoise shading). (A) Hi-C data from the NHEK cell line (bold pink bars) placed the enhancer in the same topological region with the *FDFT1* and *CTSB* genes. (B) CTCF binding sites in the keratinocyte (NHEK, pink peaks) and the breast cancer (MCF-7, black peaks) cell lines are highly similar. (C) Using the CTCF binding sites identified in NHEK cells, CTCF interaction loops (and hence subdomains) in these cells were predicted.[4] These data indicate a larger subdomain extending from the strong CTCF binding site at the centromeric end of the NHEK Hi-C domain to the *GATA4* gene, and a smaller subdomain, including only the enhancer and *CTSB*. (D) ChIA-PET CTCF data from MCF-7 cells showing that *CTSB*, *FDFT1* and *NEIL2* may occur in the same topological domain with the enhancer (Rep 1 (Repetition 1)). A smaller subdomain exists that only includes the *CTSB* gene and the enhancer (Rep 1 and 2). (E) MCF-7 Pol2 ChIA-PET interaction data show interaction between the enhancer and the *CTSB* promoter, but not with the promoters of *FDFT1 or NEIL2* (Rep 3 and 4). Hi-C, CFCF-binding sites and ChIA-PET CTCF data are from the ENCODE data.[5]

Table S1. Primers used to verify South African and Norwegian tandem duplication breakpoints

Primer pairs	Primer Sequences (5' – 3')		Amplicon size
	Forward	Reverse	
SA_Junction*	CTAGGCTTGCAGTGTTGGTC	GTTAAATCAGGCTGGGCGAG	375 bp
SA_Control*	AGCCAAGGCAAAATTGAGG	TCCAGCCGATCTCTCTGTTC	250 bp
N_Junction**	GCCTGGCCACTTTCTTTCTT	GGTCATATGCTCAGGCAGGT	505 bp
N_Insertion**	CCGCATCCAGCATTTTTATT	CTGCTCCAAGTCACCCTCTC	624 bp

*Forward and reverse primers were selected for verification of the breakpoints in the South African duplication (SA_Junction) along with a control primer set (SA_control) overlapping the *LEP* gene. The SA_control primer pair was included in a multiplex PCR reaction to determine PCR efficiency. The breakpoint and the control amplicon were amplified using 1X KAPA TaqReadMix, 0.4 µM of each primer and 0.1 µg of genomic DNA. Cycling conditions included a 3 minute initial denaturation at 95°C, followed by 35 cycles of 30 seconds denaturation (95° C), 30 seconds annealing (55°C) and a 30 seconds extension at 72°C with a 1 minute final extension at 72°C.

**In the Norwegian affected individuals, the duplication breakpoints were verified by WGS in two individuals, and Sanger sequencing demonstrated a tandem duplication (N_Junction primers) with a 95 bp insertion (i.e. triplication; N_Insertion primers) in between. PCR conditions: 1xATG 360 Mastermix, 0.5 µM of each primer and 100 ng genomic DNA. Cycling conditions were as described above, except for a 10 min initial denaturation at 95°C.

Table S2: Relative gene expression and immunohistochemistry findings for *CTSB*

Sample	Status*	Sex	*CTSB*_CT**	*RPLP0*_CT**	Fold change	# of granular layers	Intensity***
NOR _Ctr1	N	F	26.21	21.86	1.39	2	0
NOR _Ctr2	N	F	26.02	21.72	1.45	3	0
NOR _Ctr3	N	F	25.73	21.72	1.78	3-4	1
NOR _Ctr4	N	F	25.75	21.78	1.81	4	1
SA_Ctr1	N	F	28.70	22.48	0.38	4	0
SA _Ctr2	N	F	27.45	22.24	0.77	4	0
SA _Ctr3	N	M	27.96	22.19	0.52	3	0
NOR_DII-1	A	F	24.55	22.62	7.46	6	2.5
NOR_DII-6	A	F	24.65	22.77	7.77	4-5	2.5
NOR_DIII-6	A	F	24.22	22.05	6.29	4	1
NOR_EI-2	A	F	23.83	22.12	8.67	6	0.5
SA_KWE2	A	M	24.57	21.55	3.49	5	2.5
SA_KWE3	A	M	26.19	22.37	2.02	5	2.5
SA_KWE1	A	F	no result	no result	no result	4	0.5

	Average fold change	Standard errors of the mean
Control (N)	1.16	0.19
Affected (A)	5.95	0.93

* N=Not affected, A=Affected

** CT= Cycle threshold

*** Intensity of CTSB staining in the granular layer of the epidermis

Table S3: Relative gene expression *FDFT1*

Sample	Status*	Sex	*FDFT1*_CT**	*RPLP0*_CT**	Fold change
NOR _Ctr1	N	F	29.40	21.86	1.33
NOR _Ctr2	N	F	28.74	21.72	1.90
NOR _Ctr3	N	F	28.81	21.72	1.82
NOR _Ctr4	N	F	28.74	21.78	1.98
SA_Ctr1	N	F	31.45	22.48	0.49
SA _Ctr2	N	F	31.24	22.24	0.48
SA _Ctr3	N	M	31.25	22.19	0.46
NOR_DII-1	A	F	29.55	22.62	2.02
NOR_DII-6	A	F	30.09	22.77	1.54
NOR_DIII-6	A	F	29.19	22.05	1.74
NOR_EI-2	A	F	29.05	22.12	2.01
SA_KWE2	A	M	28.98	21.55	1.43
SA_KWE3	A	M	30.33	22.37	0.99

	Average fold change	Standard errors of the mean
Control (N)	1.21	0.22
Affected (A)	1.62	0.13

* N=Not affected, A=Affected

** CT= Cycle threshold

Table S4: Relative gene expression for *NEIL2*

Sample	Status*	Sex	*NEIL2*_CT**	*RPLP0*_CT**	Fold change
NOR _Ctr1	N	F	29.91	21.86	2.17
NOR _Ctr2	N	F	29.75	21.72	2.21
NOR _Ctr3	N	F	29.24	21.72	3.15
NOR _Ctr4	N	F	30.70	21.78	1.19
SA_Ctr1	N	F	32.85	22.48	0.44
SA _Ctr2	N	F	32.54	22.24	0.46
SA _Ctr3	N	M	33.20	22.19	0.28
NOR_DII-1	A	F	30.86	22.62	1.90
NOR_DII-6	A	F	30.97	22.77	1.95
NOR_DIII-6	A	F	30.33	22.05	1.85
NOR_EI-2	A	F	29.64	22.12	3.11
SA_KWE2	A	M	29.55	21.55	2.24
SA_KWE3	A	M	32.01	22.37	0.73

	Average fold change	Standard errors of the mean
Control (N)	1.41	0.30
Affected (A)	1.96	0.22

* N=Not affected, A=Affected

** CT= Cycle threshold

Supplemental References

1. Thiele, H., and Nurnberg, P. (2005). HaploPainter: a tool for drawing pedigrees with complex haplotypes. Bioinformatics 15, 1730-1732.
2. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol 29, 24-26.
3. Bralten, J., Arias-Vasquez, A., Makkinje, R., Veltman, J.A., Brunner, H.G., Fernandez, G., Rijpkema, M., and Franke, B. (2011). Association of the Alzheimer's gene *SORL1* with hippocampal volume in young, healthy adults. Am J Psychiatry 168, 1083-1089.
4. Oti, M., Falck, J., Huynen, M.A., and Zhou, H. (2016). CTCF-mediated chromatin loops enclose inducible gene regulatory domains. BMC Genomics 17, 252.
5. ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74.