

1 Alopecia areata susceptibility variant identified by MHC risk haplotype sequencing
2 reproduces symptomatic patched hair loss in mice

3

4 Akira Oka^{a*}, Atsushi Takagi^b, Etsuko Komiyama^b, Shuhei Mano^c, Kazuyoshi Hosomichi^d, Shingo
5 Suzuki^e, Nami Motosugi^e, Tomomi Hatanaka^{e,f}, Minoru Kimura^{a,e}, Mahoko Takahashi Ueda^g, So
6 Nakagawa^{e,g}, Hiromi Miura^{e,h}, Masato Ohtsuka^{a,e,h}, Yuko Haida^e, Masayuki Tanakaⁱ, Tomoyoshi
7 Komiyama^j, Asako Otomo^e, Shinji Hadano^{a,e}, Tomotaka Mabuchi^k, Stephan Beck^l, Hidetoshi Inoko^e,
8 Shigaku Ikeda^{b*}

9

10 ^aThe Institute of Medical Sciences, Tokai University, 143 Shimokasuya, Isehara, Kanagawa, 259-1193,
11 Japan

12 ^bDepartment of Dermatology and Allergology, and Atopy (Allergy) Research Center, Juntendo
13 University Graduate School of Medicine, 2-1-1 Hongo, Bunkyo, Tokyo 113-8421, Japan

14 ^cDepartment of Mathematical Analysis and Statistical Inference, The Institute of Statistical
15 Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

16 ^dDepartment of Bioinformatics and Genomics, Graduate School of Advanced Preventive Medical
17 Sciences, Kanazawa University, Takara-machi 13-1, Kanazawa, Ishikawa 920-8640, Japan

18 ^eDepartment of Molecular Life Sciences, Division of Basic Medical Science and Molecular Medicine,
19 Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan

20 ^fFaculty of Pharmaceutical Sciences, Josai University, Sakado, Saitama, 350-0295, Japan

21 ^gMicro/Nano Technology Center, Tokai University, Kanagawa, Japan

22 ^hCenter for Matrix Biology and Medicine, Graduate School of Medicine, Tokai University, 143
23 Shimokasuya, Isehara, Kanagawa 259-1193, Japan

24 ⁱDepartment of Bioinformatics, Support Center for Medical Research and Education, Tokai University,
25 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan

26 ^jDepartment of Clinical Pharmacology, Tokai University School of Medicine, 143 Shimokasuya,
27 Isehara, Kanagawa 259-1193, Japan

1 ^kDepartment of Dermatology, Tokai University School of Medicine, 143 Shimokasuya, Isehara,
2 Kanagawa 259-1193, Japan

3 ^lMedical Genomics, UCL Cancer Institute, University College London, London WC1E 6BT, UK

4

5 *To whom correspondence should be addressed. Email: oka246@is.icc.u-tokai.ac.jp or
6 ikeda@juntendo.ac.jp

7

8 Akira Oka: oka246@is.icc.u-tokai.ac.jp, Atsushi Takagi: t-attsu@juntendo.ac.jp, Etsuko Komiyama:

9 etsuko-k@juntendo.ac.jp, Shuhei Mano: smano@ism.ac.jp, Kazuyoshi Hosomichi:

10 khosomic@med.kanazawa-u.ac.jp, Shingo Suzuki: ss079143@tsc.u-tokai.ac.jp, Nami Motosugi:

11 motosugi@is.icc.u-tokai.ac.jp, Tomomi Hatanaka: tmmhtnk@tokai-u.jp, Minoru Kimura:

12 kimura@is.icc.u-tokai.ac.jp, Mahoko Takahashi Ueda: mahoko@tokai.ac.jp, So Nakagawa:

13 so@tokai.ac.jp, Hiromi Miura: hiromi@tokai-u.jp, Masato Ohtsuka: masato@is.icc.u-tokai.ac.jp,

14 Yuko Haida: y-haida@oregano.ocn.ne.jp, Masayuki Tanaka: matanaka@tokai-u.jp, Tomoyoshi

15 Komiyama: komiyama@tokai-u.jp, Asako Otomo: asako@tokai-u.jp, Shinji Hadano:

16 shinji@is.icc.u-tokai.ac.jp, Tomotaka Mabuchi: mabuchi@is.icc.u-tokai.ac.jp, Stephan Beck:

17 s.beck@ucl.ac.uk, Hidetoshi Inoko: hinoko@is.icc.u-tokai.ac.jp, Shigaku Ikeda:

18 ikeda@juntendo.ac.jp

19

20

21

22

23

24

25

26

27

1 **Abstract**

2 **Background:** Alopecia areata (AA) is a highly heritable multifactorial and complex disease. However,
3 no convincing susceptibility gene has yet been pinpointed in the major histocompatibility complex
4 (MHC), a region in the human genome known to be associated with AA as compared to other regions.

5 **Results:** By sequencing MHC risk haplotypes, we identified a variant (rs142986308, p.Arg587Trp) in
6 the coiled-coil alpha-helical rod protein 1 (*CCHCR1*) gene as the only non-synonymous variant in the
7 AA risk haplotype. Using CRISPR/Cas9 for allele-specific genome editing, we then phenocopied AA
8 symptomatic patched hair loss in mice engineered to carry the *Cchcr1* risk allele. Skin biopsies of
9 these alopecic mice showed strong up-regulation of hair-related genes, including hair keratin and
10 keratin-associated proteins (KRTAPs). Using transcriptomics findings, we further identified
11 CCHCR1 as a novel component of hair shafts and cuticles in areas where the engineered alopecic mice
12 displayed fragile and impaired hair.

13 **Conclusions:** These results suggest an alternative mechanism for the aetiology of AA based on
14 aberrant keratinization, in addition to generally well-known autoimmune events.

15

16 **Keywords:** alopecia areata, CRISPR/Cas9, MHC, haplotype, association, sequencing

17

18 **Background**

19 Alopecia areata (AA) is a complex disease defined by focal or universal hair loss, most commonly
20 occurring on the scalp[1]. Although benign, AA can seriously impair quality of life, such as negative
21 effects on self-perception and self-esteem[2], and even attempted suicide in rare cases[3]. The
22 prevalence of AA is approximately 0.1% to 0.2% in the United States, with a lifetime risk of 1.7%[4],
23 while a twin study suggested a 55% concordance rate in identical twins with a significant occurrence
24 of AA in families[5]. In addition, the prevalence rate of AA in families has been shown to be higher
25 than that in the general public, though the rate varied in each study and population examined[6-10].

26 AA is driven by cytotoxic T lymphocytes and was found to be reversed by Janus kinase (JAK)
27 inhibition in model mice by clinical treatment[11], demonstrating that the immunological pathway is

1 a factor in this multifactorial disease. Previous genome-wide association study results have implicated
2 a number of immune and non-immune loci in the aetiology of AA[12-14], though none has yet been
3 demonstrated to be causative for the disease. Thus, no variants among those have provided
4 experimental evidence for biological functions between alleles and AA pathogenesis. Alleles of the
5 human leukocyte antigen (HLA) genes within the major histocompatibility complex (MHC) region on
6 chromosome 6p21.3 have thus far shown the strongest associations with AA among the human
7 genome[12-14] in observations of different ethnic groups. However, the strongest associations with
8 AA have not been supported by functional evidence.

9 The genetic architecture of the MHC region shows that multiple haplotypes with the highest
10 degree of diversity are often maintained in a population by balancing selection, and that positive
11 selection can occasionally generate long-range haplotypes[15, 16] . The strong linkage disequilibrium
12 (LD) observed in such haplotypes can mask the ability to discriminate between a *bona fide* variant
13 associated with disease and a variant influenced by LD. This limitation can be addressed by analysis
14 of microsatellites that have higher mutation rates than SNPs, thus leading to breakup of apparently
15 invariant SNP haplotypes into lower frequency haplotypes for functional analysis[17]. Analysis of
16 multi-allelic microsatellites may therefore be an effective strategy for identifying rare disease-
17 associated haplotypes in the MHC.

18 With this background in mind, we implemented a 4-step study design. First, we performed
19 association analysis using microsatellites for the entire MHC region with AA patients and healthy
20 controls to identify risk haplotypes associated with AA. Second, we sequenced representative risk and
21 control haplotypes to identify variants that were present only in identical risk haplotypes based on all
22 of the variants detected. Third, we performed whole exome sequencing of the variant-associated
23 candidate gene, haplotype estimation, and extended haplotype homozygosity (EHH) analysis in all
24 subjects. Finally, we engineered mice carrying the human risk allele using allele-specific genome
25 editing with the CRISPR/Cas9 system and performed functional evaluations.

26

27

1 Results

2 Association analysis and risk haplotype sequencing

3 A total of 171 AA patients and 560 healthy controls were enrolled for association analysis using
4 22 microsatellites spanning the human leukocyte antigen (HLA) class I and II regions (chr6:
5 30407655- 32854116, hg19), including the *HLA-C* locus, which was previously implicated to be
6 involved in AA[18]. We detected a single microsatellite, *D6S2811* (allele 208, OR = 3.41, CI 95% =
7 1.94-5.99, $P = 3.39 \times 10^{-5}$), which was shown to be significantly associated with AA after Bonferroni
8 correction (**Fig. 1a and Additional file1: Table S1 and S2**). Pair-wise evaluation of these multi-allelic
9 loci indicated that a strong long-range LD was maintained across the assayed region of the MHC
10 (**Additional file1: Figure S1**). Furthermore, estimation of haplotypes in 3 loci from *D6S2811* to
11 *D6S2930* showed that the risk haplotype was defined as the segment that displayed a significant
12 association with AA (MShap01, OR = 3.78, CI 95% = 2.00 - 7.16, $P = 6.57 \times 10^{-5}$) (**Table 1**).

13 To move from the identified risk haplotypes to causal AA variants, we next sequenced 5
14 individuals with MShap01 and 7 individuals with the other non-risk haplotypes spanning the entire
15 MHC (chr6:28477797-33451433, hg19) (**Additional file1: Table S3**). As all risk haplotypes were
16 heterozygous, any AA causal variant(s) would be expected to be heterozygous as well. Therefore,
17 variants were accordingly filtered and only variants found to be identical between risk haplotypes were
18 retained. Following this strategy, we extracted 3895 heterozygous risk variants from the 77,040
19 variants identified in the 12 individuals (**Additional file1: Table S4**). Of these, only 16 variants were
20 identical between the 5 AA risk haplotypes (**Fig. 1a, Additional file1: Figure S2, and Table S5 and**
21 **S6**) and only one was a non-synonymous coding SNV (rs142986308), defining a p.Arg587Trp
22 substitution and mapping to *CCHCR1* (**Additional file1: Figure S3**). Using pair-wise LD analysis,
23 we further established that a haplotype composed of 16 extracted variants in a 651.7-kb region between
24 *CCHCR1* and *VARS* displayed strong LD despite including segments with a high recombination rate
25 (**Fig. 1b**), implying that the haplotype was likely to be younger and/or has undergone positive selective
26 pressure[16]. These results also suggested that our strategy used for stratification, sequencing, and
27 filtering was effective for discovering risk haplotypes and novel MHC variants associated with AA.

1

2 **Sequencing *CCHCR1* gene and haplotype analysis**

3 To check for further variations in *CCHCR1*, we sequenced all coding exons in all subjects and
4 found 2 additional nonsense and 22 non-synonymous variants, though only SNV rs142986308 was
5 shared between all 5 patients and demonstrated a significant association with AA (OR = 3.41, CI 95%
6 = 1.94-5.99, $P = 3.39 \times 10^{-5}$) (**Additional file1: Table S7**). These two nonsense variants were mapped
7 to the coding regions in alternative transcripts 1 and 2, which are common in Japanese and European
8 Americans, and to the non-coding region in alternative transcript 3 (**Additional file1: Figure S4 and**
9 **Table S7**). SNV rs3130453 gave rise to the shortest of the 3 alternative transcripts previously reported
10 for *CCHCR1*[19] and SNV rs72856718, which has not been reported and is predicted to result in the
11 same alternative transcript (**Additional file1: Figure S4**).

12 To confirm the T allele of variant rs142986308 as an AA susceptibility allele, we estimated
13 haplotypes for 24 SNVs. Haplotype 26 (Hap26) harboring the T allele rs142986308 showed a
14 statistically significant association with AA (OR = 3.41, CI 95% = 1.94-5.99, $P = 3.39 \times 10^{-5}$), and
15 rs142986308 was the only SNV associated with AA in Hap26 (**Table 2**). Thus, rs142986308 was
16 determined to be the primary variant associated with AA.

17 Next, we investigated decay of the risk haplotype by recombination events that have occurred
18 in evolutionary history. Extended haplotype homozygosity (EHH)[20] analysis with 5 of the identified
19 risk alleles (designated as core alleles in **Fig. 1c**) was conducted. The core allele T of rs142986308
20 tagged the largest (93.7 kb) LD block with all values between rs1576 and *D6S2931* showing
21 $EHH=1.00$ (**Fig. 1c**), indicating that Hap26, exclusively shared by the patients, had recently increased
22 in frequency. Adjacent alleles also showed some frequency increase in the patient group, possibly by
23 hitchhiking (**Additional file1: Table S7**). Although the exact selection mechanism operating for the T
24 allele of rs142986308 remains unknown and the overall number of haplotypes analyzed is modest, it
25 is plausible that this allele is the primary target of selection for AA. Hence, we excluded the other
26 nonsense and non-synonymous variants as causal for AA, despite being in strong LD with the causal
27 rs142986308 variant identified by haplotype and EHH analyses.

1 **Evaluation of impact on function of CCHCR1 by allele T of rs142986308**

2 We next examined whether the variant rs142986308 had influence on the function of CCHCR1,
3 which is predicted to contain several coiled-coil domains (**Fig. 2a and Additional file1: Table S7**)
4 [21]. The domain including the AA-associated variant p.Arg587Trp is well conserved in parts across
5 many species (**Fig. 2a and Additional file1: Table S8**). Using structure prediction, the probability of
6 coiled-coil conformation of this domain was notably reduced in only Hap26 harboring the AA-
7 associated T allele of rs142986308 (**Fig. 2b and Additional file1: Figure S5 and S6**). Aromatic
8 substitutions are known to be more disruptive towards coiled-coil domains than alanine, glutamic acid,
9 lysine, leucine, and arginine, which favor coiled-coil domain formation[22], adding weight to our
10 speculation that the variant rs142986308, which substitutes arginine with aromatic tryptophan, does
11 indeed impair coiled-coil conformation of CCHCR1 (**Fig. 2a**).

12 We also performed homology searches using the PDB database and obtained the template
13 structure for partial CCHCR1 protein around the p.Arg587Trp substitution, the crystal structure of the
14 human lamin-B1 coil 2 segment (**Fig. 2c**) [23]. We then generated a partial CCHCR1 structure using
15 the template and evaluated the effect of p.Arg587Trp substitution on protein stability by performing
16 molecular dynamics simulations using a CHARMM force field. The simulation showed that
17 p.Arg587Trp substitution reduced protein stability (mutation energy: 1.03 kcal/mol) (**Fig. 2d**) and
18 changed the CCHCR1 structure around the residue at 587 (**Fig. 2e and f**). This indicates that
19 p.Arg587Trp substitution may alter the protein-protein interaction of CCHCR1 (**Fig. 2c**).

20

21 **Allele-specific genome editing using CRISPR/Cas9 in mice**

22 Next, we functionally evaluated the AA-associated variant by *in vivo* phenocopying
23 p.Arg591Trp in murine *CCHCR1* using allele-specific genome editing with the CRISPR/Cas9 system.
24 Mice were generated with the risk allele concordant with p.Arg587Trp derived from the T allele of
25 rs142986308 in humans (**Fig. 3**), then we established mouse strains with homozygous risk alleles (AA
26 mice). In this study, 15 mice (55.5%) among 27 AA mice displayed patched dorsal hair loss until 10
27 months after birth, thus successfully phenocopying the AA phenotype. The incidence of hair loss was

1 higher in those mice as compared to C3H/HeJ mice, which show spontaneously development of AA
2 with age.[24] Over time, the initial area of hair loss expanded in the majority of the AA mice (**Fig. 4a**),
3 though constant and recovered hair loss was observed in some of those mice (**Additional file1: Figure**
4 **S7 and S8**). The male to female ratio of alopecic AA mice was nearly equal, and their surface displayed
5 black spots, while the hairs appeared to be broken and tapering (**Fig. 4b**), similar to the conditions
6 seen in humans with AA and specific for AA-associated hair loss[25]. Thus, an altered hair shaft
7 structure may be a feature concordant between human AA[26] and the present engineered AA mice.
8 All of the AA mice showed retained hair follicles in the area of hair loss and no signs of lymphocyte
9 infiltration were seen in microscopic observations (**Fig. 4c**), implying involvement of a non-immune
10 mechanism.

11

12 **Microarray analysis of alopecic mice skin biopsies**

13 To investigate the presence of a possible biological function underlying the observed hair loss
14 in the AA mice, we performed gene expression microarray analysis of dorsal and ventral skin biopsies
15 (**Additional file1: Figure S9 and S10**), which resulted in identification of 265 probes (246 genes)
16 with 2-fold or greater up- or down-regulation as compared to wild-type mice. The full list of genes is
17 shown Table S11. Clustering analysis of these probes uncovered a strongly up-regulated gene cluster
18 in the AA mice (**Fig. 5a**), including hair-related genes (**Fig. 5b**). Nearly all of those up-regulated with
19 a greater than 25-fold change value were keratin[27] (n=12) and KRTAP (n=31) genes (**Additional**
20 **file1: Table S9**). Hair keratins and KRTAPs are the major structural components of the hair shaft, and
21 specifically expressed in the medulla, cortex, and cuticle layers of the shaft[28], and interaction
22 between hair keratins and KRTAPs suggest their contribution to its rigidity[29].

23 Other up-regulated genes included peptidyl arginine deaminase type III (*Padi3*), S100 calcium
24 binding protein A3 (*S100A3*), trichohyalin (*Tchh*)[30], and homeobox C13 (*Hoxc13*)[31] (**Additional**
25 **file1: Table S9**). In cuticular cells, S100A3 is a substrate of PADI3, while in inner root sheath (IRS)
26 cells of hair follicles it is a substrate of TCHH[30]. Ca²⁺-dependent modifications of S100A3 and
27 TCHH by PADI3 play important roles in shaping and mechanically strengthening hair with hair

1 keratin[30, 32]. *Hoxc13* is unique among the Hox genes, as it is expressed in the outer root sheath
2 (ORS), matrix, medulla, and IRS of hair follicles in a hair cycle-dependent manner, and it has a role
3 in hair shaft differentiation[31]. Thus, the majority of genes shown to be strongly up-regulated in AA
4 mice are involved in the hair shaft and its formation.

5 Hairless model mice, nude mice lacking forkhead box N1 (*Foxn1*), have a pleiotropic mutation
6 that leads to 2 independent phenotypic effects, which are disturbed development of hair follicles and
7 dysgenesis of the thymus[33]. The *Foxn1* expression has been shown to be regulated by *Hoxc13* in
8 hair follicles[34], and both *Hoxc13*-null and *Hoxc13*-overexpressing transgenic mice were found to be
9 an alopecic phenotype[34]. These three alopecic mice lines display defective hair shafts or aberrant
10 hair cuticles[33, 35, 36]. In the present microarray analysis, *Foxn1* expression in alopecic skin of AA
11 mice was also upregulated by 6.5-fold as compared to the wild-type mice, though this gene was ruled
12 out by the filtering criteria employed (**Additional file1: Figure S11**). A comparison between gene
13 profiling results of *Hoxc13*-null[34] and AA mice indicated that the regulated genes were quite similar
14 between them (**Fig. 6** and **Additional file1: Table S10**). Thus, among 113 down-regulated genes with
15 a gene symbol in *Hoxc13*-null mice[34], 89 were upregulated in the AA mice (**Additional file1: Table**
16 **S10**), the majority of which were hair keratins or KRTAPs. Moreover, we observed a significant
17 inverse correlation of fold change values for these genes between these mouse lines (**Fig. 6**), despite
18 differences in strain and mutant genes. Therefore, *Cchcr1* may be involved in the hair shaft
19 differentiation regulatory network controlled by *Hoxc13*, *Foxn1*, keratins, and KRTAPs, though the
20 regulation was in opposite directions.

21 Enrichment analysis of these 246 genes using the Database for Annotation, Visualization and
22 Integrated Discovery (DAVID) tool[37] showed Gene Ontology[38] (GO) terms relating to only
23 keratin and the intermediate filament, but not to any immunological pathways (**Additional file1: Table**
24 **S11**).

25 To confirm the microarray results, we performed quantitative PCR (qPCR) analysis of 7 of the
26 up-regulated genes and *Cchcr1* using a *comparative C_T method*[39]. Those results confirmed
27 significant differences between dorsal skin biopsies from areas of hair loss in the AA and wild-type

1 mice (**Fig. 7a and Additional file1: Figure S12**). All AA mice showed highly concordant rank orders
2 of expression levels of the examined genes (**Fig. 7b and Additional file1: Figure S13**), implicating
3 involvement of the regulatory network in hair shaft differentiation[34].

4 5 **Hair shaft and follicles**

6 Finally, we evaluated the localization of CCHCR1 using immunostaining and electron
7 microscopy to delineate potential mechanisms underlying the observed hair loss in our mouse model.
8 Immunostaining revealed CCHCR1 to be located in the mid-to-upper hair shaft but not in the cell and
9 hair shaft within hair bulb (**Fig. 8a-i**), suggesting that CCHCR1 is a structural component, similar to
10 keratin, of the hair shaft. CCHCR1 was co-localized with the hair cortex and also found to be localized
11 in the hair medulla (**Fig. 8b, e and h**). Expression of this protein in the skin of AA mice was stronger
12 as compared to that of the wild type (**Fig. 8a-i**), supporting our findings of expression analysis using
13 RNAs (**Fig. 5a, b**). Hair keratins are known to be organized in bundles with structures dominated by
14 alfa-helical coiled-coils[40]. Using low-resolution immunostaining, we were not able to identify any
15 difference between the AA and wild-type mice. However, high-resolution scanning electron
16 microscopy (SEM) results demonstrated hair abnormalities, including turbulence in cuticle formation
17 and flattened shafts in alopecic AA mice (**Fig. 8p-x**). Alopecic AA mice displayed aberrant hair not
18 only in areas of hair loss (**Fig. 8x-t**), but also in normal areas (**Fig. 5s-u**). Moreover, AA mice that did
19 not yet show hair loss also displayed aberrant hair (**Fig. 8p-r**), suggesting that all AA mice are affected
20 by abnormal hair keratinization. These aberrations are similar to hair cuticle aberrations observed in
21 AA patients[41], as well as in *Foxn1*-null, *Hoxc13*-null, and *Hoxc13*-overexpression mice[33, 35, 36].

22 23 **Discussion**

24 Our finding that *CCHCR* is an AA-associated locus in the HLA class I region differs from that
25 of a previous genome-wide association study, which suggested that *HLA-DR* in the HLA class II region
26 is a key driver of AA aetiology[13]. One possible explanation for this seeming contradiction is that
27 our study may not have been sufficiently powered to detect the weak association with AA within the

1 HLA-class II region, though the microsatellites utilized covered the entire class II region as well (**Fig.**
2 **1a**). On the other hand, consistent with our finding, studies performed in Chinese using HLA genes as
3 markers suggested that the locus associated with AA maps to the class I rather than the class II
4 region[42-44]. Moreover, variant rs142986308 is very rare in Caucasians (**Additional file1: Table S8**)
5 and has undergone positive selection (**Fig. 1c**), at least in Japanese, suggesting population-specific
6 differences regarding AA risk haplotypes. This is the first study to show that an HLA class I allele,
7 rs142986308, can be functionally linked to the AA phenotype, which has not been demonstrated for
8 any other variant, including *HLA-DR*.

9 Previous studies have provided evidence of a relationship between *CCHCR1* and hair keratin-
10 related genes. For example, a risk haplotype (*CCHCR1*WWCC*) was previously implicated to be
11 involved in psoriasis in Europeans[21, 45]. Transgenic mice with the risk haplotype appeared normal,
12 though it was shown that overexpression of *CCHCR1* affected keratinocyte proliferation[46], and
13 hyperproliferation of keratinocytes is a hallmark for psoriasis. However, keratin-related genes showed
14 altered expression in mice at risk and most of the genes with significantly lower expression were those
15 encoding hair keratins or KRTAPs[47]. These observations suggested that expression of these genes
16 is dependent on the *CCHCR1* haplotype.

17 AA shows clinical heterogeneity[48], thus findings from the lymphocyte infiltration model as
18 well as the AA mice are not able to fully explain the pathogenesis. Those obtained with AA mice may
19 be more suitable to explain aberrant hair shaft results, such as black dots (cadaverized hairs), tapering
20 hairs, and broken hairs, which are specific AA-associated hair loss conditions[25]. Therefore,
21 investigation of the relationship between autoimmunity response and aberrant keratinization in hair
22 follicles may lead to a better understanding of AA pathogenesis. Conversely, each model may have an
23 independent pathway, leading to AA heterogeneity.

24 The current study has several limitations thus further studies are needed. First, more detailed
25 investigation of immunological function in AA mice must be performed, because the present study did
26 not completely exclude immunological aspects. Second, additional examination of factors that
27 eventually lead to hair loss is needed, though all of the AA mice displayed abnormal hair shafts. In

1 addition, investigation of how the mutation alters the function of the CCHCR1 protein is necessary.
2 Finally, it is important to investigate hairs and hair follicles obtained from patients with and without
3 the risk allele in order to fully understand the functions of CCHCR1 in human hair.

4 In summary, we identified a novel AA susceptibility variant (T allele of rs142986308,
5 p.Arg587Trp) in the human MHC class I region and established an alopecic mouse model. Findings
6 of functional analyses indicate that CCHCR1 is a novel component in hair shafts and its variation
7 alters the coiled-coil formation of CCHCR1, resulting in abnormalities of hair shafts and cuticles,
8 along with up-regulation of hair keratin, KRTAPs, and other relevant genes, ultimately leading to AA
9 pathogenesis, shown by fragile and impaired hair as well as hair loss. In addition, our findings suggest
10 an alternative pathway for that pathogenesis based on aberrant keratinization. Together with the
11 present engineered mice, our results provide a valuable resource for continued research of AA
12 pathogenesis and development of potential future treatments.

13

14 **Conclusions**

15 This study genetically identifies an AA susceptibility variant by association and sequencing analysis
16 within the MHC region, and functionally confirms that the variant involves hair abnormalities using
17 engineered mice by CRISPR/Cas9 for allele-specific genome editing.

18

19 **Materials and methods**

20 **Patients and controls**

21 Upon approval of the experimental procedures from the relevant ethical committees of Juntendo
22 University and Tokai University, we obtained informed consent from all unrelated AA and healthy
23 individuals prior to collection of DNA samples. A total of 171 individuals affected with AA (MAA-
24 multiple alopecia areata: 115, AT-alopecia totalis: 18, AU-alopecia universalis: 38) and 560 unrelated
25 individuals of Japanese origin participated in this study. All cases were diagnosed and treated at a
26 Juntendo University Hospital in Japan. DNA was extracted using a QIAamp DNA blood kit (QIAGEN,
27 Hilden, Germany) under standardized conditions to prevent variations in DNA quality. For additional

1 quality control, we used 0.8% agarose gel electrophoresis to check for DNA degradation and/or RNA
2 contamination, and performed optical density measurements to check for protein contamination. The
3 final DNA concentration was determined with 3 successive measurements using a PicoGreen
4 fluorescence assay (Molecular Probes, Thermo Fisher Scientific, Rockford, IL, USA).

5

6 **Microsatellite genotyping.**

7 We selected 22 microsatellites spanning 2.44 Mbp (from *HLA-E* to *PSMB9* gene) in the HLA
8 region (**Additional file1: Table S12**) harbouring the HLA class I, II, class III regions, and genotyped
9 all patients and control subjects. Forward primers of the primer sets to amplify microsatellites were
10 labeled by 5' fluorescent FAM. Oligonucleotides were obtained from Greiner Bio-one. PCR and
11 fragment analyses were performed with capillary electrophoresis using an Applied Biosystems 3730
12 Genetic Analyzer. Allele assignment was determined with GeneMapper Software (Thermo Fisher
13 Scientific) and conducted as previously described[49]. Fragment sizes were assigned to allele names
14 in the corresponding microsatellites. In the *MICA* locus, 5 *MICA* polymorphisms (*A4*, *A5*, *A5.1*, *A6*,
15 *A9*) were determined based on the number of alanine (GCT) repeats. The *A5.1* allele contained 5 GCT
16 repeats, plus 1 extra guanine nucleotide (GCT)₂G(GCT)₃.

17

18 ***HLA-C* locus genotyping**

19 A LABType[®] SSO typing test produced by ONE LAMBDA (Inc., Canoga Park, CA) was
20 utilized. This product is based on the reverse SSO method for use with a suspension array platform
21 with microspheres as a solid support to immobilize oligonucleotide probes. Target DNA is amplified
22 by PCR, then hybridized to the bead probe array, followed by flow analysis using a LABScan[™] 100
23 flow analyzer (ONE LAMBDA). *HLA-C* locus genotype data from 156 AA patients and 560 controls
24 obtained in our previous study[18] were used and we also genotyped an additional 15 AA cases as part
25 of the present study.

26

27 **Genomic library construction and sequencing**

1 For HLA region capture and sequencing, genomic DNA (2 µg) was sheared to approximately
2 500 bp in size using a Covaris Acoustic Adaptor. Genomic libraries were prepared using a TruSeq
3 DNA Sample Preparation kit, v.2 (Illumina, San Diego, CA, USA) following the manufacturer's
4 instructions, which involved size selection of DNA fragments of 550–650 bp in length on 2% agarose
5 gels. HLA region enrichment was performed with an adaptor-ligated DNA sample library using the
6 SeqCap EZ Choice Library Human MHC Design system (Roche NimbleGen, Madison, WI, USA)
7 [50], according to the manufacturer's instructions. To quantify and verify the genomic libraries, all
8 samples were analyzed with a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) using
9 an Agilent DNA 1000 kit prior to sequencing (**Additional file1: Figure S16**). Sequence analysis was
10 performed with the Illumina Genome Analyzer Iix platform, using a paired-end sequencing protocol
11 (2×100 bp). We sequenced 5 risk and 7 non-risk haplotypes to a mean depth of 249 reads (**Additional**
12 **file1: Table S3**), covering 95.2% (mean) of the 4.97-Mb MHC region (chr6:28477797-33451433,
13 hg19) with at least 10 reads (**Additional file1: Table S3**).

14

15 **Next generation sequencing (NGS) data analysis**

16 Fastx-toolkit, v. 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/index.html), was used for
17 quality control of the sequencing reads. Reads that passed quality control were mapped to the human
18 reference genome (UCSC Genome Browser assembly GRCh37/hg19, <http://genome.ucsc.edu/>) using
19 Burrows-Wheeler Aligner (BWA), v. 0.5.9, with the default parameters[51]. After alignment,
20 Sequence Alignment/Map (SAMtools), v. 0.1.17, was used to convert the .sam to .bam files[52], and
21 potential PCR duplicates were flagged with Picard MarkDuplicates (v. 1.88;
22 <http://picard.sourceforge.net/>). A Genome Analysis Toolkit (GATK, v. 2.2-8) was used to perform local
23 realignment, map quality score recalibration, and variant detection[53]. SNVs and indels were then
24 annotated for functional consequences at the gene and protein sequence levels using ANNOVAR[54].
25 Finally, we manually checked the raw sequencing data using Tablet, a sequence assembly visualization
26 tool[55] (**Additional file1: Figure S2**).

27

1 **Variant discovery and genotyping of *CCHCR1* by Sanger sequencing**

2 Coding exons of *CCHCR1* were sequenced using PCR-based capillary Sanger sequencing.
3 Oligonucleotides were purchased from Greiner Bio-One (**Additional file1: Table S13**). PCR was
4 performed in a reaction volume of 10 μ l containing 5 ng of genomic DNA, 0.2 U of KOD FX Neo
5 (TOYOBO Life Science, Osaka, Japan), 5 μ l of 2 \times PCR Buffer, 2 μ l of dNTP (2 mM each), and 0.2
6 μ M (final concentration) of each of the primers. The thermal cycling profile was as follows: initial
7 denaturation at 94°C for 2 minutes and 35 rounds of amplification at 98°C for 10 seconds, then 59-
8 65°C (depending on primer set; see **Additional file1: Table S13**) for 30 seconds and 68°C for 1 minute.
9 PCR products were purified using an AMPure XP (Beckman Coulter, Fullerton, CA, USA), according
10 to the manufacturer's protocol. Purification and sequencing of the PCR products were carried out using
11 a BigDye Terminator v3.1 Cycle Sequencing kit (Thermo Fisher Scientific) and BigDye XTerminator
12 Purification Kit (Thermo Fisher Scientific), following the manufacturer's instructions. Automated
13 electrophoresis was performed with an ABI PRISM 3730 Genetic Analyzer (Thermo Fisher Scientific).
14 Sequencing data were analyzed using Sequencher (v. 5.1, Gene Codes Corporation, Ann Arbor, MI,
15 USA). Genomic coordinates for all variants were called using the UCSC Genome Browser assembly
16 GRCh37/hg19 (<http://genome.ucsc.edu/>).

17

18 **Statistical analysis**

19 Logistic regression models were used to assess the genetic effects of multi-allelic loci, SNVs,
20 and AA risk haplotypes. Comparisons of genotype and haplotype frequency differences were done by
21 regression analysis for log-additive models[56]. Unadjusted odds ratio (OR) and 95% confidence
22 intervals (95% CI) were calculated. Analysis was carried out using the SNPassoc R library[56]. For
23 these association analyses, we used Bonferroni-corrected values to account for the problem of multiple
24 testing to a threshold P value of 1.98×10^{-04} , after accounting for multiple testing of 252 alleles in 23
25 multi-allelic loci for the first microsatellite analysis. An exact P-value test of Hardy–Weinberg
26 proportion and evaluation of LD (linkage disequilibrium) for multi-allelic loci were simulated by the
27 Markov chain method within Genepop[57]. To evaluate the degree of LD for bi-allelic loci (SNVs)

1 we used Haploview 4.2.[58] To estimate haplotypes for SNVs and multi-allelic loci, we used PHASE
2 v2.1.1.[59] For EHH analysis, haplotype data were generated from 22 nonsynonymous SNVs, 2 stop
3 gain SNVs, and 19 multi-allelic loci with fastPHASE v1.2.[60] Moreover, for this estimation, each
4 multi-allelic locus was regarded as a SNV[15]. Thus, we were able to extract the allele demonstrating
5 LD ($D' \geq 0.5$) for the AA risk allele (rs142986308: T allele) from all alleles in each locus, if such an
6 allele was detected (**Additional file1: Figure S15**), and then merged it with the other alleles. EHH
7 was calculated using the “rehh” package[61]. To estimate the power of this study design to detect
8 associated loci with AA, we performed statistical power calculations using the Genetic Power
9 Calculator web application[62] (<http://pngu.mgh.harvard.edu/~purcell/gpc/>) assuming a type I error
10 of alpha (α) = 0.05. The statistical power for a significance level of $\alpha = 0.05$ in our sample was
11 calculated for the *D6S2930* locus under several assumptions, as follows: a) high risk allele frequency
12 of 0.144 in the general population, b) prevalence for AA of 0.001, c) heterozygote genotype relative
13 risk of 1.76, d) homozygote genotype relative risk of 4.66, and e) a control group of subjects comprised
14 of unselected individuals from the general population not screened for AA. Based on these conditions,
15 the statistical power for the *D6S2930* locus was calculated to be 0.781.

16

17 **Prediction of coiled-coil domain and structural analyses**

18 Multiple sequence alignment was performed with CLUSTALW (<http://www.genome.jp/>). To
19 predict coiled-coil domains within the amino acid sequence of CCHCR1, we used 2 different programs,
20 COILS[63] (http://embnet.vital-it.ch/software/COILS_form.html) and Paircoil2[64, 65]
21 (<http://groups.csail.mit.edu/cb/paircoil2/>), with the following recommended default settings: COILS
22 v2.2 with a window size of 28 and the MTIDK table, and Paircoil2 with a window size of 28 and *P*-
23 score cutoff of 0.025.

24 The amino acid sequence of CCHCR1 (GenBank accession: NP_061925.2) was obtained from
25 the NCBI GenBank database (<https://www.ncbi.nlm.nih.gov>) and used as a target for homology
26 modeling. After PSI-BLAST searches[66] of protein databank (PDB) sequence entries
27 (<http://www.rcsb.org/pdb/>) using the CCHCR1 sequence, the crystal structure of the human lamin-B1

1 coil 2 segment (PDB ID: 3TYY[23]) was selected as the best template for homology modeling. The
2 partial 3TYY structure was optimized and used as a template to generate 50 homology models using
3 the Build Homology Models protocol. Probability Density Function total energy and Discrete
4 Optimized Potential Energy scores were used to select the best model. Protein stability for mutants
5 was calculated using the Calculate Mutation Energy (Stability) protocol. All molecular modeling and
6 simulation were performed with Discovery Studio, v. 4.1, from BIOVIA (Accelrys Inc. San Diego,
7 USA) with the default parameter setting.

8

9 **Genome editing by CRISPR/Cas9 in mouse embryos**

10 To generate mice carrying the p.Arg587Trp disease-associated missense variant, we edited the
11 Cchcr1 codon sequence of amino acid 591 in mice using the following protocol (**Fig.3**). Cas9 mRNA
12 was prepared using a pBGK plasmid, as previously described.[67] The plasmid was linearized with
13 XbaI and used as the template for *in vitro* transcription with an mMESSAGING mMACHINE T7 ULTRA
14 kit (Ambion, Foster City, CA, USA). Single guide RNAs (sgRNAs) (guide1:
15 GCTGTGTCAGCTCCTGACGG[AGG], guide2: GGAAGCTGCCAGCCTCCGTC[AGG]) were
16 designed using CRISPR design (CRISPR.mit.edu). The templates for sgRNA synthesis were PCR
17 amplified with primer sets (5'-
18 TAATACGACTCACTATAGGGCTGTGTCAGCTCCTGACGGGTTTTAGAGCTAGAAATAGCA
19 AG-3' / 5'-AAAAAAGCACCGACTCGG-3' for sgRNA1, 5'-
20 TAATACGACTCACTATAGGGGAAGCTGCCAGCCTCCGTCGTTTTAGAGCTAGAAATAGCA
21 AG-3' / 5'-AAAAAAGCACCGACTCGG-3' for sgRNA2) using pUC57-sgRNA vector (Addgene
22 number: #51132) as a template. Then, 400 ng of gel-purified PCR products were subjected to RNA
23 synthesis with a MEGAshortscript T7 Kit (Ambion), according to the manufacturer's instructions.[68]
24 Both Cas9 mRNA and sgRNA were purified with a MEGAclear kit (Ambion) and filtered by passing
25 through an Ultrafree-MC filter (HV; 0.45 µm pore size; Millipore, Billerica, MA) to avoid clogging
26 during microinjection. The single-stranded oligodeoxynucleotide (ssODN) (5'-
27 CAGCAGTTGGAGGCAGCACGTCGGGGCCAGCAGGAGAGCACGGAGGAAGCTGCCAGC

1 CTCtGgCAGGAGCTGACACAGCAGCAGGAAATCTACGGGCAAGGTGTGGGGGCGTGGCG
2 GTGTGTG-3') was synthesized by IDT (Coralville, IA, USA). C57BL/6N strain mouse zygotes were
3 obtained using *in vitro* fertilization. One-cell stage fertilized mouse embryos were injected with 10
4 ng/ μ l of Cas9 mRNA, 10 ng/ μ l of sgRNA, and 20 ng/ μ l of ssODN. Injected eggs were cultured
5 overnight in KSOM medium and the resulting two-cell embryos were transferred into the oviducts of
6 pseudo-pregnant ICR females, as previously described[68].

7

8 **Mice**

9 C57BL/6N mice were purchased from CLEA (Shizuoka, Japan) and maintained under specific
10 pathogen-free conditions. Wild-type mice (8- to 10-month-old females and males) were used as control
11 group and/or for calibration. This study used 27 mice (8- to 12-month-old females and males) carried
12 the mutation generating amino-acid substitution. Randomization and blinding tests were not
13 performed in this study. All animal procedures were done according to protocols approved by the
14 Institutional Animal Care and Use Committee of Tokai University.

15

16 **Genotyping allele of target locus for generated mice with alkaline lysis method**

17 Mouse tissue obtained by ear punch was added to 180 μ l of a 50-mM NaOH solution and
18 incubated at 95°C for 10 minutes. The lysate for PCR was obtained by neutralizing with 20 μ l of 1 M
19 Tris-HCl (pH 8.0) and centrifugation. PCR was performed in a reaction volume of 10 μ l containing 1
20 μ l of lysate, 0.2 U of KOD FX Neo (TOYOBO), 5 μ l of 2 \times PCR buffer, 2 μ l of dNTP (2 mM each),
21 and 0.2 μ M (final concentration) of each primer. Forward (AGCTGAGTGCCACCTGAT) and
22 reverse (TGTGTCTCAGTGCTGCCTTC) primers were used for sequencing (Greiner Bio-one). The
23 thermal cycling profile was as follows: initial denaturation at 94°C for 2 minutes, then 35 rounds of
24 amplification at 98°C for 10 seconds and 60°C for 25 seconds. The protocol for sequencing was
25 performed as previously described in as above in 'Variant discovery and genotyping of *CCHCR1* by
26 Sanger sequencing'.

27

1 **RNA isolation**

2 Total RNA was isolated from sections of mouse skin using ISOGEN (Nippon Gene, Tokyo,
3 Japan), according to the manufacturer's protocol, and treated twice with TURBO DNase (Ambion) to
4 eliminate contaminating DNA. RNA was quantified using a NanoDrop 2000 (Thermo Fisher
5 Scientific) and the quality of the extracted RNA was evaluated with a Bioanalyzer 2100 (Agilent
6 Technologies).

7

8 **Microarray analysis**

9 Fluorescent cRNA synthesis derived from skin RNAs was performed using a Low RNA Input
10 Linear Amplification kit (Agilent Technology) and subjected to DNA microarray analysis with single-
11 color microarray-based gene-expression analysis (SurePrint G3 Mouse GE, v. 2.0, 8x60 K, Agilent
12 Technology). All procedures were performed according to the manufacturer's instructions. Data from
13 samples that passed the QC parameters were subjected to 75th percentile normalization and analyzed
14 using Genespring GX (version 12, Agilent Technologies). Gene ontology enrichment analysis was
15 performed with DAVID 6.8 (<https://david.ncifcrf.gov/>) using 3 functional database (Cellular
16 Component, Biological Process and Molecular Function). Significance values were calculated
17 between the AA and wild-type mice using Spearman's rank correlation coefficient (paired, two-tailed).
18

19 **Quantitative PCR**

20 cDNAs were synthesized using 1 µg of total RNA in a 20-µl total volume using SuperScript®
21 VILO™ MasterMix (Thermo Fisher Scientific) and random hexamers. Quantitative PCR was
22 performed using a StepOnePlus™ Real-Time PCR System, TaqMan® Universal Master Mix, and
23 TaqMan gene expression assay (Thermo Fisher Scientific), according to the manufacturer's protocol.
24 The primer-probe sets were as follows: Mm00652053_g1 (*Krt34*), Mm02345064_m1 (*Krt73*),
25 Mm04208593_s1 (*Krtap3-3*), Mm04336701_s1 (*Krtap16-1*), Mm00478075_m1 (*Padi3*),
26 Mm01214103_g1 (*Sl00a3*), Mm00461542_m1 (*Cchcr1*), and Mm99999915_g1 (*Gapdh*) (Thermo
27 Fisher Scientific). All PCR reactions were performed in triplicate. Relative quantification of gene

1 expression was performed using the $2^{-\Delta\Delta Ct}$ method[39]. Fold change values were calculated using
2 *Gapdh* as the internal control and a dorsal skin sample from a wild-type mouse was used as a calibrator.
3 Significance values were calculated between the AA and wild-type mice using Student's t-test
4 (unpaired, two-tailed).

5

6 **Immunohistochemistry**

7 Anti-CCHCR1 (rabbit polyclonal) and anti-hair cortex cyokeratin antibodies (mouse
8 monoclonal [AE13]) were obtained from Novus Biologicals (NBP2-29926) (Littleton, CO, USA) and
9 Abcam (ab16113) (Cambridge, MA, USA), respectively. This CCHCR1 antibody is generated from
10 rabbits immunized with a KLH conjugated synthetic peptide between 599-627 amino acids from the
11 central region of human CCHCR1. The homology of the 29 peptides between human and mouse is
12 97% (28/29).

13 Deparaffinized skin sections were boiled in 10 mM of citrate buffer (pH 5.0) for antigen
14 unmasking. Sections were incubated in Blocking One Histo (Nacalai tesque, Kyoto, Japan) for 30
15 minutes at room temperature, then incubated with primary antibodies [anti-CCHCR1 (1:100), anti-
16 hair cortex Cytokeratin (1:200)] in PBS containing 5% Blocking One Histo and 0.05% Triton-X 100
17 overnight at 4°C. Sections were then washed and incubated with secondary antibodies [anti-mouse
18 IgG-Alexa488 (1:500), anti-rabbit IgG-Alexa594 (1:500)] for 2 hours at room temperature. Controls
19 for all immunostaining were simultaneously performed by omitting the primary antibody. Sections
20 were cover-slipped using Vectashield with 4',6-diamidino-2-phenylindole dihydrochloride (DAPI)
21 (Vector Laboratories, Burlingame, CA, USA) for nuclei counterstaining and analyzed with Keyence
22 BZ X-700 (Keyence, Tokyo, Japan).

23

24 **Morphology of hair shafts**

25 Hairs were plucked from mice and analyzed by SEM using a JSM 6510LV (Jeol Co., Tokyo,
26 Japan).

27

1 **Additional files:** Supplementary figures, tables and sequences.

2 **Figure S1.** Evaluation of pair-wise LD between 23 multi-allelic loci in MHC region. **Figure S2.**

3 Evaluation of 16 variants extracted from all variants detected using NGS. **Figure S3** Sanger

4 sequencing confirmation of rs142986308. **Figure S4** Schematic overview of CCHCR1 gene structure

5 and variants with amino acid substitution. **Figure S5** Coiled-coil structure prediction of CCHCR1 in

6 26 haplotypes using COILS, v. 2.2. **Figure S6** Coiled-coil structure prediction of CCHCR1 in 5

7 selected haplotypes using Paircoil2. **Figure S7** Fixed hair loss area in representative AA mouse after

8 onset of hair loss. **Figure S8** Hair loss area recovered in representative AA mouse. **Figure S9** Scatter

9 plot between dorsal and ventral skin areas in same individual. **Figure S10** Scatter plots of comparisons

10 between AA and wild-type skin. **Figure S11** Filtering scheme used for microarray analysis. **Figure**

11 **S12** Analysis of expression in mouse skin using quantitative PCR and a comparative CT method.

12 **Figure S13** Gene expression trends in skin biopsies from AA mouse 01 and 02. **Figure S14** Pair-wise

13 LD between 24 variants with amino acid substitution in *CCHCR1*. **Figure S15** Analysis of pair-wise

14 LD between 24 loci for investigating LD. **Figure S16** Quantification and verification of genomic

15 libraries. **Table S1** Allelic association analysis of 23 loci spanning 2.45 Mb in the MHC region. **Table**

16 **S2** Allelic association analysis of *HLA-C* gene. **Table S3** Overview of sequencing output by NGS.

17 **Table S4** Filtering of variants identified by MHC region sequencing. **Table S5** Heterozygous variants

18 identical in individuals with risk haplotype. **Table S6** Overview of 16 variants identical in 5 individuals

19 with risk haplotype. **Table S7** SNV discovery and allelic association with *CCHCR1*. **Table S8**

20 Homology of amino acid sequence to Hap26 in various species. **Table S9** List of 265 probes showing

21 ≥ 2 -fold up- or down-regulation in AA mice as compared to wild type. **Table S10** Inverse correlation

22 of expression fold change values between Hoxc13- and AA mice. **Table S11** Enrichment analysis by

23 DAVID in 246 genes showing ≥ 2 -fold change in gene expression. **Table S12** Microsatellites and the

24 primers. **Table S13** Primer sets for PCR direct sequencing in all exons of *CCHCR1* gene.

25 **Supplementary sequences** Amino acid sequences were generated from haplotype nucleotide

26 sequences (Table 2) as transcript variant3 (NM_019052) (Additional file1: Figure S4).

27

1 **List of abbreviations**

2 Alopecia areata (AA), the major histocompatibility complex (MHC), the coiled-coil alpha-helical rod
3 protein 1 (*CCHCR1*), keratin-associated proteins (KRTAPs), the human leukocyte antigen (HLA),
4 linkage disequilibrium (LD), extended haplotype homozygosity (EHH), peptidyl arginine deaminase
5 type III (*Padi3*), S100 calcium binding protein A3 (*S100A3*), trichohyalin (*Tchh*), homeobox C13
6 (*Hoxc13*), inner root sheath (IRS), outer root sheath (ORS), forkhead box N1 (*Foxn1*), the Database
7 for Annotation, Visualization and Integrated Discovery (DAVID), and quantitative PCR (qPCR)

8

9 **Declarations**

10 **Ethics approval and consent to participate**

11 Upon approval of the experimental procedures from the relevant ethical committees of Juntendo
12 University (reference number:2013097) and Tokai University (reference number: 131-07), we
13 obtained informed consent from all unrelated AA and healthy individuals prior to collection of DNA
14 samples. All animal procedures were done according to protocols approved by the Institutional Animal
15 Care and Use Committee of Tokai University.

16

17 **Consent for publication**

18 Not applicable.

19

20 **Availability of data and materials**

21 Complete DNA microarray data set is deposited at the Gene Expression Omnibus (GEO) database
22 under GSE100630 (<https://www.ncbi.nlm.nih.gov/geo/info/linking.html>). To whom correspondence
23 and material requests should be addressed: 143 Shimokasuya, Isehara, Kanagawa, 259-1193, Japan
24 Tel: +81 463931121; Fax: +81 463964137; Email: oka246@is.icc.u-tokai.ac.jp

25

26 **Competing interests**

27 The authors declare that they have no competing interests.

1 **Funding**

2 This work was supported by JSPS KAKENHI (grant number JP16K10177). S. B. was supported by
3 the NIHR UCLH Biomedical Research Centre (BRC84/CN/SB/5984).

4

5 **Authors' contributions**

6 A.Oka and S.I. conceived the project. A.Oka, M.O., and S.I. designed the research. A.T., E.K., and S.I.
7 were involved in sample collection and clinical interpretation. A.Oka, K.H., S.S., N.M., T.H., M.K.,
8 H.M., M.O., Y.H., A.Otomo, S.H., and T.M. performed laboratory experiments. A.Oka, M.T.U., S.N.,
9 M.T., T.K., and S.M. contributed to the data analysis and statistical support. A.Oka, M.O., A. Otomo,
10 S.H., H.I., S.B., and S.I. wrote the manuscript with contributions from all coauthors.

11

12 **Acknowledgments**

13 We express our gratefulness to Wang Ting, Hisako Kawada Masayuki Tanaka, and Hideki Hayashi of
14 the Support Center for Medical Research and Education, Tokai University.

15

16 **References**

- 17 1. Wasserman D, Guzman-Sanchez DA, Scott K, McMichael A. Alopecia areata. *Int J Dermatol.*
18 2007;46:121-31.
- 19 2. Dubois M, Baumstarck-Barrau K, Gaudy-Marqueste C, Richard MA, Loundou A, Auquier P, et
20 al. Quality of Life Group of the French Society of Dermatology. Quality of life in alopecia areata: A
21 study of 60 cases. *J Invest Dermatol.* 2010;130:2830-33.
- 22 3. Sinclair RD. Alopecia areata and suicide of children. *Med J Aust.* 2014;200:145.
- 23 4. Safavi K. Prevalence of alopecia areata in the first national health and nutrition examination
24 survey. *Arch Dermatol.* 1992;128:702.

- 1 5. Jackow C, Puffer N, Hordinsky M, Nelson J, Tarrand J, Duvic M. Alopecia areata and
2 cytomegalovirus infection in twins: Genes versus environment? *J Am Acad Dermatol.* 1998;38:418-
3 25.
- 4 6. Yang S, Yang J, Liu JB, Wang HY, Yang Q, Gao M, et al. The genetic epidemiology of alopecia
5 areata in china. *Br J Dermatol.* 2004;151:16-23.
- 6 7. Tan E, Tay YK, Goh CL, Chin Giam Y. The pattern and profile of alopecia areata in singapore--a
7 study of 219 asians. *Int J Dermatol.* 2002;41:748-53.
- 8 8. van der Steen P, Traupe H, Happle R, Boezeman J, Strater R, Hamm H. The genetic risk for
9 alopecia areata in first degree relatives of severely affected patients. an estimate. *Acta Derm*
10 *Venerol.* 1992;72:373-5.
- 11 9. Shellow WV, Edwards JE, Koo JY. Profile of alopecia areata: A questionnaire analysis of patient
12 and family. *Int J Dermatol.* 1992;31:186-9.
- 13 10. Xiao FL, Yang S, Liu JB, He PP, Yang J, Cui Y, et al. The epidemiology of childhood alopecia
14 areata in china: A study of 226 patients. *Pediatr Dermatol.* 2006;23:13-8.
- 15 11. Xing L, Dai Z, Jabbari A, Cerise JE, Higgins CA, Gong W, et al. Alopecia areata is driven by
16 cytotoxic T lymphocytes and is reversed by JAK inhibition. *Nat Med.* 2014;20:1043-9.
- 17 12. Petukhova L, Duvic M, Hordinsky M, Norris D, Price V, Shimomura Y, et al. Genome-wide
18 association study in alopecia areata implicates both innate and adaptive immunity. *Nature.*
19 2010;466:113-7.
- 20 13. Betz RC, Petukhova L, Ripke S, Huang H, Menelaou A, Redler S, et al. Genome-wide meta-
21 analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. *Nat*
22 *Commun.* 2015;6:5966.
- 23 14. Forstbauer LM, Brockschmidt FF, Moskvina V, Herold C, Redler S, Herzog A, et al. Genome-
24 wide pooling approach identifies SPATA5 as a new susceptibility locus for alopecia areata. *Eur J*
25 *Hum Genet.* 2012;20:326-32.

- 1 15. Kawashima M, Ohashi J, Nishida N, Tokunaga K. Evolutionary analysis of classical HLA class I
2 and II genes suggests that recent positive selection acted on DPB1*04:01 in japanese population.
3 PLoS One. 2012;7:e46806.
- 4 16. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, et al. A high-resolution
5 HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat
6 Genet. 2006;38:1166-72.
- 7 17. Gulcher J. Microsatellite markers for linkage and association studies. Cold Spring Harb Protoc.
8 2012;2012:425-32.
- 9 18. Haida Y, Ikeda S, Takagi A, Komiyama E, Mabuchi T, Ozawa A, et al. Association analysis of
10 the HLA-C gene in japanese alopecia areata. Immunogenetics. 2013;65:553-7.
- 11 19. Tervaniemi MH, Siitonen HA, Soderhall C, Minhas G, Vuola J, Tiala I, et al. Centrosomal
12 localization of the psoriasis candidate gene product, CCHCR1, supports a role in cytoskeletal
13 organization. PLoS One. 2012;7:e49920.
- 14 20. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, et al. Detecting recent
15 positive selection in the human genome from haplotype structure. Nature. 2002;419:832-7.
- 16 21. Asumalahti K, Veal C, Laitinen T, Suomela S, Allen M, Elomaa O, et al. Coding haplotype
17 analysis supports HCR as the putative susceptibility gene for psoriasis at the MHC PSORS1 locus.
18 Hum Mol Genet. 2002;11:589-97.
- 19 22. Gromiha MM, Parry DA. Characteristic features of amino acid residues in coiled-coil protein
20 structures. Biophys Chem. 2004;111:95-103.
- 21 23. Ruan J, Xu C, Bian C, Lam R, Wang JP, Kania J, et al. Crystal structures of the coil 2B fragment
22 and the globular tail domain of human lamin B1. FEBS Lett. 2012;586:314-8.
- 23 24. McElwee KJ, Hoffmann R. Alopecia areata - animal models. Clin Exp Dermatol. 2002;27:410-7.
- 24 25. Inui S, Nakajima T, Nakagawa K, Itami S. Clinical significance of dermoscopy in alopecia
25 areata: Analysis of 300 cases. Int J Dermatol. 2008;47:688-93.
- 26 26. Miteva M, Tosti A. Hair and scalp dermoscopy. J Am Acad Dermatol. 2012;67:1040-8.

- 1 27. Langbein L, Rogers MA, Praetzel-Wunder S, Helmke B, Schirmacher P, Schweizer J. K25
2 (K25irs1), K26 (K25irs2), K27 (K25irs3), and K28 (K25irs4) represent the type I inner root sheath
3 keratins of the human hair follicle. *J Invest Dermatol.* 2006;126:2377-86.
- 4 28. Fujikawa H, Fujimoto A, Farooq M, Ito M, Shimomura Y. Characterization of the human hair
5 shaft cuticle-specific keratin-associated protein 10 family. *J Invest Dermatol.* 2013;133:2780-2.
- 6 29. Fujikawa H, Fujimoto A, Farooq M, Ito M, Shimomura Y. Characterization of the human hair
7 keratin-associated protein 2 (KRTAP2) gene family. *J Invest Dermatol.* 2012;132:1806-13.
- 8 30. Kizawa K, Takahara H, Unno M, Heizmann CW. S100 and S100 fused-type protein families in
9 epidermal maturation with special focus on S100A3 in mammalian hair cuticles. *Biochimie.*
10 2011;93:2038-47.
- 11 31. Qiu W, Lei M, Tang H, Yan H, Wen X, Zhang W, et al. Hoxc13 is a crucial regulator of murine
12 hair cycle. *Cell Tissue Res.* 2016;364:149-58.
- 13 32. U Basmanav FB, Cau L, Tafazzoli A, Mechin MC, Wolf S, Romano MT, et al. Mutations in
14 three genes encoding proteins involved in hair shaft formation cause uncombable hair syndrome. *Am*
15 *J Hum Genet.* 2016;99:1292-304.
- 16 33. Mecklenburg L, Nakamura M, Sundberg JP, Paus R. The nude mouse skin phenotype: The role
17 of Foxn1 in hair follicle development and cycling. *Exp Mol Pathol.* 2001;71:171-8.
- 18 34. Potter CS, Pruett ND, Kern MJ, Baybo MA, Godwin AR, Potter KA, et al. The nude mutant
19 gene Foxn1 is a HOXC13 regulatory target during hair follicle and nail differentiation. *J Invest*
20 *Dermatol.* 2011;131:828-37.
- 21 35. Godwin AR, Capecchi MR. Hoxc13 mutant mice lack external hair. *Genes Dev.* 1998;12:11-20.
- 22 36. Tkatchenko AV, Visconti RP, Shang L, Papenbrock T, Pruett ND, Ito T, et al. Overexpression of
23 Hoxc13 in differentiating keratinocytes results in downregulation of a novel hair keratin gene cluster
24 and alopecia. *Development.* 2001;128:1547-58.
- 25 37. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the
26 comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37:1-13.

- 1 38. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool
2 for the unification of biology. the gene ontology consortium. *Nat Genet.* 2000;25:25-9.
- 3 39. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat*
4 *Protoc.* 2008;3:1101-8.
- 5 40. Yang FC, Zhang Y, Rheinstadter MC. The structure of people's hair. *PeerJ.* 2014;2:e619.
- 6 41. Peereboom-Wynia JD, Koerten HK, Van Joost T, Stolz E. Scanning electron microscopy
7 comparing exclamation mark hairs in alopecia areata with normal hair fibres, mechanically broken
8 by traction. *Clin Exp Dermatol.* 1989;14:47-50.
- 9 42. Xiao FL, Zhou FS, Liu JB, Yan KL, Cui Y, Gao M, et al. Association of HLA-DQA1 and DQB1
10 alleles with alopecia areata in chinese hans. *Arch Dermatol Res.* 2005;297:201-9.
- 11 43. Xiao FL, Yang S, Yan KL, Cui Y, Liang YH, Zhou FS, et al. Association of HLA class I alleles
12 with alopecia areata in chinese hans. *J Dermatol Sci.* 2006;41:109-19.
- 13 44. Xiao FL, Yang S, Lin GS, Gao M, Cui Y, Yin XY, et al. HLA haplotypic association with
14 different phenotype of alopecia areata in chinese hans. *J Dermatol Sci.* 2007;45:206-9.
- 15 45. Asumalahti K, Laitinen T, Itkonen-Vatjus R, Lokki ML, Suomela S, Snellman E, et al. A
16 candidate gene for psoriasis near HLA-C, HCR (Pg8), is highly polymorphic with a disease-
17 associated susceptibility allele. *Hum Mol Genet.* 2000;9:1533-42.
- 18 46. Tiala I, Wakkinen J, Suomela S, Puolakkainen P, Tammi R, Forsberg S, et al. The PSORS1
19 locus gene CCHCR1 affects keratinocyte proliferation in transgenic mice. *Hum Mol Genet.*
20 2008;17:1043-51.
- 21 47. Elomaa O, Majuri I, Suomela S, Asumalahti K, Jiao H, Mirzaei Z, et al. Transgenic mouse
22 models support HCR as an effector gene in the PSORS1 locus. *Hum Mol Genet.* 2004;13:1551-61.
- 23 48. Dudda-Subramanya R, Alexis AF, Siu K, Sinha AA. Alopecia areata: Genetic complexity
24 underlies clinical heterogeneity. *Eur J Dermatol.* 2007;17:367-74.
- 25 49. Tamiya G, Shinya M, Imanishi T, Ikuta T, Makino S, Okamoto K, et al. Whole genome
26 association study of rheumatoid arthritis using 27 039 microsatellites. *Hum Mol Genet.*
27 2005;14:2305-21.

- 1 50. Cao H, Wu J, Wang Y, Jiang H, Zhang T, Liu X, et al. An integrated tool to study MHC region:
2 Accurate SNV detection and HLA genes typing in human MHC region using targeted high-
3 throughput sequencing. *PLoS One*. 2013;8:e69388.
- 4 51. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform.
5 *Bioinformatics*. 2009;25:1754-60.
- 6 52. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map
7 format and SAMtools. *Bioinformatics*. 2009;25:2078-9.
- 8 53. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome
9 analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.
10 *Genome Res*. 2010;20:1297-303.
- 11 54. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from
12 high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
- 13 55. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, et al. Using tablet for visual
14 exploration of second-generation sequencing data. *Brief Bioinform*. 2013;14:193-202.
- 15 56. de Cid R, Riveira-Munoz E, Zeeuwen PL, Robarge J, Liao W, Dannhauser EN, et al. Deletion of
16 the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat*
17 *Genet*. 2009;41:211-5.
- 18 57. Rousset F. Genepop'007: A complete re-implementation of the genepop software for windows
19 and linux. *Mol Ecol Resour*. 2008;8:103-6.
- 20 58. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of LD and
21 haplotype maps. *Bioinformatics*. 2005;21:263-5.
- 22 59. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from
23 population genotype data. *Am J Hum Genet*. 2003;73:1162-9.
- 24 60. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype
25 data: Applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet*.
26 2006;78:629-44.

- 1 61. Gautier M, Vitalis R. Rehh: An R package to detect footprints of selection in genome-wide SNP
2 data from haplotype structure. *Bioinformatics*. 2012;28:1176-7.
- 3 62. Purcell S, Cherny SS, Sham PC. Genetic power calculator: Design of linkage and association
4 genetic mapping studies of complex traits. *Bioinformatics*. 2003;19:149-50.
- 5 63. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science*.
6 1991;252:1162-4.
- 7 64. McDonnell AV, Jiang T, Keating AE, Berger B. Paircoil2: Improved prediction of coiled coils
8 from sequence. *Bioinformatics*. 2006;22:356-8.
- 9 65. Berger B, Wilson DB, Wolf E, Tonchev T, Milla M, Kim PS. Predicting coiled coils by use of
10 pairwise residue correlations. *Proc Natl Acad Sci U S A*. 1995;92:8259-63.
- 11 66. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and
12 PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res*.
13 1997;25:3389-402.
- 14 67. Harms DW, Quadros RM, Seruggia D, Ohtsuka M, Takahashi G, Montoliu L, et al. Mouse
15 genome editing using the CRISPR/cas system. *Curr Protoc Hum Genet*. 2014;83:15.7.1-27.
- 16 68. Miura H, Gurumurthy CB, Sato T, Sato M, Ohtsuka M. CRISPR/Cas9-based generation of
17 knockdown mice by intronic insertion of artificial microRNA using longer single-stranded DNA. *Sci*
18 *Rep*. 2015;5:12799.

19

20 **Figure legends**

21 **Fig. 1 Identification of AA susceptibility variant by association and sequencing analysis. a**

22 Association analysis and risk haplotype resequencing of the MHC region. Red diamonds indicate P-
23 values ($-\log_{10}$ scale) and locations. Three diamonds refer to haplotypes used for downstream analysis.
24 Blue circles indicate individuals with variants identical between risk haplotype cases and the variant
25 locations. **b** Pair-wise LD between 16 variants identified by NGS using genotype data from 89
26 Japanese individuals obtained with the 1000 Genomes Browser. Upper track shows recombination rate

1 (cM/Mb) estimated from Phase II HapMap data (release 21), middle track the gene map (RefSeq
2 genes) generated with the UCSC Genome Browser, and lower track the pair-wise LD between the 16
3 variants in R-squared. **c** EHH analysis of core alleles at 43 loci displaying LD with the T allele of
4 rs142986308. An estimated 43 loci haplotypes encompassing 24 SNVs of CCHCR1 and 19 multi-
5 allelic loci (2.32Mbp) were used for this investigation. The 7 selected core alleles were as follows:
6 rs142986308 allele T, rs142986308 allele C for the internal control, 4 SNVs that displayed LD with
7 rs142986308 (**Additional file1: Figure S14**), and *HLA-C*04:01* (**Additional file1: Figure S15**) as
8 functional variants.

9

10 **Fig. 2 Prediction of impact on function of CCHCR1 by amino acid substitution.** **a** Multiple amino
11 acid sequence alignment of CCHCR1 showing evolutionarily conserved amino acids. The sequences,
12 except for Hap01 and Hap26, were NP_001009009 (Pan troglodytes), NP_001108422 (Macaca
13 mulatta), XP_532064 (Canis lupus familiaris), NP_001019707 (Bos taurus), NP_666360 (Mus
14 musculus), and NP_001116918 (Xenopus tropicalis). Blue indicates residues that prefer to form
15 coiled-coil domains (Ala, Glu, Lys, Leu, Arg) and red indicates aromatic residues that do not prefer to
16 form coiled-coil domains. Arrow and box indicate the position of substitution of p.Arg587Trp
17 (rs142986308). **b** Coiled-coil structure prediction of CCHCR1 in the AA-associated haplotype and
18 different species using COILS v2.2. The Y axis indicates the probability of coiled-coil conformation
19 and the X axis amino acid residue number. Full amino acid sequences are described in **Additional**
20 **file1: Supplementary sequences**. Multiple amino acid sequence alignments were assigned to the
21 probabilities of coiled-coil conformation for each haplotype and specie. Line brakes correspond to
22 gaps in the multiple alignments. Arrow shows position of p.Arg587Trp (rs142986308). **c** Ribbon and
23 molecular surface representations of partial CCHCR1 structure. Surface models are shown in 2
24 orientations. Residue 587 is represented in red. **d** Effects of Arg587Trp mutation on CCHCR1 protein
25 structure at different pH values. A change of >0.5 cal/mol, which destabilized the structure, was
26 classified as a significant change. Surface maps showing **e** hydrophobicity and **f** electrostatic potential
27 of the wild type and Arg587Trp are presented.

1

2 **Fig. 3 Sequence alignments noted in human and mouse samples around AA-susceptibility**
3 **variant.** Upper portion shows the mouse sequence concordant for Arg587 amino acid in human and
4 mouse sequences aimed for allele-specific genome editing. Lower portion shows electropherograms
5 from sequencing for genotyping the allele.

6

7 **Fig. 4 Phenotypic traits of skin biopsy samples from AA risk allele mice (AA model mice). a**
8 Expansion of hair loss area in representative AA mouse. **b** Morphology of hair loss area in AA mouse.
9 Red arrows show broken hair, black arrows show black spots, white arrows show tapering hair. **c**
10 Microscopic features of hair loss area in representative AA mouse. Paraffin section of skin from
11 representative AA mouse after staining with hematoxylin and eosin.

12

13 **Fig. 5 Microarray analysis of skin biopsy samples from hair loss areas in 2 AA mice and 1 wild-**
14 **type mouse. a** Heat map of 265 probes showing ≥ 2 -fold change in gene expression. The list of genes
15 is shown in Table S11. The cluster displaying high expression in the dorsal hair loss area in both AA
16 mice was defined as the 'core cluster'. **b** Heat map of core cluster genes. The color code depicts
17 KRTAP family (red), keratin family (blue), other hair-related (black), and non hair-related (grey) genes.

18

19 **Fig. 6 Absolute fold change values of 89 concordant genes showing ≥ 2 -fold change in *Hoxc13***
20 **null and AA mice as compared to wild type.** All genes demonstrated down-regulation in *Hoxc13*
21 null mice and up-regulation in AA mice. Correlation coefficients are indicated for each mouse strain.
22 Correlation coefficients and statistical significance in comparisons between *Hoxc13* null and AA mice
23 were determined using Spearman's rho.

24

25 **Fig. 7 Validation of upregulated gene expression for *Cchcr1* and 7 selected genes. a** Mouse skin
26 biopsy samples were subjected to expression analysis by qPCR and a comparative CT method. Bars
27 reflect 95% confidence intervals. Fold change values were normalized to dorsal hair loss in a wild-

1 type mouse as a calibrator, thus the fold change value of the calibrator was always 1. Statistical
2 significance was determined using Student's t test. **b** Gene expression trends in skin biopsy samples
3 from dorsal hair loss areas in AA mice. Correlation coefficients are indicated for each mouse.
4 Statistical significance was determined using Pearson's product-moment correlation between each
5 mouse and AA mouse 01.

6

7 **Fig. 8 Investigations of hair shafts and follicles. a-l** Co-localization of CCHCR1 with hair cortex
8 keratin in follicles from AA and wild-type mouse skin. Paraffin sections were stained with anti-
9 CCHCR1 and anti-hair cortex keratin antibodies, and subjected to fluorescent microscopy. Left panels
10 (**a, e, i**) show CCHCR1 (green). Middle panels (**e, g, k**) show hair cortex keratin (red). Right panels
11 (**c, g, k**) show merged sections with DAPI nuclear staining (blue). Upper panels (**a-d**) show
12 longitudinal section of skin including subcutaneous tissue from AA mouse. Middle panels (**e-h**) show
13 longitudinal section of upper hair follicle from AA mouse. Lower panels (**i-l**) show longitudinal section
14 of upper hair follicle from wild-type mouse. (**m-x**) Scanning electron microscopy (SEM) imaging of
15 hair shafts on dorsal skin was performed. In each panel, hair orientation is shown with the bottom to
16 the left and top to the right. Non-alopecic indicates AA mice that have not yet displayed hair loss (**p-**
17 **r**). Alopecic indicates AA mice that have already displayed hair loss (**s-x**). Alopecic column also
18 indicates hairs from non-hair loss (**s-u**) and hair loss (**v-x**) areas in the same mouse. Data shown in 3
19 columns are from 3 different mice.

20

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Table 1 Haplotype association analysis of 3 loci around *HLA-C* gene.

Haplotype ID	Haplotype			Haplotype frequency		Haplotype association		
	<i>D6S2930</i>	- <i>HLA-C</i>	- <i>D6S2811</i>	Case	Control	OR	(95% CI)	<i>P</i>
MShap01	441	- <i>C*04:01</i>	- 208	0.0614	0.0179	3.78	2.00 - 7.16	6.57 x 10 ⁻⁵
MShap02	433	- <i>C*07:02</i>	- 192	0.0117	0.0009	13.4	1.49 - 121	7.24 x 10 ⁻³
MShap03	437	- <i>C*07:02</i>	- 192	0.0292	0.0125	2.42	1.06 - 5.56	4.37 x 10 ⁻²
MShap04	441	- <i>C*15:02</i>	- 192	0.0175	0.0018	10.2	2.03 - 50.7	1.91 x 10 ⁻³

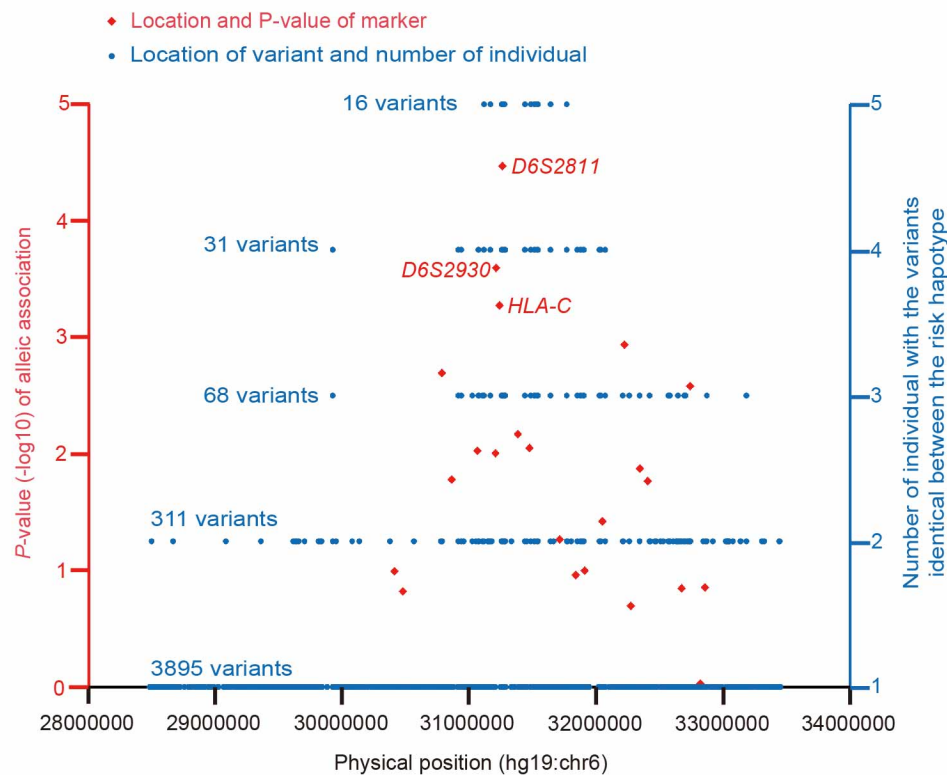
Haplotypes in case subjects with a *P*-value <0.05 and haplotype frequency >0.01.

Table 2 Haplotype analysis of 24 SNVs with amino acid substitution in CCHCR1

Haplotype	SNV*																Frequency		Haplotype association										
	rs72856718	rs3130453	rs130075	rs130065	rs130076	rs130066	rs11540822	rs130067	rs2027937	rs130068	rs130069	SNV01	rs147733073	rs2073720	SNV02	rs130079	rs142986308	SNV03	rs202129359	rs130072	rs130074	rs116969494	rs73397100	rs1576	Case	Control	OR	(95% CI)	P
Hap01	G	G	G	C	C	C	T	A	G	C	G	A	G	A	C	G	C	G	G	G	G	A	C	C	0.1667	0.1223	1.43	1.02 - 2.00	3.97 x 10 ⁻²
Hap02	G	G	A	C	C	C	T	A	G	C	G	A	G	A	C	G	C	G	G	G	G	A	C	C	0.0906	0.1295	0.68	0.45 - 1.01	5.02 x 10 ⁻²
Hap03	G	A	G	C	C	G	T	A	G	C	G	A	G	A	C	G	C	G	G	G	G	A	C	C	0.0029	0.0000	-	- - - -	2.34 x 10 ⁻¹
Hap04	G	A	G	C	C	G	T	C	G	C	G	A	G	A	C	G	C	G	G	G	G	A	C	C	0.2865	0.3170	0.87	0.66 - 1.13	2.86 x 10 ⁻¹
Hap05	T	A	G	C	C	G	T	C	G	C	G	A	G	A	C	G	C	G	G	G	G	A	C	C	0.0058	0.0000	-	- - - -	5.45 x 10 ⁻²
Hap06	G	A	G	C	C	G	T	C	A	T	G	A	G	A	C	G	C	G	G	G	G	A	C	C	0.0439	0.0223	2.06	1.06 - 4.00	3.95 x 10 ⁻²
Hap07	G	A	G	C	C	C	T	A	G	C	A	A	G	A	C	G	C	G	G	G	G	A	C	C	0.0029	0.0045	0.65	0.08 - 5.63	6.85 x 10 ⁻¹
Hap08	G	A	G	C	C	G	T	C	G	C	G	C	G	A	C	G	C	G	G	G	G	A	C	C	0.0029	0.0009	3.29	0.20 - 52.9	4.14 x 10 ⁻¹
Hap09	G	G	G	C	C	C	T	A	G	C	G	A	C	A	C	G	C	G	G	G	G	A	C	C	0.0468	0.0652	0.71	0.41 - 1.23	5.41 x 10 ⁻¹
Hap10	G	G	A	C	C	C	T	A	G	C	G	A	C	A	C	G	C	G	G	G	G	A	C	C	0.0000	0.0009	-	- - - -	1.00
Hap11	G	A	G	C	C	G	T	C	G	C	G	A	G	A	A	G	C	G	G	G	G	A	C	C	0.0117	0.0027	4.45	0.99 - 20.1	5.43 x 10 ⁻²
Hap12	G	A	G	C	C	G	T	C	G	C	G	A	G	A	C	G	C	C	G	G	G	A	C	C	0.0088	0.0018	4.98	0.83 - 30.1	7.94 x 10 ⁻²
Hap13	G	G	G	C	C	C	T	A	G	C	G	A	G	A	C	G	C	G	A	G	G	A	C	C	0.0000	0.0027	-	- - - -	1.00
Hap14	G	G	G	C	C	C	T	A	G	T	G	A	G	G	C	G	C	G	G	A	G	A	C	C	0.0029	0.0000	-	- - - -	2.34 x 10 ⁻¹
Hap15	G	G	G	C	C	C	T	A	G	C	G	A	G	A	C	G	C	G	G	G	T	A	C	C	0.0029	0.0000	-	- - - -	2.34 x 10 ⁻¹
Hap16	G	G	G	C	C	C	T	A	G	T	G	A	G	A	C	G	C	G	G	G	T	A	C	C	0.0585	0.0670	0.86	0.51 - 1.45	7.67 x 10 ⁻¹
Hap17	G	A	G	C	C	C	T	A	G	T	G	A	G	A	C	G	C	G	G	G	T	A	C	C	0.0029	0.0000	-	- - - -	2.34 x 10 ⁻¹
Hap18	G	G	G	C	C	C	T	A	G	C	G	A	C	A	C	G	C	G	G	G	T	A	C	C	0.0029	0.0000	-	- - - -	2.34 x 10 ⁻¹
Hap19	G	G	G	C	C	C	T	A	G	T	G	A	G	G	C	G	C	G	G	G	T	A	C	C	0.0965	0.1241	0.75	0.5 - 1.12	1.51 x 10 ⁻¹
Hap20	G	G	G	C	C	C	T	A	G	T	G	A	C	G	C	G	C	G	G	G	T	A	C	C	0.0000	0.0009	-	- - - -	1.00
Hap21	G	G	G	C	C	C	T	A	G	T	G	A	G	G	C	G	C	G	G	A	T	A	C	C	0.0029	0.0000	-	- - - -	2.34 x 10 ⁻¹
Hap22	G	G	G	C	C	C	T	A	G	C	G	A	G	A	C	G	C	G	G	G	G	T	T	C	0.0789	0.0946	0.82	0.53 - 1.27	3.74 x 10 ⁻¹
Hap23	G	G	G	C	C	C	T	A	G	C	G	A	C	A	C	G	C	G	G	G	G	T	T	C	0.0000	0.0009	-	- - - -	1.00
Hap24	G	A	G	T	T	G	T	A	G	T	G	A	G	A	C	T	C	G	G	G	G	A	C	G	0.0000	0.0054	-	- - - -	3.45 x 10 ⁻¹
Hap25	T	G	G	C	C	C	A	A	G	T	G	A	G	A	C	G	C	G	G	A	G	A	C	G	0.0058	0.0125	0.46	0.10 - 2.05	2.66 x 10 ⁻¹
Hap26	T	G	G	C	C	C	A	A	G	T	G	A	G	A	C	G	T	G	G	A	G	A	C	G	0.0760	0.0250	3.41	1.94 - 5.99	3.39 x 10 ⁻⁵

*All SNVs with amino acid substitution used in analysis of haplotypes. Gray area indicates relationship between rs142986308 and Hap26.

a



b

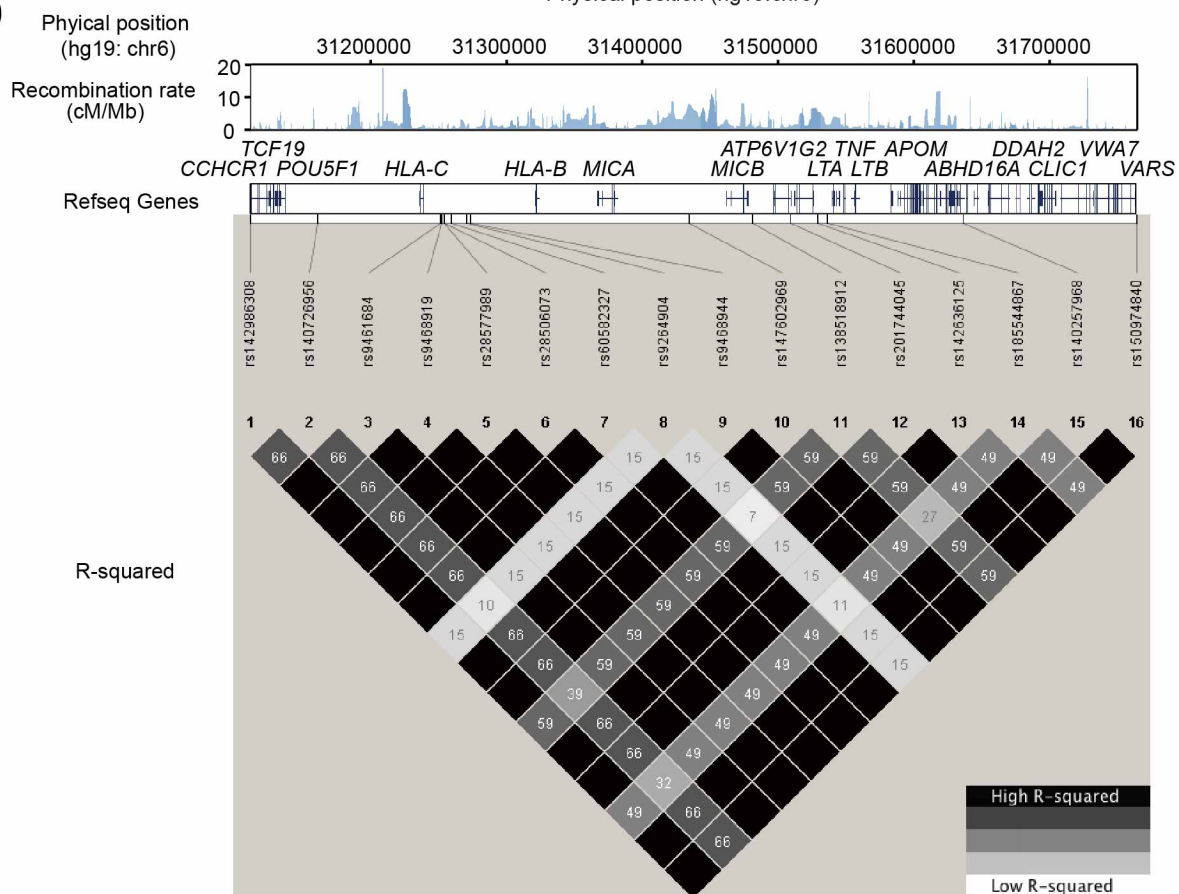


Fig. 1

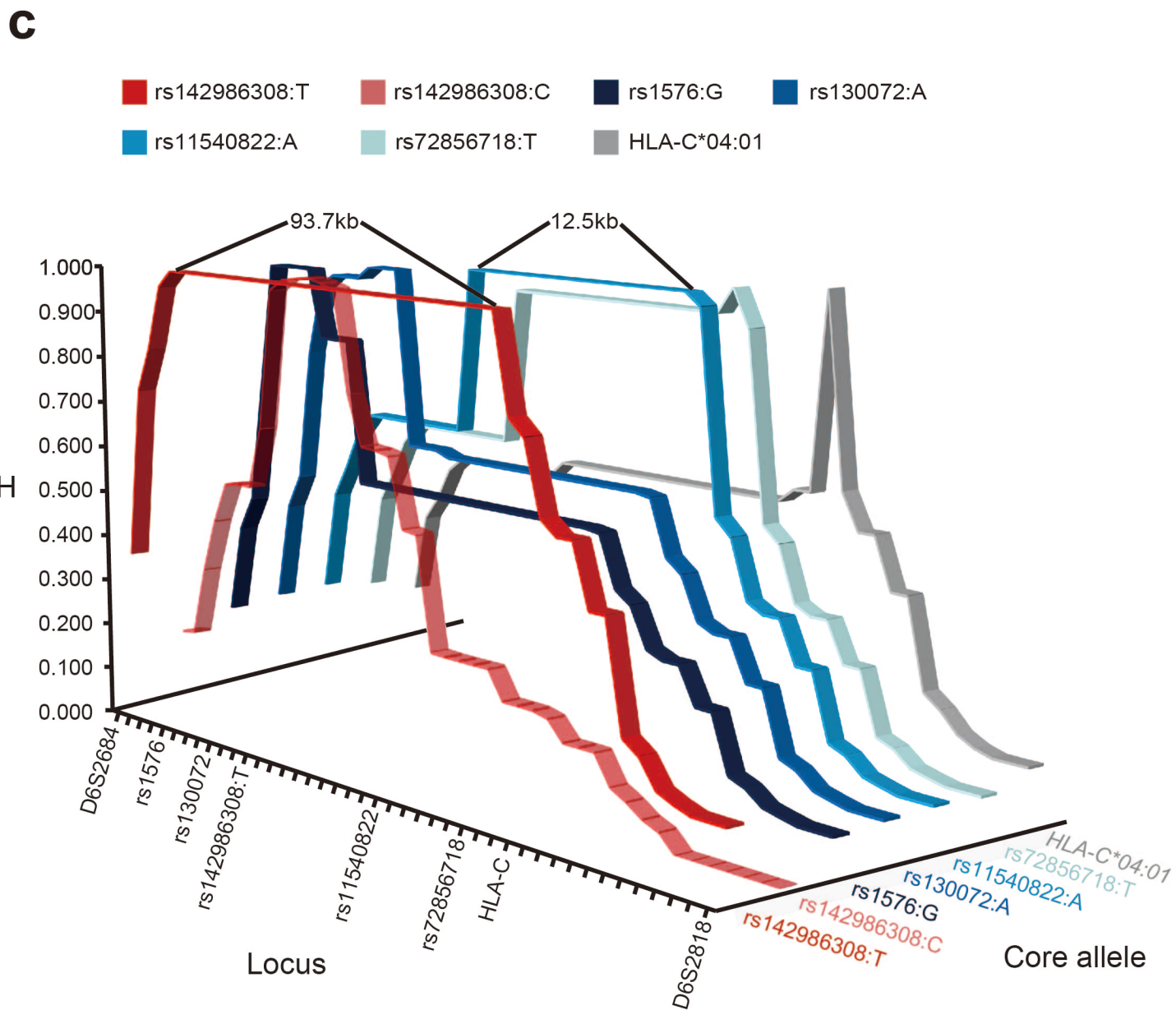


Fig. 1

a

p.Arg587Trp
↓

Hap26 TQESLASLGLQLEVARQGQQESTEEAASLWQELTQQQELYGQALQEKVAEVETRLREQLS
Hap01 TQESLASLGLQLEVARQGQQESTEEAASLRQELTQQQELYGQALQEKVAEVETRLREQLS
Pan troglodytes TQESLASLGLQLEVARQGQQESTEEAASLRQELTQQQELYGQALQEKVAEVETRLREQLS
Macaca mulatta TQESLASLGLQLEVARQGQQESTEEAASLRQELTQQQDLYRQALQEKVAEVETRLREQLS
Canis lupus familiaris TQESLASVGLQLEAARQGQQESTEEAANLRQELTQQQELYGQALQEKVAEVETRLREQLL
Bos taurus TQESLASLGLQLEAARQGQQESTAEAASLRKELTQQQEIYGQALQEKVAEVEGRLREQLS
Mus musculus AQESLASVGGQLEAARRQGQQESTEEAASLRQELTQQQEIYGQALQEKVAEVETRLREQLS
Xenopus tropicalis NEKSRSELKEQLGVTERELQESCENTAELREQLLGTKAEYEKELQRKVNELADHHVQQLA

b

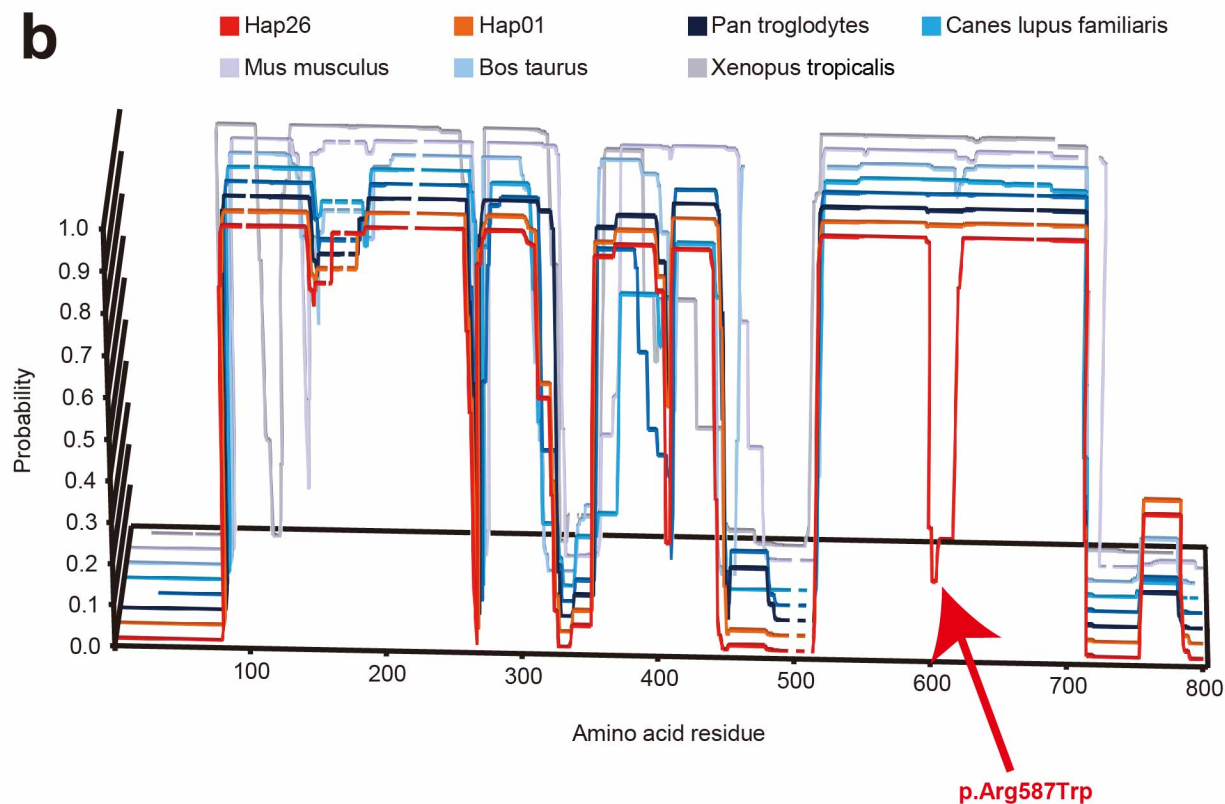
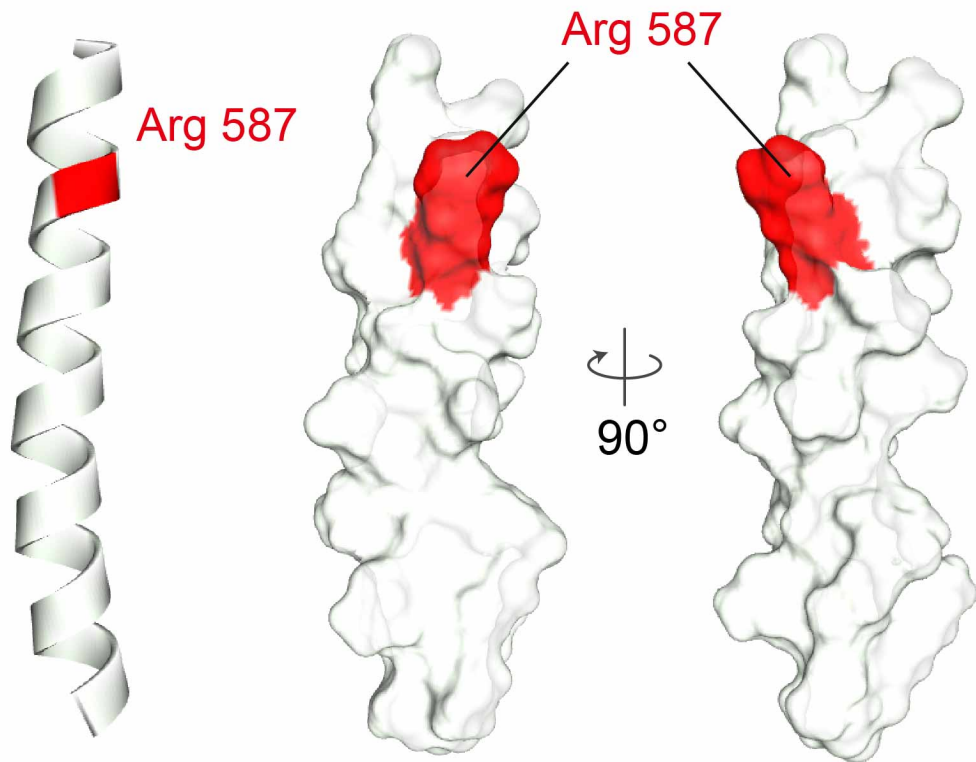


Fig. 2

c



d

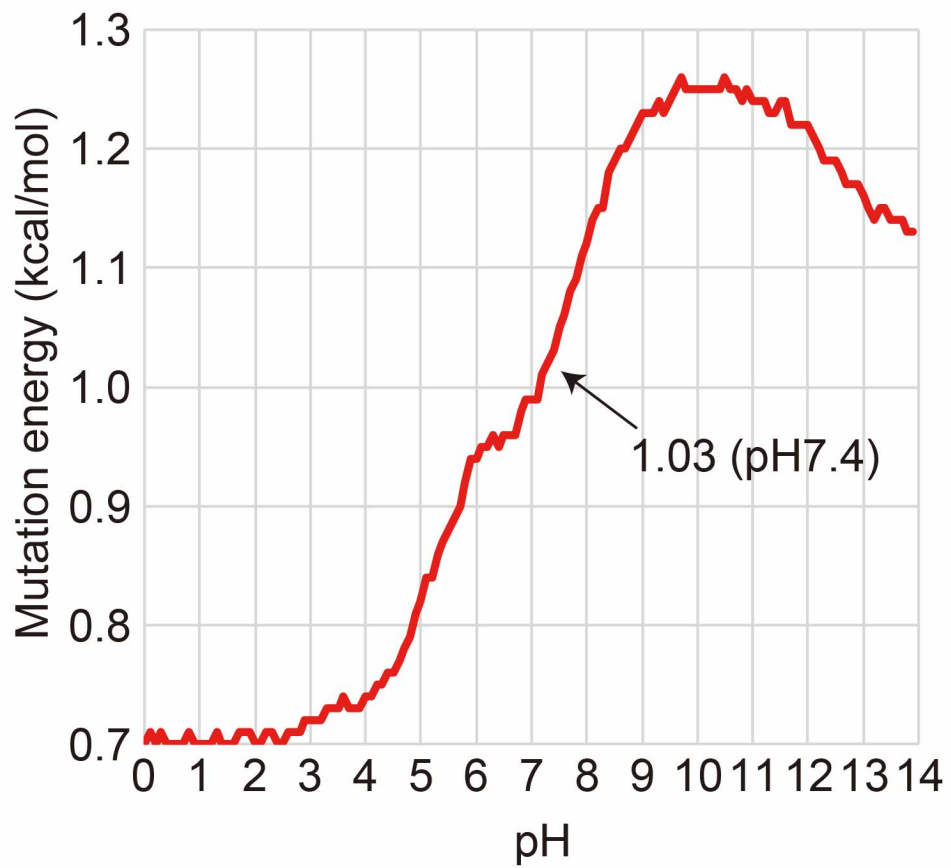


Fig. 2

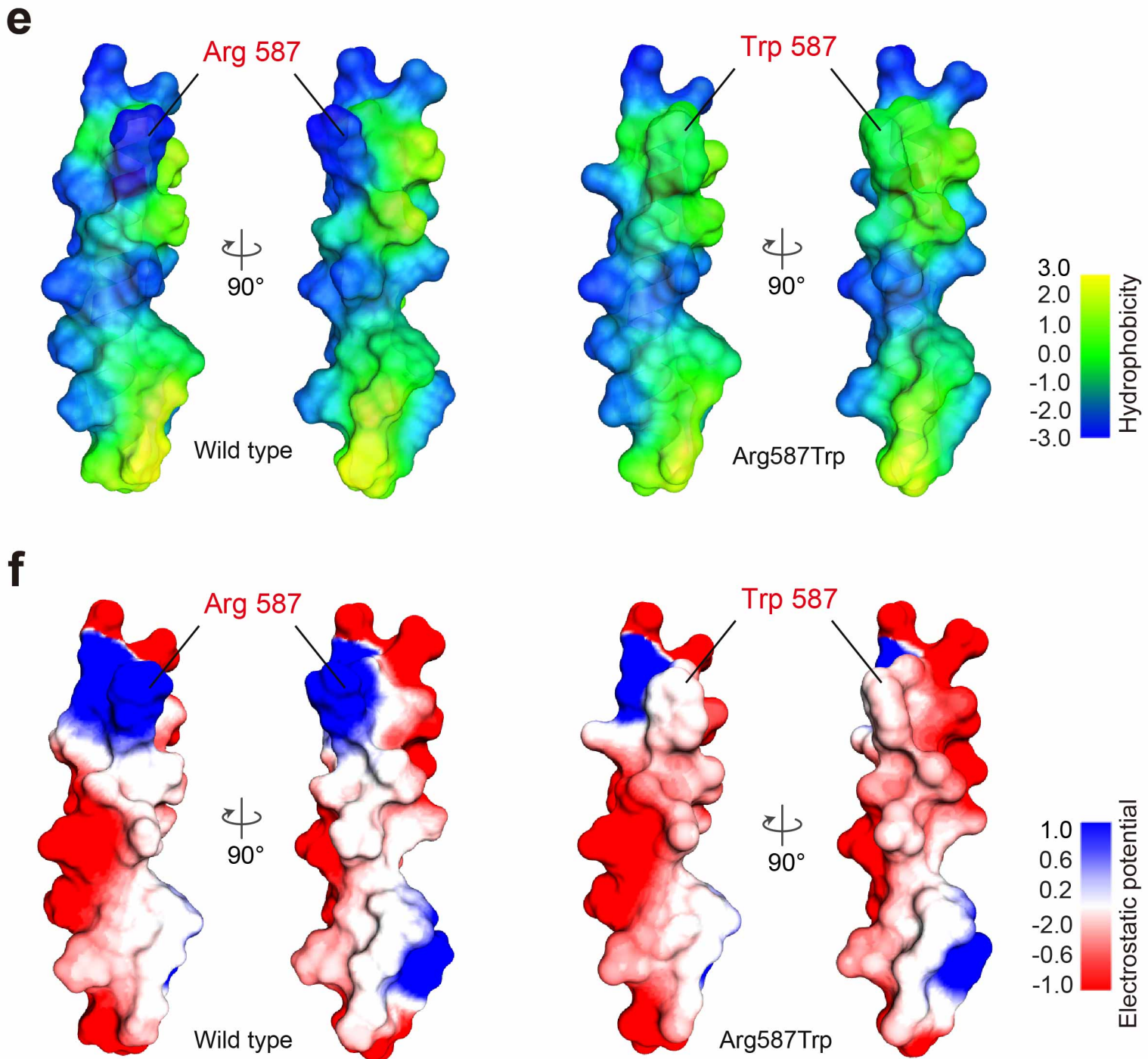


Fig. 2

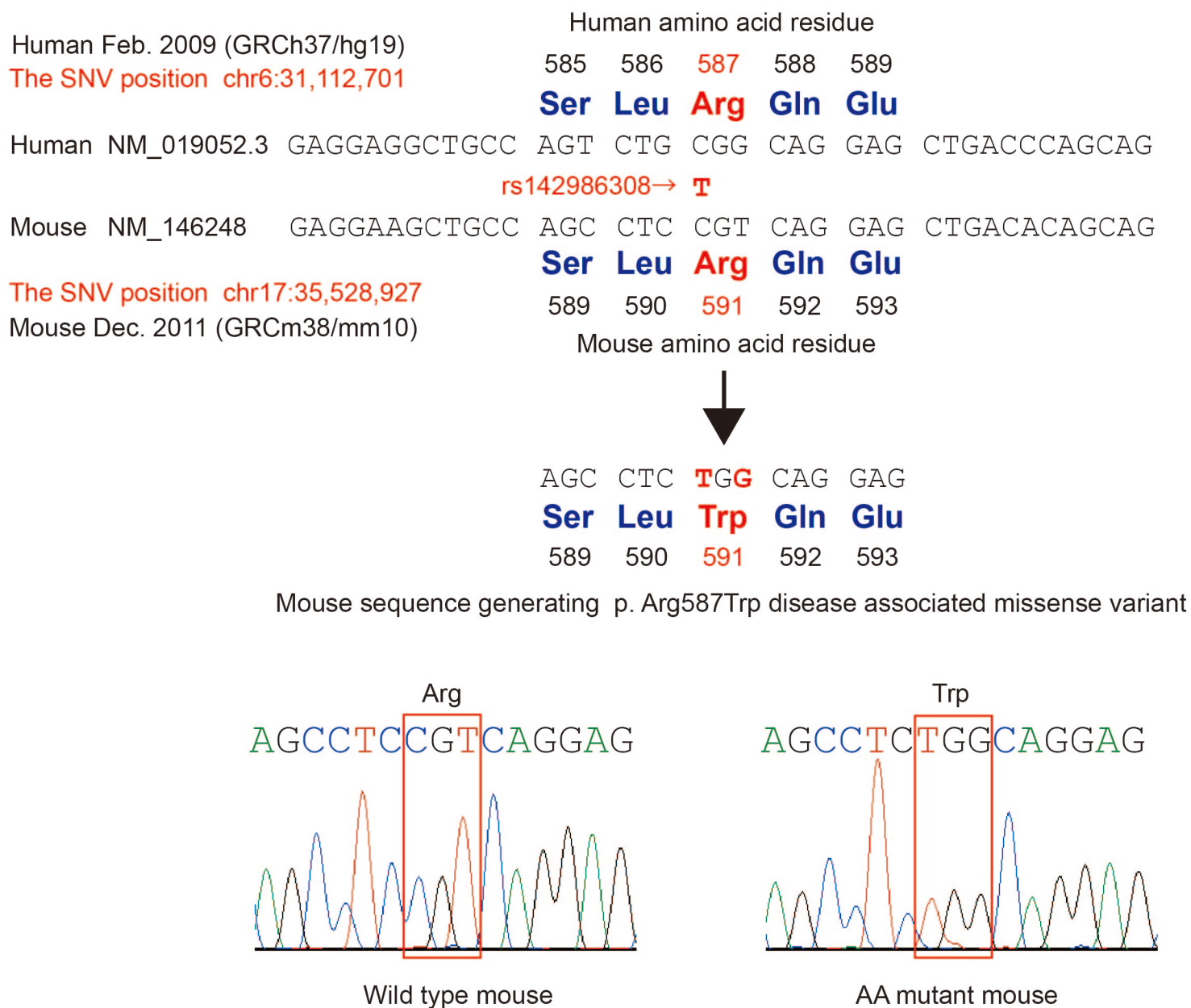


Fig. 3

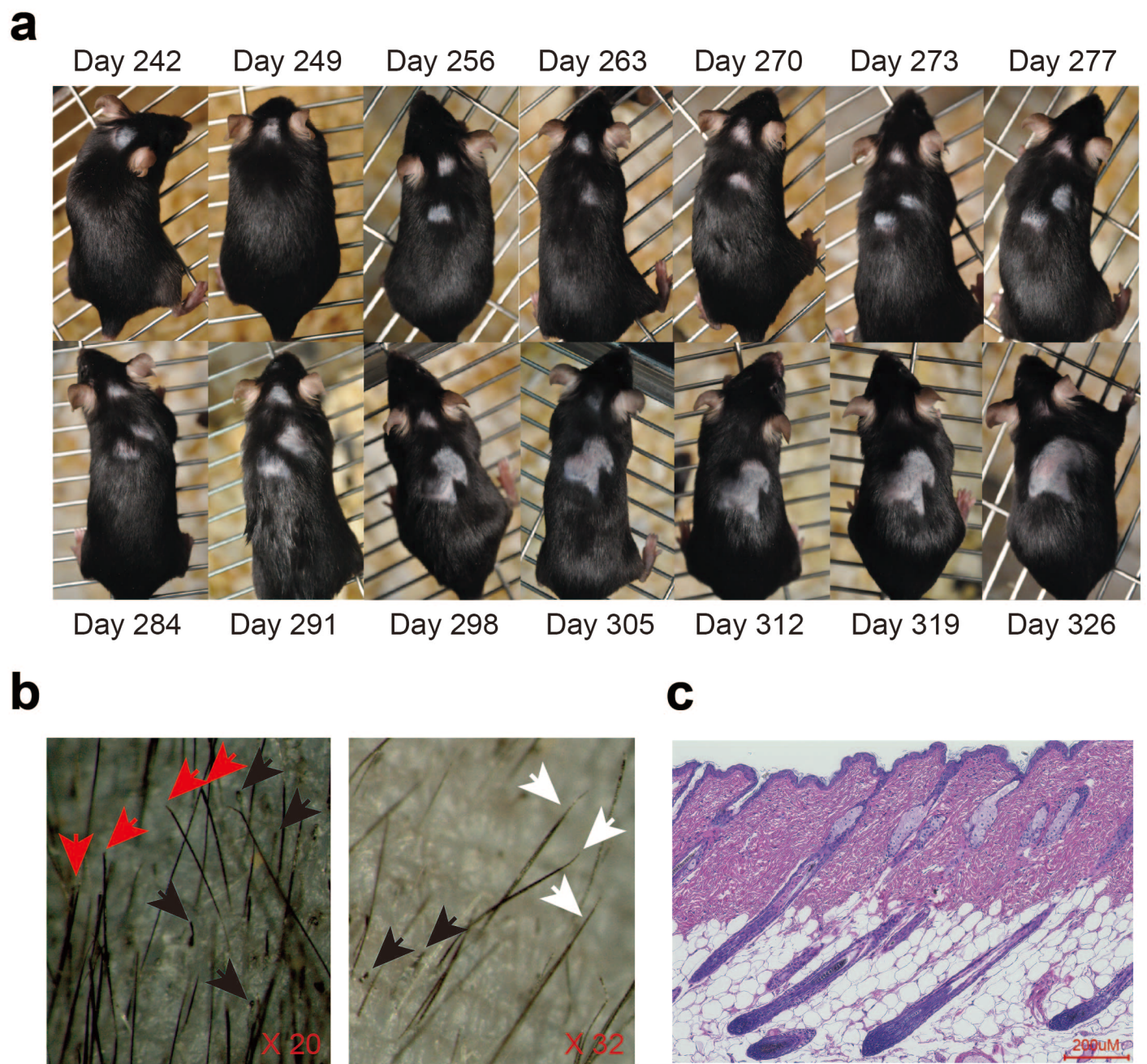


Fig. 4

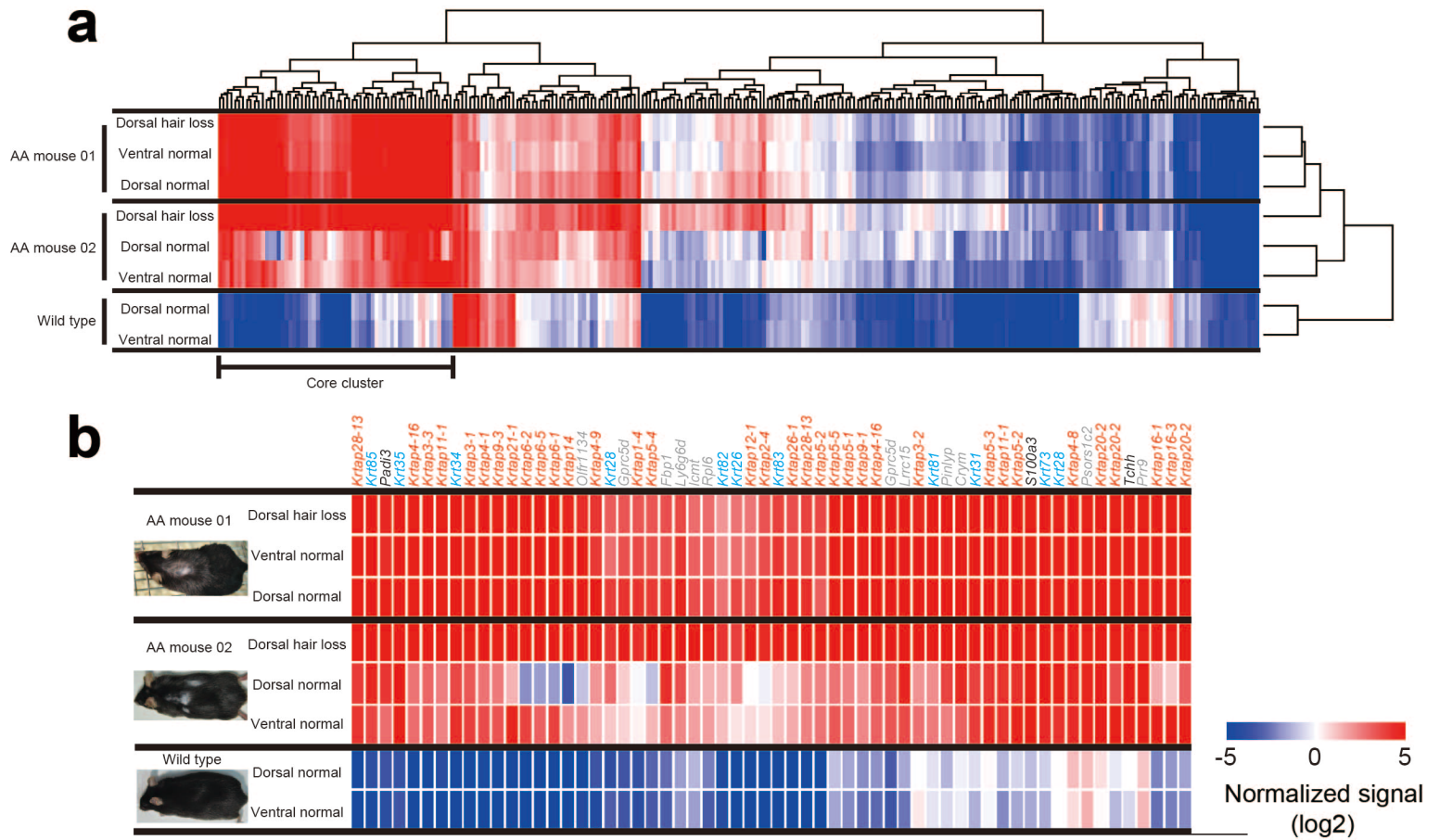


Fig. 5

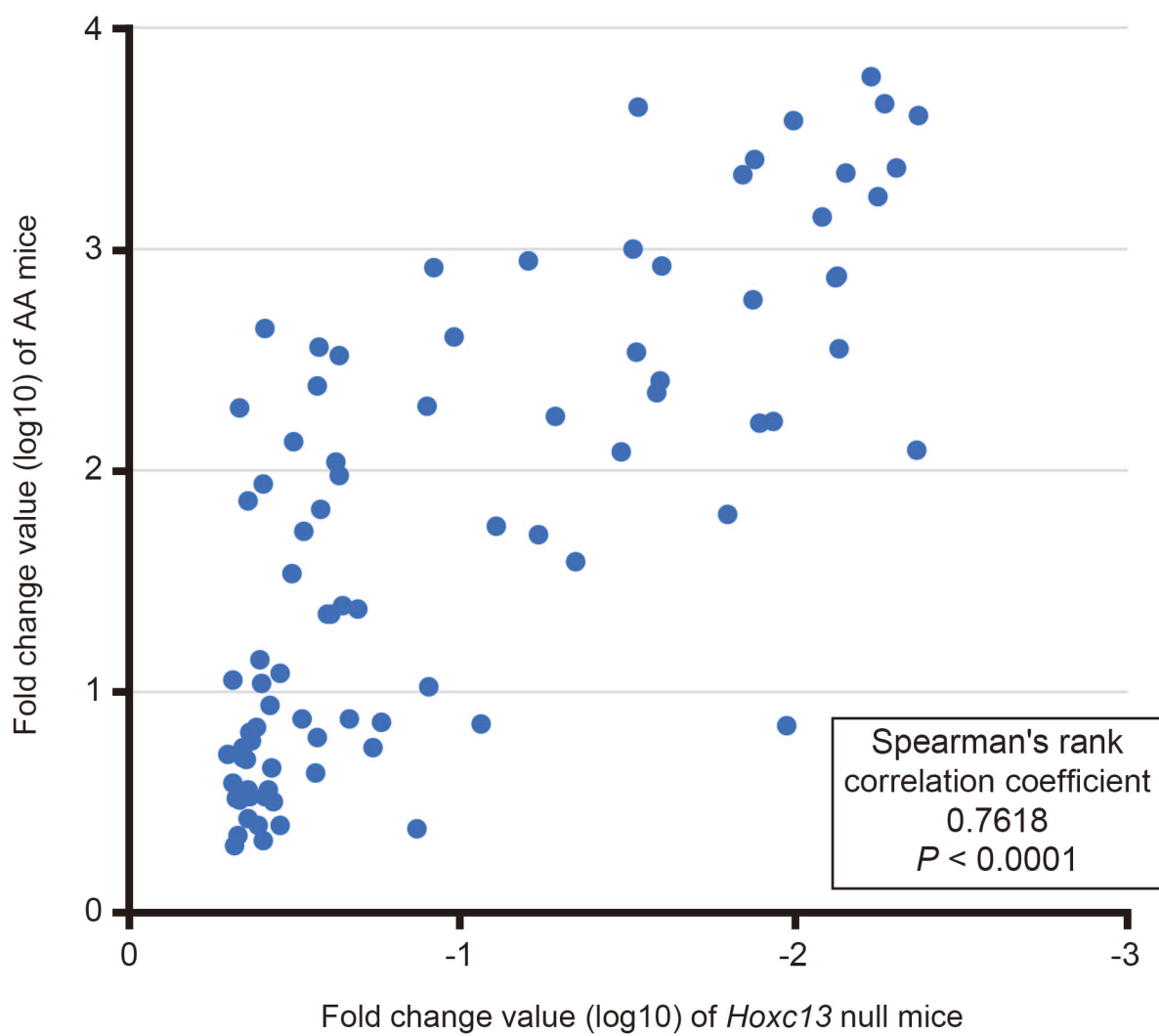


Fig. 6

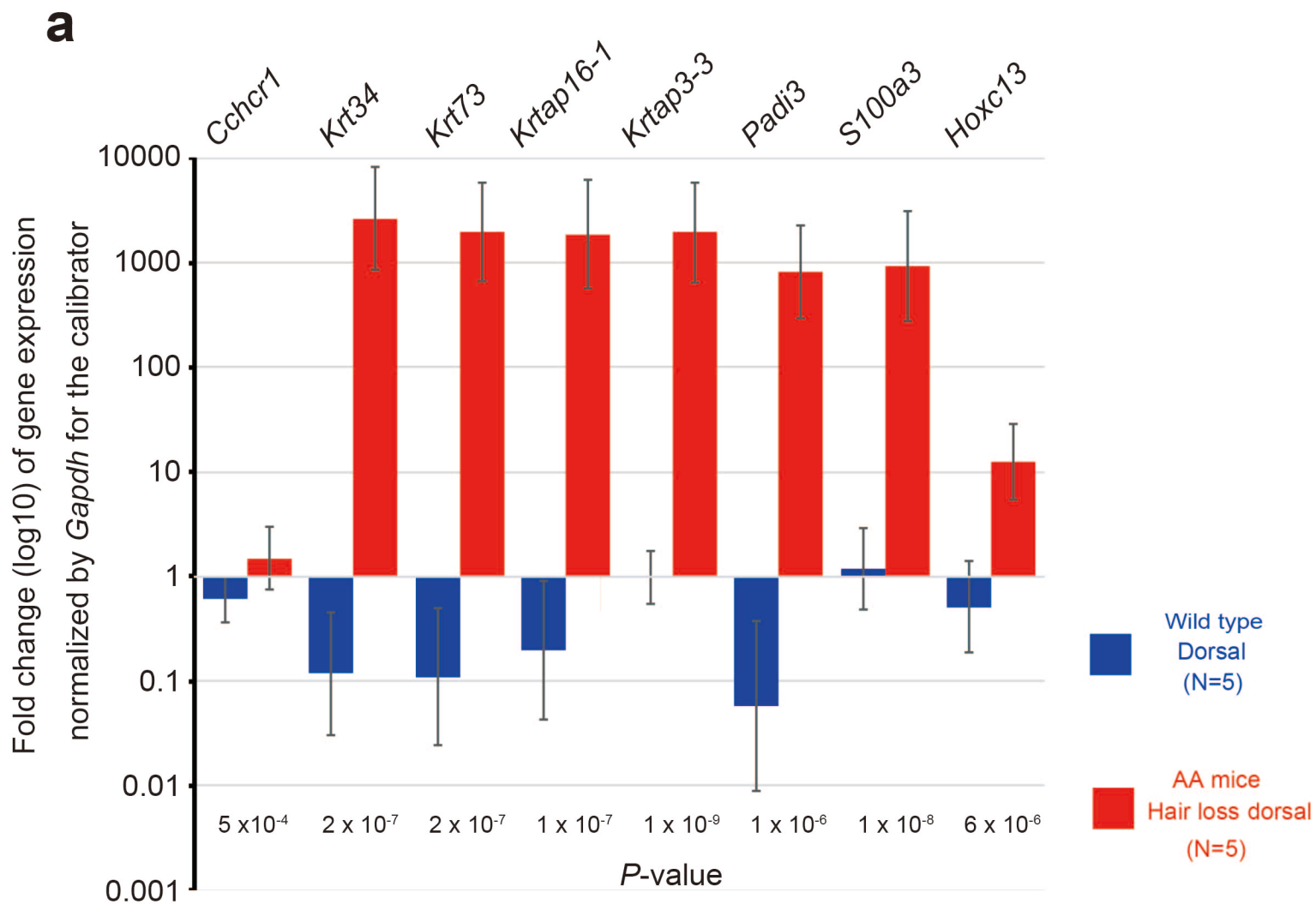


Fig. 7

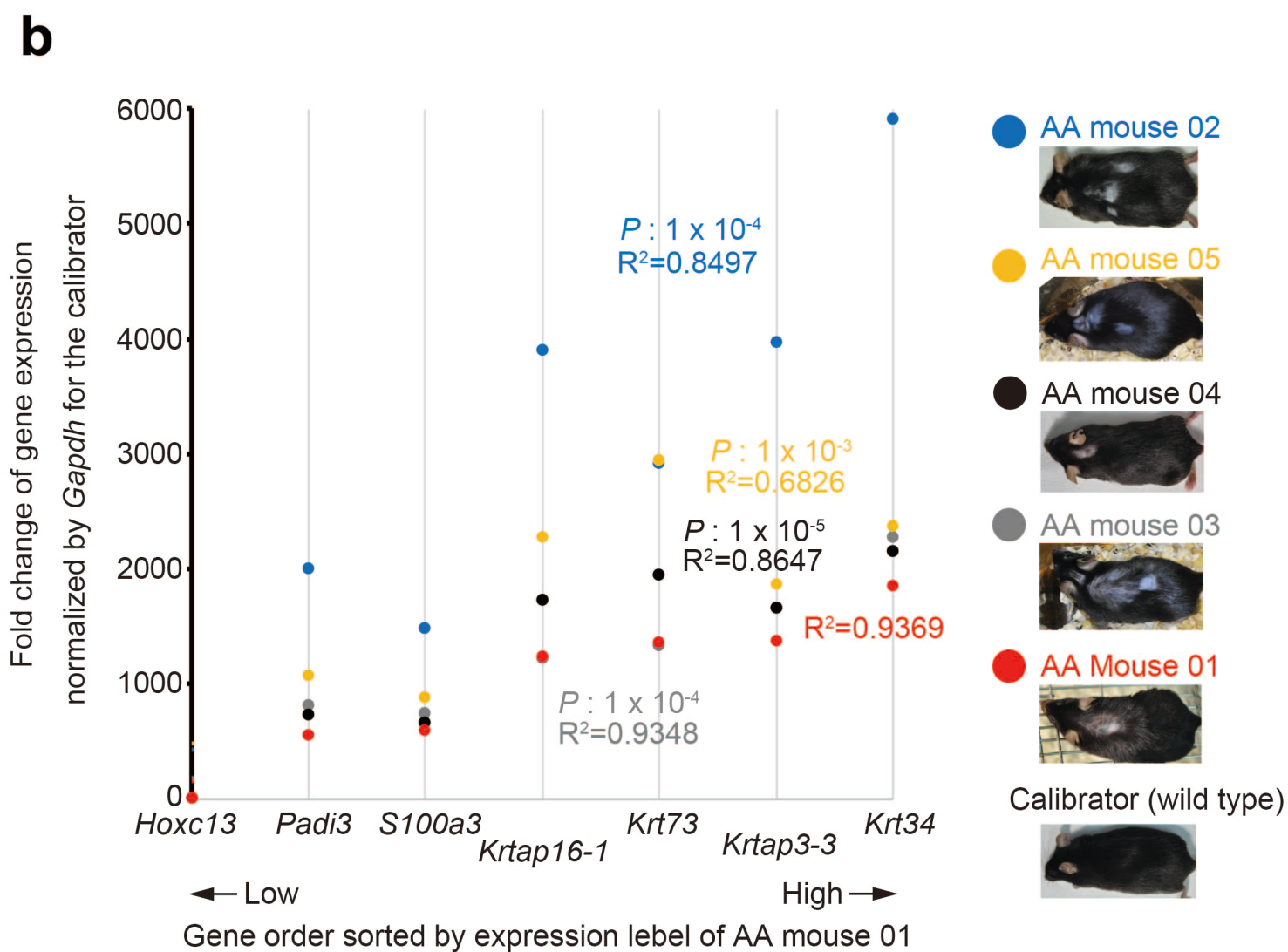


Fig. 7

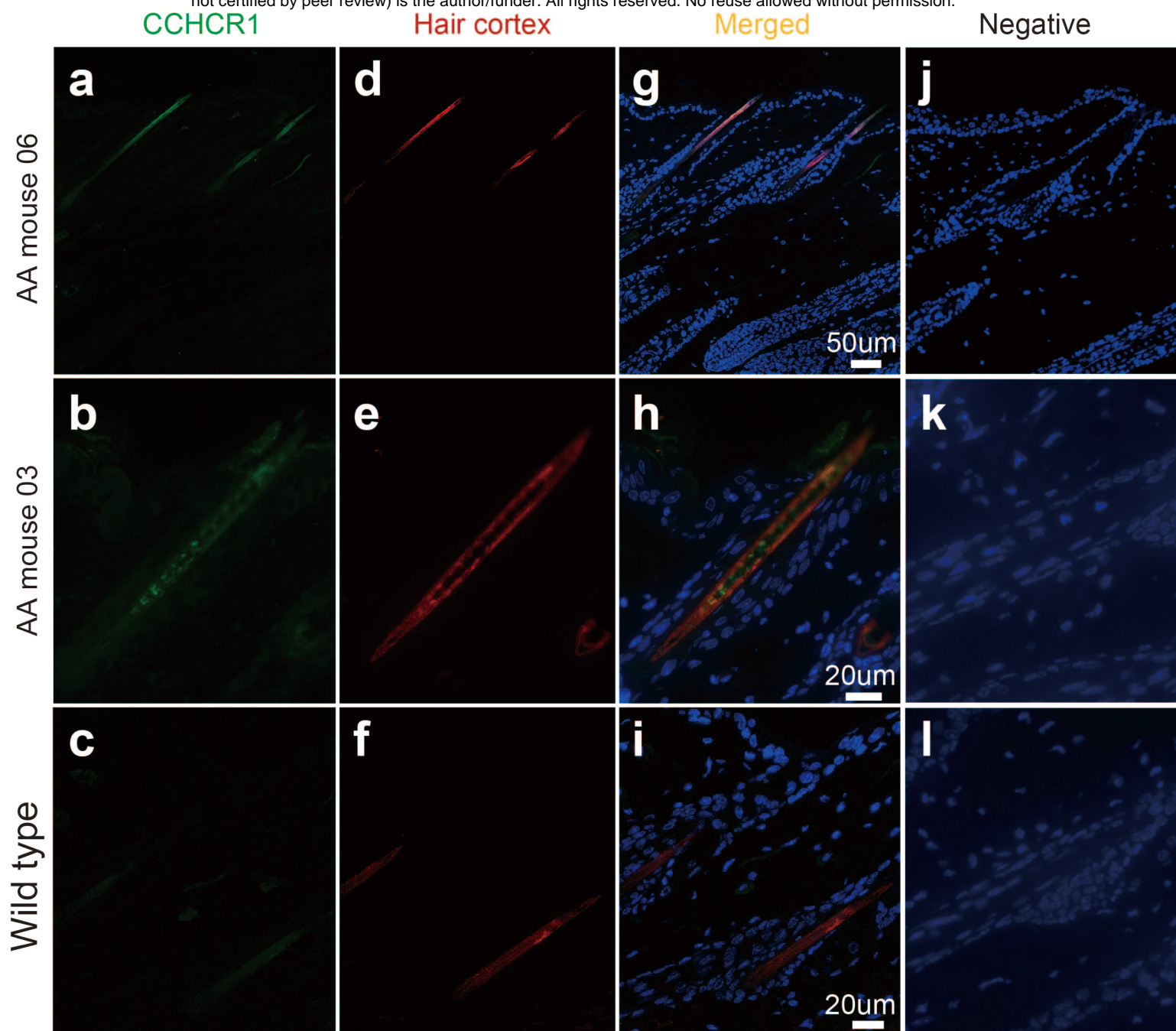


Fig. 8

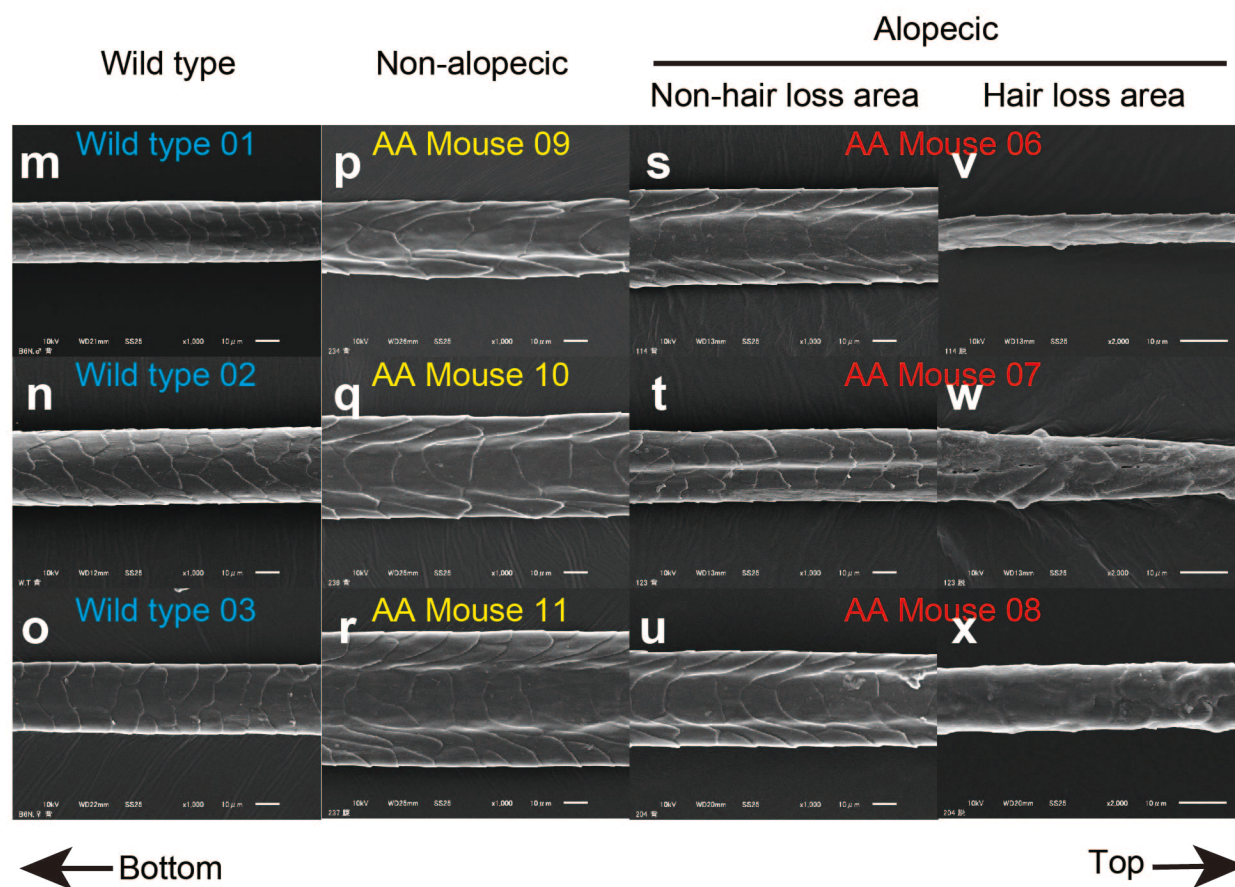


Fig. 8