See related Commentaries on pages vi and xv

Characterization of New Members of the Human Type II Keratin Gene Family and a General Evaluation of the Keratin Gene Domain on Chromosome 12q13.13

Michael A. Rogers,* Lutz Edler,† Hermelita Winter,* Lutz Langbein,‡ Iris Beckmann,* and Jürgen Schweizer*
*Section of Normal and Neoplastic Epidermal Differentiation; †Biostatistics Group; ‡Division of Cell Biology, German Cancer Research Center, Heidelberg, Germany

The recent completion of a reference sequence of the human genome now allows a complete characterization of the type II keratin gene domain on chromosome 12q13.13. This, domain, approximately 780 kb in size, is present on nine bacterial artificial chromosome clones sequenced by the Human Genome Sequencing Project. The type II keratin domain contains 27 keratin genes and eight pseudogenes. Twenty-three of these genes and four pseudogenes have been previously reported. This study describes, in addition to the genomic sequencing of the *K2p* gene and the bioinformatic identification of four keratin pseudogenes, the characterization of cDNA corresponding to three previously undescribed keratin genes *K1b*, *K6l*, and Kb20, as well as cDNA sequences for the previously described keratin genes *hHb2*, *hHb4*, and *K3*. Northern analysis of the new keratins K1b, K6l, K5b, and Kb20 using mRNA of major organs as well as of specific epithelial subtypes shows singular expression of these keratins in skin, hair follicles and, for K5b and Kb20, in tongue, respectively. In addition, the obvious discrepancies between the current reference sequence of the human genome and the previously described gene/cDNA sequences for *K6c*, *K6d*, K6e, K6f, *K6h* are investigated, leading to the conclusion that *K6c*, *K6d* as well as K6e, K6f are probably polymorphic variants of *K6a* and *K6h*, respectively. All 26 human type II keratins found on this domain as well as K18, dtype 1 Keratin, are identified at the genomic and transcriptional level. This appears to be the total complement of functional type II keratins in humans.

Key words: cDNA/epidermis/gene expression/keratin 6/polymorphism
J Invest Dermatol 124:536 –544, 2005

Keratins consist of two families of intermediate filament proteins which make up one of the the major structural fibers of epithelial cells (Steinert and Roop, 1988; Fuchs and Weber, 1994; Rogers, 2003). The 8–10 nm keratin intermediate filaments consist of higher ordered copolymers of specific type I and type II keratin protein members, and keratins have gained increasing attention during the last couple of decades, because of their causal involvement in the etiology of a large heterogeneous group of hereditary epithelial diseases (Irvine and McLean, 1999; Smith, 2003). These paired members are usually expressed in a tissue-specific manner (Moll *et al*, 1982; for a review, see Moll, 1993). In humans, keratin genes are clustered as families in two regions of the genome, the type I genes on chromosome 17q21.2, those encoding type II keratins on chromosome 12q13.13 (Lessin *et al*, 1988; Romano *et al*, 1988, 1992; Rosenberg *et al*, 1988; Waseem *et al*, 1990a; Rogers *et al*, 1995). In the recent past, large portions of both the type I and type II keratin gene domains have been characterized and a preliminary bioinformatic analysis of the human genome provided evidence for several putatively new keratin genes within the type I and type II keratin gene clusters, as well as for a large number of keratin pseudogenes, found dispersed on either of the domains or spread out across the entire genome (Rogers *et al*, 1998, 2000; Hesse *et al*, 2001; Langbein *et al*, 2002, 2003).

This report describes the bioinformatic analysis of the 780 kb region of chromosome 12q13.13, containing the human type II keratin gene domain. Based on literature analysis and the expression data presented here, we show the presence of 27 functional keratin genes as well as eight keratin pseudogenes. As a part of this analysis we also present the genomic sequence of the *K2p* gene, and the cDNA sequences for three novel keratin genes previously designated *K1b*, *K6l*, and *Kb20*, respectively (Hesse *et al*, 2001, 2004) as well as for three previously described keratin genes, *K3*, *hHb2*, and *hHb4* (Klinge *et al*, 1987; Rogers *et al*, 2000). In addition, northern analyses are shown for the expression of K1b, K6l, Kb20 as well as K5b, another new keratin cDNA sequence recently isolated by the Human Genome Project. Moreover, partial DNA sequence analyses of the postulated K6 isoforms *K6a–f* and *K6h* (Takahashi *et al*, 1995; Langbein *et al*, 2002) in several individuals, as well as a complete scan of the Human Genome Project DNA sequences, indicates that genes *K6c*, *K6d*, and *K6e*, *K6f* appear to be polymorphisms of the *K6a* and *K6h* genes, respectively.

Abbreviations: BAC, bacterial artificial chromosome; ORF, open reading frame; PAC, P1 artificial chromosome

Copyright © 2005 by The Society for Investigative Dermatology, Inc.

Results

Bioinformatic analysis of the human type II keratin domain Screening of the EBI/Genebank database using a helix termination motif sequence, which is highly conserved in all type II keratin genes, led to the identification of nine bacterial artificial chromosome (BAC) clone DNA sequences that formed a single contig on chromosome 12q13 (Fig 1). This contig was approximately 780 kb in size and contained 35 putative keratin gene/pseudogene loci. The majority of this loci (from *K7* to *K2e*, see Fig 1) have been described previously (Rogers *et al*, 2000; Langbein *et al*, 2002, 2003). In addition, after the initial submission of this study, a bioinformatic analysis of the entire type II keratin domain (Hesse *et al*, 2004) was reported. This included the identification of a novel type II keratin, named *Kb20* using a new unified system of keratin nomenclature. This novel keratin gene had escaped our attention during our original bioinformatic evaluation. In total, the type II keratin contig harbors 21 epithelial keratin genes, 17 of which have been at least partially characterized, as well as six hair keratin genes, all of which have already been described (a list of the keratin gene loci, their cDNA sequences and publications relevant to these sequences can be found in Table S2). In addition to the 27 functional type II keratin genes, eight pseudogenes were also elucidated, four of which have been previously reported (Table S3). There are no unified directions of gene expression on the type II keratin gene domain (Fig 1). In general, the keratin gene loci range from 3.8 to 15.9 kb in size, and possess 3.6–44 kb of intervening sequences.

Two of the new, functional keratin genes, *K1b*, which corresponds to a previously described gene locus lying upstream of the K1 gene (Popescu *et al*, 1989) and *K6l* have previously only been analyzed bioinformatically (Hesse *et al*, 2001). Three further genes, *K5l*, *Kb20*, and *K2p* were present as published cDNA sequences in the EBI/Genebank database (Table S2) (Collin *et al*, 1992; Strausberg *et al*, 2002; Ota *et al*, 2004). In addition, at the beginning of our studies, the genomic sequence for *K2p* was not known. Therefore, a human genomic P1 artificial chromosome (PAC) clone containing this gene was isolated using the 3′-non-coding region of the K2p cDNA (Collin *et al*, 1992) as a probe, and the gene sequenced. This PAC clone (PAC24, see Fig 1 and Table S2) covered a region from upstream of the *K8* gene to downstream of the *K2p* gene and overlapped with BAC clones AC107016, AC139763, and AC068988 (Fig 1), which were submitted to the Genebank database at approximately the same time as the *K2p* gene sequence presented in this study. *Nde*I subclones of the relevant *K2p* containing regions from PAC24 were subcloned and sequenced (Table S2). The *K2p* gene is 9.1 kb in size, and consists, like other previously described type II keratin genes, of nine exons and eight introns, with seven exons (2–8) exhibiting position and size conservation.

Initial expression analysis of the new K1b, K5b, K6l, and Kb20 keratin genes Northern analysis of a range of human tissues using 3′-non-coding region probes derived from the *K1b*, *K5b*, *Kb20*, and *K6l* genes showed the mRNA expression of all of these keratins confined largely to single epithelial tissues. The approximately 3 kb K1b mRNA was found to be expressed weakly in skin (Fig 2*b*, *lane A*), but not in human scalp (*lane C*). In contrast, approximately 1.9 kb transcripts of K5b were clearly found in human tongue (Fig 2*c*, *lane E*). In addition, weak 3.5 kb transcripts of Kb20 were also found in this tissue (Fig 2*c*, *lane E*). In the case of K6l, moderate amounts of approximately 2.1 kb transcript occurred in human scalp (Fig 2*b*, *lane C*). In addition, a very weak expression was seen both in human skin and skeletal muscle (see Fig 2*b*, *lane A*, *asterisk* and Fig 2*a*, *lane 3*, *asterisk*). No other expression sites for K1b, K5b, K6l, or Kb20 were seen in any other of the tissues analyzed.

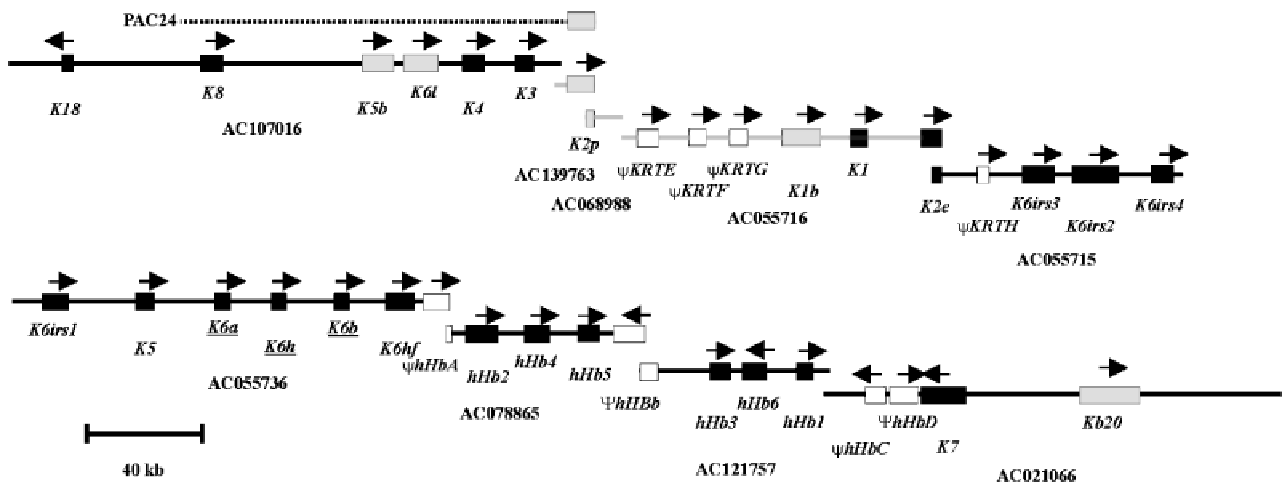

Figure 1
Physical map of the type II keratin gene domain. Horizontal black and gray lines represent bacterial artificial chromosome (BAC) clones sequenced to completion by the Human Genome Sequencing Project. Black lines represent positive strand sequence; gray lines are negative strand sequence. Black boxes designate previously described keratin gene loci; Gray boxes new keratin gene loci; white boxes keratin pseudogenes. Horizontal arrows show the direction of transcription. The accession numbers of the respective BAC clones are given below the gene names. The dotted line represents the *K2p* gene sequence (PAC24) described in this paper. The names of the "conventional" K6 isoforms are underlined. *Note*: The transcriptional orientation of the *hHb1* gene was incorrect in the original publication of the type II hair keratin domain (Rogers *et al*, 2000). We apologize for any inconvenience this might have caused.

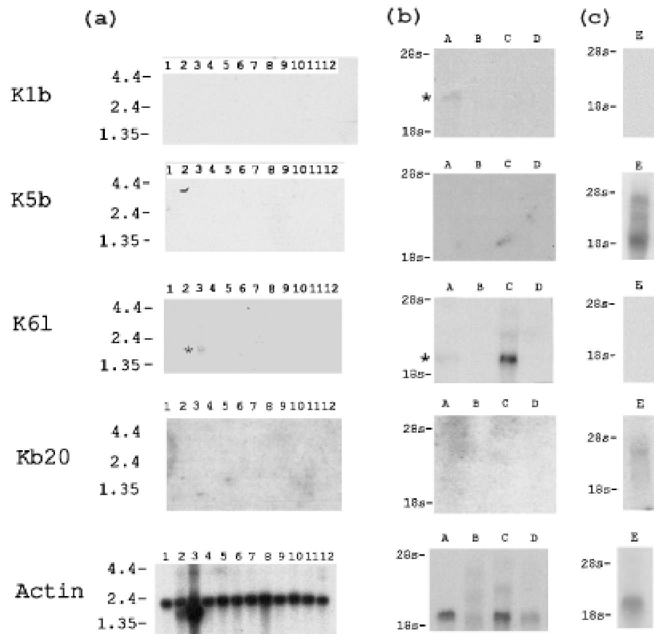

Figure 2
Northern blot analysis of K1b, K5b, and K6l. Panel (*a*) shows autoradiographs (K1b and K5b, 6 d; K6l, 3 d; Kb20, 8 d; actin 18 h) of the multi-tissue northern blot II (Clontech). The vertical numbering in (*a*) designates bands of an RNA molecular weight marker (kb). All tissues are of human origin. 1, brain; 2, heart; 3, skeletal muscle; 4, colon; 5, thymus; 6, spleen; 7, kidney; 8, liver; 9, small intestine; 10, placenta; 11, lung; 12, peripheral leukocytes. Panels (*b*) and (*c*) shows autoradiographs of human skin (A), total eye (B), scalp (C), breast (D), and total tongue mRNA (E); vertical number in (*b*, *c*) indicates position of 28s and 18s ribosomal RNA bands. Asterisks indicate weak bands of K1b in skin, and K6l in skeletal muscle and skin.

cDNA and deduced protein characterization of the K1b, K5b, K6l, and Kb20 keratins Using mRNA from human skin (K1b, K6l) or scalp (Kb20), it was possible to obtain full-length cDNA clones via RT-PCR for K1b, K6l, and Kb20 despite their weak, or in the case of Kb20, undetectable, northern blot expression, in these tissues (see Figs S1, S3, S4). Together with the database entry for K5b (Fig S2, see Table S2), a cDNA previously isolated from human tongue, this allowed the description of the four new keratins at the protein level (Figs S1–S4).

The K1b protein consists of 578 amino acids with a calculated molecular weight of 61.8 kDa. K5b, K6l, and Kb20 are somewhat smaller, consisting of 520, 535, and 452 amino acids, with a calculated molecular weight of 56.8, 57.8, and 50.5 kDa, respectively. Like all known keratins, the four new type II family protein members could be subdivided into amino- and carboxyterminal domains and a central rod domain consisting of four α-helical regions, separated by three non-helical linker sequences. Moreover, similar to other epidermal keratins (Korge *et al*, 1992a, b), K1b and K6l possessed numerous GGG and GGX repeats in their head and tail domains, whereas these motifs were distinctly less abundant in the corresponding domains of K5b and Kb20 (Figs S1–S4).

In addition to these new keratins, the hitherto lacking full-length cDNA for the type II hair keratins hHb2 and hHb4, as well as the corneal keratin K3, were isolated from human scalp and eye mRNA, respectively (Table S2).

Evolutionary analysis Comparison of the α-helical regions of the new keratins with the corresponding regions of the other type II keratins found on the gene domain described here, allowed the construction of a paralogous evolutionary tree, which depicted their placement within the entire human type II keratin family (Fig 3). In general, the type II keratins divided into several groups that often showed clustering of keratins exhibiting expression in distinct epithelial tissues. Striking was the grouping of keratins of the piliary apparatus into those of the hair-forming compartment, the inner root sheath, the companion layer, and the outer root sheath. The same held true for the keratins of simple and glandular epithelia and those of the differentiated epidermis. K1b showed good co-segregation with the epidermal keratins K1 and K2e. In contrast, K5b and K6l did not seem to co-segregate well with K5 or with the "conventional" K6 isoforms (see Discussion), but rather formed distinct, independent branches. Kb20 also showed a very long and distinctive branch length, reflecting its relatively low homology to the other type II keratins. Furthermore, the branch lengths of the type II keratins, in general, demonstrated that hair follicle associated epithelial and hair keratins arose at a much later date than the other keratins in this family (Fig 3).

K6 isoform analysis The genes postulated for *K6c*, *K6d*, *K6e*, and *K6f* (see Takahashi *et al*, 1995) could not be found in the region of chromosome 12 described in this report and were also absent in the bioinformatic evaluation of the same region by Hesse *et al* (2004). In addition, exact sequences of these genes could not be identified on any further region of the Human Genome Sequencing Project (Hesse *et al*, 2004,and this report). An attempt was made, therefore, to analyze the K6 isoforms in order to see if the above named genes might simply be polymorphisms of the three K6 gene loci (*K6a*, *K6b*, *K6h*) presented in Fig 1. The *K6a–f*, and *K6h* isoforms were divided into three groups (group 1, *K6a*, *K6c*, *K6d*; group 2, *K6b*; group 3, *K6e*, *K6f*, *K6h*) in a manner similar to that described previously (Takahashi *et al*, 1995, see Materials and Methods), and these amplification groups were investigated in four individuals. These experiments were performed under the auspices of the German Cancer Research center and conforms to the principles of the Declaration of Helsinki. An approximately 430 bp region from exon 9 of the various *K6* isoforms was analyzed. This region was chosen because it covers, genomically, the largest exon of these genes, which can be amplified using specific primer pairs for the *K6* isoforms. Exon 9 is located, in part, in the 3′-non-coding region of type II keratin genes, which should allow the detection of more diversity because of the lower evolutionary pressure on non-coding regions.

The upstream primer used for the analysis of the members of all three groups was identical. The downstream primers for groups 1 and 2 members were specific for each individual group. In the case of group 3 members, *K6e,f,h*, DNA sequences for *K6e* and *K6f* were not completely present in the region to be analyzed (accession numbers for these sequences are presented in the legend to Fig 4). Therefore, the sequence for *K6h* was used to prepare the downstream primer, under the assumption that the primer pair would probably amplify all three isoforms (see Discus-

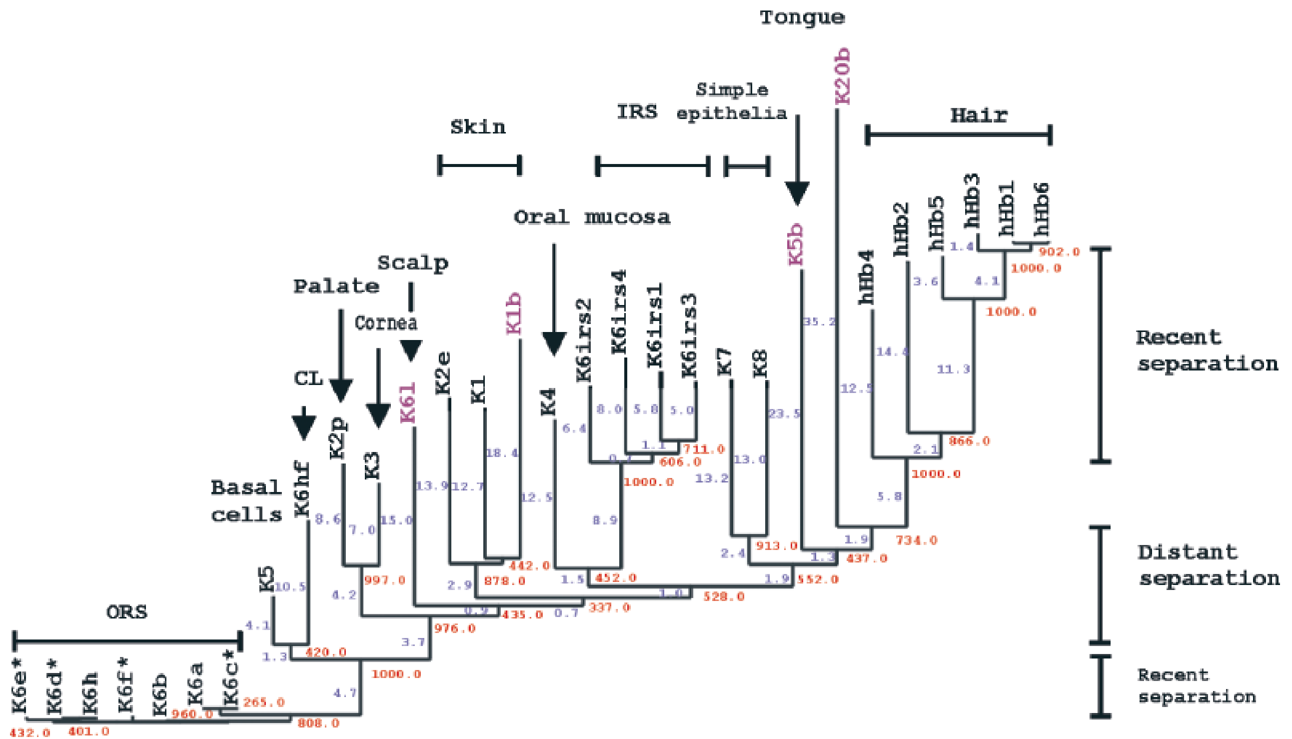

Figure 3
Paralogous evolutionary analysis of human type II keratins. The central approximately 310 amino acid region encompassing the rod domain of all known type II keratins was compared using the CLUSTAL program. Graphical analysis was performed using CLUSTREE. Numbers in blue show distance values; numbers in red represent statistical boot strapping values. Previously described keratins are labeled in black, the new keratin described in this paper are shown in violet. Asterisks beside the names delineate possible K6 isoform polymorphisms (see Discussion). Text above the tree mirrors correlation between keratin clustering and regions of common keratin expression. ORS, outer root sheath; IRS, inner root sheath; CL, companion layer.

sion). As can be seen in Fig 4 and as described below, *K6e* and *K6h* could be amplified using this approach (compare K6e and K6h with individuals 1 and 4, respectively, Fig 4*b*). No sequences were seen for *K6f*, however.

A moderate degree of variation was seen in group 1 between the four individuals analyzed. With minor variations, individual 2 showed only the DNA sequence for *K6a*, whereas individuals 1 and 4 possessed sequences for both *K6a* and *K6d*. Individual 3 exhibited two sequences, one of which contained aspects of both *K6a* and *K6d*. The single base pair deletion reported for *K6c* (Fig 4*a*, *yellow box*; see Takahashi *et al*, 1995) could not be found in any of the individuals analyzed. In addition several, previously undescribed nucleotide exchanges were found (Fig 4*a*, *black letters*), in addition to a 2 bp TT insertion (Fig 4*a*, *green box*) in single sequences of all individuals analyzed (Fig 4*a*).

In the group 2 analysis, individuals 1, 2, and 4 contained one single DNA sequence completely identical to that of *K6b*. Individual 3 possessed two DNA sequences, one of which was completely identical to K6b, and one which varied in three positions to *K6b* (data not shown).

In the group 3 analyses, individuals 2–4 contained two DNA sequences; one being completely identical to *K6h* and one in individuals 2 and 3 containing one A–G exchange when compared with *K6h*. Individual 4 also possessed the *K6h* sequence as well as an additional A–C exchange. Moreover, individual 1 exhibited two DNA sequences, one similar to the 1 bp *K6h* variation seen in individuals 2 and 3 but also containing the A–C exchange seen in individual 4, and one DNA sequence that showed complete identity to K6e (Fig 4*b*). No sequence homologies to *K6f* could be found in the four individuals investigated.

In order to expand and confirm the data presented above, group 1 amplification and cloning of a larger region of *K6a,c,d* was performed using, again, a conserved upstream primer from exon 7 in combination with the downstream primer from the previous exon 9 analysis, and the genomic DNA from individuals 2 and 4 were analyzed (Fig 4*c*). This new PCR product covered an approximately 1100 bp region encompassing exons 7–9 of *K6a,c,d*, and included the regions previously analyzed in Fig 4*a*. The reamplification/sequencing of 10 isolated clones from each individual resulted in the complete affirmation of the exon 9 data found in Fig 4*a* for individuals 2 and 4. This confirmed that the sequence variations in exon 9 found in these individuals were not because of early errors in PCR amplification. Analysis of the 683 bp region upstream of this sequence showed very little variation in the two individuals analyzed. The Human Genome Project sequence for K6a was used as reference sequence in this case, since it covered the entire region amplified. Both individuals 2 and 4 possessed one DNA sequence that was completely identical to the K6a (HG) reference sequence in this region. Individual 2 possessed a second DNA sequence showing a 1 nucleotide A–C deviation that is also found in K6d. Individual 4 possessed a second sequence that also showed this deviation as well as two additional nucleotide exchanges that were not similar to that of K6c or K6d. Several single

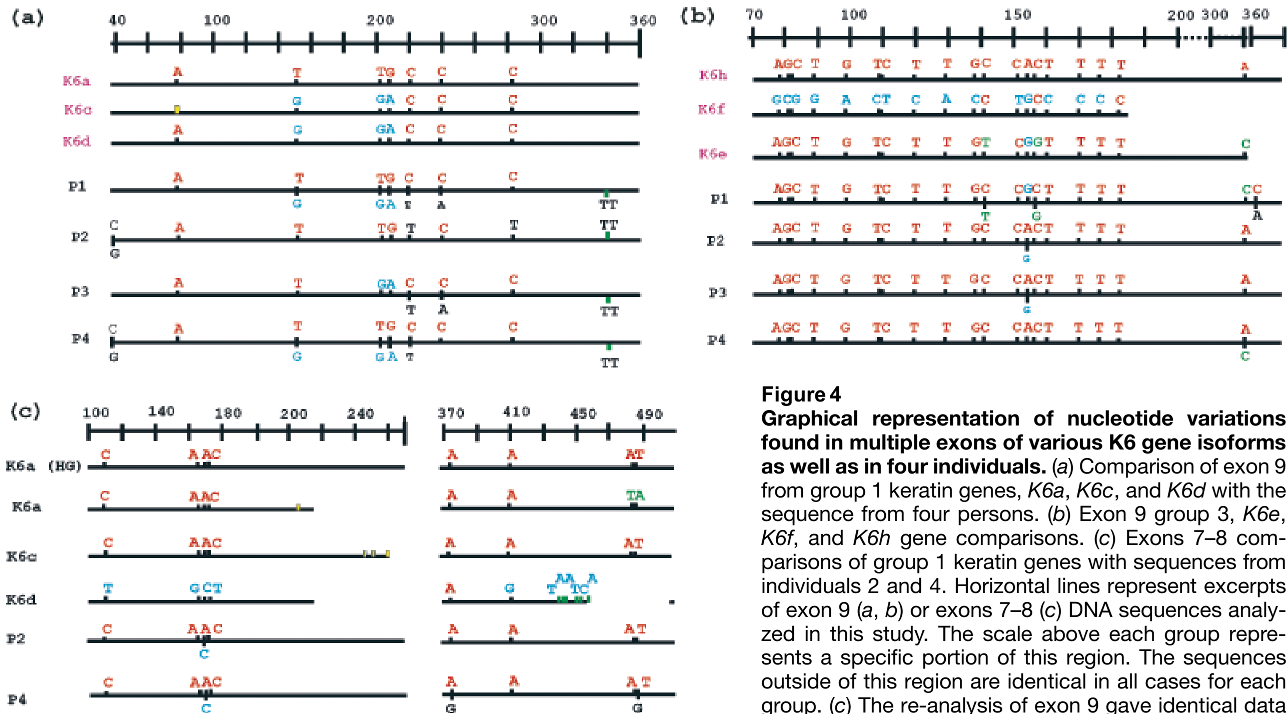

Figure 4
Graphical representation of nucleotide variations found in multiple exons of various K6 gene isoforms as well as in four individuals. (*a*) Comparison of exon 9 from group 1 keratin genes, *K6a*, *K6c*, and *K6d* with the sequence from four persons. (*b*) Exon 9 group 3, *K6e*, *K6f*, and *K6h* gene comparisons. (*c*) Exons 7–8 comparisons of group 1 keratin genes with sequences from individuals 2 and 4. Horizontal lines represent excerpts of exon 9 (*a*, *b*) or exons 7–8 (*c*) DNA sequences analyzed in this study. The scale above each group represents a specific portion of this region. The sequences outside of this region are identical in all cases for each group. (*c*) The re-analysis of exon 9 gave identical data for individuals 1 and 2 as that seen in (*a*) and, as such were not shown. The variations in the sequences are shown as black boxes. Boxes above the line indicate a single variation found. Boxes above and below the line show an additional variation found in the individuals analyzed. Yellow boxes shows deletions, green boxes represent insertions, the sequence of these insertions are shown above the box. Letters above the line show the first variation. Letters below the line show an additional variation. All variations are coupled on one DNA sequence. Letters in red show variations characteristic for *K6a* (*a*, *c*), and *K6h* (*b*) as well as similar sequences found in the groups and patients analyzed. Since (*c*) includes sequence data not found in the original article by Takahashi *et al* (1995), the Human Genome database sequence for K6a was used as reference sequence in this case. *K6a* and *K6h* are used as reference sequences in each group. Blue sequences show nucleotides that deviate from the reference sequence, but are found in another other, previously described group member. Letters in green shows variations specific for the putative *K6e* gene. Letters in black designate nucleotide variations that deviate from the originally published "conventional" sequences. The nucleotide sequences for the exon 9 K6 isoform analysis were taken from the accession numbers: *K6a*, hskrt6a09.emhum5; K6a (Human Genome Project), AC055736 bp 75310–76419, K6b, hskrt6b09.emhum5; *K6c*, hskrt6c09.emhum5; *K6d*, hskrt6d09.emhum5; *K6h*, AC055736 bp 94650–95082 (+strand); K6e cDNA, hskrt6e.emhum5; K6f cDNA, hskrt6f.emhum5.

nucleotide exchanges and deletions occurring in exon 7 of K6c when compared with K6a (HG) as well as multiple insertions occurring in exon 8 of K6d were not seen (Fig 4*c*). A 1 bp deletion in the intron following exon 7 in the original K6a sequence (Takahashi *et al*, 1995) was not seen in the human genome reference sequence or in the individuals analyzed.

Discussion

The characterization of human keratin genes and the localization of keratin gene expression by, among other methods, the identification, isolation, and characterization of keratin cDNA sequences from various epithelia, have involved the work of many laboratories over more than 20 years in time. The beginning of the end of these efforts appears to have occurred in 2001, and again recently in 2004, with reports of a bioinformatic analysis of the type I and type II keratin gene domains in the nearly complete, and then complete sequence of the human genome (Hesse *et al*, 2001, 2004). Furthermore, the identification and physical characterization of larger regions of these domains (Rogers *et al*, 1998, 2000; Langbein *et al*, 2002, 2003) has also occurred during this time period. In addition, a new aspect was added to this issue by the discovery of a domain of high/ultrahigh sulfur hair keratin-associated protein genes, which was imbedded into the type I keratin gene domain on chromosome 17q21.2 (Rogers *et al*, 2001). Although the final papers describing the sequencing of chromosomes 12 and 17 are still pending, the recent announcement of the completion of a reference sequence of the human genome (see http://www.ncbi.nlm.nih.gov/genome/seq/HsHome.shtml from 17 April 2003 and, Hesse *et al*, 2004), has allowed an extensive analysis of the type II keratin gene locus on chromosome 12q13.13, resulting here in the isolation of cDNA for the novel type II keratins K1b, K6l, and Kb20 (Table S2; Figs 1, S1, S3, S4), described previously by bioinformatic analysis in the reports of Hesse *et al* (2001, 2004). Together with the recent characterization at the cDNA/protein level of four type II keratins specific for the inner root sheath of the hair follicle (K6irs1–K6irs4, Langbein *et al*, 2003), as well as an additional keratin cDNA isolated from human tongue during the course of the human genome cDNA sequencing project (K5b, see Table S2 and Fig S2), this has resulted in the expansion of the number of functional keratin genes on chromosome 12 to 27 (note that K6irs3 was not seen in the original bioinformatic screen by Hesse *et al*, 2001). One of these genes encodes the type I keratin K18, known previously, as an exception, to be located on the type II keratin domain (see Fig 1, Waseem *et al*, 1990b; Yoon *et al*, 1994).

It is perhaps not surprising that the detection and characterization of keratins K1b, K5b, K6l, and Kb20 (as well as the inner root sheath keratins) has taken so long. Although this study has shown that K1b occurs in human skin, a tissue extensively analyzed in keratin research, the low degree of mRNA expression seen in Fig 2 implies that this keratin is expressed either in low quantities or in limited regions of the skin or in skin associated structures such as hairs, sebaceous, or sweat glands. Similarly, the presence of both K6l, or K5b and Kb20 in specialized regions such as the scalp or tongue, respectively, is also a reason for the difficulty encountered in identifying these keratins. Of special interest is Kb20. Initial cDNA sequences for this gene have been previously isolated by the Human Genome Project from tumor material. Since the normal tissue of expression was not known, and discrepancies existed in the cDNA sequence encoding the carboxyterminal portion of this keratin, it was decided to reisolate this cDNA sequence from non-tumor tissue. Both the low degree of homology of this protein to other keratins (see distance length of Kb20, Fig 3), as well as the positioning of this keratin in the evolutionary tree analysis between the epithelial and trichocytic keratins might lead to the assumption that this keratin possesses characteristics of both epithelial- and hair-forming cells (Fig 3, and preliminary expression data). Currently, further studies have been initiated to determine the precise expression profiles of all the keratins analyzed in the report via *in situ* hybridization, western blot analysis and immunocytochemistry.

At the beginning of our investigations, no DNA sequence existed for the *K2p* gene, a keratin found in the human palate epithelium (Collin *et al*, 1992). At this time, the type II keratin gene domain consisted of two partially sequenced BAC contigs separated by a gap. The isolation of a PAC clone containing the locus for the *K2p* gene allowed us to sequence this gene and to close the gap between the contigs (Fig 1 and Table S2).

Multiple *K6* gene isoforms have previously been shown to occur in a variety of mammalian species (Tyner *et al*, 1985; Blessing *et al*, 1987; Takahashi *et al*, 1995; Rothnagel *et al*, 1999; Langbein *et al*, 2002). In humans, the *K6* isoforms can basically be subdivided into two subsets. One comprises the originally described "conventional" isoforms, termed *K6a–f* (Takahashi *et al*, 1995, see also Wang *et al*, 2003), and *K6h*, respectively (Hesse *et al*, 2001; Langbein *et al*, 2002). All these genes encode proteins, which are strictly conserved in length and exhibit a very high degree of homology among each other. The second subset comprises the recently described "new" *K6* isoforms *K6hf*, and *K6irs1–4* (in addition to *K6l* from this report), which were named according to both their expression in the hair follicle and their co-localization with the members of the conventional K6 isoforms in western blot analyses (Winter *et al*, 1998; Langbein *et al*, 2002, 2003), in accordance to the original nomenclature concept of Moll *et al* (1982). The "new" K6 isoforms do not show length conservation and have a lower degree of homology when compared with the "conventional" isoforms.

Relying on the initial characterization of the human *K6b* gene and a K6a cDNA (Hanukoglu and Fuchs, 1983; Tyner *et al*, 1985), Takahashi *et al* (1995) described the isolation and exon–intron boundary sequencing of genes for four "conventional" *K6* isoforms, *K6a*, *K6b*, *K6c*, and *K6d*, in addition to two further putative isoforms, termed K6e and K6f, isolated as cDNA sequences. Based on the hybridization characteristics of the genes, they were divided into three groups. Group 1 comprised the *K6a*, *K6c*, and *K6d* genes; group 2 contained the *K6b* gene and group 3 consisted of non-analyzed genes, which did not fall into the other two groups, but possibly contained the putative genes for K6e and K6f, because of the high sequence homology of all "conventional" K6 isoforms. Recently, one further "conventional" *K6* gene isoform, *K6h*, was reported that could be located between the *K6a* and *K6b* genes on BAC clone AC055736 (Hesse *et al*, 2001; Langbein *et al*, 2002, see Fig 1). *K6h* possessed a high degree of homology to both the K6e and K6f cDNA and was therefore added to group 3 in this study.

Bioinformatic analyses of the nearly completed (Hesse *et al*, 2001), as well as completed human genome sequence (Hesse *et al*, 2004, and this paper), reveals the presence of only three "conventional" K6 isoform gene loci; namely those originally identified on the BAC clone AC055736, (K6a, K6b, and K6h, Langbein *et al*, 2002). Because of the high homology of the K6a,c,d and K6e,f,h isoforms, respectively, one answer to the discrepancy could be that each of these two groups could constitute polymorphic variants of a single gene. Another possible, but probably unlikely, answer might be that, because of the high degree of homology, difficulties arose in the assembly of the BAC sequences analyzed by the Human Genome Project leading to the discrepancies stated above. As such we analyzed the genomic sequence of a portion of exon 9 from all the "conventional" *K6* isoforms in four individuals, according to their subdivision into the three groups mentioned above, and, in a further experiment, analyzed exons 7–9 in two of these individuals. The original assumption associated with our analysis was that the differences observed in exon 9 of the various "conventional" isoforms, although minor, would be large enough to produce a pattern for each individual gene. Random polymorphic variation was assumed to be minimal (approximately 1 exchanges per kb) over such a stretch based on previous data taken from the literature (Brookes, 1999). With little variation, gene sequences continually present in all gene loci of the genome should be found in all persons analyzed. If, however, one or more of the *K6* isoform sequences was due to polymorphic variation of individual genes, then members might be absent in certain individuals, but present in others. Co-amplification of the DNA sequences present in the three groups using primers conserved for each group followed by cloning of the PCR product and sequencing of 10 single clones should, statistically, result in the identification of all isoforms present in the groups (see Materials and Methods). It should be noted, in general, that the determination of the exact gene number of recently divergent and highly conserved multiple gene loci by polymorphism analysis, without the help of supplemental bioinformatic and gene sequence data derived from the literature (as presented here for the "conventional" K6 isoforms) would be problematic, at best.

The data presented in Fig 4 led to the conclusion that several of the K6 isoforms are indeed polymorphic variants. In the group 1 exon 9 analyses, two individuals, 1 and 4

possessed two sequences, one containing the pattern for *K6a*, the other exhibiting the *K6d* pattern. Individual 2 displayed only a single sequence strongly similar to that of *K6a*, whereas individual 3 possessed sequences that appeared to be a mixture of *K6a* and *K6d*. In addition, all four individuals showed additional single nucleotide changes (indicated as *black letters* in Fig 4*a*) and an insertion (Fig 4*a*, *green bar*) not seen in the original DNA sequences (Takahashi *et al*, 1995). No sequence for *K6c* was seen in any of the individuals analyzed. The single nucleotide deletion in *K6c*, which is the only distinguishing feature between *K6c* and *K6d* in exon 9 (Takahashi *et al*, 1995), however, might not be adequate enough for discrimination between the two genes. The analysis of exons 7–9 in two individuals, however, showed the absence of strong sequence homologies for both K6c and K6d (see insertions and deletions in Fig 4*c*). Nevertheless, the presence of only one sequence showing strong identity to K6a in individual 2, and the presence of a maximum of two sequences in the other persons, leads to the conclusion that only one gene is probably present under the group 1 keratin genes analyzed. Furthermore, the mixed sequences in exon 9 of individual 3 would also support the idea that the changes between *K6a*, *K6c*, and *K6d* were polymorphic in nature.

The group 2 (*K6b*) amplification and sequening showed that three of the individuals, 1, 2, and 4, contained single sequences completely identical to *K6b*. Individual 3 possessed three nucleotide variations from this sequence (data not shown). This points to *K6b* being the single group 2 gene present on the human genome.

The analysis of exon 9 from group 3 (*K6e,f,h*) was more difficult. The downstream oligonucleotide primer used for the analysis of the three genes was taken exclusively from the *K6h* gene sequence, for no downstream sequences were available for *K6e* and *K6f* (see line lengths in Fig 4*b*). In addition, although *K6f* was quite different from *K6h* and *K6e* in the region analyzed (nucleotides in blue in Fig 4*b*), not even a downstream primer for a shorter *K6f* version could be made, for the originally described *K6f* sequence was 100% identical to that of the group 2 member *K6b* in this region. Therefore, in the best-case scenario, the primer pairs selected for the group 3 genes would amplify all three genes; in the worst-case scenario it would amplify only *K6h*. The data from Fig 4*b* point toward the amplification of at least two of the three sequences. Individuals 2–4, possessed two sequences, one identical to that of *K6h*, and one with a single nucleotide variation to *K6h*. Individual 1 also possessed two DNA sequences, one nearly identical to that of *K6h* and the other completely identical to that of *K6e*. In the best-case scenario, this would lead to the conclusion that the group 3 consists of a single *K6h* locus, whereas in the worst-case scenario, two loci, *K6h* and *K6e*, would be present.

Collectively, however, despite the inability to clarify all of the discrepancies found between this study and that of Takahashi *et al* (1995), and despite the uncertainties that remain concerning the number of group 3 *K6* isoforms, the fact that individual 2 in exon 9 possessed only sequence identities to *K6a*, *K6b*, and, with 1 nt difference, to *K6h*, coupled with the bioinformatic data presented in Fig 1, would indicate that only three "conventional" *K6* genes were present in any one individual. This is further confirmed in individual 2 by analysis of exons 7 and 8, which show only a single nucleotide variation between the reference K6a sequence and individual 2, as well as an additional 1 bp deletion occurring between the human genome K6a reference sequence and the exon 7 sequence of Takahashi *et al* (1995). Several deletions seen in the intron region of K6c following exon 7 and several insertions in the region following exon 8 of K6d are not present in the individuals analyzed, providing evidence that only the K6a isoform is present as a separate gene locus in these individuals. In total, the analysis of exon 9 in individuals 1–4 and the further analysis of exons 7 and 8 in individuals 2 and 4, coupled with the inability to find K6c, K6d, K6e, and K6f in the human genome reference sequence strongly suggests that the postulated *K6c* and *K6d* genes represent polymorphisms of *K6a* whereas the *K6e* and *K6f* genes are probably polymorphism of *K6h*.

In summary, the bioinformatic analysis of the region encompassing the human type II keratin domain coupled with the detection of three novel keratin genes, *K1b*, *K6l*, and *Kb20*, the sequencing of the *K2p* gene and the characterization of all previously outstanding cDNA sequences for distinct keratins leads to the conclusion that the number of genes presented here are complete at the genomic and transcriptional level. As further bioinformatic analysis outside of this region, at least in our hands, has not resulted in the identification of any further, functional type II keratin gene sequence, we feel, therefore, that this domain harbors the total number of functional type II keratin genes on the human genome.

Materials and Methods

Bioinformatic analysis of the type II keratin locus BLASTN2 homology analysis of type II keratin genes using the conserved helix termination motives found in exon 7 of several previously described type II keratin gene loci allowed the identification of nine BAC clone sequences characterized recently by the Human Genome Sequencing Project, which formed a single contig at chromosome 12q13 on the UCSC Human Genome Sequence map (http://genome.cse.ucsc.edu) (see Fig 1). Further bioinformatic analyses of these BAC clones were performed with the SIMILARITY program using known keratin gene/pseudogene/cDNA sequences as a query, which allowed the localization of previously described type II keratin genes on this genomic contig. Eight further, previously undescribed, keratin gene/pseudogene sequences were also found. The intron–exon boundaries of these novel genes were identified by comparison with known type II keratin cDNA as well as with expressed sequence tag sequences found in the EMBO/Genebank database. In addition, the intactness of the open-reading frames (ORF) of these putative genes/pseudogenes were analyzed using the α-helical regions of known type II keratin cDNA. Gene loci were designated pseudogenes when either single exons were missing or frame shifts were found in the ORF of the putative cDNA sequences. Evolutionary analyses of the type II keratin proteins were performed using the CLUSTAL homology analysis program. Graphical representation of this data was carried out using the CLUSTREE program. DNA sequence analysis was performed using the STADEN program. The majority of these programs are found in the Heidelberg Unix Sequence Analysis Resource (HUSAR).

Isolation of the K2p gene A 0.2 kb 3′-non-coding region PCR product from the K2p cDNA was used to screen an arrayed human genomic PAC library (see Table S1; library number 709 is available from the German Human Genome Resource Center (RZPD), Berlin, Germany) according to previously described methods (Rogers

et al, 1998). Thereafter, the isolated PAC clone DNA was restriction enzyme digested with *Nde*I, purified of small DNA fragments using a S400HR column (Amersham Biotech, Freiburg, Germany), ligated into the pGEM5Z cloning vector and transformed into XL1 blue competent cells. The correct *K2p* gene containing subclones were isolated by colony hybridization using the complete K2p cDNA (Collin *et al*, 1992) as a probe and sequenced.

Northern analysis Fifteen micrograms of total mRNA from human skin, eye, breast, placenta, thymus, tongue (all obtained from Bio-Cat/Biochain, Heidleberg, Germany) or human scalp (Rogers *et al*, 1995) were separated on a 1% agarose–formaldehyde gel and blotted onto Hybond N + nylon membrane as previously described (Fregien *et al*, 1983; Rogers *et al*, 1995). Pre-hybridization of this northern blot, as well as a further commercially obtained northern blot containing 2 μg of poly A + selected mRNA from human brain, heart, skeletal muscle, colon, thymus, spleen, kidney, liver, small intestine, placenta, lung, and peripheral blood leukocyte (MTNII-Clonetech, Heidelberg, Germany) were performed at 62°C for 2 h using Express Hyb hybridization solution (Clontech) supplemented with 100 μg per mL denatured Herring Sperm DNA (Sigma, Taufkirchen, Germany). Hybridization was performed overnight at 62°C using the same hybridization solution supplemented with 1×10^7 c.p.m. per mL of the respective ^{32}P-dCTP random labeled 3′-non-coding region probe described in Table S1. Thereafter the blot was washed five times in 2 × SSC/1% SDS followed by two times in 0.3 × SSC/1% SDS, all at 65°C. The blots were then autoradiographed on Kodak BMR1 film (Amersham Biotech, Freiburg, Germany) using intensifying screens for the times indicated in Fig 2.

Isolation of the cDNA for K1b, K3, K6l, and Kb20 via RT-PCR Exon–intron boundaries and ORF for the putative genes were determined by bioinformatic analysis and primer pairs were generated from the putative 5′- and 3′-non-coding regions of each gene (see Table S1). First strand syntheses of total human skin RNA (K1b, K6l), human scalp (Kb20), or total human eye (K3, all mRNA obtained from BioCat/Biochain) were performed using Superscript II AMV reverse transcriptase (Invitrogen, Karlsruhe, Germany). RT-PCR was performed using an Expand Long PCR kit (Roche, Mannheim, Germany) under the conditions detailed by the manufacturer (see also Table S1). The PCR products were separated on a 1% agarose gel, the appropriate bands excised and purified using a DNA gel extraction kit (Roche), incorporated into the PCR sequencing vector pCR4.1 (Invitrogen) and sequenced. In addition, the cDNA sequence for two previously described RT-PCR products of the type II hair keratins hHb2 and hHb4 were submitted to the Genebank/EBI database (Rogers *et al*, 2000).

DNA sequencing Plasmid DNA sequencing was performed using fluorescence dye terminator cycle sequencing using a Big Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Darmstadt, Germany) as previously described (Rogers *et al*, 2000). The sequence reactions were separated on an ABI310 DNA capillary sequencing apparatus (Applied Biosystems) using a 47 cm capillary.

Accession numbers for Gene/cDNA sequences analyzed in this study *K2p* gene, AJ564103; K3 cDNA, AJ628418; hHb2 cDNA, AJ628419; hHb4 cDNA, AJ628420; K1b cDNA, AJ564104; K6l cDNA, AJ564105; Kb20 cDNA, AJ717743.

Analysis of keratin 6a–f isoforms by genomic PCR An obvious discrepancy exists between the number of the originally described "conventional" *K6* isoforms (Takahashi *et al*, 1995; Langbein *et al*, 2002; Hesse *et al*, 2004, seven putative members—see Discussion for details) and the *K6* gene loci described here. Since one possibility might be that the four remaining isoforms are polymorphic variants of the three *K6* gene loci underlined in Fig 1, we decided to analyze these isoforms at the genomic level. As such, the *K6a–f*, and *K6h* isoforms were divided into three homology groups (group 1, *K6a*, *K6c*, *K6d*; group 2, *K6b*; group 3, *K6e*, *K6f*, *K6h* in a manner similar to that described previously by Takahashi *et al* (1995), and these amplification groups were analyzed in four individuals. Initially, an approximately 400 bp section of exon 9 of all *K6a–f* and *K6h* isoforms was chosen for PCR amplification and DNA sequence analysis because it allowed the largest region for comparison with the previously describe gene/cDNA sequences (Takahashi *et al*, 1995), as well as enabling the complete coverage of the region to be analyzed by end sequencing of the cloned PCR products. Since *K6a*, *K6c*, *K6d* were nearly, but not completely, identical, conserved primer pairs were made which co-amplified all three isoforms. In a similar manner, the amplificates of *K6e*, *K6f*, and *K6h* also possessed conserved upstream primers; the downstream primer was obtained from the *K6h* gene (see Discussion for details). All of the amplified PCR products (groups 1–3) were cloned into the CR4.1 vector and transformed into One Shot competent cells (Invitrogen). Plasmid DNA from 10 to 12 clones per group were prepared and sequenced from both ends. In order to confirm the data found during the exon 9 sequencing of the group 1 members, a conserved PCR primer was identified in exon 7 of all of the "conventional" K6 and a PCR analysis of of exons 7–9 was performed in two individuals using this primer as well as the same downstream primer found in the exon 9 analysis (see Table S1). This resulted in a 1110/1112 bp PCR product that was analyzed in a manner similar to that of exon 9 (see Fig 4*c*). In all cases, sequence correction was performed manually using the STADEN program. Discrepancies or regions of poor DNA sequence were resequenced, which allowed, in the end, approximately 1.7-fold coverage of each region. DNA homology analysis was performed using the CLUSTAL program. Besides the DNA sequence patterns of the seven isoforms described here, as well as the additional previously undescribed polymorphisms found in the groups 1 and 3 members during the course of analysis (Fig 3, see Discussion), additional single nucleotide exchanges in single cloned DNA sequences were seen in all three groups analyzed and were assumed to be reading errors of the Taq/PWO polymerase. In general, they appeared in less than half of the sequences analyzed and occurred with a frequency of 0–2 exchanges per sequence (421–432 bp).

Probability calculations Separately, for each of the group 1 (K6a, K6c, K6d) and group 3 (K6e, K6f, K6h) members, we investigated per individual $n = 10$ single PCR product clones under the hypothesis that each of the three respective DNA sequences in each group would have the same chance of developing into a clone, given it was present in the individual. The null hypothesis was, therefore, that there exists three DNA sequences in each group with each having the same chance of being identified by DNA sequencing. We calculated as statistical significance the probability p(n) of observing only one sequence in n replications given that there exist three distinct possible sequences (this problem is similar to discerning the presence or absence of three types of balls in an urn using n drawings). In this case $p(n) = (1/3)^{n-1}$. In the case of $n = 10$, we obtained a value of $p(n) = 0.000051$, which would give a statistical significance of $p < 0.001$ for this procedure.

We wish to thank Dr Paul Bowden, University of Wales College of Medicine, Cardiff, UK for pointing out an error in the previously reported orientation of the *hHb1* gene (Rogers *et al*, 2000). We also wish to thank Dr Francis J. D. Smith and Professor W. H. Irwin McLean, Ninewells Hospital and Medical School, University of Dundee, Dundee, Scotland with providing us with preliminary unpublished data concerning the *K2p* gene, and the German Human Genome Resource Center (RZPD), Berlin for their continuing help.

Supplementary Material

The following material is available from http://www.blackwellpublishing.com/products/journals/suppmat/JID/JID23530/JID23530sm.htm

Figure S1

cDNA- and amino acid sequence of K1b. Black arrowheads define the central rod domain as well as the region used for evolutionary analysis. Bent arrows delineate the α-helical and linker regions found within the rod domain. An asterisk indicates the stop codon.

Figure S2
cDNA- and amino acid sequence of K5b. For further details see legend to Fig S1.
Figure S3
cDNA- and amino acid sequence of K6l. For further details see legend to Fig S1.
Figure S4
cDNA- and amino acid sequence of Kb20. For further details see legend to Fig S1.
Table S1. Oligonucleotide primers and conditions used for genomic/cDNA PCR amplification. *, oligonucleotides conserved in K6 isoform sequences; #, downstream primer taken exclusively from the *K6h* gene (see Discussion for details). Note that the primer and PCR amplifying conditions for the hHb2 and hHb4 cDNA can be found in Rogers *et al* (2000)
Table S2. Genes and cDNA sequences found on the type II keratin domain. Only the genes currently found on the portions of the Human Genome Sequencing Project described in this paper are presented. See the Discusssion with regard to the *K6c*, *K6d*, *K6e*, and *K6f* genes. Relevant publications contain the first description of the most complete gene or cDNA sequence. * and #, gene or cDNA sequences completed by the Human Genome Sequencing Project which do not have a current literature citation. ∼ , Keratin genes for which no current Human Genome Nomenclature Committee (HGNC) designation exists
Table S3. Pseudogene sequences found on the type II keratin domain. *, These sequences are third party annotations (TPA) of the Human Genome Sequencing Project DNA sequence, which were submitted by the authors for this publication. HGNC, Human Genome Nomenclature Committee designation

DOI: 10.1111/j.0022-202X.2004.23530.x

Manuscript received March 16, 2004; revised July 23, 2004; accepted for publication August 24, 2004

Address correspondence to: PD Dr Michael A. Rogers, Research Program A145, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. Email: m.rogers@dkfz.de

References

Blessing M, Zentgraf H, Jorcano JL: Differentially expressed bovine cytokeratin genes. Analysis of gene linkage and evolutionary conservation of 5′-upstream sequences. EMBO J 6:567–575, 1987

Brookes AJ: The essence of SNPs. Gene 234:177–186, 1999

Collin C, Ouhayoun JP, Grund C, Franke WW: Suprabasal marker proteins distinguishing keratinizing squamous epithelia: Cytokeratin 2 polypeptides of oral masticory epithelium and epidermis are different. Differentiation 51:137–148, 1992

Fregien N, Dolecki GJ, Mandel M, Humphreys T: Molecular cloning of five individual stage- and tissue-specific mRNA sequences from sea urchin pluteus embryos. Mol Cell Biol 3:1021–1031, 1983

Fuchs E, Weber K: Intermediate filaments: Structure, dynamics, function and disease. In: Richardson CC (ed). Annual Review of Biochemistry, 63 edn. Palo Alto, CA: Annual Reviews Inc., 1994; p 345–382

Hanukoglu I, Fuchs E: The cDNA sequence of a type II cytoskeletal keratin reveals constant and variable structural domains among keratins. Cell 33:915–924, 1983

Hesse M, Magin TM, Weber K: Genes for intermediate filament proteins and the draft sequence of the human genome: Novel keratin genes and a surprisingly high number of pseudogenes related to keratin genes 8 and 18. J Cell Sci 114:2569–2575, 2001

Hesse M, Zimek A, Weber K, Magin TM: Comprehensive analysis of keratin gene clusters in humans and rodents. Eur J Cell Biol 83:19–26, 2004

Irvine AD, McLean WHI: Human keratin diseases: The increasing spectrum of disease and subtlety of the phenotype–genotype correlation. Br J Dermatol 140:815–828, 1999

Klinge EM, Sylvestre YR, Freedberg IM, Blumenberg M: Evolution of keratin genes: Different protein domains evolve by different pathways. J Mol Evol 24:319–329, 1987

Korge BP, Compton JG, Steinert PM, Mischke D: The two size alleles of human keratin 1 are due to a deletion in the glycine-rich carboxyl-terminal V2 subdomain. J Invest Dermatol 99:697–702, 1992a

Korge BP, Gan S-Q, McBride W, Mischke D, Steinert PM: Extensive size polymorphism of the human keratin 10 chain residues in the C-terminal V2 subdomain due to variable numbers and sizes of glycine loops. Proc Natl Acad Sci USA 89:910–914, 1992b

Langbein L, Rogers MA, Praetzel S, Aoki N, Winter H, Schweizer J: A novel epithelial keratin, hK6irs1, is expressed differentially in all layers of the inner root sheath, including specialized Huxley cells (Flugelzellen) of the human hair follicle. J Invest Dermatol 118:789–799, 2002

Langbein L, Rogers MA, Praetzel S, Winter H, Schweizer J: K6irs1, K6irs2, K6irs3, and K6irs4 represent the inner-root-sheath-specific type II epithelial keratins of the human hair follicle. J Invest Dermatol 120:512–522, 2003

Lessin SR, Huebner K, Isobe M, Croce CM, Steinert PM: Chromosomal mapping of human keratin genes: Evidence of non-linkage. J Invest Dermatol 91:572–578, 1988

Moll R: Cytokeratine als Differenzierungsmarker: Expression von Epithelien und epithelialen Tumoren. Jena: Gustav Fischer Verlag, 1993

Moll R, Franke WW, Schiller DL, Geiger B, Krepler R: The catalog of human cytokeratins: Patterns of expression in normal epithelia, tumors and cultured cells. Cell 31:11–24, 1982

Ota T, Suzuki Y, Nishikawa T, *et al*: Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet 36:40–45, 2004

Popescu NC, Bowden PE, DiPaolo JA: Two type II keratin genes are localized on human chromosome 12. Hum Genet 82:109–112, 1989

Rogers MA: Keratins and keratin diseases. In: Cooper DN (ed). Nature: Encyclopedia of the Human Genome. London: Nature Publishing Group, 2003; p 606–610

Rogers MA, Langbein L, Winter H, Ehmann C, Praetzel S, Korn B, Schweizer J: Characterization of a cluster of human high/ultrahigh sulfur keratin associated protein (KAP) genes imbedded in the type I keratin gene domain on chromosome 17q12–21. J Biol Chem 276:19440–19451, 2001

Rogers MA, Winter H, Langbein L, Wolf C, Schweizer J: Characterization of a 300 kbp region of human DNA containing the type II hair keratin gene domain. J Invest Dermatol 114:464–472, 2000

Rogers MA, Winter H, Wolf C, Heck M, Schweizer J: Characterization of a 190-kilobase pair domain of human type I hair keratin genes. J Biol Chem 273:26683–26691, 1998

Rogers MA, Nischt R, Korge B, *et al*: Sequence data and chromosomal localization of human type I and type II hair keratin genes. Exp Cell Res 220:357–362, 1995

Romano V, Bosco P, Rocchi M, Costa G, Leube RE, Franke WW, Romano G: Chromosomal assignments of human type I and type II cytokeratin genes to different chromosomes. Cytogenet Cell Genet 48:148–151, 1988

Romano V, Raimondi E, Bosco P, *et al*: Chromosomal mapping of human cytokeratin 13 gene (KRT13). Genomics 14:495–497, 1992

Rosenberg M, Raychaudhury A, Shows TB, Le Beau MM, Fuchs E: A group of type I keratin genes on human chromosome 17 characterization and expression. Mol Cell Biol 8:722–736, 1988

Rothnagel JA, Seki T, Ogo M, *et al*: The mouse keratin 6 isoforms are differentially expressed in the hair follicle, footpad, tongue and activated epidermis. Differentiation 65:119–130, 1999

Smith F: The molecular genetics of keratin disorders. Am J Clin Dermatol 4:347–364, 2003

Steinert PM, Roop DR: Molecular and cellular biology of intermediate filaments. Annu Rev Biochem 57:593–625, 1988

Strausberg RL, Feingold EA, Grouse LH, *et al*: Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Natl Acad Sci USA 99:16899–16903, 2002

Takahashi K, Paladini RD, Coulombe PA: Cloning and characterization of multiple human genes and cDNAs encoding highly related type II keratin 6 isoforms. J Biol Chem 270:18581–18592, 1995

Tyner AL, Eichman MJ, Fuchs E: The sequence of a type II keratin gene expressed in human skin: Conservation of structure among all intermediate filament genes. Proc Natl Acad Sci USA 82:4683–4687, 1985

Wang Z, Wong P, Langbein L, Schweizer J, Coulombe PA: Type II epithelial keratin 6hf (K6hf) is expressed in the companion layer, matrix, and medulla in anagen-stage hair follicles. J Invest Dermatol 121:1276–1282, 2003

Waseem A, Alexander CA, Steel JB, Lane EB: Embryonic simple epithelial keratins 8 and 18: Chromosomal location emphasizes difference from other keratin pairs. New Biol 2:464–478, 1990a

Waseem A, Gough AC, Spurr NK, Lane EB: Localization of the gene for human simple epithelial keratin 18 to chromosome 12 using polymerase chain reaction. Genomics 7:186–194, 1990b

Winter H, Langbein L, Praetzel S, *et al*: A novel human type II cytokeratin, K6hf, specifically expressed in the companion layer of the hair follicle. J Invest Dermatol 111:955–962, 1998

Yoon S-J, LeBlanc-Straceski J, Ward D, Krauter K, Kucherlapati R: Organization of the human keratin type II gene cluster at 12q13. Genomics 24:502–508, 1994