Identification of Human Epidermal Differentiation Complex (EDC)-Encoded Genes by Subtractive Hybridization of Entire YACs to a Gridded Keratinocyte cDNA Library

Ingo Marenholz, Maja Zirra, David F. Fischer, et al.

Genome Res. 2001 11: 341-355
Access the most recent version at doi:10.1101/gr.114801

Article

Identification of Human Epidermal Differentiation Complex (EDC)-Encoded Genes by Subtractive Hybridization of Entire YACs to a Gridded Keratinocyte cDNA Library

Ingo Marenholz,[1] Maja Zirra,[1] David F. Fischer,[2,4] Claude Backendorf,[2] Andreas Ziegler,[1,5] and Dietmar Mischke[1,3]

[1]*Institut für Immungenetik, Universitätsklinikum Charité der Humboldt-Universität zu Berlin, 14050 Berlin, Germany;*
[2]*Laboratory of Molecular Genetics, Institute of Chemistry, Leiden University, 2300RA Leiden, The Netherlands*

The epidermal differentiation complex (EDC) comprises a large number of genes that are of crucial importance for the maturation of the human epidermis. So far, 27 genes of 3 related families encoding structural as well as regulatory proteins have been mapped within a 2-Mb region on chromosome 1q21. Here we report on the identification of 10 additional EDC genes by a powerful subtractive hybridization method using entire YACs (950_e_2 and 986_e_10) to screen a gridded human keratinocyte cDNA library. Localization of the detected cDNA clones has been established on a long-range restriction map covering more than 5 Mb of this genomic region. The genes encode cytoskeletal tropomyosin TM30nm (*TPM3*), HSI-binding protein Hax-I (*HAXI*), RNA-specific adenosine deaminase (*ADARI*), the 34/67-kD laminin receptor (*LAMRL6*), and the 26S proteasome subunit p31 (*PSMD8L*), as well as five hitherto uncharacterized proteins (*NICE–1, NICE–2, NICE–3, NICE–4,* and *NICE–5*). The nucleotide sequences and putative ORFs of the EDC genes identified here revealed no homology with any of the established EDC gene families. Whereas database searches revealed that *NICE–3, NICE–4,* and *NICE–5* were expressed in many tissues, no EST or gene-specific sequence was found for *NICE–2.* Expression of *NICE–1* was up-regulated in differentiated keratinocytes, pointing to its relevance for the terminal differentiation of the epidermis. The newly identified EDC genes are likely to provide further insights into epidermal differentiation and they are potential candidates to be involved in skin diseases and carcinogenesis that are associated with this region of chromosome 1. Moreover, the extended integrated map of the EDC, including the polymorphic sequence tag site (STS) markers DIS1664, DIS2346, and DIS305, will serve as a valuable tool for linkage analyses.

[The sequence data reported in this paper have been submitted to EMBL under the accession nos. AJ243659–AJ243673.]

The chromosomal band 1q21 has been shown to harbor three gene families involved in terminal differentiation of the human epidermis within 2 Mb of genomic DNA (Volz et al. 1993). They encode precursor proteins of the cornified cell envelope (CE), intermediate filament-associated proteins, and calcium-binding proteins. The clustered organization of these genes and their evolutionarily conserved structural relationship together with functional interdependence of the encoded proteins led to their designation as a gene complex; the epidermal differentiation complex (EDC) (Mischke et al. 1996).

Loricrin, involucrin, and small proline-rich proteins (SPRRs) are the major precursors of the CE (Steinert et al. 1998), a highly insoluble and rigid structure that is essential for the barrier function of the skin. In terminally differentiating keratinocytes, the CE is assembled at the intracellular surface of the plasma membrane by transglutaminase-mediated cross-linking of these proteins. The corresponding genes *LOR* (Hohl et al. 1991), *IVL* (Eckert and Green 1986) and 10 *SPRR* genes belonging to 3 subgroups (2 *SPRR1,* 7 *SPRR2,* and 1 *SPRR3*) (Gibbs et al. 1993) are characterized by a similar gene structure, homologies in the terminal protein domains, and a variable number of internal tandem repeats. They constitute a cluster that most likely evolved from a common ancestor (Backendorf and Hohl 1992).

Profilaggrin, which is processed to functional filaggrin monomers, and trichohyalin are the main constituents of the keratohyalin granules in the epidermis and the hair follicle, respectively (Steven et al. 1990;

[3]Deceased September 11, 1999.
[4]Present address: Netherlands Institute for Brain Research, 1105 AZ Amsterdam, The Netherlands.
[5]Corresponding author.
E-MAIL andreas.ziegler@charite.de; FAX +49-30-450-53953.
Article and publication are at www.genome.org/cgi/doi/10.1101/gr.114801.

Fietz et al. 1993). They serve as keratin filament matrix and are also cross-linked to the CE (Steinert and Marekov 1995; Steinert et al. 1998). Their multifunctional structure combines sequence repeats, similar to the CE precursors, with two calcium-binding EF-hand domains that are typical features of the S100 proteins (see below) (Lee et al. 1993; Markova et al. 1993; Presland et al. 1995). The gene loci *FLG* and *THH* are located close together, centromeric to the CE precursor genes (Volz et al. 1993).

All of these structural genes of the EDC are flanked by 13 members of the *S100* family (*S100A1* to *S100A13*) (Schaefer et al. 1995; Wicki et al. 1996a,b) encoding calcium-binding proteins with two EF-hands (a calcium-binding motif named after the E- and F-helices of parvalbumin). S100 proteins are primarily regulatory proteins involved in different steps of the calcium signal transduction pathway. They mediate effects on cell shape, cell cycle progression, and differentiation (Schaefer and Heizmann 1996). In addition, incorporation of S100A10 and S100A11 into the CE was reported (Robinson et al. 1997), suggesting functional cooperation between calcium-binding and structural proteins in terminal differentiation of human keratinocytes.

Several inherited skin diseases have been associated with the EDC. Mutations in the loricrin gene accompanying abnormal CE formation are responsible for Vohwinkel's syndrome, a palmoplantar hyperkeratosis with ainhum-like constrictions of the fingers (Maestrini et al. 1996; Korge et al. 1997), and for progressive symmetric erythrokeratoderma (PSEK), which is characterized by a similar phenotype with expanded erythematous hyperkeratotic plaques (Ishida-Yamamoto et al. 1997). In addition, low levels of profilaggrin have been detected in ichthyosis vulgaris, a mild hyperkeratosis (Nirunsuksiri et al. 1995), and coordinate overexpression of *S100A7, S100A8, S100A9, SPRR1,* and *SPRR2* has been shown in chronic inflammatory and hyperproliferative psoriasis (Hardas et al. 1996), in line with a psoriasis susceptibility locus within the 1q21 region (Capon et al. 1999).

Altered expression of certain *S100* genes has also been observed in other diseases, such as chronic inflammation (Rammes et al. 1997) and cardiomyopathy (Remppis et al. 1996), as well as in different tumors, such as breast cancer (Lee et al. 1992; Pedrocchi et al. 1994; Moog-Lutz et al. 1995; Albertazzi et al. 1998) and malignant melanoma (Maelandsmo et al. 1997). Furthermore, chromosomal aberrations of the 1q21 region are often implicated in tumorigenesis (Gendler et al. 1990; Hoggard et al. 1995; Weterman et al. 1996; Forus et al. 1998).

In summary, identification of further genes located within the EDC should aid (1) to resolve the composition of biological structures in the epidermis, (2) to reveal potential control elements and signaling pathways governing differentiation of keratinocytes, and (3) to uncover genes and processes implicated in skin diseases or tumors associated with this region of chromosome 1. To reach this goal, a gridded keratinocyte cDNA library was constructed and successively hybridized with two entire YAC probes from the 1q21 region. Identified cDNA clones representing potentially new EDC genes were sequenced and their localization was confirmed on the integrated map of the EDC by use of additional 1q21-specific hybridization markers. Furthermore, cDNA sequences were analyzed with regard to their protein-coding regions and their functional domains. Finally, expression of the corresponding new genes during differentiation of cultured human epidermal keratinocytes was investigated.

RESULTS

Hybridization of the Gridded cDNA Library with a YAC

To identify new genes from chromosomal region 1q21 involved in epidermal differentiation, ^{32}P-labeled DNA of YAC 986_e_10 (1440 kb) that covers the central part of the EDC (Marenholz et al. 1996) was hybridized to the 184,320 clones of the gridded human keratinocyte cDNA library. Despite excessive competition with total genomic human DNA, almost 12,000 cDNA clones were detected (Fig. 1, 986_e_10), 50 of which were randomly picked and sequenced (Table 1, 1st approach). The hybridization succeeded in identifying several cDNA inserts originating from EDC genes. They encoded SPRR1A/1B and SPRR2A/2B, demonstrating inclusion of transcripts expressed late during terminal differentiation within the library. In addition, one previously unidentified cDNA could be assigned to distinct YACs of region 1q21, representing the *NICE-1* gene (for newly identified cDNA from the EDC).

However, as anticipated from the complexity of the probe, >70% of the clones contained repetitive sequences, like diverse short and long interspersed elements (SINEs and LINEs). Hybridization of 10 such inserts back to the YACs of the contig resulted in either non-specific signals (i.e., detection of multiple restriction fragments of several YACs) or gave no signal at all. Therefore, all clones containing repetitive elements were excluded from further analyses. A second source of non-specific hybridization were cDNA clones containing ribosomal DNA that were detected by YAC 986_e_10 (Table 1, 1st and 2nd approach). On a YAC filter, the respective inserts showed strong hybridization signals with yeast DNA due to the homology in ribosomal sequences between the yeast and human genomes. Detection of such clones was likely caused by contamination of the isolated YAC probe with yeast DNA.

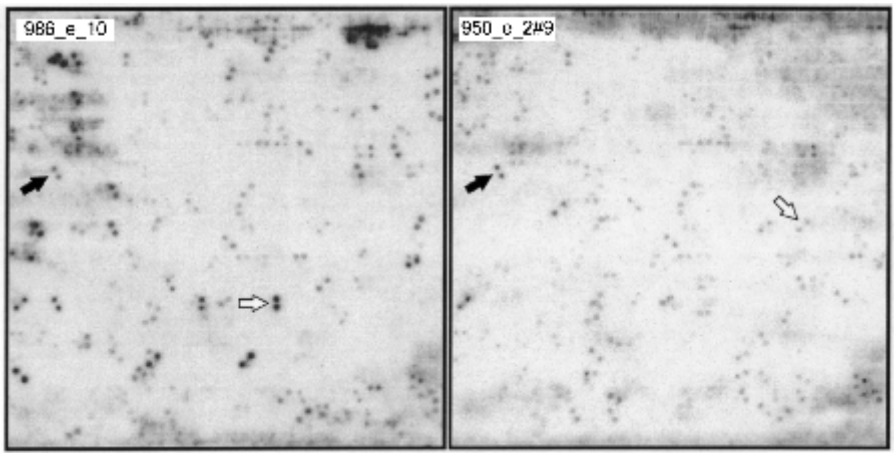

Figure 1 Hybridization of YACs 986_e_10 and 950_e_2#9 to a high-density filter of the gridded keratinocyte cDNA library. Autoradiographs show a quarter of a blot containing 18,432 double-spotted cDNA clones, which was successively hybridized with ^{32}P-labeled DNA derived from YACs 986_e_10 and 950_e_2#9, respectively. Positive clones are detected by two spots arranged in a specific pattern. Clones identified with both YACs were subtracted. Arrows indicate examples of a subtracted clone (black arrows) and of potentially EDC-specific cDNAs (white arrows).

Subtractive Hybridization

To increase the specificity of the method, the number of cDNA clones that were false-positive because of their repetitive sequences had to be reduced. Because YACs mapping in close vicinity within the human genome usually have a similar content of repetitive elements, YAC 950_e_2#9 (690 kb) was used as a probe for a second hybridization. This YAC is located in the distal part of the EDC but does not overlap with YAC 986_e_10 (Marenholz et al. 1996). Subsequently taking advantage of the gridded nature of the library, all cDNA clones hybridizing with both YACs could be easily identified to be non-specific and were disregarded (Fig. 1).

As a result, only 2091 (YAC 986_e_10) and 744 (YAC 950_e_2#9) of the initially detected clones remained that were expected to carry unique cDNA sequences from the EDC. Again, 50 clones were analyzed for each YAC (Table 1, 2nd approach). Among the cDNA clones detected by YAC 986_e_10 and left after subtraction, not only sequences corresponding to *NICE-1, SPRR1A/1B,* and *SPRR2A/2B,* but also to *SPRR3* and *IVL,* were identified as a consequence of the known gene content of the YAC. Comparable results were obtained for YAC 950_e_2#9, because genes known to be expressed in keratinocytes were retrieved, that is, *S100A4, S100A6,* and *S100A7*. Moreover, cDNA sequences of nine genes were detected that could be newly assigned to YACs of the EDC region. They represented *TPM3* (cytoskeletal tropomyosin TM30nm), *HAX1* (HS1-binding protein Hax-1), *ADAR1* (RNA-specific adenosine deaminase), *LAMRL6* (34/67-kD laminin receptor-like), *PSMD8L* (26S proteasome subunit p31-like), *NICE-2, NICE-3, NICE-4,* and *NICE-5*.

Table 1. Evaluation of Detected cDNA Clones

		1st Approach	2nd Approach (subtractive)	
		YAC 986_e_10	YAC 986_e_10	YAC 950_e_2#9
cDNA Clones	Detected	~12000	~12000	~10000
	YAC-specific	~12000	2091	744
	Sequenced	50	50	50
	Repetitive	36	3	—
	Ribosomal	1	3	—
	Assigned to 1q21	9	40	45
	Others	4	4	5
Established EDC genes		4	6	3
New-assigned EDC genes		1	1	9
EDC gene-specific clones among selected clones[a]		15%	85%	92%

[a]Determined by hybridization of gene-specific probes back to the gridded library.

After subtraction, only a few clones remained that could not be mapped to the YAC contig of region 1q21. Such clones were attributed to rare repetitive elements that are not contained in both of the YAC probes or might indicate chimerism of a YAC, as has been shown recently for YAC 986_e_10 that carries a rearranged centromeric arm (Lioumi et al. 1998). Further misleading hybridization signals were due to chimeric cDNA sequences or to clones contaminated by others, of which ~5% were identified each.

For a comparison of the results before and after subtractive evaluation, we used a mixture of the EDC-specific cDNA inserts identified by YAC 986_e_10 (corresponding to *SPRR1, SPRR2, SPRR3, IVL, NICE-1*) to probe the gridded library; 1777 clones were detected, that is, only 15% of the 12,000 clones initially found, but 85% of the 2091 clones left after subtraction. YAC 950_e_2#9 surpassed even this result; back hybridization of the 12 cDNA inserts originating from the EDC to the gridded library detected 685 of the 744 YAC-specific clones, which corresponds to a success rate of 92% (Table 1, last row).

Restriction Mapping of the Newly Identified EDC Loci

Because region 1q21 is frequently involved in chromosomal aberrations (Hoggard et al. 1995; Weterman et al. 1996; Forus et al. 1998), and assembly of the YAC contig had identified several rearranged YACs (Marenholz et al. 1996; Lioumi et al. 1998), we combined genomic and YAC restriction mapping to confirm the localization and to determine the order of the 10 newly detected EDC genes. Accordingly, 16 YACs of the established contig as well as 11 overlapping YACs (950_e_2#1 to 950_e_2#11) isolated from the unstable original culture 950_e_2, were used for construction of a *Sal*I restriction map. To achieve a higher resolution, additional probes for *S100A1, S100A2, S100A11, S100A13,* and for the nicotinic acetylcholine receptor β2-subunit gene *CHRNB2,* as well as five newly generated 1q21-specific markers (24f6, 24f15, 24f32, 24f39, and 24f57) were assigned to defined YAC restriction fragments. Subsequently, selected markers were positioned on the genomic long-range restriction map of 1q21 to compare the distances in YACs with genomic distances obtained from the H2LCL cell line.

Mapping results and sizes of the *Sal*I restriction fragments for the *S100* genes and *CHRNB2* agreed with previous data (Schaefer et al. 1995; Wicki et al. 1996a,b; Lueders et al. 1999). In addition, hybridization with defined fragments of distinct YACs unambiguously established the order of eight newly detected EDC genes. Only the arrangement of *NICE-4* and *HAX1,* both located on the same fragments, could not be resolved (Fig. 2B). Genomic *Not*I, *Nru*I, *Mlu*I, and *Bsi*WI restriction fragments detected by *S100A6* and *S100A4* were the most distal fragments of the previously established map (Volz et al. 1993). The most telomeric markers hybridized to hitherto undetected fragments and extended the physical map of the EDC ~1 Mb toward 1q22 (Fig. 2A). The assignment of *NICE-1, NICE-2, NICE-3, NICE-4, NICE-5* and *HAX1* to the EDC was unambiguously established by hybridization of the corresponding cDNA inserts exclusively to 1q21-specific restriction fragments. Likewise, localization of *ADAR1* and *TPM3* on these fragments refined their cytogenetic assignment to the chromosomal region 1q21.1-q21.2, and not 1q22-q23 (Weier et al. 1995; Wilton et al. 1995). In contrast to *ADAR1,* the *TPM3* probe hybridized with multiple genomic restriction fragments, indicating the presence of several related sequences within the human genome, in line with previous studies (MacLeod et al. 1986). A similar result was obtained by the *LAMRL6* probe due to pseudogenes on chromosomes 3, 12, 14, and X (Bignon et al. 1991; Richardson et al. 1998), and the functional laminin receptor gene (*LAMR1*) on chromosome 3p21.3 (Jackers et al. 1996). Finally, the *PSMD8L* probe also gave rise to extra bands, presumably due to cross-reaction with the *PSMD8* gene on chromosome 19 (accession no. AC005789).

Except for the lower resolution, the order of loci on the genomic map agreed with that of the YAC map. However, a discrepancy was observed in the distances between *NICE-2* and *ADAR1;* 1.3 Mb in genomic DNA as compared with the size of YAC 950_e_2#9 (690 kb), which had detected the corresponding cDNAs (Fig. 2A,B). Because at least 10 of 11 YACs isolated from 950_e_2 indicated the presence of an internal deletion, the genomic distance is more reliable. Correspondingly, the largest YAC from this series, 950_e_2#4, is expected to lack ~400 kb of genomic DNA.

The Extended Integrated Map of the EDC

The polymorphic STS markers D1S305 and D1S1664 from the initial YAC contig as well as D1S2346 from the genetic map of chromosome 1 (Hudson et al. 1995) were used to integrate physical and genetic mapping data. The positions of D1S305 and D1S1664 could be refined and assignment to various YACs of the contig established the localization of D1S2346, as illustrated on the integrated map (Fig. 2).

Apart from the above genetic markers, the map comprises all loci that were assigned to the EDC in this report, that is, 10 cDNA sequences from human keratinocytes and 5 1q21-specific hybridization markers, as well as the STS markers SHGC 57801 (accession no. G41921), SHGC-33740 (accession no. G29465), and SHGC-11135 (accession no. G13549), which coincide with 3 of the novel genes. In addition, the localization of all genes and hybridization markers mapped previously to the EDC (Marenholz et al. 1996; Mischke

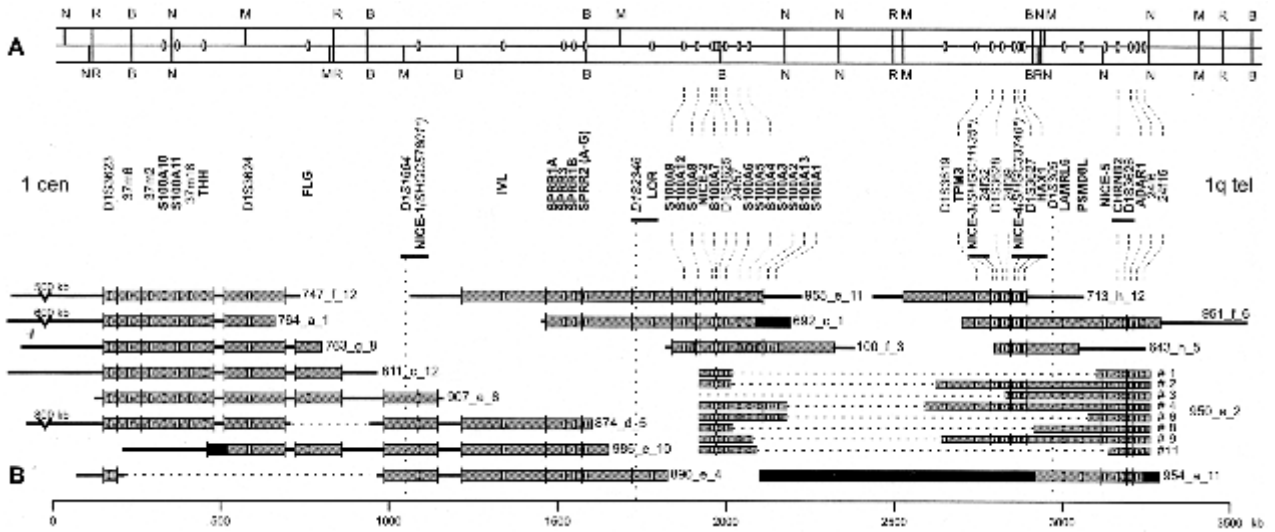

Figure 2 Integrated map of the EDC. Genomic and YAC restriction mapping results yielded the order of genes (in bold), STSs (in italics), and other loci within the EDC as shown. STS markers coinciding with the newly assigned genes were not tested and are marked with an asterisk. Precise localization of 37m2/37m16, of the two *SPRR1* genes and of *S100A1/S100A13* was resolved previously (South et al. 1999). The order of loci underlined with a horizontal bar could not be resolved. (*A*) Genomic restriction map of the EDC. A continuous map spanning 3.5 Mb of region 1q21 with *Not*I (N), *Nru*I (R), *Mlu*I (M), and *Bsi*WI (B) restriction sites of the two haplotypes present in the H2LCL cell line (represented by the two lanes) is shown. White ovals indicate hybridization of the probe specified below to the corresponding restriction fragments that resulted from single and double restriction enzyme digests. Refined mapping results of *S100A10, THH, FLG, IVL, SPRR3, SPRR1, SPRR2, LOR, S100A9, S100A8,* and *S100A6* obtained by digestion with additional restriction enzymes (not shown) were adopted from the established map (Mischke et al. 1996). (*B*) Refined YAC contig of the EDC. Sixteen YACs from the 6-Mb contig covering the EDC (Marenholz et al. 1996) are shown with their addresses. Eight overlapping YACs of 950_e_2 that were used to resolve the order in the distal region are included. Sizes of YACs as determined previously by rotating-field gel electrophoresis (ROFE) are represented by the sum of the length of the gray and black boxes (detected fragments) and of the black horizontal bars (unidentified DNA sequences). White ovals represent hybridization of the respective fragment with the corresponding probe, derived from the locus above. *Sal*I restriction sites are indicated by vertical lines. A polymorphic restriction site in *NICE-1* was detected for 907_e_6 and 874_d_5. Gray boxes correspond to continuously covered human DNA as deduced from marker content and from identical fragments in several independent YACs. Broken horizontal lines indicate deletions and black boxes indicate other rearranged fragments identified by aberrant size and/or marker content. Restriction fragments of distinct size that were only detected in a single YAC are potential end fragments or could be rearranged and lack the second *Sal*I site marker. Arrangement of detected *Sal*I restriction fragments with respect to the unidentified DNA sequences depended on YAC sizes and genomic mapping results. *S100A3, S100A5,* and *S100A12* represented by gray ovals were mapped previously to YAC 100_f_3 (Schaefer et al. 1995; Wicki et al. 1996b). D1S1664, D1S2346, and D1S305 (vertical broken lines) were assigned to the corresponding YACs by PCR and therefore could not be positioned on single restriction fragments. The *NICE-2, S100A7,* and D1S3625 probes each hybridized to the same two *Sal*I fragments, most likely indicating a duplication of sequences. Their order was resolved on the basis of signal intensities.

et al. 1996; Wicki et al. 1996a,b; Lueders et al. 1999; South et al. 1999) has been verified on the *Sal*I restriction map.

Analysis of the cDNA Sequences Corresponding to EDC Genes

Of the 150 clones sequenced from the gridded human keratinocyte cDNA library, 94 could be unambiguously assigned to region 1q21 (Table 1). These represented a total of 19 different genes, 9 established EDC genes serving as positive controls for subtractive hybridization, 5 genes coding for known proteins, which could be identified here as new members of the EDC, and 5 novel EDC-encoded genes (i.e., the *NICE* loci, see below).

Comparison of the cDNA sequences encoding the known proteins with the corresponding database entries revealed >99% similarity (Table 2). In the case of *SPRR3,* a second transcript was isolated that was distinguished by a missing repeat and defined nucleotide substitutions (Fig. 3). The deduced amino acid sequence was identical with SPRC, a protein identified previously in oral mucosa (Robinson et al. 1994).

Several cDNAs did not match a gene of known function in the databases. Because of their identification in the keratinocyte library and their assignment to the EDC, these were designated as *NICE.* Alignment studies revealed transcripts of five different *NICE* genes, two of which exhibited alternative splicing. The corresponding cDNAs were analyzed with regard to overlapping nucleotide sequences in the databases and to their putative protein coding regions, allowing the characterization of amino acid sequences in four cases (Table 3). In addition, as a first approximation, the tissue distribution of the transcripts was determined by EST analysis, followed by the direct analysis of differ-

Table 2. Characterization of EDC Genes Encoding Known Proteins

YAC probe	Detected transcript (gene)	Size of cDNA	Homologous to: acc. no. (identities)	Positive YACs
	Established EDC genes			
986_e_10	Involucrin (*IVL*)	2139 bp	M13902 / M13903 (99%)	874_d_5, 986_e_10, 890_e_4, 955_e_11
	Small proline-rich protein 3 (*SPRR3*)	890 bp / 859 bp	AF077374 (99% / 95%)	874_d_5, 986_e_10, 890_e_4, 955_e_11, 692_c_1
	Small proline-rich protein 1A/1B (*SPRR1A/1B*)	769 bp / 595 bp	L05187 (99%) / M84757 (99%)	874_d_5, 986_e_10, 890_e_4, 955_e_11, 692_c_1
	Small proline-rich protein 2A/2B (*SPRR2A/2B*)	674 bp / 661 bp	M20030 (99%) / M21302 (99%)	874_d_5, 986_e_10, 890_e_4, 955_e_11, 692_c_1
950_e_2#9	S100 Calcium-binding protein A7 (*S100A7*)	416 bp	M86757 (99%)	955_e_11, 692_c_1, 100_f_3, 950_e_2 (#1, #2, #4–#11)
	S100 Calcium-binding protein A6 (*S100A6*)	420 bp	M14300 (100%)	955_e_11, 692_c_1, 100_f_3, 950_e_2 (#4, #6, #7, #9, #11)
	S100 Calcium-binding protein A4 (*S100A4*)	496 bp	M80563 (99%)	955_e_11, 692_c_1, 100_f_3, 950_e_2 (#4, #6, #7, #9, #11)
	New-assigned EDC genes			
	Adenosine deaminase, RNA-specific (*ADAR1*)	2625 bp[a]	U18121 (99%)	950_e_2 (#1–#11), 951_f_6, 954_a_11
	Cytoskeletal tropomyosin TM30nm (*TPM3*)	2075 bp	X04588 (99%)	950_e_2 (#2, #4, #7, #9), 713_h_12, 951_f_6
	26S Proteasome subunit p31 (*PSMD8L*)	973 bp	D38047 (100%)	950_e_2 (#2–#4, #7–#10), 951_f_6, 954_a_11
	HAX-1, HS1-binding protein (*HAX1*)	1084 bp	U68566 (99%)	950_e_2 (#2–#4, #7, #9), 713_h_12, 951_f_6, 643_h_5
	34 / 67 kd Laminin receptor (*LAMRL6*)	1025 bp	J03799 (99%)	950_e_2 (#2–#4, #7–#10), 951_f_6, 643_h_5, 954_a_11

[a]All sequenced cDNA inserts originated from the 3′ untranslated region (UTR) of the gene.

entiation-specific expression of the *NICE* genes by Northern blot hybridization.

Three alternatively spliced cDNA sequences were obtained for *NICE-3* (Fig. 4). The corresponding transcripts of this gene revealed identical 3′-terminal sequences including the STS markers SHGC-11135 and TIGR-A002G29 from chromosome 1. They lacked up to two internal exons, and two more missing exons were identified in ESTs from various human tissues. One of the proteins deduced from *NICE-3* was identical with the predicted human protein HSPC012 (accession no. 4689120).

NICE-4 was represented by two different cDNA sequences in keratinocytes. Another cDNA originating from the same gene is KIAA0144, which had been isolated from a myeloid cell line (Nagase et al. 1995). Although identical in the major part of the sequence, each of the transcripts carried a different 3′-terminal sequence, resulting in modified carboxy-termini of the encoded proteins (Fig. 5). Similarity search with the *NICE-4* sequences detected highly conserved ESTs from mouse and rat, confirming previous results of Vos (1997) who had identified mouse ESTs homolog to KIAA0144 as well as alternatively spliced human sequences that overlap KIAA0144 and include STS 33740 from the chromosomal region 1q21.

Single consensus sequences with distinct ORFs were identified for the cDNAs corresponding to *NICE-1* and *NICE-5,* respectively. A large number of clones contained the *NICE-1* sequence (Fig. 6), which includes the STS marker SHGC-57801. The predicted ORF resembled (36% similarity) the skin-specific protein (accession no. 2589188) expressed from the gene *xp5* (accession no. AF005080), which had been mapped previously to the 1q21 region (Zhao and Elder 1997). It is possible that NICE-1 and xp-5 are the first members of a novel EDC gene family.

The protein translated from the predicted ORF of *NICE-5* is related (~50% similarity) to two proteins of unknown function, one deduced from the *Drosophila melanogaster* gene EG:25E8.2 (accession no. AL009196) and one from the *Caenorhabditis elegans* gene F25H2.8 (accession no. Z79754). Northern blot hybridization yielded a transcript size of 2000 nucleotides and human ESTs, which overlap the 5′-end of the consensus sequence (880 bp) indicate a partial cDNA (Fig. 7).

Several cDNA clones were isolated from the *NICE-2* gene. Although varying in size, they obviously origi-

nated from the same mRNA that was converted into cDNA to different extents. Database search identified D1S3625, a subcloned sequence of a YAC from the contig, with 85% similarity, but no EST or gene-specific sequence. Because the *NICE-2* sequence contained no significant ORF, the detected clones presumably represent a large 3′-untranslated region (Fig. 7).

A Northern blot containing RNA samples of human skin, keratinocytes cultured under different conditions, and various tumor cell lines, as well as primary fibroblasts and melanocytes was used to investigate expression of the *NICE* genes depending on the tissue and the differentiation stage of keratinocytes. Low RNA levels in all samples tested were detected for *NICE-2, NICE-3, NICE-4,* and *NICE-5* (data not shown). In contrast, *NICE-1* expression was only found in keratinocytes, with the highest RNA level in keratinocytes induced to differentiate by calcium addition (Fig. 8). The expression pattern for *NICE-1* was similar to the expression of the keratinocyte-specific terminal differentiation marker *SPRR2*. However, *NICE-1* ESTs were also detected in a heart library, indicating that its expression is not restricted to keratinocytes (Table 4).

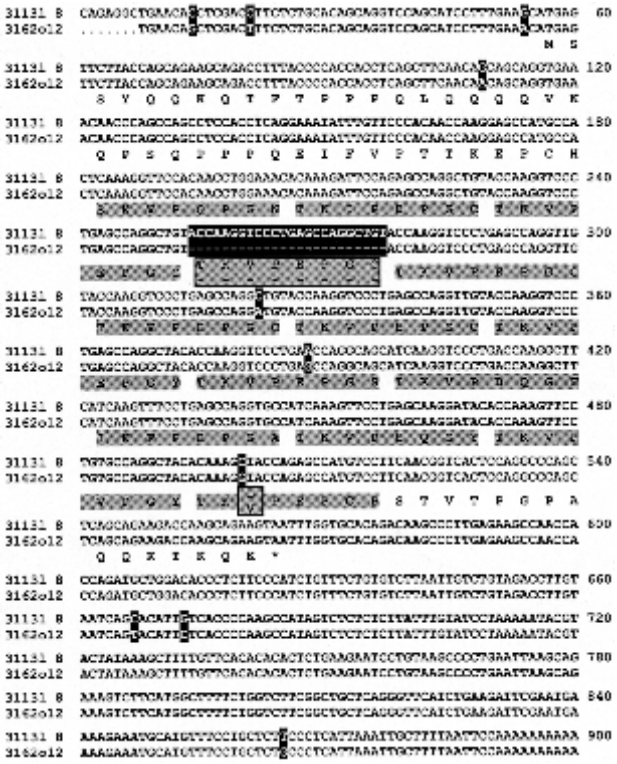

Figure 3 Alignment of the cDNA clones 31131 8 and 3162o12 representing two different alleles of the *SPRR3* gene. Black boxes indicate divergences in the nucleotide sequences, open boxes indicate divergences in the deduced amino acid sequences that contain 14 (clone 31131 8) and 13 (clone 3162o12) octapeptide repeats (shaded gray), respectively. Among the four mutations within the coding region, only one results in a substituted amino acid residue, with higher homology to the consensus repeat sequence TKVPEPGC of clone 3162o12.

DISCUSSION

In this work, we describe a highly efficient and straightforward method for identifying expressed genes from a targeted genomic region, by use of entire YACs to screen a gridded cDNA library. Whereas previous investigations reported on problems due to nonspecific hybridization by use of such complex DNA probes (Elvin et al. 1990; Boultwood et al. 1997), the novelty of the present work relies on the use of a subtractive approach. A gridded keratinocyte cDNA library allowing cross-referencing of data was successively hybridized with two non-overlapping YACs from the well-characterized contig covering the EDC (Marenholz et al. 1996; Lioumi et al. 1998). After subtracting non-specific cDNA clones as indicated by hybridization with both of the YACs, characterization of the remaining clones yielded transcripts of 19 EDC genes, 10 of which were assigned to this gene complex for the first time. Although sequencing and mapping was restricted to 50 cDNA clones per YAC, the isolated EDC-specific sequences covered ~90% of all 2835 clones specifically detected by one YAC, surpassing most positional cloning methods that yielded a maximum of 15% positives (Hisama et al. 1998). Compared with this success rate, the only drawback, namely the loss of those cDNA clones that contain genuine repetitive sequences despite their localization within the region of interest, should constitute a minor problem. Considering that 10,000 clones were disregarded because of nonspecific hybridization, ~5% of the 186,432 cDNAs in the library might be lost by this method. Because publically available data obtained by large-scale sequencing and mapping of ESTs are increasing rapidly, we compared our results with the respective database entries for sequences that were expressed from the *NICE* genes (Table 4). No information was available for the *NICE-2* sequence. Single-splicing products of *NICE-3* and *NICE-4* had not been detected either, and the *NICE-1* sequence was present only partially. ESTs originating from the *NICE-1, NICE-3, NICE-4,* and *NICE-5* genes had already been located on the human transcript map, between the markers D1S514 in region 1q21 and D1S2635 in 1q22. However, as this interval spans ~13 cM of chromosome 1, this localization on the gene map was too imprecise for the assignment of any of these loci to the EDC.

Genomic restriction mapping established the localization of the *NICE* genes, of *HAX1, ADAR1,* and *TPM3* within the EDC. The new loci for a 26S proteasome subunit p31 gene (*PSMD8L*) and for a laminin receptor gene (*LAMRL6*) in region 1q21, however, require further characterization to resolve their functional properties and significance. In line with genetic data (Hudson et al. 1995), the polymorphic STS mark-

Table 3. Characterization of the *NICE* Genes

YAC probe	Gene	Size of mRNA[a]	BLASTN (DNA): acc. no. (size), identities	BLASTP (protein): acc. no. (size), identities	Putative ORF	Putative protein domains	Positive YACs
986_e_10	*NICE-1*	694 bp	EST AI096376 (482bp), 99%	Skin-specific protein, 2589188 (110aa), 36%	99 aa	Metallothionein, prokaryotic membrane lipoprotein lipid attachment site	907_e_6, 874_d_5, 986_e_10, 890_e_4
950_e_2#9	*NICE-2*	5500 bp	—	—	—[b]	—	955_e_11, 692_c_1, 100_f_3, 950_e_2 (#1, #2, #4–#11)
	NICE-3	1710 bp 1608 bp 1554 bp	HSPC012 mRNA, AF077036 (1636bp), 99%	HSPC012 4689120 (219aa), 100%	253 aa 219 aa 201 aa	N-4 cytosine-specific DNA methylases signature	950_e_2 (#2, #4, #7, #9), 713_h_12, 951_f_6
	NICE-4	3976 bp 3908 bp	cDNA KIAA0144, D63478 (3411bp), 99%	Hypothetical protein of KIAA0144, 2495711 (983aa), 100%	1167 aa 1068 aa	Bipartite nuclear localization signal, ubiquitin-associated domain, prokaryotic membrane lipoprotein lipid attachment site (1167 aa protein only)[c]	950_e_2 (#2–#4, #7, #9), 713_h_12, 951_f_6, 643_h_5
	NICE-5	2000 bp	EST AA769605 (701bp), 99%	Hypothetical protein of *Drosophila melanogaster*, 2827495 (394aa), 52%	>322aa[d]	Biotin-requiring enzymes attachment site, glutamine synthetase class-I adenylation site	950_e_2 (#1–#10), 951_f_6, 954_a_11

[a]Size as determined by Northern hybridization and confirmed by cDNA sequencing; sizes of *NICE-2* and *NICE-4* were only determined by Northern hybridization, because all cDNA inserts sequenced were incomplete.
[b]No significant ORF was detected, implying that the available sequence represents the 3′ untranslated region (UTR) of the gene.
[c]Because of space limitations, only the results of the PROSITE pattern search are indicated.
[d]The available sequence is likely to lack the 5′ end of the ORF.

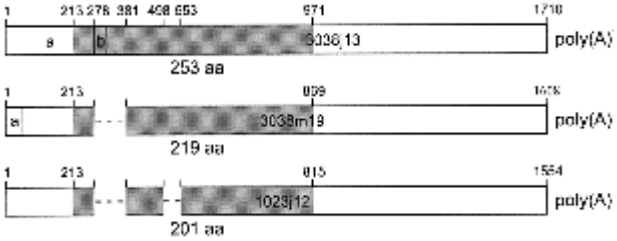

Figure 4 Alternatively spliced gene products of *NICE-3*. The nucleotide sequences of the three cDNA clones were identical except for the lacking exons (broken lines) of clones 3038m19 (exon 279–380) and 1023j12 (exons 279–380 and 499–552). Incomplete 5′-termini of clones 3038m19 and 3038j13 were filled up with overlapping sequences of clone 1023j12 (a) and EST T27537 (b). Sequence information derived from the respective clones is represented by open boxes; ORFs are shaded gray. The first nucleotide of the putative initiation methionine codon, exon boundaries, and the last nucleotide of the coding sequence, as well as the size of the predicted protein are indicated for each transcript. Additional spliced exons and combinations of them were identified in ESTs by similarity search: exon 202–278 in AA354455 from Jurkat T-cells, exons 202–278 and 279–380 in AA463392 from total fetus, exon 381–498 in AA057488 from colon, exons 279–380 and 381–498 in Z42265 from infant brain.

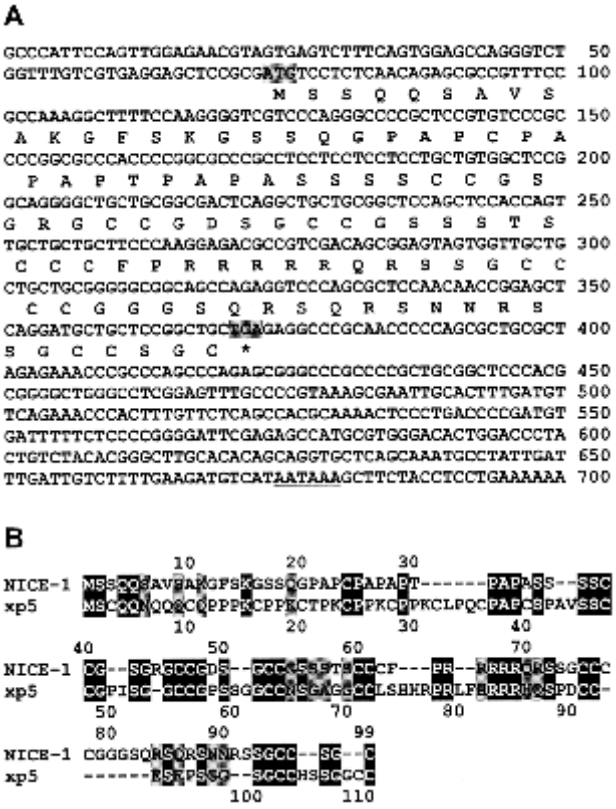

Figure 6 Sequence analysis of *NICE-1* cDNA. (*A*) Nucleotide sequence of *NICE-1* and the amino acid sequence deduced from its largest ORF. The putative initiation methionine codon ATG and the termination codon TGA are shaded gray. The polyadenylation signal AATAAA is underlined. (*B*) Comparison of the protein sequences of *NICE-1* and *xp5*. Black boxes indicate identical, gray boxes indicate similar amino acids. Dashes indicate missing amino acids. Pairwise sequence alignment was performed at http://vega.crbm.cnrs-mop.fr/bin/align-guess.cgi.

ers D1S1664, D1S2346, and D1S305 were placed on the physical map. In the telomeric region of the EDC, the high density of markers revealed deletions in all YACs carrying the neighboring loci *S100A1* and D1S3619. The identification of these rearrangements lead to a corrected order of the loci distal to *S100A6* in the initial contig. Interestingly, *TPM3,* which is involved in several oncogenic chromosomal aberrations (Butti et al. 1995; Lamant et al. 1999) maps close to this unstable region. In the majority of YACs, *TPM3* is a part of the deletion, suggesting that instability of YACs might concur with rearrangements in the human genome.

Sequence analyses of the genes newly assigned to the EDC indicated no overlap with the characteristic protein domains of the known EDC genes, such as repeat structures or calcium-binding sites (Mischke et al. 1996), and the putative functions of the encoded proteins have been mainly elucidated outside of the skin so far. For example, Hax-1 interacts with HS1, an actin cytoskeleton-associated protein, regulating clonal expansion and deletion of lymphoid cells (Suzuki et al. 1997), the RNA-specific adenosine deaminase is involved in modification of transcripts encoding glutamate and serotonin receptors in the central nervous system (Keller et al. 1999), and the yeast homolog of the 26S proteasome subunit p31 is necessary for the degradation of proteins regulating the cell cycle

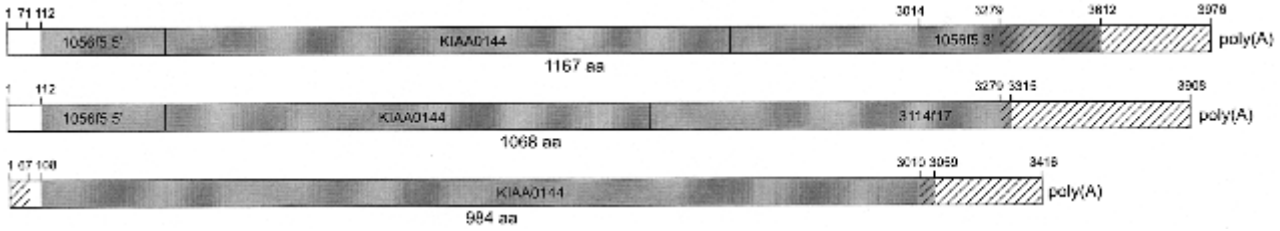

Figure 5 Alternative 3′-termini of the *NICE-4* gene products. Sequence information derived from cDNA clones 1056f 5, 3114f17, and cDNA KIAA0144 is represented by open boxes. Clone 3114f17 only contains the 3′-terminal sequence of the transcript and was filled up with overlapping sequences of clone 1056f 5 and cDNA KIAA0144. Nonhomologous regions are cross-hatched, ORFs are shaded gray. The first and last nucleotide of the putative coding sequence and of identical regions are indicated above, the size of the predicted protein below the corresponding transcripts.

(Kominami et al. 1995). However, for the homologs of TM30nm and of the 34/67-kD laminin receptor, which are also ubiquitously expressed, a role in epidermal cells has already been shown; both are involved in malignancy of tumors of the mouse skin (Tennenbaum et al. 1992; Miyado et al. 1996). Similar observations, a broad tissue distribution and specific functions in the epidermis, have been made for other EDC-encoded proteins, for example, certain members of the S100 family (Mischke et al. 1996; Schaefer and Heizmann 1996).

In the central region of the EDC, a novel transcript of the *SPRR3* gene was identified. Apart from the established sequence with 14 internal repeats of 24 nucleotides yielding a protein size of 169 amino acids (Gibbs et al. 1993), cDNA clones with 13 repeats and divergences in defined positions of the nucleotide sequence were isolated (Fig. 3). The encoded 161-amino-acid protein that, in addition, carries one substituted amino acid, most likely reflects a polymorphism in the singular *SPRR3* gene, similar to the variable number of repeats that have been described for the human involucrin gene (Simon et al. 1991). According to the evolution of the SPRR gene family by intra- and intergenic duplications (Gibbs et al. 1993), which suggests that the inner repeating units were generated most recently, the allele with 13 repeats would indicate an earlier origin.

The only novel gene that was mapped to this part of the EDC is *NICE-1*. Its localization between genes encoding the structural proteins filaggrin and involucrin together with its expression pattern similar to *SPRR2* suggests a role in terminal differentiation of the epidermis. This conclusion is supported by the predicted protein of *NICE-1;* the amino acid sequence contains several glutamine and lysine residues, characteristics of the transglutaminase substrates of the CE, and like loricrin it is serine, glutamine, and cysteine rich (Fig. 6). As a putative new precursor of the CE, it displays a weak similarity to another predicted protein from human skin, which is encoded by the *xp5* gene (Zhao and Elder 1997). This gene is expressed in normal and psoriatic skin, but not in undifferentiated keratinocytes, and it has also been mapped close to *IVL* and the *SPRR* genes (Zhao and Elder 1997). However, like profilaggrin, *xp5* has not yet been detected within the gridded keratinocyte cDNA library. Low abundance or even absence of the corresponding transcripts in the library could be attributed to gene expression restricted to the latest steps of epithelial differentiation, which might be unattainable in cultured keratinocytes. This could also be the reason for the low density of genes identified between *FLG* and *IVL*.

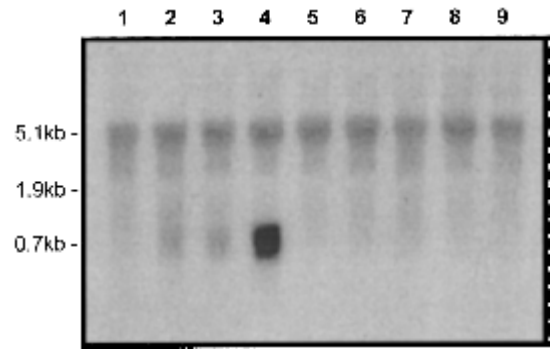

Figure 8 Northern blot analysis of *NICE-1*. A Northern blot containing RNA from different cell populations was hybridized with a specific probe for *NICE-1*. Total RNA (20 µg) in each lane was isolated from the following tissues or cultured cells: (1) human skin; (2) stripped human keratinocytes maintained in medium without calcium (Fischer et al. 1999); (3) stripped human keratinocytes maintained in a proliferative state in medium with 1.8 mM strontium (Praeger et al. 1987); (4) stripped human keratinocytes induced to differentiate in medium with 1.8 mM calcium; (5) HaCaT cells; (6) HeLa cells; (7) SCC4 cells; (8) primary human fibroblasts; (9) primary human melanocytes. The *top* band is due to cross-reaction with 28S ribosomal RNA.

It has been proposed that expression patterns of EDC genes change gradually from a broad tissue distribution and limited differentiation specificity in the telomeric region, including the *S100* genes, toward a strong tissue- and differentiation-specific expression in the more centromeric region, in which *THH* and *FLG* are located (Zhao and Elder 1997). The expression data for the novel EDC genes clearly support this conclusion. For instance, the *NICE-2, NICE-3, NICE-4,* and *NICE-5* genes, which extend the EDC toward the telo-

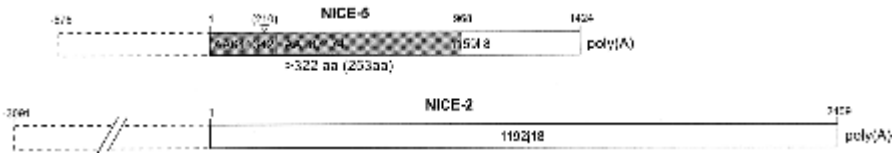

Figure 7 Sequence analysis of the *NICE-5* and *NICE-2* gene products. cDNA clones 11591 8 (*NICE-5*) (*top*) and 1192j18 (*NICE-2*) (*bottom*) contain the 3′-termini of the corresponding transcripts. Overlapping ESTs AA287174 and AA641342 were used to extend the *NICE-5* sequence. Sequence information derived from the respective clones is represented by open boxes; broken line boxes indicate lacking sequences as deduced from the approximate size of the entire mRNA, which was determined by Northern hybridization. The ORF of *NICE-5* (shaded gray) is likely to be incomplete. No significant ORF was detected within the *NICE-2* sequence. Numbering is relative to the start of the available sequence (+1). In the case of *NICE-5,* the last nucleotide of the putative coding sequence and the minimum protein size are indicated. The size of the protein resulting from the first initiation methionine codon within the available sequence and the position of the initial adenine (△) are in parenthesis.

Table 4. EST Analyses for the *NICE* Genes

Gene	UniGene EST-cluster	Map position (GeneMap'99)	EST sources
NICE-1	Hs.110196	D1S514-D1S2635	Heart
NICE-2	—	—	—
NICE-3	Hs.31989	D1S514-D1S2635	Adipose, adrenal gland, aorta, blood, brain, breast, colon, ear, eye, foreskin, gall bladder, germ cell, heart, kidney, liver, lung, lymph, muscle, ovary, pancreas, parathyroid, placenta, prostate, skin, stomach, synovial membrane, testis, tonsil, uterus, whole embryo
NICE-4	Hs.8127	D1S514-D1S2635	Adrenal gland, aorta, blood, bone, brain, breast, colon, ear, germ cell, heart, kidney, lung, lymph, omentum, ovary, pancreas, placenta, prostate, smooth muscle, stomach, testis, thymus, tonsil, uterus, whole embryo
NICE-5	Hs.107538	D1S514-D1S2635	Aorta, brain, breast, CNS, colon, eye, foreskin, gall bladder, germ cell, heart, kidney, lung, lymph, ovary, pancreas, placenta, prostate, stomach, testis, thyroid, tonsil, uterus, whole embryo

meric side, appear to be ubiquitously expressed at a low level in many cell populations. Furthermore, the *NICE-5* gene appears to be highly conserved from *C. elegans* to man. In contrast, the *NICE-1* gene, which we have mapped in the more centromeric region of the EDC, is strongly up-regulated in differentiated keratinocytes. Although it is possible that a locus control region determines appropriate expression of these genes during terminal differentiation, conserved regulatory elements within the *SPRR* gene family suggest a major control function for individual promotors in coordinated, differentiation-specific expression of the respective genes (Fischer et al. 1996, 1999; Sark et al. 1998).

The identification of 10 additional genes within the EDC that are expressed in keratinocytes strengthens the role of this genomic region as a gene complex (Mischke et al. 1996). Now harboring 37 genes within 3 Mb, the gene density is certainly still far from the >50 genes including pseudogenes/Mb of the human major histocompatibility complex (MHC) on chromosome 6 (The MHC sequencing consortium 1999), but in contrast to the MHC, most of the genes of the EDC participate in one major goal, the building-up of the epidermis as the major barrier between the human body and the environment. It can be expected that the number of EDC genes will still increase, as the whole centromeric part of the EDC, the region between *SPRR2A* and *S100A7* including the loricrin gene, *S100A9, S100A12,* and *S100A8,* as well as the deleted part of YAC 950_e_2#9 between *S100A1* and D1S3619 were excluded from our investigations. Moreover, further evaluation of the YAC-specific cDNAs might reveal additional genes of the examined region, like profilaggrin or rare transcripts that have not yet been recovered. Finally, the results reported here have shifted the telomeric border of the EDC toward 1q22, indicating that the extension of the EDC is so far unknown.

In conclusion, subtractive hybridization as described here is a powerful tool to isolate specific transcripts from large chromosomal regions, which is essential, for example, for the identification of disease genes that have been localized to an Mb-sized candidate region by lineage analyses. A small contig that includes the corresponding genetic marker and a gridded cDNA library from the affected tissue are the only tools necessary for the identification of specific transcripts after just two hybridization experiments. By applying this technique to the EDC, we succeeded in the identification of 23 different transcripts originating from 19 genes, 10 of which were assigned to this human gene complex for the first time.

METHODS

Construction of the Gridded Keratinocyte cDNA Library

RNA was isolated from human primary keratinocytes cultured in vitro with different calcium concentrations and at different degrees of confluence and stratification (Fischer et al. 1996). Poly(A)$^+$-RNA from different cultures was pooled and size fractionated on sucrose

gradients. Four libraries were prepared by use of oligo(dT)-primers for first strand cDNA synthesis and the high efficiency Lambda ZAP II cloning system (Stratagene). Subsequently, the cloned fragments were excised in vivo from the λ vector into the phagemid pBluescript. The original complexity of each library was ~10^6 clones. Libraries 1 and 2 (corresponding to larger inserts) were pooled as well as 3 and 4 (corresponding to smaller inserts) and both of the pools were processed separately. The libraries were gridded at the Ressourcenzentrum in Berlin, basically following the protocol of Nizetic and Lehrach (1995). A total of 184,320 cDNA clones were picked by a robot into 480 384-well microtiter plates. After replication, one copy of the library was used for spotting the clones onto 10 high-density filters. Microtiter plates containing a second copy of the library were stored at −80°C to retrieve positively identified clones for analyses.

YACs and DNA Probes

All YACs have been characterized extensively (Marenholz et al. 1996; Lioumi et al. 1998). YACs 950_e_2#1 to #11 represent single colonies from the unstable CEPH YAC culture 950_e_2. Specific markers for *S100A1, S100A2, S100A8, S100A9, S100A10, S100A11, S100A13, SPRR1B, SPRR2A, SPRR3, LOR, IVL, FLG, THH, CHRNB2,* D1S305, D1S1664, D1S2346, D1S3619, D1S3623, D1S3624, D1S3625, D1S3626, D1S3627, D1S3628, 37m2, 37m6, and 37m16 were the same as described previously (Engelkamp et al. 1992; Volz et al. 1993; Pedrocchi et al. 1994; Hudson et al. 1995; Marenholz et al. 1996; Mischke et al. 1996; Wicki et al. 1996a,b; Lueders et al. 1999; South et al. 1999). Cloned cDNA inserts or fragments of them used as probes that were isolated from the gridded library are listed in Table 5. Whole YAC DNA was isolated as described (Marenholz et al. 1996) for subtractive hybridization (YACs 986_e_10 and 950_e_2#9) and for generation of plasmid subclones 24f6, 24f15, 24f32, 24f39, and 24f57 (YAC 950_e_2#9).

Table 5. cDNA Fragments Used as Probes

Gene	Clone	*Eco*RI/*Xho*I fragment
S100A4	1007n11	327 bp
S100A6	1023f1	402 bp
S100A7	1010b14	425 bp
ADAR1	1007n 1	1530 bp
HAX1	3027d12	949 bp
LAMRL6	30381 4	1030 bp
PSMD8L	3017o23	923 bp
TPM3	3006f16	2108 bp
NICE-1	3042e22	728 bp
NICE-2	1192j18	1587 bp
NICE-3	3038m19	1586 bp
NICE-4	3114f17	1820 bp
NICE-5	1025p19	752 bp

Northern and Southern Hybridization

Total RNA isolation and Northern blotting (Lohman et al. 1997), preparation of Southern blots for genomic (H2LCL cell line), and YAC restriction mapping, labeling of probes, and hybridization (Volz et al. 1993; Marenholz et al. 1996) were performed essentially as described. Modified parameters for rotating-field gel electrophoresis (ROFE) (Ziegler and Volz 1992), were used for genomic restriction mapping; duration three times 25 h, interval (pulse time) 80–240 s log, angle 115–130° log, voltage 110–130 V log, temperature 12°C. To block nonspecific hybridization, ^{32}P-labeled DNA of YACs 986_e_10 and 950_e_2#9 was prehybridized extensively with 100 µg/mL sonicated total genomic human placenta DNA for 1 h at 66°C.

Sequencing

Sequencing reactions were performed using 200 ng of plasmid DNA/kb template from selected cDNA clones, 4 pmole of IRD-800 or IRD-700-labeled M13 (−20) or M13 reverse primers and the Thermo Sequenase fluorescent labeled primer cycle sequencing kit with 7-deaza-dGTP (Amersham) following the protocol provided by the manufacturer of the Li-Cor DNA sequencer model 4200. In general, sequence editing yielded a reliable sequence (>98% identity, as determined by alignment to known nucleotide sequences) for ~600 bp from each primer and for additional 200 bp with decreasing reliability. The continuous sequence of transcripts >1200 bp was determined by constructing contigs of overlapping 5′-terminal sequences that were derived from incomplete cDNAs originating from the same gene. Remaining gaps were filled with overlapping DNA sequences identified in databases by BLASTN search (see below). To rule out chimerism, several independent cDNA inserts of the same gene were sequenced. Applying this strategy yielded a single consensus sequence in most cases, but also indicated alternatively spliced forms.

Computational Analyses

The following uniform resource locators (URL) and subdirectories were used for sequence alignment, similarity and protein domain searches: http://www.toulouse.inra.fr/multalin.html (Corpet 1998); http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nphnewblast (Altschul et al. 1997); http://www.expasy.ch/tools/scnpsit1.html (Hofmann et al. 1999) (PROSITE patterns with a high probability of occurrence were excluded); http://www.isrec.isb-sib.ch/software/PFSCAN_form.html (PROSITE profiles, Pfam, and Gribskov collection, including weak matches).

ACKNOWLEDGMENTS

We thank Dr. Susan Gibbs for constructing the lambda-ZAP cDNA libraries and the Ressourcenzentrum Berlin for giving us the opportunity to prepare the gridded filters. The probe for *CHRNB2* was kindly provided by Dr. Kira Lueders prior to publication. The contribution of probes for *S100A1, S100A2, S100A11,* and *S100A13* by Drs. Claus W. Heizmann and Beat Schäfer and the contribution of ideas by Dr. Armin Volz is also appreciated. Furthermore, we are grateful to the reviewers of this manuscript for their helpful comments. This work was supported by the J.A. Cohen Institute (Leiden), by the Sonnenfeld-Stiftung (Berlin), and by a grant from the European Union (BMH4-CT96-0319).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Albertazzi, E., Cajone, F., Leone, B.E., Naguib, R.N., Lakshmi, M.S., and Sherbet, G.V. 1998. Expression of metastasis-associated genes h-mts1 (S100A4) and nm23 in carcinoma of breast is related to disease progression. *DNA Cell Biol* **17:** 335–342.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Backendorf, C. and Hohl, D. 1992. A common origin for cornified envelope proteins? *Nat. Genet.* **2:** 91.

Bignon, C., Roux-Dosseto, M., Zeigler, M.E., Mattei, M.G., Lissitzky, J.C., Wicha, M.S., and Martin, P.M. 1991. Genomic analysis of the 67-kDa laminin receptor in normal and pathological tissues: Circumstantial evidence for retroposon features. *Genomics* **10:** 481–485.

Boultwood, J., Fidler, C., Soularue, P., Strickson, A.J., Kostrzewa, M., Jaju, R.J., Cotter, F.E., Fairweather, N., Monaco, A.P., Muller, U., et al. 1997. Novel genes mapping to the critical region of the 5q-syndrome. *Genomics* **45:** 88–96.

Butti, M.G., Bongarzone, I., Ferraresi, G., Mondellini, P., Borrello, M.G., and Pierotti, M.A. 1995. A sequence analysis of the genomic regions involved in the rearrangements between TPM3 and NTRK1 genes producing TRK oncogenes in papillary thyroid carcinomas. *Genomics* **28:** 15–24.

Capon, F., Novelli, G., Semprini, S., Clementi, M., Nudo, M., Vultaggio, P., Mazzanti, C., Gobello, T., Botta, A., Fabrizi, G., et al. 1999. Searching for psoriasis susceptibility genes in Italy: Genome scan and evidence for a new locus on chromosome 1. *J. Invest. Dermatol.* **112:** 32–35.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16:** 10881–10890.

Eckert, R.L. and Green, H. 1986. Structure and evolution of the human involucrin gene. *Cell* **46:** 583–589.

Elvin, P., Slynn, G., Black, D., Graham, A., Butler, R., Riley, J., Anand, R., and Markham, A.F. 1990. Isolation of cDNA clones using yeast artificial chromosome probes. *Nucleic Acids Res.* **18:** 3913–3917.

Engelkamp, D., Schaefer, B.W., Erne, P., and Heizmann, C.W. 1992. S100 alpha, CAPL, and CACY: Molecular cloning and expression analysis of three calcium-binding proteins from human heart. *Biochemistry* **31:** 10258–10264.

Fietz, M.J., McLaughlan, C.J., Campbell, M.T., and Rogers, G.E. 1993. Analysis of the sheep trichohyalin gene: Potential structural and calcium-binding roles of trichohyalin in the hair follicle. *J. Cell Biol.* **121:** 855–865.

Fischer, D.F., Gibbs, S., van De Putte, P., and Backendorf, C. 1996. Interdependent transcription control elements regulate the expression of the SPRR2A gene during keratinocyte terminal differentiation. *Mol. Cell. Biol.* **16:** 5365–5374.

Fischer, D.F., Sark, M.W., Lehtola, M.M., Gibbs, S., van de Putte, P., and Backendorf, C. 1999. Structure and evolution of the human SPRR3 gene: Implications for function and regulation. *Genomics* **55:** 88–99.

Forus, A., Berner, J.M., Meza-Zepeda, L.A., Saeter, G., Mischke, D., Fodstad, O., and Myklebost, O. 1998. Molecular characterization of a novel amplicon at 1q21-q22 frequently observed in human sarcomas. *Br. J. Cancer* **78:** 495–503.

Gendler, S.J., Cohen, E.P., Craston, A., Duhig, T., Johnstone, G., and Barnes, D. 1990. The locus of the polymorphic epithelial mucin (PEM) tumour antigen on chromosome 1q21 shows a high frequency of alteration in primary human breast tumours. *Int. J. Cancer* **45:** 431–435.

Gibbs, S., Fijnemann, R., Wiegant, J., van Kessel, A.G., van de Putte, P., and Backendorf, C. 1993. Molecular characterization and evolution of the SPRR family of keratinocyte differentiation markers encoding small proline-rich proteins. *Genomics* **16:** 630–637.

Hardas, B.D., Zhao, X., Zhang, J., Longqing X., Stoll, S., and Elder, J.T. 1996. Assignment of psoriasin to human chromosomal band 1q21: Coordinate overexpression of clustered genes in psoriasis. *J. Invest. Dermatol.* **106:** 753–758.

Hisama, F.M., Oshima, J., Yu, C.E., Fu, Y.H., Mulligan, J., Weissman, S.M., and Schellenberg, G.D. 1998. Comparison of methods for identifying transcription units and transcription map of the Werner syndrome gene region. *Genomics* **52:** 352–357.

Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. 1999. The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27:** 215–219.

Hoggard, N., Brintnell, B., Howell, A., Weissenbach, J., and Varley, J. 1995. Allelic imbalance on chromosome 1 in human breast cancer. II. Microsatellite repeat analysis. *Genes, Chromosomes, Cancer* **12:** 24–31.

Hohl, D., Mehrel, T., Lichti, U., Turner, M.L., Roop, D.R., and Steinert, P.M. 1991. Characterization of human loricrin. Structure and function of a new class of epidermal cell envelope proteins. *J. Biol. Chem.* **266:** 6626–6636.

Hudson, T., Stein, L., Gerety, S., Ma, J., Castle, A., Silva, J., Slonim, D., Baptista, R., Kruglyak, L., Xu, S., et al. 1995. An STS-Based Map of the Human Genome. *Science* **270:** 1945–1954.

Ishida-Yamamoto, A., McGrath, J.A., Lam, H., Iizuka, H., Friedman, R.A., and Christiano, A.M. 1997. The molecular pathology of progressive symmetric erythrokeratoderma: A frameshift mutation in the loricrin gene and perturbations in the cornified cell envelope. *Am. J. Hum. Genet.* **61:** 581–589.

Jackers, P., Minoletti, F., Belotti, D., Clausse, N., Sozzi, G., Sobel, M.E., and Castronovo, V. 1996. Isolation from a multigene family of the active human gene of the metastasis-associated multifunctional protein 37LRP/p40 at chromosome 3p21.3. *Oncogene* **13:** 495–503.

Keller, W., Wolf, J., and Gerber, A. 1999. Editing of messenger RNA precursors and of tRNAs by adenosine to inosine conversion. *FEBS Lett.* **452:** 71–76.

Kominami, K., DeMartino, G., Moomaw, C., Slaughter, C., Shimbara, N., Fujimuro, M., Yokosawa, H., Hisamatsu, H., Tanahashi, N., and Shimizu, Y. 1995. Nin1p, a regulatory subunit of the 26S proteasome, is necessary for activation of Cdc28p kinase of Saccharomyces cerevisiae. *EMBO J.* **14:** 3105–3115.

Korge, B.P., Ishida-Yamamoto, A., Punter, C., Dopping-Hepenstal, P.J., Iizuka, H., Stephenson, A., Eady, R.A., and Munro, C.S. 1997. Loricrin mutation in Vohwinkel's keratoderma is unique to the variant with ichthyosis. *J. Invest. Dermatol.* **109:** 604–610.

Lamant, L., Dastugue, N., Pulford, K., Delsol, G., and Mariame, B. 1999. A new fusion gene TPM3-ALK in anaplastic large cell lymphoma created by a (1;2)(q25;p23) translocation. *Blood* **93:** 3088–3095.

Lee, S.C., Kim, I.G., Marekov, L.N., O'Keefe, E.J., Parry, D.A.D., and Steinert, P.M. 1993. The structure of human trichohyalin: Potential multiple functions as an EF-hand-like calcium binding protein, a cornified cell envelope precursor and an intermediate filament associated (crosslinking) protein. *J. Mol. Biol.* **268:** 12164–12176.

Lee, S.W., Tomasetto, C., Swisshelm, K., Keyomarsi, K., and Sager, R. 1992. Down-regulation of a member of the S100 gene family in mammary carcinoma cells and reexpression by azadeoxycytidine treatment. *Proc. Natl. Acad. Sci.* **89:** 2504–2508.

Lioumi, M., Olavesen, M.G., Nizetic, D., and Ragoussis, J. 1998. High-resolution YAC fragmentation map of 1q21. *Genomics* **49:** 200–208.

Lohman, F.P., Medema, J.K., Gibbs, S., Ponec, M., van de Putte, P., and Backendorf, C. 1997. Expression of the SPRR cornification genes is differentially affected by carcinogenic transformation. *Exp. Cell Res.* **231:** 141–148.

Lueders, K.K., Elliott, R.W., Marenholz, I., Mischke, D., DuPree, M., and Hamer, D. 1999. Genomic organization and mapping of the human and mouse neuronal β2-nicotinic acetylcholine receptor genes. *Mamm. Genome* **10:** 900–905.

MacLeod, A.R., Houlker, C., Reinach, F.C., and Talbot, K. 1986. The mRNA and RNA-copy pseudogenes encoding Tm30nm, a human cytoskeletal tropomyosin. *Nucleic Acids Res.* **14:** 8413–8426.

Maelandsmo, G.M., Florenes, V.A., Mellingsaeter, T., Hovig, E., Kerbel, R.S., and Fodstad, O. 1997. Differential expression patterns of S100A2, S100A4 and S100A6 during progression of human malignant melanoma. *Int. J. Cancer* **74:** 464–469.

Maestrini, E., Monaco, A.P., McGrath, J.A., Ishida-Yamamoto, A., Camisa, C., Hovnanian, A., Weeks, D.E., Lathrop, M., Uitto, J., and Christiano, A.M. 1996. A molecular defect in loricrin, the major component of the cornified cell envelope, underlies Vohwinkel's syndrome. *Nature Genet.* **13:** 70–77.

Marenholz, I., Volz, A., Ziegler, A., Davies, A., Ragoussis, I., Korge, B.P., and Mischke, D. 1996. Genetic analysis of the epidermal differentiation complex (EDC) on human chromosome 1q21: Chromosomal orientation, new markers, and a 6-MB YAC contig. *Genomics* **37:** 295–302.

Markova, N.G., Marekov, L.N., Chipev, C.C., Gan, S.-Q., Idler, W.W., and Steinert, P.M. 1993. Profilaggrin is a major epidermal calcium binding protein. *Mol. Cell. Biol.* **13:** 613–625.

MHC Sequencing Consortium. 1999. Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401:** 921–923.

Mischke, D., Korge, B.P., Marenholz, I., Volz, A., and Ziegler, A. 1996. Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex ("epidermal differentiation complex") on human chromosome 1q21. *J. Invest. Dermatol.* **106:** 989–992.

Miyado, K., Kimura, M., and Taniguchi, S. 1996. Decreased expression of a single tropomyosin isoform, TM5/TM30nm, results in reduction in motility of highly metastatic B16-F10 mouse melanoma cells. *Biochem. Biophys. Res. Commun.* **225:** 427–435.

Moog-Lutz, C., Bouillet, P., Regnier, C.H., Tomasetto, C., Mattei, M.G., Chenard, M.P., Anglard, P., Rio, M.C., and Basset, P. 1995. Comparative expression of the psoriasin (S100A7) and S100C genes in breast carcinoma and co-localization to human chromosome 1q21-q22. *Int. J. Cancer.* **63:** 297–303.

Nagase, T., Seki, N., Tanaka, A., Ishikawa, K., and Nomura, N. 1995. Prediction of the coding sequences of unidentified human genes. IV. The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res.* **2:** 167–174.

Nirunsuksiri, W., Presland, R.B., Brumbaugh, S.G., Dale, B.A., and Fleckman, P. 1995. Decreased profilaggrin expression in ichthyosis vulgaris is a result of selectively impaired posttranscriptional control. *J. Biol. Chem.* **270:** 871–876.

Nizetic, D. and Lehrach, H. 1995. Gridding and handling the cosmid libraries. In *DNA cloning 3: A practical approach,* 2nd ed. (eds. D.M. Glover and B.D. Hames), pp. 60–67. Oxford University Press, Oxford, UK.

Pedrocchi, M., Schaefer, B.W., Mueller, H., Eppenberger, U., and Heizmann, C.W. 1994. Expression of Ca^{2+}-binding proteins of the S100 family in malignant human breast-cancer cell lines and biopsy samples. *Int. J. Cancer* **57:** 684–690.

Praeger, F.C., Stanulis-Praeger, B.M., and Gilchrest, B. 1987. Use of strontium to separate calcium-dependent pathways for proliferation and differentiation in human keratinocytes. *J. Cell. Physiol.* **132:** 81–89.

Presland, R.B., Bassuk, J.A., Kimball, J.R., and Dale, B.A. 1995. Characterization of two distinct calcium-binding sites in the amino-terminus of human profilaggrin. *J. Invest. Dermatol.* **104:** 218–223.

Rammes, A., Roth, J., Goebeler, M., Klempt, M., Hartmann, M., and Sorg, C. 1997. Myeloid-related protein (MRP) 8 and MRP14, calcium-binding proteins of the S100 family, are secreted by activated monocytes via a novel, tubulin-dependent pathway. *J. Biol. Chem.* **272:** 9496–9502.

Remppis, A., Greten, T., Schaefer, B.W., Hunziker, P., Erne, P., Katus, H.A., and Heizmann, C.W. 1996. Altered expression of the Ca^{2+}-binding protein S100A1 in human cardiomyopathy. *Biochim. Biophys. Acta* **1313:** 253–257.

Richardson, M.P., Braybrook, C., Tham, M., Moore, G.E., and Stanier, P. 1998. Molecular cloning and characterization of a highly conserved human 67-kDa laminin receptor pseudogene mapping to Xq21.3. *Gene* **206:** 145–150.

Robinson, N.A., Lapic, S., Welter, J.F., and Eckert, R.L. 1997. S100A11, S100A10, annexin I, desmosomal proteins, small proline-rich proteins, plasminogen activator inhibitor-2, and involucrin are components of the cornified envelope of cultured human epidermal keratinocytes. *J. Biol. Chem.* **272:** 12035–12046.

Robinson, P.A., Marley, J.J., High, A.S., and Hume, W.J. 1994. Differential expression of protease inhibitor and small proline-rich protein genes between normal human oral tissue and odontogenic keratocysts. *Arch. Oral Biol.* **39:** 251–259.

Sark, M.W., Fischer, D.F., de Meijer, E., van de Putte, P., and Backendorf, C. 1998. AP-1 and ets transcription factors regulate the expression of the human SPRR1A keratinocyte terminal differentiation marker. *J. Biol. Chem.* **273:** 24683–24692.

Schaefer, B.W. and Heizmann, C.W. 1996. The S100 family of EF-hand calcium-binding proteins: Functions and pathology. *Trends Biochem. Sci.* **21:** 134–140.

Schaefer, B.W., Wicki, R., Engelkamp, D., Mattei, M.G., and Heizmann, C.W. 1995. Isolation of a YAC clone covering a cluster of nine S100 genes on human chromosome 1q21: Rationale for a new nomenclature of the S100 calcium-binding protein family. *Genomics* **25:** 638–643.

Simon, M., Phillips, M., and Green, H. 1991. Polymorphism due to variable number of repeats in the human involucrin gene. *Genomics* **9:** 576–580.

South, A.P., Cabral, A., Ives, J.H., James, C.H., Mirza, G., Marenholz, I., Mischke, D., Backendorf, C. Ragoussis, J., and Nizetic, D. 1999. Human Epidermal Differentiation Complex in a single 2.5 Mbp long continuum of overlapping DNA cloned in bacteria integrating physical and transcript maps. *J. Invest. Dermatol.* **112:** 910–918.

Steinert, P.M. and Marekov, L.N. 1995. The proteins elafin, filaggrin, keratin intermediate filaments, loricrin, and small proline-rich proteins 1 and 2 are isodipeptide cross-linked components of the human epidermal cornified cell envelope. *J. Biol. Chem.* **270:** 17702–17711.

Steinert, P.M., Kartasova, T., and Marekov, L.N. 1998. Biochemical evidence that small proline-rich proteins and trichohyalin function in epithelia by modulation of the biomechanical properties of their cornified cell envelopes. *J. Biol. Chem.* **273:** 11758–11769.

Steven, A.C., Bisher, M.E., Roop, D.R., and Steinert, P.M. 1990. Biosynthetic pathways of filaggrin and loricrin—two major proteins expressed by terminally differentiated epidermal keratinocytes. *J. Struct. Biol.* **104:** 150–162.

Suzuki, Y., Demoliere, C., Kitamura, D., Takeshita, H., Deuschle, U., and Watanabe, T. 1997. HAX-1, a novel intracellular protein, localized on mitochondria, directly associates with HS1, a substrate of Src family tyrosine kinases. *J. Immunol.* **158:** 2736–2744.

Tennenbaum, T., Yuspa, S.H., Grover, A., Castronovo, V., Sobel,

M.E., Yamada, Y., and De Luca, L.M. 1992. Extracellular matrix receptors and mouse skin carcinogenesis: Altered expression linked to appearance of early markers of tumor progression. *Cancer Res.* **52:** 2966–2976.

Volz, A., Korge, B.P., Compton, J.G., Ziegler, A., Steinert, P.M., and Mischke, D. 1993. Physical mapping of a functional cluster of epidermal differentiation genes on chromosome 1q21. *Genomics* **18:** 92–99.

Vos, H.L. 1997. Characterization of ESTs mapped to the Whitehead Institute's RH map of chromosome 1q21 (Release 2.0). http://linkage.rockefeller.edu/chr1/data/1q21est/index.shtml or http://sunny.ebi.ac.uk/EBI/Mapping/Chr1/data/1q21est/index.shtml.

Weier, H.U., George, C.X., Greulich, K.M., and Samuel, C.E. 1995. The interferon-inducible, double-stranded RNA-specific adenosine deaminase gene (DSRAD) maps to human chromosome 1q21.1-21.2. *Genomics* **30:** 372–375.

Weterman, M.A.J., Wilbrink, M., and Geurts van Kessel, A. 1996. Fusion of the transcription factor TFE3 gene to a novel gene, PRCC, in t(X;1)(p11;q21)-positive papillary renal cell carcinomas. *Proc. Natl. Acad. Sci.* **93:** 15294–15298.

Wicki, R., Schaefer, B.W., Erne, P., and Heizmann, C.W. 1996a. Characterization of the human and mouse cDNAs coding for S100A13, a new member of the S100 protein family. *Biochem. Biophys. Res. Commun.* **227:** 594–599.

Wicki, R., Marenholz, I., Mischke, D., Schaefer, B.W., and Heizmann, C.W. 1996b. Characterization of the human S100A12 (calgranulin C, p6, CAAF1, CGRP) gene, a new member of the S100 gene cluster on chromosome 1q21. *Cell Calcium* **20:** 459–464.

Wilton, S.D., Eyre, H., Akkari, P.A., Watkins, H.C., MacRae, C., Laing, N.G., and Callen, D.C. 1995. Assignment of the human alpha-tropomyosin gene TPM3 to 1q22-q23 by fluorescence in situ hybridization. *Cytogenet. Cell Genet.* **68:** 122–124.

Zhao, X.P. and Elder, J.T. 1997. Positional cloning of novel skin-specific genes from the human epidermal differentiation complex. *Genomics* **45:** 250–258.

Ziegler, A. and Volz, A. 1992. Rotating field gel electrophoresis (ROFE). In *Methods in molecular biology: Pulsed field gel electrophoresis* (eds. M. Burmeister and L. Ulanovsky), Vol. 12, pp. 63–72. Humana Press, Totowa, NJ.

Received June 15, 2000; accepted in revised form December 21, 2000.