Human Epidermal Differentiation Complex in a Single 2.5 Mbp Long Continuum of Overlapping DNA Cloned in Bacteria Integrating Physical and Transcript Maps

Andrew P. South, Adriana Cabral,* Jane H. Ives, Colin H. James, Ghazala Mirza,† Ingo Marenholz,‡ Dietmar Mischke,‡ Claude Backendorf,* Jiannis Ragoussis,† and Dean Nizetic

Center for Applied Molecular Biology, School of Pharmacy, University of London, London, U.K.; *Department of Molecular Genetics, Leiden Institute of Chemistry, University of Leiden, The Netherlands; †St Thomas' Hospitals (UMDS), Guy's Hospital, London, U.K.; ‡Institute for Immunogenetics Charite, Humbolt University, Berlin, Germany

Terminal differentiation of keratinocytes involves the sequential expression of several major proteins which can be identified in distinct cellular layers within the mammalian epidermis and are characteristic for the maturation state of the keratinocyte. Many of the corresponding genes are clustered in one specific human chromosomal region 1q21. It is rare in the genome to find in such close proximity the genes belonging to at least three structurally different families, yet sharing spatial and temporal expression specificity, as well as interdependent functional features. This DNA segment, termed the epidermal differentiation complex, contains 27 genes, 14 of which are specifically expressed during calcium-dependent terminal differentiation of keratinocytes (the majority being structural protein precursors of the cornified envelope) and the other 13 belong to the S100 family of calcium binding proteins with possible signal transduction roles in the differentiation of epidermis and other tissues. In order to provide a bacterial clone resource that will enable further studies of genomic structure, transcriptional regulation, function and evolution of the epidermal differentiation complex, as well as the identification of novel genes, we have constructed a single 2.45 Mbp long continuum of genomic DNA cloned as 45 p1 artificial chromosomes, three bacterial artificial chromosomes, and 34 cosmid clones. The map encompasses all of the 27 genes so far assigned to the epidermal differentiation complex, and integrates the physical localization of these genes at a high resolution on a complete *Not*I and *Sal*I, and a partial *Eco*RI restriction map. This map will be the starting resource for the large-scale genomic sequencing of this region by The Sanger Center, Hinxton, U.K. ***Key words: 1q21/cornified envelope/gene complex/sequence-ready map. J Invest Dermatol 112:910–918, 1999***

Terminal differentiation in stratified squamous epithelia involves the sequential expression of several major proteins which can be identified in four distinct cellular layers within the mammalian epidermis and are characteristic for the maturation state of the keratinocyte (Resing and Dale, 1991). Terminal differentiation of keratinocytes involves the cessation of mitotic activity and the migration of cells through the four layers with the final differentiation step being the formation of the cornified cell envelope (CE), a 15 nm thick, highly insoluble structure formed on the inner surface of the plasma membrane (Williams and Elias, 1993). The CE is assembled by keratinocyte transglutaminase-catalyzed cross-linking of stratum corneum precursor proteins. At the same time, intermediate filament-associated proteins such as filaggrin and trichohyalin are activated from their precursors and lead to dense packing of keratin filaments inside cornified cells (Steinert and Roop, 1988; Dale *et al*, 1994). Genes encoding the majority of the above mentioned structural precursor proteins are clustered in a stretch of genomic DNA at 1q21 corresponding to approximately 1% of the length of the largest human chromosome. This DNA region, named the epidermal differentiation complex (EDC) (Mischke *et al*, 1996) harbors 27 hitherto identified genes, including genes for many of the CE precursor proteins, such as involucrin (Eckert and Green, 1986), loricrin (Hohl *et al*, 1991), and 10 members of the three subfamilies of small proline-rich proteins (SPRR) (Gibbs *et al*, 1993), intermediate filament-associated protein precursors profilaggrin and trichohyalin (Markova *et al*, 1993; O'Kefee *et al*, 1993), as well as 13 genes of the S100 family (S100A1–13) (Schaefer and Heizmann, 1996; Wicki *et al*, 1996a, b), encoding small calcium-binding proteins containing two EF-hand motifs with likely roles in calcium-mediated signaling during cell cycle progression and/or cellular differentiation. Co-localization of genes encoding calcium binding proteins with those encoding for structural proteins expressed during terminal epidermal differentiation is particularly interesting as calcium levels tightly control not only epidermal and general epithelial cell differentiation, but also the expression of specific genes encoding structural epidermal proteins (Yuspa *et al*, 1989; Presland *et al*, 1995). An intriguing

Manuscript received November 5, 1998; revised January 14, 1999; accepted for publication March 7, 1999.

Reprint requests to: Dr. D. Nizetic, Center for Applied Molecular Biology, School of Pharmacy, University of London, 29/39 Brunswick Square, London WC1N 1AX, U.K.

Abbreviations: BAC, bacterial artificial chromosome; EDC, epidermal differentiation complex; MTP, minimal tiling path; PAC, p1 artificial chromosome; SPRR, small proline rich proteins; YAC, yeast artificial chromosome.

hypothesis is that the spatial and temporal expression of structural genes in the EDC could be tightly and coordinately regulated, possibly by shared locus control elements operating at the level of the whole gene complex (Townes and Behringer, 1990; Hardas *et al*, 1996; Mischke *et al*, 1996), as has been shown for other gene complexes such as the globin locus (Dillon and Grosveld, 1993).

Many of these 27 genes within all three families have been implicated, or are known to be directly involved, in a number of wide-ranging disorders. Vohwinkel's keratoderma, a rare dominantly inherited genodermatosis, was the first disease directly related to a mutation in an EDC gene – a frameshift and delayed termination codon in loricrin (Maestrini *et al*, 1996; Korge *et al*, 1997). Autosomal recessive pycnodysostosis has also been mapped to 1q21 (Gelb *et al*, 1996; Polymeropoulos *et al*, 1995). Disorders of keratinization, such as various types of ichthyosis are marked with a drastic decrease in profilaggrin expression and/or abnormal CE (Resing and Dale, 1991), whereas S100A7, S100A8, and S100A9 (formerly known as psoriasin, calgranulin A, and calgranulin B, respectively), have been found highly upregulated in psoriatic epidermis (Hardas *et al*, 1996). Moreover, S100 proteins have been associated with tumor development and the metastatic behavior of breast cancer cells (Englekamp *et al*, 1993). In addition, amplification of the region 1q21–q22 has been detected in sarcomas (Forus *et al*, 1998). The breakpoint of a papillary renal cell carcinoma associated translocation (X;1) has been mapped to about S100A4, with the presence of related sequence in the region being suggested as a cause (Weterman *et al*, 1993).

So far, a long range physical map of the EDC using genomic DNA Southern blots, as well as dense yeast artificial chromosome (YAC) and YAC-fragmentation maps have been constructed (Marenholz *et al*, 1996; Mischke *et al*, 1996; Zhao and Elder, 1997; Lioumi *et al*, 1998). YAC clones, although very useful in bridging long distances, frequently suffer from chimerism, deletions, and rearrangements making them imperfect for reliable gene detection and unsuitable for high resolution, faithful, "sequence ready" representation of genomic DNA (Cheng *et al*, 1994; Groet *et al*, 1998).

Here we present a single continuum of 2.45 Mbp of overlapping genomic DNA fragments (contig) cloned in bacterial vectors, spanning the entire known EDC. The map integrates the localization of all 27 genes so far placed within the EDC on a high resolution restriction map of P1 (PAC), bacterial artificial chromosome (BAC), and cosmid clones. This map provides an appropriate set of DNA clones to be used as tools for the detailed expression studies of single genes and smaller groups of genes within the EDC using PAC/BAC retrofitting technology in the transient transfection studies (Mejia and Monaco, 1997; Kim *et al*, 1998). It also represents a resource for the positional derivation of potentially many additional genes, as well as promoters and regulatory elements of single genes and gene families, and for the studies aimed at detecting shared transcriptional locus control elements. The presented map will also be the main substrate for the long-range nucleotide sequence determination of genomic DNA from this region of chromosome 1, as agreed by The Sanger Center, Hinxton, U.K.

MATERIALS AND METHODS

Libraries used Cosmid ICRFc112 flow-sorted chromosome 1 library (Nizetic *et al*, 1994). PAC Library RPCI-1 (Ioannou *et al*, 1994). Total Human BAC library (Shizuya *et al*, 1992) (Research Genetics).

PAC/BAC insert isolation and hybridizations Hybridization were carried out as described by Baxendale *et al* (1991) with some modifications. PAC/BAC clones were digested at 37°C with *Not*I (NEB). Pulse field gel electrophoresis was carried out using a CHEF-DR II system (Biorad) according to manufacturers recommendations, with 0.5 × TBE (10 × TBE: 108 g Tris base, 55 g boric acid, 40 ml 0.5 M ethylene diamine tetraacetic acid (pH 8.0) per litre), and 1% low melting point agarose (Life Technologies). Electrophoresis conditions for the 4–200 kbp window: 5.2 V per cm, 14 h, switch time 3–15 s. DNA size markers used: Lambda ladder (Biorad), high molecular weight DNA markers (Life Technologies), and 1 kb DNA ladder (Life Technologies).

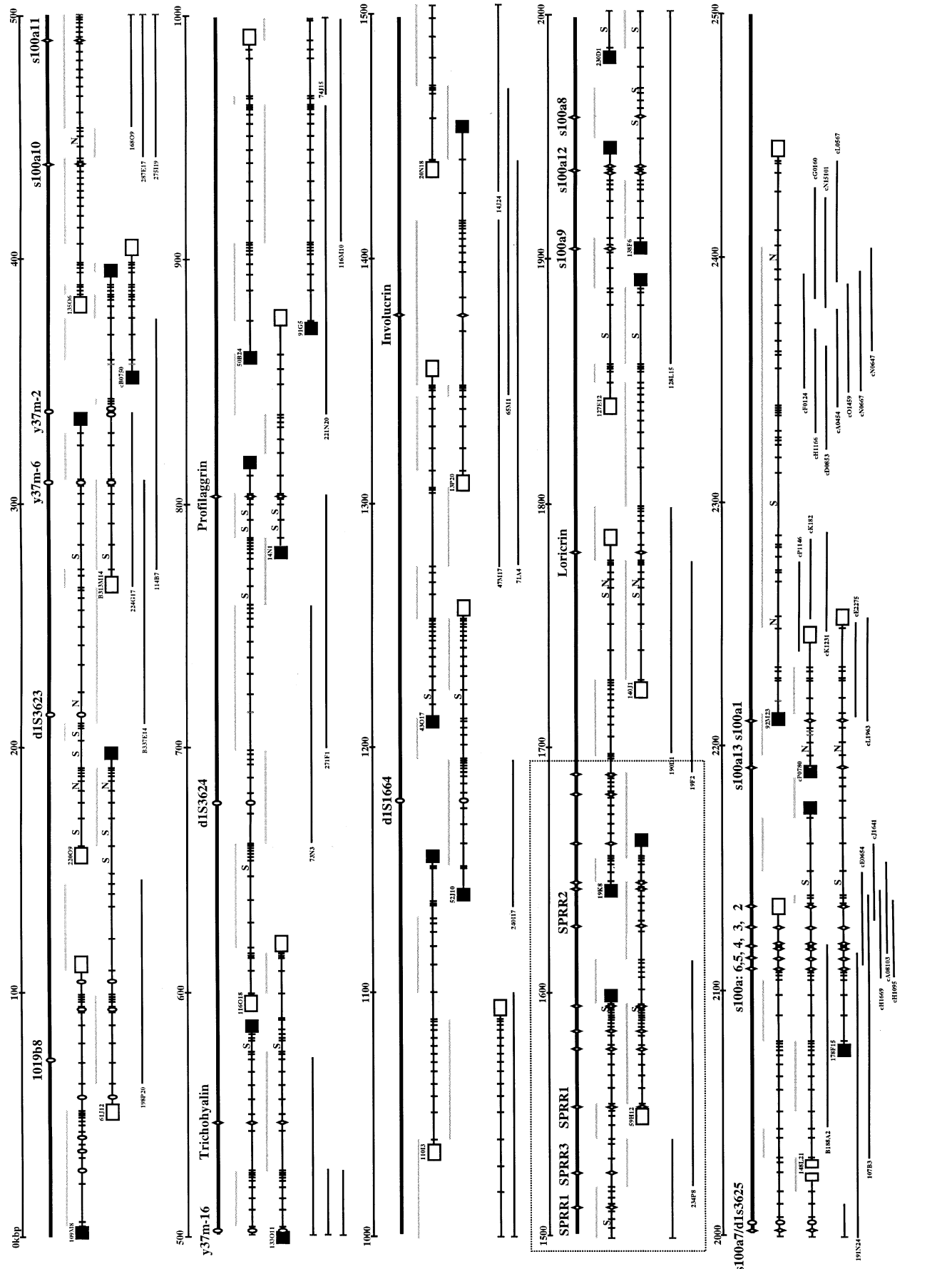
Inserts were cut out of the gel and weighed. To an average slice that weighed 250 mg, 150 μl of H_2O, 40 μl of 1 M NaCl, and 8 μl of 0.5 M ethylene diamine tetraacetic acid (pH 8.0) were added. The agarose slice was dissolved at 68°C for 15 min, followed by incubation at 37°C for 15 min. Two microliters of β-Agarase I was added (Calbiochem 1000 U per ml) and incubated at 37°C for 14 h. After phenol and chloroform: iso-amylalcohol (24:1) extraction, 360 μl of sample was precipitated with 5.4 μl 10 mg per ml Dextrane T40, 95 μl of 1 M NaCl, and 1062 μl 100% ethanol. The pellet was dissolved in 20 μl H_2O. YAC DNA was isolated as described previously (Nizetic and Lehrach, 1995). Picking and spotting of PAC clones, generation of cosmid insert probes and Southern blotting were carried out as recently described (Groet *et al*, 1998). Fluorescence *in situ* hybridizations were carried out as described previously (Marenholz *et al*, 1996)

Pool construction and pool polymerase chain reaction (PCR) One set of duplicated microtiter dishes containing the 1q21 PAC sublibrary was divided in two. One half of the well volume was frozen after adding glycerol whereas the other half was pooled in following way: The wells in columns from plate 1 were pooled and labeled pools 1–12 whereas the remaining nine clones in plate 2 were pooled as one and labeled 13. Glycerol was added to constitute 25% and the resulting volumes were aliquotted and frozen down. *Escherichia coli* cells were used as templates in PCR reactions performed using amplimers from the 3′ untranslated regions of the following genes: trichohyalin, profilaggrin, involucrin S100A10, S100A9, S100A8, and S100A6 as given in the Genome Database (Johns Hopkins University, Baltimore). PCR conditions are described below except 2 μl of cells from the PAC pools were used as a template in the first round of screening, whereas cells scraped from individual frozen microtiter plate wells were used in conjunction with 2 μl of sterile water as a template for the final round of screening. The PCR products from these reactions were run on appropriate agarose gels. The bands were purified with Prep-A-Gene (BioRad) according to the manufacturer's specifications and used in subsequent hybridization experiments.

Cultured keratinocyte library construction cDNA clone 1019b8 (**Fig 1**) has been detected using the YAC 764_a_1 as a hybridization probe against high density membranes of a cultured keratinocyte library. Poly(A) + RNA isolated from primary human keratinocytes cultured according to Fischer *et al* (1996) in both proliferating and differentiating culture conditions was size fractionated on a sucrose gradient. Four pooled fractions ranging from larger than 5 kb to smaller than 1500 bp were reversed transcribed with SuperScript II RNase H-reverse transcriptase (Gibco BRL) and cloned in the Uni-ZAP XR lambda vector (Stratagene). The primary complexity of these libraries varied from 1.3 to 2 million plaques. Equal plaque-forming units of libraries 1 and 2 and libraries 3 and 4 were mixed before *in vivo* excision with ExAssist helper phage (M13). Single colonies of the two resulting pBluescript SK(–) phagemid libraries (in strain SOLR) were scraped from agar plates and stored at –80°C before picking and filter spotting. Gridding of the library into microtiter plates and the production of high density membranes for screening will be discussed elsewhere (Marenholz *et al* in preparation).

cDNA and sequence tagged site isolation Plasmid containing cDNA representing the loricrin gene was kindly donated by Daniel Hohl. Clones from the cultured keratinocyte cDNA library were used to isolate probes for S100A11, S100A4, S100A7, and 1019b8. IMAGE consortium clones 153992, 398802, 377441, and 342548 were obtained from the HGMP resource center and used as probes for S100A1, S100A2, S100A3, and S100A12, respectively. We designed PCR primers for S100A1, S100A5, and S100A13**:** A1For: AGACAGCCACATTGGGCAGCGC; A1Rev: GATGAGTTGCAGGCTTGGACCGC; A5For: GCATCGATGACTTGATGAAG; A5rev: CTGCGATGGAAACTTTATTTC; A13For: GAAGTAAAGCCGCCTGGCTGAGATG; A13Rev: GAGGAAGCTTTATTTGGGAAGAGTGCG.

Conditions as given below except for the case of the A1 primers where 10 ng of the clone ICRFc112P0780 was used as starting material. Subcloned portions of YAC 764_a_1 were used for probes representing Y37 m-2, Y37 m-6, and Y37 m-16. Clones containing d1S3623 and d1S3624 were used as probes. Probes for the sequence tagged site markers d1S3625 and d1S1664 were produced by PCR with amplimers as given in Marenholz *et al*, 1996 and the Genome Database, respectively. Twenty nanograms of total human placental DNA (Sigma) was used as starting material for the reactions. PCR was carried out under standard conditions for 32 cycles (95°C, 60 s; 60°C, 60 s; 72°C, 120 s). The initial denaturation was 5 min at 95°C, and the final extension was 7 min at 72°C.


0kbp
1019b8
d1S3623
y37m-6
y37m-2
s100a11
s100a10
y37m-16
Trichohyalin
d1S3624
Profilaggrin
d1S1664
Involucrin
SPRR1 SPRR3 SPRR1
SPRR2
Loricrin
s100a9
s100a12
s100a8
s100a: 6,5, 4, 3, 2
s100a13 s100a1
s100a7/d1s3625

SPRR cosmid mapping Membranes containing a high density gridded chromosome 1 specific cosmid library, constructed in Lawrist 4 (Nizetic *et al*, 1994) were screened with a mixture of *SPRR1*, *SPRR2*, and *SPRR3* subclass-specific probes (Gibbs *et al*, 1993). Overlapping clones were detected by cosmid walking, using labeled RNA end probes, synthesized with either SP6 or T7 RNA polymerase and [^{32}P]-UTP (Amersham) essentially as described (Ivens and Little, 1995). Restriction mapping was performed on cosmids linearized either with *Nru*I or *Cpo*I (which did not cut inserts) by partial digestion with *Eco*RI. Restriction fragments were separated by pulse field gel electrophoresis (Gibbs *et al*, 1993), in the presence of DNA size markers and detected by either one of the two labeled end-probe oligonucleotides (see below), which allowed the mapping of each insert in both directions. Cosmids, totally digested with the above mentioned restriction enzymes, were hybridized with either subclass-specific *SPRR* probes or, as a control, with the respective cosmid DNA, labeled by random priming.

The following primers were labeled with T4 polynucleotide kinase and γ^{32}P-ATP and were used as end-probes: bio-1095: TTCAGCTGC-TGCCTGAGGCTGGACGACCTCGCGG and bio-1096: TTCCACC-ATGATATTCGGCAAGCAGGCATCGCCA.

RESULTS

Bacterial clone identification Four YAC clones from the previously published 6 Mb contig of the region (Marenholz *et al*, 1996) were used as starting material. The YAC clones used were 764_a_1, 907_e_6, 955_e_11, and 950_e_2; of which 764_a_1 and 950_e_2 had been previously mapped to 1q21.3 by two color fluorescence *in situ* hybridization on extended metaphase chromosomes displaying the given orientation (Marenholz *et al*, 1996). Artificial chromosomes (YAC) were separated from the host yeast chromosomes by excision from the gel after pulse-field gel electrophoresis, their DNA purified and used as a hybridization probe against a library representing the whole human genome 5-fold, cloned as PAC in *E. coli* (Ioannou *et al*, 1994) and displayed on high-density membranes. After eliminating PAC clones which hybridized to unrelated YAC clones from other regions of the human genome (J. Groet, personal observation) a total of 105 PAC clones, grouped in seven pockets (determined by their pattern of hybridization to individual overlapping YAC clones), were regrid-ded in two 96 well microtiter plates as a sublibrary of PAC clones enriched for the 1q21 region. These plates were duplicated and spotted on to nylon membranes where they were grown as colonies, processed, and used as an immobilized DNA target for subsequent hybridization experiments. In addition, *E. coli* cells from one of the duplicate sublibrary sets of plates were used to provide template DNA in PCR screening, as described in *Materials and Methods*.

Genes and sequence tagged site markers already placed on to YAC clones and genomic DNA of the region (Volz *et al*, 1993; Marenholz *et al*, 1996) were assigned to individual PAC clones within the 1q21 sublibrary using hybridization (with the 1q21 sublibrary membranes as targets) and/or PCR (using the *E. coli* cells as templates). Purified DNA of inserts from positive PAC clones (*Materials and Methods*) were used as probes in further rounds of hybridization experiments to the gridded PAC sublibrary membranes. This detected further overlapping clones from the sublibrary which were digested with *Eco*RI restriction enzyme, resolved on agarose gels, transferred to membranes by Southern blotting and rehybridized with PAC clone insert DNA probes. This helped establish the initial sets of overlapping PAC clones (PAC contigs) and identified the outermost (gap-facing) clones from initial contigs. The insert DNA from the gap facing clones was used in additional rounds of hybridization experiments to the sublibrary membranes. Remaining PAC clones from the 1q21 sublibrary that had not been recognized by established markers or characterized PAC clones were then randomly selected and their insert DNA isolated and hybridized back to the 1q21 PAC sublibrary membranes. Where no further clones within the sublibrary could be identified, inserts of gap facing PAC clones were hybridized back to the main PAC library (Ioannou *et al*, 1994), a total human BAC library (Shizuya *et al*, 1992) and a chromosome 1 specific cosmid library (Nizetic *et al*, 1994).

Fluorescent *in situ* hybridizations have been performed using the following six PAC clones: 61 J12, 135O6, 14 N1, 43O17, 71 A4, and 148L21 confirming their specific localization to the 1q21 region (results not shown).

Figure 2 summarizes the results of the hybridization and PCR experiments performed to define the map. Owing to spatial constraints all probe target combinations have not been shown. For easier interpretation of the results, we will give an example. Looking at the probes (first left-hand column, **Fig 2**) the horizontal row labeled 43O17 represents the results obtained using this PAC as a probe. We can see that five target PAC clones have been detected (gray fields in five vertical columns: 52J10, itself, 47M17, 71A4, and 13P20). All five were detected using labeled 43O17 insert DNA in hybridization to colonies (cells processed and immobilized on a nylon filter), code "0" (see **Fig 2** legend). Only two PAC clones (itself and 52J10) were positive when an RNA T7-end probe from 43O17 was used (code "3" in **Fig 2** legend), indicating that the PAC 52J10 overlaps the T7 end of the PAC 43O17 (see positioning in **Fig 1**). Of the remaining three code "0" positive clones, the clone 13P20 was detected with involucrin and also the SPRR cluster containing PAC 20N18 used as a probe. This established the PAC clones 52J10 and 13P20 as the minimum necessary to extend the contig (minimal tiling path or MTP) left and right, respectively, from the starting PAC in this example (43O17). Therefore, only these three PAC clones (52J10, 43O17, and 13P20) were used in the next round of hybridizations to Southern blots (see below). Hence, only these three PAC clones have an additional code "1" positive result in **Fig 2** in the row with which we started this example, the 43O17 horizontal row.

In the case of probes representing a number of S100 genes (A2–A6, first column in **Fig 2**), all known individual S100 genes from this group have been positioned on to PAC clones 148L21 and 178F15, and further mapped on to individual restriction fragments (see below). This was achieved prior to obtaining additional clones (right hand side end columns in **Fig 2**) that were left not tested (nt) for these S100 gene probes. The testing was considered unnecessary because the probes were already positioned within the MTP. The reason the additional clones were obtained was to bridge the one gap that remained near the right hand side of the map.

Restriction mapping All clones were initially digested with *Eco*RI restriction enzyme and separated by electrophoresis on 20 cm long, 0.8% agarose gels. The overlaps, as seen by shared restriction

Figure 1. Restriction and transcriptional map of the 2.45 Mbp bacterial contig encompassing the entire EDC. The top horizontal scale bar is in kilobase pairs, 0 being the most centromeric end of the contig. The second, thicker, horizontal bar is a schematic representation of chromosomal DNA showing the localization of markers, genes and EST in the region. Rhomboid signs represent genes; ellipsoid signs represent markers, sequence tagged sites and EST. The y37m-2, m-6 and m-16 are new markers and their GenBank accession numbers are X94412, AJ009641, and X95103, respectively. 1019b8 is a new keratinocyte cDNA clone mapping to the region, which has the accession number AJ009640. Underneath this are the bacterial clones represented by dark horizontal lines. Clones that make up the MTP are shown with orientation, restriction enzyme sites and localization of genes/markers or EST. Filled in boxes represent T7 clone ends whereas open larger boxes represent SP6 clone ends. Small vertical bars show *Eco*RI restriction enzyme sites whereas a large N or S represents *Not*I and *Sal*I restriction enzyme sites within particular *Eco*RI fragments, respectively. Polymorphic *Eco*RI sites in our map are represented by a small gray vertical bar. The light gray horizontal bars above the bacterial clones represent *Eco*RI "bins"—the precise order of fragments within the bins has not been determined. BAC clones are shown with the prefix "B" whereas cosmid clones have the prefix "c" instead of ICRFc112. The remaining clones are PACs. A more detailed and enlarged map of the region marked by the dashed rectangle can be seen in **Fig 4**, showing the precise ordering of the *SPRR* gene cluster.

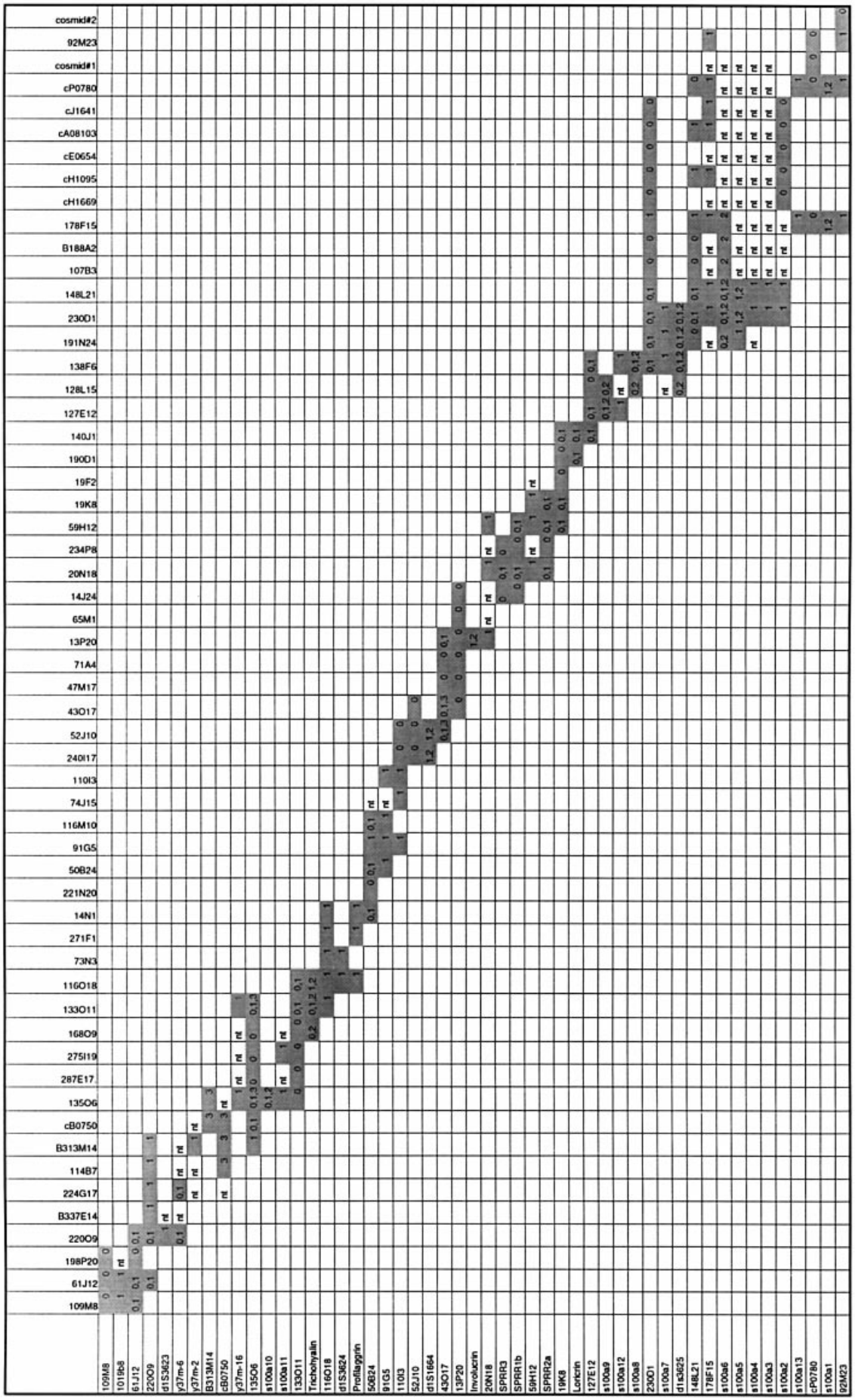


Figure 2. Summary of the hybridization and PCR based confirmations that establish the overlap shown in Fig 1. Left: sources of probes. Top: target clones. The exact nature of probe used in each case was (0) hybridization to colony, (1) hybridization to Southern blot, (2) gene/marker by PCR, (3) T7 clone end riboprobe. In the case of BAC and PAC probes hybridization refers to labelled insert, in the case of cosmid clone cP0780 the probe used was isolated from internal *Sau*3AI fragments. Probes used for genes/markers/EST are indicated in text along with relevant primer sequences. Markers are taken from the Genome Database (GDB). Addresses of clones are as **Fig 1** except cosmid 1: ICRFc112L1963 and ICRFc112E2275, whereas cosmid 2 refers to ICRFc112P1146, ICRFc112K182, ICRFc112K1231, ICRFc112D0853, ICRFc112H1166, ICRFc112A0454, ICRFc112O1459, ICRFc112N0667, ICRFc112F124, ICRFc112G0160, ICRFc112N15101, and ICRFc112 L0567. "nt" indicates clone not tested.

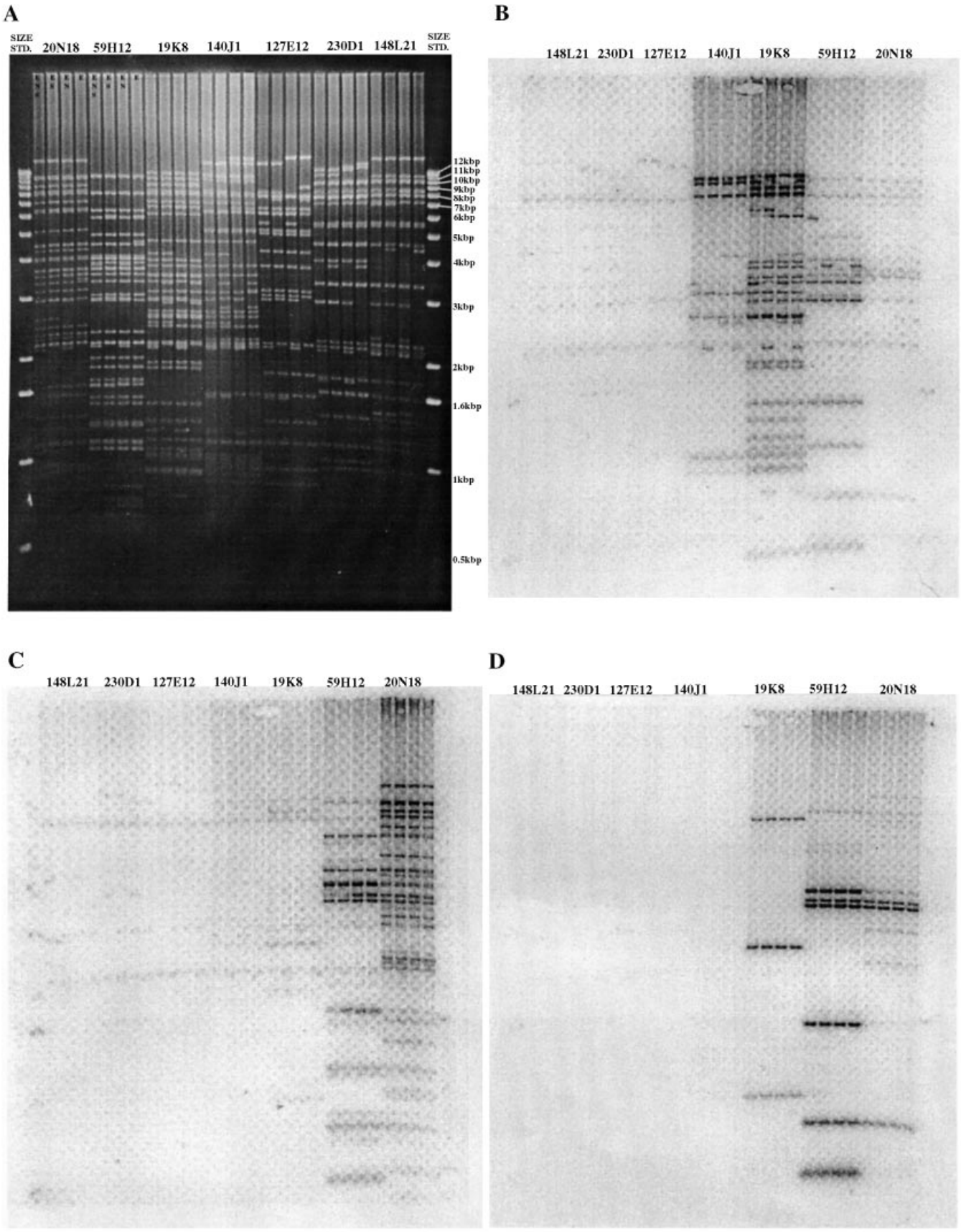


Figure 3. Example of a 0.8% agarose gel and the resulting Southern blot hybridized with various ^{32}P-labelled probes from the EDC. (*A*) agarose gel of seven clones from the MTP (see **Fig 1**), digested with four different combinations of restriction enzymes and stained with ethidium bromide and photographed under ultraviolet. The restriction enzyme combinations for all clones are shown for the first two: E = *Eco*RI, N = *Not*I, and S = *Sal*I. The first and last lane contain a 1 kb ladder size standard, the size of which is indicated for the last. (*B*) Southern blot of gel shown in (*A*), hybridized with insert of the PAC 19K8 used as a probe. (*C*) Southern blot of gel shown in (*A*), hybridized with insert of the PAC 20N18 used as a probe. (*D*) Southern blot of gel shown in (*A*), hybridized with an *SPRR2* cDNA probe. Southern blots display mirror image of the gel in (*A*).

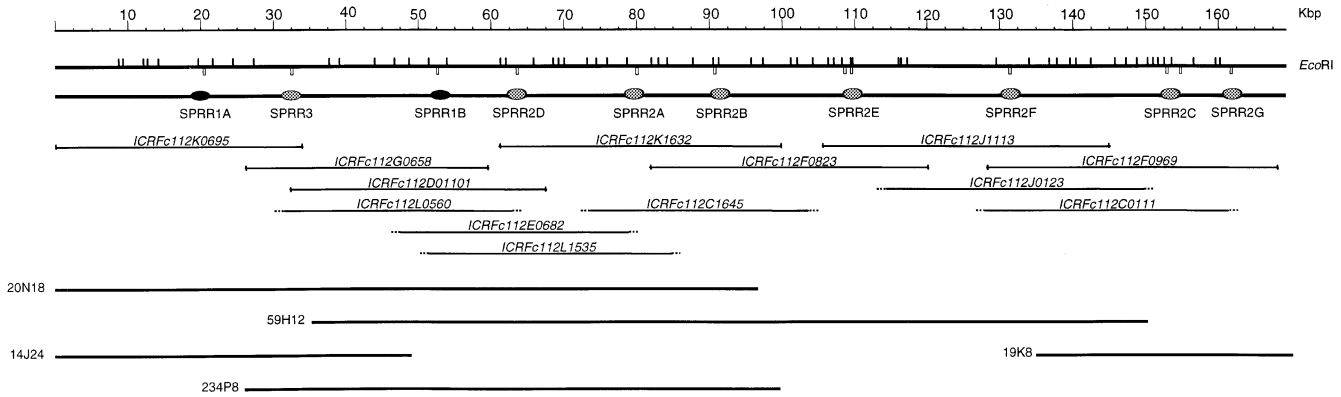


Figure 4. Contiguous *Eco*RI restriction map of the *SPRR* cosmid contig (covering the region within the dashed rectangle around 1500–1690 kbp of the 1q21 map in Fig 1). PACs from the region are shown underneath the cosmid contig. The three different *SPRR* subclasses are indicated by black (*SPRR1*), gray (*SPRR2*), and white (*SPRR3*) ovals. The sequences of *SPRR1A* (L05187), *SPRR1B* (M84757), *SPRR2*A (X53064), *SPRR2*B (L05188), *SPRR2*C (L05189, L05190), and *SPRR3* (AF077374) are available from the Genome Database (GenBank ID in brackets). Sequence analysis of the other *SPRR2* genes, which have been denominated according to their respective localization within the map, is in progress (Cabral *et al* in preparation). Restriction fragments hybridizing to the *SPRR2* probe are indicated with a small open square beneath the bar representing the restriction site pattern. The finding that two genes (*SPRRC* and *SPRRE*) are identified by two hybridizing fragments is due to the fact that the probe contains an internal *Eco*RI site.

fragments determined from these gels, coupled with the results from PAC insert hybridization experiments identified a tentative MTP of overlapping clones. Clones from this MTP were digested with four combinations of restriction enzymes: *Eco*RI only, *Eco*RI and *Not*I, *Eco*RI and *Sal*I, and *Eco*RI, *Not*I, and *Sal*I together. Fragments were again resolved on identical agarose gels which were used to produce Southern blots. Inserts from MTP clones were then systematically hybridized back to the Southern blots in order to define the precise degree of overlap shared by clones. *Eco*RI fragments of the MTP were sized and grouped together in bins (light gray horizontal bars in **Fig 1**) depending on which clones contain the fragments, which clones hybridize to the fragments and other restriction enzyme sites limit their order and position.

Figure 3 displays an example of one such gel and the resulting Southern blot hybridized with a number of radiolabeled probes. Seven clones from the MTP have been resolved on the agarose gel in **Fig 3**(***A***). A number of shared fragments between adjacent clones can clearly be seen in the ethidium bromide stained ultraviolet photo, except in the case of 127E12 and 230D1 which do not overlap. In some instances clones can contain *Eco*RI bands of the same size by coincidence, which is why it is necessary perform hybridizations with adjacent clone inserts as probes. This serves the purpose of both verifying the overlapping fragments and measuring the total length of the overlapping segment. **Figure 3**(***B***) clearly defines the degree of overlap between 19K8 (insert used as probe) and adjacent clones. Another example of this overlap conformation can be seen in **Fig 3**(***C***) where the insert of clone 20N18 has been used as a probe. **Figure 3**(***D***) displays the same Southern blot hybridized with the cDNA derived from the 930 bp probe corresponding to *SPRR2E*. Owing to the highly conserved nature of the *SPRR2* genes (90% in the coding region Gibbs *et al*, 1993) a number of positive *Eco*RI fragments can be seen (nine in total) corresponding to the seven *SPRR2* genes (see also **Fig 4**). These positive fragments have been placed in *Eco*RI bins accordingly: in each of three separate bins (59H12 and 20N18 shared, 59H12 specific and 19K8 specific), three *SPRR2* positive fragments can be placed. As this procedure does not allow the positioning of the three different fragments in one bin and their assessment to one specific gene, the whole SPRR region was mapped in more detail by using a cosmid contig (see below).

Probes representing S100A6, S100A5, S100A4, and S100A3 localize to PAC clones 230D1, 148L21, and 178F15 (see **Fig 1**). S100A6 and S100A5 localize to a 10.6 kb fragment present in all three clones. S100A4 localizes to this same fragment in all three clones but also to a 17 kb fragment in clones 148L21 and 178F15 along with the 14 kb SP6-end fragment of 230D1, thus indicating an *Eco*RI site within the S100A4 sequence. S100A3 localizes to the 17 kb fragment of 148L21 and 178F15 as well as the 14 kb SP6 fragment of 230D1, whereas S100A2 is only present on the 17 kb fragment from 148L21 and 178F15. This establishes the order S100A5 + S100A6–S100A4–S100A3–S100A2 and is in full agreement with the genomic sequence data (Englekamp *et al*, 1993). This sequence also contains the internal *Eco*RI site of S100A4 and, coupled with our map, unambiguously defines the order as S100A6–S100A5–S100A4–S100A3–S100A2.

Recombinant PAC clones are constructed (Ioannou *et al*, 1994) in such a way that the random genomic DNA insert (100–150 kb on average) has been inserted in the cloning site between the two RNA polymerase promoters T7 and SP6. The specific orientation of the PAC inserts in relation to the vector could be determined as the restriction enzyme *Not*I digests the PAC vector either side of the cloning site, whereas the restriction enzyme *Sal*I digests the vector on the SP6 side only. In the case of the one BAC clone, the two cosmid clones, and where PAC end-fragments could not be easily identified within the MTP (due to co-running with internal fragments), T7-RNA end-probes were hybridized to Southern blots, defining clone orientation and the exact positioning of certain *Eco*RI fragments.

Gene and transcript map Hybridization probes were produced for all genes and markers shown in **Fig 1** and used against the Southern blots of the MTP, thus positioning the respective genes and markers to specific *Eco*RI fragments (for an example, see **Fig 3*D***).

In the centromeric half of the map bacterial clones confirm the order of genes and markers and the respective distances between them from previous mapping studies (Marenholz *et al*, 1996; Mischke *et al*, 1996; Lioumi *et al*, 1998).

Telomeric of S100A9, our map for the first time establishes the order and distances between all known genes and markers in this segment. The S100 genes A11, A12, and A13 have been precisely positioned in relation to the previously published order of S100A1–10. We establish the order as follows (**Fig 1**): cen–S100A10–S100A11–S100A9–S100A12–S100A8–S100A7–S100A6–S100A5–S100A4–S100A3–S100A2–S100A13–S100A1–tel.

The region occupied by the *SPRR* gene family (a 170 kbp segment marked by a hashed line rectangle in **Fig 1**) has been mapped in more detail by assembling a contiguous *Eco*RI map, using chromosome 1 specific cosmids (**Fig 4**). This allowed the assignment of the nine *SPRR2* specific hybridizing bands (**Figs 1** and **3*D***) to seven individual *SPRR2* genes and the determination of the relative position of the various members of the three

subfamilies in the *SPRR* multigene family. A more detailed study is addressing the exon–intron organization, transcriptional orientation, and evolution of the various members of this gene family, and will be published elsewhere (Cabral *et al* in preparation).

The three new markers y37 m-2, y37 m-6, and y37 m-16 were produced from subcloning YAC 764_a_1, increasing marker density within the region. One new cDNA clone, 1019b8 was found from YAC hybridizations to a cultured keratinocyte cDNA library, and has been mapped approximately 350 kb telomeric of S100A10.

DISCUSSION

The overall size of the region mapped in **Fig 1** is 2450 kbp. Of the total contig length 59.3% is covered by overlapping clones with a depth ranging from 3- to 9-fold, and only 15.6% is covered by single-clone-deep parts of the contig. The average degree of overlap in the MTP is 47.2% of clone length. The sum of the lengths of the nonredundant restriction fragments shown in the map in **Fig 1** gives the measurement of the physical distance between the markers for various intervals of the map which matches well with the previously published physical maps of large genomic DNA fragments separated by pulse-field gel electrophoresis (Volz *et al*, 1993; Mischke *et al*, 1996). These maps found the genes S100A10 and S100A6 on genomic DNA fragments with minimum sizes 1760 kbp and 1750 kbp, respectively, and maximum sizes of 2050 kbp and 1900 kbp, respectively. We find the same genes occupying a maximum distance of 1700 kbp which is in good agreement with the data from direct genomic physical mapping. The small difference of 50–60 kb could be easily attributed to the difference in resolution between the two experimental approaches.

One small refinement to previously published work localizing genes and markers to YAC clones can be seen. The previously published 6 Mb YAC contig (Marenholz *et al*, 1996) was unable to resolve the position of the marker d1S3625 with respect to S100A6 and gave the following order: cen–S100A9–S100A6–d1S3625–tel (see **Fig 1**, Marenholz *et al*, 1996). We see from the hybridization pattern in the bacterial contig the order: cen S100A9–d1S3625–S100A6–tel.

The genes S100A6, S100A5, S100A4, and S100A3 have previously been localized to a 15 kb region in genomic DNA-containing phages (Englekamp *et al*, 1993). In agreement we see these S100 genes localizing to a region of similar size. S100A13 is shown to hybridize to an *Eco*RI fragment from digested YAC DNA of about 10 kb (Wicki *et al*, 1996a). This is in line with our finding, an 8.9 kb *Eco*RI fragment positive on hybridization of an S100A13 probe.

The study of a YAC containing S100 genes A9–A1 (Schaefer *et al*, 1995) gives a distance between these two genes of 320 kbp. Our contig is in agreement, with a maximum distance of 330 kbp and also confirms the position of *Sal*I restriction enzyme sites in the region. A mapping study using mainly YAC clones from Zhao and Elder, 1997 gives a distance from involucrin to S100A2 of 780 kbp. Our map is in agreement with a maximum distance of 790 kbp and also confirms the position of a single *Not*I restriction enzyme site over this region. Clones 220O9 and 92M23 could potentially contain as yet uncharacterized CpG islands (both contain two *Not*I restriction enzyme sites) and are currently the subject of an ongoing investigation.

A multigene family clustered in a small genomic region is not a novelty for complex genomes, but very few of them include structurally different (though distantly related in evolution) classes of genes. Best examples are the major histocompatibility complex, and other gene complexes belonging to the immunoglobulin superfamily. Genomic regions occupied by such complexes are usually transcriptionally very active, and truly "gene-rich", including numerous additional transcripts whose direct evolutionary, structural, and functional link to the multigene family (if any) is not apparent. For example, the genomic segment harboring the human major histocompatibility complex in 6p is one of the "gene-richest" regions in the entire genome with gene densities exceeding 1 per 10 kbp spreading over several Mbp (Milner and Campbell, 1992). This allows for prediction that the 1q21 region may harbor novel transcripts unrelated to the EDC as well as new members of the current gene families. The bacterial clones presented in this map provide optimal material for further isolation of genes using exon trapping, cDNA selection and the screening of cDNA libraries. This study reports one new cDNA, which has been mapped approximately 350 kb telomeric of S100A10 and was obtained by hybridization of a YAC to the keratinocyte cDNA library. This position is outside the current EDC *in sensu stricto*. This cDNA does not show sequence similarity to any of the three EDC gene families. If the expression of the gene underlying this cDNA is found to be regulated during keratinocyte differentiation, it may add further genes (gene families) to the EDC, expanding its physical borders. This is currently an ongoing investigation.

It has been suggested that EDC gene expression could be coordinated from a shared locus control region (Hardas *et al*, 1996), raising the question of evolutionary history and biologic role. Are the EDC genes kept together coincidentally or are there regulatory and functional constraints keeping the genes in close physical distance over millions of years and across species barriers? What could the study of the EDC gene analogs throughout vertebrate evolution tell us about the role of specific proteins in the evolution of epidermis? A high resolution physical map of ordered bacterial clones is an essential resource in experimental approaches addressing these and other questions. An integrated restriction map, in particular with *Not*I, renders the clones specially useful for direct transient transfection approaches using PAC/BAC retrofitting technology (Mejia and Monaco, 1997; Kim *et al*, 1998).

In conclusion, we believe the high resolution bacterial contig spanning the EDC will be a very helpful resource in further studies of this intriguing gene complex. This contig will also be the main starting substrate for the long range sequencing of this region by The Sanger Center, Hinxton, U.K.

We thank Beat Schaefer and Daniel Hohl for some gene-specific probes, Maria Lioumi for sharing corroborative mapping information, the MRC-Human Genome Mapping Project Resource Center U.K. for access to various molecular resources and Juergen Groet for advice in experimental procedures. Thanks to Susan Gibbs and David Fischer (Leiden University) for essential contribution in the construction and propagation of the human keratinocyte cDNA library. We also thank David Bentley, Simon Gregory, and Mark Vaudin for advice and co-ordination to help incorporate this map into the overall chromosome 1 mapping and sequencing project at The Sanger Center. A.P.S. has been partly supported by the Constance Bequest Fund. This work has been mainly supported by the BMH4-CT96-0319 grant from the Commission of European Communities to the European 1q21 Consortium, and partly by the Human Genome Mapping Project Strategic grant G9422614 from the Medical Research Council, U.K.

REFERENCES

Baxendale S, Bates GP, MacDonald ME, Gusella JF, Lehrach H: The direct screening of cosmid libraries with YAC clones. *Nucleic Acids Res* 19:6651, 1991

Cheng JF, Boyartchuk V, Zhu Y: Isolation and mapping of human chromosome 21 cDNA. Progress in constructing a chromosome 21 expression map. *Genomics* 23:75–85, 1994

Dale BA, Resing KA, Presland RB. Keratohyalin granule proteins. In: Leigh I, Land B, Watt F (eds). *The Keratinocyte Handbook*. Cambridge: Cambridge University Press, 1994, pp 323–350

Dillon N, Grosveld F: Transcriptional regulation of multigene loci: multilevel control. *Trends Genet* 9:134–137, 1993

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

Englekamp D, Schafer BW, Mattei MG, Erne P, Heizmann CW: Six S100 genes are clustered on human chromosome 1q21: identification of two genes coding for the two previously unreported calcium binding proteins S100D and S100E. *Proc Natl Acad Sci USA* 90:6547–6551, 1993

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

Forus A, Berner JM, MezaZepeda LA, Saeter G, Mischke D, Fodstad O: Molecular characterisation of a novel amplicon at 1q21-q22 frequently observed in human sarcomas. *Br J Cancer* 78:495–503, 1998

Gelb BD, Spencer E, Obad S, Edelson GJ, Faure S, Weissenbach J, Desnick RJ: Pycnodysostosis: Refined linkage and radiation hybrid analyses reduce the critical region to 2 cM at 1q21 and map two candidate genes. *Hum Genet* 98:141–144, 1996

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

Groet J, Ives JH, South AP, *et al*: Bacterial contig map of the 21q11 region associated with Alzheimer's disease and abnormal myelopoiesis in Down syndrome. *Genome Res* 8:385–398, 1998

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

Ioannou AP, Amemiya J, Garnes J, *et al*: A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet* 6:84–89, 1994

Ivens AC, Little PFR. Cosmid clones and their application to genome studies. In: Glover DM, Hames BD (eds). *DNA Cloning 3 A Practical Approach*. Oxford: IRL Press, 1995, pp 1–46

Kim SY, Horrigan SK, Altenhofen JL, Arbieva ZH, Hoffman R, Westbrook CA: Modification of bacterial artificial chromosome clones using Cre recombinase: introduction of selectable markers for expression in eukaryotic cells. *Genome Res* 8:404–412, 1998

Korge BP, Ishida-Yamamoto A, Punter C, *et al*: Loricrin mutation in Vohwinkel's keratoderma is unique to the variant with ichthyosis. *J Invest Dermatol* 109:604–610, 1997

Lioumi M, Olavesen MG, Nizetic D, Ragoussis J: High-resolution YAC fragmentation map of 1q21. *Genomics* 409:200–208, 1998

Maestrini E, Monaco AP, McGrath JA, *et al*: A molecular defect in loricrin, the major component of the cornified cell envelope, underlies Vohwinkel's syndrome. *Nature Genet* 13:70–77, 1996

Marenholz I, Volz A, Ziegler A, Davies A, Ragoussis I, Korge BP, Mischke D: 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, 1996

Markova NG, Marekov LN, O'Keefe EJ, Parry DAD, Steinert PM: Profilaggrin is a major epidermal calcium binding protein. *Mol Cell Biol* 13:613–625, 1993

Mejia JE, Monaco AP: Retrofitting vectors for *E. coli*-based artificial chromosomes (PACs and BACs) with markers for transfection studies. *Genome Res* 7:179–186, 1997

Milner CM, Campbell RD: Genes, genes and more genes in the major histocompatibility complex. *Bioessays* 14:565–571, 1992

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

Nizetic D, Lehrach H: Chromosome specific cosmid libraries: construction, handling and use in parallel and integrated mapping. In: Glover DM, Hames BD (eds). *DNA Cloning 3. A Practical Approach*. Oxford: IRL Press, 1995, pp 49–79

Nizetic D, Monard S, Cotter F, Young BD, Lehrach H: Construction of cosmid libraries from flow sorted material of human chromosomes 1, 6, 7, 11, 13 and 18 for use as reference libraries. *Mamm Genet* 5:801–802, 1994

O'Kefee EJ, Hamilton EH, Lee SC, Steinert P: Trichohyalin—A Structural protein of hair, tongue, nail, and epidermis. *J Invest Dermatol* 101:65s–71s, 1993

Polymeropoulos MH, De Ortiz Luna RI, Ide SE, Torres R, Rubenstein J, Francomano CA: The gene for pycnodysostosis maps to human chromosome 1cen-1q21. *Nature Genet* 10:238–239s, 1995

Presland RB, Bassuk JA, Kimball JR, Dale BA: Characterization of two distinct calcium-binding sites in the amino-terminus of human profilaggrin. *J Invest Dermatol* 104:218–223, 1995

Resing KA, Dale BA. Proteins of keratohyalin. In: Goldsmith LA (ed.). *Physiology, Biochemistry and Molecular Biology of the Skin*. New York: Oxford University Press, 1991, pp 148–167

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

Schaefer BW, Wicki R, Englekamp D, Mattei MG, Heizmann CW: 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, 1995

Shizuya H, Birren BI, Kim U-J, Mancino B, Slepak T, Tachiiri Y, Simon MI: Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci USA* 89:8794–8797, 1992

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

Townes TM, Behringer RR: Human globin locus activation region (LAR): Role in temporal control. *Trends Genet* 6:219–229, 1990

Volz A, Korge BP, Compton JG, Zeigler A, Steinert PM, Mischke D: Physical mapping of a functional cluster of epidermal differentiation genes on chromosome 1q21. *Genomics* 18:92–99, 1993

Weterman MAJ, Wilbrink M, Dijkhuizen T, van den Berg E, van Kessel AG: Fine mapping of the 1q21 breakpoint of the papillary venal cell carcinoma-associated (X; 1) translocation. *Hum Genet* 98:16–21, 1993

Wicki R, Schaefer BW, Erne P, Heizmann CW: 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, 1996a

Wicki R, Marenholz I, Mischke D, Schaefer BW, Heizmann CW: 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, 1996b

Williams ML, Elias PM: From basket weaver to barrier—unifying concepts for the pathogenesis of the disorders of cornification. *Arch Dermatol* 192:626–629, 1993

Yuspa SH, Kilkenny AE, Steinet PM, Roop DR: Expression of murine epidermal differentiation markers is tightly regulated by restricted extracellular calcium concentrations in vitro. *J Cell Biol* 109:1207–1217, 1989

Zhao XP, Elder JT: Positional cloning of novel skin-specific genes from the human epidermal differentiation complex. *Genomics* 45:250–258, 1997