Haplotype analysis of distantly related populations implicates corneodesmosin in psoriasis susceptibility


Psoriasis (MIM *177900) is a hyperproliferative skin disorder, characterised by inflammatory cell dermal infiltration, disruption of keratinocyte terminal differentiation, and premature desquamation of the stratum corneum. Although the disease pathogenesis is poorly understood, it is well recognised that genetic factors underlie psoriasis susceptibility, as documented by family clustering of the disease and increased concordance rates in monozygotic twin pairs. To date, genome wide scans have identified seven distinct psoriasis susceptibility regions (PSORS 1-7) and have provided compelling evidence for a major role of the PSORS1 locus on chromosome 6p21. Linkage disequilibrium (LD) based studies of microsatellite maps have refined the PSORS1 boundaries to a 150 kb interval harbours three sets of coding polymorphisms that have repeatedly shown association with psoriasis: HLA-Cw*0602, HCR*269, and CDSN*TTC.

In a recent analysis of a UK family cohort, we have shown that the most common psoriasis susceptibility chromosomes (cluster E) carry HCR*269, CDSN*TTC, and two SNPs (defined as n7 and n9) lying proximal to HLA-C and showing extremely significant disease association. We also observed a rare haplotype (cluster D), marginally over-transmitted to affected patients and likely to have originated from cluster E by a double recombination event, replacing HCR*269 while preserving SNPs 7/9 and CDSN*TTC. This observation suggested that CDSN SNPs might be required to confer psoriasis susceptibility, in conjunction with determinants lying in the HLA-C genomic region. CDSN is also a very attractive biological candidate, since corneodesmosin is a desmosomal protein involved in keratinocyte cohesion/desquamation.

With this study, we have sought to validate the hypothesis of CDSN involvement in psoriasis by: (1) confirming that cluster D is a risk haplotype, through the analysis of an independent population, originating from the Gujurat region of northern India; and (2) defining CDSN genetic variation by an association study of intragenic SNP haplotypes. This analysis identified a 16 SNP chromosome segment that contains three genes carrying disease associated alleles: HLA-C, α-helix coiled coil rod homologue (HCR), and corneodesmosin (CDSN).

SNP identification

Four PCR primer pairs were designed based on GenBank sequence L28015, in order to amplify the CDSN coding region and the entire 3′UTR. A further primer pair was designed on the CDSN upstream region contained in genomic sequence AP000510, in order to amplify 785 bp of putative gene promoter. All primer sequences are available on our web site (www.leicester.ac.uk/ge/rt7/index2.html). PCR products were sequenced using Big Dye Terminators (Applied Biosystems) and electrophoresed on an ABI 377 automated sequencer (Applied Biosystems). SNPs were identified by visual inspection of chromatograms.

SNP typing

Samples were amplified using the primer pairs reported on our web site and the resulting PCR products were blotted onto Hybond N nylon membranes (Amersham). For each SNP, a pair of allele specific oligonucleotides (ASOs) was synthesised, end labelled, and used to probe dot blots of the corresponding PCR products, as described elsewhere. Synonym nucleotide changes and SNPs presenting a minor allele frequency <0.25 were excluded from the genotyping panel.
### Statistical analyses

Family based association analysis of SNP haplotypes was performed by using the TRANSMIT program. Owing to computational limitations, overlapping segments of 7-9 SNPs were first determined. These were then pieced together in succession by identification of unique SNPs in overlapping regions until full length haplotypes were generated.

The PHASE program was used to derive haplotypes from the diploid genotypes of the Gujarati subjects. The frequencies of haplotypes predicted for cases and controls were compared using a chi-square test with 1 degree of freedom.

Analysis of LD conservation among controls was performed using ad hoc software written in True Basic 4.1 by Alec Jeffreys. The haplotypes segregating in the two three generation families were derived manually, by minimising recombination events.

### Bioinformatic analyses

The sequence of the minimal promoter region was predicted using a Neural Network tool (available at http://www.fruitfly.org/seq_tools/promoter.html). The putative transcription factor binding sites were assessed by analysing 900 bp of upstream CDSN sequence with the TESS software (www.chil.upenn.edu/cgi-bin/tess/tess33). A 900 bp segment upstream of the Mus musculus Cdsn gene was similarly analysed, and the results of human and mouse TESS analyses were compared, searching for conserved binding sites. The human sequence was used for this analysis (nucleotides 117010-117910 in GenBank sequence AF111103) was identified by a BLAST search (www.ncbi.nlm.nih.gov/Blast) of the most likely positional candidate for the region shared by clusters D and E, and defining an interval that encompasses the entire CDSN and STG genes, together with the first two exons of SEEK1 (fig 1). However, resequencing of STG has failed to identify any high frequency SNP while SEEK1 exons 1 and 2 lie within the gene 5’UTR (see GenBank sequence AP000510). Thus, CDSN appears to be the most likely positional candidate for the region shared by clusters D and E.

### CDSN gene resequencing

Resequencing of eight unrelated patients carrying PSORSI risk chromosomes identified 17 coding variants, all of which had been previously described. We also observed 10 polymorphisms in the putative promoter and the 3’UTR. Eight of these variants are novel, this study being the first systematic survey of CDSN non-coding regions. A complete list of the polymorphisms observed in resequenced patients is available on our web site.

### LD conservation patterns

Fig 2 shows the results of diploid LD analysis between 14 SNP loci presenting a minor allele frequency >0.15. LD appears to be maintained along the upstream region and the proximal CDSN coding sequence (D’<0.8, columns –636 to 619 in fig 2), but it tends to decay with distance in the distal portion of the gene (D’< 0.4, columns 1243 to 1748).

### Bioinformatic analyses

In order to generate a framework model allowing a preliminary assessment of SNP functional impact, we used bioinformatic tools to analyse the CDSN genomic sequence. Analysis by neural networks predicted a 41 bp polymerase II promoter site (nucleotides 21608-21648 in Genbank sequence AP000510; score = 0.96), containing a CATAAA box. This interval lies within a conserved 101 bp sequence showing 88% identity with the homologous mouse region. Analysis of human and murine CDSN upstream sequences highlighted conservation of potential binding sites for AP1 (-73/-67, -823/-817), Sp1 (-251/-246, -313/-308, -588/-593, -909/-915), CACCC box (-1061/-101, -301/-296, -459/-454) and AP2-alpha (-851/-844) transcription factors, all of which have been implicated in keratinocyte differentiation.

Secondary structure analyses of the coding sequence identified an alternation of short beta-sheets and longer coil regions. Searches of domain database failed to identify any match to consensus sequences. An analysis with the ScanProsite program predicted glycosylation sites corresponding to Asn residues 134 and 156. These results were confirmed by the NetNGlyc server.

### CDSN haplotype analysis

We typed a panel of 16 SNPs in our original UK data set and in 30 Gujarati parent-offspring trios. The resulting CDSN haplotypes are shown in table 2. Six chromosomes occurring with

<table>
<thead>
<tr>
<th>Table 1</th>
<th>PSORSI haplotypes carrying SNP9 and/or CDSN risk alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP9C</td>
<td>HCR*269T</td>
</tr>
<tr>
<td>Cluster D</td>
<td>+</td>
</tr>
<tr>
<td>Cluster E</td>
<td>+</td>
</tr>
<tr>
<td>Cluster G1</td>
<td>+</td>
</tr>
<tr>
<td>Cluster G2</td>
<td>+</td>
</tr>
</tbody>
</table>

In order to exclude the possibility that the CDSN*TTC association might reflect LD with a coding variant located in a downstream gene, we typed two SNPs within the GTF2H4 gene lying 200 kb distal to CDSN. The analysis of our original 171 trios UK sample showed that these two marker alleles freely associate with all PSORSI haplotypes, placing GTF2H4 outside the segment shared by clusters D and E, and defining an interval that encompasses the entire CDSN and STG genes, together with the first two exons of SEEK1 (fig 1). However, resequencing of STG has failed to identify any high frequency SNP while SEEK1 exons 1 and 2 lie within the gene 5’UTR (see GenBank sequence AP000510). Thus, CDSN appears to be the most likely positional candidate for the region shared by clusters D and E.
a frequency >2% accounted for ~90% of haplotypes, both in the UK and the Gujurati population. Haplotype I was significantly over-transmitted to affected offspring in both data sets ($p<10^{-10}$ and $p=3.9 \times 10^{-3}$, respectively). Haplotype II, which is likely to have originated from I by ancestral mutation, was over-transmitted in the UK sample ($p=3.9 \times 10^{-3}$), while showing a non-significant transmission increase in the smaller Gujurati data set. The cumulative frequency of haplotypes I and II was remarkably similar in the two populations, summing up to 40.1% in the UK and 41.9%
among the Gujarati. Haplotype V showed the highest number of substitutions with respect to I (11/16 variants), and was significantly under-transmitted to patients in the UK sample (p=4×10^-4), while showing a lack of transmission of borderline significance (p=5×10^-2) among the Gujarati.

TRANSMIT analysis of SNP9, HCR*269, and CDSN genotypes showed that haplotype I is found on disease bearing clusters D and E, together with SNP9C (table 3). This combination is unique to risk chromosomes as neutral clusters G1 and G2 respectively carry SNP9T + haplotype I and SNP9C + haplotype VI.

To validate experimentally the haplotypes derived by TRANSMIT, we typed SNP9, HCR*269, and the 16 CDSN variants in two extended three generation pedigrees, where we were able to confirm the segregation of all the haplotypes observed in the UK data set.

DISCUSSION

CDSN alleles have previously been implicated in psoriasis pathogenesis, but disease associations have generally been ascribed to LD with HLA-Cw*0602, which is currently recognised as the marker conferring the highest psoriasis risk. However, we recently showed that HLA-C and CDSN SNPs do not show linkage disequilibrium when non-disease bearing chromosomes are analysed. This suggested that the associations with HLA-C and CDSN SNPs might reflect a requirement of both sets of alleles on risk haplotypes. This hypothesis was also supported by the conservation of CDSN*TTC on cluster D. Here, we have validated cluster D as a risk chromosome, through the analysis of an Indian case-control data set. Studying this distantly related white population disclosed a distinctive repertoire of ancestral haplotypes. Thus, the identification of a common, neutral chromosome carrying SNP9C in the absence of CDSN*TTC (cluster G2) and the observation of significant cluster D increase among affected subjects concur to support a requirement of CDSN risk alleles on psoriasis susceptibility haplotypes.

The corneodesmosin gene being extremely polymorphic, we have derived SNP haplotypes spanning 5.3 kb of CDSN genomic sequence, in order to assess the contribution of additional variants that might be in LD with CDSN*TTC (cluster G2) and the observation of significant cluster D increase among affected subjects concur to support a requirement of CDSN risk alleles on psoriasis susceptibility haplotypes.

The corneodesmosin gene being extremely polymorphic, we have derived SNP haplotypes spanning 5.3 kb of CDSN genomic sequence, in order to assess the contribution of additional variants that might be in LD with CDSN*TTC (cluster G2) and the observation of significant cluster D increase among affected subjects concur to support a requirement of CDSN risk alleles on psoriasis susceptibility haplotypes.

| Table 2A | CDSN chromosomes occurring with >2% frequency; haplotype alignment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5' region | Coding | 3' UTR |
| Allele | −712 | −646 | −461 | −31 | −22 | 9 | 442 | 619 | 1215 | 1236 | 1243 | 1593 | 1606 | 1675 | 1739 | 1747 |
| I | c | t | t | a | c | g | (Ser) | t | a | c | g | (Ser) | t | g | t | t |
| II | c | t | t | a | c | g | (Ser) | t | a | c | g | (Ser) | t | g | t | t |
| III | c | t | t | a | c | g | (Ser) | c | a | t | c | g | (Ser) | t | g | t | t |
| IV | t | g | c | a | c | t | g | (Ser) | a | t | t | g | aag | t | c | c |
| V | c | a | t | g | t | a | t | (Asn) | a | t | t | a | aag | t | c | c |
| VI | t | g | c | a | c | t | g | (Ser) | a | g | c | g | (Ser) | (Asp) | t | t | t |

*Allele number according to the classification of CDSN coding haplotypes proposed by Guerrin et al.

| Table 2B | CDSN chromosomes occurring with >2% frequency; haplotype transmission to patients |
|---|---|---|---|---|---|---|
| Haplotype | Freq (%) | Status | UK | Gujarati | UK | Gujarati | χ² |
| I | 31.8 | 35.8 | Over-transmitted | Over-transmitted | 42.8 | 8.3 |
| II | 8.3 | 6.1 | Over-transmitted | Neutral | 8.3 | 8.3 |
| III | 12.9 | 2.0 | Neutral | Neutral | 6.1 | 6.1 |
| IV | 16.2 | 13.3 | Under-transmitted | Neutral | 12.4 | 3.8 |
| V | 13.6 | 6.9 | Under-transmitted | Under-transmitted | 5.1 | 5.1 |
| VI | 6.8 | 23.2 | Under-transmitted | Neutral | 5.1 | 5.1 |

| Table 3 | CDSN haplotypes present on PSORS1 clusters |
|---|---|---|---|
| SNP9C | HCR*269T | CDSN haplotype |
| Cluster A | − | − | V |
| Cluster B | − | − | III, VI |
| Cluster C | − | + | IV |
| Cluster D | + | − | I |
| Cluster E | + | + | I, II |
| Cluster G1 | − | − | I |
| Cluster G2 | + | + | VI |
shorter haplotype (allele 1.4), which shows disease association in the Japanese population. The presence of different CDSN*619 and CDSN*1215 alleles on the otherwise identical haplotypes I, II, and III suggests that these SNPs do not play a major pathogenic role.

Haplotype IV, V, and VI display a higher degree of diversity, each carrying at least five substitutions with respect to the other two. Haplotype V shows the highest number of substitutions compared to I (11/16 variants), and is under-transmitted in both the UK and the Gujarati sample. Interestingly, the frequency of CDSN allele 2.21, which is contained in haplotype V, also appears to be increased among Japanese controls, if compared to patients. Altogether, these observations suggest that haplotype V may have a protective effect in a wide range of populations. As CDSN being over-expressed in psoriatic lesions, it is of note that haplotypes I and V differ at 8/10 variants localised to potential regulatory regions. These include two SNPs (CDSN*–31 and CDSN*–22) lying within the predicted RNA polymerase II promoter and four variants from the 3′UTR. This latter group of alleles might alter gene expression levels by affecting mRNA stability, as shown for a number of non-coding SNPs involved in complex disease pathogenesis.12 13

As we have only resequenced 785 bp upstream of the CDSN coding region, we cannot exclude a contribution of additional regulatory variants lying in the distal genetic promoter. However, the pattern of LD decay observed in UK controls suggests that such SNPs would be unlikely to be in strong linkage disequilibrium with CDSN*TTC, which lies in the distal portion of the gene.

Haplotype I and V also diverge at three (CDSN*442, CDSN*1243, and CDSN*1593) of the four coding variants that are common to chromosomes I, II, and III. CDSN*442 appears to introduce a potential site for N-glycosylation, a process which is known to take place in vivo, as N-linked oligosaccharides account for ~10% of CDSN molecular weight. The potential impact of CDSN*1243 and CDSN*1593 cannot be easily predicted, since both SNPs occur on open coil regions matching no conserved functional domain. However, a corneodesmosin 3D structural model, recently generated by our group using fold recognition methods, predicted that the amino acid affected by CDSN*1243 localises to an exposed region of the protein (Allen et al, manuscript in preparation), where a change in residue polarity might hamper the action of the proteases that digest CDSN during skin desquamation.22 Clearly, bioinformatic predictions have to be interpreted with caution; nonetheless these analyses have generated a number of working hypotheses that can now be assessed experimentally, in order to dissect the functional contribution of regulatory and coding SNPs lying on CDSN risk haplotypes.

The localisation of haplotype I to clusters D and E, both of which also contain HLA-C related risk alleles, is consistent with our initial two locus model of the PSORS1 interval. Interestingly, haplotype III only occurs on cluster B, which lacks the second set of disease alleles. This may explain why haplotype III does not show psoriasis association among the white population, while conferring disease susceptibility in the Japanese, where it occurs on the background of an ancestral haplotype bearing an HLA-C risk allele.25

Clustering of complex disease loci in regions defined by a single peak of linkage has previously been described, with accumulating evidence for multiple MHC genes conferring susceptibility to type 1 diabetes26 and multiple sclerosis.27 The occurrence of separate, closely linked disease susceptibility loci adds to the complexity of common disease analysis and warrants the study of divergent populations presenting with different patterns of linkage disequilibrium conservation.

ACKNOWLEDGEMENTS
The authors wish to thank all the psoriasis families whose participation made this project possible. The authors would also like to thank Professor Sir Alec Jeffreys for access to software for LD diploid analysis. This research was supported by a Wellcome Trust grant to RC and JWNB (No 056713/Z/99/Z) and a Medical Research Council (MRC) UK Cooperative Group Grant. FC is a recipient of a Wellcome Trust Travelling Research Fellowship.

Supplementary tables 1 and 2 can be found on our web site, www.jmedgenet.com/supplemental

Authors’ affiliations
F Capon, J K Tool, J C Evans, S Patel, R C Trembath, Division of Medical Genetics, University of Leicester, Leicester, UK
F Capon, Division of Human Genetics, “Tor Vergata” University of Rome, Italy
M H Allen, J N W N Barker, St John’s Institute of Dermatology, Kings College, London, UK
D Tillman, D Burden, Department of Dermatology, Western Infirmary, Glasgow, UK

Correspondence to: Professor R Trembath, Division of Medical Genetics, Department of Genetics and Medicine, Adrian Building, University Road, Leicester LE1 7RH, UK; rtremba@hgmpr.mrc.ac.uk

REFERENCES

www.jmedgenet.com


Haplotype analysis of distantly related populations implicates corneodesmosin in psoriasis susceptibility


J Med Genet 2003 40: 447-452
doi: 10.1136/jmg.40.6.447

Updated information and services can be found at:
http://jmg.bmj.com/content/40/6/447

These include:

**Supplementary Material**
Supplementary material can be found at:
http://jmg.bmj.com/content/suppl/2003/07/01/40.6.447.DC1

**References**
This article cites 36 articles, 13 of which you can access for free at:
http://jmg.bmj.com/content/40/6/447#BIBL

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**
Articles on similar topics can be found in the following collections

- Dermatology (240)
- Molecular genetics (1254)
- Ethics (220)
- Epidemiology (630)
- Genetic screening / counselling (886)
- Immunology (including allergy) (604)

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/