

ORIGINAL ARTICLE

Haplotypic diversity in human CEACAM genes: effects on susceptibility to meningococcal disease

MJ Callaghan¹, K Rockett², C Banner², E Haralambous³, H Betts⁴, S Faust⁵, MCJ Maiden⁶, JS Kroll⁴, M Levin⁴, DP Kwiatkowski² and AJ Pollard¹

¹Department of Paediatrics, Centre for Clinical Vaccinology and Tropical Medicine, University of Oxford, Churchill Hospital, Headington, Oxford, Oxon, UK; ²Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; ³Department of Clinical Molecular Genetics, Institute of Child Health, London, UK; ⁴Department of Paediatrics, Faculty of Medicine, Wright Fleming Institute, Imperial College London, London, UK; ⁵Wellcome Trust Clinical Research Facility, University of Southampton School of Medicine, Southampton General Hospital, Southampton, UK and ⁶Department of Zoology, Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, UK

Adhesion between the opacity-associated adhesin (Opa) proteins of Neisseria meningitidis and human carcino-embryonic antigen cell adhesion molecule (CEACAM) proteins is an important stage in the pathogenesis of meningococcal disease, a globally important bacterial infection. Most disease is caused by a small number of meningococcal genotypes known as hyperinvasive lineages. As these are also carried asymptotically, acquisition of them alone cannot explain why only some hosts develop meningococcal disease. Our aim was to determine whether genetic diversity in CEACAM is associated with susceptibility to meningococcal disease. Frequency distributions of alleles, genotypes and haplotypes were compared in four CEACAM genes in 384 case samples and 190 controls. Linkage disequilibrium among polymorphic sites, haplotype structures and relationships were also analysed. A number of polymorphisms were observed in CEACAM genes but the diversity of CEACAM1, to which most Opa proteins bind, was lower, and a small number of high-frequency haplotypes were detected. Dose-dependent associations of three CEACAM haplotypes with meningococcal disease were observed, with the effect of carrying these haplotypes amplified in homozygous individuals. Two haplotypes were protective while one haplotype in CEACAM6 was associated with a twofold increase in disease susceptibility. These data imply that human CEACAM may be one determinant of human susceptibility to meningococcal disease.

Genes and Immunity (2008) 9, 30–37; doi:10.1038/sj.gene.6364442; published online 25 October 2007

Keywords: CEACAM; SNP; haplotype; meningococcal; susceptibility

Introduction

Adhesion of *Neisseria meningitidis* to the human nasopharyngeal epithelium is critical for establishment of invasive meningococcal disease,¹ a globally important infection that cannot be effectively controlled due to the lack of a comprehensive vaccine.² In Europe 1–6/100 000 people contract the disease annually, with a case fatality rate of approximately 8%.³ Asymptomatic carriage is more frequent however, with meningococci present in approximately 10% of the general population and up to 40% in 15–24 year olds.⁴ Both colonization and invasion require the specific interaction of Opacity-associated adhesin (Opa) proteins on the meningococcal surface with the carcino-embryonic antigen cell adhesion molecule (CEACAM) family of cell surface proteins on the

human nasopharyngeal epithelium.⁵ Most individuals are colonized asymptotically but, for incompletely understood reasons, adhesion in some hosts is followed by invasion of the mucosa, entry of meningococci into the blood stream and meningococcal disease. Most global meningococcal disease over the past half century has been caused by less than ten genotypes, known as the hyperinvasive lineages, as defined by multilocus sequence typing.^{6,7} As these genotypes are also carried asymptotically, their acquisition alone cannot explain all cases of invasive disease.

CEACAM proteins are encoded by 7 members of a 29-member gene family (the remainder being pseudogenes), in the q13.2 region of chromosome 19.⁸ Their usual biological functions range from cell adhesion, facilitation and regulation of signal transduction to a possible role in innate immunity.^{9–11} Structurally, these proteins belong to the immunoglobulin superfamily of surface proteins, consisting of an IgV-like N-terminal domain followed by a number of IgC-like conserved domains. In some members of the CEACAM family, dependent on variation in splicing, these are linked to the cell surface by a GPI anchor whereas in others, a transmembrane region and a cytosolic signalling domain are present. The

Correspondence: Dr MJ Callaghan, Department of Paediatrics, Centre for Clinical Vaccinology and Tropical Medicine, University of Oxford, Churchill Hospital, Headington, Oxford, Oxon OX3 7LJ, UK.

E-mail: martin.callaghan@paediatrics.ox.ac.uk

Received 1 June 2007; revised 24 September 2007; accepted 24 September 2007; published online 25 October 2007

extramembranous loop regions of Opa proteins, which exhibit high levels of diversity¹² interact with the nonglycosylated face of a β -pleated-sheet in the N-terminal domain of CEACAM proteins.^{5,13,14} Meningococcal Opa proteins have been shown to bind to HeLa cells expressing CEACAM1 (formerly known as BGP or CD66a), CEACAM3 (formerly known as CGM1a or CD66d), CEACAM5 (formerly known as CEA or CD66e) and CEACAM6 (formerly known as NCA or CD66c) and to soluble chimeric CEACAM1 N-terminal domains.^{14,15}

Despite their diversity, the majority of meningococcal Opa proteins maintain the ability to bind to at least one member of the CEACAM family and in some receptor-ligand pairs, this interaction leads only to attachment, whereas uptake into the host cells ensues in others.^{5,13–16} Upregulation of CEACAM expression *in vitro* leads to increased uptake of meningococci, even when encapsulated,¹⁷ further indicating the importance of CEACAM proteins to the meningococcal life cycle and pathogenesis of meningococcal disease.

Host genetics is thought to contribute one-third of meningococcal disease susceptibility¹⁸ but little is known about the population genetic diversity of the CEACAM family and its influence on susceptibility, which may be manifested at the level of interaction between bacterium and host. Genetic diversity in CEACAM may result in particular members of the human population being more easily colonized by meningococci, perhaps placing them at an increased risk of invasive disease. Furthermore, diversity in CEACAM may provide an evolutionary selection pressure driving the high diversity of the Opa proteins¹² for specificity to individual CEACAM variants. This study was undertaken to determine whether genetic differences in human CEACAM genes are associated with susceptibility to meningococcal disease.

Results

CEACAM genetic diversity

During curation of data from the initial 94 samples from each of the case and control groups, assays producing genotypes falling within Hardy–Weinberg equilibrium (HWE) and which failed in less than 15% of these initial samples were accepted. A total of 26 samples (18 case, 8 control) were removed due to their high assay failure rates, possibly reflecting DNA quality. Subsequent removal of assays across the remaining 76 cases and 86 controls left 9 single nucleotide polymorphisms (SNPs) in CEACAM1, 10 in CEACAM3 and 16 in each of CEACAM5 and CEACAM6. Monomorphic (noninformative) SNPs were removed, leaving a total of 38 informative SNPs. CEACAM1 was considerably less polymorphic than the other three genes with 6 of 9 SNPs monomorphic compared to 2 of 10 in CEACAM3, 2 of 16 in CEACAM5 and 3 of 16 in CEACAM6. A total of 13 tag SNPs, 2 in CEACAM1, 3 in CEACAM6 and 4 in each of CEACAM3 and CEACAM5 were chosen. These SNPs were used to investigate the genetic diversity in a final total of 384 cases and 190 controls, reflecting removal during data curation of 75 samples (59 case and 16 control) due to their high assay failure rates, again possibly due to poor sample quality. No significant difference in allele or genotype frequency was detected

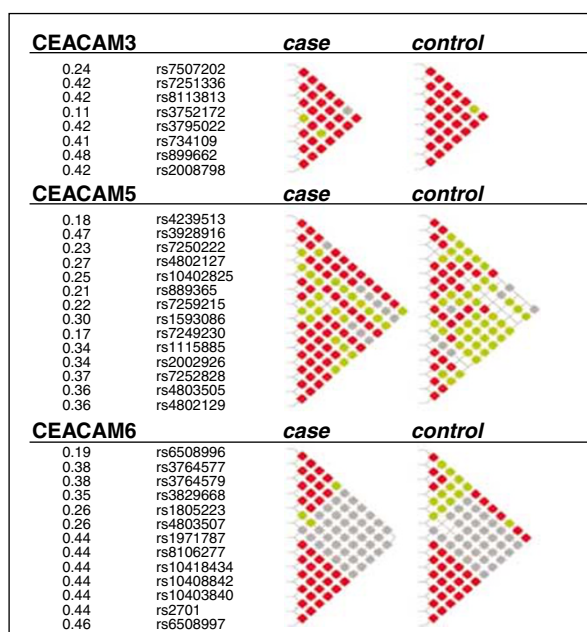


Figure 1 Linkage disequilibrium (LD) plots for SNPs in CEACAM3, 5 and 6. LD plots were constructed using the program MARKER and based on absolute D' values (a standard measure of LD based on the normalized deviation of the observed allele frequencies from the expected).¹⁹ Gene direction is 5'–3' from the top to the bottom of each map. The level of LD is indicated by color with high LD between sites (absolute D' of >0.9) indicated by red squares, through intermediate to high LD (absolute D' of $0.7–0.9$) indicated in yellow, intermediate LD (absolute D' of $0.5–0.7$) in gray and low LD (absolute D' below 0.5) in white. An LD map for CEACAM1 was not shown as only three polymorphisms were detected, which were in complete LD (data not shown). Numbers in the left hand column are major allele frequencies.

at any SNP between the final set of case and control samples (full data set available from authors upon request). N-terminal-encoding domain SNPs, which could directly affect Opa binding were also analysed. In CEACAM5, the SNP rs3815780 was monomorphic in both cohorts, whereas rs1805223 in CEACAM6 was polymorphic, but no significant differences in the allele frequency between cases and controls were observed.

Linkage disequilibrium (LD) analysis

LD within each gene was analysed using the 38 informative polymorphisms in the initial set of samples (Figure 1). The majority of sites in CEACAM1 and CEACAM3 were in complete LD. In CEACAM5, SNPs in cases exhibited apparently higher LD than in controls, especially 3' of the SNP rs7249230. The most striking feature of the LD plot for CEACAM6 was the appearance of two distinct high LD blocks (SNPs rs6508996–rs4803507 and rs1971787–rs6508997) with the majority of SNPs in either half of the gene not in linkage disequilibrium with those of the other half. In controls but not in cases, SNP rs6508996 was in higher LD with 6 of the 7 SNPs towards the 3' end of the CEACAM6 gene and in lower LD with its closest neighbours.

Haplotype analysis

Haplotype structure, phylogenetic relationships and frequencies were analysed and compared in the final case/control set (Figures 2a–d). In CEACAM1, four

major haplotypes were observed (Figure 2a), accounting for all case and control samples. The haplotypic diversity of CEACAM1 was dominated by haplotype A, observed in 94.1% of cases and 89.9% of controls. Haplotype D clustered together with the low frequency haplotypes B and C, representing a further 5.3% cases and 8.9% controls.

In CEACAM3, six major haplotypes were observed (Figure 2b), accounting for 99.4% of cases and 99.8% of controls. Haplotype A clustered separately and accounted for 40.3% of cases and 37.6% of controls. Three other haplotypes, B (23.4% of cases, 22.3% of controls), C (13.4% of cases, 21.5% of controls) and F (20.2% of cases, 17.1% of controls) were also common, whereas haplo-

types D and E together accounted for 2.1% of cases and 1.3% of controls.

In CEACAM5, nine major haplotypes were observed (Figure 2c) accounting for 95.8% of cases and 94.7% of controls. Two haplotypes in CEACAM5 accounted for approximately half of the haplotypic diversity in both cases and controls. Haplotype A accounted for 26.0% of cases and 24.5% of controls whereas haplotype E was observed in 26.2% cases and 27.0% of controls. Phylogenetic analysis revealed that the nine haplotypes were grouped into six subclades clustering into two major clades. The first major clade (haplotypes A–D) accounted for 35.7% of case haplotypes and 31.9% of control haplotypes. The second clade (haplotypes E–I)

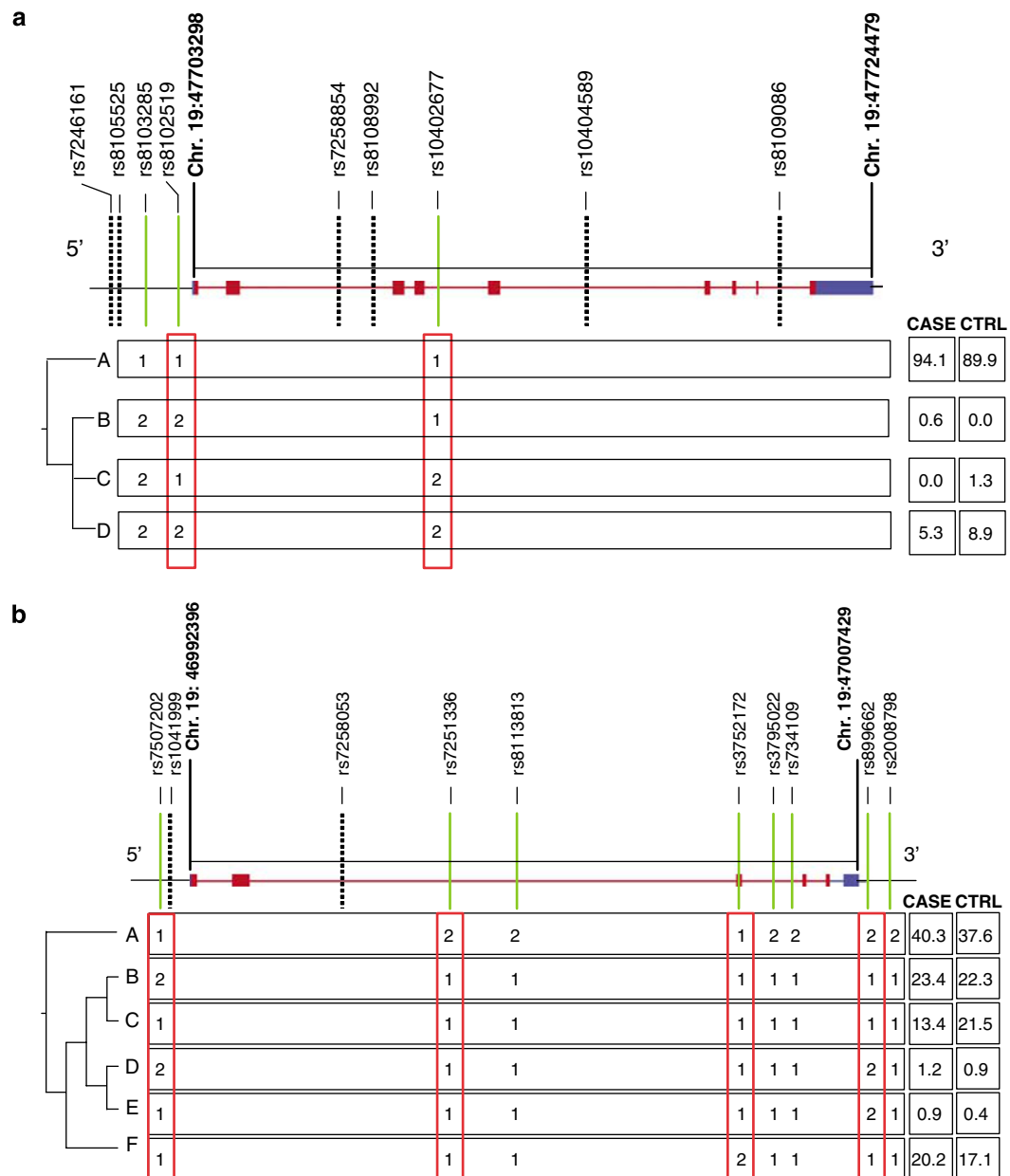
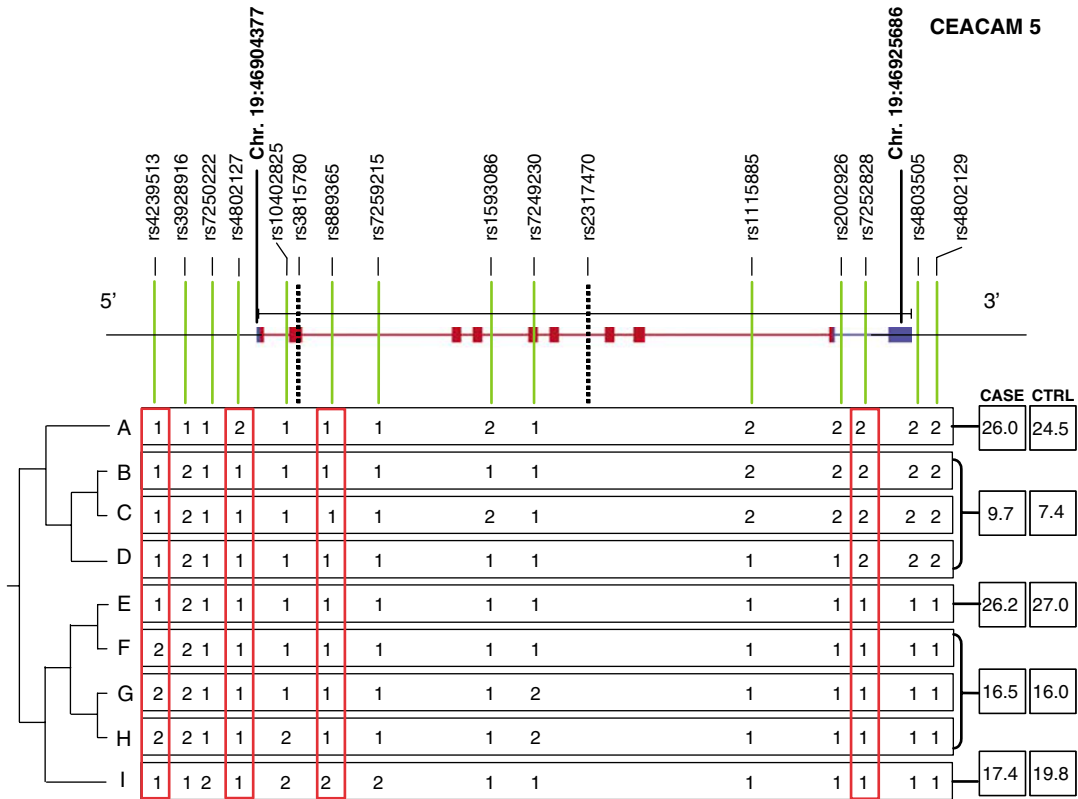


Figure 2 (a–d) Genetic diversity of CEACAM1, 3, 5 and 6 (a–d, respectively). Exons appear as solid red boxes on the gene architecture schematic while untranslated regions are shown as solid blue boxes. All SNPs analysed are indicated: solid green lines were polymorphic sites, whereas dashed lines were monomorphic in our cohort. Diversity at tag SNPs is indicated by red boxes. Phylogenetic relationships, indicating topology only, and case vs control frequencies are shown for haplotypes present at greater than 1% frequency in the sample collection. Groups of haplotypes detected by the chosen sets of tag SNPs are indicated by curved black lines.

c



d

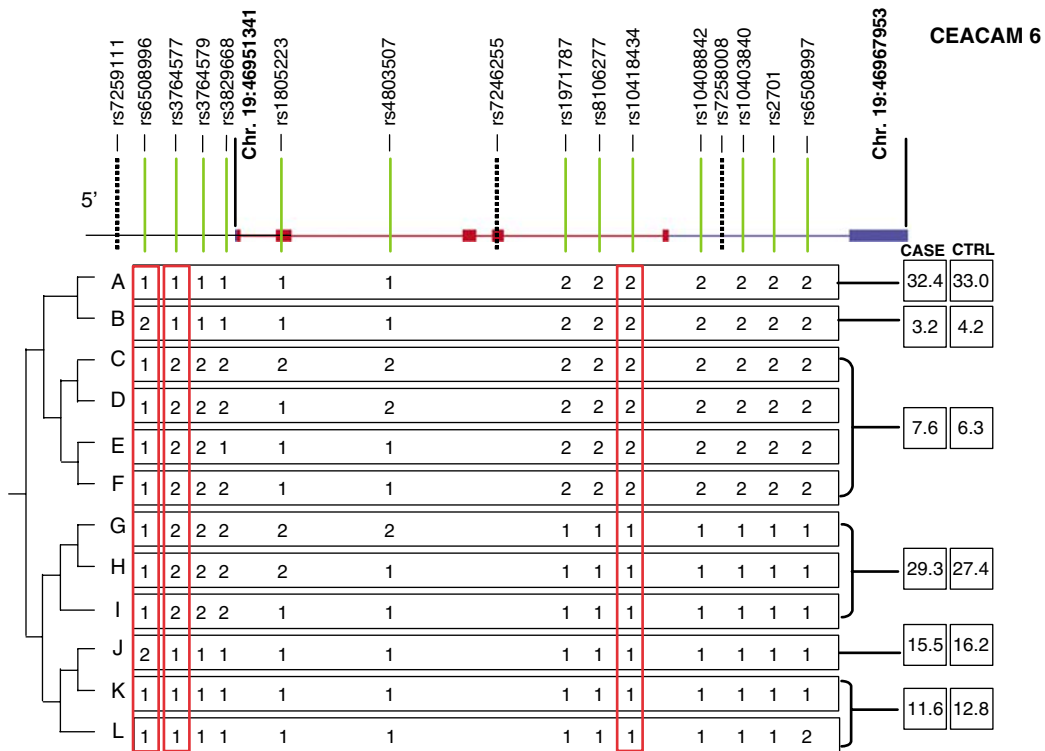


Figure 2 Continued.

accounted for the remaining 60.1% of case haplotypes in 62.8% of control haplotypes. This phylogenetic structuring was most likely caused by the approximately mirrored allelic differences towards the 3' end of the

gene after SNP rs7249230, consistent with LD analysis of this gene.

In CEACAM6, 12 major haplotypes (Figure 2d) accounted for 99.6% of cases and 99.9% of controls, with

haplotype A accounting for 32% of cases and 33% of controls. Haplotypes were grouped into 4–7 subclades forming two major clades. The clade containing haplotypes A–F accounted for 43.2% of case haplotypes and 43.5% of control haplotypes whereas the clade including haplotypes G–L accounted for the remaining 56.4% of haplotypes in both cohorts. As in CEACAM5, the phylogenetic differences in CEACAM6 were consistent with LD analysis of this gene, reflecting the mirrored allelic diversity in its 3' half, downstream of, and including SNP rs1971787.

Association of CEACAM diversity with meningococcal disease

The frequency distribution of haplotype C in CEACAM6 was significantly different between cases and controls (χ^2 $P=0.018$) and the effect of carrying this haplotype on meningococcal disease was dose dependent (Pearson's χ^2 $P=0.017$). Possession of this haplotype, which was observed in 7.6% of cases, was associated with an OR of 2.01 with 95% CI of 1.13–3.6 (RR of 1.21 with 95% CI of 1.01–1.46) for increased susceptibility to meningococcal disease. The frequency distribution of haplotype B in CEACAM6 (observed in 3% of cases) and haplotype C in CEACAM3 (observed in 13% of cases) was also significantly different between cases and controls (χ^2 both $P<0.001$) and their effects were also dose dependent (Pearson's χ^2 both $P<0.001$). CEACAM6 haplotype B had an OR of 0.29 with 95% CI of 0.14–0.61 and an RR of 0.57 with 95% CI of 0.36–0.9. CEACAM3 haplotype C had an OR of 0.52 with 95% CI of 0.35–0.75 and an RR of 0.79 with 95% CI of 0.64–0.96. No association between susceptibility to meningococcal disease and genetic diversity at either CEACAM1 or CEACAM5 was detected. The P -values presented are uncorrected. Use of the highly conservative Bonferroni correction would be inappropriate here since this correction assumes no linkage disequilibrium between SNPs.

Discussion

Here we describe the genotypic diversity of four CEACAM genes in a case/control study cohort of Caucasians to determine whether the diversity in these genes influences susceptibility to meningococcal disease. Three haplotypes in two CEACAM genes were found to affect susceptibility, which may exert their effects through two mechanisms. However, our data suggest that diversity in CEACAM is unlikely to drive the high diversity of the Opa proteins, adhesins that are important in mediating meningococcal pathogenesis.

Genotyping of the CEACAM genes revealed differences in the extent of their diversity relative to each other, each with a limited number of high-frequency haplotypes circulating in the Caucasian population. Based on the SNPs typed in this study, CEACAM1 was less polymorphic than CEACAM3, 5 and 6. This could reflect a necessity to maintain the structure and function of the CEACAM1 protein, and explain why the majority of Opa proteins tested bind to CEACAM1, whilst varying in their abilities to bind to those of other CEACAM proteins. Meningococcal Opa proteins have been shown to bind to HeLa cells expressing CEACAM1, CEACAM3, CEACAM5 and CEACAM6 and to soluble chimaeric CEACAM1 N-terminal domains.^{5,14–16} Further-

more, the comparatively lower level of polymorphism in CEACAM1 may also have led to its targeting by the adhesins of other bacterial pathogens including other *Neisseria* species (which also express Opa proteins), UspA1 of *Moraxella catarrhalis*,²⁰ the P5 proteins of *Haemophilus influenzae*²¹ and the Afa/Dr adhesins of *Escherichia coli*²² during adhesion to human hosts.

Our data indicate that CEACAM diversity contributes little in driving the high diversity of meningococcal opa genes,¹² suggesting that immunological selection pressures on these antigens, as in other meningococcal outer membrane proteins, including the PorA and PorB porins are more influential.^{23,24} Future functional studies aimed at more fully understanding the role of Opa proteins in meningococcal pathogenesis, together with that of CEACAM-binding adhesins in other bacterial species, may be aided by our observation that the genetic variation in their receptors is low in the Caucasian population.

Haplotype C in CEACAM6 was significantly associated with meningococcal disease in 7.6% of cases while CEACAM3 haplotype C and CEACAM6 haplotype B were significantly associated with protection against meningococcal disease. The effects of these haplotypes were dose dependent, amplified in homozygous individuals. Although uncorrected P -values were presented for the reasons given in the results section, these data warrant further investigation in a larger set of samples and of the functional role of these CEACAM haplotypes in susceptibility to meningococcal disease. Phylogenetic and LD analyses revealed striking features of the genetic structure of the CEACAM6 gene, with phylogenetically distinct mirror haplotypes and a block of high LD in its 3' half indicating a likely recombination breakpoint between SNPs rs4803507 and rs1971787. CEACAM5 also exhibited a similar, but less clear, pattern.

CEACAM diversity may influence human susceptibility to meningococcal disease in two ways. CEACAM proteins on the nasopharyngeal epithelium and on cells of the immune system are recognized as receptors for meningococcal Opa proteins,⁵ so susceptibility may be influenced by the effects of CEACAM diversity on meningococcal adhesion. Since CEACAM is also expressed on endothelia,¹⁵ it may play a role in mediating meningococcal entry into the meninges. In this study, there was no evidence of polymorphism in the Opa binding site, encoded by the first and second exons of each CEACAM gene, affecting disease susceptibility. Polymorphisms in the 5' extragenic regions upstream of CEACAM6 may, however, be associated with effects on the regulation of gene expression but further functional analyses would be required to test this hypothesis. There is clear evidence, however, that increased surface density of CEACAM proteins leads to increased Opa-mediated uptake of even encapsulated meningococci and expression levels of different CEACAMs on the same cells may also be important in differential mediation of cell signalling.^{17,25}

An alternative means by which CEACAM-mediated disease susceptibility may arise involves subversion of the human immune response. Interaction of gonococcal Opa proteins with CEACAM1 switches off and prevents proliferation of CD4+ T cells.²⁶ The effect of CEACAM diversity on immunity may not reflect a role in binding Opa, but may indicate an underlying immunological

defect, perhaps involving intercellular communication or contact. CEACAM3 and CEACAM6, along with other CEACAM proteins are thought to play a role in innate immunity.^{10,27} CEACAM3, which consists only of an amino-terminal IgV-like domain, is exclusively expressed on granulocytes and is involved in opsonin-independent phagocytosis and oxidative killing of human-specific bacterial pathogens.^{11,28} Taken in conjunction with our data, this may indicate that individuals with protective haplotypes in CEACAM3 and CEACAM6 have improved innate immunity against the meningococcus, whereas the disease-associated haplotype in CEACAM6 may indicate a defect in innate immunity. The underlying molecular mechanisms are unclear however, and would require further analysis of CEACAM polymorphism and function. Different expression levels of alternatively spliced CEACAM isoforms or alternatively posttranslationally processed CEACAM proteins may also play a role in mediating different signalling pathways and in innate immunity²⁹ and it may be useful in the future to investigate levels of CEACAM3 expression on granulocytes and CEACAM6 on other cells of the immune system.

A number of examples of polymorphisms in cell surface receptor-encoding genes affecting human disease susceptibility to a variety of pathogens have been reported. Promoter polymorphisms in CD209, encoding the C-type lectin DC-SIGN, a major *Mycobacterium tuberculosis* receptor, are associated with decreased risk of developing tuberculosis.³⁰ These polymorphisms appear to be more common in Eurasian populations than Africans, possibly reflecting a longer history of Eurasian exposure to tuberculosis. Homozygosity of tandem repeats in the CLEC4M gene, also known as CD209L and encoding the protein L-SIGN, is protective in severe acute respiratory syndrome coronavirus infection.³¹ Homozygotes for a 32bp deletion in CCR5, an HIV1-coreceptor gene, are highly resistant to the virus.³²

Among a range of suggested environmental, host and bacterial factors,³³ host genetic polymorphism is thought to contribute approximately a third of the total risk for invasive meningococcal disease.¹⁸ Previous studies have detected the influence of polymorphisms in genes involved in both the acquired and innate immune response, the inflammatory response and the coagulation/fibrinolysis pathway on aspects of meningococcal disease including susceptibility, severity and outcome.³⁴ Secondary familial cases of meningococcal disease³⁵ are also known to be associated with genetic polymorphism.^{18,36} Our data suggest that haplotype C in CEACAM3 and B and C in CEACAM6 could present additional targets for future investigations assessing familial risk.

Most meningococcal disease in the latter half of the 20th century was caused by a small number of genotypes known as the hyperinvasive lineages.^{6,7} In developed countries, these represent a small proportion of the asymptotically carried meningococcal population, causing endemic disease, localized outbreaks and epidemics. In the developing world, hyperinvasive lineages are responsible for periodic, large-scale epidemics and pandemics. It is tempting to speculate that genetic polymorphisms, perhaps including those detected in this study, contribute to the observed epidemiological differences by increasing the risk of meningococcal disease in particular populations.

Materials and methods

Patients and controls

A total of 387 Caucasian patients were recruited following either admission to the pediatric intensive care unit at St Mary's Hospital, London, United Kingdom between 1992 and 2002, or recruited from a UK meningococcal disease study overseen by the Royal College of Paediatrics and Child Health in which all fatal pediatric meningococcal disease cases between 1 December 1997 and 28 February 1999 were investigated. Confirmation of diagnosis was made by positive meningococcal culture from blood or cerebrospinal fluid (CSF), detection of increased meningococcal antibodies, or by PCR detection of meningococcal DNA in the blood or in CSF. In patients with no microbiological confirmation, meningococcal disease was diagnosed clinically upon presentation with petechial or purpuric rash and fever and features of systemic sepsis or meningitis where no other pathogen could be isolated. A further 56 samples from survivors of meningococcal disease were enrolled *via* the Meningitis Research Foundation (MRF) charity between 1996 and 1999. The patients' general physician or hospital consultant confirmed the diagnosis. There was no overlap between the three sources of patient recruitment, which together yielded 443 samples. Reliable statistical analysis of the influence of individual polymorphisms and haplotypes on the clinical severity of meningococcal disease was prohibited by the low number of samples from cases for which clinical disease severity scores were known.

A total of 206 DNA samples were extracted from the blood of healthy Caucasian individuals for use as controls, enrolled *via* both St Mary's Hospital ($n=42$) and the MRF ($n=164$). These individuals originated from throughout the United Kingdom and were non-related contacts of patients at the time of meningococcal disease onset, but who had not themselves contracted the disease.

All DNA samples were prepared using established techniques as previously described.³⁷ Ethical approval for the investigation of CEACAM diversity in this sample collection was obtained from Oxfordshire Local Research Ethics Committee A (OXREC number 04.OXA.024) and the St Mary's Local Research Ethics Committee (St Mary's LREC number EC3263) under whose approval the samples were originally collected.

Choice of SNPs

A total of 20 known single nucleotide polymorphisms (SNPs) per gene in the CEACAM 1, 3, 5 and 6 genes, which encode proteins that interact with meningococcal Opa proteins,^{14,15} were initially chosen from dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) and ENSEMBL (<http://www.ensembl.org/>) databases covering an ~832 kb region of chromosome 19. The SNPs appeared only once in the human genome sequence and first round PCR products gave only single bands when analysed by agarose electrophoresis (data not shown). SNPs in CEACAM4, 7 and 8 were not included due to the lack of evidence of their interaction with meningococcal Opa proteins at the time the study was designed. The possibility that some Opa proteins interact with these cannot however be discounted completely since the number of Opa proteins investigated is relatively low.

CEACAM1 SNPs were located on average once every 1409 bp. CEACAM3 SNPs were located on average once every 1507 bp. CEACAM5 SNPs were located on average once every 1880 bp. CEACAM6 SNPs were located on average every once every 1353 bp. A full list of SNPs analysed in this study is available upon request from the authors.

Genotyping

High-throughput genotyping was performed using the Sequenom MassARRAY system using the Homogeneous MassEXTEND assay as previously described.³⁸ Briefly, oligonucleotide primer pairs (Metabion, Martinsried, Germany) were designed using the Sequenom Spectro-DESIGNER software and used to amplify short sequences surrounding the chosen SNPs. Nonincorporated dNTPs were removed from amplicons using arctic shrimp alkaline phosphatase, before a third (universal extension) primer specific to each SNP allowed determination of allelic differences by primer extension reaction in the presence of SNP specific dNTP/ddNTP termination mixes. Allelic differences were determined by mass spectrometry using the Sequenom SpectroPOINT/SpectroCHIP system in conjunction with a Bruker Biflex III Mass Spectrometer. Raw data were autocurated electronically using the Sequenom SpectroTYPER software and manually checked to ensure accuracy and consistency of allele assignments.

Data analysis

Initially, a set of 94 samples from each of the case and control cohorts was investigated to choose informative assays for further haplotype and allele frequency analysis of the remaining cohort. Major and minor allele frequencies and overall genotype frequencies were calculated for all SNPs. HWE at each SNP and linkage disequilibrium within each CEACAM gene was analysed using MARKER (<http://www.gmap.net/marker>). HWE in cases and controls was tested separately and SNPs with $P < 0.001$ HWE were excluded. The program SNP-HAP (written by David Clayton, University of Cambridge, United Kingdom, accessed through a Pise interface at www.gmap.net) was used to construct and analyse haplotypes occurring at greater than 1% frequency in each CEACAM gene. Phylogenetic trees indicating relationships among haplotypes were constructed using the neighbour-joining algorithm implemented in the NEIGHBOUR program, part of the PHYLIP software package (Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, WA, USA).

In conjunction with manual analysis, the program ENTROPY (<http://www.well.ox.ac.uk/~rmott/SNPS>) in MARKER (www.gmap.net/marker), with default settings, was used to generate 'tag' SNPs that were informative of the haplotypic diversity in each CEACAM gene. These SNPs were used to investigate the genetic diversity of CEACAM genes in the remaining case and control samples and to investigate their associations with meningococcal disease.

The program SPSS was used to perform statistical analyses on the full data set. Allele, genotype and haplotype frequencies were compared between case and control cohorts by χ^2 analysis. Pearson χ^2 analysis

was then used to determine whether any associations were dose dependent. Odds ratio (OR) and relative risk (RR), both with 95% confidence intervals (CI), were calculated to determine the degree of association with meningococcal disease.

Acknowledgements

This study was funded by the Meningitis Research Foundation (project 02/02). DPK and KR are funded by the Medical Research Council. We wish to thank the staff at the core genomics and genotyping group at the Wellcome Trust Centre for Human Genetics, University of Oxford for their aid in Sequenom genotyping. AJP and MCJM are named as inventors, and MJC as a contributor, on a patent for the use of Opa proteins in meningococcal vaccines. All other authors declare no conflicts of interest.

References

- Nassif X. Interaction mechanisms of encapsulated meningococci with eucaryotic cells: what does this tell us about the crossing of the blood-brain barrier by *Neisseria meningitidis*? *Curr Opin Microbiol* 1999; **2**: 71–77.
- Jodar L, Feavers IM, Salisbury D, Granoff DM. Development of vaccines against meningococcal disease. *Lancet* 2002; **359**: 1499–1508.
- Trotter C, Samuelsson S, Perrocheau A, de Greeff S, de Melker H, Heuberger S *et al*. Ascertainment of meningococcal disease in Europe. *Euro Surveill* 2005; **10**: 247–250.
- Caugant DA, Hoiby EA, Magnus P, Scheel O, Hoel T, Bjune G *et al*. Asymptomatic carriage of *Neisseria meningitidis* in a randomly sampled population. *J Clin Microbiol* 1994; **32**: 323–330.
- Virji M, Makepeace K, Ferguson DJ, Watt SM. Carcinoembryonic antigens (CD66) on epithelial cells and neutrophils are receptors for Opa proteins of pathogenic neisseriae. *Mol Microbiol* 1996; **22**: 941–950.
- Caugant DA. Population genetics and molecular epidemiology of *Neisseria meningitidis*. *APMIS* 1998; **106**: 505–525.
- Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R *et al*. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 1998; **95**: 3140–3145.
- Thompson J, Zimmermann W, Osthus-Bugat P, Schleussner C, Eades-Perner AM, Barnert S *et al*. Long-range chromosomal mapping of the carcinoembryonic antigen (CEA) gene family cluster. *Genomics* 1992; **12**: 761–772.
- Hammarstrom S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol* 1999; **9**: 67–81.
- Gray-Owen SD, Blumberg RV. CEACAM1: contact-dependent control of immunity. *Nat Rev Immunol* 2006; **6**: 433–446.
- Kuespert K, Pils S, Hauck CR. CEACAMs: their role in physiology and pathophysiology. *Curr Opin Cell Biol* 2006; **18**: 565–571.
- Callaghan MJ, Jolley KA, Maiden MCJ. The opacity-associated adhesin repertoire in hyperinvasive *Neisseria meningitidis*. *Infect Immun* 2006; **74**: 5085–5094.
- Virji M, Watt SM, Barker S, Makepeace K, Doyonnas R. The N-domain of the human CD66a adhesion molecule is a target for Opa proteins of *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *Mol Microbiol* 1996; **22**: 929–939.
- Virji M, Evans D, Hadfield A, Grunert F, Teixeira AM, Watt SM. Critical determinants of host receptor targeting by *Neisseria meningitidis* and *Neisseria gonorrhoeae*: identification of

- Opa adhesiotopes on the N-domain of CD66 molecules. *Mol Microbiol* 1999; **34**: 538–551.
- 15 Muenzner P, Dehio C, Fujiwara T, Achtman M, Meyer TF, Gray-Owen SD. Carcinoembryonic antigen family receptor specificity of *Neisseria meningitidis* Opa variants influences adherence to and invasion of proinflammatory cytokine-activated endothelial cells. *Infect Immun* 2000; **68**: 3601–3607.
 - 16 de Jonge MI, Hamstra HJ, van Alphen L, Dankert J, van der Ley P. Mapping the binding domains on meningococcal Opa proteins for CEACAM1 and CEA receptors. *Mol Microbiol* 2003; **50**: 1005–1015.
 - 17 Bradley CJ, Griffiths NJ, Rowe HA, Heyderman RS, Virji M. Critical determinants of the interactions of capsule-expressing *Neisseria meningitidis* with host cells: the role of receptor density in increased cellular targeting via the outer membrane Opa proteins. *Cell Microbiol* 2005; **7**: 1490–1503.
 - 18 Haralambous E, Weiss HA, Radalowicz A, Hibberd ML, Booy R, Levin M. Sibling familial risk ratio of meningococcal disease in UK Caucasians. *Epidemiol Infect* 2003; **130**: 413–418.
 - 19 Lewontin RC. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 1964; **49**: 49–67.
 - 20 Hill DJ, Virji M. A novel cell-binding mechanism of *Moraxella catarrhalis* ubiquitous surface protein UspA: specific targeting of the N-domain of carcinoembryonic antigen-related cell adhesion molecules by UspA1. *Mol Microbiol* 2003; **48**: 117–129.
 - 21 Hill DJ, Toleman MA, Evans DJ, Villullas S, Van Alphen L, Virji M. The variable P5 proteins of typeable and non-typeable *Haemophilus influenzae* target human CEACAM1. *Mol Microbiol* 2001; **39**: 850–862.
 - 22 Berger CN, Billker O, Meyer TF, Servin AL, Kansau I. Differential recognition of members of the carcinoembryonic antigen family by Afa/Dr adhesins of diffusely adhering *Escherichia coli* (Afa/Dr DAEC). *Mol Microbiol* 2004; **52**: 963–983.
 - 23 Gupta S, Maiden MC, Feavers IM, Nee S, May RM, Anderson RM. The maintenance of strain structure in populations of recombining infectious agents. *Nat Med* 1996; **2**: 437–442.
 - 24 Urwin R, Holmes EC, Fox AJ, Derrick JP, Maiden MC. Phylogenetic evidence for frequent positive selection and recombination in the meningococcal surface antigen PorB. *Mol Biol Evol* 2002; **19**: 1686–1694.
 - 25 Rowe HA, Griffiths NJ, Hill DJ, Virji M. Co-ordinate action of bacterial adhesins and human carcinoembryonic antigen receptors in enhanced cellular invasion by capsule serum resistant *Neisseria meningitidis*. *Cell Microbiol* 2006; **9**: 154–168.
 - 26 Boulton IC, Gray-Owen SD. Neisserial binding to CEACAM1 arrests the activation and proliferation of CD4+ T lymphocytes. *Nat Immunol* 2002; **3**: 229–236.
 - 27 Hammarstrom S, Baranov V. Is there a role for CEA in innate immunity in the colon? *Trends Microbiol* 2001; **9**: 119–125.
 - 28 Schmitter T, Agerer F, Peterson L, Munzner P, Hauck CR. Granulocyte CEACAM3 is a phagocytic receptor of the innate immune system that mediates recognition and elimination of human-specific pathogens. *J Exp Med* 2004; **199**: 35–46.
 - 29 Chen D, Iijima H, Nagaishi T, Nakajima A, Russell S, Raychowdhury R *et al*. Carcinoembryonic antigen-related cellular adhesion molecule 1 isoforms alternatively inhibit and costimulate human T cell function. *J Immunol* 2004; **172**: 3535–3543.
 - 30 Barreiro LB, Neyrolles O, Babb CL, Tailleux L, Quach H, McElreavey K *et al*. Promoter variation in the DC-SIGN-encoding gene CD209 is associated with tuberculosis. *PLoS Med* 2006; **3**: e20.
 - 31 Chan VS, Chan KY, Chen Y, Poon LL, Cheung AN, Zheng B *et al*. Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. *Nat Genet* 2006; **38**: 38–46.
 - 32 Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R *et al*. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 1996; **86**: 367–377.
 - 33 Cartwright K. Meningococcal carriage and disease. In: Cartwright K (ed). *Meningococcal Disease*. John Wiley & Sons Ltd: Chichester, 1995, pp 115–146.
 - 34 Emonts M, Hazelzet JA, de Groot R, Hermans PW. Host genetic determinants of *Neisseria meningitidis* infections. *Lancet Infect Dis* 2003; **3**: 565–577.
 - 35 Cooke RP, Riordan T, Jones DM, Painter MJ. Secondary cases of meningococcal infection among close family and household contacts in England and Wales, 1984–7. *Br Med J* 1989; **298**: 555–558.
 - 36 Bathum L, Hansen H, Teisner B, Koch C, Garred P, Rasmussen K *et al*. Association between combined properdin and mannose-binding lectin deficiency and infection with *Neisseria meningitidis*. *Mol Immunol* 2006; **43**: 473–479.
 - 37 Haralambous E, Hibberd ML, Hermans PW, Ninis N, Nadel S, Levin M. Role of functional plasminogen-activator-inhibitor-1 4G/5G promoter polymorphism in susceptibility, severity, and outcome of meningococcal disease in Caucasian children. *Crit Care Med* 2003; **31**: 2788–2793.
 - 38 Ross P, Hall L, Smirnov I, Haff L. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat Biotechnol* 1998; **16**: 1347–1351.