Confidential

Protocol

Title of Investigation:

Genomic advances in sepsis (GAinS)

A multi-centre, UK based, genomic association study of the genetic determinants of the susceptibility to, and outcome from severe life-threatening infection and sepsis

Organising body:

UK Critical Care Genomics (UKCCG)

MEMBERSHIP OF THE UKCCG

Principal-Investigators

Prof. C.J. Hinds (Intensive Care Unit, St. Bartholomew's Hospital, London)
Dr C.S. Garrard (Intensive Care Unit, John Radcliffe Hospital, Oxford)

Co-Investigators

Dr G. Bellingan (Intensive Care Unit, University College Hospital, London)
Professor M. Caulfield (Director, Genomics Centre, William Harvey Research Institute, London)
Professor G.A. Hitman (Department of Metabolic Medicine, Barts and The London Queen Mary School of Medicine and Dentistry, London)
Dr N.J. Klein (Institute of Child Health, Great Ormond Street Hospital, London)

Collaborators

Professor A. Hill (Wellcome Trust Centre for Human Genetics, Oxford)
Professor D. Kwiatkowski (Wellcome Trust Centre for Human Genetics, Oxford)
Dr M. McDermott (Department of Metabolic Medicine, Barts and The London Queen Mary School of Medicine and Dentistry, London)
Dr R. Mott (Wellcome Trust Centre for Human Genetics, Oxford)
Professor D. Perrett (Department of Bio-Analytical Science, Barts and The London Queen Mary School of Medicine and Dentistry)

Investigators and Members

Ms E. Aganna (Department of Metabolic Medicine, Barts and The London Queen Mary School of Medicine and Dentistry)
Dr J. Bion (Intensive Care Unit, Queen Elizabeth University Hospital, Birmingham)
Dr S. Brett (Intensive Care Unit, Hammersmith Hospital, London)
Professor T. Evans (Intensive Care Unit, Royal Brompton Hospital, London)
Dr G. Findlay (Intensive Care Unit, Cardiff)
Dr H. Galley (Intensive Care Unit, Aberdeen Royal Infirmary)
Dr A.C. Gordon (Intensive Care Unit, Royal Brompton Hospital)
Ms P. Hutton, Intensive Care Unit, John Radcliffe Hospital, Oxford)
Dr A. Kapila (Intensive Care Unit, Royal Berkshire Hospital, Reading)
Dr A. McLuckie (Intensive Care Unit, Guys/St Thomas' Hospital, London)
Dr R. Marshall (Intensive Care Unit, University College Hospital, London)
Dr J. Millo (Intensive Care Unit, John Radcliffe Hospital, Oxford)
Dr H. Montgomery (Intensive Care Unit, University College Hospital, London)
Dr G. Quinlan (Intensive Care Unit, Royal Brompton Hospital, London)
Dr J. Pappachan (Intensive Care Unit, Southampton General Hospital,

Hampshire)
Dr A. Short (Broomfield Hospital, Chelmsford, Essex)
Professor. David Stuart, (Wellcome Trust Centre for Human Genetics, Oxford)
Dr D. Watson (Intensive Care Unit, St. Bartholomew's Hospital, London)
Professor N. Webster (Intensive Care Unit, Aberdeen Royal Infirmary)
Dr D. Young (Intensive Care Unit, John Radcliffe Hospital, Oxford, Research Director, Intensive Care Society)

IMPORTANCE

Sepsis is the systemic inflammatory response to infection and is the commonest cause of death in adult intensive care units. There are up to 750,000 cases of severe sepsis a year in the USA [1], and around 21,000 cases per annum in the UK, while the incidence of severe sepsis in hospitals has been conservatively estimated at 2 per 100 admissions [2] and the incidence amongst patients in intensive care units at around 6-10 per 100 admissions [1,3]. In the UK severe sepsis accounts for about 46% of all bed days in intensive care units. Mortality rates, which are closely related to the severity of illness and the number of organs which fail are high (20-60%) and there may be more than 200,000 deaths from severe sepsis every year in the USA. [1], and more than 1400 deaths per day worldwide. The impact on health care expenditure and resource utilisation has been considerable (annual total hospital costs for these patients in the USA have been estimated at approximately $17 billion and in Europe at 7.6 billion euros).

Early attempts to combat the high mortality associated with sepsis concentrated on cardiovascular and respiratory support. Despite some success, [4] mortality rates remained unacceptably high and often death was merely postponed until they were overwhelmed by a dysfunctional host response (characterised by persistent or recurrent sepsis, intractable hypotension and failure of vital organs). Current evidence suggests that this response is largely independent of the site of infection and the responsible organism. Efforts to further reduce mortality by manipulating haemodynamics or modulating the host response have generally proved disappointing [5-7], perhaps in part because of our limited understanding of the complex mechanisms which regulate innate immunity and the inflammatory response. Importantly, also such interventions have usually been applied unselectively to heterogeneous groups of patients, without considering the potential influence of their genetic diversity on the response to treatment.

It has long been recognised that individuals vary considerably in their susceptibility to infection as well as in their ability to recover from apparently similar infectious illnesses. In 1988 it was clearly demonstrated that premature death in adults, especially when due to infectious and vascular causes is a strongly heritable trait [8]. By studying adoptees these investigators showed that the death of a biological parent from an infectious cause before the age of 50 years was associated with a relative risk of 5.81 for premature death of the natural child from infection, and that even for parental deaths at over 70 years the relative risk was 5.0. Conversely the death of an adoptive parent from infection resulted in relative risks in the adoptees that were close to unity. Since then a number of relatively small studies have indicated that individual variations in susceptibility to, and outcome from severe sepsis/septic shock may be explained in part by polymorphisms of the genes encoding proteins involved in mediating and controlling the innate immune response and inflammatory cascade [9-11]. There is also evidence that such polymorphisms play a role in the development of certain organ failures. For example members of our group have reported that a polymorphism of the angiotensin converting enzyme gene is associated with susceptibility and outcome in acute respiratory distress

syndrome (ARDS)[12]. It is, therefore highly likely that susceptibility to infection and the development, progression and outcome of sepsis/septic shock is influenced by genetic variants [9-16] that might, in turn also influence the response to pharmacological interventions in sepsis.
Such simple candidate gene association studies must, however, be interpreted cautiously, not least because sepsis is a complex polygenic disorder and the functional importance of these polymorphisms is frequently uncertain; indeed many may simply represent a genomic marker for other more functionally relevant genetic variations with which they are in linkage disequilibrium (LD). Moreover results to date have often been inconsistent and failure to replicate positive findings has been frequent [17,18]. For example some studies have failed to confirm the association between the TNF-308 single nucleotide polymorphism (SNP) and outcome from severe sepsis/septic shock [17] and, in a study more than three times larger than any reported to date, the applicants and their collaborators were unable to demonstrate any associations between several individual TNF/LTA polymorphisms, or their haplotypes, and susceptibility or outcome in such patients [18].
It is now widely accepted that these inconsistent results can be explained in part by the limited power of the studies performed so far, over-interpretation of marginal results in small samples (sometimes compounded by multiple testing), the heterogeneity of the patient populations, the complexity of some haplotypic structures, unrecognised confounding effects, inadequate definition of phenotypes, ethnic differences and variable quality control of genotyping techniques and statistical analysis [19,20]. Of particular relevance to this application the central major histocompatibility complex (MHC) has a very high density of genes, many of which are involved in immunity and inflammation, and a complex haplotypic structure [21] which considerably complicates interpretation of SNP associations in this region.

Aims

There is an overwhelming need for large collections to permit adequately powered genomics research in septic patients. The aims of this proposal are:
1.To deliver
- A high quality phenotypic resource of large, well-characterised, homogeneous groups of patients with, or at risk of, sepsis.
- A repository of DNA from these patients, (which will include whole genome amplification) for national and international collaboration.
- A resource of blood and urine samples and cryopreserved cell lines for functional genomic, proteomic and metabonomic analysis.

2.To perform high quality, functional association studies to determine the influence of a discrete number of candidate genes, alone and in combination on the development, progression and outcome of sepsis/severe sepsis/septic shock.
3.To undertake genome wide association studies using this resource when practicable and affordable.
The long term objective is to substantially reduce the morbidity, mortality and costs associated with overwhelming infection and systemic inflammation by yielding new insights into the pathogenesis of sepsis and organ failure,

discovering new targets for therapy or prevention, and identifying individuals who might benefit from specific treatments or preventative measures.

SCIENTIFIC POTENTIAL

People and track record

This collaboration brings together internationally recognised research teams already active in the field to form the **UK Critical Care Genomics** group (UKCCG), thereby generating the critical mass necessary to achieve our objectives. The UKCCG group has worked effectively as a steering committee to build a unique collaborative consortium of 30 major ICUs and 4 co-ordinating centres throughout the UK. Several of these can offer substantial field experience and are partnered by 2 established Genomics Centres with a strong portfolio of complex trait research programmes. The proposed studies will therefore be undertaken by a cohesive, committed group of participating units with a common approach to treatment. The applicants include clinical academics with many years experience of research in the intensive care environment, including the conduct of large clinical trials and smaller genetic association studies. Importantly the UKCCG group also includes highly regarded infectious diseases experts, geneticists, biochemists, epidemiologists and bio- statisticians with extensive experience of performing large, genetic association studies.

The group has established close links, and has collaboration agreements already in place with investigators performing allied studies in the USA (GenIMS) and Europe (the European Society of Intensive Care Medicine, European Critical Care Research Network GenOSept Study). Our proposed study is endorsed by the European Society of Intensive Care Medicine. Participation in such large international studies will allow pooling and replication of data, thereby considerably increasing the power of our collection to detect smaller relative risks. Collaboration will also provide access, if required, to additional analytical, epidemiological and statistical expertise.

Environment

We will use existing laboratory resources at the Barts and the London Genome Centre in the William Harvey Research Institute (WHRI) and the Wellcome Trust Centre for Human Genetics (WTCHG). Importantly this proposal will also provide a platform from which to pursue the international collaborations just mentioned as well as many allied research and training initiatives.

Proposed management structure: The study will be co-ordinated by Hinds and Garrard on behalf of the UKCCG; they will review progress, set priorities on a day to day basis, ensure patient confidentiality and will have immediate responsibility for management of the centralised collection of DNA, blood and urine samples. Custodianship of the resource will rest with the host institution. A *Management Committee* was recently convened and consists of the named applicants/collaborators, and the lead investigator (or deputy) from each of the co-ordinating centres, as well as an independent invited expert in intensive care research. The committee will meet regularly to review progress, establish priorities, ensure targets are being met and implement forward planning initiatives. An administrator has been appointed to co-ordinate data collection,

provide quality control and report to the management committee. A representative of the research nurses will also attend, and report to, this committee. A *Steering Committee*, chaired by an independent expert in complex trait mathematical genetics and the establishment of large-scale genetic resources, will meet as required to approve arrangements for access to the resource, and the exploitation of intellectual property and will operate in accordance with MRC Human Sample Collection Guidelines. At these meetings participating centres will have the opportunity to submit proposals for investigations to be pursued by the whole group or to request access to elements of the resource. This committee would also review and adjudicate external applications to access the collection, including those emanating from international colleagues. All projects utilising the resource will first be subject to peer review and time limits will be set for the completion of such projects, as well as for copies of the resulting data to be deposited into the common database. The administrator will maintain a record of the distribution of precise quantities of samples and users will be expected to account for, and return, material surplus to their requirements. Released material will not be used for additional studies or passed on to others. There will be meetings at least once a year for all study participants.

RESEARCH PLANS

General approach

Our hypothesis is that functional polymorphisms of genes involved in regulating the host response to infectious insults, and gene-gene interactions, will influence the levels and/or activity of key proteins and hence the onset, progression and severity of an infectious illness, the development of organ failures and outcome. The proposed data collection is designed to enable us to identify with confidence genomic influences on:

1. The susceptibility of patients with predisposing infectious diseases to the development of severe sepsis/septic shock and specific organ failures.
2. The outcome from sepsis/severe sepsis/septic shock in these patients.
3. The outcome from specific organ failures in these patients.

To address the difficulties encountered in previous studies we propose to:

a) Collect sufficient patient numbers over 3 years to ensure that the study is adequately powered and exploit our agreements to access international DNA collections when required (eg. when relative risks are very low and for subgroup analyses).

b) Recruit patients with rigorously defined infectious aetiologies.

c) Focus on the differences (qualitative and quantitative) in the host response to these infections and the consequent outcomes (development of organ failures, length of stay, time of death), as well as death or survival.

d) Establish an accurate, relevant and robust clinical database.

e) Structure candidate genes into key elements of the host response (e.g. pro- and anti-inflammatory, innate immunity, programmed cell death, coagulation, metabolic) and identify candidate systems within these groups (see genotyping strategy).

f) Use robust, “state of the art” statistical genetics methods.

g) Provide sufficient genotypic and phenotypic material to enable complex, extended haplotype analysis, which will be supported and informed by confirmatory functional data. Local centres will provide additional functional data in their particular field of interest.

h) Positive associations identified by analysing an initial cohort will later be replicated using the remainder of the collection and the European and North American DNA resources. This approach will also allow subsequent expansion of targeted subgroups.

Patient populations

It is recognised that accurate description of phenotype (including the infecting organism where possible) will be crucial to the success of this project. For this reason we intend to recruit patients with easily identified underlying infectious diseases (community acquired pneumonia - CAP, faecal peritonitis - FP) predisposing to the development of sepsis and organ failure. Additional advantages of studying these patient populations include a time of onset that is usually readily identified (especially for FP), and the availability of equivalent animal models (caecal ligation and puncture in rodents and direct bacterial innoculation, including transgenic/knockout mice). In a small subgroup of patients with FP from our pilot collection [15,18] we were able to confirm our observation [15] of a high incidence of Mannose Binding Lectin (MBL) intron 1 allelic variants in patients with sepsis compared to the normal population. Our choice of these patient groups is complementary but distinct from those chosen for the EU funded GenOSept study (nosocomial pneumonia, pancreatitis, meningococcal disease and peritonitis). Although a genetic association study in patients with CAP is in progress in the USA findings are likely to reflect the particular demography of their population and their local prevalence of bacterial pathogens. Also our proposed methodologies differ in some important respects (in particular the US study includes CAP patients not admitted to ICU but the proportion of patients with severe sepsis is comparatively small). We have established formal agreements with the US and European investigators in order to maximise the potential of these complementary databases.

Inclusion criteria

We will recruit patients more than 18 yrs of age admitted to the high dependency units (HDU) or intensive care units (ICU) in the participating centres, with community acquired pneumonia (CAP) [defined as in reference 22 – febrile illness associated with cough, sputum production, breathlessness, leucocytosis and radiological features of pneumonia acquired in the community or within less than 2 days of hospital admission.] or faecal peritonitis (FP) [defined as inflammation of the serosal membrane that lines the abdominal cavity secondary to contamination by faeces as diagnosed at laparotomy]. The use of standard diagnostic criteria for CAP and an unambiguous definition of FP, together with adjudication by two experienced clinicians will ensure accurate diagnosis. The commonest causes of FP are diverticular disease, colonic malignancy and surgical anastomotic breakdown. More unusual causes, such as penetrating trauma, would also be included if such cases were encountered. Peritonitis due to perforation of the upper gastrointestinal tract is excluded from this definition.

For each case of FP the underlying cause will be documented and for CAP the causative organism (where known) will be documented.
These patients may be admitted with, or may later develop, and progress through the spectrum of responses to infection, sepsis/severe sepsis/septic shock defined according to the extensively validated and widely accepted "Society of Critical Care Medicine/American College of Chest Physicians" consensus criteria definitions [23] and may be admitted with, or later develop specific organ failures (e.g. Acute lung injury/Acute respiratory distress syndrome as defined in reference 24) which will be scored using the extensively validated and widely accepted "Sepsis-related Organ Failure Assessment" (SOFA) methodology[25].

Exclusion criteria

Patients will be excluded when: patient, next-of-kin or legal representative unwilling or unable to give informed consent; patient <18 yrs of age; patient already enrolled in an interventional research study of a novel / unlicensed drug / therapy (patients enrolled in interventional studies examining the clinical application or therapeutic effects of widely accepted, "standard" treatments are not excluded); patient pregnant; advanced directive to withhold or withdraw life sustaining treatment or admitted for palliative care only; patient immunocompromised (known regular systemic corticosteroid therapy, exceeding 7mg/kg/day of hydrocortisone or equivalent, within three months of admission and prior to acute episode, known regular therapy with other immunosuppressive agents, e.g. azathioprine, known to be HIV positive or have acquired immunodeficiency syndrome as defined by the Centre for Disease Control, neutrophil count less than 1000 mm^{-3} due to any cause including metastatic disease and haematological malignancies or chemotherapy, but excluding severe sepsis; organ or bone marrow transplant receiving immuno-suppressive therapy.

Patient numbers

FP represents 2.8% of all admissions to ICUs in the UK according to the most comprehensive Intensive Care National Audit and Research Centre (ICNARC) data set. On the basis of admissions data collected over the last few years in each centre it is estimated that the 30 participating units will admit between them more than 2000 cases of FP over the three year period of the collection. Pneumonia represented 7.2% of all admissions in the ICNARC data set with a mortality of 40% and we estimate that the group will admit more than 5000 cases of CAP over the three year period of the collection. It is anticipated that further recruiting centres will come on board as the study gathers momentum. Based on these conservative estimates the final resource will therefore be large and

adequately powered and will be many times greater than the largest sepsis association studies reported to date. Its particular strength is that well-defined phenotypes will be recruited, together with the establishment of a comprehensive database and functional correlates (see below for discussion of control populations).

End Points

Primary end points will be based on:

1.Susceptibility to:

- Development of sepsis/severe sepsis/septic shock in patients with CAP or FP
- Development of specific organ failures identified by the SOFA score [25].

2.Outcome as measured by:

- Death or survival (in intensive care, in hospital and at 6 months) in all patients

Secondary end points will include:

1.Severity of illness using APACHE II on admission and SAPS II (scoring systems which assess the degree of physiological derangement)[26].

2.Duration of respiratory and other organ support.

3.Shock reversal

4.Duration of ICU and hospital stay.

Identifying comorbidity

Pre-defined comorbidities, such as respiratory or renal disease and malignancy will be recorded and scored using a Medical History Questionnaire (Charleson Index based)

Microbiological considerations

Faecal Peritonitis: Heterogeneity attributable to differences in the infecting organism will be minimised in those with FP by the fact that in all cases infection will be with mixed intestinal flora. Except in patients who have previously been exposed to prolonged antibiotic therapy the faecal bacterial flora remains largely predictable in profile and quantity. In all cases every effort will be made to identify the organism(s) causing bacteraemia from blood cultures (routine investigation in FP).

Community Acquired Pneumonia The majority of cases of CAP are caused by a small range of key pathogens [27]. The predominant pathogen is *Streptococcus pneumoniae*, which accounts for about two thirds of all cases of bacteraemic pneumonia. In patients admitted to the ICU with severe pneumonia the most common causes are *S. pneumoniae, Legionella spp*., *H. influenzae*, gram-negative bacilli and *S. aureus*. Overall a causative organism can be identified in around 50% of cases using conventional tests (sputum, blood cultures and serology). Importantly S. pneumoniae has also been found to be responsible for one third of cases not documented by conventional testing [27].

Recruitment and Phenotyping.

All patients fulfilling the inclusion criteria will be eligible for recruitment to the study unless exclusion criteria are met. The applicants, and some other members of the UKCCG, have a unique experience in recruiting critically ill patients with sepsis to genetic association studies. Furthermore the 30 recruiting ICUs will function according to standard operating procedures to ensure rigorous quality control of phenotyping. The research nurses will ensure that all eligible patients

are identified and considered for enrolment at each participating hospital by daily telephone and e-mail contact with the operating theatre and HDU/ICU, as well as by regular site visits. Recruitment rates in individual centres will be monitored throughout the study by the research nurses and by the administrator. Eligible patients who are not recruited will be recorded, together with the reason for failure to recruit. All patients recruited with FP or CAP will be formally reviewed on days 1,2,3,5 and 7 for the development of sepsis/severe sepsis/septic shock or specific organ failures. Phenotypic data will also include severity of illness scoring and co-morbidities at recruitment, organ failure scores for days 1,2,3,5 and 7, microbiological findings, duration of organ support and discharge and outcome information.

Databases

Local investigators will enter demographic and phenotypic data onto a bar-coded paper CRF at the bedside. This CRF has been developed and tested in some of the participating units. The CRFs will be securely stored locally and copied to the genome centre, where they will be archived, and the data entered independently by two part time data entry clerks, with reconciliation by the administrator, into a secure, central web-based electronic database for storage of clinical data and the calculation of derived values such as APACHE II, SAPS II and SOFA. The patient database will be anonymised and maintained in the WHRI on secure servers with password access only and local automated digital tape backup, as well as regular remote backup to Oxford. Transferred data will be encrypted. The system will be written in mySQL and run on a UNIX server. The patient codes for genetic analysis will be derived directly from the clinical database. However, those undertaking genotyping will be blinded to the phenotype and these two databases will be brought together at the time of analysis only under the direct supervision of one of the applicants. For the large proportion of units participating in the UK Intensive Care National Audit and Research Centre (ICNARC) programme additional case mix adjustment, including all critically ill patients admitted to those units, will be possible. In addition the clinical phenotype database will provide an important resource for studying the epidemiology of sepsis in patients with FP and CAP; and will be made available for use in such studies.

Sample handling and storage.

In developing our sample handling strategy we have drawn on the UK Biobank Sample handling and storage subgroup protocol and recommendations (http://www.ukbiobank.ac.uk). Following consent or assent, two 10 ml whole blood samples will be drawn into barcoded EDTA tubes and stored at room temperature until dispatch to the genome centre. Bar-coded 2d unique study number labels will be used to anonymously label all blood samples, consent forms and CRFs with centre specific codes and individual IDs. DNA will be extracted from the anticoagulated whole blood using standard techniques and quantified using the picogreen assay. Samples will be normalised to a standard concentration using an automated protocol that is designed to reduce the possibility of sample misassignment. Samples for DNA can be drawn at any time during the patients' admission. In 15 of the most research active ICUs, additional

whole blood (10 ml) and urine (20 ml) will also be obtained on days 1, 3, and 5 of ICU admission, for functional studies including proteomics, metabonomics and determination of circulating levels of the relevant protein mediator by radioimmunoassay. Additional blood (in 8.5ml ACD tubes) will be obtained in the "hub" units for cryopreservation. Blood samples will be cold-centrifuged at 2,500 g for ten minutes. The resulting plasma will be immediately (within 15 minutes) frozen and stored at -80^{o}C as eight, 500 µl aliquots, Urine will be stored in 10 ml aliquots at -80^{o}C. The red/ white cell plug will be stored at –20^{o}C as a back up sample for DNA extraction.

Sample tracking and data warehousing: We will use bar coded sample tracking from acquisition label packs through to genotype generation. Split samples will be securely stored at two distinct sites. A computer database of stored samples will be maintained for logging and ease of identification.

Generating a repository for National and International collaboration: There has been rapid progress in the reliability of whole genome amplification (WGA) methods, especially for single nucleotide polymorphism analysis. We have favourable experience of the reliability and cost of the GenomiPhi DNA Amplification kits (Amersham) and Molecular Staging Kits for 800 fold amplification from genomic template within our collaboration and believe we can deliver this very competitively in house. WGA will enable us to maintain a substantial but not inexhaustible supply of DNA for our group and other collaborators. For functional genomics and proteomic studies the lymphocyte is an ideal tool for investigating genes that may affect immunological responses. We therefore propose cryopreserving 200 cell lines from each sepsis phenotype that could also be valuable for haplotype definition in candidate gene studies.

Quality control of the database and sampling. In order to achieve and maintain consistency in phenotyping, database entry and sample collection across all sites standard operating procedures will be developed, guided by expertise derived from the diabetes and hypertension genetic programmes, accompanied by regular training updates for the research nurses and local participants

Genotyping.

Our genotyping strategy is based on complementary experience at two of the genome centres involved in this consortium (Barts and the London Genome Centre and WTCFHG, Oxford) and a commercial company (KBioscience). The two genome centres are already at the forefront of large UK based studies (including The Diabetes UK/MRC Warren 2 type 2 diabetes study and the MRC hypertension BRIGHT study) and have complementary areas of experience, techniques and expertise. By using this established genome centre infrastructure, and drawing upon experience shared with KBioscience we can ensure that the group has access to the full range of genotyping techniques using microfluidics at a very economic price. These would include the Taqman (ABI 7900HT system), SNAPSHOT (on ABI3700s), pyrosequencing; and the SEQUENOM MASSARRAY technology, as well as the necessary bioinformatics, quality control and statistical support. Both genome centres have an active programme of technology evaluation and have access to Affymetrix, ABI and

Illumina platforms that are being developed for whole genome association methods. It is recognised that in this rapidly developing field new techniques are likely to be introduced during the life of this project. Where appropriate the group intends to exploit such developments.

In the first instance the resource will be used to investigate a defined set of 20 candidate genes (see below) involved in recognised aetiological mechanisms using automated high-throughput genotyping. We will identify a list of candidate genes and polymorphisms that have been shown to have a direct effect on gene function (based in part on the results of smaller hypothesis generating association and functional studies including a number recently conducted by members of the group) and then compile a SNP inventory *in silico* (dbSNP and other informatic tools) and supplement this with SNP detection by dHPLC or direct sequencing as necessary. For each locus these would be used to define haplotype blocks and the minimum number of tag SNPs that could be used to define 95% of the haplotypes. Given the current state of knowledge our first priority will be to resolve the uncertainty regarding the importance of some of the more promising candidate genes previously investigated only in small studies. Although in all cases the final decision will rest with the Management Committee, these may include: TNF/LTA and the central MHC region, TNF receptors, IL-1, IL-6, IL-10, bactericidal/permeability increasing protein, lipopolysaccharide binding protein, CD-14, TLR-2, TLR-4, plasminogen-activator-inhibitor 1, ACE, MBL and complement. Novel genes to be investigated may include those involved in fibroproliferation and the resolution of inflammation, apoptosis, mitochondrial function, lipoprotein production and intracellular signalling pathways.

For the nested case control studies (examining genetic determinants of outcome) all SNPs will be typed directly. For individual genotyping we envisage the main genotyping platforms being TAQMAN and SEQUENOM. TAQMAN-based typing is well established at the WHRI, utilising the 5' exonuclease assay with the new modified minor grove binding (MGB) probes and we have recently introduced microfluidic robotic dispensing. This system, coupled with the robotic assay set up in the Genome Centre can theoretically achieve approximately 96000 genotypes per week at a highly competitive cost (8.5p per genotype). Similarly at the WTCFHG the MASSARRAY system is capable of 200,000 SNPs a week.

Quality control of genotyping.

Currently, there is no agreed standard for quality control that should be applied to such large-scale genotyping studies. We will include 10% duplicates (blinded to the reader) for all markers in the total study population for the primary genotyping method. Software programmes will be used to run automated data quality checks on genotype consistency and accuracy. All genotypes will be scored independently and blindly by 2 technicians. To verify validity of SNP allele calling 25 subjects will be directly forward and reverse sequenced using an ABI 3700. Any markers that are not in Hardy-Weinberg Equilibrium, that are difficult to interpret or for which there is a discrepancy in duplicates, will be repeated by an additional method (currently Ampliflour or SEQUENOM). For this study a 0.5% error rate will be above the threshold to use an additional genotyping method.

HLA typing will be performed to address the issue of genetic diversity. Population specific alleles will be used to test for admixture in the resource and exclude false positive associations. Geographic differences will be monitored.

Genomic databases and bio-informatics.

Recent important advances in genomic databases and bioinformatics offer tremendous opportunities for sophisticated interrogation and display of genomic information. It is now possible to electronically search public domain databases to define and order contents of a genomic region, search using sequence, examine transcript maps, expressed sequence tags (ESTs), microsatellite and SNP databases. The core bioinformatics team within the Barts and The London genome centre can exploit the full range of publicly available genome Annotation Tools (e.g. the "Golden Path" – **http://genome.ucsc.edu**), "EnsEMBL" – http//www.ensembl.org and "Celera" – http://www.celera.com). Local copies of some of these resources sit alongside custom built in-house tools that together provide a rich source of analysis and data flow. ENSEmbl contains pooled SNP data from a number of SNP resources including the "Human Genome Variation Database" (HGVbase – **http://hgvbase.cgb.ki.se/**) and dbSNP. In addition to this, Celera contains the "Human Genome Mutation Database" (HGMD – http//www.hgmd.org) resource that contains mutations thought to affect phenotype. The combined use of these tools will provide a valuable SNP knowledge base. Use of the distributed annotation system (DAS) within ENSEmbl will give researchers the option to collate and display information in a single view over the ENSEmbl annotation.

Functional studies

We recognise the importance of determining the functional significance of genotypic variations and of using the findings of functional studies to inform genotyping strategies. Therefore, although this application is primarily directed at the creation of a resource of DNA and subsequent genotyping, samples will also be collected simultaneously for determination of functional correlates. Approaches we intend to exploit, and which will form the basis of further applications for funding from this group, include:

Plasma levels of encoded proteins: Where appropriate we will use radioimmunoassay to determine circulating levels of encoded proteins.

Gene expression: We envisage using our cryopreserved cell lines for functional genomic studies using microarray. This approach will allow us to identify differentially expressed genes in patients who do or do not develop sepsis/severe sepsis/septic shock or organ failures and in those who die compared to those who live. Findings may then inform the selection of candidate genes for genotyping.

Proteomics/metabonomics: The recent developments in proteomic (and most recently metabonomic) analysis provide an important opportunity to explore the clinical functionality of genotypic variations. Information from the genotyping may be used to identify novel proteins as potential expression markers for diagnostic and treatment targets and to evaluate protein-protein interactions. Conversely findings from the proteomic or metabonomic analyses may prompt a search for polymorphisms of the relevant gene.

Data Analysis, Power Calculations and Statistical Methods

The outcome analysis will be a nested case control study, with additional "case control" comparisons with disease specific and population-based controls. In this way we can investigate genetic determinants of outcome from and susceptibility to sepsis/severe sepsis/septic shock and organ failure in the predefined categories of infectious diseases (FP and CAP). Secondary end points will be subjected to a qualitative trait analysis. The strength of the association between the presence of specific polymorphisms and primary and secondary end points will be assessed using two-by-two tables and Chi-square analyses and more recently developed neural networks. Logistic regression will be used to adjust for confounding variables and clinical features, including the infectious organism when identified.

Since high-throughput techniques allow multiple comparisons it is essential to ensure that the study is adequately powered. Power is enhanced by avoiding diagnostic ambiguity for patients with CAP and FP. To further reduce heterogeneity in the CAP cases, a subset of microbiologically confirmed *Strep. Pneumoniae* pneumonias will also be analysed (estimated 1000 cases). Our power calculations take into account allele frequency and the level of functional expression of the polymorphisms and will vary depending on the candidate gene selected. Many of the known candidate genes we will target have allele frequencies that exceed 20%. We have calculated, for example, that in the case of the influence of the TNF-308 A allele on outcome from septic shock analysis of 2000 cases would give a more than 80% probability of detecting a relative risk of 1.5 with a p value <0.01.

Our analytical strategy allows for the large number of genetic markers to be tested. Nevertheless, to restrict the number of comparisons to be made, the collection has been scaled to allow interim analysis to identify promising candidate genes. This approach will allow us to generate hypotheses that will then be tested on the remainder of the population, together with relevant functional information, as well as to identify and expand targeted subgroups. Where appropriate we will exploit other DNA resources, in particular those of our collaborators in Europe and North America to achieve increased power and for replication.

Statistical Approach to Haplotype Analysis

Two methods will be used. The first employs a traditional approach, using SPSS for Windows (v10), and the second is based on an artificial neural network, recently developed at Barts and The London. In the first, differences in frequencies of individual SNPs and haplotype combinations will be computed using SPSS for Windows, with Exact or Monte-Carlo methods as necessary. HAPLOTYPER will be used for haplotype inference: this uses a Bayesian algorithm and is available on the Harvard web site. The use of artificial neural networks to detect associations between a disease and multiple genetic markers has recently been publicised and validated. This programme is based on a pattern recognition system and makes no *a priori* assumptions on population history or marker map and does not require pre-specification of derived haplotypes. Detailed marker maps will be required to resolve the causal origin of

disease associations, especially in the MHC region (e.g. at the TNF locus) [21].

Controls and comparator groups

To determine the susceptibility to the conditions under study distinct controls and comparators will be required. Susceptibility to the development of severe sepsis/septic shock (in CAP and FP patients) will be tested by comparing the genotype of those who do, or do not progress to severe sepsis/septic shock (including CAP patients in the US study who are not admitted to ICU). Susceptibilty to organ failures will be tested by comparing those who do, or do not develop organ failures. Outcome will be analysed by comparing the genotype of survivors and non-survivors. Outline permission has been obtained for use of the 1958 Birth Cohort that comprises 10,000 subjects, (2000 of which should be available for release after scientific scrutiny in 2004) to determine population based allele frequencies to compare with our subjects. During the subsequent years DNA will become available from the remaining 8000 subjects; this will provide the opportunity to match the numbers in our phenotype subgroups. In addition, as the MRC/Wellcome UK Biobank evolves we would apply to access this resource in order to derive age and gender matched controls. It is important to note that the UK Biobank has been powered to study the five most common cancers and cardiovascular phenotypes and will not provide sufficient numbers of the incident sepsis, with contemporaneously derived phenotypes, to elucidate the questions addressed by this proposal.

REFERENCES

1. Angus DC et al Crit Care Med 2001; 29: 1303-1310.
2. Sands KE et al JAMA 1997; 278: 234-240.
3. Brun-Buisson C et al JAMA 1995; 274: 968-974.
4. Ledingham IM McArdle C.S. Lancet 1978;1: 1194-1197.
5. Hayes MA et al N Engl J Med 1994; 330: 1717-1722.
6. Ziegler EJ et al N Engl J Med 1991; 324: 429-436.
7. Fisher CJ Jr et al for The Soluble TNF Receptor Sepsis Study Group. N Engl J Med 1996; 334: 1697-1702.
8. Sorensen TIA et al N Engl J Med 1988; 318: 727-732.
9. Stuber F et al Crit Care Med 1996; 24: 381-384.
10. Mira JP et al JAMA 1999; 282: 561-568.
11. Hermans PW et al for The Meningococccal Research Group. Lancet 1999; 354: 556-560.
12. Marshall RP et al Am J Respir Crit Care Med 2002; 166: 646-650.
13. Hibberd ML et al for The Meningococcal Research Group. Lancet 1999; 353: 1049-1053.
14. Roy S et al for the Oxford Pneumococcal Surveillance Group. Lancet 2002; 359: 1569-1573.
15. Gordon AC et al Presented at the 100th American Thoracic Society International Conference, Orlando, USA, May 2004.
16. Fidler KJ et al Intensive Care Med 2004; 30: 1438-1445.
17. Stuber F, Udaleva IA, Book M Journal of Inflammation 1996; 46: 42-50.
18. Gordon AC et al Genes and Immunity 2004; 5: 631-640.
19. Colhoun HM, McKeigue PM, Smith GD Lancet 2003; 361: 865-872.

20. Carden LR, Palmer LJ Lancet 2003; 361: 598-604.
21 Ackerman HC et al Genes and Immunity 2003; 4: 476-486.
22. Angus C et al Am J Respir Crit Care Med 2002; 166: 717-723.
23. Bone RC, Sibbald WJ, Sprung CL Chest 1992; 101: 1481-1483.
24. Bernard GR et al Am J Respir Crit Care Med 1994; 149: 818-824.
25. Vincent JL et al Intensive Care Medicine 1996; 22: 707-710.
26. Le Gall J-R, Lemeshow S, Saulnier FA JAMA 1993; 270: 2957-2963.
27. File TM. Lancet 2003; 362: 1991-2001**.**