

Executive Summary – Final publishable summary report

EUCLIDS aimed to investigate the genetic basis of the major childhood bacterial infections to address the question “are there genetic differences which explain why some children are more susceptible to infection than others, and why some infections are mild while others due to the same pathogen are severe or fatal”. The EUCLIDS consortium recruited over 5,000 children with severe infections in 9 European countries and West Africa and included over 1,500 patients with meningococcal disease and 5,000 controls previously recruited. Detailed clinical investigation and data on severity of illness and outcome was collected, enabling patients to be assigned as having definite bacterial infections, viral infections, mixed bacterial/viral infections and those where accurate assignment was not possible.

EUCLIDS undertook 3 genome wide association studies (GWAS) on patients with meningococcal disease (UK, Western European and Spanish cohorts); a multi-country cohort of children over 2000 children with life threatening bacterial infections due to a range of pathogens; a vaccine response GWAS of 4000 infants undergoing childhood immunisation to identify genes underlying differences in response to vaccines.

Analysis of the UK, Spanish, and Western European meningococcal GWAS identified a gene region controlling susceptibility to meningococcal disease containing the complement Factor H (FH) and five Factor H related protein (FHR1-5) genes. Deep sequencing of the FH/FHR region, and analysis of the effect of each variant on serum protein concentrations established that susceptibility to meningococcal disease was determined by FH plasma concentration controlled but variants in the FH related proteins. Functional studies in differentiated stem cells established that FH related proteins regulates production of Factor H through long range interactions. Analysis of the same gene variants in the multi-country cohort GWAS established that same genetic variants determine susceptibility to meningococcal infection and other bacterial infections. The same four GWAS data sets were analysed to identify genes controlling severity of disease, and novel gene variants were identified. To identify genes determining persistence of antibodies following vaccination EUCLIDS undertook GWAS on infants undergoing vaccination, leading to identification of novel genes controlling persistence of antibody responses to childhood vaccines. To identify rare genetic variants underlying bacterial infections deep sequencing of the coding region of all genes was undertaken in 600 patients with extreme phenotypes or familial disease. Novel genetic mutations predisposing to meningococcal disease and other infections were identified suggesting that that about 10% of all childhood infections may be caused by rare genetic variants.

EUCLIDS has generated a large biobank of patient samples and clinical data, and genomic and transcriptomic data which will be available to the scientific community for further analysis. Identification of the major genes controlling meningococcal disease susceptibility as well as other bacterial infections provides new insight to susceptibility and may be relevant to many other infectious and inflammatory diseases in which regulation of complement activation is involved. The important role of rare genetic variants identified by deep sequencing in controlling meningococcal disease and other bacterial infections provides important new information on pathways underlying susceptibility to infection, and will be of benefit to individual families carrying these rare mutations. EUCLIDS has provided a unique example of how collaborative research on infectious diseases can be achieved by combining patient cohorts and scientific resources across Europe. Information derived from EUCLIDS has already been the basis for award of an ongoing Horizon 2020 research grant(PERFORM) aiming to distinguish bacterial from viral infection, and including partners from the EUCLIDS consortium

1. Concept and objectives

Bacterial infection is the major infectious cause of death in young children, accounting for over a quarter of all child deaths globally. Death from bacterial infection in children has persisted despite availability of antimicrobial agents and current childhood vaccines, highlighting the need for better understanding of the inflammatory response to infection, for novel treatments of acute infection, for new methods to identify those at risk, and for better preventative strategies.

Importance of genetic factors in childhood infection

A remarkable feature of all known infections is the variability of response to infection among individuals within the population. For common childhood bacterial infections such as *S. pneumoniae*, *H. Influenzae*, *S. aureus*, *N. meningitidis* or *M. tuberculosis*, the majority of the population appears to be innately resistant to disease, and these organisms behave as harmless residents of the nose and throat. Invasive disease represents a rare outcome of the otherwise harmless host-bacterium commensal relationship. Among the small proportion of the population who develop invasive disease, there is also great variability in outcome. For example some patients with meningococcal or pneumococcal bacteraemia present with self-limiting febrile illnesses, while others develop meningitis, shock, multi-organ failure and death.

There is now clear evidence that the variability in response to infection between individuals and within populations is genetically determined, with both host and pathogen genome contributing to the outcome of the interaction. Early epidemiological studies documented multi-case families, and concordance for common infections in monozygotic twins [reviewed by (Sirugo et al. 2008)]. Furthermore, the high incidence of disease and death from imported infection in isolated island communities, and in indigenous populations such as Native Americans suggested an important contribution of genetic factors to disease occurrence and outcome. The seminal study of Sorenson documented that risk of death from infection was inherited from biological parents in adoptees living in different environmental background (Sorensen et al. 1988). These studies established an important genetic contribution to infectious disease occurrence and outcome.

Contribution of Mendelian single gene defects, major genes, and polygenic inheritance to infection

The genetic contribution to infectious diseases is likely to be more important in childhood than later in life, as genetically determined differences in host response are likely to manifest on first exposure of the human immune system to the pathogen (Alcais et al. 2009). This is particularly true of severe immune defects inherited on a classical Mendelian basis (Casanova and Abel 2007). Five main forms of genetic effect are likely to contribute jointly to the overall genetic predisposition to infection in children (Figure 1).

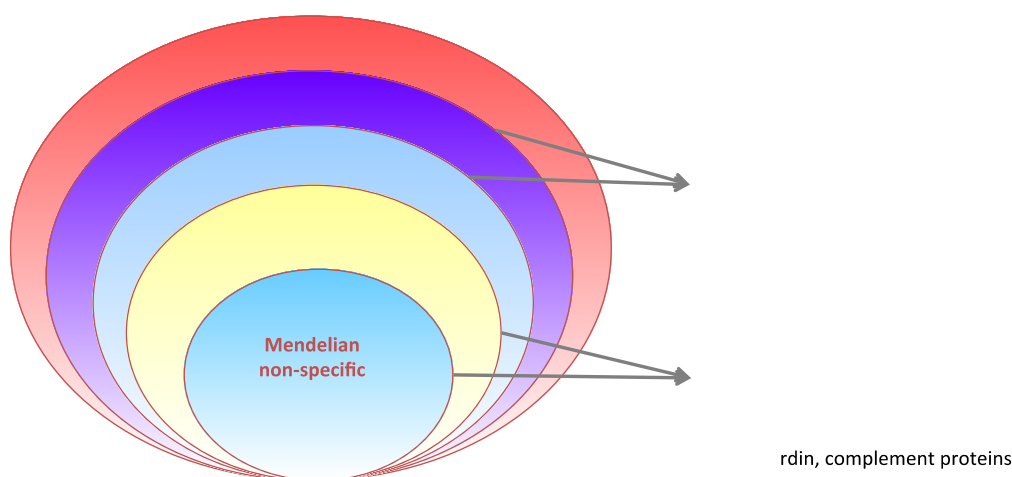


Figure 1: Forms of genetic effects

Firstly, rare single gene Mendelian mutations predispose to infection acting non-specifically to increase susceptibility to multiple infections. These defects include immunoglobulin deficiency, complement deficiencies, various forms of T and B cell deficiency. (Casanova and Abel 2007) and generally present early in life with severe or recurrent infections with multiple different pathogens. Secondly there are a growing number of **Mendelian defects predisposing to specific infections such as** interferon gamma/IL12 pathway defects leading to susceptibility to mycobacterial and

salmonella infections (Newport 1996, Casanova and Abel 2002). The IRAK4 and MYD88 defects leading to pneumococcal infection (Picard et al.2006) complement and properdin defects leading to meningococcal infection (Wright et al. 2009); UNC 93B and TLR3 defects leading to Herpes simplex encephalitis (Abel et al.2010).

It is likely that the number of single gene defects specific for individual pathogens will increase, as only a small proportion of Mendelian susceptibility is explained by current known defects. The third mechanism is that of major genes. Major genes loci have now been implicated in a number of complex traits including infectious diseases such as leprosy (Mira *et al.* 2004), and schistosomiasis (Marquet *et al.* 1996). Polygenic predisposition plays a role in most complex traits including infectious diseases, where multiple different genes contribute to disease occurrence or disease severity (Burgner *et al.* 2006; Brouwer *et al.* 2009; Wright, Hibberd et al. 2009). Finally, epigenetic processes are increasingly recognised to be important determinants of gene regulation (Feinberg and Irizarry 2009). EUCLIDS aimed to comprehensively identify all five types of genetic effects contributing to overall genetic predisposition to childhood bacterial infection, by using a range of state-of-the-art and “beyond state-of-the-art” technologies and approaches.

Susceptibility and severity genes: The genes that control disease susceptibility is likely to be different from those determining disease severity for each infectious disease. Using the example of meningococcal or pneumococcal infection (Figure 2), different genes may control adherence of the bacteria to the nasopharynx, invasion into the blood stream, or survival within the blood. These genes act as **susceptibility genes**, and may have no effect on the severity of the disease. Other genes may determine the intensity and duration of the inflammatory response, and thus the clinical manifestations and outcome of the infection (**severity genes**). Comparison between cases with an infection and uninfected controls is required to identify genes controlling susceptibility, whereas comparison of cases with different phenotypes (such as mild and severe) is required.

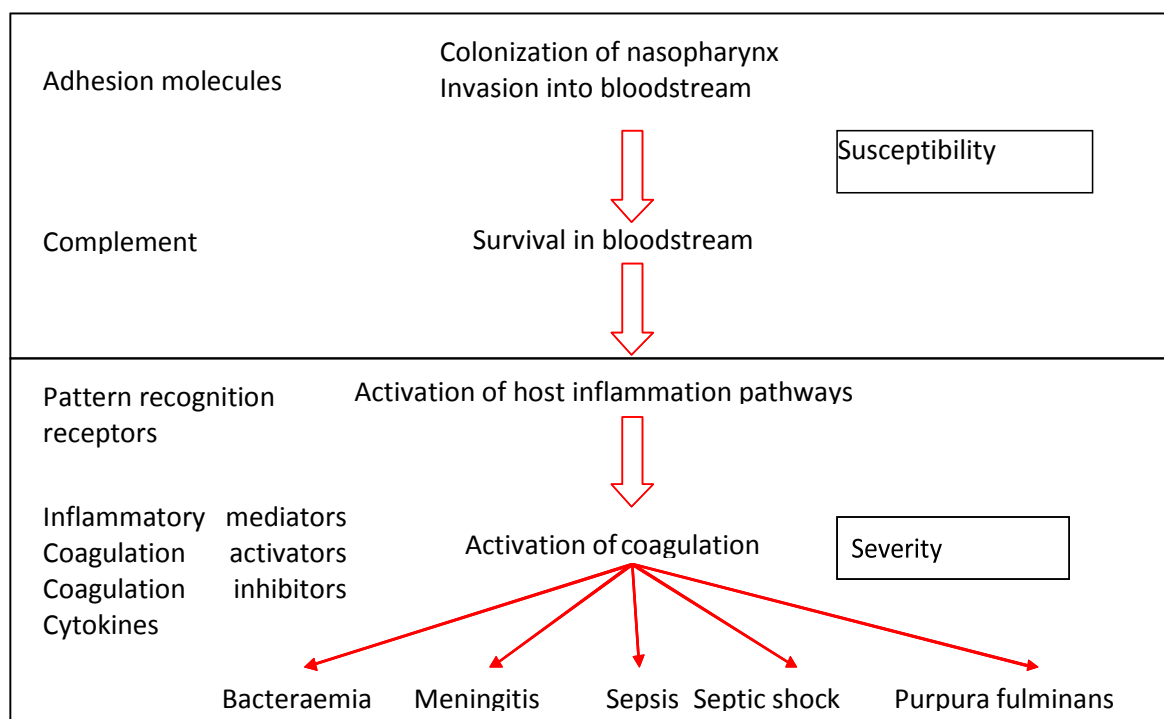


Figure 2: Different mechanisms controlling Susceptibility and Severity

Meningococcal disease (MD) as a Model to study genetics of infection. MD is not only one of the most important life threatening infections in children, but is a unique model through which to study the genetic basis of infectious diseases as it is easily recognised by the characteristic rash and shock syndrome, notifiable, and common in many countries. **Severe cases also serve as a paradigm of Gram negative septic shock, a leading cause of mortality in adults as well as children.** MD shows remarkable variability in its severity and clinical manifestations, which range from asymptomatic nasopharyngeal colonisation (found in 5-20% of children and young adults, to minimally symptomatic bacteraemia, fulminant and rapidly progressive shock and multi-organ failure, meningitis, or the extreme phenotype of purpura fulminans (Nadel and Kroll 2007). The clinical features and remarkable spectrum of severity make meningococcal disease an ideal disease-model through which to study genetic factors controlling susceptibility and severity of infectious diseases.

Genetic control of response to immunisation

Vaccination has been the most successful strategy for control of infectious diseases including childhood bacterial infection. For most vaccines now in use, a varying proportion of the population are “**vaccine failures**” either because of poor initial response to the vaccine, or rapid waning of the response with time. Vaccine failures have been identified for most vaccines in current use, including highly successful vaccines such as *Haemophilus influenzae* type b (Hib), *Neisseria meningitidis* serotype C (Men C) and pneumococcal conjugate vaccines (Ladhani et al. 2010). Furthermore, rapid decline of protective antibodies occurs after immunisation in some individuals receiving Men C and pneumococcal vaccines, making older children susceptible years after initial vaccination (Campbell 2010). The immune response to vaccination is known to be under genetic influence, and there are data from family and twin studies suggesting that heritability of vaccine response ranges from 35-90% depending on the vaccine (Poland 2007; Sirugo, Hennig *et al.* 2008). Differences in vaccine efficacy also occur between different ethnic groups further suggesting a role for genetic factors (Kimman et al. 2007) it is likely immune response induced by Hib, Pneumococcal conjugate and Men C vaccines may be under similar genetic control as all are polysaccharide-protein conjugate vaccines.

Overall approach in EUCLIDS to identify the genetic basis of life threatening Childhood infection

EUCLIDS followed a staged approach using meningococcal disease as the model, through which to identify the genes controlling susceptibility and severity of bacterial infection and response to vaccines and to understand the biological processes involved (Figure 3). The approach was then extended to the other major childhood bacterial infections.

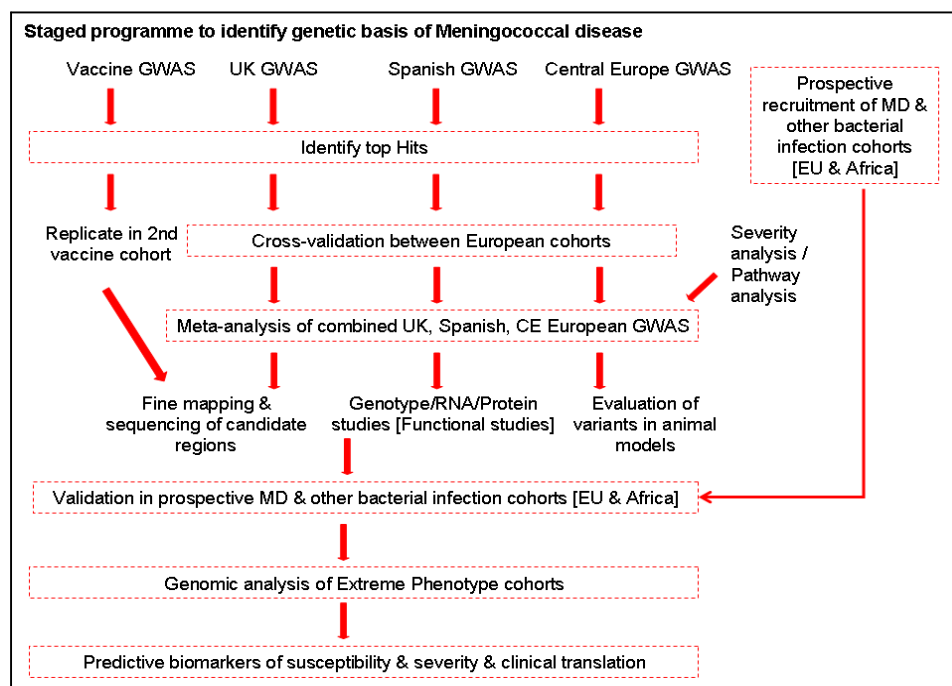


Figure 3: Overall approach in EUCLIDS

Stage 1: Genome-wide case control study of three European cohorts EUCLIDS undertook genome wide SNP (single nucleotide polymorphism) genotyping for over 600,000 genetic variants in 3 European cohorts (UK, Western Europe and Southern Europe) including over 1500 cases of meningococcal disease and 5000 healthy controls to identify genes controlling susceptibility to infection.

Stage2: Identification of genes controlling severity of MD. To identify genes determining severity and outcome of infection we used detailed phenotype and severity markers to compare the frequency of each SNP in meningococcal cases differing in severity and outcome, using novel gene and pathway based approaches to identify the most significant genes.

Stage3: Validation of Hits and fine mapping, with high-throughput sequencing to identify rare and causal variants. To identify the causal variants in the genes controlling susceptibility, we used deep sequencing of the candidate gene regions, and functional studies to determine the effect of the identified variants on protein production and cellular function.

Stage 4: Identification of functional mechanisms underlying genetic association. To identify the mechanisms by which gene variants controlled disease susceptibility and severity we used RNA expression analysis, plasma protein measurements, in vitro bactericidal assays and animal models, relating these to identified gene variants

Stage 5: Identification of interaction (“fit”) of bacterial genomic variation and human variation. We postulated that the genetic variation in both host and pathogen would interact to determine the outcome of infection. We used the most significant findings from the GWAS analysis- the Complement factor H region- which interacts with a meningococcal Factor H binding protein, to study the interaction of host and pathogen genome.

Stage 6: Identification of functional consequences of genetic variants in murine models. To define the role of gene variants identified (for example in the fH/fHrp genes) we assessed murine models to examine meningococcal pathogenesis and immune responses. After initial studies, we replaced animal models with in vitro human systems.

Stage 7: Study of the influence of genetic variation on meningococcal vaccine responses. Waning of antibody responses to vaccines is a major problem in protecting children by vaccination. Identification of genes underlying vaccine responses may help develop longer acting vaccines. We therefore undertook a GWAS of antibody persistence following vaccination and RNA expression studies of response to the meningococcal group B vaccine to understand genetic differences in response to vaccines.

Stage 8: Identification of rare Mendelian variants underlying susceptibility and severity of infection. EUCLIDS used Next Generation Sequencing to study an Extreme Phenotype cohorts with meningococcal disease and other bacterial infection, and then functional studies to understand the role of identified variants.

Stage 9: Application of model of MD to other life threatening childhood infections: pneumococcal, staphylococcal cohorts (EU and African cohorts) and salmonella (African cohort). We aimed to extend the approach developed for meningococcal infection to the other main causes of life threatening bacterial infection. We prospectively recruited cohorts with other infections during years 1-3 of the project, and commence analysis in year 4 once the well-defined patient cohorts had been assembled. Many of the genes controlling susceptibility or severity of MD are also likely to be involved in other bacterial infections. For example, several other pathogens use fH or fHR to evade complement-mediated damage,

Similarly, genes controlling the intensity of the inflammatory response, which also influence disease severity may control outcome of a range of bacterial infections. We explored the gene variants identified in the MD GWAS for their role in susceptibility and severity using the cohorts of patients with a range of other bacterial infections.

Dissemination and Training and ongoing work: Through the collaborative program of work the EUCLIDS consortium has brought together an international team of clinicians and scientists, and provided training to many young doctors and scientist across Europe and West Africa. The biobank of samples collected, and extensive genomic data will continue to be used for collaborative research for many years to come.

2. Scientific & technological methodology and associated work plan

Organisation of EUCLIDS work: The overall organisation of work in EUCLIDS was carried out in work packages (WP), each lead by a different partner. Imperial Consultants (ICON) worked as **project coordinator** to ensure communication between components of the project. **The Bioinformatics and Data Handling work package (WP8), and Management (WP9), together with Clinical work package (1 and 2) operated throughout the project to ensure consistency of data, clinical information and samples, and to provide samples and information for the other WPs.**

Initial GWAS was undertaken on existing cohorts of meningococcal disease (WP3). While on-going **clinical recruitment** in year 1-3 established a prospective cohort of MD patients and cohorts with a range of bacterial infection across Europe and Africa. **The clinical groups in London, Liverpool, Amsterdam, Rotterdam, Nijmegen, Austria and Spain as well as the West African cohort coordinated through the MRC Gambia** followed identical criteria for patient recruitment, phenotype assignment, severity assessment and used a common database and sample handling procedures (WP1 and WP8). **High-throughput GWAS of the meningococcal cohorts** was undertaken in year 1 (WP3), with analysis of the individual data from the three European cohorts and the combined meta-analysis by single SNP, pathway analysis and severity, undertaken jointly by GIS, Spanish node with bioinformatics statistical genomic group at Imperial College (WP8). Validation and fine mapping of target hits from each of these analyses was undertaken by **GIS, AMC, and Santiago de Compostela, (WP 3)**. High-throughput sequencing of the fH region as well as of other target regions identified from the GWAS was undertaken in years 2 and 3 at GIS (WP4). The **meningococcal vaccine cohort (WP2) recruited by the Oxford group**, underwent genotyping in GIS in year 1 together with completion of functional analysis of vaccine responses (WP2 and 3). **Extreme phenotype analysis of meningococcal disease, other infections and vaccine recipients** was undertaken by our SME partner Oxford Gene Technology and Oxford University in year 4 through application of Next Generation Sequencing, RNAseq, on highly characterised Extreme Phenotype

cohorts with meningococcal disease, each other infection, and extremes of vaccine response (100 high and low responders). Bioinformatic analysis of this data was undertaken in collaboration with the Imperial College group in years 4 and 5 of the programme (WP7 and 8). **Functional studies of the fH and fHrp** required development of monoclonal antibodies to FH and fHrp1-5 by AMC Amsterdam (WP5). Exploration of the effect of genetic variants on fH and fH-related proteins and of bacterial variation using **animal models** and in vitro models was undertaken by Nijmegen and AMC groups as well as Oxford (WP6). Complete sequencing of the meningococcal Factor H binding protein (fHbp) on isolates from patients in the cohort was undertaken by SME partner Micropathology and GSK (WP5) **Analysis of the Extreme Phenotype cohorts of MD, pneumococcal, staphylococcal other infection** using Next Generation Sequencing, RNA expression and epigenetic was undertaken in year 5 (WP7 and 8). **Biomarkers of disease susceptibility and severity, and of vaccine failure or response**, were identified by bioinformatic analysis of the combined data (year 4 and 5) (WP7, 8, 3). Publication dissemination and translation of the findings have taken place throughout the study and will continue in the coming year (WP10).

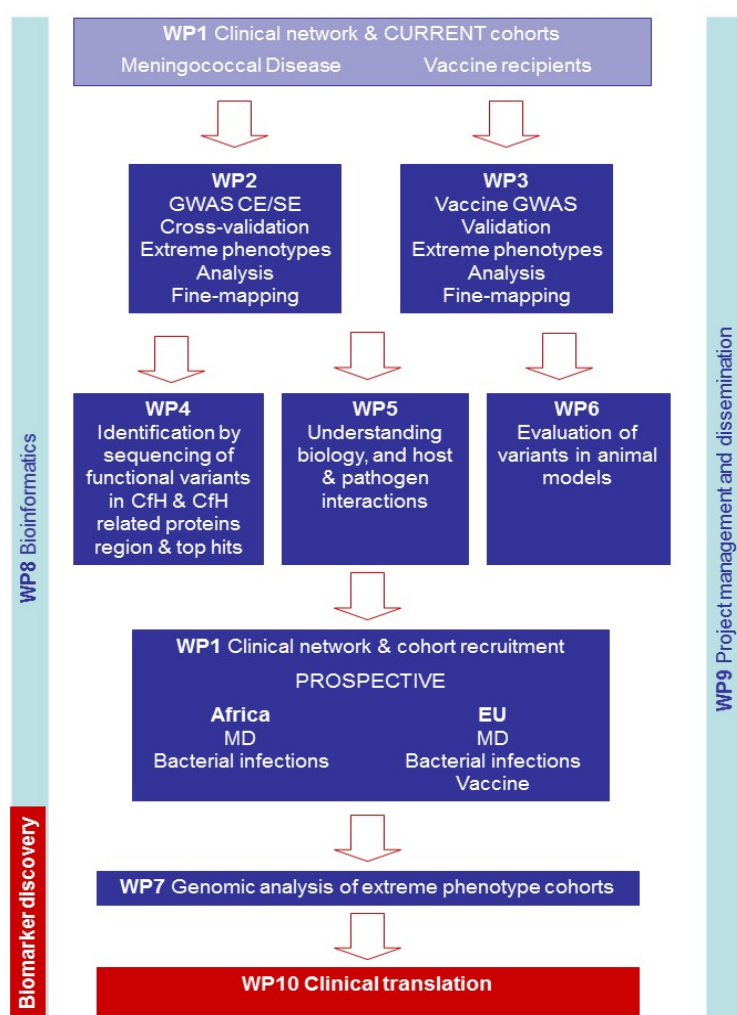


Figure 4 Overall organisation of the EUCLIDS consortium into 10 work packages

3. Principal achievements and Results

3.1. Recruitment of the EUCLIDS prospective cohort.

EUCLIDS recruited children aged 1 month-to-18 years old with sepsis or severe febrile illness admitted to hospitals in Europe and West Africa (Gambia) Europe from July-2012 to December-2016. A total of 84800 eligible patients with complete data were recruited (Figure 5, Table 1).

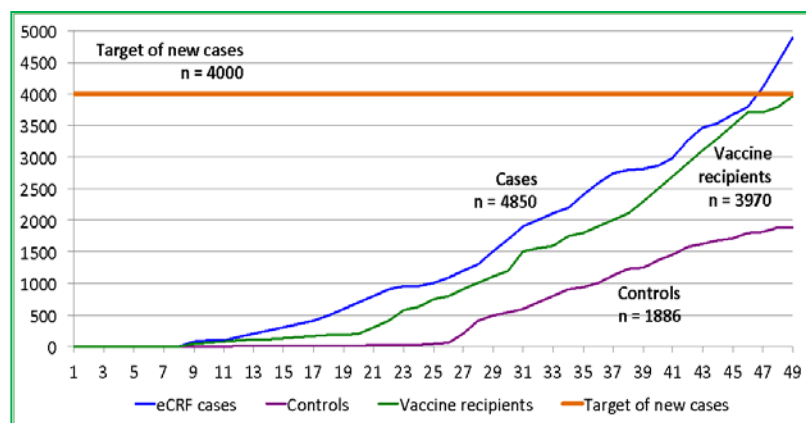


Figure 5: recruitment into EUCLIDS over course of the study

Recruitment was conducted in 10 Countries each with a network of hospitals. A total of 139 hospitals contributed to the EUCLIDS recruitment network EUCLIDS (www.euclids-project.eu). Data was collected on age, sex, comorbidities, prior antibiotic use, duration of hospital and PICU stay, respiratory and inotropic support required, level of disability at discharge and death. These data were used to assign patients to diagnostic groups according to whether an aetiological pathogen was confidently identified. In addition to investigations undertaken at each hospital, EUCLIDS applied **molecular genotyping undertaken by partner Micropathology**, to identify both bacterial (blood) and nasopharyngeal pathogens. The study concluded that molecular pathogen detection methods can enhance diagnosis (Figure 6).

NODE	Recruitment		
	CASES	CONTROLS	TOTAL
AUSTRIAN NODE			705
AUSTRIA	520	—	520
GERMANY	164	—	164
SERBIA & MONTENEGRO	16	—	16
LITHUANIA	5	—	5
GAMBIAN NODE	483	1382	1865
NETHERLANDS NODE	530	—	530
SPANISH NODE	563	357	920
SWISS NODE	665	—	665
UNITED KINGDOM NODE	1904	147	2051
Summary	4850	1886	6736

Table 1: Global Recruitment

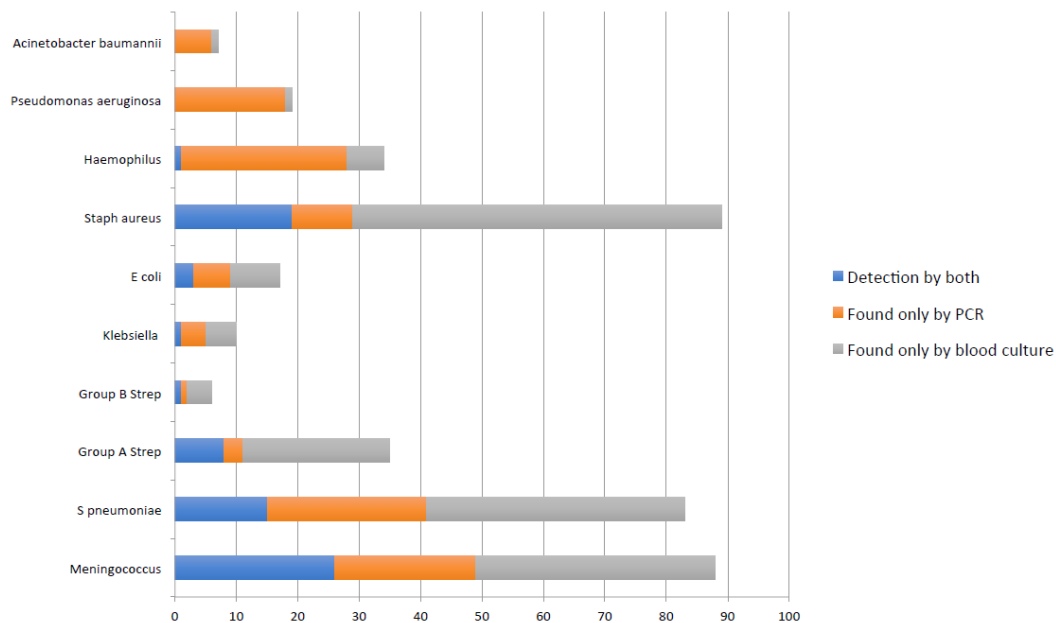


Figure 6: Comparison of molecular diagnosis using PCR and blood culture

Using all the available clinical and laboratory data, Patients were assigned based on a diagnostic algorithm developed by the EUCLIDS partners (Figure 7).

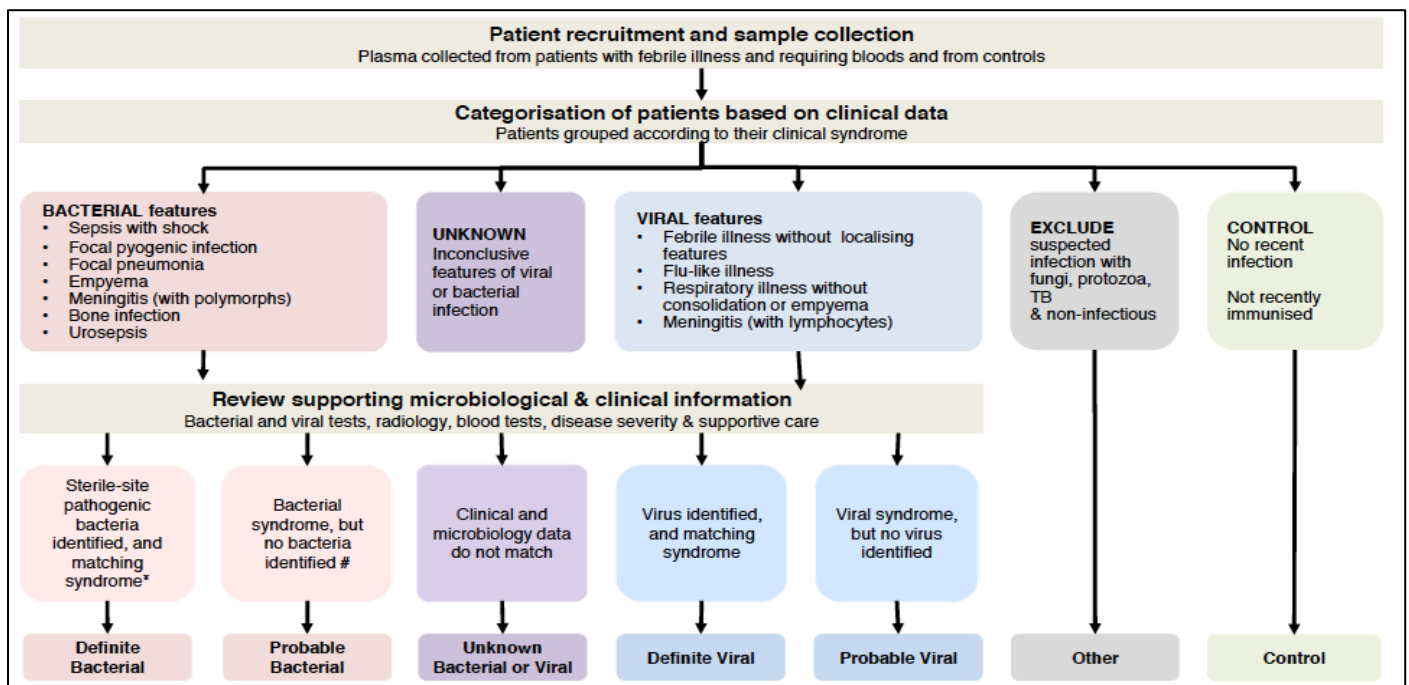


Figure 7: Diagnostic Algorithm

As for the demographic features of the EUCLIDS cohort, the Median age was 39.05 months (IQR = 12.4-93.9) and 53.2% of the patients were male. A total of 43.2% (n = 1229) had sepsis and 56.8% (n = 1615) had severe focal infection (SFI) without matching sepsis criteria. The mortality rate was 2.2% (n = 57). The main focal syndromes diagnosed were pneumonia (n = 511, 18%), central nervous system infection (n = 469, 156.5%) and skin and soft tissue infection (n = 247, 8.7%) (Figure 8). Sepsis was diagnosed predominantly in younger children and SFI in older children. Patients with a causative organism identified had overall more severe disease.

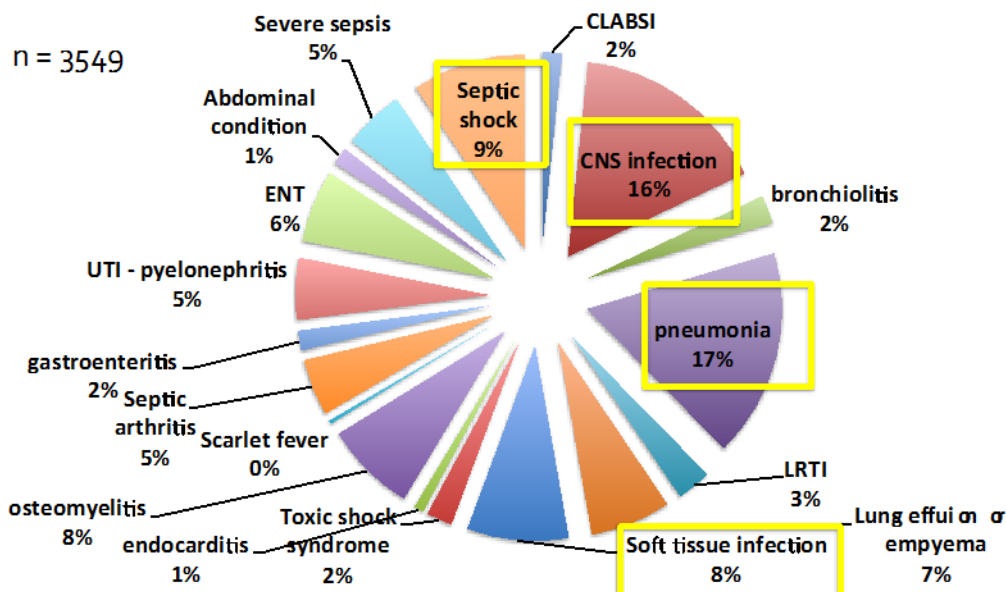


Figure 8: Clinical syndromes in EUCLIDS

Microbiological Causes of Infection. A causal microorganism was identified in 55.7% (n=1978) of the cases (Figure 9). The most prevalent bacterial causative agent was *Neisseria meningitidis* in 9.1% (n = 259) of the SFI and 33.8% (n=53) of the severe sepsis patients. Other commonly reported microorganisms included *Staphylococcus aureus* (7.8%; n = 222), *Streptococcus pneumoniae* (7.7%; n = 219) and *Streptococcus pyogenes* (5.7%; n = 162). Viruses were detected in 185 (6.5%) of the patients as the causative agent. A total of 37.6% (n = 1070) patients required PICU admission with a median duration of stay of 4 days (IQR = 2-9). During hospitalization, 36.3% (n = 923) of the children required supplemental oxygen, 25% (n = 649) invasive ventilation and 11.8% (n = 304) inotropic support.

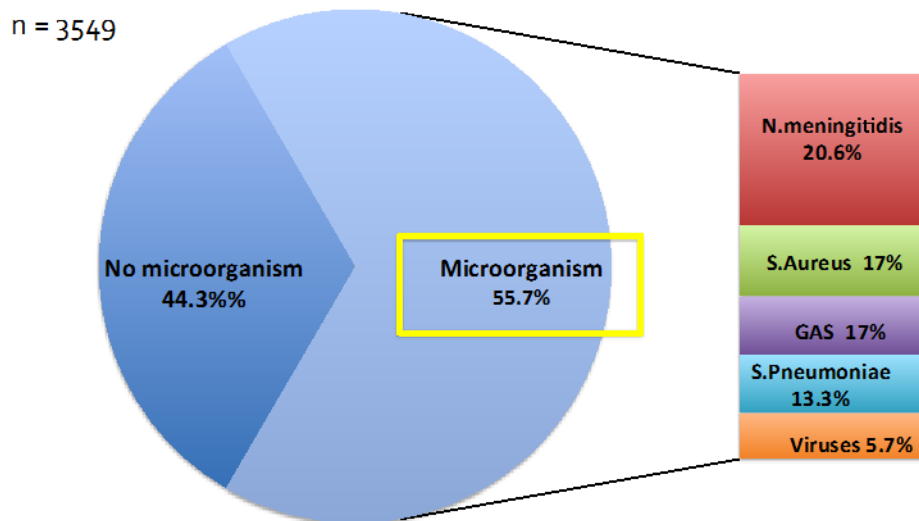


Figure 9: Major Pathogens identified in EUCLIDS

An important finding was the high frequency of patients in whom no pathogen could be identified despite intensive diagnostic approaches, although the proportion of patients with identifiable pathogens varied between syndromes (Figure 10).

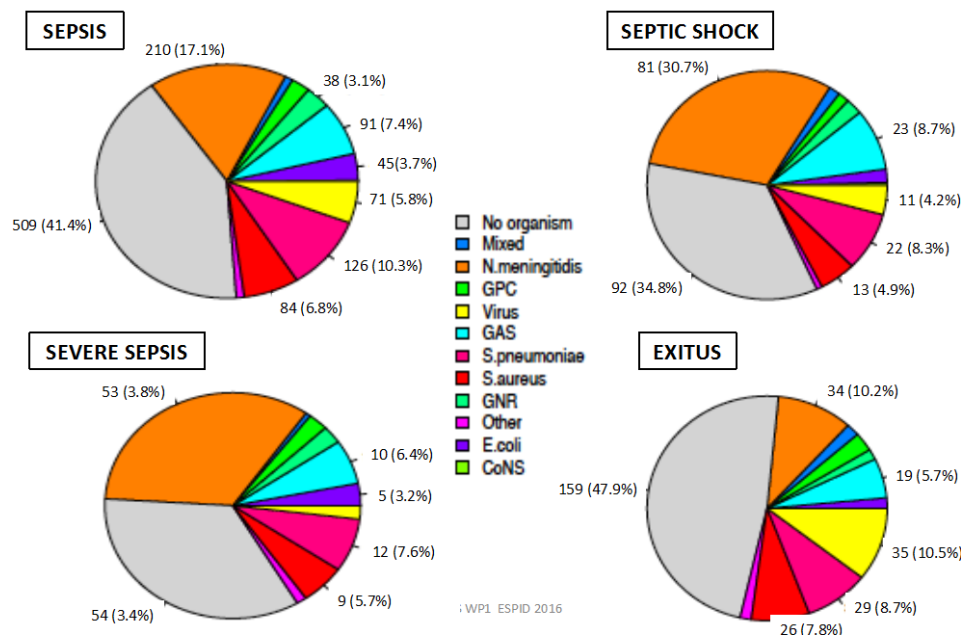


Figure 10: Major pathogens identified in each clinical syndrome

In **patients requiring intensive care**, the most common identified pathogen was meningococcus. 31% of patients had septic shock (Figure 11).

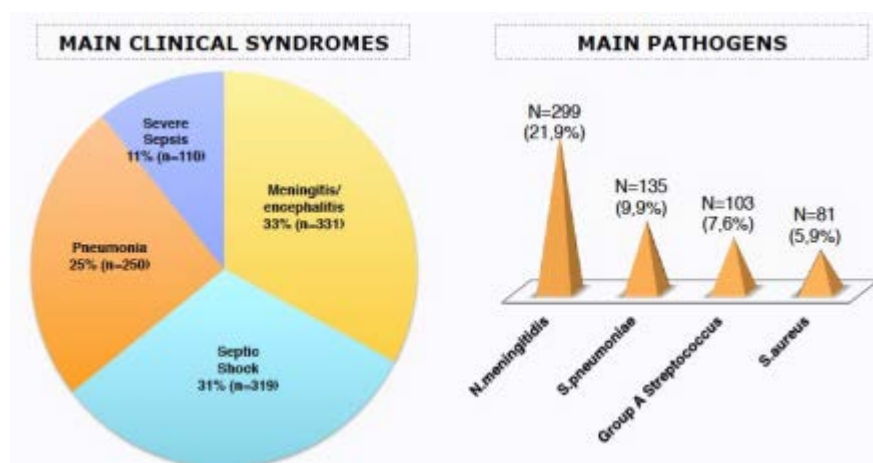


Figure 11: Patients requiring intensive care. The major pathogens remain vaccine preventable infections.

3.2. Indicators of genetic predisposition

An important finding from the genetic point of view was the high prevalence of a family history of infection and previous admission for infection among recruited patients (Figure 12).

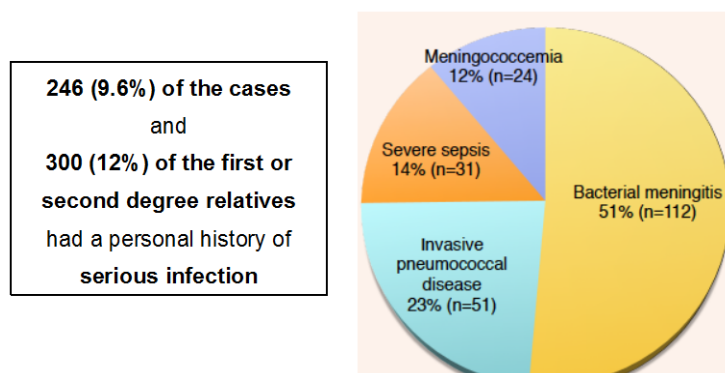


Figure 12: Risk factors for Infection in EUCLIDS. A surprising finding was the high proportion of cases with severe infection that had previous infection or family history of infection pointing to a role for host genetics

3.3. Pathogens identified in EUCLIDS by clinical syndrome severity and outcome

The EUCLIDS cohort included a high proportion of critically ill children requiring Intensive care (Table 2).

2/1641)

Table 2: Severity of the illness

Although overall mortality was low, a significant proportion of children had long term consequences including neurological disorder, deafness, amputations and skin grafting (Table 3).

4/1449)

3.3% (32/979)

Table 3: Outcome

3.4. EUCLIDS Biobank.

EUCLIDS has created a unique biobank of carefully curated and phenotyped patients, with clinical data stored on an anonymised data base, and samples of DNA, RNA, Plasma, Serum and throat swabs available for the planned genetic studies and future analyses (Figure 13).

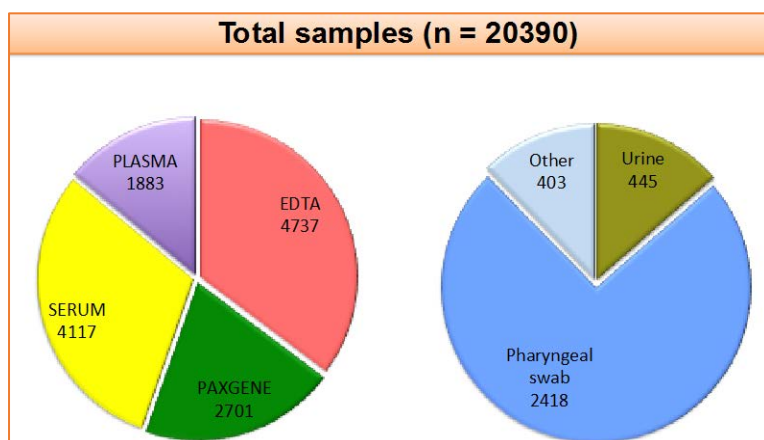


Figure 13: Number and nature of samples in EUCLIDS biobank

3.5. Database and Sample bank

The EUCLIDS biobank not only provided all the samples for the genetic studies described in the next sections, but is available for future research. EUCLIDS has established a procedure through which the scientific community can apply for and access the biobank and the data generated (Figure 14).

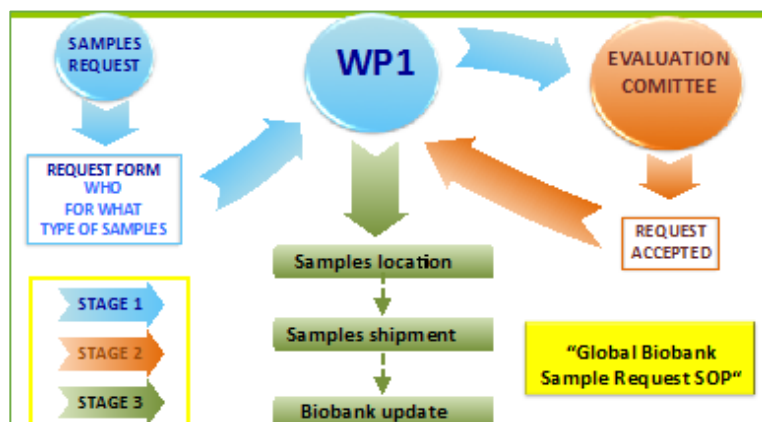


Figure 14: Access to the biobank

4. Main Findings and Results

4.1. Genome wide association studies - Identification of the causative genetic variants underlying meningococcal and other bacterial infections

Our previous genome-wide association study (GWAS) identified an association between MD and a broad genomic region spanning Complement Factor H (*CFH*) and the Complement Factor H-Related protein (*CFHR*) genes (Davila Nature Genetics 2010). Identification of the causal gene, and characterization of the functional variant(s) has been difficult due to the complexity of the region as *CFH* shows sequence similarity to five adjacent *CFHR* genes on chromosome 1.

Factor H (FH) is a serum glycoprotein, synthesized mostly in the liver, which acts as a negative regulator of alternative complement pathway activation. FH is a crucial factor in preventing host cell damage by uncontrolled complement activation and genetic variation in *CFH* or *CFHR* genes are associated with several diseases including systemic lupus erythematosus, glomerulonephritis, IgA nephropathy, atypical hemolytic uremic syndrome and age-related macular degeneration. *N. meningitidis* is highly sensitive to complement-mediated killing. Binding of *N. meningitidis* FH-binding protein (fHbp) to FH facilitates evasion of complement-mediated killing, as FH binds to meningococcal surface and inhibits complement mediated killing (Figure 15).

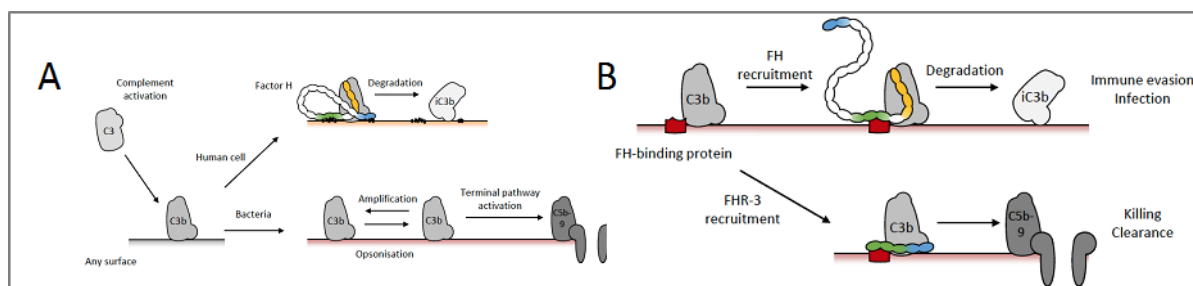


Figure 15: The complement system attacks pathogens but Complement Factor H (CFH) inhibits the activated complement from causing damage to host cells (A). Factor H binding protein (fHbp) expressed by meningococcus binds to CFH, to prevent meningococcus being killed by complement (B)

Host variation in the *CFH-CFHR* region may influence the success of this “Trojan horse” evasion by *N. meningitidis*. Furthermore, inhibition of complement by “hijacking” FH has been adopted as an immune evasion strategy by several bacteria as well as plasmodia. To identify the causative variant in the FH region determining susceptibility to meningococcal disease we first meta-analysed the UK, Austrian, and Spanish GWAS, and confirmed that the FH Region was the principle genomic region with strong statistical association with Meningococcal disease. We then fine-mapped the *CFH-CFHR* region and identified the causal variants (Figure 16).

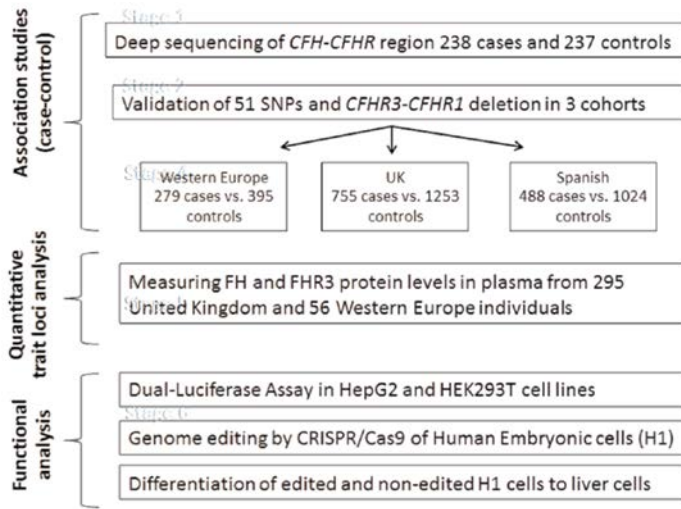


Figure 16: The meningococcal GWAS indicated the importance of the *CFH*-*CFHR* region. Fine sequencing of the region permitted validation of GWAS findings. Variants were compared to with complement protein concentrations, and genome editing was used to investigate the function of the variants

The FH region is complex with a high degree of sequence homology between FH and the 5 FHR genes (Figure 17).

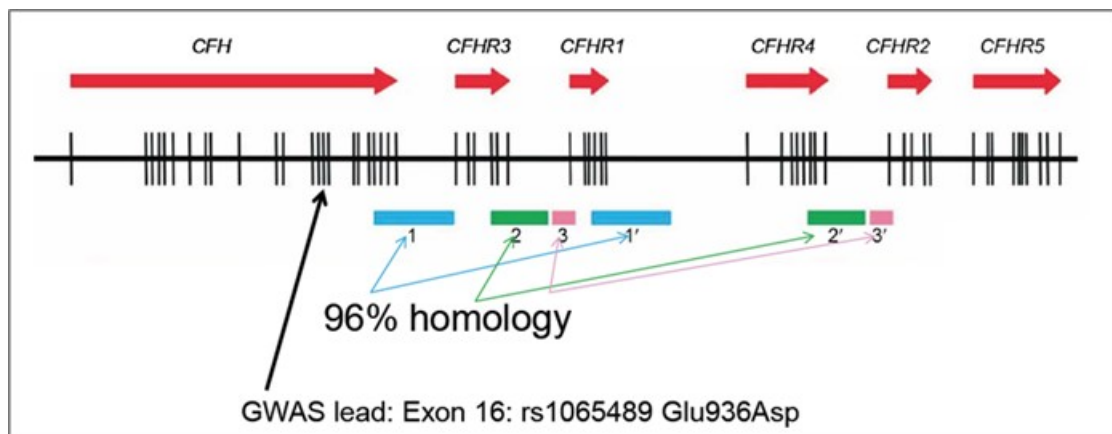


Figure 17: The organisation of the FH region on chromosome 1.

Deep-sequencing of the *CFH*-*CFHR* region was undertaken in 238 MD patients and 237 controls from the Western European cohort. Replication was undertaken in 755 cases and 1,253 controls from the UK, 279 cases and 395 controls from Western Europe, and 488 cases and 1,024 controls from Spain. Convalescent serum was available from 367 patients and 124 healthy unrelated controls for measurement of FH and FHR-3 levels, and 295 UK patients and 56, healthy unrelated controls from Western Europe were used in protein quantitative trait loci (pQTL) analysis.

4.2. Sequencing and genotyping of the *CFH*-*CFHR* region

To identify functional variants driving the association with MD susceptibility we devised a capture-targeted sequencing strategy (Nimblegen™ Roche design) with tiling arrays covering over 85% of the *CFH*-*CFHR* region on chromosome 1 spanning ~360 kb (chr1:196722205-196808506) followed by sequencing with Illumina HiSeq 2000 using 100bp paired-end reads (Figure 18).

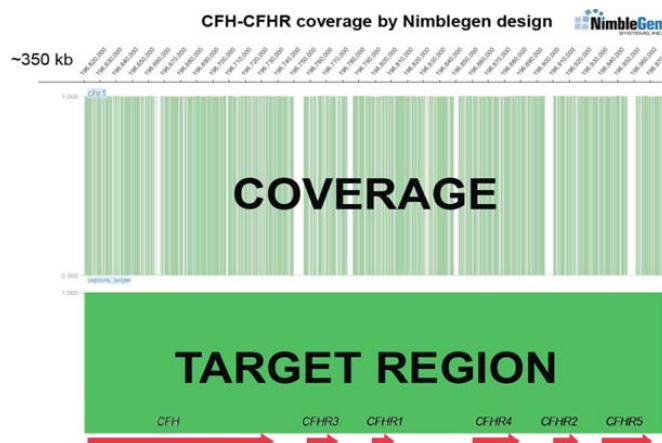


Figure 18: The coverage of the *CFH*-*CFHR* region using a capture-targeted sequencing strategy.

Deep sequencing of the *CFH-CFHR* region identified 4,369 SNPs after applying stringent quality control filters. The strongest signal was identified on one of the *CFHR*'s in a region with high linkage disequilibrium (LD, $D'=0.92$) with the previously reported lead variant, rs1065489 in *CFH*. The 51 SNPs with the strongest association with MD were selected for validation and 45 SNPs were successfully typed in the UK, Spanish and Western European cohorts (13 SNPs, in a tight LD block within *CFHR3*, achieved genome-wide significance in the meta-analysis), confirming the genetic association ($P=1.1 \times 10^{-16}$)

4.3. Serum concentrations of FH, but not FHR-3, are higher in MD than in controls

To explore the relationship between serum concentrations of *FH*, *FHR* and genotype, we needed to measure plasma concentrations of *FH* and *CFHR1-5*. As there were no specific reagents for this task Partner AMC developed monoclonal antibodies to each of the proteins which were then used in studies of the gene/protein interaction (Figure 19).

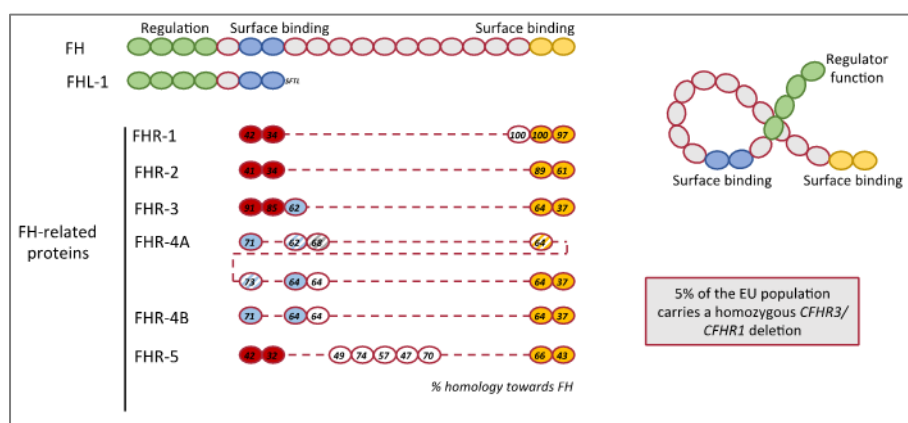


Figure 19: Structural homology between CFH and the 5 CFHR proteins. Specific antibodies were raised to each protein

The EUCLIDS West African cohort enabled us to also examine the role of *FH* in meningococcal disease in Africa, where epidemics of the disease are common. Serum concentrations of *FH* and *FHR* were measured in survivors of MD at least 6-24 months after the acute illness. Serum concentrations of *FH* were significantly increased in MD survivors as compared with healthy controls.

4.4. Serum concentrations of FH are associated with both SNP and CNV in *CFHR3*

When genotype in the *FHR* region was related to protein concentration we found that the plasma concentration was influenced by both SNP and copy number variation, resulting in 6 genotypes controlling FH concentration.

4.5. *CFHRs* controls CFH expression through epigenetic long range interaction

Having established the correlation between “protective genotypes” and lower serum levels of FH, and “risk genotypes” with higher levels of FH, we investigated the epigenetic histone marks in various cell lines to provide information on the putative regulatory role of the potential functional SNP *CFHRs*. Histone marks (H3K4me3 and H3K9ac) from the Roadmap epigenomics database indicated that all investigated hepatic cell lines have an active regulatory site within *CFHR3*. To confirm that the deletion and SNP are functionally active in regulating *CFH* production we used genome editing using CRISPR. As no cell lines were available that expressed FH (which is produced in the liver), we used stem cells differentiated into hepatic cells to show that deletion of the candidate region resulted in increased CFH expression. The results of these experiment establish that plasma concentrations of CFH are controlled by both copy number variation in the CFH related genes, and by the identified SNP through long range genetic interactions.

Conclusions

This extensive body of work has established that risk of MD is determined by FH blood levels, which are regulated by a distal variant in the CFH related genes gene rather than in the CFH gene. As complement inhibition by FH is used for immune evasion by several bacteria, viruses and parasites, and is involved in inflammatory diseases, genetic regulation of FH through *CFHR3* may be relevant to many diseases

4.6. Gene variants controlling susceptibility to meningococcal disease also affect susceptibility to other bacterial pathogens

Many bacterial pathogens other than N. meningitis have CFH binding proteins and may use CFH to evade killing by complement (Figure 20).

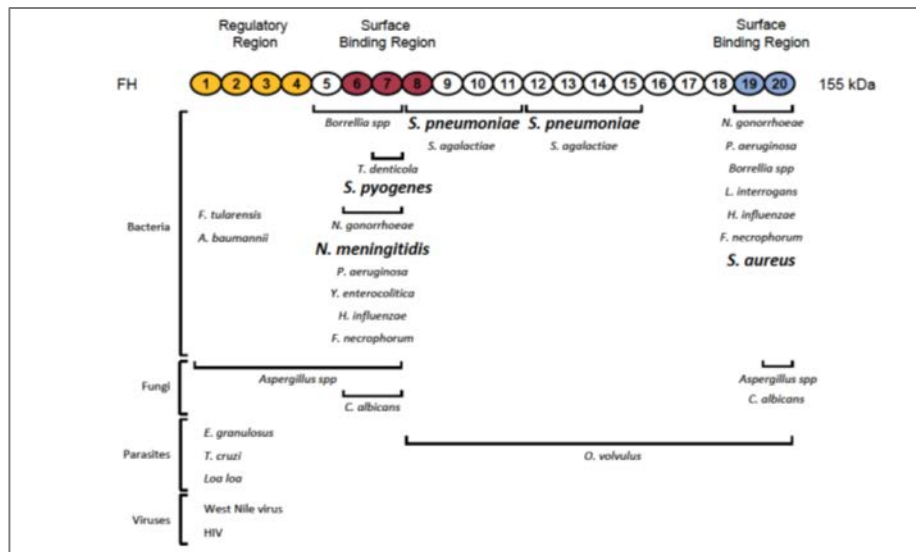


Figure 20: FH in other pathogens

In order to investigate whether the CFH region we had found to be important determinants of meningococcal disease also played a role in other infections, we genotyped the prospectively recruited EUCLIDS cohort, comprising 1,500 patients genotyped using the Illumina Core exome array. 509,164 SNPs passed QC. After quality control and exclusions for non-European ancestry we have genotype and phenotype data for 1012 individuals. Genotype was also obtained by Sequenom genotyping of 55 SNPs selected by association with susceptibility or severity in the original '3 GWAS' including the *CFH*-*CFHR* SNPs showing association with meningococcal disease. We found significant association of the same SNPs with definite and probable bacterial infections. Although the numbers of individual in the individual pathogen groups was too low for high levels of statistical significance, in all cases the Odds ratios and direction of effect was in the same direction. These results will be confirmed on the larger numbers of samples currently being genotyped but suggest a major effect of the *CFH* region on multiple bacterial pathogens (Figure 21).

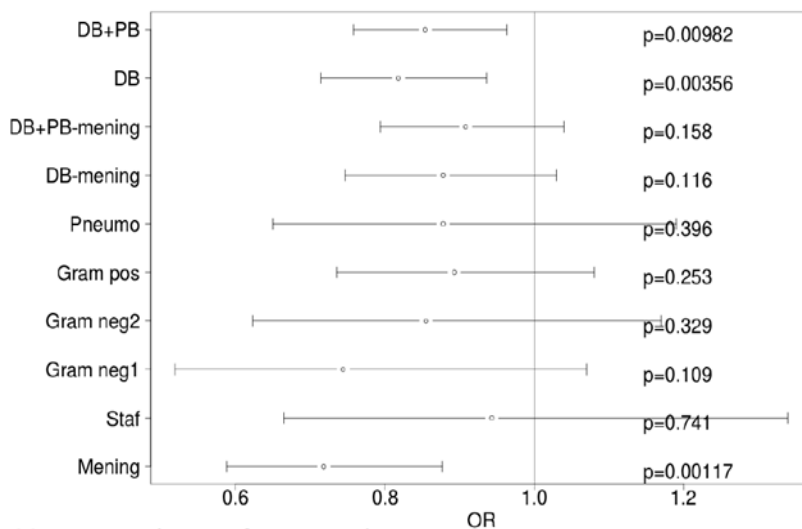


Figure 21: Association of CFH region with multiple bacterial illnesses. The odds ratios for the lead SNP for susceptibility to definite and probable bacterial infection and individual pathogens in EUCLIDS cohort; including aggregated definite bacterial (DB) and probable bacterial (PB) infections, pneumococcal infection (pneumo), gram positive and negative infections, staphylococcus. Meningococcal patients (mening) showed the strongest effect.

Conclusion The role of CFH in susceptibility to meningococcal & other bacterial infections

EUCLIDS has identified and validated the major gene locus determining susceptibility to meningococcal disease as the *CFH* region. We have established that plasma concentrations of *CFH* are the main determinant of susceptibility, and this is controlled through a long range genetic effect by variants in *CFHR*'s. The *CFH*-*CFHR* association also plays a role in susceptibility to other bacterial infections.

5. Meningococcal variants and FHR-1-5 binding: the microbial side

Susceptibility to disease might arise from genetic variation in both host and pathogen. The meningococcal Factor H binding protein FHBP exhibits extensive genetic heterogeneity with over 860 different amino acid sequences which have been divided in three main variants. A mass spectrometry technique (Selected Reaction Monitoring Mass Spectrometry) was developed to measure the absolute quantity of fHbp expressed in each of a panel of over 100 representative strains. Overall, as shown in **Figure 22** below, statistical analysis confirmed that strains carrying var1 fHbp subvariants expressed significantly higher amounts of protein compared to var2 or var3 strains (Mann-Whitney p-values <0.0001).

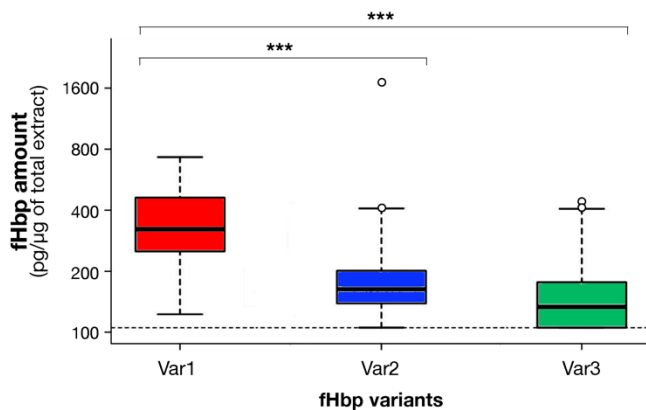


Figure 22: Expression of FHBP by each variant

Analysis of the genetic diversity in the upstream intergenic or promoter sequences of the strains allowed the identification of 8 major promoter clades using data originally generated by EUCLIDS partner Novartis (now GSK). When the absolute amount of fHbp measured in each strain was plotted against the clades of the intergenic region, we found that the protein expression was associated with the promoter clade. The susceptibility to complement-mediated killing correlated with the amount of protein expressed by the different meningococcal strains and this could be predicted from the nucleotide sequence of the promoter region. Therefore, expression of fHbp is genetically determined and we have identified a fHbp locus sequence typing system which can be used to predict fHbp levels in clinical isolates from the sequence of the fHBP gene and upstream intergenic region.

Lysis of bacterial strains carrying the fHbp was investigated by the EUCLIDS partner Tang and published previously (Caesar *et al.*, 2014). The conclusion was that the Var1 fHbp used as the background for the study (V1.1.) revealed a degree of specificity for FH over FHR-3 (in the order of ~ 20 -fold) in contrast to Var2 and Var3 fHbp, suggesting that certain fHBPs can discriminate between FH and FHR-3. To examine whether this had any functional consequence, complement killing of isogenic strains expressing fHbp sequences with different abilities to discriminate between FH and FHR-3 have been examined. Alternative complement pathway (AP) killing was assessed to exclude any confounding influence of antibodies in sera preferentially recognizing different fHBPs. Bacteria expressing Var3 (V3.28) fHbp, which has identical affinities for FH and FHR-3, were the most sensitive to AP killing, while the strain expressing the Var1 (V1.1) fHbp sequence which has ~ 20 -fold tighter binding to FH than FHR-3 was the least sensitive; bacteria with fHBPs with intermediate FH specificity displayed intermediate levels of protection. The ability of fHbp to favour FH binding promotes bacterial survival and offers a potential explanation for the prevalence of strains expressing Var1 fHbp causing invasive disease.

Recent data on fHbp sequences have been collected from the EUCLIDS cohort (isolated by MCP). The results of the subset sequenced so far showed a distribution of the variants of fHbp in the strains as 69.1% variant 1, 29.1% variant 2 and 1.8% variant 3. Analysis of the *fHbp* intergenic region identified 9 of the 11 main fHBP intergenic region (fIR) alleles present. Therefore, sequencing data indicate that variant 1, which exhibits the highest expression levels and binds with higher avidity to FH with respect to FHR-3, is most prevalent in the EUCLIDS cohort. The whole genome sequencing data once finalized will allow us a better understanding of any possible correlations between the pathogen/host genotype and will allow the investigation of a 'fit' between the levels of fHbp and FH in each case of IMD. This will take place in the coming months. Further analysis of FHR protein binding to meningococcal strains allowed us to detect recombinant FHR-5 binding but this is to be confirmed with plasma-derived FHR-5 using normal human serum and our well characterized anti-FHR-5 mAbs. Since we did not identify the capsular structure to which FHR-5 could bind, or whether this truly happens in the presence of serum, containing high levels of FH, it remains uncertain what role this minor component may infer on complement-mediated lysis of meningococcal strains.

We can conclude that our data collectively indicate that the competition of FHR-3 with FH for fHbp binding is unlikely to be an important factor to rescue the host from meningococcal disease. The main reason for susceptibility to meningococcal disease relates to the levels of FH as was further unravelled at the genetic level described above.

5.1. Binding of CFH and CFHR1-5 to pneumococci.

The most commonly studied pneumococcal strains in our EUCLIDS cohort led to the confirmation that FH and FHR-1 were able to bind to pneumococcal strains derived from patients. None of the other FHR proteins interacted with the pneumococcal strains. Binding of FHR-1 did neither compete with FH binding nor did it depend on the presence of the FH-binding protein PspC. One possible binding protein is the recently characterized pneumococcal protein Tuf, as reported by Mohan *et al.*, (*Mol. Immunol.* 2014) to bind FHR-1. Recently, PspC was reported to cause an intramolecular switch in FH increasing its binding to C3b and thus complement regulation. A similar potentiating activity, as suggested for the PspC:FH complex, is also induced by a monoclonal antibody that we developed in a recent project unrelated to EUCLIDS (Pouw *et al.*, unpublished).

Apart from the common pneumococcal serogroups, additional studies will be initiated in the near future on FH and FHR protein binding to *Staphylococcus aureus* and beta-hemolytic *Streptococcus pyogenes* Group A as the other major invasive pathogens in paediatrics.

Our overall conclusion is that FH binding (and thus genetically controlled levels of FH) is highly likely to impact on susceptibility to other invasive infections. This hypothesis will be addressed by further analyses using the clinical, genetic and biological data that we have generated throughout the past 5 years.

6. Genetic determinants of Severity of meningococcal and other bacterial infections

To identify genetic determinants of severity and outcome, we first undertook GWAS SNP analysis of the UK, Western European and Southern European meningococcal cohorts, relating SNPs to clinical markers of severity and outcome (Table 4). The findings were then validated in the prospective EUCLIDS cohorts. For this analysis it was necessary to have clear clinical definitions of disease severity. We used clinical outcomes and mechanical ventilation as robust severity markers across different diseases (Table 5).

Cohort	Cases	Controls	Total	SNPs (maf>0.5%)
UK	475	4703	5178	5,632,096
Austrian+Dutch	344	2587	2931	5,620,712
Spanish	417	882	1332	5,766,568
Total / Overlap	1236	8172	9441	4,336,327

Table 4: The GWAS genotype data used for the meningococcal severity analysis

Endpoint	UK		Spain		Austrian+Dutch		Total	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Severity endpoints:								
Death	18	424	8	354	19	314	45	1092
Skin graft/ amputation	18	424	9	264	15	318	42	1006
Skin graft/ amputation/ death	36	406	16	257	34	299	86	962
Mechanical ventilation	321	121	62	185	75	245	458	551

Table 5: the number of patients for each severity endpoint

In addition to the clinical end points shown above we used laboratory markers as intermediate markers of meningococcal disease severity that are strongly associated with outcome of meningococcal disease, as shown in Table 6 below. The intermediate phenotypes are potentially closer to underlying biology, and are continuous measures rather than binary, so they potentially contain more and more powerful information.

Measure	UK	Spain	Austrian+Dutch	Total	Correlation with outcome	
					Death	Skin graft/Amputation
Plasma measures:						
aPTT	353	211	281	845	0.150	0.230
Bases excess	371	218	240	829	-0.210	-0.160
CRP	350	0	288	638	-0.100	-0.027
Fibrinogen	311	181	231	723	-0.210	-0.240
INR	344	165	127	636	0.140	0.200
Lactate	224	0	115	339	0.130	0.240
Platelets	391	265	313	969	-0.190	-0.220
Potassium	388	246	308	942	-0.060	-0.045
WCC	393	267	324	984	-0.120	-0.170
Severity scores:						
PRISM	276	217	43	536	0.120	0.110
GMSPS	395	0	255	650	0.200	0.200

Table 6: Intermediate markers of meningococcal disease severity.
Measures marked in blue are negatively associated with endpoints.

Genome-wide association analyses (GWASs) were run for all sixteen severity measures separately in the three cohorts and in analyses pooling the three cohorts using linear and logistic regression where appropriate. The continuous outcomes were inverse-normal transformed to ensure normality. SNP selection for replication was based on the pooled analyses, analyses in the individual cohorts were used to detect between cohort heterogeneity, removing those SNPs showing significant heterogeneity in effect sizes at $p < 0.0001$. For each selected locus, regional association plots were inspected for evidence of multiple independent signals/causal variants; evidence for additional signals within loci was defined as SNPs with $r^2 < 0.2$ with the lead SNP and association p -value $< 10^{-4}$. The number of loci selected for each phenotype is shown in Table 7.

Phenotype	N	No. sig. Loci	Lowest maf Locus			Highest p locus		
			p	BF	maf	p	BF	maf
WCC	972	8	2.3e-7	3789	0.02	7.4e-6	3112	0.29
Platelets	957	8	4.8e-7	13440	0.047	7.6e-6	3018	0.23
aPTT	846	5	4.6e-6	3222	0.098	4.8e-6	4387	0.17
INR	635	2	4.6e-7	17477	0.12	6.4e-6	3389	0.40
Fibrinogen	725	5	1.8e-7	7633	0.033	6.1e-6	3572	0.27
CRP	635	5	2.0e-6	4659	0.077	4.4e-6	4828	0.36
Potassium	932	0						
Base Excess	826	5	3.0e-6	4133	0.091	5.2e-6	3331	0.11
Lactate	338	1	2.7e-6	3411	0.13			
PRISM	530	8	2.6e-6	4519	0.088	5.7e-6	3583	0.23
GMSPS	646	4	5.9e-6	3462	0.17			
BEP	818	6	1.3e-6	5086	0.056	6.8e-6	3148	0.19
Skin Gr/Amp	162	0						
Death	172	0						
Mechanic Ventilation	1000	1	1.6e-6	13404	0.44			
Death/Amp	316	2	3.7e-8	75497	0.21	2.1e-6	3934	0.46
Skin Gr.								
Susceptibility	4294	10	7.7e-6	3015	0.12			

Table 7: Sample size (effective sample size for binary outcomes) and number of loci with at least one SNP with Bayes Factor >3000. Also shown for each phenotype is the lowest minor allele frequency and highest p-value of all SNPs selected.

6.1. Validation of genetic associations with severity of meningococcal infection

Association between long non-coding RNA and white cell count and risk of mechanic ventilation

The most significant severity finding from discovery GWAS was between SNPs in a long non-coding RNA (lncRNA) and white cell count (WCC) in acute meningococcal patients.

We examined the association of the leading SNP in the replication cohort. The rare allele is associated with higher levels of aPTT ($p=0.009$); lower levels of WCC ($p=0.149$) and increased risk of mechanic ventilation in the pooled group of all definite and probable bacterial infections ($p=0.019$, $OR=1.32$) and in particular group A streptococcus patients ($p=0.004$, $OR=3.5$) (Figure 23).

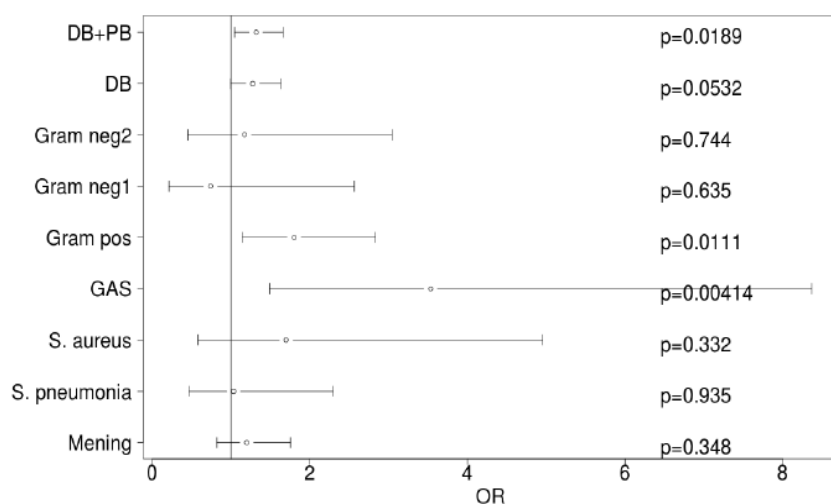


Figure 23: Estimated odds ratios, confidence intervals and p-values for association between long non coding RNA SNP and risk of mechanic ventilation in different groupings of bacterial infections. DB – definite bacterial, PB – probable bacterial, Gram neg1 – E coli + gram negative rods, Gram neg2 – E coli + gram negative rods + H influenza, Gram pos – group A Streptococcus + S. aureus + S. pneumonia, GAS – group A Streptococcus, Mening – meningococcus.

6.2. Association between risk of death, skin graft or amputation and novel variant

In the discovery cohort we found an SNP cluster that is associated with death, skin graft or amputation (severe phenotype) meeting Genome wide significance. This locus is currently undergoing functional validation.

Conclusion

The severity analysis has identified SNPs associated with severity and outcome of meningococcal disease and other life threatening infections. The strength of these associations is likely to increase with the additional genotyping of EUCLIDS samples currently being undertaken. The work establishes the principle that severity and outcome of bacterial infections is under genetic control. Several other genes show association with outcome and will be explored in detail in the coming months. Functional studies on the role of the long non-coding RNA in determining disease severity are under way.

7. Extreme phenotype analysis by exome sequencing

Whole exome sequencing was used to identify rare gene variants predisposing to each infection. We sequenced 229 meningococcal disease cases, and varying numbers each of pneumococcal, Group A streptococcal, staphylococcal, and other gram negative bacteria. Exomes were jointly genotyped using established methods and a customised pipeline was developed to annotate and filter the detected genomic variants. Variants that appeared to be common in the general population (allele frequency $>5\%$ in ExAC or the 1000 Genomes Project) were filtered out. Similarly, we excluded variants that did not affect the resulting protein sequence (based on Ensembl's Variant Effect Predictor). The resulting filtered variant set was used as the starting point for all downstream analyses.

Familial cases were examined separately using a novel statistical approach, which identifies and prioritises genomic segments that are identical by descent (IBD) in the affected family members. The remaining meningococcal cases sequences were analysed as a cohort and put through 1) a gene burden test, 2) pathway burden test and 3) candidate gene/pathway analysis to identify any potentially relevant genes. For the burden analysis we used ~400 ethnically matched exomes from the 1000 Genomes Project as controls and investigated whether any genes or pathways were enriched for pathogenic variants in our IMD cases. For the candidate approach, we narrowed our focus to previously implicated genes and pathways and sought to identify mutations of clinical relevance (Figures 24, 25).

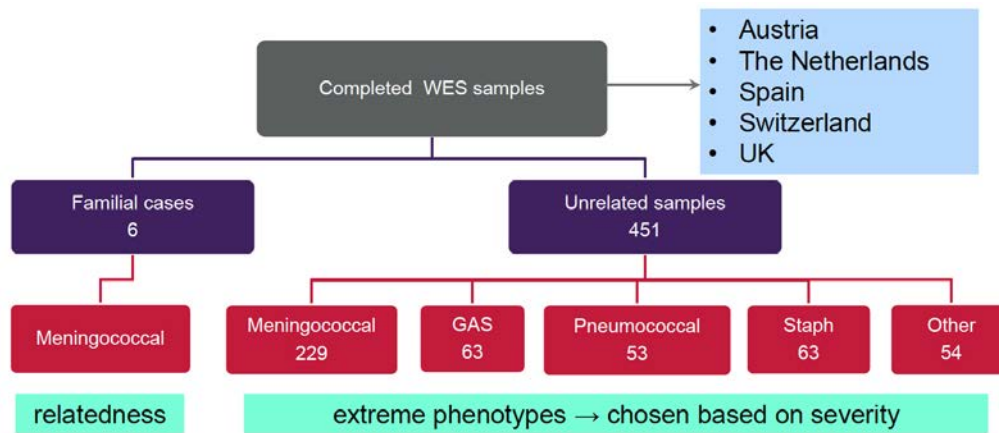


Figure 24: Whole Exome Sequencing study – breakdown of WES dataset

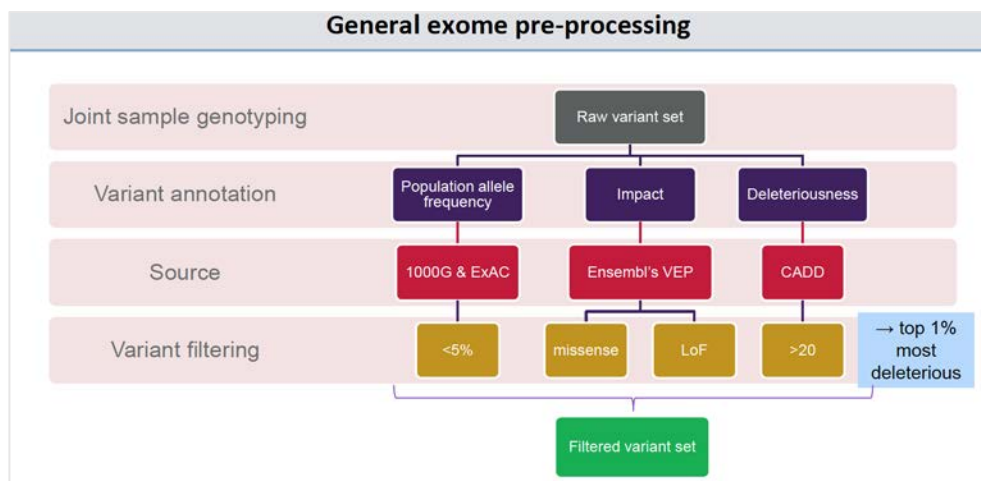


Figure 25: Process to identify genetic variants from exome sequence data. Specific pipelines were used for the analysis of the cohort and the familial cases. The analysis of the familial cases using Mendelian segregation led to identification of a number of novel genes associated with disease within families.

7.1. Familial Analysis

Whole exome sequences of 3 cousins with IMD (Figure 26) was analysed from the ‘familial’ cohort of patients, leading to the identification of a novel heterozygous predicted deleterious mutation. The mutation was Sanger sequence confirmed in all affected patients and all family members are being sequenced for the mutation. One additional IMD patient that died from meningococcal sepsis was found to carry a different rare heterozygous of the same gene. This is pending Sanger sequence confirmation.

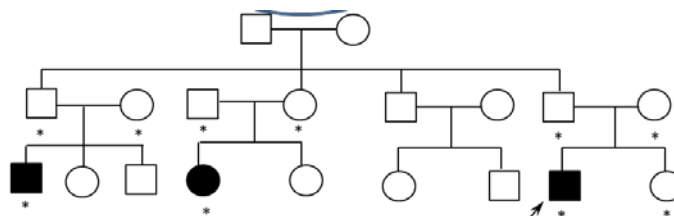


Figure 26: Pedigree of extended family with meningococcal disease novel rare mutation

7.2. Familial Analysis: novel mucosal secreted protein

The genetic analysis of one family in which two siblings suffered from IMD, revealed a novel non-synonymous missense mutation in a *novel mucosal secreted protein gene*. This gene encodes a protein specifically expressed in saliva, nasal lavage fluid, nasopharyngeal and tracheobronchial epithelial cells¹¹. The function of this protein is not well characterised but it has been suspected to play an important role in host innate immune defence against respiratory infectious pathogens, owing to its structural homology with LPS-binding protein (LBP) and bacterial permeability increasing (BPI) related proteins. In the context of IMD, we hypothesised that this mutation may contribute to

susceptibility to IMD. The novel missense heterozygous mutation was confirmed in two affected siblings and unrelated 3rd IMD case by Sanger sequencing (). This mutation was not found in 57 healthy adult controls tested in our laboratory or in any other publically available databases mentioned previously (Figure 27).

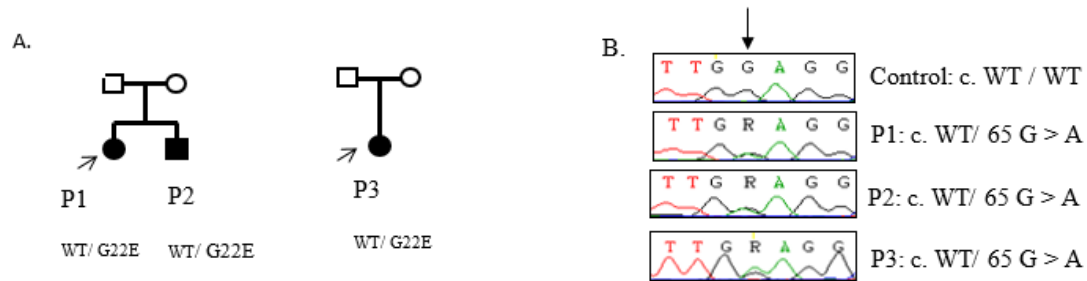


Figure 27: Heterozygous mutation in IMD patients.

Previously, the gene product has been shown to inhibit biofilm formation by gram-negative bacteria such as *K.pneumoniae*. We assessed whether mutant protein could have the same function on *Nm* biofilm formation using a standard microtitre plate based biofilm biomass assay. The assay measured wild type and pili deficient (*pilE*-) MC58 *Nm* biofilm biomass formation on the microplate surface in the absence and presence of wild type and mutant *SPLUNC1* treatments. The *pilE*- MC58 strain of *Nm* was included as control, since it lacks the ability to form intricate biofilm compared with wild type²⁹ strain. The treatment with wild type significantly inhibited early biofilm biomass formation of wild type *Nm* MC58 strain and the biofilm biomass produced was comparable to that of *pilE*- strain. Both the treatments with G22E allelic mutant were able to inhibit MC58 biofilm biomass, but were partially inhibitory compared to wild type protein. The effect was determined to be specific, as treatment with control protein, BSA at the same concentration (10 µg / ml) had no effect on MC58 biofilm biomass.

As this gene has been shown to encode a secreted protein, its possible role in NM adherence and invasion into the epithelial cells were tested, in the context of IMD. The number of bacterial CFU per 16HBE14 monolayer was markedly decreased (~1 log reduction in growth) compared to the number seen with the control at the same concentration. Similarly, the invasion ability of NM into 16HBE14 cells was greatly reduced (~1/2 log reduction in growth) in the presence of the mutant protein when compared with control. Overall, these results indicate novel role for the encoded protein in inhibiting adhesion and invasion of *Neisseria* into epithelial surface, which in turn may lead to containment of pathogen at colonisation and prevention of invasive disease from occurring.

In conclusion, we have shown that our exome sequencing candidate gene and encoding a novel mucosal protein plays an essential role in preventing NM infection through inhibition of early biofilm formation, which in turn leads to increased adhesion and invasion of the airway epithelial cells. We further demonstrated that the novel G22E allelic mutation identified in 3 IMD patients may confer susceptibility to IMD by impeding the antimicrobial activity of the wild type protein in carriers.

7.3. Exome cohort analysis

Due to the relatively small sample size, traditional rare variant association analysis could not yield statistically significant results in our cohort. Thus, to increase the statistical power of our analysis, we designed a novel pipeline that aggregates variants at the gene level and at the pathway level using pathways from GO, KEGG & Reactome. To determine whether certain genes or pathways were overrepresented in our cohort, we estimated the burden of pathogenic variants in our cases versus 400 ethnically controls from the 1000 Genomes Project. To ensure robustness, we evaluated the statistical significance of our results using permutation tests.

Pathway for cohort analysis of exome sequences

Preliminary examination did not reveal any novel genes enriched in our cohorts. However, the data revealed novel variants in known primary immunodeficiency (PID) genes among a minority of meningococcal patients that were originally diagnosed as immunocompetent. The meningococcal patients also presented with an enrichment of variants in pathways associated with coagulation and ciliary function (Figure 28).

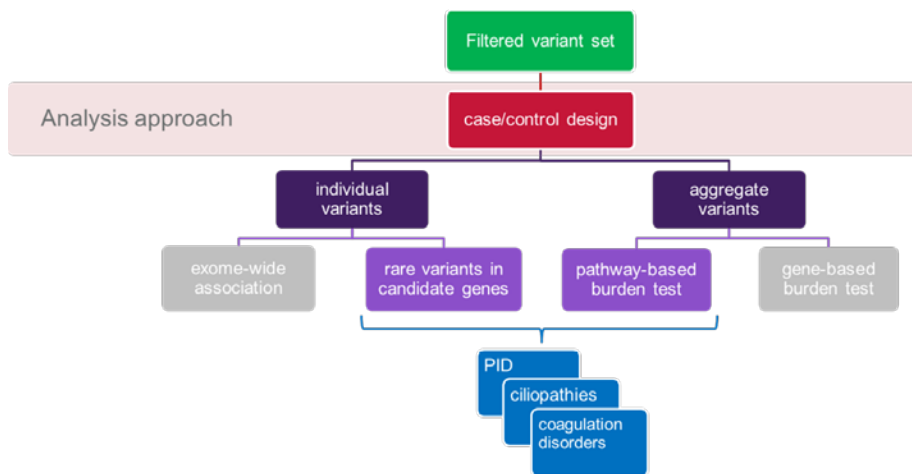


Figure 28: Cohort analysis of exome sequences

Preliminary examination did not reveal any novel genes enriched in our cohorts. However, the data revealed novel variants in known primary immunodeficiency (PID) genes among a minority of meningococcal patients that were originally diagnosed as immunocompetent. The meningococcal patients also presented with an enrichment of variants in pathways associated with coagulation and ciliary function. We identified a number of biological pathways in which there was over representation of rare deleterious genes (Figure 29).

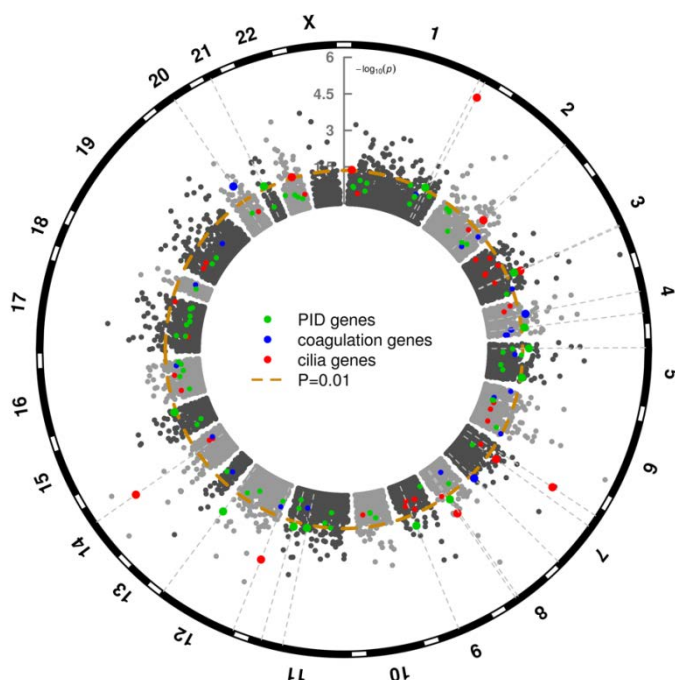


Figure 29: Gene-wise pathogenic variant burden. Each point represents a gene, arranged by genomic coordinate. The higher the variant burden of a gene in cases vs. controls the larger the radius. Chromosomes are arranged sequentially clockwise.

7.4. Complement pathway mutations in meningococcal disease.

An excess of complement mutations was found in the meningococcal cohort, comprising both homozygous mutations and compound heterozygous mutations in multiple genes in the complement cascade (Figure 30).

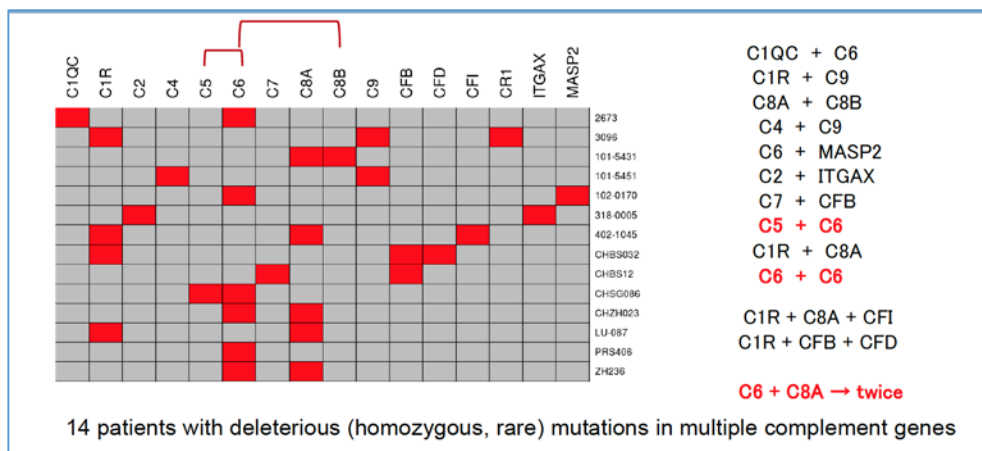


Figure 30: Combined complement mutations

7.5. Cohort Analysis: Primary immune deficiency genes

All sequences were tested for the presence of rare damaging mutations in any of the known primary immunodeficiency genes (homozygous MAF<0.01). A number of mutations were identified however none have been described as being pathogenic and requires further functional and clinical analysis.

7.6. Cohort Analysis: Mucosal surface functional genes

Unbiased gene-wise burden testing revealed multiple mucosal surface genes to be highly enriched in IMD patients versus controls. To reinforce this finding, we tested the pathway burden of ciliopathy genes in our IMD cohort and found a significant enrichment ($P<0.05$) for rare, deleterious variants.

7.7. Cohort Analysis:

Seven mutations in a gene controlling platelet numbers have been identified in 10 unrelated IMD patients by whole exome sequencing. The presence of these specific mutations in the patients has been confirmed by Sanger sequencing. Autosomal dominant (AD) somatic gain of function mutations have been identified in patients with adult onset ET and myeloproliferative diseases. Functional assays are being conducted for complete molecular characterisation of the mutations and to study patient cells to observe the impact of the mutations *ex vivo*. Overexpression of some of the mutations in relevant cell lines have shown one of the mutations constitutively activates the candidate protein. Cytokine production in patient cells *ex vivo* has been tested to understand the role of the mutations in the natural context of IMD infection. Normal or slightly high IL6 production was observed in response to PAM2, IL1 β and heat-killed *N. meningitidis* stimulation in one of the IMD patients. This patient had abnormally low platelet levels during acute infection and abnormally high levels in convalescence. Further studies are in progress to clarify the role of the novel mutations identified.

7.8. Cohort analysis Coagulation genes.

Exome analysis has identified an excess of deleterious rare variants in the coagulation pathway. The patients carrying these variants appear to have extreme phenotypes with very severe disease and high incidence of amputations and skin grafting

Conclusions

Exome sequencing has proved to be a powerful tool to identify the contribution of rare highly deleterious mutations to childhood disease. The vast amount of exome data is still in the process of being analysed, as variants have to be validated by Sanger sequencing, and functionally characterised before final confirmation as disease associated. However, from the analysis to date it is clear that novel and rare mutations contribute to disease together with the common variants described in earlier sections.

8. Genetic control of response to vaccine

The World Health Organization estimates childhood immunisations prevent 2.5million deaths annually. However, considerable inter-individual variability in the magnitude and persistence of immunity following vaccination is observed. Heterogeneous vaccine-induced immunity in childhood is particularly concerning, as infants generally have lower magnitude immune responses that wane more rapidly than adults. Maintenance of protective vaccine-specific serum antibody is essential for continuity of protection against rapidly invading pathogens, such as encapsulated bacteria.

To identify genes controlling persistence of vaccine responses, we undertook the first genome-wide association study of the persistence of immunity after immunisation with three routine childhood vaccines: capsular group C meningococcal (MenC), *Haemophilus influenzae* type b (Hib), and tetanus toxoid (TT) vaccines. We conducted a two-stage study, performing genome-wide genotyping on an initial cohort of 2000 European children, with replication data in a further 2000 individuals. Genotyping of the discovery cohort ($n=2061$) was conducted using the Illumina Omniexpress-12v1 or Omniexpress-12v1.1 microarray; following genotype imputation, using the 1000 Genomes Phase I integrated variant set release (March 2012) as the variant reference set, approximately 6.2 million SNPs were included in association analyses (31). We conducted quantitative trait association analyses for four log₁₀ normalised vaccine-induced immunological measures: (i) MenC-specific IgG concentrations (ii) MenC-specific serum bactericidal

antibody (SBA) titres (functional antibody), (iii) Hibpolyribosylribitol phosphate (PRP)-specific IgG concentrations and (iv) TT-specific IgG concentrations (Figure 31).

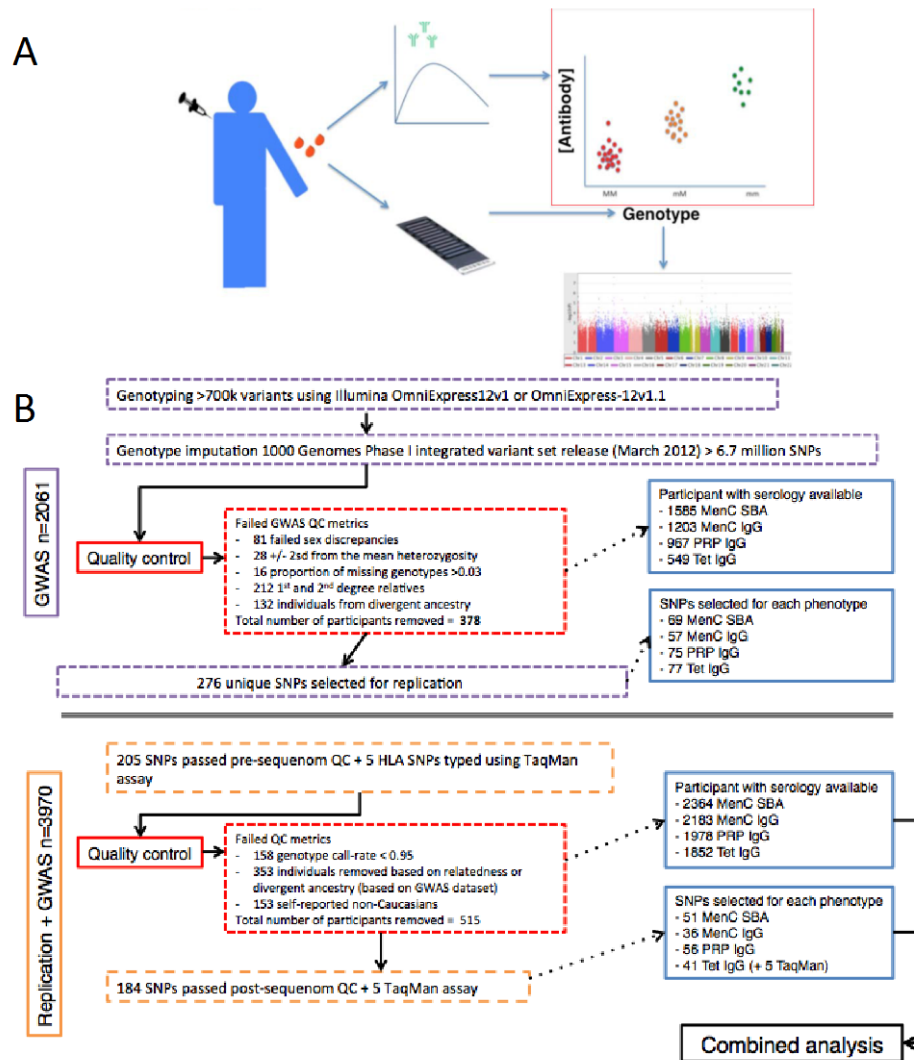


Figure 31: The overall plan of the GWAS

8.1. Significant results

Capsular group C meningococcal conjugate vaccine

A single locus was found to contain SNPs that were associated with the persistence of MenC-specific SBA titres, at the level of genome-wide significance, in the combined analysis of discovery and replication cohorts. These SNPs were within a genomic region containing a family of signal-regulatory proteins. Moreover, the sequences surrounding these SNPs were evaluated for potential transcription factor (TF) binding sites and the allelic variants were considered for their impact on TF motifs, using the R package 'motifbreakR'. The lead SNP at this locus (i.e. most statistically associated with MenC SBA titres), was predicted to substantially alter the motifs of four TFs: serum response factor (SRF), zinc finger protein 410 (ZNF410), retinoic acid receptor gamma (RAR γ) and RAR-related orphan receptor alpha (ROR α). Moreover, this SNP is an expression quantitative trait locus (eQTL) SIRP genes.

Haemophilus influenzae type b conjugate vaccine

Seventy-five variants were selected for Hib IgG persistence replication analysis. Six of these SNPs failed Sequenom[®] iPLEX design; none of the remaining 69 SNPs surpassed the level of genome-wide significance in the combined analysis of discovery and replication cohorts.

Tetanus toxoid vaccine

Seventy-seven SNPs were selected for tetanus toxoid IgG persistence replication. Twenty-two of these SNPs failed Sequenom[®] iPLEX design, leaving 55 SNPs for replication genotyping. As anticipated the replication SNPs selected within the HLA locus failed Sequenom[®] iPLEX design, due to the polymorphic nature of this locus, iPLEX primer design

is problematic; therefore, conditional linear regression analysis was used to select five HLA SNPs for genotyping using the TaqMan® SNP genotyping methodology. None of the Sequenom genotyped SNPs surpassed the level of genome-wide significance in combined analysis. However, a SNP in HLA typed on the TaqMan® platform, did surpass the level of genome-wide significance in the combined analysis. This SNP is annotated as an eQTL for HLA-DRB1 and HLA-DRB5 on the HaploReg database (<http://compbio.mit.edu/HaploReg>).

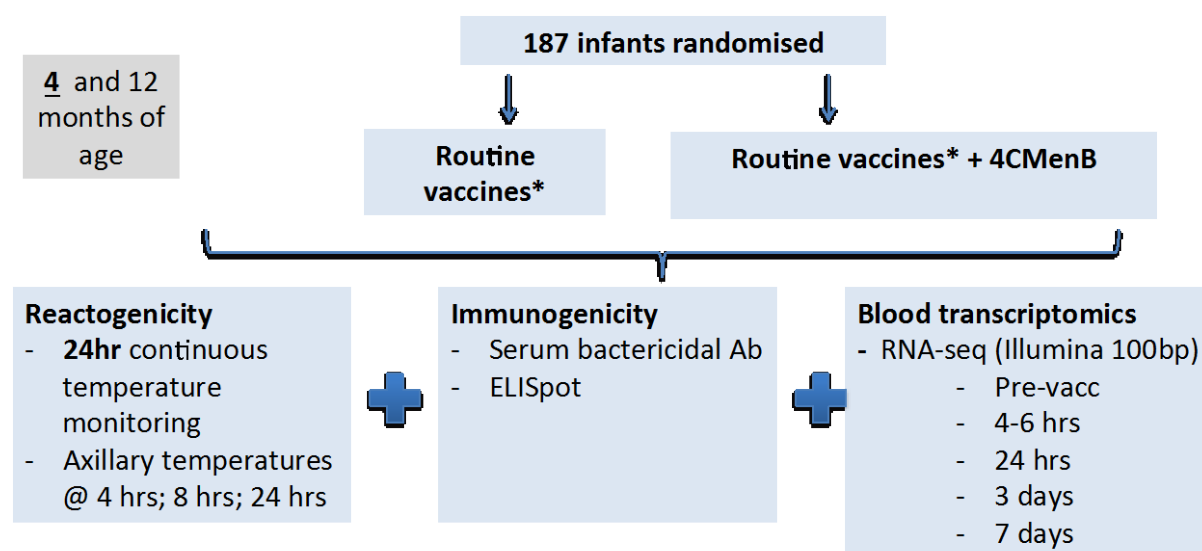
Conclusion

The genome-wide association study of persistence of immunity to three routine childhood vaccines: MenC conjugate vaccine, Hib conjugate vaccine and tetanus toxoid vaccine, identified two loci associated ($p < 5 \times 10^{-8}$) with the persistence of vaccine-induced immunity. Further studies are in progress to characterise these associations functionally.

8.2. RNA expression following immunisation with Meningococcal group B vaccine

EUCLIDS partner Oxford conducted a clinical trial assessing immunological and physiological responses, following infant immunisation with group B meningococcal vaccine (4CMenB, Bexsero®) in healthy infants. Bexsero® was licensed by the European Medicines Agency in January 2013, and was introduced into the routine infant immunisation schedule in the UK in late 2015. Recent UK data has shown a reduction in group B meningococcal disease in those eligible for the vaccine, since its introduction (Parikh et al., 2016). However, pre-licensure data showed this vaccine to be associated with significant reactogenicity, with some studies showing post-vaccination fever rates of up to 60% (Gossger et al., 2012). Public Health England deemed this vaccine to be reactogenic enough to recommend paracetamol (acetaminophen) usage, when this vaccine is administered to infants. This study utilised RNA sequencing to describe immunological and physiological responses to this 4CMenB when given concomitantly with routine infant immunisation, and compare these with control infants who received routine immunisations alone.

One hundred and eighty-seven infants were randomized to receive routine immunisations +/- 4CMenB vaccine. (Figure 32). Peripheral blood samples were taken prior to a second dose of vaccine (2+1 schedule), and 6 hours, 24 hours, 3 days and 7 days post-vaccination. Gene expression profiles were assessed, on RNA extracted from whole blood samples stored in PAXgene reagent, using Illumina® 100bp paired-end RNA-sequencing.



*4 months: PCV 13 (*Prevenar 13*®) and DTaP-IPV-Hib (*Pediacel*®)

12 months: PCV 13 (*Prevenar 13*®), Hib-Men C (*Menitorix*®)

Figure 32: Experimental design for vaccine study with Bexsero

A continuous temperature monitoring device, iButton®, was used to measure temperature for the first 24 hours after vaccination; in addition, repeated axillary temperatures were taken for the first week post-vaccination. Vaccine immunogenicity was assessed 7 days post-vaccination by ex vivo B-cell ELISpots, and serum bactericidal assay (functional antibody) titres were measured 28 days post-vaccination (Figure 33).

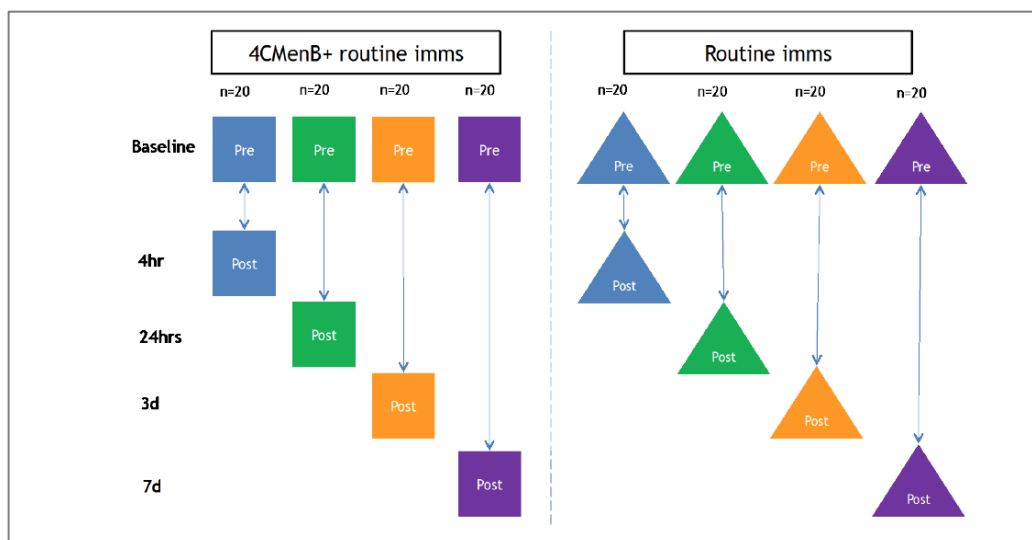


Figure 33: Collection of samples for whole blood transcriptomic analysis

8.3. Results

Peak differential gene expression 24 hours after infant immunisation

When the routine and routine + 4CMenB groups were analysed together, the peak in differential gene expression was seen 24 hours after vaccination (Figure 34). These may be the first data looking at the kinetics of early peripheral blood transcriptional profiles following infant immunisation, but are consistent with published data following adult immunisation.

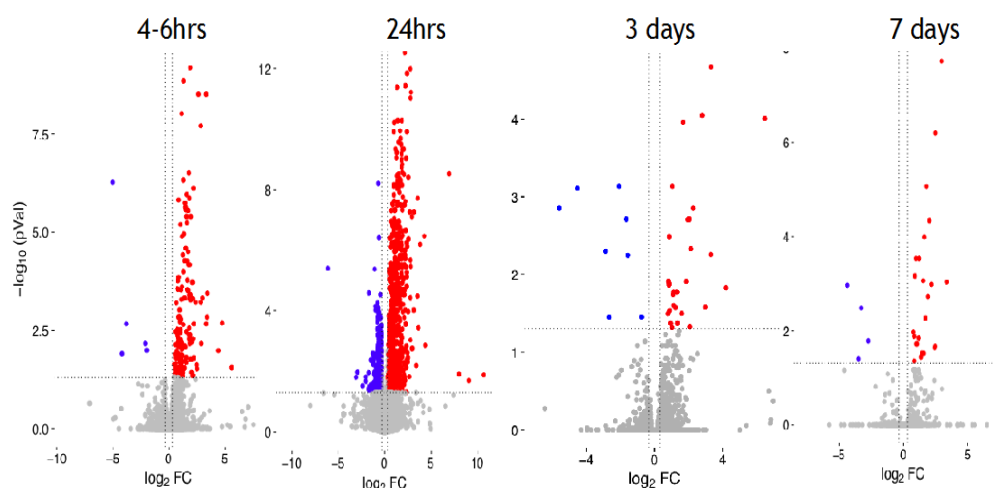


Figure 34: Plot of differentially gene expression at study time points

When the vaccine groups were analysed separately and the gene profiles compared, we observed a significant overlap in differentially expressed gene within the first 24 hours (Figure 35).

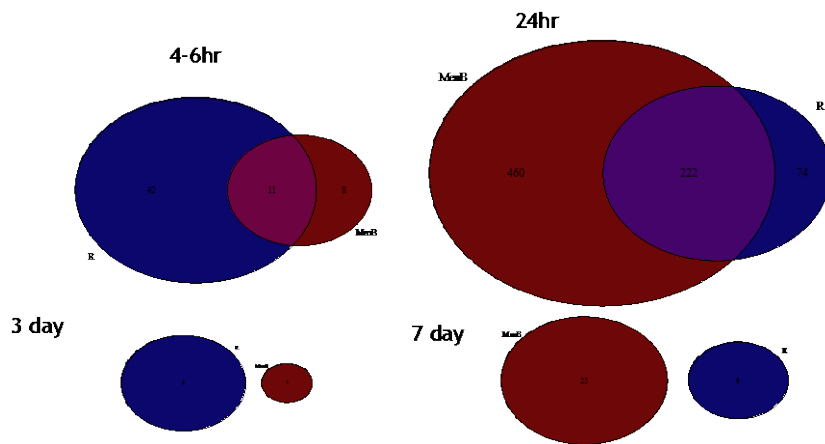


Figure 35: Common set of genes differentially expressed early following infant immunisation with routine or routine + 4CMenB vaccine

Differential regulation in vaccinees receiving 4CMenB vaccine

Differences were observed in gene expression 4-6 hours after vaccination in those who received the 4CMenB compared with those who received routine immunisations alone.

Differential gene expression in vaccinees who experienced a febrile event

We described differences in gene expression 24 hours after vaccination in those who experienced a fever ($>38^{\circ}\text{C}$) within the first 24 hours post-vaccination compared with those who did experience a fever. Significantly, these molecular differences may be causally implicated in the physiological changes resulting in fever; therefore, may give us important biological insight into this important clinical phenotype.

4CMenB vaccine is immunogenic after two primary doses

We analysed the SBA titre against the *Neisseria meningitidis* strains 44/76-SL at 5/6 (Table 8 and Table 9) and 13 months of age in children receiving 4CMenB at 2, 4 and 12 months of age (**MenB2,4**), as well as the control group who received 4CMenB at 6,8 and 13 months of age (**MenB6,8**). A titre of $\geq 1:4$ post-vaccination, was considered a “responder”. In the **MenB2,4 group**, one month after two primary doses (at 2 and 4 months), **97.3%** of the participants had a $\geq 1:4$ titre compared to 28.4% in the unvaccinated control group.

hSBA titre $\geq 1:4$				
Visit	Group	N	n (%)	95% CI
5/6 M	Control	74	21 (28.4%)	18.5% to 40.1%
	MenB2,4	73	71 (97.3%)	90.5% to 99.7%

Table 8: Proportion of participants with hSBA $\geq 1:4$ titre, classified by group and age, against the ref strain 44/76-SL

hSBA Geometric Mean Titre			
Visit	Group	N	Mean (95% CI)
5/6 M	Control	74	2.1 (1.7 to 2.4)
	MenB2,4	73	48.6 (37.2 to 63.4)

Table 9: hSBA geometric mean for all participants divided by time points and groups

Conclusions

The vaccine GWAS genetic polymorphisms associated with persistence of immunity to three infant immunisations, namely group C meningococcal (MenC) conjugate vaccine, *Haemophilus influenzae* type b (Hib) conjugate vaccine and tetanus toxoid vaccine. These data included the description of two loci associated with the persistence of vaccine-induced immunity, at the level of genome-wide significance ($p < 5 \times 10^{-8}$). Currently, the value of these individual genetic polymorphisms as biomarkers of immune responses appears limited, as they have modest effect sizes – so will not on their own be able to well classify infants whose immunity is likely to wane rapidly. However, they do present extremely valuable foundations with which to design further experiments to describe the mechanisms underlying persistent of infant immunity, which will inform future vaccine design and development. We have also described detailed gene

expression kinetics following infant immunisation, including gene signatures that appear to be specific to those the 4CMenB vaccine. **Importantly, we also described gene signatures that were associated with episodes of post-vaccination fever.** While such episodes are mild and self-limiting, fever in young infants can result in invasive procedures to rule out infection, and may ultimately influence public perception of vaccines and influence vaccine uptake. Therefore understanding the molecular mechanisms underlying fever following vaccination may ultimately result in the formation of less reactogenic vaccine formulations, while preserving immunogenicity.

9. RNA expression in bacterial and viral infections

EUCLIDS aimed to use transcriptomic analysis to understand the complex host response to infection. We performed peripheral whole blood gene expression analysis on over 1000 patients using RNA sequencing. Full analysis of this data is still in progress and will be reported over the coming months. Expression data were analysed using R version 3.1.2 (R Project for Statistical Computing). The culture confirmed bacterial patients had a different transcriptional profile to the viral patients and the healthy controls (Figure 36) and the significantly differentially expressed transcripts distinguishing bacterial infection from viral infection were identified.

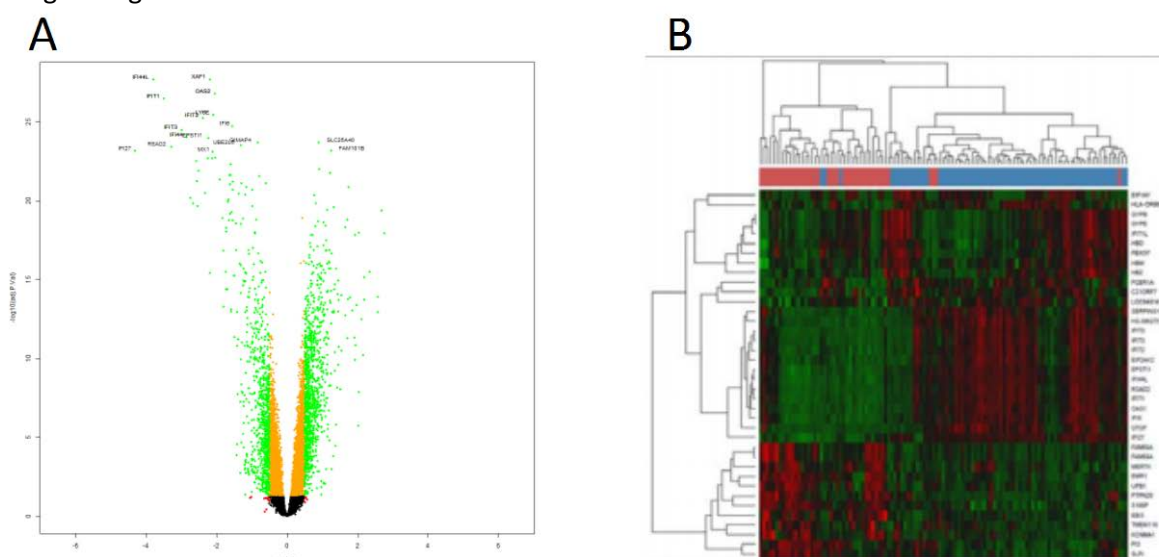


Figure 36: (A) Volcano plot for the bacterial vs. viral comparison using the microarray gene expression data. Significantly differentially expressed transcripts between patients (absolute log2-fold change >1 and 2-sided $P < .05$) are shown in green. (B) Clustering of bacterial and viral patients based on 38-transcript signature. Patients are represented as columns (red are patients with Definite Bacterial infection, blue are patients with Definite Viral infection) and individual RNA transcripts are shown in rows (RNA transcripts shown in red are up-regulated and those in green are down-regulated). The dendrograms for samples and transcripts are shown on the top and left of the heatmap respectively, indicating hierarchical clusters of the data.

9.1. Validation of 2 gene signature to distinguish bacterial from viral infection

Of the 8565 transcripts differentially expressed between bacterial and viral infections, 285 were identified as potential biomarkers after applying filters based on log fold change and statistical significance. Variable selection using elastic net identified 38 transcripts as best discriminators of bacterial and viral infection in the discovery test set, with sensitivity of 100% (95% CI, 69%-100%) and specificity of 95% (95% CI, 84%-100%). In the validation group, this signature had an AUC of 98% (95% CI, 94%-100%), sensitivity of 100% (95% CI, 85%-100%), and specificity of 86% (95% CI, 71%-96%) for distinguishing bacterial from viral infection (Figure 37).

After using the novel forward selection process to remove highly correlated transcripts, a 2-transcript signature was identified that distinguished bacterial from viral infections, including interferon-induced protein 44-like (IFI44L, RefSeq NM_006820.1), and family with sequence similarity 89, member A (FAM89A, RefSeq NM_198552.1). Both transcripts were included in the 38-transcript signature. The expression data of both genes in the signature was combined into a single DRS for each patient, which simplifies application of multi-transcript signatures as diagnostic tests. The sensitivity of the DRS was 86% (95% CI, 74%-95%) in the discovery group training set, 90% (95% CI, 70%-100%) in the

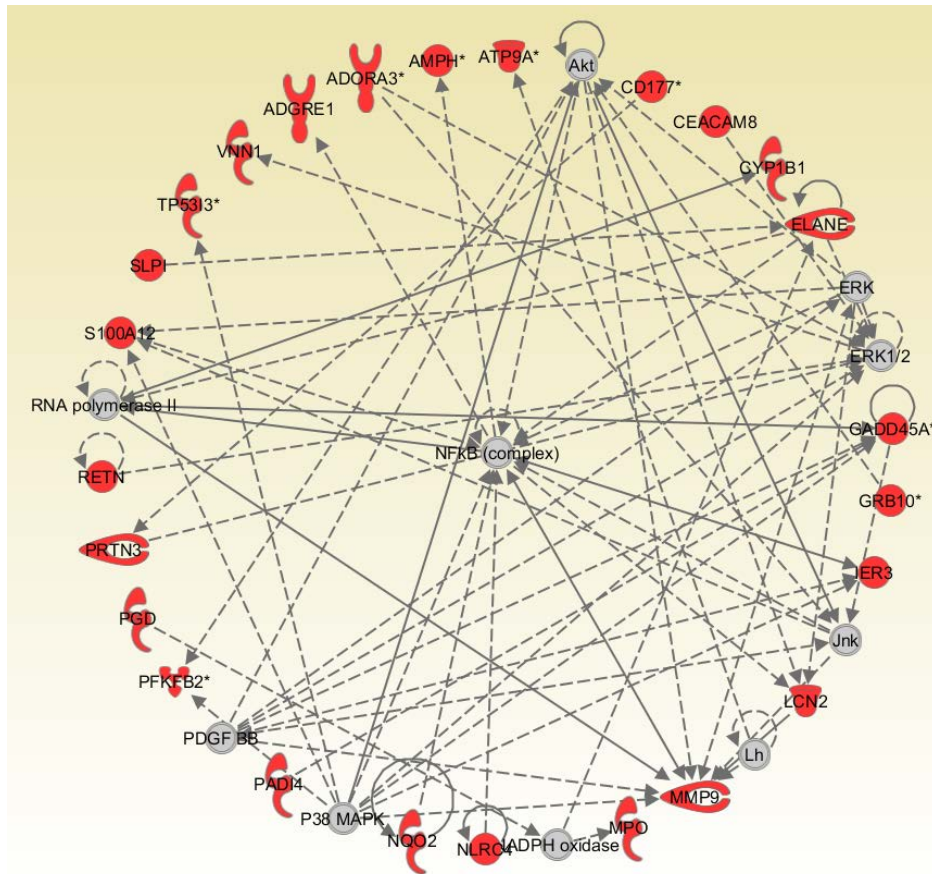


Figure 39: Genes significantly up-regulated in bacterial infection relatively to viral infection.

Conclusions

The full data from the RNA expression studies is still being analysed. Initial findings are that RNA expression can be used to distinguish different classes of infection such as bacterial and viral infection, and the gene expression data will be a valuable resource for studying the complex host response to each pathogen. On-going work is exploring the effect of host genetic variation on RNA and protein expression.

10. Integration of Genomic data and development of a bioinformatics resource for future research

EUCLIDS has generated a vast amount of clinical, microbiological, genetic, transcriptomic and epigenetic data on thousands of patients and healthy controls.

EUCLIDS has also generated a valuable resource of samples linked to carefully curated anonymised clinical data. The data has been linked and coordinated by Work Package 8 providing linkage of all genomic, clinical and protein data. In addition to providing long term storage and access to the data, the EUCLIDS bioinformatics team has developed novel bioinformatic methods for analysis of genomic data, and is developing new approaches to integration and analysis of the complex data sets.



Figure 40: Databases and integration

Integration of genomic, transcriptomic and phenotypic data

We have developed methodology to integrate genotype, gene-expression and phenotype data. This work was motivated by our discovered association between SNPs in a long non-coding RNA (lncRNA) gene on chromosome 22 and severity of disease. The developed methodology integrates the available genotype, gene expression and phenotype data in a Bayesian model by jointly modelling the effect of the top SNP in the lncRNA on the expression of each gene, and the effect of that gene's expression on severity variables.

Integration of genetic severity associations and genetic determinants of inflammatory disorder /Identification and genotyping of structural variation

We have developed two novel CNV detection and genotyping methods for capture sequencing experiments *cnvCapSeq* and *cnvOffSeq*. *cnvCapSeq* was designed specifically to address the challenges presented by current capture sequencing technologies when applied to contiguous genomic regions. This approach unmasks CNV evidence that was previously swamped by noise, while eliminating signal artefacts that would lead to spurious CNV calls. Our approach utilises information at the population level to achieve sensitive CNV discovery and high genotyping accuracy without the need for a control.

cnvOffSeq was designed to uncover CNV evidence outside the predetermined targets of capture sequencing experiments. Current WES technologies generate large amounts of sequence off-target, which is routinely discarded from all downstream analyses. This is due to the fact that off-target data arises through multiple distinct mechanisms and is thus highly heterogeneous and difficult to interpret. *cnvOffSeq* adopts an adaptive data-driven normalisation approach to de-noise and enhance off-target signal according to the underlying coverage pattern. *cnvOffSeq* is the first algorithm specifically designed for off-target CNV detection and was tested using public WES data from the 100 Genomes Project. It was compared against methods designed for on-target WES CNV calling as well as traditional WGS methods. In this benchmark, *cnvOffSeq* was shown to outperform existing methods by a wide margin, achieving 90% sensitivity for CNVs larger than 5kb, while maintaining a high specificity of 98%. As a result, this new algorithm constitutes a valuable tool for WES-based CNV detection that can tap into the unexplored potential of off-target datasets. The methodology paper describing *cnvOffSeq* was published in *Bioinformatics* ([8] Bellos & Coin, 2014). We have also developed novel approaches for identifying small signatures from complex genomic data (Kaforou PLOS Medicine 2013; Anderson NEJM 2014; Herberg JAMA 2016).

In the coming year the EUCLIDS bioinformatics team will integrate the multiple levels of data to describe how genetic variation influences RNA and protein expression and ultimately contributes to disease phenotype.

The data and sample bank generated during the work of the Consortium will continue to provide a unique resource for researchers worldwide aiming to improve the diagnosis, treatment and prevention of infection.

11. Expected impacts listed in the work programme

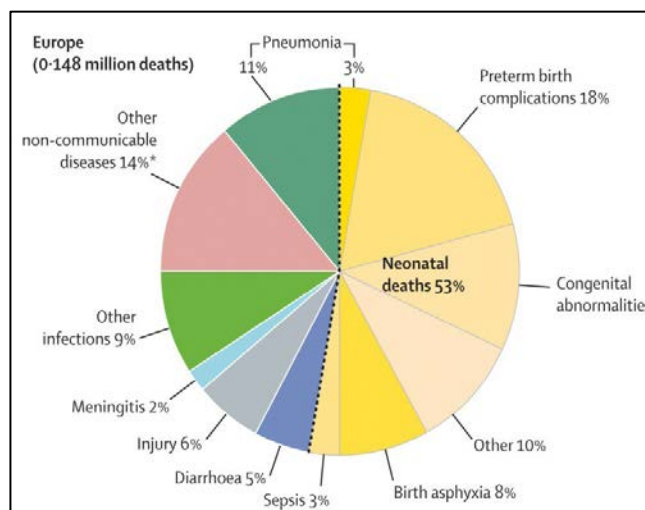


Figure 1: Child deaths in Europe

Our proposal to identify the genetic basis of susceptibility, severity and vaccine response to the major bacterial infections of childhood, and to determine the genes controlling the success or failure of immunisation, addresses an issue of fundamental importance to the health and wellbeing of children not only in the EU member states, but also in the global population. **The wellbeing of children, maintenance of child health, and prevention of disease, disability and death in children are issues of fundamental concern to all societies.**

Although the highest burden of death from infection is in the countries of sub-Saharan Africa, bacterial infections are major causes of hospital admission and death in the EU. Figure 1 shows that a third of child deaths continue to be due to infection in Europe (Black, Lancet 2010) with sepsis, pneumonia and meningitis accounting for almost a fifth of

child deaths. Our proposal, to identify the genetic basis of susceptibility to bacterial infection, and the genes that control severity of disease and response to vaccines, thus addresses an issue of fundamental importance to EU and Global community. **Our project will have impact on (a) socio-economics (b) scientific knowledge (c) clinical practice (d) therapeutics (e) public health and vaccinology (f) translation of research into clinical benefit (g) global health (h) European business and (i) European and African scientific and technical training.**

11.1. Socio-Economic Impact

The full societal cost of invasive bacterial infections such as meningitis, pneumonia, septicaemia, osteomyelitis and septic arthritis are difficult to establish, as accurate figures are not available to describe the full spectrum of these diseases in either EU or developing countries. Although estimates are available for individual diseases such as pneumococcal meningitis (McIntosh, Vaccine 2003), or meningococcal meningitis (Bilukha, CDC MMWR report 2005), the full economic and societal impact of these diseases is difficult to describe accurately due to their long-term consequences. Using standard estimates from EU countries such as UK or the Netherlands, mortality from pneumococcal meningitis occurs in 10% of cases, and long-term neurological disability occurs in a further 15-20% of cases, with deafness requiring cochlear transplantation in a significant proportion, and others requiring long-term educational support for learning disability or institutional care. The full costs of acute care including intensive care support, as well as the cost of managing long-term disability are significant in all EU member countries. For meningococcal disease, intensive care support may include requirements for dialysis, ventilation and myocardial support, whilst the long-term consequences of the neurological and musculoskeletal damage are great both in terms of economics, but also in terms of social and educational effects on the child and their families. Purpura fulminans (Figure 2) affects 5-10% of children with meningococcal sepsis, and leads to gangrene requiring amputation and/or skin grafting in 5-10% of survivors. Our own study of the long-term consequences for children surviving meningococcal disease (Allport, Paediatrics 2008) documented long-term effects on quality of life and physical functioning.



Figure 2: Purpura fulminans

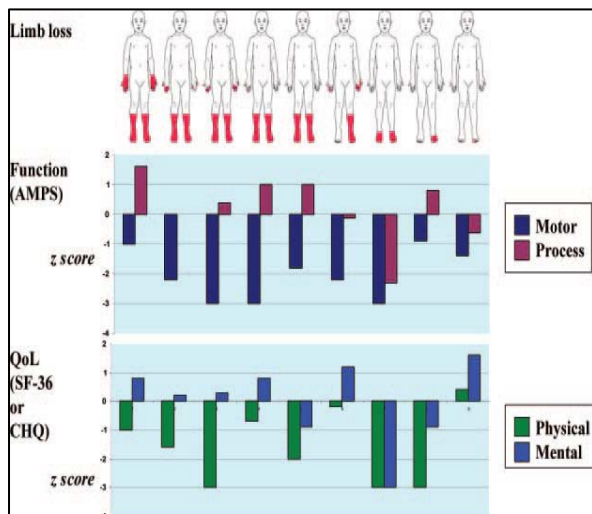


Figure 3: Functional scores for children with different levels of amputation in survivors of meningococcal sepsis

Figure 3 shows functional scores (AMPS, and quality of life (SF36 score) for children with different levels of amputation in survivors of meningococcal sepsis.

There are no accurate figures for the socioeconomic costs of staphylococcal infection. However staphylococcal osteomyelitis, septic arthritis, pneumonia and empyema are common causes of hospital admission. Cases are admitted on a daily basis to all major paediatric centres in the EU. These patients have prolonged requirements for intravenous antibiotics, orthopaedic intervention to drain bone abscesses, intensive care support, and long-term follow up.

The societal toll and economic cost of acute illness, and long-term disability from bacterial infection is thus immense in all EU countries, as well as in the developing world. EUCLIDS has helped to identify the genes and ultimately those at risk of severe invasive bacterial infection and severe disease, to improve prevention of bacterial infection by vaccination and to improve

treatment. It therefore contributes to reduction of global toll of death and long-term disability from these diseases. The UK Meningitis Research foundation has estimated the cost of Acute care, and longer term management of a typical case of severe meningococcal sepsis as being in excess of 1 million euro (www.meningitis.org.)

11.2. Scientific knowledge

Many of the most important scientific clues to the workings of the human immune system have come from identification of rare genetic defects. Our understanding of the importance of antibodies followed the identification of children with X-linked Bruton's agammaglobulinemia. Similarly the importance of T and B cells followed identification of the various forms of severe combined immune deficiency, and the contribution of neutrophils was revealed through identification of patients with chronic granulomatous diseases. In each case identification of the genes responsible was a crucial step in understanding immune function, in explaining the disease and in developing new forms of treatment such as antibody replacement, bone marrow transplantation or genetic therapy. The work by the co-ordinating applicant (Professor Levin) to identify the genetic basis of Mendelian susceptibility to mycobacteria opened up the field of Mendelian defects and led to the understanding of the roles of interferon gamma and IL12 in resistance to mycobacterial infections (Newport, NEJM 1996, Juanguiy NEJM 1996).

EUCLIDS has identified some of the Mendelian defects, common polymorphisms and major genes controlling susceptibility to and severity of life-threatening bacterial infection in children and is contributing to **a new understanding of the host-pathogen interaction for these important diseases**. For example, our identification of host variation in the the CFH region as key to meningococcal disease susceptibility has immense implications for understanding the interaction of the meningococcus with the host immune system, as the meningococcus is known to express a fH binding protein (fHbp), which is currently a major target for the new group B meningococcal vaccines. EUCLIDS has shown that both host and pathogen genomics contribute to disease susceptibility. **Furthermore Identification of genes and molecular pathways influencing susceptibility and severity of bacterial diseases is likely to reveal the factors that explain the pathophysiology of critical illness caused by infection; these would be key targets for therapeutic intervention.**

EUCLIDS has generated a vast amount of genetic and transcriptomic data on childhood infectious diseases cohorts, as well as healthy controls. In keeping with other major genomic efforts, such as the Wellcome Trust's Case Control Consortium (<https://www.wtccc.org.uk/info/access>) or the Malariagen consortium (www.malariagen.net), we will make all genetic and transcriptomic data publically available and open access to the international scientific community from the date of publication of the primary manuscript. Genotype, RNA expression, sequence and clinical data on these large cohorts of children with infectious diseases and those undergoing vaccination will represent a **resource to the scientific community for long-term use.**

11.2.1. Clinical practice

Unfortunately antimicrobial agents and current supportive treatments have failed to prevent death from childhood infection and mortality rates for paediatric septic shock remain 10-20%. For meningitis, mortality rates range from 5-30% depending on the pathogen (Tyler, Neurology 2008). Long-term neurological disability is common after

meningitis, whilst vascular complications of meningococcal disease, and bone and joint consequences of staphylococcal infection cause permanent disability in many children. The persistent toll from bacterial infection has highlighted the need for a better understanding of the inflammatory response to microorganisms, and for novel treatments to reverse the process leading to organ failure and death in severe infection.

Why does death occur despite administration of antimicrobial agents? Despite commencing anti-microbial agents, the progression of illness and deterioration to multi-organ failure and death is due to persistence of the inflammatory process triggered by the infecting agent, and progression of the physiological derangement that results from this inflammatory process (Van der Pol, Lancet Infectious Diseases 2008).

11.2.2. Failure of previous anti-inflammatory treatments

The past two decades have seen an intense effort by the scientific community and the pharmaceutical industry to develop immuno-modulatory treatments for severe infections, sepsis and meningitis. Although experimental treatments directed against early triggers of the inflammatory process, or individual mediators such as TNF and IL1, were shown to be highly effective in animal models of sepsis, in humans clinical trials of a large range of inhibitors of inflammation have proven largely disappointing, with no benefit demonstrated in the majority of these trials (Cunnington, Curr Top Med Chem 2008)). It is likely that this failure to improve outcome in human disease is due to a limited understanding of how human disease differs from experimental sepsis. The recent explosion in knowledge on the inflammatory response induced by microorganisms makes the earlier thinking appear simplistic. We now appreciate that following recognition of a range of pathogen associated molecules by host pattern recognition receptors, including the Toll-like receptors, a cascade of intracellular signalling triggers the induction of host inflammatory genes and the release of a wide range of inflammatory mediators including cytokines, chemokines, proteolytic enzymes and anti- microbial peptides (Opal, Int J Med Microbiol 2007). The intensity and duration of these inflammatory processes are likely to be determined by each individual's genetic makeup, and are likely to differ between individuals depending on the polymorphic variants carried by each individual. The failure of immuno-modulatory treatments targeted at specific mediators in clinical trials now appears to be largely due to the hitherto unappreciated complexity of the inflammatory process in human diseases, including the timing and role of each inflammatory component in the overall disease pathophysiology, and the importance of individual variation in response. The growing evidence that the host inflammatory response to infection is genetically determined (Alcais JCI 2009) suggests that genetically controlled differences in the host response result in different inflammatory pathways being activated in different individuals.

In view of the extensive evidence for genetic control of the host response to infection, identification of the genes controlling susceptibility, intensity of the host response, and response to vaccines will be a powerful method for understanding the biological processes involved in infectious disease. Identification of individuals highly susceptible to infection because of an underlying severe Mendelian defect, single gene defect or complex pattern of polymorphisms will allow focused treatment and follow up of such highly susceptible individuals. For example patients with complement deficiencies require long-term penicillin prophylaxis to prevent meningococcal disease. If genomic screening is able to identify the small proportion of the population who are at greatly increased risk of infection, then these individuals could be targeted either for antibiotic prophylaxis, for more careful follow up or for enhanced vaccination treatments.

11.2.3. Therapeutics

The genetic approach linked to transcriptomic and functional analyses which we are undertaking is likely to improve understanding of the immunopathogenesis of life threatening infection and the role of specific inflammatory mediators and their pathways. By identifying new targets for immunomodulatory treatment based on individual patterns of gene expression we will contribute to improved understanding of the immunopathogenesis of life threatening infections. **EUCLIDS has identified gene and transcriptome-based biomarkers that could ultimately employed to predict disease severity, identify immunological pathways involved in disease pathogenesis, and allow individually target treatments including immunomodulatory therapy. Further analyses of the huge amount of genomic data generated by EUCLIDS to identify predictive markers is currently underway.**

11.2.4. Public health and immunisation

Immunisation has been the most effective public health strategy for prevention of infectious diseases. Our study seeks to identify the cause of differing vaccine responses within the population and therefore **may allow individualised vaccination depending on genetic makeup.** Individuals genetically determined to have sub-optimal vaccine responses

could receive long-term antibiotic prophylaxis, or targeted vaccine approaches to overcome their immune defects. **Additionally, understanding the genetic differences between high vaccine responders and low vaccine responders is likely to inform the development of improved vaccines.** For example the use of different protein targets or adjuvants may help to overcome an impaired vaccine response in individuals with genetic variants of their pattern recognition receptors.

EUCLIDS' analysis of the genetic basis of vaccine response and persistence will have implications for the development and implementation of childhood vaccines. A common problem increasingly recognised for several childhood vaccines in the waning immunity with time following immunisation. As a result, diseases such as *Haemophilus Influenzae* type-B meningitis (Hib), and pneumococcal meningitis, may re-emerge in older children as the effect of infant vaccines declines over time. Understanding the genetic factors involved in vaccine-induced antibody persistence will enable introduction of more effective vaccines. Alternatively, identification of individuals likely to receive only short-term protection from childhood vaccines will enable targeted re-vaccination strategies. The potential cost-benefits of reduced need to re-vaccinate large portions of the population who are genetically determined to maintain long-term vaccine responses may enable targeted re vaccination of only those genetically determined to show poor long-term protection. **EUCLIDS has contributed to improved understanding of the genetic basis of vaccine response, and will ultimately help to develop individually targeted vaccination strategies.**

11.3. Translation of research into clinical benefit

Our programme of research has linked together a number of leading academic groups in Europe with SME partners, industrial partners and leading research groups in the developing world. The proposal is likely to identify biomarkers that can be used to predict those who are susceptible to serious invasive bacterial infection at a genetic level, those who may fail to respond to vaccines, and those at risk of severe outcomes from infection. Furthermore, based on RNA or genetic expression profiles, individuals with different patterns of inflammatory response who may benefit from specific immunomodulatory treatments may be identified. The EUCLIDS data on host response, genetic basis of susceptibility, severity and vaccine responses is currently being mined to identify findings which can be translated into clinical benefit by identification of at risk patients or understanding of the inflammatory response and identifying targets from intervention.

11.3.1. Global health

At a global level, life-threatening bacterial infections are priority diseases for developing countries. We have included West African cohorts, and African institutions as partners so that the impact of our studies is not confined to EU countries, but extends to the populations of Sub-Saharan Africa. EUCLIDS has identified a susceptibility locus and mechanism explaining susceptibility to meningococcal disease, which has major implications for the meningococcal belt of sub Saharan Africa where large epidemics of the disease continue to occur.

Our work to define the reason for susceptibility to infection, variations in vaccine protection and to understand the host response to infection are likely to inform global vaccination strategies and targeted treatments of these important diseases, and will contribute to the EU and Global effort to reduce childhood mortality in developing countries.

11.3.2. European business

The biotechnology and pharmaceutical industry are important strengths of the European economy. EUCLIDS included two SMEs as partners as well as major pharmaceutical and biotechnology companies. The partnership of health services, academia, SMEs and industry is likely to be a model for future joint research. The potential benefit arising from this collaborative research on EU biotechnology is thus significant, and the consortium has been a stimulant of European excellence in this area.

11.3.3. Why was a European approach needed to achieve these impacts?

In order to undertake a genome-wide study of major infectious diseases in children, and to enable results to be cross-validated on independent cohorts, a multinational approach was essential. It would have been impossible to establish cohorts of sufficient size with specific infections such as meningococcal disease without combining resources and utilising several EU member states for recruitment. However in addition to the requirement for several national patient cohorts, EUCLIDS required pooling of expertise not available in the individual participant countries. For example the expertise of AMC (Amsterdam, The Netherlands) in protein studies on complement, the expertise of our Austrian partner, MUG, in coagulation studies of meningococcal disease, has complemented the genomic and bioinformatic expertise of the UK partners. By combining patient cohorts, clinical and scientific expertise, EUCLIDS

was able to focus greater expertise and clinical resource on the problem than could have been achieved by any individual partner acting solely in their own country.

Impact on future research

Over the course of the five year EUCLIDS project the consortium successfully established a unique resource for ongoing research of patient information from seriously ill children with life threatening infections drawn from across Europe and West Africa. In addition to information on the illness, hospital course and outcome of each child there is detailed information on the laboratory findings, bacterial and viral pathogens causing and contributing to the illness and demographic data on predisposing factors and the patterns of illness across Europe. In addition to the patient information, detailed genetic analyses have been undertaken with genotyping of both DNA and RNA utilising the most sophisticated current methods for genome wide analysis of single nucleotide polymorphisms and sequencing. Furthermore additional analysis at a cellular and protein level particularly in patients with unusual or extreme phenotypes, provides detailed biological information on each disease. A biobank with many thousands of samples of DNA, RNA, plasma and surface swabs as well, as micro-organisms, has been established. The combined resource of patient information, genetic and transcriptomic information and samples for further research provide an invaluable resource to the scientific community worldwide for further research.

11.4. Implications of key scientific findings

The study has definitively identified the Factor H gene region as underlying susceptibility to meningococcal disease. Previous studies had implicated the Factor H region but the EUCLIDS programme has defined the genetic variants and mechanisms responsible for alterations in susceptibility to the disease. Our finding, that plasma concentrations of the complement inhibitor Factor H determine differences in susceptibility, and that the mechanism by which this occurs is through binding to the Factor H binding protein on the meningococcal surface, enabling the bacteria to evade killing by complement within the bloodstream has provided new understanding of how the meningococcus escapes the human immune response and why some individuals become infected. The detailed genetic analysis which has shown that plasma concentrations are regulated by a long range interaction from the Factor H related genes provides new insights which are relevant to many other infectious and inflammatory diseases. Complement Factor H is an important regulator implicated in many inflammatory diseases as well as infectious diseases. Our unravelling of the mechanisms controlling plasma levels of Factor H will thus be applicable to many other infections.

Meningococcal disease is a serious threat to populations in sub-Saharan Africa. EUCLIDS' initial finding through studies in West Africa that patients who had suffered meningococcal disease have higher levels of Factor H has great implications for understanding the epidemics of meningococcal disease in sub-Saharan Africa. The genome wide association studies of severity of disease have identified novel genetic associations and new mechanisms which may explain why some children succumb to devastating infections while others recover uneventfully. Our finding of a long non coding RNA controlling multiple severity genotypes and which appears to influence the outcome not only of meningococcal disease but of other bacterial infections has opened the way to a new avenue of research as have the other novel associations with disease severity we have identified.

EUCLIDS has explored the genetic underpinning not only of meningococcal disease but of multiple other bacterial infections. As many bacteria express Factor H binding proteins and utilise plasma Factor H to evade host responses, it was to be expected that our finding of the role of Factor H in meningococcal disease may also be applicable to other bacterial infections. The prospective EUCLIDS cohort has confirmed that the Factor H region is associated with severity of multiple other pathogens and thus our findings of the importance of the Factor H region and the mechanisms of its control genetically provide new insights into the major childhood pathogens including Group A streptococcus, pneumococcus and other gram negative pathogens.

EUCLIDS has applied next generation sequencing to undertake detailed analysis of patients with the most extreme phenotypes of infection. This beyond state of the art analysis has already identified novel rare Mendelian defects which underlie susceptibility to meningococcal disease and other bacterial infections. While this work is still in progress and full analysis of the extensive genomic data is likely to take a number of years to complete the picture emerging is that rare Mendelian defects in genes controlling adherence of bacteria to mucosal surfaces, ciliary function, invasion and survival in the bloodstream will explain an increasing proportion of children and adults who develop life threatening infections.

The identification of Mendelian gene defects underlying childhood infection has important implication for the individual families. The prospectively recruited EUCLIDS cohort has found a surprisingly high frequency of familial disease and previous infection in children admitted with life threatening infections. Our exome sequencing analysis

has identified rare highly deleterious mutations in approximately 10% of all patients. Many of the genes identified within families are likely to predispose to recurrent infection, or infection in other family members. The data is likely to require careful clinical and ethical implications of the finding of rare Mendelian mutations and how to counsel and protect individual families who may be at risk of repeated infections.

EUCLIDS' genetic analysis aiming to identify genes underlying vaccine response and focusing on persistence of antibodies is likely to have important implications for understanding why some children fail to be protected by vaccines and why others only have short term protection. The RNA expression analysis of response to novel vaccines has shown that the type and nature of vaccine administered results in many different patterns of gene expression. This data may help to guide our understanding of the nature of the inflammatory response required for long term protection by vaccines. Vaccines containing constituents such as outer membrane vesicles may elicit very different responses from purified proteins or polysaccharides and the RNA expression data generated will be of value in understanding the complexity of host response to vaccines.

A multi-level analysis was undertaken which has linked genetic and transcriptomics to understand childhood infection. The large RNA sequencing experiment which has generated RNA sequence data on over 1,000 patients is only just beginning to be fully analysed. Our initial analysis has confirmed that bacterial infections can be accurately distinguished from viral infection raising the possibility that small gene signatures (as small as two genes as described by EUCLIDS) can accurately distinguish bacterial from viral infection. Preliminary data suggests that each pathogen may elicit its own unique signature and thus the RNA transcriptomics may be an important means of diagnosing bacterial infection. The development of diagnostic tests based on host transcriptomics is an exciting area and the importance of this approach is highlighted by our finding that a high proportion of all patients admitted to hospitals remain undiagnosed with conventional microbiological investigations. Identification of the host response rather than the pathogen offers a new approach to diagnosis and is likely to be developed as a diagnostic method in future years.

Initial interrogation of the host transcriptomic response is providing new understandings of the complex biology underlying different disease presentations. The RNA transcriptomic data will be available to scientists worldwide for further interrogation of the biology of infectious diseases.

During the course of EUCLIDS novel monoclonal antibodies have been developed to Factor H and the five Factor H related proteins. The complexity of the region and the high genetic similarity between Factor H related proteins has previously made accurate measurement of these proteins difficult. In view of their importance not only in infectious diseases but also in inflammatory diseases the reagents developed through EUCLIDS provide researchers worldwide with powerful new tools for studying the role of the Factor H region in regulating the inflammatory response.

EUCLIDS has generated multi-level data ranging from DNA, RNA, micro RNA, protein and clinical laboratory data on many thousands of children. The Complexity of this data has required a novel bioinformatics, mathematical and statistical approach to understand how gene variants interact to regulate the biology of infectious and inflammatory diseases. The EUCLIDS consortium has developed novel analytical approaches and a community of computational biologists who will continue to work to fully analyse and interpret a huge amount of data that has been generated through the study.

11.5. Training and capacity building

During the course of the EUCLIDS project a large number of young clinicians, scientists, nurses and laboratory workers have been recruited to the project and trained. Many have achieved MSc or PhD and others have developed into promising scientists working in the field of paediatric infection. The study has thus had huge benefit to the development of European science and has contributed to strengthening the research base, expertise and capacity for paediatric research in Europe and West Africa.

EUCLIDS has provided an example of how successful, collaborative research across multiple European countries can contribute to solving important scientific problems. None of the individual partners or countries who contributed to EUCLIDS could have had access to sufficient numbers of patients with life threatening infections nor had the scientific capacity and expertise to conduct a study of this nature. It was only by pooling the patient resources of over 139 hospitals across Europe, the scientific expertise of different research groups and countries that a project of this complexity and nature could have been successfully concluded. Partnerships developed through EUCLIDS are likely to continue with collaborative research for many years to come. EUCLIDS provided the basis for the award of a major Horizon 2020 programme grant (PERFORM) to investigate the use of RNA expression and protein expression for

diagnosis of bacterial infection. This programme, which aims to improve the management of febrile children across Europe, has developed out of the successful EUCLIDS collaboration.

11.6. Benefits to the worldwide scientific community

The vast amount of clinical, genetic and transcriptomic data developed during the course of EUCLIDS and currently being analysed will be made Open Access and available to the international scientific community from the time of publication. Large scale data sets including exome sequence, genome wide SNP analysis and RNA sequencing data sets will provide a resource to science and the industry worldwide to explore the biology of life threatening infection. The transcriptomic studies and genetic studies of vaccine responses are likely to be used for further exploration of the processes underlying persistence of protective responses following vaccines.

11.7. Dissemination of the findings to the scientific community and public

Throughout the five years of the EUCLIDS project members of the consortium have presented their findings at major national and international paediatric and infectious diseases meetings including the annual European Society of Paediatric Infectious Diseases meeting, the European Academy of Paediatrics meetings and the World Society of Paediatric Infectious Diseases meetings. Members of the consortium have been keynote speakers at these and other international congresses and meetings. In addition the progress in the study and the major findings have been disseminated to the meningitis charities who have a particular interest in the field of meningococcal disease and other bacterial infections. Members of the EUCLIDS consortium have presented at bi-annual meetings of the Meningitis Research Foundation and members of the charity have served on the Advisory and Ethical Boards of the EUCLIDS project. Much of the data from the project is currently being prepared for publication while a number of publications have already presented the results to the scientific community the major papers arising from the work are currently being written and are likely to be published over the course of the coming year. Much of the major scientific impact from EUCLIDS will therefore appear in future years.

Finally, at a time when the world is facing complex political interactions between countries and numerous international conflicts, EUCLIDS has provided a wonderful example of how paediatricians, scientists and researchers from multiple countries can interact in a collegial and harmonious way to help improve our scientific understanding of some of the most serious illnesses facing the worldwide population.

12. Dissemination and/or exploitation of project results, and management of intellectual property

The institutions that are partners in EUCLIDS are academic institutions with a commitment to publishing research data in the highest quality publications. The major publications arising from EUCLIDS are currently being finalised for submission or are under review. We expect the publication output from this programme of work to be published in the leading international peer-reviewed journals. In order to ensure access for clinicians and scientists working in the poor regions of the world, we have a commitment to publication of our research findings in open access journals or in a manner which will ensure open access. In addition to publication in peer-reviewed journals, the work has been and will in future as analyses proceed be presented at both national and international medical research meetings to inform doctors, scientists, programme managers and allied health care workers of our findings.

All the partner institutions have press offices, with extensive experience in dissemination of scientific and medical information to both the scientific and lay public. The co-ordinating institution (Imperial College) has a track record of working with the responsible media to disseminate scientific advances to the public.

The partners in this application have established links with **patient support groups and charities**. These include the UK Meningitis Research Foundation, Meningitis Trust and Meningitis UK charities. Each of these organisations has established public information and media programmes, and we will work with these patient groups to present the findings from our studies in a publicly accessible, sensitive and ethically acceptable manner. The findings from EUCLIDS will be presented to the public in close discussion with these patient support groups.

Types of Dissemination

Table 1 shows the types of dissemination activities and the target groups.

Type of impact	Target	Countries involved	Partner
Corporate identity of EUCLIDS available through EUCLIDS website and partner institution websites	Public Worldwide	All	All
Training and site induction workshops; Exchange visits, educational workshops	Young clinicians and scientists employed on the project; MSc / PHD students	Europe and West Africa	All
Presentations at international meetings	Scientific community	All	All
Press releases, interviews, television and radio	National for each partner	All	All
Open Access of genomic data	Scientific community	Worldwide	1,9,10
Publication in high impact journal	Scientific community	Worldwide	All
Trainee and participant meetings. Opportunity to present work in progress	Students, young scientists employed on the project	Partner countries	All
Annual symposia with open and closed sessions inviting external and internal speakers. Publicised on web and press release	Public and scientific community, young scientists	Partner countries	All
Feedback to patients and patient support groups through newsletters and website	Public and patients	All partners	All

Although this programme of work is conducted with the goal of elucidating the genetic factors involved in bacterial infections, there will be extensive research data produced by each component of the study, which may enhance understanding of the immunology, molecular genetics and pathophysiology, in clinical management of infectious diseases. Extensive data on gene expression and proteomics linked to detailed clinical findings will provide researchers throughout the world with access to novel data relevant to their understanding of many aspects of the biology of infection. Publications are now in progress **in the field of gene expression, the inflammatory response to the pathogen, the inflammatory response that distinguishes bacterial infections from other infectious agents, and the protein and body fluid responses to each pathogen.** We will endeavour to publish each component of the study as quickly as possible after completion to make the scientific data from this programme widely available.

The major goal of this study will be to identify genetic and gene expression profiles that can be used as predictive markers of susceptibility, severity and vaccine response to bacterial infection, within Europe and also in Sub-Saharan Africa and other developing countries. Findings relevant to this objective would be published preferably in widely circulated general journals and would be used to inform the development of a simple, robust and affordable diagnostic test suitable for use in both developed and developing country settings.

12.1. Project website

The EUCLIDS project website provides information to the scientific community, public, partners, and EU on the corporate identity and aims of EUCLIDS and the progress of the project. The website has sections with different levels of access to enable public, scientific community and partners to access.

Website link: www.euclids-project.eu

12.2. Training of staff and young scientists

All partners in EUCLIDS were committed to strengthening the scientific and clinical base of their institutions and countries. Our dissemination activities included dissemination of scientific and technical training to all staff employed on the project. The strength of our consortium is the wide range of expertise and technology brought together by the partnership. Each partner employed young staff who used their participation on the project for personal professional development, including undertaking PhDs, MS's or acquiring specific technical qualifications and skills. All staff received training through the annual scientific meeting, annual workshops and exchanges, and through invited international speakers who will participate in the annual scientific meetings.

12.3. Intellectual property arrangements and commercial development of results

The partner institutions in EUCLIDS have well established intellectual property departments with experience of protecting scientific discoveries and translation of discoveries into commercial development and health gain. The co-ordinating institution (Imperial College) has a particular experience in intellectual property protection and exploitation of major scientific advances in collaboration with healthcare institutions and industry. The underlying principle in the arrangements within our consortium is that all intellectual property arising from any discoveries is shared jointly between the contributing partners and formal agreements to this was established following the award of the grant in the consortium agreement.

As a component of the work aims to improve understanding of life threatening infection in developing countries, the partner institutions are all committed to development of any clinically important discoveries or those which affect vaccination in a manner that maximises benefit to developing countries in a non-profit manner. **Our arrangements for intellectual property and exploitation of results include the principle that any profits made in developed countries will be used to make available advances from this work in developing countries at cost. Furthermore, where possible industrial exploitation of any findings will utilise partners in Europe as well as in developing countries.**

12.4. Exploitable knowledge arising from EUCLIDS

The project has identified genes controlling the susceptibility, severity and outcome of life-threatening infection, as well as to identify genes controlling the success or failure of immunisation. The findings have potential for commercial development of products for clinical and public health use (for EU member countries), or non-profit development (for those outputs relevant to Sub-Saharan Africa)

Table 2 shows the types of exploitable output likely to arise from EUCLIDS

Type of output	Potential use	Exploitable by	Target population
Identity of genes controlling susceptibility	Test for identification of at-risk individuals	Biotechnology company; SME	Europe and developed world
Identity of genes controlling severity	Rapid identification of patients for early treatment	SME; Biotechnology company	Hospitalised patients with infection
RNA or genetic biomarkers of pathways involved in sepsis and meningitis	Targeted treatment with immune-modulating agents	Pharmaceutical Industry, Biotechnology	Patients needing intensive care
Genomic marker of vaccine failure/poor response	Identification of patients needing targeted vaccines, or non-vaccine treatment	Biotechnology, SMEs and pharmaceutical industry	Populations of EU member states and internationally
Identification of genes and pathways for optimal vaccine response	Improving vaccines incorporating better proteins or adjuvants	Vaccine companies	World-wide population

The partners in EUCLIDS have all signed a consortium agreement to adhere to REGULATION (EC) No 1906/2006 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 18 December 2006, which sets out the principle by which the EU encourages dissemination of results and protection of the intellectual property generated in the course of the project. Prior to commencement of the project we developed a consortium agreement, with assistance from experts in intellectual property and exploitation of scientific advances. As our consortium includes partners from national health services, SMEs, academia and industry, as well as developing countries, the IP issues are complex and we were defined in advance through relevant “background” information and IPR contributed by partners. The “foreground” which is arising from the project is currently being identified as analysis of the data progresses. Any potential IP that requires protection will follow the agreed plan as set out in the consortium agreement. **Imperial College Innovations** (the Imperial College-owned company responsible for exploitation of scientific advances) will interact with the IP divisions of all partner institutions to maximize the shared benefit to all partners and the EU funders of the action.

12.5. Ongoing identification of exploitable findings

Much of the genomic data developed during EUCLIDS is currently being analysed. As the work progresses any potential IP needing protection will be evaluated by the EUCLIDS consortium management group and taken forward for protection.

12.6. Policy on publication

Academic IP will be recognised in authorship of publications. In general we will seek to acknowledge jointly the contribution of all partners fully, and will assign authorship after joint consultation between all partners and in accordance with the Vancouver rules defined by the editors of major scientific journals.