PNEUMOPATH FINAL REPORT

Grant Agreement number:	HEALTH-F3-2009-222983
Project acronym:	PNEUMOPATH
Project title:	A comprehensive dissection of pneumococcal-host interactions
Funding Scheme:	CP
Period covered:	1 March 2009 to 31 August 2012

Project Co-ordinator: Professor Peter Andrew
University of Leicester

Tel: +44 (0)116 252 2951
Fax: +44 (0)116 252 5030
E-mail: pwa@le.ac.uk

Project website: www.le.ac.uk/projects/pneumopath

Contents

Final Publishable Report

1. Executive Summary

PNEUMOPATH was a wide-ranging study of host-pathogen interactions during disease due to *Streptococcus pneumonia*. The interdisciplinary approach was made possible by the PNEUMOPATH consortium, involving academic and industrial partners, gave rise to innovative ideas and results. The project resulted in entirely new insights into the pathogenesis of pneumococcal disease, with implications for diagnosis, vaccination and treatment.

The project was founded on the premise that because pneumococcal isolates vary in their repertoire of genes, the contribution of an individual factor to the infectious process will vary according to the other host and pathogen factors that are present. Previous studies tended to consider the contribution of each virulence factor in single pneumococcal strains or host factors in isolation, but in contrast, in PNEUMOPATH a panel of pneumococcal isolates associated with carriage or disease was used to challenge a range of in vivo and in vitro models.

The project confirmed that clonal types, exhibiting different invasive disease potential, populate pneumococcal serotypes. Furthermore, even intra-clonal variants showed major differences in pneumococcal surface antigens. This observation has implications for the design of protein-based vaccines. A DNA microarray chip covering all of the PNEUMOPATH strains was created and it was used to identify genes associated with invasive disease: results with high diagnostic and treatment implications. It also was found that most *S. pneumoniae* virulence factor genes are present in *S. oralis* and *S. mitis* and that only a few loci have *S. pneumoniae*-specific virulence genes. This observation is a very significant lesson for those seeking new diagnostics.

In vivo studies made important discoveries. It was found that cytokine levels could predict the outcome of pneumococcal disease in mice early in infection and were characteristic of host susceptibility. The prognostic significance of this observation should now be tested in humans. The *in vivo* studies also gave new insights into the process of pneumococcal sepsis development. It was ascertained that the action of spleen macrophages during the first minutes of infection determines the clinical outcome. Furthermore, invasive pneumococcal disease was shown to result from multiplication of a single pneumococcal cell.

PNEUMOPATH also investigated the genetics of susceptibility to invasive pneumococcal disease. Three loci associated with susceptibility were localised on two mouse chromosomes but none of the genes identified in the mouse showed a high association with invasive diseases caused by encapsulated bacteria in humans. In humans, two rare sequence variants disrupting two genes were discovered to have large effects on risk of IPD, when homozygous. These are by far the biggest genetic risk factors of IPD described to date.

An intriguing observation was that a pneumococcal isolate had an entirely different pattern of pathogenesis, with death occurring without sepsis. This isolate appeared to be invisible to the host, with an atypical absence of inflammatory cellular and cytokine inflammatory responses. *In vitro* studies supported the *in vivo* work, with this strain showing none of the typical cytotoxicity of pneumococci and failing to induce cytokine responses from cells in culture. This work emphasises that multifaceted approaches to treatment and prophylaxis may be required.

Other *in vitro* cellular studies were very informative. Significant differences between isolates in binding to cells and in susceptibility to uptake and killing by macrophages were observed. A direct correlation between *in vitro* phagocytosis by spleen macrophages and *in vivo* bacterial clearance from the blood stream was demonstrated, providing for the first time an *in vitro* assay that may predict pneumococcal virulence in the host.

A combination of genomic methods identified new factors involved in adherence. This work underlined the importance of nutritional balance during adherence and identified new potential therapeutic targets. Comparison of carbon and nitrogen regulatory networks revealed several overlapping targets. Various regulatory elements were discovered to be

important for growth: the global regulator CcpA affected expression of many C and N metabolic genes; two other transcriptional factors regulated arginine acquisition. One regulator was essential for pneumococcal survival and constitutes a new therapeutic target. The importance of nutritional factors in virulence was supported by the observation that better growth *in vitro* was linked to loss of virulence. The*se* studies reinforce the idea that a pathogen's physiology, a long neglected research area, could be a rich area of new therapeutic targets.
Several bioinformatic advances, including development of dedicated data processing services and an Application Programming Interface and deployment of a web-based semantic integration environment enabled computational analysis of the multidisciplinary experimental data.

2. Summary description of project context and objectives

Streptococcus pneumoniae is a major cause of human disease in all countries of the world. It is the main cause of bacterial pneumonia and sepsis accompanies many cases of pneumococcal pneumonia. In many countries, *S. pneumoniae* is the leading cause of bacterial meningitis. Issues related to anti-pneumococcal antibiotics and vaccines exacerbate the impact of burden of pneumococcal diseases.

Antibiotic resistance in *S. pneumoniae* is an increasing global phenomenon. Disturbingly, most penicillin resistant pneumococci also are resistant to cephalosporins. The increasing incidence of antibiotic-resistant pneumococci complicates empirical therapy of community-acquired pneumonia. The increasing prevalence of β-lactam/macrolide resistance means more reliance on newer agents, such as fluoroquinolones, but resistance to these is emerging The incidence of multidrug resistant strains also is driving more reliance on vancomycin, yet about some pneumococcal isolates already may be vancomycin tolerant. Resistant strains also can affect adversely the outcome of treatment, with mortality being higher in disease due to antibiotic resistant strains.

New pneumococcal conjugate vaccines appear to be successful they are unlikely to be a long-term solution. Although sometimes useful, pneumococcal polysaccharide vaccines are not generally effective, especially in children. The limitations of polysaccharides as vaccines led to the introduction of a polysaccharide-protein conjugate vaccine, which have been a great success, especially in reducing the incidence of pneumococcal diseases in children. It seems, however, that this success may diminish. Conjugates are resulting in changes in pneumococcal epidemiology, with increasing disease from new strains of non-vaccine serotypes and by existing strains that have acquired new capsular polysaccharides. Furthermore, even if conjugate vaccines do decrease the incidence of pneumococcal diseases, new protein vaccine components will be needed to maintain this trend and to provide vaccines for geographical regions with serotype prevalence not covered properly by the conjugates.

The issues outlined in the preceding paragraphs point to the need for further research. However, almost all previous studies of pneumococci in models of disease had been done with a very limited number of strains, and often only one strain. This is because over the past twenty years study of infection has tended to take a reductionist approach, with the contribution of individual virulence factors or host factors being considered in isolation This may give a distorted picture because pneumococcal isolates have differing potentials to cause diseases. Different pneumococcal isolates have differing propensities to cause invasive disease or to be associated with carriage. This suggests that the mechanisms for evoking pathogenicity differ between strains. It is very well accepted that the outcome of an infection is determined by the interplay of the collection of attributes of the host and of the pathogen. Also, it is known that pneumococcal isolates vary in the repertoire of genes that they possess and hence the contribution of an individual factor to the infectious process may

vary according to the other host and pathogen factors that are present. Thus, it was deduced that studies with a much wider set of isolates were required.

The overarching hypothesis of the PNEUMOPATH project was that through a systems approach to host-pneumococcal interaction, the most important and consistently involved host and pneumococcal factors would be identified. These could be targets for diagnosis, therapy, enhancement of host defence and prophylaxis. Four individual hypotheses were tested:

1. The carriage rate, invasive disease potential, epidemiology and spread of antibiotic susceptible and resistant strains depend upon the repertoire of genes and the expression of genes for individual virulence factors that will vary between pneumococcal isolates of different clonal types.

2. The influence on the infectious process of individual virulence factors depends on the pneumococcal strain. In other words, the importance of each virulence factor in pathogenesis and transmission will depend on the level of its synthesis and its structure, plus it will depend on the total gene repertoire of the pneumococcal strain and the overall pattern of gene expression.

3. The importance of individual pneumococcal virulence factors and the response of the host to these virulence factors will depend on attributes of the host, particularly the innate susceptibility of the host and the site of infection.

4. The details of key physiological processes, for example competence, biofilm formation, central carbon metabolism and energetics, will vary between pneumococcal strains. Differences in these physiological processes may influence *in vivo* behaviour.

It was intended that the scientific and technological objectives of the project would be achieved by testing these hypotheses. Compilation and computational analysis of the multidisciplinary experimental data from the studied collection of pneumococcal strains and disease models will identify the most important and consistently involved host and pneumococcal factors. To test the hypotheses, a range of techniques in molecular and systems biology was used, including determination of transcriptomes, proteomes and metabolomes and state of the art bioinformatics and data management tools. Three *in vivo* models of invasive disease or asymptomatic colonisation were used, as well as four *in vitro* tissue models. The role of host factors was explored by identification of genes conferring innate resistance or susceptibility to pneumococcal disease in humans and mice deficient in selected aspects of innate immunity. The objective was that through comparative data analysis and cross-referencing of the data from the individual work packages, a comprehensive dissection of pneumococcal-host interactions would be provided. This knowledge was to enable the consortium to deliver new knowledge for diagnosis of disease, diagnosis of risk, vaccination and treatment of pneumococcal diseases.

3. Scientific and Technical Results / Foregrounds

Workpackage 1 – Molecular Epidemiology and Genetic Characterisation of Pneumococcal Isolates

Workpackage Objectives

Well-characterised, antibiotic susceptible and resistant, pneumococcal isolates, as well as the laboratory strain, TIGR4, were provided for use in other WPs. The genetic content and genetic relationships of clinical pneumococcal isolates were determined with molecular

epidemiological tools, microarray technology and sequencing. This comparative genomic study was done with pneumococcal isolates from carriage and disease. Based on the resultant data, the expression of selected genes studied. The genomes of the pneumococcal isolates were compared with those of *Streptococcus mitis* and *Streptococcus oralis*.

Progress Towards Objectives

All objectives were attained

Provision of strains and genome sequences. Nine pneumococcal strains were selected for use within PNEUMOPATH. These were distributed to the partners. The serotypes and clonal types were selected to be representatives of important clonal lineages in invasive disease and in carriage. Three strains of serotypes 3, 4 and 14 were previously sequenced and the sequences are available on the Internet. Two were chosen from an epidemiological study where invasive disease isolates and carriage isolates were collected from children in the same geographic area (Stockholm) during the same time period (see below). The invasive disease potential (IDP) was calculated by Partner 4 and it was found that the there were differences in IDP not only between serotype, but also between clonal types and even within clonal types, as determined using MLST. Two serotype 6B isolates representative of two pneumococcal clones by PFGE, but the same by MLST, and belonging to the same serotype but with different invasive disease potential were included to be further characterised in PNEUMOPATH. These two strains were fully sequenced and genomic analysis was performed (see below). Partner 8 delivered the two 19F serotype strains. The strains are of ST179 and were isolated in Portugal from healthy carriers in 1998 and 2007. While the first isolate is fully susceptible to penicillin, the latter one is resistant. Both belong to a clone that appears to have a superior capacity to colonise and transmit and be maintained in the population even with widespread use of pneumococcal conjugate vaccines as described in two previous studies conducted by this partner. Genome sequencing of these two strains was performed as planned. Genome assembling and comparison was done.

In collaboration with WP6, genomic sequences of the sequenced strains were assembled and put on the project webpage so that all partners within the project can access the data. A separate meeting was held in Lisbon in June 2010 to discuss these issues.

Development of a pneumococcal microarray chip from genome sequences

To be able to perform various microarray experiments with flexible and custom design, Partner 3 installed the Nimblegen platform. Nimblegen has considerable experience with prokaryotic array design, and their experience in prokaryotic genomics and the flexibility of their array formats was the deciding factors in this choice. The number of features on Nimblegen arrays can be either 72,000 (a “4x72k” array format; 5 µm) or, on their latest high-density arrays, 135,000 (a “12x135k” format, 2 µm). The features consist of 60-mer oligonucleotide probes, with the possibility of separate design for specific applications such as expression analysis, comparative genomic hybridisation, and genomic array footprinting (GAF). The design itself was performed in collaboration with the Nimblegen design team.

Several expression designs have been made during the project. For a single genome expression array (e.g., for *S. pneumoniae* TIGR4), on average 7-8 probes were designed for each gene, each spotted 3-4 times (4x72k format). For a multiple-genome array, ideally the pneumococcal core genome (i.e. genes common to all strains) is represented by one reference (TIGR4), after which probes for strain-specific transcripts are added for each additional strain. Initial attempts to accomplish this were undertaken using six strains. To this end, the Nimblegen design team were supplied with either the publicly available genome (in case of TIGR4, D39, and Sp14-BS69) or a list of transcripts for three strains sequenced within the consortium (BHN100, BHN191, BHN418). Probes were designed for all transcripts, after which redundant probe sets (i.e., matching more than one transcript) were removed.

Unfortunately, this resulted in a higher number of “strain-specific” transcripts (>6,000) than expected based on the known pneumococcal core genome. As a result, this “pan-genomic” array covered 8,325 transcripts, each represented by five probes spotted in triplicate (126k probes), as such already reaching the limits of the 12x135k format for six of the ten strains. Therefore, the following alternative approach was explored. Together with Partners 4, 8 and 11, annotation and blast analysis of all PNEUMOPATH strains (publicly available or sequenced within the consortium) was performed, using the TIGR4 strain as reference. This resulted in a more condensed list of strain-specific transcripts for each of the nine additional strains. This list, together with the complete TIGR4 gene list, was then used for a new pangenomic array design that has been used within the consortium. The Nimblegen system has been installed and extensively tested by Partner 3. To optimise standardisation of microarray-based experiments, this platform was accessible to all PNEUMOPATH partners who decided to run and analyse their microarray experiments with Partner 3.

Characterisation of the genetic content and genetic relationships of pneumococcal isolates

Partners 4 and 14 have characterised 715 invasive and carriage isolates from children, from the same region (Stockholm) and time period, using pulsed-field-gel-electrophoresis (PFGE) and multi-locus-sequence-typing (MLST). Different serotypes were associated with different invasive disease potential (IDP), with serotypes not covered by the 13-valent pneumococcal vaccine (PCV13) collectively showing a low IDP. Examples of clonal types of the same serotype with different IDPs were found. Serotype 6B isolates of common clonal complex (CC) 138 could be divided into several PFGE patterns, partly explained by number, location and type of temperate bacteriophages. One PFGE pattern was associated with a high IDP. Four isolates representing different invasive disease potential were included in the strain collection of PNEUMOPATH and these strains were further characterised with whole genome sequencing. These four CC138 isolates represented PFGE clones with different IDP and revealed intra-clonal variations in the sequence of virulence associated proteins PspA, and PspC, and in the presence / absence of PcpA and in the presence of phages. Interestingly differences were found with regard to phage content and an in depth analysis of functional (mutanolysin assay) and non-functional (by PCR) presence of phages in other pneumococcal strains belonging to the same clone as the 6B isolates that were sequenced were performed to study if there are associations between phage content and disease.

Partner 7 made an analysis of members of a 23F clone. This was done by comparative genomic hybridisation using a *S. pneumoniae* R6/TIGR4-specific oligonucleotide microarray, and genomic sequence data obtained from six of these isolates. The numbers indicate a high degree of conservation even between isolates from different continents, and those isolated over 15 years apart.

The microarray chip, described above, was used for the identification of genes associated with pneumococcal invasive disease potential. Strains used in this study were selected from a well-characterised collection of carriage and invasive isolates belonging to serotypes and clones with different invasive disease potential. The identification of genes was done by comparative genomic hybridisation (CGH). The study was designed and prepared by Partner 8. A student from Partner 8 visited Partner 3 for three months, to perform the CHG experiments. The pneumococcal collection used in this study was selected from an initial collection of 1,244 strains isolated in Portugal between 2001 and 2003 from either carriage (n=475) or invasive disease (n=769). This large collection had previously been characterised by molecular methods, grouped into clones, and the invasive disease potential of the clones determined (Sá-Leão *et al.* 2011. J Clin Microbiol 49:1369-75). This knowledge enabled the selection of a group of representative strains in terms of invasive disease potential at the same time including most of the serotypes that are more frequently found colonising or causing invasive disease. A total of 175 strains were selected, 74 from invasive disease and 101 from carriage.

To improve understanding of pneumococcal genetic relationships, Partner 8 conducted four other studies that are summarised below:

1. The impact of the 7-valent pneumococcal conjugate vaccine (PCV7) on the pneumococcal flora was mostly studied without evaluating multiple colonisation and the mechanism(s) leading to serotype replacement. The effect of a single dose of PCV7 on colonisation was studied. A group of children received one PCV7 dose just after nasopharyngeal sampling, with the control receiving no vaccine and both groups being sampled again a month later. Up to ten pneumococcal colonies were recovered per colonised child — 1,224 isolates were serotyped and representative ones were analysed by pulsed-field gel electrophoresis. In vaccinated children, serotype replacement between vaccine (VT) and non-vaccine (NVT) types occurred in single and multiple carriers, and VTs were less prone to be *de novo* acquired. NVT unmasking was only detected in the vaccinated group. It was concluded that one month after vaccination with a single dose, PCV7 prevents VT *de novo* acquisition and promotes NVT unmasking.
2. A collection of atypical pneumococci isolated from healthy children was characterised. While performing surveillance studies in Oeiras, Portugal, aimed to describe the impact of pneumococcal conjugate vaccine on colonisation, an increase from 0.7% in 2003 to 5% in 2006 was observed in the prevalence of penicillin resistance (MIC 2-6 mg/L) among presumptively identified pneumococci. Although 15 of the 22 penicillin-resistant isolates obtained in 2006 were optochin-resistant, they were bile soluble and thus considered to be bona fide pneumococci. Clarification on the nature of these isolates was obtained by using a combination of phenotypic and genotypic approaches that included routine strategies for pneumococcal identification, multilocus sequence analysis (MLSA), and comparative genomic hybridisation (CGH). CGH analysis was done in collaboration with Partner 3. By MLSA all isolates were classified as “streptococci of the mitis group” that, however, were distinct from typical *S. pneumoniae* or *S. mitis*. A single isolate was identified as *S. pseudopneumoniae*. CGH confirmed these findings and further indicated that a considerable part of the proposed pneumococcal core genome is conserved in these isolates, including several pneumococcal virulence genes (e.g. *pavA*, *spxB*, *cbpE*, and *cbpD*). These results suggested that among pneumococci and closely related streptococci, universal unique phenotypic and genetic properties that could aid on species identification are virtually impossible to establish. In pneumococcal colonisation studies, when atypical strains are found, MLSA and CGH are informative tools that can be used to complement routine tests (when full genome sequencing is not an alternative). These findings are in line with those by Partner 7 described below. In the study, after correct identification of the penicillin-resistant true pneumococci, it was found that penicillin resistance levels among pneumococci remained stable from 2003 through 2006.
3. As non-capsulated pneumococci (also referred to as non-typeable pneumococci or NTPn) are difficult to identify, as their differentiation from closely related species such as *Streptococcus pseudopneumoniae* and other streptococcus of the mitis group is not always straight forward, a low cost and easy assay was developed to detect and quantify NTPn in primary samples (which may contain multiple species) obtained from nasopharyngeal swabs. The strategy was based on a multiplex PCR targeting *lytA*, *cpsA*, *aliB*-like ORF2 and 16S rDNA genes, plus a RFLP assay to differentiate typical from atypical *lytA.* The application of the proposed methodology to over 500 nasopharyngeal samples found that the prevalence of NTPn in colonisation was three-fold higher than estimated by routine methods (from 2.9% to 8.6% in the study collection). The international clone Norway^{NT}ST344 was the major clone identified.
4. The introduction of the seven-valent pneumococcal conjugate vaccine (PCV7) in Portugal led to extensive serotype replacement among carriers of pneumococci, with a marked decrease of PCV7-types. Although antimicrobial resistance was traditionally

associated with PCV7-types, no significant changes were observed in the rates of non-susceptibility to penicillin, resistance to macrolides or multidrug resistance. Investigation of the mechanisms leading to maintenance of antimicrobial resistance despite marked serotype replacement was undertaken. Through molecular typing, 252 antibiotic resistant pneumococci recovered from young carriers in 2006 and 2007 (era of high-PCV7 uptake) were compared with collections of isolates from 2002-2003 (n=374, low-PCV7 uptake era) and 1996-2001 (n=805, pre-PCV7 era). It was observed that the group of clones that account for antimicrobial resistance since 1996 is essentially the same as the one identified in the PCV7 era. The relative proportions of such clones have, however, evolved substantially overtime. Notably, widespread use of PCV7 led to an expansion of two PMEN clones expressing non-PCV7 capsular variants of the original strains: Sweden^{15A}ST63 (serotypes 15A and 19A) and Denmark^{14}ST230 (serotypes 19A and 24F). These variants were already in circulation in the pre-PCV7 era, although they have now become increasingly abundant. Emergence of novel clones and *de novo* acquisition of resistance contributed little to the observed scenario. No evidence of capsular switch events occurring after PCV7 introduction was found. In the era of PCVs antimicrobial resistance remains a problem among the carried pneumococci. Continuous surveillance is warranted to evaluate serotype and clonal shifts leading to maintenance of antimicrobial resistance.

In Sweden, Partner 4 performed a national surveillance of invasive pneumococcal infections before and after the introduction of the conjugated pneumococcal vaccine in the childhood vaccination program in 2009. All invasive isolates from the clinical microbiological laboratories in Sweden were sent for serotyping and selected have also been characterized using molecular typing, PFGE and MLST. Clinical information about the patients has also been collected through responsible physician. A decreased incidence of invasive pneumococcal disease in Sweden was observed as well as an increase of non-vaccine types for example of serotype 22F. All data are in the process of being analysed and two manuscripts are being written, one on the situation in the largest area, Stockholm, and one comprising all data on a national level.

Progress on comparison of pneumococcal and *S. mitis* and *S. oralis* genomes

Comparison of *S. mitis* / *S. oralis,* using an *S. mitis* B6 specific microarray, was performed. For comparative genomic hybridisation, DNA isolated from 10 *S. mitis* and 10 *S. oralis* strains were used. Mobile elements and RNA-coding genes were not considered in this analysis, leaving 1684 features to be considered. The positive signals obtained with the *S. mitis* DNAs ranged between 86 - 75 %, whereas *S. oralis* DNAs hybridised with 68 - 37 % of the *S. mitis* B6 specific oligonucleotides; *S. pneumoniae* R6 DNA which was included as a pneumococcal reference hybridised with 72 %. Whereas 58 % of these genes hybridised with DNA from all *S. mitis* strains, and 27 % hybridised with all *S. oralis*; 23 % of all features hybridised with both, all *S. mitis* and *S. oralis* DNAs.

In addition to the *S. mitis* B6 genome, the first *S. oralis* Uo5 genome was finished during the funding period. On the basis of the *S. mitis* B6 genome, an oligonucleotide microarray was designed, which in addition to the *S. pneumoniae* R6/TIGR specific microarray, is now available for comparative genomic hybridisation (CGH). The major task was the analysis of the core and accessory genome that is common between these three species, both using the microarray and the genome data for *in silico* analysis.

The CGH data confirmed that all three *Streptococcus* spp. share much of the accessory genome, which causes a 'smooth transition' between these species, as had been proposed previously. Moreover, most genes known as virulence factors in *S. pneumoniae* are also commonly found in *S. oralis* and *S. mitis*, leaving only a few loci, including the capsule

cluster, the hyaluronidase, and some surface proteins as *S. pneumoniae* specific virulence genes. Moreover, large clusters of the accessory genome, such as the large transposable element that includes the tetM determinant, a gene cluster encoding a large Ser-rich LPXTG surface protein (monX), and a cluster that includes a V-type ATPase occur in *S. pneumoniae*, and in the other species as well.

Specific to *S. oralis* was a cluster that encodes a novel export system not known previously in *S. pneumoniae*, but which has been discussed as virulence factors in other bacteria. The genetic and biochemical analysis of this cluster and its elements has been characterised, suggesting a complex regulatory mechanism of this machinery (Becker *et al*, manuscript in preparation).

Workpackage 2 – Host Susceptibility

Workpackage Objectives

- The overall objective was to identify new host loci conferring innate resistance (or susceptibility) to invasive pneumococcal disease.
- To use the diverse susceptibility of the different strains of mice to pneumonia to map susceptibility QTL (quantitative trait locus) by *in silico* methods.
- Cross comparison of the mouse pneumococcal susceptibility or resistance QTLs with human pneumococcal susceptibility or resistance loci identified through linkage and/or genome-wide association analysis of human sample sets from invasive pneumococcal disease (IPD).
- Cross species validation of genetic markers, either identified through mouse QTL analysis or linkage and association analysis in the IPD sample set, for identification of novel sequence variants that predispose to pneumococcal susceptibility or resistance.
- Identification of new human pneumococcal susceptibility or resistance genes by analysis and cross species comparison of genome-wide gene expression data derived from blood and adipose tissue of 1000 human individuals and lung tissue of different mouse strains.
- Identification of genetic variation in pathways contributing to susceptibility or resistance from IPD, using tagging SNPs across genes within candidate pathways.

Progress Towards Objectives

All objectives were attained

The susceptibility to infection with *S. pneumoniae* is associated with host genetic background. It is well described for humans, as well as in mice, that some individuals are less likely to develop invasive pneumococcal disease, even after confirmed exposure. In murine models there are well-described strains displaying permissive or resistant phenotypes. Despite these observations, the genetic factors conferring resistance to pneumococcal infection remain unknown. Recent advances in genetics of humans and mice and increasing availability of dense genotyping data offer a new opportunity to dissect the genetics behind complex diseases. In the PNEUMOPATH project the genetics underlying susceptibility to pneumonia in mouse and humans was investigated.

A panel of 26 genetically defined mouse strains from Jackson Laboratories (Jax) was phenotyped, by Partner 1, for their susceptibility to invasive pneumococcal disease after intranasal infection with serotype 2 pneumococcal strain D39. The phenotyping data were then used to look for genetic association with observed clinical outcomes. The data confirmed that susceptibility to pneumonia strongly depends on host genetic background, as

tested mouse strains displayed different degree of susceptibility to the infection. There were strains that survived the challenge, had no bacteria in any of the tested tissues and only mild to moderate signs of disease were recorded. Other mouse strains had bacteria in the blood and in tissues but either cleared the infection or established carriage. For some strains the survival rate was around 50% and strains with severe clinical signs, high bacteria count in tissue and very short survival time.

Recently the density of SNP coverage of mouse genomes increased significantly and with this increase far greater sensitivity is obtained in genome-wide association studies. From the panel of 26 Jax-mouse strains, three significant QTLs were identified which strongly associated with resistance to pneumococcal disease. One of these QTLs was located within the Spir1 region previously identified by Denny *at al* (2003) as conferring resistance to pneumococcal infection in independent linkage studies using CBA/Ca x BALB/c crosses. Within the loci identified in WP2 55, potential candidate genes were located. The combined analysis of GWAS and gene expression data in lungs at 6h post-infection selected 14 genes with expression changed during D39 infection in resistant (BALB/c) and/or susceptible (CBA/Ca) mice. Among those 14 genes, five were up-regulated at least two-fold as compared to control animals. Some candidate genes identified have described polymorphism associated with pneumonia in humans, namely *Fas* and *Nfkbib*.

The strongest gene candidates were *Fas* and *Ch25h*, which play important roles in lung injury and macrophage function. *Fas* is expressed by neutrophils, macrophages, dendritic and epithelial cells. The myeloid-cells respond to Fas activation primarily by secretion of proinflammatory cytokines while epithelial cells undergo apoptosis. It was also reported that Fas can control Cxcl1 release by lung epithelium in MyD88-dependent manner. During resolution of lung inflammation Fas selectively targets recruited macrophages for apoptosis. The Ch25h is involved in oxysterol 25HC productions. Oxysterols are short-lived derivates of cholesterol. Besides their role in lipid metabolism they have important function in macrophages function. Oxysterols bind to Liver X nuclear receptors (LXR) expressed by alveolar macrophages, epithelial class II cells, dendritic cells and neutrophils. Ch25h has been associated with increased macrophages survival and inhibition of lymphocytes T and B proliferation.

A polymorphism within the mouse *Fas* gene was also identified in this study. SNPs that separated resistant mouse strains from susceptible ones were located within *Fas* coding and regulatory sequence. One of those SNPs resulted in an amino acid change. A similar polymorphism was observed upstream of the *Ch25h* coding sequence. It was previously reported that located upstream of the *Ch25h* coding sequence were consensus IFN-γ activation-site elements and INF-stimulated response elements.

Among genes located within chromosome 7 QTLs, the most interesting gene candidates, with changed expression during infection, were *Nfkbib*, *Map4k1* and *Kcnk6*.

The results of the analysis allowed selection of a short list of candidate genes that may play an important role in resistance to pneumococcal disease. The claim was supported by results of other studies in mouse and humans that pointed to those proteins or loci as associated with pneumonia in humans (*Fas*, *Nfkbib*) or mice (Spir1). The candidate genes selected also have strong links to immune responses in lungs (*Fas*, *Ch25h*).

Partner 5 used bioinformatic tools to map twelve genes identified as associating with resistance to pneumococcal infection in the mouse. The coding region of each gene and 1 MB on either site was used to look for association of sequence variants (SNPs and Indels) existing in the database dbSNP, with invasive pneumococcal disease (IPD) and/or invasive disease caused by encapsulated bacteria (*S. pneumoniae, N. meningitidis* or *H. influenzae*) (IBD_encaps) in humans. The associations in the mouse were not replicated in humans, as no sequence variant in the genes or within 1 MB flanking regions of the genes in these loci showed genome wide significant or suggestive association with IPD or IBD_encaps in

humans. No genome wide significant or suggestive association of non-synonymous variants (causing missense mutations, or amino acid change) with IPD or IBD_encaps was found. Also there was no association of loss of function variants (splice, stopgain, frameshift) with IPD or IBD_encaps found in humans.

Partner 5 also mapped 146 genes in the ten QTL loci that showed significant or suggestive association with susceptibility to pneumococcal pneumonia in the mouse and identified the location of each gene plus/minus 1MB. Assessment was made of whether sequence variants in dbSNP associated with IPD or IBD_encaps in humans. No sequence variant showed genome wide significant or suggestive association with IPD or IBD_encaps in humans. Two very rare variants associated with IBD_encaps. Both are in the Dcc locus on chr 18, but have very low functional impact as one is downstream of the gene and the other is intergenic. The only missense variant showing some association with IPD in humans was in the *ZNF740* gene, encoding a zinc finger protein, and corresponding to the *Zfp808* on chr 13 in the mouse. Thus, association of the QTLs associating with susceptibility to pneumonia in the mouse was not replicated in humans.

Workpackage 3 – Animal Models

Workpackage Objectives

- Comparative virulence testing in inbred mouse strains permissive to pneumococcal infection.
- Comparative virulence testing in inbred mouse strains resistant to pneumococcal infection.
- Definition of host inflammatory responses to infection with different pneumococcal strains.
- Definition of host genetic responses to infection with different pneumococcal strains.
- Definition of genetic responses of different pneumococcal strains in innately resistant and susceptible mice and in different tissues in these mice.
- Definition of active and passive protection models for pneumococcal infection
- Evaluation in an appropriate animal model of the hypotheses developed from use of *ex vivo* and *in vitro* models in WP4 and WP5.

Progress Towards Objectives

All objectives were attained

Comparative virulence testing in inbred mouse strains either permissive or resistant to pneumococcal infection.

The susceptibility to pneumococcal disease is influenced by host genetic background. Numerous studies had showed that some individuals are able to withstand the infection while others show severe clinical signs or even mortality. The host responses to infection were studied with nine pneumococcal strains belonging to five different serotypes and six sequence types. Two mouse strains, resistant (BALB/c) and sensitive (CBA/Ca) to infection with serotype 2 D39 strain were chosen for this study. After infection with nine pneumococcal strains BALB/c mice showed resistance to 7 out of 9 pneumococcal strains. Two of the seven non-virulent strains (both 6B serotypes) were able to enter the blood stream of BALB/c animals but were later cleared. Only two strains: TIGR4 (ST4) and BS71 (ST3) were virulent in BALB/c with rapid onset, high disease severity and the experiment endpoint reached within 28h. Surprisingly this strain had very poor dissemination to the blood and the majority of the bacterial load was contained in the lungs. CBA/Ca animals showed higher susceptibility as compared to BALB/c mice. Five out of nine pneumococcal strains tested

(BS71, D39, TIGR4 and both 6B strains: BHN191 and BHN418) caused 70%-100% mortality in CBA/Ca mice. Therefore three pneumococcal strains (serotype 2 and 6Bs) showed different disease phenotype depending on the host genetic background. Further difference between BALB/c and CBA/Ca mice could be seen after infection with LgtSt215 (19f) strain displaying non-virulent phenotype (100% survival in both mouse strains). LgtSt215 established successful carriage in the lungs of CBA/Ca but not BALB/c mice. In conclusion we were able to divide tested pneumococcal strains into three groups based on their virulence type: non-virulent (BHN100, CBR206, LgtSt215 and BS69), host-dependent virulence (D39, BHN191 and BHN418) and highly virulent (TIGR4 and BS71). It is worth noticing that all tested serotype 19F pneumococci were placed in the non-virulent group (together with BS69 serotype 14).

Intranasal challenge of C57/BL6 mice was performed using ten (plus additionally 2) strains of the PNEUMOPATH project. The comparative intranasal virulence assays revealed statistically significant differences in the pathogenicity between the bacterial strains. There was a significant difference in the overall survival and the bacterial burden in the blood stream. Strains of serotype 14 (one strain), 19F (three strains) were less virulent than strains of type 4 (one strain), 6B (four strains) and 3 (one strain). Two variants of the same strain D39-L and D-39-S showed a big difference in virulence where D39-L was much more virulent than D39-S. Sequence analyses within the project have identified genomic differences between these two strains. Furthermore the data suggest that among the isolates assayed there may be two ways to cause disease. In a set of strains with a similar time of onset of disease some bacterial strains (D39-L, BHN-191, BHN-418) seemed to be more prone to go to the blood stream (bacteraemia) while Sp3-BS71 was primarily found in the lungs and found in the blood only at an advanced stage of the disease. The latter strain was also consistently found in the lungs of asymptomatic mice at the experimental endpoint.

<u>Definition of host inflammatory and genetic responses to infection with different pneumococcal strains.</u>

Pneumococcal infection causes strong immune and inflammatory responses with large number of recruited neutrophils being the hallmark of the infection. In PNEUMOPATH project we investigated host genetic and inflammatory responses to infection with strains from different pneumococcal serotypes. The results showed a complex picture of the interplay between the host and the pathogen. Host whole lungs gene expression, as early as 6h post infection, showed surprisingly uniformed picture with very similar gene regulation in both mouse strains. Unsurprisingly the gene expression strongly reflected lung inflammatory responses when the highest regulated genes were directly involved in pathogen recognition (CD14 among top 5 up-regulated genes) neutrophil recruitment (e.g.: Cxcl1 was the highest up-regulated gene), cytokines and interleukins production (IL4i1, IL1b, CCl4, CCl3 and CCl17 among the top 15 up-regulated genes), regulation of cytokine production (Socs3, among top 10 regulated genes) or apoptosis (up-regulation of Fas and Bcl3). This pattern of gene expression was observed for D39-, BHN191- and LgtSt215-infected BALB/c and CBA/Ca mice but not for BS71-infected mice. The BS71 strain failed to efficiently induce gene expression, with only a couple of genes showed change in expression level as compared to sham-infected control animals. A similar picture was seen when host cytokine level was measured (IL1β, IL6, IL10, IL17A, INFγ and TNFα). At 6h time point, BS71-infected mice had all cytokines unchanged (with the exception of IL1β in CBA/Ca), however 24h post infection IL1β, IL6 and TNFα reached the highest level as compared to the mice infected with other pneumococcal strains. BS71-infected BALB/c mice also had IL10 and INFγ unchanged while infection with other strains usually caused further decrease in these cytokines at 24h.

Variation in cytokine levels was observed between animals infected with different pneumococcal strains. However the impact of these differences on animal survival was complex. Only combined analysis of all tested cytokines revealed correlation with observed clinical outcome of the infection. To summarise, our results indicated that changes in IL6,

IL10, IL17, INFγ and TNFα but not IL1β significantly correlated with animal survival. IL6 and IL10 seemed to have opposite effect since higher level of IL6 and lower of IL10 at 6h time point and decrease in IL6 with increase in IL10 at 24h correlated with better survival. Decrease in INFγ and TNFα at 24h also seemed beneficial. The Pearson R^2 in our model was moderate (r2=0.6, p<0.00001) and perhaps could be improved if more cytokine were included or different set of cytokines.

A very interesting aspect of this study was the unique phenotype of the BS71 serotype 3 strain. Its hypervirulence and its simultaneous failure to induce host responses at an early time post-infection suggests that either this strain adopted some unique hiding strategy or was able to block host signalling pathways. It was interesting to check whether this phenomenon is serotype specific. To do so BS71 was compared with another serotype 3 strain of the same sequence type: BHN35. Surprisingly BHN35 differed significantly from strain BS71. Survival of BS71-infected animals was extremely short (below 28h) despite a very poor dissemination of the bacteria into blood. In contrast, BHN35-infected animals survived longer (mean 104h) and developed high levels of bacteremia. The BHN35 strain gave rise to the typical pathology of tissue damage, enhanced cytokine production and inflammatory cell influx but despite its extremely high virulence BS71 did not damage epithelial tissue in an *in vitro* assay and failed to activate the usual host inflammatory responses. This study suggests that strains of similar serotype are capable of activating distinct disease pathways, which reinforces the opinion that factors additional to capsule determine the pathology of pneumococcal disease.

Gender differences in susceptibility to pneumococcal disease have been noticed before. In humans this results in a slight predominance in invasive disease in male with respect to females. In PNEUMOPATH, it was demonstrated that this occurs in rodents. The influence of gender on the ability to control systemic infection with *S. pneumoniae* was investigated in male and female mice, in both pneumonia and sepsis. Males were found to be more susceptible to invasive infection. In sepsis, male mice showed greater weight loss, more marked decrease of body temperature and a significantly higher mortality rate. In pneumonia, there were significant differences in survival, with females able to clear their infection over time.

Using a Bioplex suspension array, the cytokine and chemokine profile was determined during infection. Overall, pneumonia was characterised by a more pronounced increase in cytokines indicative of neutrophil and macrophage involvement, while in sepsis Th1 related cytokines were up-regulated. In both experimental infections male mice exhibited significant increases in several pro-inflammatory cytokines during both sepsis and pneumonia compared to female mice. Increases in IL-6, IL-12(p70), IL-17A, IFNγ, RANTES, KC and G-CSF were seen during sepsis, while increases in IL-1β, IL-15, IL-17A, IL-18, TNFα, RANTES, MIP2, KC and GM-CSF during pneumonia. The data obtained indicate a detailed cytokine profile for pneumococcal pneumonia and sepsis in mice with a surprising difference in cytokines, clinical signs, and survival between male and female mice. The data were published in Kadioglu *et al* (2011).

Definition of genetic responses of different pneumococcal strains in innately resistant and susceptible mice and in different tissues in these mice**.**

The capacity to clear bacteria in the initial minutes and hours after invasion of the bloodstream has been found to correlate with clearance of the infection. Susceptible and resistant mice have been shown to have drastically different capacities to carry out these early innate events. The early events have been in depth characterised in resistant mice. Analysis of macrophage-depleted mice clearly indicated that macrophage phagocytosis is the main innate mechanism responsible for the initial clearance of pneumococci after intravenous challenge. In order to identify the cellular mechanism and define appropriate *in vitro* or *ex vivo* assays to investigate this a methodology was developed to efficiently obtain primary splenic macrophage cells. In FACS analysis these cells presented the markers

characteristic of marginal zone macrophages, both when cultured from susceptible and resistant mice. When running phagocytosis assays with these cells, surprisingly both were equally efficient in ingesting non-virulent pneumococci and also showed equally low efficiency in ingestion of virulent pneumococci. Thus, despite the significant difference observed in macrophage phagocytosis in susceptible and resistant mice, once cultured *in vitro*, the relative primary cells did not show this phenotype. These data highlight the complexity of the innate early response to pneumococci.

Definition of active and passive protection models for pneumococcal infection.

The early events in the host response to pneumococcal infection were investigated the by following the fate of an intravenous inoculum of pneumococci comprised of roughly equal numbers of three isogenic variants. Sequential analysis of blood samples over the first hours and days indicated that most episodes of invasive infection were monoclonal events. This finding, based on mathematical and statistical analysis, can only be reasonably explained if there is a single bacterial cell bottleneck. This allowed definition of a restricted timeframe when host clearance mechanisms removed all but one organism of the original challenge inoculum. It was shown that macrophages are the major cellular mediators of early clearance of pneumococci, with neutrophils having importance at later stages. However, selection of phenotypic variants, for example phase switching of capsule expression, also plays a role in purging the population of less fit pneumococci. The extreme bottleneck demonstrated in these and other experimental infections indicates the complexity of the pathogenesis of invasive bacterial infections. Understanding the mechanisms of the clonal origin of systemic infection is relevant to undertaking research on virulence, for example using high-throughput genetic screens, or in considering host and microbial factors that are relevant to selection of vaccine escape variants following immunisation.

Evaluation in an appropriate animal model of the hypotheses developed from use of *ex vivo* and *in vitro* models in WP4 and WP5.

Choline (Cho) is an essential nutrient for growth and virulence of the pneumococcus. One of the main operons involved in Cho metabolism is the essential *lic1* operon, which was unambiguously established to be induced in conditions of Cho deprivation *in vitro* in WP5. To test whether this Cho-based regulation was important *in vivo*, a mutant in the Cho-regulated promoter, unable to respond to Cho deprivation, was compared with its wildtype parent in a mouse model of nasal colonization and virulence. The ability to respond to Cho deprivation by inducing expression of *lic1* was found to be important for nasal colonisation, as the mutant was out-competed by the wild-type strain in a co-infection setup (see also WP5 section).

Neuraminidases are upon the best known pneumococcal virulence factors. We have set up carriage model in different mouse strains and confirmed that intranasal administration of sialic acid increases pneumococcal carriage by a factor of 10 to 1000 depending on the mouse strain. Neuraminidase inhibitors are drugs licensed for the treatment of influenza infection. Here we confirm that neuraminidase inhibitors can be efficiently tested in different mouse strains to evaluate their efficacy for reduction of pneumococcal carriage. The carriage model thus has been confirmed to be robust and suitable to assay for the *in vivo* effect of carbohydrates on pneumococcal carriage and to be suitable also for the assay of the efficacy of drugs based to be used in treatment of pneumococcal carriage.

Workpackage 4 – Cell Culture Models

Workpackage Objectives

The aims of the workpackage were to explore, in a standardised and inter-laboratory fashion, various *in vitro* cell culture models.

- Identification of the species- and strain-specific characteristics of the pneumococcal isolates with respect to adherence, invasion, transcytosis, cytotoxicity, uptake/killing, cytotoxicity and chemotaxis.
- Identification and characterisation of the species- and strain-specific pneumococcal genes involved in adherence, invasion, transcytosis, cytotoxicity, uptake, cytotoxicity and chemotaxis.
- Characterisation of the species- and strain-specific molecular response of pneumococci during adherence, invasion, transcytosis, cytotoxicity, uptake/killing, cytotoxicity and chemotaxis.
- Characterisation of the molecular response of the host during pneumococcal adherence, invasion, transcytosis, cytotoxicity, uptake/killing, cytotoxicity and chemotaxis.
- Characterisation of the pneumococcus-specific pathways involved in adherence, invasion, transcytosis, cytotoxicity and chemotaxis.

Progress Towards Objectives

All objectives were attained

Phenotypic characterization of pneumococcal isolates in various *in vitro* cell culture models.

Adherence to epithelial cells. To ensure standardisation between laboratories, inocula used for all *in vitro* assays were prepared in the same manner by all participants in this work package. Kinetics of adhesion and invasion of epithelial cells by the ten PNEUMOPATH strains was studied using two human cell lines: pharyngeal epithelial Detroit 562 and lung epithelial A549 cells. Clear adherence kinetics were observed using Detroit 562 cells, with highest levels of adherence reached after 4h for all strains. Furthermore, strain-specific adherence phenotypes were identified, but no clear correlation between adherence ability and strain origin (carriage or disease). Highest adherence levels were obtained with the non-encapsulated R6 strain, while of the encapsulated strains, the serotype 6B strains showed the greatest capacity to adhere, and the thickly encapsulated serotype 3 strain (Sp3-BS71) the lowest. Adherence levels of an additional serotype 3 strain, BHN35, were comparable to those of Sp3-BS71. Adherence kinetics of pneumococcal strains using A549 cells were similar to Detroit cells. Invasion of both epithelial cell lines by pneumococcal strains was limited, with only minor differences between strains. Given the known variation across capsular serotypes in their ability to adhere to and colonise epithelial surfaces, the observed phenotypic differences between the strains could, at least partly, be due to their difference in serotypes. To examine this in more detail, capsular switch mutants were constructed, i.e., strains that are isogenic except for the capsular type produced (Trzcisńki *et al.*, 2003, Appl. Environ. Microbiol, 69:7364-7370), in which the serotype 4 capsule locus (*cps*) was removed and / or replaced by serotype 2, 6B, 14, and 19F *cps* loci. Adherence kinetics of these switch mutants clearly showed a similar adherence pattern between serotypes as obtained with the wild-type strains. Interestingly, the presence of a foreign *cps* locus did appear to confer some (energetic) cost, as adherence levels did not increase much after 1-2h.

Adherence, phagocytosis and killing by immune cells. To analyse the interaction between the ten PNEUMOPATH strains and antigen presenting cells, a standard bacterial uptake / killing assay was used with different immune cells. First, human monocytes isolated from buffy coats were differentiated into macrophages for ten days using three different growth factors: Macrophage colony-stimulating factor (M-CSF), Granulocyte / macrophage-CSF (GM-CSF), and Interleukin 3 (IL-3). Results obtained showed enormous variability between different macrophage donors, as well as between different strains. All strains were able to bind to the macrophages, most prominently after differentiation with GM-CSF and IL3. All strains were phagocytosed and showed some survival, with only very low levels for D39, Sp3-BS71, and LgSt215. Second, PMA-differentiated THP-1 cells (a human monocyte leukemia cell line) were used. Low variability during the assay and detectable differences between strains were

observed, suggesting a higher reproducibility for differentiated THP-1 cells. Finally, bacterial uptake/killing assay using Raw 264.7 mouse macrophage cells were performed. Statistically significant differences in binding capability, phagocytosis and killing susceptibility of the ten PNEUMOPATH strains was observed. In particular, two strains stood out against all: Sp14-BS69, which exhibited a significant higher number of cell-associated, intracellular and killed CFU compared to most strains, and Sp3-BS71, which exhibited a significant lower number of cell-associated CFU and no intracellular CFU. In addition to Sp3-BS71, four other serotype 3 strains (with the same and different sequence types) displayed similar phenotypes as Sp3-BS71 with Raw 264.7 mouse macrophage cells: a significant lower number of cell-associated CFU compared to the serotype 2 D39-L strain, and no intracellular CFU compared to 10^4 CFU found for D39-L.

Phagocytosis by murine spleen macrophages. Spleen macrophages (Sp-Mφ) were isolated from BALB/c mice, essentially as described before (Alatery *et al.* 2008, J Immunol Meth), and mature macrophages were obtained after seven days of culture in medium containing M-CSF, confirmed to be positive for CD11b, CD11c, F4/80 and SIGLEC-1 by FACS analysis. Phagocytosis assays using different pneumococcal strains showed that the unencapsulated strain was efficiently phagocytosed *in vitro*, while phagocytosis of the encapsulated type 2 strain D39 were at very low levels. These results were in accordance with the important role of spleen macrophages in the clearance of bacteria, as observed by the inability of macrophage-depleted mice to resist pneumococcal infection (WP3). Furthermore, these data show a direct correlation between *in vitro* phagocytosis and *in vivo* capacity of bacterial clearance from the blood stream, thus providing for the first time an *in vitro* assay able to predict pneumococcal virulence in the host.

Cytotoxicity. The ciliary beat frequency (CBF) of rat ependymal was used to assess toxicity of selected pneumococcal strains essentially as described (Hirst *et al*. 2000, Infect. Immun. 68:1557-62). All strains, except serotype 3 strain BS71, were able to significantly inhibit the CBF rat epyndemal cells, reaching complete stasis within the first 1-4h of incubation. Surprisingly, the efficiency of cilia beating inhibition did not correlate with bacterial virulence as BS71 was the most virulent of all strains tested (see WP3).

Chemotaxis. The ability of selected pneumococcal strains to attract human CD4 T-lymphocytes isolated from peripheral blood of healthy human donors was assessed as described before (Kadioglu *et al.* 2004, Infect. Immun. 72: 2689–2697). Chemotaxis of CD4 T cells was only induced by TIGR4 and not by any of the other *in vitro* grown pneumococci. Interestingly, *in vivo* grown (mouse-passaged) D39 was much more efficient in activation of CD4 T-cells than any *in vitro* grown strain, with an apparent effect of the route of infection used for the mouse passage: 50% migration for intraperitoneally injected bacteria compared to 13% after intranasal infection.

Comparison of two D39 strains. Comparison of the D39 strain from Leicester (D39-L) and the D39 strain from Stockholm (D39-S) with regard to uptake/killing by Raw 264.7 macrophages cells and adhesion to A549 and Detroit epithelial cells showed a significantly lower number of macrophage-associated, as well as intracellular killing of D39-L than D39-S, as well as a significantly lower binding ability to A549, but not Detroit, epithelial cells. Furthermore, FACS analysis using anti-type 2 serum or purified antibodies suggested slight differences in the amount of their capsule. No difference in induction of host gene expression was observed between the two D39 strains.

<u>Transcriptional response of pneumococci during adherence</u>

To be able to examine the transcriptional response of pneumococci during adherence, a pan-genomic Nimblegen expression array was designed, covering common and strain-specific genes of all ten strains used within the consortium (see WP1). After 4 hours of adherence to Detroit 562 cells, the expression of adherent (cell-attached) bacteria were compared to the expression of non-adherent (planktonic) bacteria present in the same well to exclude effects

on pneumococcal gene expression of secreted factors by the Detroit cells. Array signal were subjected to extensive data filtering to exclude signal due to cross-hybridisation of Detroit RNA (present in the fraction of cell-attached bacteria despite enrichment for microbial RNA) to pneumococcal gene probes. A variable number of genes were found to be statistically significantly differentially expressed in all strains, ranging from 158 genes for the unencapsulated, highly-adherent R6 strain to >600 for strains D39, Sp14-BS69 and CBR206. To filter out possible false-positives due to technical limitations in the amount of RNA recovered from the less adherent bacteria, the focus of further analyses on only those genes that showed a significant change in expression of at 2-fold or more in at least 5 of the 9 strains examined (due to technical reasons, we excluded the serotype 3 strain from further analysis). In total, this led to the identification of 170 genes that were consistently upregulated in cell-attached pneumococci, and 221 genes that were down-regulated in the cell-attached fraction. Among the genes with higher expression in the cell-attached fraction were several encoding known adhesion factors, such as *pspA, cbpA, hyl,* and *slrA*. Furthermore, different sets of transporter genes were observed to be differentially expressed, highlighting the importance of nutritional balance during adherence. Due to the selection criteria (five out of nine strains), serotype-specific capsular genes were not included in the final gene set, but no significant down-regulation of capsular genes in selected strains was noticed, in line with published inverse correlation of capsule expression and adherence ability. Similarly, differential regulation of known (strain-specific) adherence factors such as *rlrA* islet genes was observed.

Identification of pneumococcal genes affecting adherence

To identify (novel) pneumococcal genes essential for adherence, the high-throughput genome-wide negative selection screenings method TnSeq (van Opijnen *et al.*, 2009, Nat. Methods 6:767-772) was employed, where transposon mutant libraries are subjected to a particular challenge, after which next-generation sequencing technology (NGS) is used to identify mutants that are negatively selected from the library population due to the challenge. To be able to perform an adherence TnSeq screen, generation of mutant libraries, and thus efficient transformability of a strain, is of utmost importance. Since this turned out to be more problematic than anticipated for most PNEUMOPATH strains, mutant libraries in the R6 (non-encapsulated, $8x10^6$ adherent CFU) and TIGR4 (serotype 4, $2x10^5$ adherent CFU) strains were generated of 15,000 and 3,000 mutants, respectively. After 2h adherence, chromosomal DNA extracted from the adherent (cell-attached), non-adherent (planktonic), and total library fractions, was used to generate mutant-specific DNA probes. Mutants with decreased capacity to adhere to Detroit 562 cells were identified by Tn-Seq and the web-based interface ESSENTIALS recently developed at RUNMC (Zomer *et al.*, 2012, PLoS One 7(8):e43012). A total of 63 genes were found to be essential for adherence in R6 and 59 for TIGR4, while 16 and 26 genes appeared to be unfavorable for adherence in R6 and TIGR4, respectively. Surprisingly very little overlap in the genes identified between R6 and TIGR4, was observed: only two genes, encoding a phosphate transporter ATP-binding protein PstB and an acetyltransferase, were found to be essential for adherence in both strains, and two genes; coding for the rRNA methyltransferase RsmB and a hypothetical protein, were found to be disadvantageous for adherence. Only very few genes were identified encoding surface-exposed proteins, with no overlap between the two strains. This could indicate that the cell surface of both strains is different, or that the TIGR4 library was too small and thus not representative of all genes. Importantly, the TnSeq adherence screen identified several genes whose products have been described to be either directly or indirectly involved in adherence, such as the *ami* locus, *ppmA*, *bgaA*, *nanA*, and various transporter genes. Even though not all genes that are required for adherence will be differentially expressed, and, reversely, differential expression does not always directly translate into gene essentiality, it is of interest to note that expression of most of the known adherence factors identified by TnSeq appeared to be highest in the planktonic, non-adherent fraction, while expression of other known adherence factors, such as *pspA* and *cbpA*, was found to be significantly

upregulated in cell-attached bacteria in most strains. These results might potentially be explained by the following "priming hypothesis": when the pneumococcus senses the presence of epithelial cells, it will prime itself for adherence by upregulation of several adherence factors such as *ppmA* and the *ami* locus. Once bound to epithelial cells, expression of these early adhesion factors is switched off, while expression of others, such as *pspA* and *cbpA*, is required for maintained adherence. Further experiments are required to test this hypothesis.

Pathway analysis of identified genes.

The genes identified as factors affecting adherence by both genome-wide approaches were distributed over a variety of functional categories, with predominant functional classes of 'Conserved hypothetical and hypothetical proteins' and 'Transport and binding'. KEGG pathway enrichment analysis of identified genes was performed using the web-based DAVID bioinformatics tool (Huang *et al.*, 2009, Nature Protoc. 4:44-57). This showed significant enrichment of the ABC transporters pathway (KEGG pathway spn02010) among the genes whose expression was downregulated during adherence as well as the essential genes identified by TnSeq in R6. Enrichment of the same pathway was also observed in the group of genes with increased expression during adherence and the TIGR4 TnSeq list, but this was not significant in both cases. Taken together, our results again underscore the importance of nutritional balance during adherence, and provide new insights into this process.

Characterisation of the host response during adherence.

The host response after 4h of pneumococcal adherence was examined by real-time PCR, focusing on a selected number of innate immunity genes. Upon data normalisation to GAPDH, clear differences in expression response of Detroit cells to adherent pneumococci was observed, with the highest expression levels observed for IL8, IL6, IL1B, and CXCL2. Expression of IL10 and IL17 was below detection level, while expression of a second housekeeping gene, ACTB, did not differ between the samples. The differences in host response appeared to be somewhat related to capsular type and adherence levels, as the strain with highest adherence generally induced the strongest response. No difference was observed in the response to adherence of two serotype 3 strain. Interestingly, a small but significant difference was observed in expression of Detroit cells upon adherence of the two sets strains with identical serotype and MLST genotype, BHN191/BHN418 (serotype 6B and ST138) and CBR206/LgSt215 (serotype 19F and ST179), indicating a complex and dynamic interplay between host and pneumococcus during the adherence and colonisation.

Workpackage 5 – Pneumococcal Physiology

Workpackage Objectives

The main objective of this work package was to document the central metabolism of *S. pneumoniae* to provide the basic knowledge required to correlate strain-to-strain variations in the adaptation to different host compartments with possible alterations resulting in fine-tuning changes in central metabolic fluxes.

For this goal, the aim was to determine carbon, nitrogen and energy fluxes under defined growth conditions; identify the connection(s) between carbon and nitrogen metabolism, and the interplay between carbon and nitrogen regulators; characterize the connection(s) between central metabolism and expression of the X-state (competence) and of key virulence determinants, e.g. capsule production and utilization of choline (Cho); when possible, evaluate the impact of predefined physiological conditions on expression of candidate strain-specific determinants of colonisation or virulence.

Progress Towards Objectives

All objectives were attained

Carbon, Nitrogen and Energy Fluxes.

All the clinical isolates and model strains retained for the PNEUMOPATH project have been systematically assessed for their growth under defined conditions. Carbon and energy fluxes were calculated from known pathway topography and stoechiometric matrix concepts. No significant correlation was established coupling a specific growth phenotype with virulence in the initial identified strains. The effect on pH evolution was characterised and seen to be a factor influencing carbon and energy fluxes with potential effects on virulence and capsule production.

A clear growth phenotype was observed with various D39 variants (see below). In all cases observed maximum growth rates were retained at a constant value, but for virulent strains, growth arrest occurs at a relatively low biomass concentration, while an avirulent strain continues to grow, reaching biomass levels at least three times higher. Metabolome analysis confirmed a modified metabolic flux distribution, clearly indicating that the metabolic sugar catabolism network is responding to the overall fitness of the strains rather than specific and local regulatory phenomena. Transcriptional analysis shows that during the exponential growth phase in which growth rates and sugar uptake rates are highly conserved in these strains, only a limited number of genes can be seen to be differentially expressed. These are predominantly related to pyridine biosynthesis operons (*pyrEF, pyrK/pyrDb/lytB,*) and Cho metabolism (*licCB/pck*).

Carbon and Nitrogen Metabolism.

Initial physiological analysis suggested that the expression of genes associated with capsule biosynthesis (and the rate of capsule synthesis) and Cho metabolism might be under the control of the carbon catabolite regulator CcpA, a master regulator of carbohydrate metabolism in Gram-positive bacteria. In addition, consensus sites for CcpA binding (cre sites) exist upstream of the cps (coding for capsule synthesis) and lic (coding for Cho utilization) operons in all sequenced *S. pneumoniae* strains. A ccpA mutant, however, did not show modified expression of these genes. Thus, while capsule synthesis and Cho metabolism do appear to be linked to high rates of sugar conversion, CcpA is not directly involved in this mechanism.

Genomic Analysis of D39 strains.

Three D39 strains have been sequenced and SNPs identified compared to the original D39 sequence available in the databanks. Comparison of the avirulent D39 strain with two virulent strains showed only two common SNPs and approximately 50 SNPs specific to one or more but not all of the three strains. These are potential targets that could explain the differences in virulence phenotypes. The differences include a significant proportion of mutations in transport genes linked to both sugar and N-metabolite uptake, coherent with the gain in fitness. Comparative analysis of the SNP data narrows target SNPs down to less than ten point mutations but further work is necessary to establish which of the limited number of SNPs identified is the key target for establishing improved fitness and loss of virulence.

Repertoire of Sugar Transporters and Regulation of Sugar Transport in S. pneumoniae.

A genomic survey allowed establishment the occurrence of 21 phosphotransferase systems, seven carbohydrate uptake ABC transporters, one sodium:solute symporter and a permease, emphasising an exceptionally high capacity for uptake of carbohydrate substrates. Despite high genomic variability, combined phenotypic and genomic analysis of 20 sequenced strains assigned the substrate specificity only to two uptake systems. Systematic analysis of mutants for most carbohydrate transporters enabled a phenotype and substrate specificity to be assigned to 23 transport systems. For five putative transporters for galactose, pentoses,

ribonucleosides and sulphated glycans activity was inferred, but not experimentally confirmed and only one transport system remains with an unknown substrate and lack of functional annotation. Using a metabolic approach, 80% of the 32 fermentable carbon substrates were assigned to the corresponding transporter. The complexity and robustness of sugar uptake is underlined by the finding that many transporters have multiple substrates, and many sugars are transported by more than one system. The present work permits a functional map of the complete arsenal of carbohydrate utilisation proteins of pneumococci to be drawn and allows re-annotation of genomic data and it may serve as a reference for related species. These data provide tools for specific investigation of the roles of the different carbon substrates on pneumococcal physiology in the host during carriage and invasive infection (Bidossi *et al*., *PLoS ONE* 7: e33320).

CelR-I, a transcription factor previously implicated in virulence, was found to function as an activator of a cellobiose-utilisation gene cluster. This regulator might form a key-point in the switch between the use of different carbon sources, given that several regulatory phosphorylation sites are present in the protein (Shafeeq *et al.*, 2011, *Microbiology* 157: 2854-286).

Connections between Carbon and Nitrogen Metabolism and Regulators.

Regulation of carbohydrate metabolism. CcpA was shown to govern the expression of a high number of genes in *S. pneumoniae*, in particular genes involved in carbohydrate metabolism, but also nitrogen metabolism, virulence and several other processes (Carvalho *et al*., 2011, *PLoS ONE* 6: e26707). In addition, a link between capsule production and uracil metabolism was revealed, and it was found that CcpA and the carbon source (e.g. glucose *vs* galactose) affect the expression and cell-surface attachment of pneumococcal capsule as well as the cellular metabolite profiles, which might lead to altered virulence.

Regulation of Arginine Metabolism. The putative arginine transcriptional regulators, ArgR1 and AhrC, were shown to cooperate in the regulation of arginine acquisition genes in response to the concentration of extracellular arginine (Kloosterman and Kuipers, 2011, *J Biol Chem* 286: 44594-605). Three of these genes encode components of an arginine ABC transport system. ArgR1 and AhrC, were also found to be regulated by *aliB*, which encodes an arginine-containing-oligopeptide binding lipoprotein that works together with the Ami permease to internalise oligopeptides and contributes to pneumococcal colonization of the nasopharynx (Kerr *et al*., 2004, *Infect Immun* 72: 3902-3906).

The Global Nutritional Regulator CodY. CodY is a global regulator that is highly conserved in low-G+C Gram-positive bacteria (Sonenshein, 2005). In *Bacillus subtilis*, the CodY regulon is large, encompassing nearly 200 genes (Molle *et al*., 2003), most of which are repressed during exponential growth and induced when cells experience nutrient deprivation. In several pathogenic bacteria, CodY regulates major virulence genes (for review, see Sonenshein, 2005; 2007). CodY was first established to be an essential protein in *S. pneumoniae* strain D39 because quantitative back-transformation revealed that a previously described *codY* mutant (Hendriksen *et al*., 2008, *J Bacteriol* 190: 590-601) could not be transferred into a wildtype recipient, unless it was co-transferred with two genetically independent suppressing mutations (Caymaris *et al*., 2010, *Mol Microbiol* 78: 344-360). Genome sequencing identified suppressing mutations disrupting *fatC* in the major pneumococcal iron transporter Fat/Fec (also called *piuBCDA* or *pit1*; Brown *et al*., 2002, *Infect Immun* 70: 4389-4398) and the oligopeptide transporter Ami (Alloing *et al*., 1994, *J Mol Biol* 241:44-58), both of which are repressed by CodY (Hendriksen *et al*., 2008, Ibid.). We suggested that CodY is an essential protein because inactivation results in uncontrolled iron import, which is toxic for the cell.

The finding that CodY is an essential protein in the pathogenic strain D39 and laboratory equivalent R800 was extended to three other pathogenic strains (out of three tested), suggesting that CodY is globally essential in the species, and a potential target for treatment of pneumococcal diseases.

The two-component regulatory system CiaR-CiaH (CiaRH) affects a number of physiological processes such as development of competence, autolysis and ß-lactam resistance. It is also implicated in virulence. The response regulator CiaR controls directly 15 promoters, which drive expression of 24 genes, five of which specify small noncoding RNAs (csRNAs). During this project, the presence of altered *ciaH* alleles in clinical isolates of *S. pneumoniae* was documented (Muller *et al*., 2011, *Microbiology* 157: 3104-3112). It was also established that in contrast to many other two-component regulatory systems, expression of the CiaR regulon is high under a variety of growth conditions (Halfmann *et al*., 2011, *J Mol Microbiol Biotechnol* 20: 96-104). Furthermore, studies with CiaH-deficient strains revealed that CiaH is not always necessary to achieve high level CiaR regulon expression. Thus, the response regulator CiaR, which needs to be phosphorylated to be active, must be able to receive phosphate by sources other than the kinase CiaH and this alternative phosphorylation appears to be dependent on the growth conditions. *In vitro* CiaR DNA binding studies indicated that the regulator could be phosphorylated by acetyl phosphate (AcP). Thus, this high-energy phosphoryl donor could also be the phosphate source for CiaR. Another link of CiaR to cellular physiology relies on the regulatory csRNAs. The CiaRH-related phenotypes, mentioned above, are all dependent on the csRNAs. Defining targets controlled by the csRNAs so far yielded seven genes, including *comC*, encoding the precursor of the competence stimulating peptide, that constitutes the connection of CiaR to competence control.

Overall, results obtained in this project reveal many connections between carbon- and nitrogen-dependent transcriptional regulation. These include genes involved in glycolysis, amino acid metabolism, as well as virulence genes such as *pcpA*. Further meta-transcriptome analyses could yield more clues about important physiological control points of pneumococcal nitrogen- and carbon metabolism.

<u>Connection(s) between Central Metabolism and Expression of the X-state (Competence) and of Key Virulence Determinants.</u>

CiaRH, CodY and the Regulation of X-state (Competence). The connection between CiaRH and competence control via a CiaR-dependent csRNA potentially targeting the *comC* gene, encoding the precursor of CSP, was mentioned above. Concerning CodY and competence, investigation of spontaneous competence development in the codY mutant revealed a competence-upregulated (*cup*) phenotype. However, as the suppressing mutations alone were found to confer a strong *cup* phenotype, the net effect of *codY* inactivation in this background was to attenuate this *cup* phenotype. From these data, it is therefore difficult to conclude that CodY acts as a repressor of competence. Further work using different approaches, such as a transient depletion of CodY, would be necessary to evaluate the exact role of CodY in the regulation of pneumococcal competence.

Metal Homeostasis and Virulence. Besides the documentation of iron homeostasis revealed through analysis of CodY essentiality (see above), it was shown that a number of pneumococcal genes are differentially regulated by copper, including an operon, *cop*, encoding the CopY regulator (Shafeeq *et al*., 2011, *Mol Microbiol* 81: 1255-1270). Copper homeostasis is important for pneumococcal virulence as the expression of the *cop* operon is induced in the lungs and nasopharynx of intranasally infected mice. A strain mutant for *copA*, which encodes a P1-type ATPase, CopA, which is conserved in all sequenced pneumococcal strains, had decreased growth in high levels of copper *in vitro* and showed reduced virulence in a mouse model of pneumococcal pneumonia. Furthermore, using the *copA*$^{-}$ mutant, it was observed for the first time in any bacteria that copper homeostasis also appears to be required for survival in the nasopharynx.

Zinc (Zn^{++}) is another important trace metal ion that has been shown to regulate the expression of several (virulence) genes in streptococci. The genome-wide transcriptional response of *S. pneumoniae* to intracellular Zn^{++} limitation was established, and was

demonstrated to be mainly mediated via direct regulation by the Zn^{++}-dependent regulator AdcR (Shafeeq *et al*., 2011, *Metallomics* **3**: 609-618).

Central Metabolism, Capsule Production and Cho Utilisation. As mentioned above, capsule synthesis and Cho metabolism do appear to be linked to high rates of sugar conversion, and CcpA and the carbon source (e.g. glucose vs. galactose) affect the expression and cell-surface attachment of pneumococcal capsule; in addition, a link between capsule production and uracil metabolism was revealed. However, our results suggest that CcpA is not directly involved in the control of *cps* or *lic* operons at the transcriptional level. Cho is an essential nutrient for growth and virulence of the pneumococcus. One of the main operons involved in Cho metabolism is the essential *lic1* operon. Although previous results and the presence of putative *cre* sites in *lic1* promoter region had suggested that this operon may be regulated by CcpA, mutations of the *cre* sites were found to have no effect on *lic1* expression. Further analysis of the regulation of the *lic1* operon confirmed that one of the two lic1 promoters is dependent on CiaRH. In contrast to a previous conclusion that Cho starvation induced the *licD2* gene but not the *lic1* operon (Desai *et al*., 2003, *J Bacteriol* 185: 371-373), the latter was unambiguously established to be induced in conditions of Cho deprivation, and mutant analysis indicated that an unknown regulator (provisionally named LicR) repressed the *lic1*-P1 promoter during conditions of Cho repletion, with repression alleviated once Cho is depleted. Further work is required to identify LicR. To test whether this Cho-based regulation was important *in vivo*, a mutant in the Cho-regulated promoter, unable to respond to Cho deprivation, was compared with its wildtype parent in a mouse model of nasal colonisation and virulence. The ability to respond to Cho deprivation by inducing expression of *lic1* was found to be important for nasal colonisation, as the mutant was out-competed by the wild-type strain in this model. These results show that the propensity to respond to fluctuating levels of the essential nutrient Cho, tightly regulated by two regulators, is important for pneumococcal colonization. While the Cho-regulated *lic1* promoter is highly conserved among pneumococcal isolates, the presence of altered *ciaH* alleles in clinical isolates of *S. pneumoniae* documented in this project (see above) raises the possibility of a differential regulation of this promoter by CiaRH. Likewise, differences in *licR* expression may also result in differing abilities of certain strains to respond to fluctuations in Cho levels in different host niches. Since we demonstrate that this response is important for nasopharyngeal colonisation, it would be of interest to determine whether observed differences in response to Cho deprivation are linked to carriage potential.

WP5 results led to nine publications (*J Biol Chem*, *Mol Microbiol*, *Microbiology*, *J Mol Microbiol Biotechnol, Metallomics, PLoS ONE*). Several additional publications are foreseen on the topics described above.

Workpackage 6 – Integrated Bioinformatics and Data Management

Workpackage Objectives

There were three main goals in this work package, namely: deployment of a web-based semantic integration environment (S3DB), development of dedicated data processing services and development and maintenance of an Application Programming Interface (API), which allows project participants to control the sharing and public dissemination of their contributions.

Progress Towards Objectives

All objectives were attained.

One of the main goals of this work package was to develop a semantic web data management system to store the data produced by all partners during the project. Particular attention was given to the user interfaces. Different web based interfaces for data visualisation, retrieval and querying have been developed. For this purpose the sdLink system (http:/kdbio.inesc-id.pt/sdlink/), has been created and developed to produce a completely generic integrated environment for querying, retrieving and analysing linked data, suitable for users unfamiliar with such technologies. The sdLink can be configured to any domain knowledge once the data is annotated following a given ontology. It provides data views, including graphical representations, and a user-friendly querying interface. The querying interface was developed for use by users unfamiliar with semantic technologies, so that one can, for instance, define a query by means of a simple point and click interface, which is then translated to SPARQL.

All project participants used the sdLink system, with good results both in scalability and usability by non-expert users. The system has a public project available for evaluation purposes (http:/kdbio.inesc-id.pt/sdlink/sample/), see Figure 3.6.1.

Presently, all data that were produced by the partners are modelled and stored in this system. Excel interfaces have been programmed to represent the database model and to facilitate the data collection and upload to the system. All the project participants can query and retrieve the data. Based on the results of others European projects it was decided that the best way to share and disseminate the project results was through a web interface where the users can play with and retrieve the data made available from the project. This task, which can also be seen as new research in the informatics research field, will be published. Presently this publication is being completed and is to be submitted to *BMC Bioinformatics*. Preliminary results have been presented at the main semantic web conference (http://www.iscb.org/cshals2012) supported by the International Society for Computational Biology (ISCB).

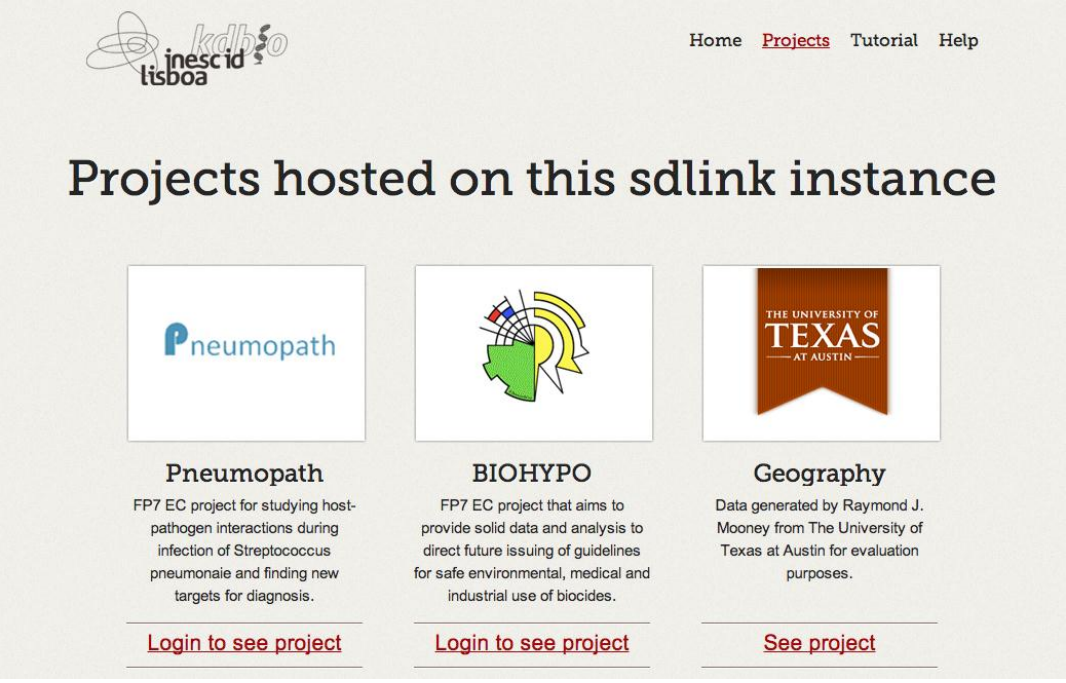

Figure 3.6.1: Interface to access the projects available in the system

sdLink Data Management System

During the last couple of years new standards and technologies have been developed to support the development of databases using semantic web principles. The development of sdLink was supported by the state-of-the-art technologies available for this type of systems.

The sdLink system was built on top of two main premises. First, it should take a minimal assumption approach with respect to data models, allowing for model extensibility and refinement, without enforcing severe restrictions on expressiveness. Second, it should be scalable, ensuring data availability, integrity and security.

With respect to the first assumption, sdLink follows the concept of linked data, a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF. Thus, data are represented following an ontology, that can be updated and which allows the definition of inference rules, an important feature for automatic data transformation, model updating and knowledge discovery.

sdLink was then built on top of well-known open technologies and was comprised of three main modules:

1) A management module for projects, users and access control. Authentication relies on the OpenID open standard, providing a decentralized mechanism, and all data about projects, users and access controls (authorization) is stored within the system itself. Figure 3.6.2 presents the new login interface where it is possible to select the OpenID provider.

2) The second module makes available endpoints for data retrieval and querying in several formats, both as REST endpoints and rich Web user interfaces, providing several views on the data. The query endpoint accepts SPARQL, with a user-friendly interface and pre-defined views and queries available. Figure 3.6.3 presents a user interface that is already available and which can be used to visualize the database concepts and the data,.

3) A third module consists of a scalable triple store, including support for transactions and a powerful query engine. This module is built on top of Virtuoso, a high performance triple store. All modules and configuration will be available as pre-configured virtual machine, suitable for both evaluation and production deployments.

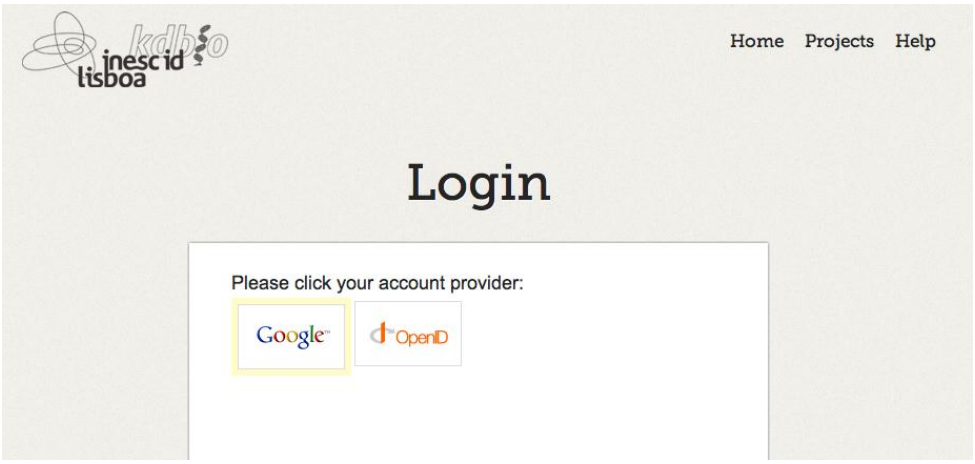

Figure 3.6.2: PNEUMOPATH new login interface

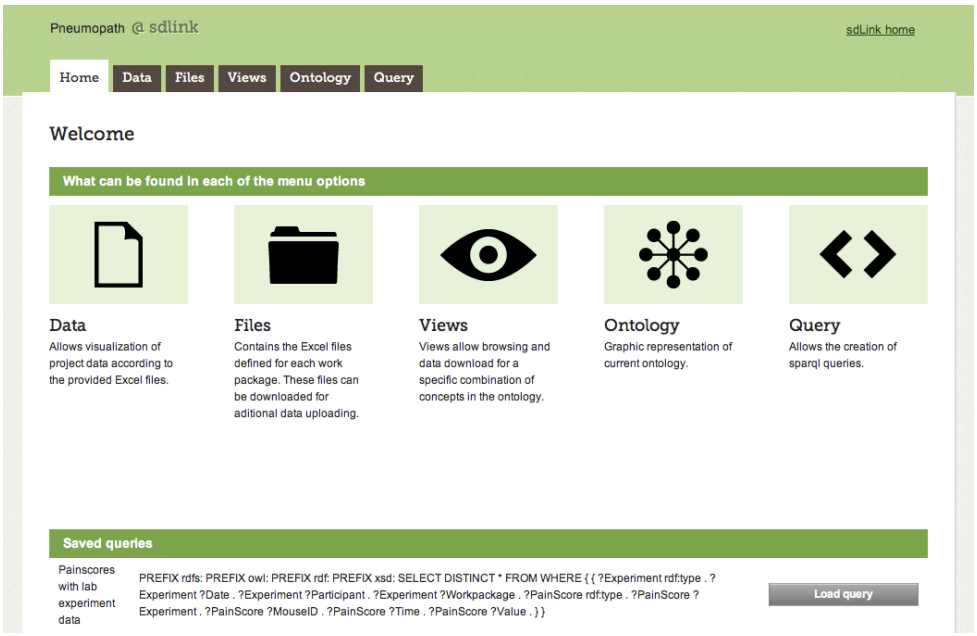

Figure 3.6.3: User interface used to visualise the database concepts and the data already available

Excel spreadsheet API

Besides the visualisation interfaces, specific interfaces have been developed to upload new data to the system. As is well-known, biologists are used to collection of the data in Excel spreadsheets and this pre-requisite was determinant to this type of interface development.

During this project, the development of API for spreadsheet client applications has been started and concluded. A new API is available at http://kdbio.inesc-id.pt/~pmreis/Setup.zip and the development site maintained at http://kdbio.inesc-id.pt/mailman/listinfo/sdl-dev.

Figure 3.6.4 shows an Excel workbook including a new tab for the sdLink API. With this new resource the user can select the database concepts and generate a personalized Excel workbook for the data that must be uploaded. When selecting the project PNEUMOPATH ontology, that describes the information system, a new window is opened in order to allow the selection of the concepts to be represented at the Excel workbook. This workbook contains one concept per worksheet and in each worksheet, each column correspond to an attribute connected to the concept. In this semantic web-based system, each concept should be seen as a table and each attribute, a column in that table. Each row of the table will be filled with the values of each attribute.

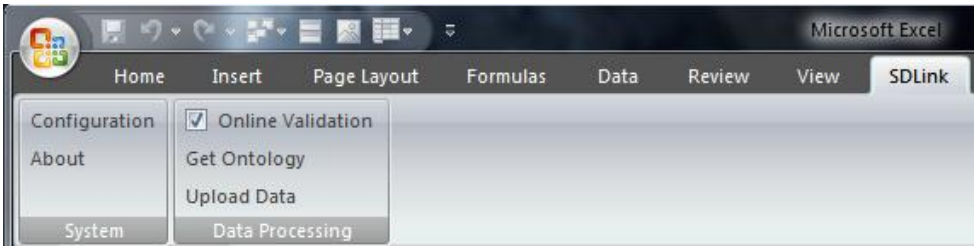

Figure 3.6.4: Excel workbook with the sdLink API included

The data considered in the context of the PNEUMOPATH project describes the host-pathogen interactions during infection by *Streptococcus pneumoniae* and thus includes:

characterisation of pneumococcal strains; typing information; data of *in vitro* and *in vivo* experiments with animal and cell models.

Some of these data were scattered across numerous information systems and repositories, each with its own terminologies, identifier schemes, and data formats. The need to share such data brought challenges for both data management and annotation, such as, the need to have a common understanding of the concepts that describes host-pneumococcal interactions. Thus, semantic annotation and interoperability became an absolute necessity for the integration of such diverse biomolecular data.

Therefore, the Spo ontology for expressing knowledge in this field has been defined, and, since contributions from all partners was taken into consideration, this approach led to a well-grounded set of concepts and annotations.

Spo provides a framework to represent the aforementioned host-pneumococcal interactions, being flexible enough to accommodate the rapid changes and advancement of research and achieve data interoperability and interchange. This has been only possible because of semantic Web recommended practices for clearly specifying names for things and relationships, expressing data using standardized and well-specified knowledge representation languages. The ontology described in OWL Lite v1.0 includes 36 classes, 24 object properties and 43 data properties.

The main contribution of the work package was not only to develop Spo, but all the approach and methodology for its construction in the context of a large research project, where many people were not aware of semantic technologies. The proposed ontology does not only describe knowledge in this field, but also allows for validating and aggregating existing knowledge, which is essential for data integration. Furthermore, the ability to accurately describe the host-pneumococcal interactions through the use of Spo has facilitated the implementation of information systems capable of coping with the heterogeneous types of data and, by using well known semantic technologies, it allowed users to query data and discover new knowledge. A manuscript is currently being finished (Vaz *et al*., to be submitted).

The analysis and integration of data obtained during the Project duration is one of the key steps to formulate new hypotheses regarding the mechanisms of infection. In particular, of special relevancy is the combination of data obtained from different work packages, namely strain genomic information, microarray data and *in vivo* experimental results in animal models, to ultimately analyse the correlation between different types of data.

The analysis included the genome wide association study of the project strains, in order to relate SNP information with specific phenotypic characteristics. This was based on comparing partitions and testing for statistically significant differences between the groups defined by those partitions, in terms of virulence. This allows sorting the SNPs according to their p-values, which might indicate some relevant biologically significant association between those mutations and virulence. This procedure can now be easily extended to other phenotypic information obtained by the partners of the consortium and inserted in the database.

Regarding the procedure proposed, the following points were explored:

1) Verification if all the relevant genes were present, otherwise use less stringent criteria for homology search. The development of the pipeline for SNP extraction from contigs data using different parameters is described in a publication (Francisco *et al*., to be submitted);

2) Calibration of the statistical tests in order to obtain more accurate results. The Kruskal-Wallis test with permutation p-value corrections was applied;

It is expected that the deliverable of these actions will be the formulation of putative associations between different types of information and hypotheses generation. These

results are being fully described in two papers, one addressing the statistical analysis in detail (Vinga. *et al*., to be submitted), the other implementing a web-application tool to allow a user-friendly environment for analysis based on new phenotypic data (Almeida *et al*., to be submitted).

Workpackage 7 – Management and Dissemination

Workpackage Objectives

- To co-ordinate the project within the consortium
- To co-ordinate contact between the consortium and the EC
- To disseminate, explain and exploit the objectives and results of the project with stakeholders such as the scientific community, public authorities and industry.

Progress Towards Objectives

All the objectives were attained.

During the lifetime of the project nine project consortium meetings were held. All partners participated in, and contributed to, all of the consortium meetings. In addition, meetings between partners took place alongside other conferences and meetings where opportunity arose.

Continuing written and verbal communication, facilitated by Partner 1, took place between the project partners in order to ensure the effective management of the project.

Following agreement from all project partners, an extension to the project from 36 to 42 months was agreed by the Commission. In addition, the re-profiling of the budget to transfer some of the budget of Partner 13 to Partner 1 to enable completion of deliverable DL3.4 and some funds from Partner 5 to Partner 11 was agreed. This re-profiling enhanced the consortium's data analysis capabilities and thus deliverables across all work packages.

Following the approval of the extension, Partners 1 and 8 made minor re-profiling of their budgets in order to appropriately finance the no-cost extension.

A final, public symposium was held to disseminate the key results of the project. Held in June 2012 in Lisbon, Portugal, the symposium was attended by around 60 people and included a key-note presentation from Staffan Normark of the Karolinska Institutet, Sweden.

The project website, http://www.le.ac.uk/projects/pneumopath was developed in the first year and maintained throughout the lifetime of the project to disseminate information about the project's activities.

4. Potential Impact, Dissemination Activities and Exploitation of Results

PNEUMOPATH was a wide-ranging study of host-pathogen interactions during disease due to *Streptococcus pneumonia*e undertaken in response to the need to increase the knowledge of mechanisms of pathogenesis across a much wider range of pneumococcal strains and host situations than has been done before now. A major impediment previously was that the breadth of technologies and model systems that were necessary to fulfill the objectives of the project were beyond the resources of a single group. The interdisciplinary effort that was

made possible by the formation of the PNEUMOPATH consortium, involving academic and industrial partners, gave rise to a large collection of ground-breaking results and ideas. The project data has resulted in entirely new perceptions of the pathogenesis of pneumococcal diseases, with implications for diagnosis, vaccination and treatment. The project also provided some tools that will benefit future work. The success of the project reaffirms the position of European scientists at the forefront of global pneumococcal research. The users and beneficiaries of the data from this project will be: microbiologists, infectious diseases specialists and other health-care professionals, the wider scientific community and the pharmaceutical industry. The general public also will benefit from access to the knowledge gained and from future exploitation of the knowledge.

All of the scientific workpackages produced data with good potential for impact. Studies conducted in WP1 improved our understanding of the evolution and genetic differences of *S. pneumoniae* and its close relatives. In particular, it may be possible to explore the genetic differences between these species to develop rapid diagnostic tests aimed to unequivocally identify *S. pneumoniae* in clinical samples. However, the data represent a 'double-edged sword': the data show that a highly specific test is a real possibility but they also provide a warning to scientists and health-care professionals that the closeness of the relationships with other species means that the difficulties in developing a unequivocal diagnostic test are high and that false positive results with the diagnostic targets that have been proposed elsewhere are real possibilities.

WP1 provided a DNA microarray chip covering ten pneumococcal strains. This was a valuable tool within the project that was used to identify genes associated with invasive disease, as well as genes expressed during specific conditions, but this microarray also will be a valuable resource to the wider pneumococcal research community. Within PNEUMOPATH use of the chip provided data of high value for the development of new diagnostic reagents and treatment strategies aimed to contain disease caused by virulent strains.

Within WP2, PNEUMOPATH investigated the genetics of host susceptibility to invasive pneumococcal disease. In the first study of its type applied to an infectious disease, a genome wide association study (GWAS) was undertaken using a large panel of inbred mice to take advantage of the increasing availability of dense genotyping data. The GWAS analysis identified three loci that associated with susceptibility to pneumonia. This type of work is in its infancy but nevertheless the data obtained here will provide a rich seam of new information for those scientists seeking to understand mammalian susceptibility to infection. In the WP2 analysis of human data, two rare sequence variants disrupting two genes were discovered to have large effects on risk of pneumococcal disease, when homozygous. These are by far the biggest genetic risk factors of pneumococcal disease described to date.

The *in vivo* studies undertaken in WP3 made important discoveries. It was found that cytokine levels could predict the outcome of pneumococcal disease in mice early in infection and were characteristic of host susceptibility. The prognostic significance of these observations should now be tested in humans. The *in vivo* studies also gave new insights into the process of development of pneumococcal bacteraemia, a condition that accompanies many cases of pneumococcal pneumonia and which is responsible for at least 10% of bacteraemia in Europe. It was ascertained that the action of spleen macrophages during the first minutes of infection determines the clinical outcome and that sustained bacteraemia develops from the replication of a single bacterial cell even when many thousands invade the blood. These data point to new routes for host defence strategies to control pneumococcal disease progression. The work provided evidence that males are more susceptible to pneumococcal disease than females, pointing to the need for further studies in humans.

A very useful tool for future studies to understand the pathogenesis of pneumonia was the identification of a strain (BS71) that was lethal to mice but was restricted to the lungs and did

not cause a bacteraemia. The potential experimental impact of this strain derives from the fact that while much human pneumonia is not accompanied by bacteraemia, all of the animal models of pneumonia are accompanied by bacteraemia. The identification of BS71 changes this situation by giving researchers a tool to study pneumonia in isolation. The knowledge gained and the realisation of the importance of this strain illustrated the impact of the interdisciplinary PNEUMOPATH consortium: work in WP1, 3, 4, 5 and 6 contributed to knowledge of this strain.

In WP4, the interplay between different pneumococcal strains and a variety of host cells in *in vitro* models were examined, representing different aspects of pneumococcal pathogenesis. This enabled identification of those aspects of the pneumococcus-host interaction that are common between strains and cell types, as well as characteristics specific for certain interactions. The knowledge of these factors, especially those that are conserved between strains, will lead to the identification of new targets for diagnosis, therapy or prevention.

Work on the physiology of the pneumococcus done in WP5 showed the importance of nutrition for the virulence of pathogens. These studies showed that the physiology of the pneumococcus could be a very productive area in the search for new therapeutic targets. An example of this potential is illustrated by the discovery that CodY is an essential protein for the pneumococcus. This finding points directly to a new target for treatment of pneumococcal diseases. A serendipitous finding within the project that has great potential for providing a definitive insight into mechanism of virulence was that two extremely related variants of the D39 strains had completely different virulence. The closeness of the genomic relationship of these two strains must have an impact on our ability to understand, and hence target, virulence. The importance of these D39 variants is another illustration the value of the interdisciplinary PNEUMOPATH consortium, as WP1, 3, 4, 5 and 6 all contributed to the work.

Several bioinformatic advances were made in WP6 that will have applicability beyond the PNEUMOPATH project, for those projects in which the organisation and querying of complex and disparately described data is required. The developments in WP6 included development of dedicated data processing services and an Application Programming Interface and implementation of a semantic web data management system. The systems were designed for use by those unfamiliar with semantic technologies and are currently available for public evaluation.

Publications

The PNEUMOPATH project was acknowledged in the following publications:

Almeida JS, Grüneberg A, Maass W, and Vinga S. 2012. Fractal MapReduce decomposition of sequence alignment. *Algorithms for Molecular Biology.* **7**:12

Almeida JS, Deus HF, Maass W. 2010. S3DB core: a framework for RDF generation and management in bioinformatics infrastructures. *BMC Bioinformatics*. **11**:387

Bidossi A, Mulas L, Decorosi F, Colomba L, Ricci S, Pozzi G, Deutscher J, Viti C, Oggioni MR. 2012. A functional genomics approach to establish the complement of carbohydrate transporters in *Streptococcus pneumoniae*. *PLoS ONE.* **7(3):**e33320

Burghout P, Cron LE, Gradstedt H, Quintero B, Simonetti E, Bijlsma JJ, Bootsma HJ, Hermans PW. 2010. Carbonic anhydrase is essential for *Streptococcus pneumoniae* growth in environmental ambient air. *Journal of Bacteriology.* **192(15):**4054-62

Camilli R, Bonnal RJP, Del Grosso M, Iacono M, Corti G, Rizzi E, Marchetti M, Mulas L, Iannelli F, Superti F, Oggioni MR, De Bellis G and Pantosti A. 2011. Complete genome sequence of a serotype 11A, ST62 *Streptococcus pneumoniae* invasive isolate. *BMC Microbiology.* **11**(1):25.

Carvalho SM, Kloosterman TG, Kuipers OP, and Neves AR. 2011. CcpA ensures optimal metabolic fitness of *Streptococcus pneumoniae*. *PloS ONE*. **6**:e26707.

Caymaris S, Bootsma HJ, Martin B, Hermans PWM, Prudhomme M, Claverys JP. 2010. The global nutritional regulator CodY is an essential protein in the human pathogen *Streptococcus pneumonia*. *Molecular Microbiology*. **78(2):**344-360

Cron LE, Stol K, Burghout P, van Selm S, Simonetti ER, Bootsma HJ, Hermans PW. 2011. Two DHH subfamily 1 proteins contribute to pneumococcal virulence and confer protection against pneumococcal disease. *Infection and Immunity*. **79(9):**3697-3710

Deus H, Correa M, Stanislaus R, Miragaia M, Maass W, de Lencastre H, Fox R, Almeida J. 2011. S3QL: A distributed domain specific language for controlled semantic integration of life sciences data. *BMC Bioinformatics*. **12:**285

Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, Oggioni MR, Dunning Hotopp JC, Hu FZ, Riley D, Covacci A, Mitchell TJ, Bentley SD, Kilian M, Ehrlich GD, Rappuoli R, Moxon ER and Masignani V. 2010. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biology*. **11(10):**R107.

Frazão N, Sá-Leão R, De Lencastre H. 2010. Impact of a single dose of the 7-valent pneumococcal conjugate vaccine on colonization. *Vaccine*. **28(19):**3445-3452

Gualdi L, Kaur Hayre J, Gerlini A, Bidossi A, Colomba L, Trappetti C, Pozzi G, Andrew PW, Ricci S, and Oggioni MR. 2012. Regulation of neuraminidase expression in *Streptococcus pneumoniae*. *BMC Microbiology*. **12(1):**2

Halfmann A, Schnorpfeil A, Muller M, Marx P, Gunzler U, Hakenbeck R, and Bruckner R. 2011. Activity of the two-component regulatory system CiaRH in *Streptococcus pneumoniae* R6. *Journal of Molecular Microbiology and Biotechnology*. **20:**96-104.

Kadioglu A, Cuppone AM, Trappetti C, List T, Spreafico A, Pozzi G, Andrew PW, Oggioni MR. 2011. Sex based differences in susceptibility to respiratory and systemic pneumococcal disease in mice. *Journal of Infectious Disease*. **204(12):**1971-9

Kaur Hayre J, Xu G, Taylor GL, Docquier JD, Andrew PW, Oggioni MR. 2012. Optimization of a direct spectrophotometric method to investigate the kinetics and inhibition of sialidases. *BMC Biochemistry*. **13:**19

Kloosterman TG, and Kuipers OP. 2011. Regulation of arginine acquisition and virulence gene expression in the human pathogen *Streptococcus pneumoniae* by transcription regulators ArgR1 and AhrC. *J Biol Chem.* **286:**44594-44605.

Mellroth P, Daniels R, Eberhardt A, Rönnlund D, Blom H, Widengren J, Normark S, Henriques-Normark B. 2012. LytA, the major autolysin of *Streptococcus pneumoniae*, requires access to the nascent peptidoglycan. *J Biol Chem*. 30;287(14):11018-29.

Molzen TE, Burghout P, Bootsma HJ, Brandt CT, van der Gaast-de Jongh CE, Eleveld MJ, Verbeek MM, Frimodt-Møller N, Ostergaard C, Hermans PW. Genome-wide identification of *Streptococcus pneumoniae* genes essential for bacterial replication during experimental meningitis. 2011. *Infection and Immunity.* 79(1):288-297

Muller M, Marx P, Hakenbeck R, and Bruckner R. 2011. Effect of new alleles of the histidine kinase gene ciaH on the activity of the response regulator CiaR in *Streptococcus pneumoniae* R6. *Microbiology.* 157: 3104-3112.

Ricci S, Janulczyk R, Gerlini A, Braione V, Colomba L, Iannelli F, Chiavolini D, Oggioni MR, Björck L, Pozzi G. 2011. Immunisation with the factor H-binding fragment of PspC protects mice from pneumococcal sepsis by multiple antibody-mediated mechanisms. *Vaccine.* 29: 8241– 8249.

Salter SJ, Hinds J, Gould KA, Lambertsen L, Hanage WP, Antonio M, Turner P, Hermans PW, Bootsma HJ, O'Brien KL, Bentley SD. 2012. Variation at the capsule locus, cps, of mistyped and non-typeable *Streptococcus pneumoniae* isolates. *Microbiology*. 158: 1560-1569

Shafeeq S, Kloosterman TG, and Kuipers OP. CelR-mediated activation of the cellobiose-utilization gene cluster in *Streptococcus pneumoniae*. 2011. *Microbiology.* 157: 2854-2861.

Shafeeq S, Kloosterman TG, and Kuipers OP. Transcriptional response of *Streptococcus pneumoniae* to Zn2+ limitation and the repressor/activator function of AdcR. 2011. *Metallomics.* 3: 609-618.

Shafeeq S, Yesilkaya H, Kloosterman TG, Narayanan G, Wandel M, Andrew PW et al. 2011. The cop operon is required for copper homeostasis and contributes to virulence in *Streptococcus pneumoniae*. *Molecular Microbiology*. 81: 1255-1270.

Simões AS, Pereira L, Nunes, S, Brito-Avô A, de Lencastre H, and Sá-Leão R. 2011. Clonal evolution leading to maintenance of antibiotic resistance rates among colonizing pneumococci in the PCV7 era in Portugal. *Journal of Clinical Microbiology.* 49(8): 2810-2817

Simões AS, Valente C, de Lencastre H, Sá-Leão R. 2011. Rapid identification of non-capsulated *Streptococcus pneumoniae* in nasopharyngeal samples allowing detection of co-colonization and re-evaluation of prevalence. *Diagnostic Microbiology of Infectious Diseases.* 71(3):208-16

Simões AS, Sá-Leão R, Eleveld MJ, Tavares DA, Carriço JA, Bootsma HJ, Hermans PWJ. 2010. Highly penicillin-resistant, multi-drug resistant, pneumococcus-like strains colonising children in Oeiras, Portugal: Genomic characteristics and implications for surveillance. *Journal of Clinical Microbiology*. 48(1):238-46.

Trappetti C, Gualdi L, Di Meola L, Jain P, Korir CC, Edmonds P, Iannelli F, Ricci S, Pozzi G, and Oggioni MR. 2011. The impact of the competence quorum sensing system on *Streptococcus pneumoniae* biofilms varies depending on the experimental model. *BMC Microbiology.* 11(1):75

Planned Publications

Almeida JS et al. A Social Computing Platform for Pneumococcal SNP phenotype-genotype association. Submission planned for October 2012.

Boianelli B, Bidossi A, Gualdi L, Mulas L, Mocenni C, Pozzi G, Vicino A, Oggioni MR. A non-linear deterministic model for regulation of diauxic lag on cellobiose by the pneumococcal multidomain transcriptional regulator CelR. *PLoS ONE.* In production

Bootsma HJ, Zomer A, Eleveld MJ, Hermans PWM. Identification of novel pneumococcal adherence factors by a combination of genome-wide approaches. Submission planned for 2013.

Browall S, Dagerhamn J, Tångroth J, Sjöström K, Hellberg C, Bättig P, Spadafina T, Norman M, Sandgren A, Örtqvist Å, Normark S, Henriques-Normark B. Intraclonal variations affect disease development in pneumococcal infections. Submission planned.

Francisco AP et al. sdLink: An integrated system for managing biological and biomedical linked data. Submission planned for October 2012.

Francisco AP et al. From sequencing data to SNPs. Submission planned for October 2012.

Frazão N et al. Drug-resistant *Streptococcus pneumoniae* clones, selected by the PCV7 vaccine, show colonization and invasive potential in mice. Submission planned for 2013.

Gerlini A, Colomba L, Braccini T, Pammolli A, Pozzi G, Ricci S, Andrew PW, Koedel U, Moxon R, and Oggioni MR. A single bacterium at the origin of invasive pneumococcal disease: macrophage clearance defines the bottleneck during the first hours of infection. Submitted.

Jonsdottir I et al. A rare frameshift mutation disrupting the complement pathway confers a 27 fold risk of invasive pneumococcal disease. Submission planned late 2012.

Jonsdottir I et al. A rare mutation affecting the NADPH oxsidase pathway confers a 10 fold risk of invasive pneumococcal disease. Submission planned late 2012.

Syk A, Fernebro J, Sandgren A, Normark S, Henriques-Normark B. Emergence of hyper-virulent spxB mutants during systemic pneumococcal serotype 1 infection in mice and man.

Tavares DA et al. Identification of genes associated with pneumococcal invasive disease potential using a comparative genomic hybridization approach. Submission planned for 2013

Vaz C et al. Spo: An ontology for describing host-pathogen interactions inherent to *Streptococcus pneumoniae* infections. Submission planned for late 2012.

Vinga S et al. SNP2Pheno - linking genome and phenotype data in bacterial strains. Submission planned for late 2012.

Dissemination Activities

Data generated during this project were disseminated by presentation at local, national and international conferences and publication in relevant high-impact, peer-reviewed journals. Dissemination of results to the wider scientific community occurred throughout the project, as soon as it was suitable for peer-review publication and after IP considerations.

Lead partner	Title	Date	Place
P1	**Project website:** www.le.ac.uk/pneumopath		
P5	**The Icelandic Centre for Research:** Experience of participation in EU-funded research projects in Health. Jonsdottir I.	2009, April 24th	Reykjavík, Iceland
P6	**International Symposium on Metabolism and Bacterial Pathogenesis:** Multiple CUT1 ABC transporter permeases, involved in the virulence of Streptococcus pneumoniae, are served by a single orphan ATP binding protein. Bidossi A., L Mulas, E Tatti, C Viti and MR Oggioni.	2009, April 4th - 7th	Hohenkammer, Germany
P6	**International Symposium on Metabolism and Bacterial Pathogenesis:** Local concentration of sialic acid as a preventable signal for pneumococcal biofilm, colonisation and invasion of the host. Metabolism Meets Virulence. Trappetti C, A Kadioglu, F Iannelli, G Pozzi, PW Andrew, MR Oggioni.	2009, April 4th - 7th	Hohenkammer, Germany
P6	**SIMGBM meeting:** Sialic acid: a signal for pneumococcal biofilm, colonisation and invasion of the host. Trappetti C., A Kadioglu, G Pozzi, PW Andrew, MR Oggioni.	2009, June 11th - 13th	Spoleto, Italy
P6	**9th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** Pneumococcal beta glucoside metabolism investigated by whole genome comparison. Mulas L., Iannelli F., Trappetti C., Bidossi A., Pozzi G., Arioli S., Mora D., Tatti E., Viti C., Hakenbeck R., and M.R. Oggioni.	2009, June 4th - 6th	Bern, Switzerland
P6	**9th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** Genome characterization of a serotype 11A, ST62 Streptococcus pneumoniae clinical isolate. Camilli R, RJP Bonnal, M Del Grosso, M Iacono, E Rizzi, F Superti, L Mulas, M Marchetti, F Iannelli, G De Bellis, MR Oggioni, A Pantosti.	2009, June 4th - 6th	Bern, Switzerland
P6	**9th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** Serum antibody and cytokine levels in patients infected with *Streptococcus pneumoniae*. Hussain S, DM Ferreira, R El-Rachkidy, PW Andrew, MR Oggioni, A Kadioglu.	2009, June 4th - 6th	Bern, Switzerland
P6	**9th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** Gender difference in the susceptibility to invasive pneumococcal disease in mice. Kadioglu A, PW Andrew, G Pozzi, and MR Oggioni.	2009, June 4th - 6th	Bern, Switzerland
P6	**9th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** Sialic acid: a preventable signal for pneumococcal biofilm, colonisation and invasion of the	2009, June 4th - 6th	Bern, Switzerland

	host. Trappetti C, A Kadioglu, M Carter, F Iannelli, G Pozzi, PW. Andrew, and MR. Oggioni.		
P6	**9th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** Biofilm and matrix: phase variant specific phenotype in pneumococci. Trappetti C,Ogunniyi AD,Oggioni MR, JC Paton.	2009, June 4th - 6th	Bern, Switzerland
P3	**9th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** Identification of essential genes in Streptococcus pneumoniae in experimental meningitis. Molzen T, Burghout P, Bootsma HJ, Brandt C, Pedersen M, Eleveld MJ, van de Gaast-de Jongh CE, Østergaard C, Hermans PWM	2009, June 4th - 6th	Bern, Switzerland
P2	**9th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** The global nutritional regulator CodY is an essential protein in the human pathogen Streptococcus pneumoniae. Stéphanie Caymaris, Bernard Martin, Hester J. Bootsma, Peter W.M. Hermans, Marc Prudhomme, and Jean-Pierre Claverys (Invited communication).	2009, June 4th - 7th	Bern, Switzerland
P6	**19th ECCMID:** Sialic acid: a preventable signal for pneumococcal biofilm, colonisation and invasion of the host. Trappetti C, A Kadioglu, F Iannelli, G Pozzi, PW Andrew, MR Oggioni.	2009, May 16th - 19th	Helsinki, Finland
P8	Highly penicillin-resistant multidrug-resistant pneumococcus-like strains colonizing children: genomic characteristics and implications for surveillance.	2009, Nov 28th – 30th	Vilamoura, Portugal
P6	**37° Congresso Nazionale della Società Italiana di Microbiologia (SIM):** Analisi genomica e fenotipica del metabolismo dei beta glucosidi in pneumococco. Laura L., S. Arioli, D. Mora, E. Tatti, C. Viti, G. Pozzi, MR. Oggioni.	2009, Oct 11th - 14th	Torino, Italy
P6	**EUROBIOFILMS:** Local concentration of sialic acid as a preventable signal for pneumococcal biofilm, colonization and invasion of the host. Trappetti C., A Kadioglu, G Pozzi, PW Andrew, MR Oggioni.	2009, Sept 2nd - 5th	Rome, Italy
P8	**ISPPD-7**: Drug resistant clones colonizing children in Portugal in the PCV7 era – how much has changed?	2010, March 14th – 18th	Tel Aviv, Israel
P6	**ISPPD-7:** Hpr dependent autorepression of beta-glucoside uptake; a novel mechanism for carbon catabolite repression in firmicutes. Bidossi A., L Mulas, E Tatti, C Viti and MR Oggioni.	2010, March 14th - 18th	Tel Aviv, Israel
P6	**ISPPD-7:** Defining the bottleneck for invasive pneumococcal infection in mice. Colomba L, T Braccini, A Gerlini, G Pozzi, S Ricci, MR Oggioni.	2010, March 14th - 18th	Tel Aviv, Israel
P6	**ISPPD-7:** Sialic acid as molecular signal in a model of pneumococcal carriage. Trappetti C, L Colomba, G Pozzi, A Kadioglu, P Andrew, MR Oggioni.	2010, March 14th - 18th	Tel Aviv, Israel

P5	**University of Iceland, School of Health Sciences:** Experience of participation in EU-funded collaborative research projects	2010, March 2nd	Reykjavik, Iceland
P6	**Biofilms 4:** Quorum sensing in pneumococcal biofilm. DiMeola L, C Trappetti, L Gualdi, G Pozzi and MR Oggioni.	2010, Sept 1st - 3rd	Winchester, UK
P6	**2nd Florence Conference on Phenotype MicroArray Analysis of Microorganisms The Environment, Agriculture, and Human Health:** Genomic overview on carbohydrate uptake in *Streptococcus pneumoniae*. Bidossi F., L. Mulas, F. Decorosi, C. Viti, G. Pozzi and MR Oggioni.	2010, Sept 13th - 15th	Florence, Italy
P6	**21st ECCMID European Congress of Clinical Microbiology and Infectious Diseases:** Chromosomal drug efflux pumps in *Streptococcus pneumoniae*. Tocci N. , A. Bidossi, M.L. Ciusa, F. Decorosi, C. Viti, G. Pozzi, S. Ricci, M.R. Oggioni.	2011, May 7th - 10th	Milan, Italy
P6	**International Conference on Gram-Positive Microorganisms:** Involvement of the EIIB domain of the beta-glucoside operon regulator exclusively in the duration of diauxic lag in pneumococci. Gualdi L.' A. Bidossi, A. Boianelli, L. Mulas, C. Mocenni, G. Pozzi, A. Vicino, MR Oggioni.	2011, June 21st - 23rd	Montecatini-Terme, Italy
P6	**International Conference on Gram-Positive Microorganisms:** Functional genomic characterization of the complete set of pneumococcal carbohydrate uptake systems. Bidossi A., L. Mulas, F. Decorosi, S. Ricci, G. Pozzi, C. Viti, MR Oggioni.	2011, June 21st - 23rd	Montecatini-Terme, Italy
P6	**International Conference on Gram-Positive Microorganisms:** A single bacterium at the origin of invasive pneumococcal disease: macrophage clearance defines the bottleneck during the first hours of infection. Colomba L., A. Gerlini, T. Braccini, A. Pammolli, G. Pozzi, S. Ricci, P. Andrew, U. Koedel, R. Moxon and MR Oggioni.	2011, June 21st - 23rd	Montecatini-Terme, Italy
P2	**10th European meeting on the molecular biology of the pneumococcus (EuroPneumo):** Analysis of mutations suppressing the essentiality of the global nutritional regulator CodY in *Streptococcus pneumoniae*. Calum Johnston, Stéphanie Caymaris, Bernard Martin, Hester Bootsma, Peter Hermans, Marc Prudhomme, Jean-Pierre Claverys (Communication).	2011, June 23rd - 26th	Amsterdam, Netherlands
P6	**10th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** Functional genomic characterisation of the complete set of pneumococcal carbohydrate uptake systems. Bidossi A., L. Mulas, F. Decorosi, S. Ricci, G. Pozzi, C. Viti, MR Oggioni.	2011, June 23rd - 26th	Amsterdam, Netherlands
P6	**10th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** Comparative evaluation of pneumococcal efflux transporters. Ricci S., Tocci N. , A. Bidossi, M.L. Ciusa, F. Decorosi, C. Viti, G. Pozzi, M.R. Oggioni.	2011, June 23rd - 26th	Amsterdam, Netherlands

P6	**10th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** Early Macrophage clearance is responsible for the control of pneumococcal invasive infection in the non-immune host. Gerlini A., L. Colomba, T. Braccini, G. Pozzi, P Andrew, U Koedel, S Ricci, and MR Oggioni.	2011, June 23rd - 26th	Amsterdam, Netherlands
P6	**10th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** A single bacterium at the origin of invasive pneumococcal disease: macrophage clearance defines the bottleneck during the first hours of infection. Colomba L., A. Gerlini, T. Braccini, A. Pammolli, G. Pozzi, S. Ricci, P. Andrew, U. Koedel, R. Moxon and MR Oggioni.	2011, June 23rd - 26th	Amsterdam, Netherlands
P3	**10th European Meeting on the Molecular Biology of the Pneumococcus (EuroPneumo):** Identification of essential genes in *Streptococcus pneumoniae* in experimental meningitis. Molzen T, Burghout P, Bootsma HJ, Brandt C, Pedersen M, Eleveld MJ, van de Gaast-de Jongh CE, Østergaard C, Hermans PWM	2011, June 23rd - 26th	Amsterdam, The Netherlands
P5	**4th Congress of European Microbiologists (FEMS).** Searching for low frequency and rare variants with high-impact on infectious diseases through whole-genome sequencing. Jonsdottir I. Invited speaker.	2011, June 26th – 30th	Geneva, Switzerland
P6	**18th World Congress of the International Federation of Automatic Control (IFAC):** A nonlinear dynamic model for diauxic growth of *Streptococcus pneumoniae* on glucose and cellobiose. Boianelli A., A. Bidossi, C. Mocenni, L. Mulas, G. Pozzi, MR. Oggioni, A. Vicino.	2011, Aug 28th - Sept 2nd	Milan, Italy
P6	**XVIII Lancefield International Symposium:** A bottleneck for invasive pneumococcal disease: infection originates from one single cell which escapes early macrophage clearance. Colomba L., A. Gerlini, T. Braccini, A. Pammolli, G. Pozzi, S. Ricci, P. Andrew, U. Koedel, R. Moxon and MR Oggioni.	2011, Sept 4th - 8th	Palermo, Italy
P6	**SIMGBM 29th International Meeting:** A single bacterium at the origin of invasive pneumococcal disease: macrophage clearance defines the bottleneck during the first hours of infection. Colomba L, Gerlini A, Braccini T, Pamolli A, Pozzi G, Ricci S, Andrei P, Koedel U, Moxon R, Oggioni MR.	2011, Sept 21st - 23rd	Pisa, Italy
P6	**1st International SystemsX.ch Conference on Systems Biology:** Systems biology approach reveals involvement of the EIIB and EIIA domains of the beta-glucoside operon regulator in regulation of the duration of diauxic lag in pneumococci. Boianelli A., A. Bidossi, L. Gualdi, C. Mocenni, L. Mulas, G. Pozzi, M.R. Oggioni, A. Vicino.	2011, October 24th - 26th	Basel, Switzerland

P8	**ISPPD-8:** *Streptococcus pneumoniae* clones selected by the 7-valent pneumococcal conjugate vaccine in healthy children show invasive potential in mice. (poster)	2012, March 11th – 15th	Iguaçu Falls, Brazil
P3	**Scientific Spring Meeting KNVM & NVMM 2012:** Identification of novel pneumococcal adherence factors by a combination of genome-wide approaches. Bootsma HJ, Eleveld ME, Burghout P, Zomer A, Hermans PWM	2012, April 17th – 18th	Papendal, Netherlands
P6	**5th International Symposium on Communications, Control and Signal Processing:** Growth of *Streptococcus Pneumoniae* on Glucose and Beta-Glucosides. Vicino A., A. Boianelli, A. Bidossi, C. Mocenni, MR Oggioni. A Nonlinear Dynamic Model For Diauxic.	2012, May 2nd - 4th	Rome, Italy
P5	**2011 Annual Research Conference:** International Research Collaboration: Opportunities for increased international research collaboration and funding. Jonsdottir I.	2012, June 7th	Reykjavik, Iceland
P3	**Streptococcus suis minisymposium:** Detection of (conditionally) essential genes of respiratory pathogens by TnSeq. Bootsma HJ, Zomer A, Burghout P, Langereis J, de Vries, S, Hermans PWM	2012, June 8th	Wageningen, The Netherlands
All	**PNEUMOPATH symposium** (conference)	2012, June 27th	Lisbon, Portugal
P1	**Innate immunity of the lung – Improving pneumonia outcome:** Genome Wide Association Study identifies novel susceptibility loci for invasive pneumococcal disease. Jonczyk M, Sylvius N, Fernandes V, Denny P and Andrew PW.	2012, Sept 19th - 22nd	Berlin, Germany
P1	**SGM Autumn Conference:** Serotype 3 pneumococcal strains have distinct disease phenotype in BALB/c mice. Jonczyk M, Smith C, Radhakrishnan P, Sylvius N, Hulland K, Haste L and Andrew PW	2012, Sept 3rd - 5th	Warwick, UK

5. Project Website and Contact

www.le.ac.uk/projects/pneumopath

Professor P.W. Andrew (co-ordinator)
Department of Infection, Immunity and Inflammation
University of Leicester

Tel: +44(0)116 252 2951
Fax: +44(0)116 252 5030
Email: pwa@le.ac.uk

6. Partners

15 partners participated in the research project; 12 academic groups, one industrial partners and two SMEs;

Partner 1 (Coordinator): ULEIC Peter W. Andrew Department of Infection, Immunity and Inflammation University of Leicester PO Box 138 University Rd Leicester, LE1 9HN (UK) Phone: +44-116 252 2951 Fax: +44-116 252 5030 e-mail: pwa@leicester.ac.uk	**Partner 2: CNRS** Jean-Pierre Claverys Laboratoire de Microbiologie et Génétique Moléculaires UMR5100 CNRS-Université Paul Sabatier 118, route de Narbonne, 31062 Toulouse cedex 09 France
Partner 3: RUNMC Peter WM Hermans Laboratory of Pediatric Infectious Diseases Radboud University Nijmegen Medical Centre PO Box 9101 6500 HB Nijmegen The Netherlands	**Partner 4: SMI** Birgitta Henriques-Normark Swedish Institute for Infectious Disease Control Dept. of Bacteriology S-171 82 Solna Sweden
Partner 5: DECODE	**Partner 6: UNISI**

Ingileif Jonsdottir Division of Infectious and Inflammatory Diseases deCODE genetics Sturlugata 8 IS - 101 Reykjavik Iceland	Marco R Oggioni MD LA.M.M.B, Dipartimento di Biologia Molecolare Università di Siena Policlinico Le Scotte (lotto 5; piano 1) 53100 Siena Italy
Partner 7: UKAIS Regine Hakenbeck Department of Microbiology University of Kaiserslautern Paul Ehrlich Str., 23, D-67663 Kaiserslautern Germany	**Partner 8: ITQB** Hermínia de Lencastre Laboratory of Molecular Genetics Instituto deTecnologia Química e Biológica (ITQB) Rua da Quinta Grande, 6 - Apt.127 2780-156 Oeiras Portugal
Partner 9: UGRON Oscar P Kuipers Groningen Biomolecular Sciences and Biotechnology Institute University of Groningen Kerklaan 30, 9751 NN Haren The Netherlands	**Partner 10: INSA** Nic D Lindley Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés Institut National des Sciences Appliquees 135 Avenue de Rangueil 31077 Toulouse cedex 4 France
Partner 11: INESC Jonas Almeida Instituto de Engenharia de Sistemas e Computadores Rua Alves Redol 9, Apartado 13069 1000-029 Lisboa Portugal	**Partner 12: IC** Andreas Meinke Pre-clinical Research and Development Intercell AG Vienna Biocenter 6 A-1030 Vienna Austria
Partner 13: MRC-MGU Infection Genetics MRC Mammalian Genetics Unit Harwell OX11 0RD UK	**Partner 14: KI** Staffan Normark Department of Microbiology, Tumour and Cell Biology Karolinska Institutet Nobelsväg 16 SE-17177 Stockholm Sweden

Partner 15:	**Partner 16: IQT**
P15 did not accede to the Grant Agreement	Herman Groen IQ Corporation BV Rozenburglaan 13A NL-9747 DL Groningen The Netherlands