Title: TB-EUROGEN: Genetic Analysis of the host-pathogen interaction in tuberculosis

Acronim: TB-EUROGEN

Duration: 01/03/2008 – 28/02/2011 (36 months)

Instrument: Collaborative Project (Small or medium-scale focused research project)

Coordinator: Prof. Francis Drobniewski MA MSc MBBS DTM&H PhD FRCPath

Health Protection Agency National Mycobacterium Reference Laboratory
Centre for Infectious Diseases, Institute for Cell and Molecular Sciences,
Barts and the London NHS Trust, Queen Mary University of London,
2 Newark Street
London, E1 2AT f.drobniewski@qmul.ac.uk
Partners : P2 – UCAM (UK), P3 – SMI (Sweden), P4 – BNITM (Germany)

Project website: www.tb-eurogen.org

The goal of our project is to better understand host-TB pathogen interactions and the factors underlying susceptibility through a genetic and phenotype analysis of host related factors leading to active pulmonary TB and the corresponding bacterial isolate in a European (Russian) population. Significant findings in this population will be evaluated in an African population.

Specifically, we are aiming to achieve the following objectives through four inter-linked work packages, including a management package: 1. Establish the world's largest and most statistically powerful resource of DNA samples from Russian TB patients (n=5000) and healthy geographically matched controls (n=5000), as well as *M. tuberculosis* isolates from the same TB patients (n=2000) and collect detailed clinical information regarding the outcome of their TB. (2) using the Russian TB sample collection and the existing Ghanaian sample collection of 2,000 TB patients and 2,000 controls perform a frontier genetic experiment aiming to identify human genes involved in susceptibility/resistance to TB; (3) understand pathogen genetic variation by characterising genetic variation in 2,000 *M. tuberculosis* isolates collected from Russian TB patients including the prevalence and distribution of known global *M.tuberculosis* families, the phylogenetic structure of *M.tuberculosis* populations and analysing polymorphisms in genes associated with virulence

We believe that our approach will lead to the discovery of genes involved in susceptibility/resistance to TB in humans. Deciphering effects of the natural variation between *M. tuberculosis* strains will help to identify virulence factors, and influence prognosis for TB patients. These results will improve our understanding of TB pathogenesis and may have impact on TB vaccine research, opening new targets for TB prevention.

TB patients and controls:

Work package 1 focuses on the establishment of human blood and mycobacterial DNA banks with detailed demographic, clinical and radiological data establishing the presence of actual pulmonary tuberculosis and healthy control within our main project site in Samara, Russia. So far all the required legal, financial and logistical mechanisms have been established including a letter of official support in the Samara Medical bulletin by the Ministry of Health & Social Development (MOH). A full protocol was finalised, formally reviewed and approved by the Samara Ethics Committee as well as the project co-ordinator's Institutional Ethics Committee.

A research team of study supervisors, and clinical site co-ordinators, bacteriologists and data managers has been established with patient recruitment (from September 2008), recruitment of control individuals, specimen and data collection, transportation and shipment procedure established. Fourteen clinics and hospitals have been networked. A password-protected project database with dual Russian and English functionality and real-time double blinded data entry with descriptive analysis has been established with anonymised data available to partners on a monthly basis.

Laboratory staff have been trained and EQA for bacterial culture and analysis completed so far. The target of 5000 control patients has been achieved and so far over 3700 TB patients have been recruited. Of those recruited within this study almost two-thirds are new TB cases, and make a medium age of 39 years, approximately 80% are of Russian /Ukrainian ethnicity, 17% reported TB amongst relatives in the past.

For this study we aim to recruit a further 900 TB patients which is challenging as we have observed a significant increase in the HIV-TB co-infection rate (study protocols excludes HIV positive individuals) leading to increasing numbers of excluded patients and almost one-quarter had been in prison previously (another possible factor possible to be included into further stratification analysis); four-fifths had a cough lasting more than two weeks and half had two or more lung regions affected radiologically with majority showing cavitation, over half of those currently in treatment had MDRTB.

Blood samples from TB patients and all controls have been shipped to Cambridge for analysis with one sample retained in Samara for analysis as indicated in the Samara Ethical Review process. At this moment the project has DNA samples available from 3078 TB patients and 5000 control individuals.

An archive of all positive cultures (in triplicates) and DNA (triplicates) is assembled: one culture and DNA is sent to the MRU for testing, one culture and DNA are

stored in the Samara archive and one culture and DNA are used for current work in Samara. Over 1100 MTB cultures have been archived.

Candidate gene studies:

Multiple SNPs in candidate genes suggested by previous studies have been genotyped in the available DNA samples of Russian TB patients and controls and tested for association with TB. Thus, in collaboration with the Genome Institute of Singapore we recently studied 1,536 polymorphisms from the candidate genes across the human genome in the Indonesian TB cases and controls and in the Russian TB collection. We found that polymorphisms in the TLR8 gene show evidence of association with TB in both populations, and that TLR8 transcript levels are significantly up-regulated in patients during the acute phase of disease, suggesting that TLR8 is involved in TB (1). Further studies of genetic variants reported to be associated with TB in the literature or emerging from our collaboration with other TB researchers are in progress.

Initially we planned a two stage experiment in which we proposed a GWAS in 1,000 TB cases and 1,000 controls using genotyping of an array comprising ~500,000 single nucleotide polymorphisms (SNPs). However, as genotyping has become cheaper we will now aim to genotyping 2,000 – 3,000 human DNA samples using the Affymetrix SNP 6.0 array. In Stage 2 (months 22-26) we proposed genotyping of 50-100 most associated SNPs in the additional 4,000 TB cases and 4,000 controls from Russia using lower scale genotyping platform, e.g. Taqman but we will now undertake GWAS in all 10,000 samples from Russian TB patients and controls. This will dramatically increase the power of the GWAS: association analysis done in all 10,000 samples and will identify TB-associated genetic polymorphisms with high statistical confidence (e.g. 80% power to detect at $P = 10^{-7}$ OR < 1.2 for common polymorphisms (see Figure 1); to avoid association due to technical artefacts that occasionally may happen during genotyping with dense arrays such as Affymetrix SNP 6.0, we will use the Taqman system to re-type in-house 50-100 most associated SNPs and confirm their association.

Pathogen related studies:

In Work Package 3, of the 1100 MTB cultures archived, spoligotyping has been performed on 1581 *M.tuberculosis* isolates and the majority of strains belong to the Euro-American and East Asian families with Beijing family strains dominating in all three groups and their prevalence increasing over the last four years(see Figure 2).

Initially it was planned to characterize phylogenetical structure and further differentiate TBs strains using detection of long-sequence polymorphisms (LSPs), single nucleotide polymorphisms (SNPs) and polymorphisms in relatively rapidly evolving repetitive sequences (minisatellites) previously reported to be phylogenetically significant and/or associated with increased virulence. However, within a recently established collaboration with the Sanger institute (Cambridge, United Kingdom) we are screening a representative panel of 200 *M.tuberculosis* strains for LSPs and SNPs using a next-generation whole genome sequencing approach, followed by a targeted detection of SNPs; 72 *M.tuberculosis* complex isolates have been selected so far for the whole genome sequencing based on spoligotyping and VNTR typing results.

Selection of a representative panel for whole genome sequencing is being implemented through a comprehensive genetic characterization of isolates; so far 1014 strains have been analysed using multilocus VNTR genotyping performed in Samara and in London; a capillary based RFLP typing methodology has been developed and is undergoing further optimization; optimisation of PCR-based assays for LSP analysis has been completed and RD150, RD105, and RD142 deletion mapping results are available or 292, 96, and 192 strains respectively.

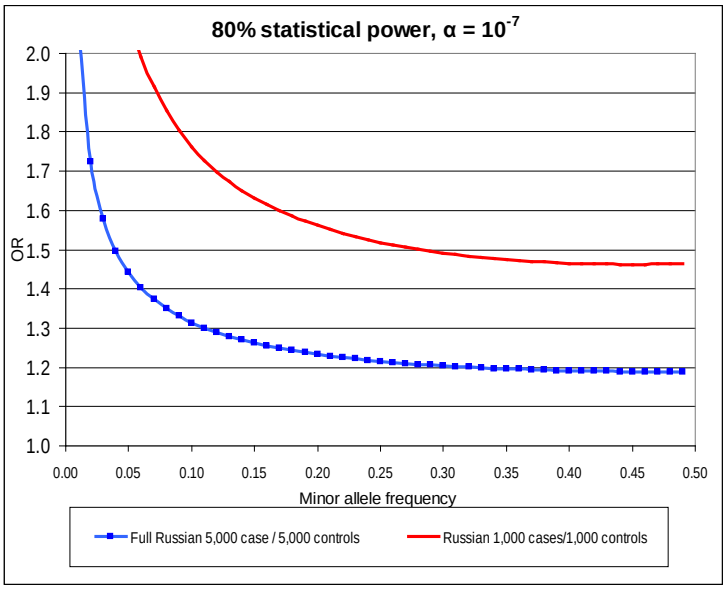

Figure 1

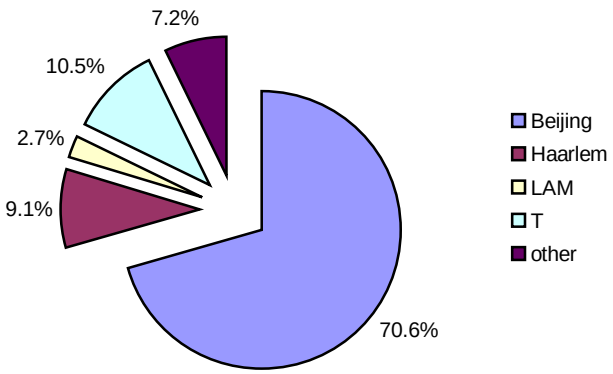

Figure 2