**SMP: DNA**

## Project: Re-sequencing and Mutation

**Richard Reinhardt - Ralf Sudbrak - MPI for Molecular Genetics, Berlin -**
**Matthias Platzer - Leibniz Institute for Age Research - Fritz-Lipmann-Inst., Jena - mplatzer@fli-leibniz.de**

### Introduction

The diversity of the human genome is a decisive factor in the causation and pathogenesis of complex disease as well as for the individual response to biologically active substances (pharmacogenomics).

A key step in all strategies for disease gene identification is the comparative sequence analysis of candidate genes in patients and controls to identify those specific sequence variations associated with complex disease. The two laboratories involved in this subproject (MPIMG, Berlin and FLI, Jena) have installed a roboter based automated high-throughput re-sequencing service platform open to all members, cooperating in NGFN (Fig.1, Fig. 2).



*Fig 1: Robotics for Automated Re-sequencing.*

This platform is specifically designed as a disease identification pipeline, to test hypotheses about the involvement of specific genes in inherited diseases, or to verify candidate genes identified by positional or functional cloning projects.



*Fig 2: Robotics for Sample Preparation*

We therefore have also implemented bioinformatic techniques (programs to predict haplotypes and haplotype block structures) to analyse complex genotype-phenotype relations. The need for such a service infrastructure has already been expressed by a number of researchers with a strong interest in candidate gene sequencing and gene-based haplotype analysis from the various clinical networks.

The coordination of the various sequencing requests from the network partners will be performed through our WEB site (http://www.resequencing.mpg.de/). This information is at all accessible to our partners and it should be obvious to all, what will be sequenced, what has been sequenced so far and who is the responsible person for the specific project.

### Project Status

**Sequencing Projects**

Currently, we are sequencing different candidate genes for complex diseases or pharmacologically relevant genes in cooperation with our clinical partners.

A successful project in the past was the analysis of the *DNAH5* gene in indiviuals with primary ciliary dyskinesia (PCD). PCD is characterized by recurrent infections of the respiratory tract due to reduced mucociliary clearance and by sperm immobility. Half of the affected offspring have situs inversus (reversed organs), which results from randomization of the left-right (LR) asymetry (Fig. 3). Sequence analysis in individuals with PCD with randomization of LR asymetry identified mutations resulting in non-functional DNAH5 proteins (Olbrich *et al.*, 2002).
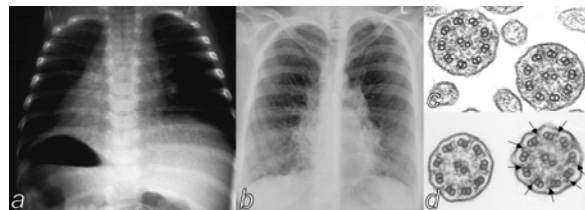


*Fig 3: a, Half of the PCD-patients show a complete situs inversus. b, Bronchiectasis. c, Electron micrograph of cross-sections of respiratory cilia from an affected individual of PCD-family. Absence of outer dynein arms is observed on all peripheral doublets whereas healthy individuals d, show normal dynein arms.*

Presently, we are analysing the whole DNAH5 gene (77 exons, 14 kb cds) in 94 individuals to identify the complete polymorphic spectrum of the gene. Until now we have found a very high genetic variability in this region, with a total of 132 SNPs and 12 indels. To indentify interactions of these variations we also use different algorithms to estimate haplotypes and haplotype block structures of the DNAH5 gene (Fig. 4.).



*Fig 4: LD and haplotype bloc structure across the DNAH5 gene. D' values for pairwise LD between each marker are represented (red).*

NGFN
Nationales
Genomforschungsnetz

Bundesministerium
für Bildung
und Forschung

Other projects at FLI were focussing on the genetic predisposition for chronic inflammatory diseases (Crohn's disease, sarcoidosis) and obesity. Detailed analysis were done for the candidate genes CARD15, BTNL2, MCHR1, MC4R, ghrelin and a gene desert on chromosome 16. Presently, the complex SNP-related sequence variation in the segmental duplicated beta-defensin cluster on 8p23 is determined with respect to inflammatory disorders.
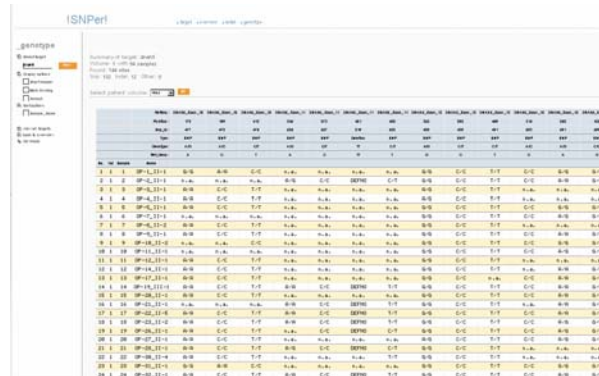


**Fig 5:** *User data access interface. Interface for our clinical partners to access sequencing data.*

### Quality Control

Presently we are working to enhance our quality management system for the whole re-sequencing process. The system integrates all steps in re-sequencing - from primer design, PCR amplification and cycle sequencing reactions, to capillary electrophoresis systems on the 3730xl DNA analyser and the data analysis. For all these steps we have established standard operating procedures (SOPs).

### Further EU-activities

The efforts to install a roboter based automated high-throughput re-sequencing service platform, specifically designed as a disease identification pipeline, that were initiated with the present proposal, lead to two funded EU 6 initiatives within the 6th framework program: **1**: STAR: a**S**NP and haplo-**t**ype m**a**p for the **r**at and **2**: NoE Marine Genomic Europe, (technical platform) The platform has been further expanded by the ability to work with plasmides, fosmids and BACs and additional bioinformatic techniques have also been implemented.

### Outlook

High-throughput re-sequencing of PCR products is the method of choice to identify unknown polymorphism. We offer our high-throughput re-sequencing service platform to all mem-

bers of the NGFN, to analyse candidate genes or genomic regions of interest.

Using nested PCR assays under stringent conditions, resulting amplicons up to 600 bp can be sequenced in both directions, to detect single nucleotide exchanges, insertions and deletions as well as micro-satellites and VNTRs.

Template preparation, purification and modification will be done using robotic systems developed for HT genomic sequencing projects. The number of patients and controls to be sequenced will depend on the allele frequency to be detected. Maximum numbers will range from 48 up to about 250 samples (alleles with an 1 % frequency). The data will be made accessible to the clinical partners (Fig. 5) and after detailed analysis will be made publicly available.

The Coordinator of the subproject is responsible for quality control of samples and communication with partners. The cost for this service pipeline was calculated to 0.045 Euro per sequenced base/sample, to be financed by the user requesting the service. This price included selection of primers, PCR amplification, sequencing reactions and data analysis. If exon-intron amplicons are already available, it is possible to send in the PCR products and primer for sequencing and data analysis (4.40 € per sequencing read). In this case we will work with the PCR products of our cooperation partners.

*Lit.: **1.** International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. (2004) 431: 931-945. **2.** Olbrich H, Häffner K, Kispert A, Völkel A, Volz A, Sasmaz G, Reinhardt R, Hennig S, Lehrach H, Konietzko N, Zariwala M, Noone PG, Knowles M, Mitchison HM, Meeks M, Eddie M.K. Chung EMK, Hildebrandt F, Sudbrak R & Omran H (2002) Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. Nature Genetics 30: 143-144. **3.** Olbrich H, Fliegauf M, Hoefele J, Kispert A, Otto E, Volz A, Wolf MT, Sasmaz G, Trauer U, Reinhardt R, Sudbrak R, Antignac C, Gretz N., Walz G., Schermer B, Benzing T, Hildebrandt F, Omran H. Mutations in a novel gene, NPHP3, cause adolescent nephronophthisis, tapeto-retinal degeneration and hepatic fibriosis. Nature Genetics (2003) 34: 455-459. **4.** Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, Nagy M, Gaede KI, Franke A, Haesler R, Koch A, Lengauer T, Seegert D, Reiling N, Ehlers S, Schwinger E, Platzer M, Krawczak M, Muller-Quernheim J, Schurmann M,Schreiber S. Sarcoidosis is associated with a truncating splice site mutation in BTNL2.Nat Genet. (2005) 37:357-64. **5.** Geller F, Reichwald K, Dempfle A, Illig T, Vollmert C, Herpertz S, Siffert W, Platzer M, Hess C, Gudermann T, Biebermann H, Wichmann HE, Schafer H, Hinney A, Hebebrand J. Melanocortin-4 receptor gene variant I103 is negatively associated with obesity.Am J Hum Genet. (2004) 74:572-81. **6.** Croucher PJ, Mascheretti S, Hampe J, Huse K, Frenzel H, Stoll M, Lu T,Nikolaus S, Yang SK, Krawczak M, Kim WH, Schreiber S. Haplotype structure and association to Crohn's disease of CARD15 mutations intwo ethnically divergent populations. Eur J Hum Genet. (2003) 11:6-16.*