HAEMOPHILIA - GENETIC AND CLINICAL STUDIES
Blödarsjuka – genetiska och kliniska studier

Research program

Research group and principal collaborators (www.phru.se):

ROLF LJUNG, MD, PhD, Professor of Paediatrics, Lund University, Department of Clinical Sciences Lund, Senior Consultant, Paediatric Clinic and Malmö Centre for Thrombosis and Haemostasis, Skåne University Hospital. Principal investigator and main applicant. The characterization of mutations in the F8/F9 genes, studies on mRNA and linkage analysis with microsatellites are done at the research laboratory of Rolf Ljung in Malmö.

CHRISTER HALLDÉN, PhD, Professor in Genetics, Dept. of Biomedicine, Kristianstad University, Kristianstad, Sweden. SNP-haplotyping, long-range PCR, droplet digital PCR, Sanger and next generation sequencing and bioinformatics are done at the research laboratory of Christer Halldén in Kristianstad. (Study 1-3)

ANNIKA MÄRTENSSON, MD, Consultant, Paediatric Clinic and Malmö Centre for Thrombosis and Haemostasis, Skåne University Hospital and PhD student, Lund University, Dept. of Clinical Sciences Malmö (planned dissertation 2015). (Study 2)

NADINE GRETKENKORT ANDERSSON, MD, PhD, Consultant, Paediatric Clinic and Malmö Centre for Thrombosis and Haemostasis, Skåne University Hospital. (Study 4)

ANNA LETELIER, Specialised laboratory technician employed 40/40 for the laboratory work in this project in Malmö. (Study 1-3)

CHRISTINA LIND-HALLDÉN, PhD, Assistant professor in Genetics, Dept. of Biomedicine, Kristianstad University, Kristianstad, Sweden. (Study 1-3)

ERIC MANDERSTEDT, PhD student, Dept. of Biomedicine, Kristianstad University, Kristianstad, Sweden. (Study 1-3)

INTERNATIONAL COLLABORATIVE NETWORK:

PEDNET, European Paediatric Network for Haemophilia management. A scientific network of 30 paediatric haemophilia centres in 22 European countries + Canada and Israel, founded 1997 by Rolf Ljung who is scientific chairman.

PEDNET REGISTRY comprises all children with haemophilia born since 2000-01-01 and diagnosed at a PedNet centre. High quality of data by regular monitoring according to GCP rules with 1110 entries (May 2014) with base line data and annual follow-up which will continue until 2019. It is the largest registry of paediatric haemophilia in the world. It is a database for research projects with the potential to correlate certain phenotype characteristics such as inhibitors with genotype in a large cohort.
Principal Investigators (PI) PedNet Registry: Rolf Ljung, Lund University, Sweden and Marijke van den Berg, University of Utrecht, The Netherlands.

INPH (International Network for Pediatric Haemophilia). An international scientific network of 15 prestigious paediatric haemophilia centres (Europe, Japan, USA and Australasia). Collaborating on the study on ICH in haemophilia (Intra Cranial Haemorrhage). Rolf Ljung is founder and scientific chairman since 2005.

Purpose and aims

The general purpose of this research program is to improve the care of patients with haemophilia and their families by translational studies on genetic and clinical issues.

AIM 1. Identification of the cause of haemophilia A in patients with mutations that are not identified using conventional mutation screening. Even after applying the full set of diagnostic tools available, some patients remain without an identified mutation. Although being a small group, these patients are of great principal interest since they may harbour mutations representing previously undescribed types of mutations and disease mechanisms. We aim to identify these hitherto unidentified mutations in the mutation-negative families by screening the whole F8 gene region for deletions, duplications and inversions by amplification of overlapping fragments using long-range PCR followed by characterization of amplification products deviating from the normal patterns. Patients where no deviations can be detected will be subjected to next generation sequencing of the pooled amplification products. The identified mutations will allow carrier- and prenatal diagnosis as well as prediction of risk of development of antibodies/inhibitors to replacement therapy with Factor VIII also in these families.

AIM 2. Determination of the frequency of mosaicism and the origin of mutations in sporadic cases. Haemophilia is not known in the family beforehand in 60% of cases, i.e. they are sporadic cases. What are the origins of these mutations and who else in the family may be at risk? Mutation screening and haplotyping will first define the mutation-carrying chromosomes and follow them in three generation pedigrees. Carrier and non-carrier status will then be confirmed using droplet digital PCR (ddPCR). The use of ddPCR for the detection of mosaics represents a paradigm shift in the analysis of mosaicism since it is highly sensitive and precise compared with all previous methods. Additional research objectives are to define the true frequency of sporadic cases and differences in sex-specific mutation rates. The improved carrier diagnosis in these families, i.e. the accurate detection of mosaics, will result in better genetic counselling and better indications for prenatal diagnosis.

AIM 3. Identification of genes and mutations contributing to the development of inhibitors following FVIII treatment. A most serious problem for the patient and an expensive complication in the treatment of haemophilia A is the development of antibodies in 30% to 40% of the patients after treatment with FVIII concentrates. Known predictors of development of antibodies are the type of mutation and treatment related factors, but is has also been argued that the haplotype of the haemophilia genes may be of importance. We aim to investigate the relationships between F8 mutations and haplotypes, phenotypes and the development of antibodies in a study using our Swedish DNA-bio bank that uniquely represent almost all families in the country. In a second step, 50 patients showing inhibitor development will be re-sequenced for their entire exomes. Candidate risk-associated variants will be identified through comparison with data from the Exome Aggregation Consortium.
harboring >60,000 exomes and with the results from the Hemophilia Inhibitor Genetics Study (HIGS). The eventual identification of risk-factors for inhibitor development may simplify the decision-making with regard to individualized introduction of FVIII concentrates and ultimately allow more efficient treatment regimens to be developed.

**AIM 4. Determination of the correlation of mode of treatment and risk of intracranial haemorrhage (ICH) in children with haemophilia A and B.** Treatment of haemophilia is one of the most expensive treatments in medicine and cost-effectiveness has almost exclusively been focused on joint outcome since joint bleeds are the typical haemorrhage. The present study aim to define the risk of ICH using different treatment approaches by longitudinally analysing 1400 children recruited from PedNet Registry and INPH. The aim is to identify optimal cost-effective treatments.

---

**Survey of the field / Background:**

Haemophilia A and B are hereditary, X-chromosomal recessive bleeding disorders caused by absent, deficient or dysfunctional factor VIII (FVIII) and factor IX (FIX), respectively. Depending on the concentration of FVIII or FIX coagulant activity in blood, the disorders may be classified as severe (<1% of normal activity), moderate (1-4%) or mild (5-25%). In Sweden we have approximately 750 patients (belonging to 375 families) with haemophilia A and approximately 170 patients (belonging to 100 families) with haemophilia B. Patients with clinically severe haemophilia receive prophylactic treatment with FVIII/FIX concentrate. Approximately 30-40% of the patients develop neutralising antibodies against FVIII and 20-25% against FIX during the first 50 exposures to FVIII/FIX, which is a most serious complication. The type of mutation influences the risk of inhibitor development. Characterisation of the type of mutation is thus an important aspect of modern haemophilia care and the basis for genetic counselling, carrier detection and prenatal diagnosis.

The F8 gene, which maps to Xq28, consists of 26 exons covering a genomic region of 186 kb and encoding a mature protein of 2332 amino acids. About 40% of all cases of severe haemophilia A are caused by an inversion created by homologous recombination involving a repeated sequence in intron 22 and related sequences outside the F8 gene (1,2). Sequencing of the 26 exons identifies a unique mutation in an absolute majority of patients, but still fails to identify mutations in about 5% of haemophilia A patients. In haemophilia B, the 33.5 kb F9 gene has 8 exons coding for a pro-peptide comprising 46 amino acids and the serine protease FIX with 415 amino acids. Almost all families have their own unique mutations (3).

Sweden (and soon Finland in collaboration with us) is the first country in the world that has a complete database of mutations causing haemophilia A and B in the population. Furthermore, we have during 2012-2014 created an extensive database of haplotypes covering the complete F8-F9 region in a subsample of 300 subjects from presumably independent families with haemophilia A, and in 100 subjects with haemophilia B and in 400 healthy controls by the use of a large collection of SNP and microsatellite markers (4,5). This is a new approach in haemophilia research and we are not aware of this being done previously on such a large scale on a complete and well-defined population. Combined use of these databases offers unique opportunities for further research and to address the scientific objectives in this project.

Prophylactic replacement of FVIII/FIX is the optimal treatment regimen for children with haemophilia. However, in most countries, due to the cost of treatment, the majority of children still receive on-demand treatment. The discussion on choice of treatment is very
much focused on the joint outcome although intracranial haemorrhage (ICH) is a significant cause of death and morbidity in haemophilia treated on-demand. There are currently no conclusive studies correlating the model of prophylactic treatment to the risk of ICH. Many countries are investigating the optimal “cost-effective - cost-benefit” prophylactic model and good quality data on serious haemorrhage is needed in the discussion.

Three of the four studies described in this research proposal involve the use of next generation sequencing or digital PCR that both represent new techniques that have not been previously applied within haemophilia research. Both techniques can be said to allow investigations that have simply not been possible to do before their introduction. Next generation sequencing will undoubtedly be used in clinical screening for mutations in single genes, sets of candidate genes and also in whole exomes or genomes in the coming years. This will allow comparatively cheap simultaneous screening of many disease-causing and modifying genes in different diseases and may change the strategies for screening of mutations in many cases. Digital PCR will revolutionize the screening for mosaics in the same way as it has already revolutionized the screening for rare mutations in the treatment of cancer.

Project description/Working plan

AIM 1. Identification of the cause of haemophilia in patients with mutations that are not identified using conventional mutation screening.

Since the F8 gene is 186 kb with >90% intronic sequence containing several highly homologous repeated sequences, a possible explanation for the missing mutations is the presence of structural rearrangements such as deletions, duplications and inversions in regions of the F8 gene that are not normally investigated for mutations (deep in introns). Several recurrent inversions in introns 1 and 22 have been previously described in addition to a number of unique structural rearrangements (Factor VIII variant database). To search for such rearrangements in the F8 gene region, a set of overlapping amplicons will be generated using long-range PCR (LRPCR). These overlapping 6-12 kb fragments will be amplified in male patients with undetectable mutations using patients with known intron 1 and intron 22 inversions and healthy individuals as positive and negative controls, respectively. Rearrangements are indicated by the lack of amplification of particular fragments. The candidate rearrangements identified will be further characterized with regard to their breakpoints using inverse PCR and LRPCR with a design derived from the predicted tentative structures. Subsequent restriction enzyme mapping and sequencing of such amplicons will allow a detailed description of the rearrangements. Potential copy number alterations will be confirmed using quantitative PCR. Patients that remain unexplained with regard to their disease mechanisms and mutations will be analysed by next generation sequencing using pooling of LRPCR fragments and individual barcoding. This will produce complete sequences for a 200 kb genomic region covering the F8 gene in these individuals and hopefully allow identification of the remaining mutations using comparisons with 1000Genomes data and other available datasets.

Methods and equipment. Long-range PCR will use a long-range enzyme mix from KAPA and high resolution agarose gels. Sequencing of deletion/inversion breakpoints will use Sanger-based sequencing and will be performed using BigDye terminator sequencing on an ABI 3130XL capillary sequencer (Life Technologies). SeqScape Software will be used to analyse the data. Copy number variants will be characterized using QPCR analysis on a CFX384 from BioRad with CopyCaller Software. Efficient robotic systems for analysing QPCR and LRPCR using low-volume 384-based PCR have been optimized using Beckman and Hydra robots.
Next generation sequencing will be performed on an Ion Torrent PGM sequencer after pooling and barcoding of LRPCR products. Fragmentation will use a Bioruptor NGS from Diagenode and sizing of fragment populations will be made on a Fragment Analyzer from Advanced Analytical. All equipment for performing the studies is available in our own laboratories in Malmö and Kristianstad and all processes have been validated.

**AIM 2. Determination of the frequency of mosaicism and the origin of mutations in sporadic cases.**

The frequency of mosaicism in haemophilia A has been analysed previously using allele-specific PCR (6). In total 13% of the 61 analysed families showed the presence of somatic mosaics. These mosaics were of varying degrees (0.2% - 25%) and were confirmed by a mutation enrichment procedure. They were present exclusively in the 32 families with point mutations (25%). The hypothesis in the present study is therefore that many of the seemingly non-carriers are partial carriers (mosaics) and thus at risk of having more children with haemophilia, *i.e.* there may be an indication for prenatal diagnosis.

Sporadic cases of haemophilia constitute 60% of all cases. Our recent studies of Swedish families with identical mutations show whether the mutations are recurrent mutations, *i.e.* are unrelated new mutations, or are identical by descent, *i.e.* share a common origin. This information will allow a 100% correct identification of a sporadic case and thus a study group without selection bias. Haplotyping is made through extensive genotyping of all individuals using large sets of SNP and microsatellite markers. In the present study the parental origins of the mutations will be determined in approximately 50 families by analysis of blood collected from 3 generations. Haplotyping will define the mutation-carrying chromosome in all cases and droplet digital PCR (ddPCR) will be used to investigate for the presence or absence of the mutation in the mother. If the mother is a carrier her parents will also be investigated. ddPCR uses allele-specific probes specific for the mutation in the family and will allow detection of very low proportions of the mutation. ddPCR is far more sensitive than all previous techniques used for detecting mosaics because ddPCR relies on the detection of the target molecules in very large populations of droplets where the individual droplets carry only very few copies of the target molecules. Thus, the mutations no longer constitute a minute fraction of the investigated targets and are readily detected by TaqMan assays for the two alleles. Also the inversions can be analysed using this approach since TaqMan assays probing the single copy flanking regions of the inversions will give rise to a signal in the same droplet also in droplet collections where limiting dilution has created droplet populations with only 0 or 1 target molecules present in each droplet. Our pilot study clearly showed that ddPCR allows detection levels in the range 0.01% and lower for all of these types of mutations, which is much more sensitive than previous studies using older techniques (6) The use of ddPCR for detection of mosaics represent a paradigm shift in the analysis of mosaicism since it is highly sensitive and precise compared with previous methods. The Swedish material will be screened for mosaics in 2016 and, pending the results of this study, can be expanded to include Denmark and Finland the coming years.

**Methods and equipment.** Microsatellite genotyping will be performed using an ABI 3130XL capillary sequencer and the data analysed using GeneMarker Software (SoftGenetics, USA). SNP genotyping will be using the Sequenom MassARRAY MALDI-TOF system and will be made at the Mutation Analysis Facility at Karolinska Institutet. Haplotyping will be performed using Phase and Haplovieview and compared with 1000Genome data. Droplet digital PCR uses the QX100 system from Bio Rad and sets of FAM- and VIC-labelled TaqMan probes for the mutation and wild-type alleles, respectively. All equipment for performing the
studies is available in our own laboratories in Malmö and Kristianstad and all processes have been validated.

**AIM 3. Identification of genes and mutations contributing to the development of inhibitors following FVIII treatment.** Inhibitor development is seen in 30% to 40% of the patients after treatment with FVIII concentrates. Known predictors of inhibitor development are the type of mutation and treatment related factors, but is has also been argued that the haplotype of the haemophilia genes may be of importance. We aim to investigate the relationships between F8 mutations, haplotypes, phenotypes and the development of inhibitory antibodies in a study using a majority of the Swedish haemophilia A patients. In a second step, 50 patients who have developed antibodies will be re-sequenced for their entire exomes. This screening will use an AmpliSeq-based enrichment of coding sequences and exome sequencing on an Ion Torrent Proton next generation sequencer. Confirmatory re-sequencing of interesting genes will be performed using Sanger sequencing or Ion Torrent PGM sequencing after selection of candidate genes using an AmpliSeq strategy. Interesting candidate mutations will be evaluated using frequency comparisons between patient data and 1000Genomes and Exome Aggregation Consortium data harboring >60,000 exomes. The results will also be compared with the results of the Ilemophilia Inhibitor Genetics Study (7). Our study will subsequently be expanded to include an even larger study group from the PcdNet Registry (potential of >700 children with severe haemophilia A of whom 32% developed inhibitor). It deserves to be mentioned that exome data is comparable between studies in much the same way as for example SNP chip data from genome-wide association studies can be compiled and evaluated in meta-studies. This means that even if we do not know at present what power we need to identify the factors responsible for inhibitor development in haemophilia A we can nonetheless confidently start analysing also comparatively small materials since exome data can be merged and compared at subsequent occasions. The eventual identification of risk-factors for inhibitor development may simplify the decision-making with regard to individualized introduction of FVIII concentrates and ultimately allow more efficient treatment regimens to be developed.

**Methods and equipment.** SNP genotyping will be made at the Mutation Analysis Facility at Karolinska Institutet and haplotyping will be performed using Phase and Haploview. Exome sequencing will use an AmpliSeq-based enrichment of coding sequences and sequencing on an Ion Torrent Proton next generation sequencer at the National Genomics Initiative, Uppsala. Re-sequencing of candidate genes will be performed using Sanger sequencing or Ion Torrent PGM sequencing after selection of candidate genes using an AmpliSeq strategy. Except where stated otherwise, all equipment for performing the studies is available in our own laboratories in Malmö and Kristianstad and all processes have been validated.

**AIM 4. What is the correlation of mode of treatment and the risk of IntraCranial Haemorrhage (ICH) in children with haemophilia?**

Hypothesis; children on regular prophylactic therapy rarely develop intracranial haemorrhage in comparison with children on-demand therapy or once/week prophylaxis; ICH in children on prophylaxis is caused by a combination of trauma and low factor VIII/IX concentrations; ICH in children on prophylaxis cause less sequel.

**Methods:** Study group: Children (n >1400), age <18 years, with severe haemophilia A or B (FVIII/IX <1%), without inhibitors who are stratified in 3 treatment groups (prophylaxis with >20 U/kg, 2 or more times/week, prophylaxis once/week or on demand treatment). The study is a 3-year prospective survey of the cohorts with a 5-year retrospective part. The study is
initiated and coordinated from our group in Malmö (Dr Gretenkort Andersson) - 31 international centres from PedNet (the European Paediatric Network for Haemophilia Management) and INPH (the International Network for Pediatric Hemophilia) are participating. Ongoing (www.ich-hemophilia.se) – prospective part finished by 2017.

**Significance/Clinical Significance:**

This research program has a “translational structure” since it bridges from basic scientific questions to clinical applications, combines information in the Swedish database with the large PedNet Registry and combines research at Lund University with Kristianstad University.

We have as the first country in the world characterised the mutations causing haemophilia A or B in the whole Swedish (and soon Finnish) population. We are not aware of any other research group which has systematically established an extensive SNP/microsatellite haplotype database of individuals with haemophilia and defined if individuals with the same mutation belong to the same family or are true *de novo* mutations. The combination of data offers unique possibilities to advance the research field and address several scientific questions on association between genotype/phenotype and clinical issues such as inhibitors.

Studying the pathogenic mechanisms in cases without detectable mutations in coding sequence has the potential to reveal novel mechanism that causes haemophilia A as well as enabling carrier- and prenatal diagnosis in these families. The experience we gain with next generation sequencing can be used to modify the current *F8/F9* gene diagnostics and possibly also include other bleeding disorders in a “gene screen”.

The frequency of mosaicism and the origin of mutation in sporadic cases will be of great importance in our genetic understanding of the disorders and, translated to the clinic, add valuable knowledge to the genetic counselling and genetic diagnosis in the families for the future. It will also be of particular interest in the developing world which has no access to DNA-diagnostics of haemophilia, *i.e.* if our hypothesis on origin of mutation is correct a simple pedigree analysis will make DNA-analysis of many members in these families unnecessary.

New knowledge is needed to address the serious clinical problem of inhibitor development to FVIII/FIX factor concentrates. We have access to suitable study groups within the Swedish and Finnish cohorts and the PedNet Registry cohort and have an established methodology for analysis of re-sequence data. The HIGS (Astermark et al. 2013) identified common variants in a set of 13 genes as associated with the development of inhibitors. The whole-exome re-sequencing described in the present proposal will complement the results of that study and may potentially find support for their results from accumulation of rare variants in the set of genes identified using common variants. The exome data gathered will also represent a resource that can be used in all future comparisons involving exome data in an inhibitor development perspective. The potential identification of rare variants with large effects will support the ongoing unravelling of this serious clinical problem.

Treatment with FVIII/FIX concentrates is extremely expensive and cost-effectiveness has a high priority and has recently specifically been addressed by the Swedish TLV (Tandvårdsoch LäkemedelsförmånsVerket). New knowledge of the impact of different treatment
regimens on serious haemorrhage such as ICH will be important in revision of treatment programs.

**Preliminary results:**

Mutations have been characterized in almost all haemophilia A and B patients in Sweden as part of a previous VR/ALF supported project. We have established a unique national database of mutations and associated phenotypic and clinical data – a prerequisite for studies on aim 1+2. We have published a study on 9 patients with haemophilia A without detectable mutation after sequencing of all coding regions of the F8 gene (8). MLPA analysis detected a duplication of exon 6 of the F8 gene in one patient. Four severe cases had different F8 haplotypes and gave no detectable RT-PCR product from different regions of the F8 mRNA, indicating failed transcription. LRPCR systems covering the complete F8 gene region have been optimized and validated using DNA from 12 control individuals. Screening of the mutation-negative patients has so far identified aberrant LRPCR amplification in 3 families. One of these was the exon 6 duplication also detected by MLPA in a previous study. Another may be identical to the rearrangement identified by Pezeshkpoor et al. (2012), whereas the third is compatible with an inversion involving repeated sequences in introns 13 and 22. We have also created a haplotype database with SNP and microsatellite data on a large subsample of the subjects included in the hemophilia A and B mutation databases. From these data it has been possible to distinguish between families that have identical mutations and determine whether the mutations are recurrent (RM) or identical by descent (IBD) (4,5). A pilot study using ddPCR of 5 different F8 mutations and artificially created mixtures of wild-type and mutant alleles has confirmed that the use of ddPCR for detection of mosaics is highly sensitive and precise. AmpliSeq-based re-sequencing using Ion Torrent next generation sequencing has produced high quality data in a number of projects performed in the Kristianstad lab. E.g. 288 patients with von Willebrand disease type 1 (VWD1) have been analysed for mutations in the VWF gene. A set of 20 candidate VWD1 modifying genes is currently under analysis in the same set of individuals. The lab has more than 3 years of experience of managing and analysing next generation re-sequencing data. A database of children (with 1400 entries Jan 2015) with haemophilia on different treatment regimens have been set up as a base for the study in aim 4 (www.ich-hemophilia.se).

**Equipment - Need for infrastructure - International and national collaboration**

As shown in the other sections we have the necessary equipment, infrastructure and collaborations for the planned studies.

**References:**


