EURATRANS Report Summary

Project ID: 241504
Funded under: FP7-HEALTH
Country: Germany

Final Report Summary - EURATRANS (European large-scale functional genomics in the rat for translational research)

Executive Summary:
Genetic studies in model organisms and humans, including human genome-wide association studies, have pinpointed genomic regions that contribute susceptibility to common disease. Prior to EURATRANS these data have provided limited insights into the genes, molecular pathways and mechanisms underlying disease pathophysiology.

EURATRANS used state-of-the-art and emerging large-scale technologies and advanced computation in an expanded multi-disciplinary approach to identify gene networks underlying common diseases. We used the rat as model organism to identify the major gene pathways for human inflammatory, cardiovascular and metabolic, and behavioural disorders. Our programme used next-generation sequencing technologies to generate genomic datasets and analyse the transcriptome across tissues in specialised rat resources. Genomic data was integrated with cutting-edge, quantitative metabolomic, proteomic, and epigenomic datasets giving significant depth of coverage across molecular components of the gene, the cell, the organ and the organism. We gathered, annotated and integrated these datasets in relational and dynamic models that can be used for comparative analysis to understand human gene function. These studies revealed new biology enabling the understanding of human genes in the context of functional genomic networks that will result in comparative medicine.

We incorporated our data into existing and new databases and establish repositories for models, reagents, discovery tools and specimens available to the scientific community. Our approach enables us to decipher gene function at large scale. Because we focus on concepts and networks (rather than on single genes) we produce results with relevance to fundamental mechanisms. Our results will contribute strategies to combat common diseases affecting human inflammatory, cardiovascular/metabolic, and behavioural disorders through identification of novel disease relevant networks that represent targets for diagnosis and drug therapy. Our future aim is to encourage rapid exchange of information between research fields and their immediate application in medicine and biotechnology.

Project Context and Objectives:
In order to understand disease and disease susceptibility and to be able to interpret personal genomic information, a systematic understanding of the functional elements in a genome and the effects of genetic variation on these elements is required. The goal of EURATRANS was to enable comparative approaches that allow investigators to understand physiological function based on conserved genetic pathways among rat models and humans in health and disease. The rat is a pivotal resource for these studies as it has been studied across the biomedical sciences and has been used for research into a broad array of human conditions. The primary motivation of this project was to leverage the deep biological history of the rat towards an in-depth understanding of the pathogenesis of common human disease of high prevalence in the EU and worldwide. Our project capitalized on the resources and technologies developed by the EU FP6 EURATools consortium and our U.S. and Japanese partners and applied those advances, coupled with further innovation across a spectrum of complimentary high throughput technologies, to the study of disease mechanism. Rather than focusing on individual genetic factors independently, we studied disease mechanisms at the level of gene networks in order to identify the pathways and networks that prominently...
influence common complex disorders. This provided a more complete mechanistic understanding rather than a snapshot view on individual factors contributing to disease. We undertook a multilevel approach (Figure 1) to identify the major regulatory pathways in selected rat models of cardiovascular/metabolic, inflammatory, and behavioural phenotypes and translate our findings to progress the understanding of common human diseases.

Enabling technologies are overarching priorities of the programme, which focused on the application of novel technologies and the development of novel strategies for a large-scale multi-disciplinary functional genomics programme in the rat. Complementary technologies included: 1) Next-generation sequencing technology and analysis of the generated data, 2) transcriptome analysis, 3) genomic variation analysis (SNP and CNV), 4) methylation-sequencing & ChIP-seq, 5) quantitative proteomics, 6) metabolomics, 7) advanced databases, 8) building regulatory networks, 9) germline manipulation of the rat genome and rat ES cell technology, 10) pathophysiology of cardiovascular/metabolic, inflammatory, and psychiatric/behavioural phenotypes, 11) translation to humans and diagnosis & therapy, 12) consistent data processing and integration.

Biological concepts: We used enabling technologies to generate large-scale multilevel datasets in recombinant inbred (RI) rat lines in which multiple phenotypes can be accumulated, combined with high resolution mapping in heterogeneous stock (HS) and congenic lines where appropriate. We aimed to extend and complement the databases of gene expression and physiological phenotypes that were generated in the EURATools programme by generating transcriptome inventories, and develop new metabolomic, proteomic, and epigenomic datasets in RI and selected HS/congenic animals. Furthermore, we planned to map the genetic determinants of these new phenotypes to the genome using existing high-resolution maps based on EURATools-generated SNP and structural variation databases, permitting identification of the cis and trans-regulatory control loci of these phenotypes. Using these multi-modality QTL datasets shall identify the most important pathways leading to clinically relevant phenotypes.

To maximise the general application of our multilevel network approaches we focused our efforts on three disease research areas in which the rat model has traditionally played a major role and in which rat biology and physiology facilitates translation to humans. Those are: i) cardiovascular/metabolic disease, ii) inflammatory disorders, and iii) psychiatric disorders. We prioritized key components in disease-related networks for functional validation by loss and gain of gene function in vivo and in vitro. To this end, we aimed at utilizing large existing and emerging resources to validate our findings: ENU-induced mutant archives, transgenesis, germline-manipulation of the rat genome by transposon-mediated mutagenesis, shRNA-mediated gene knockdown and nuclear-zinc-finger approaches and novel rat ES-cell technology.

Project Results:
Pillar 1: Large Scale Data Generation

WP 1.1 Genome Sequencing

The primary objective of this workpackage was to create a comprehensive inventory of genetic variation for the founder strains of the commonly used recombinant inbred (RI) and heterogeneous stock (HS) rat model systems. This data would form the basis for integrated genetical-genomics analyses using data generated in the other workpackages. The original objective was to analyze 9 inbred strains using whole genome sequencing. Driven by developments in next-generation DNA sequencing technologies, we have been able to extend these analyses to 40 strains and substrains in total (Figure 2). For all strains, we have created and disseminated inventories of single nucleotide variants and of structural variants. All data has been deposited in public archives and is available in the rat genome database. Finally, we explored whole genome sequencing reads that were not mapped to the reference genome and de novo assembly to identify non-reference and strain-specific genomic segments. Many thousands of such segments were identified, including segments containing coding sequences and orthologs present in other species.
WP 1.2 Transcriptome inventory - RNA sequencing

As main results of WP1.2 we report on the identification of 314 mature miRNAs in rat left ventricle and 362 mature miRNAs in rat liver, respectively. We de-novo identified 66 and 84 miRNAs in left ventricle and liver, respectively. In addition, we report on the identification of regulatory regions affecting miRNA expression in both tissues and conclude that the expression of about 20% and 16% of all expressed mature miRNAs in left ventricle and liver (including de-novo identified) is under genetic control by one or more eQTLs, with most miRNAs in both tissues being regulated in trans. A comprehensive inventory of all expressed miRNA genes in left ventricle and liver has been deposited in the Euratrans dropbox. This inventory includes all de-novo predicted rat miRNA hairpin sequences identified by the miRDeep2 algorithm. We found 30% of RNA-seq eQTL which not only affect the global gene expression level, but also cause differential ATEs in their target genes. In the RNA-seq eQTL, we observed a large number of last exons with ATEs (12%) which is also more than expected (OR=1.6 P=0.005 FET). These results imply that alternative polyadenylation events are a common mechanism to regulate eQTL transcripts.

We have identified 90 liver specific and 80 heart specific miRNAs (FDR < 0.05 log2(FC) > 1). We identified 58 and 45 differentially expressed miRNAs in left ventricle and liver at a genome-wide significance (FDR ≤ 0.05). Three miRNAs were common to both tissues, whereas 55 miRNAs in the left ventricle and 42 miRNAs in the liver data set showed a tissue-specific
differential expression. In total, 32 miRNAs in the left ventricle were up regulated in the SHR strain, with an average fold change of 2.4 and the strongest expression difference for miR-190b with 8.0 fold. 26 miRNAs were significantly lower expressed in the SHR strain with an average fold change of 1.7 and a maximal down-regulation by 4.3-fold of a de-novo detected miRNA, novel-20*. In the liver data set, 24 miRNAs were up- and 21 were down regulated in the SHR strain. On average miRNAs were 4.2-fold up- and 2-fold down regulated. The two most strongly up-regulated miRNAs with 14 fold higher expression in SHR were novel-11 and novel-18 whereas mno-let-7e* was the most strongly down regulated with 12-fold difference between the strains. 

Finally, we detected 79 and 52 QTLs with effect on the expression of 79 and 71 mature miRNAs in the left ventricle and liver dataset (FDR < 0.05). 89 % of the eQTLs (n= 70) in left ventricle were found to regulate miRNA expression in trans, whereas seven loci (~9%) harbored cis-acting variations (cis-eQTL). In the liver dataset, 96% (n= 50) of all 52 identified eQTLs were trans-regulatory, whereas only one locus exerted a cis-acting control on miRNA expression. Correlation analysis of normalized deep-sequencing counts with qRT-PCR values revealed that expression of six out of eight validated miRNAs was significantly correlated between the two technical platforms with Spearman correlation coefficients between 0.45 - 0.65 at P < 0.05.

In addition, we have generated strand-specific RNA-seq data from four DA rat ES cell lines and for comparison four non-obese diabetic (NOD) mouse ES cell lines. To identify the equivalent rat genes, we downloaded the gene models and rat-mouse orthologs from the Rat Genome Database. We initially restricted our analysis to genes with one-to-one orthologs. This set contained 14168 genes, compared to a total of 16,480 annotated rat genes, and 22,692 mouse genes. However, we observed that important mouse genes involved in pluripotency had no annotated ortholog in rat. These included Zpf42(Rex1), Sall4 and Tdgf1. To extend our set of orthologs, we took advantage of the depth of our coverage to perform de novo transcriptome assembly using Trinity. To identify the mouse orthologs of the assembled transcripts, these were mapped to the mouse genome with Blat. Using strict mapping criteria to avoid false positives, 4,380 transcripts could be mapped uniquely to mouse loci (Figure 4a,c). 78% of the inferred orthologs were already present in the annotation (Figure 4c). In all these cases, the inferred orthology was in agreement with the annotated one. Despite our stringent cut-offs, the correct ortholog was found for 58% of all genes expressed in our samples (FPKM>10), which improved to 69% when we considered highly expressed genes (FPKM>100). This shows our methodology for de novo ortholog identification is both robust and sensitive, and suggests a potential general use for our approach in transcriptome comparisons between well-annotated species and related organisms with limited annotation or without a reference genome.

We identified 291 new orthologs between known genes, and 616 non-annotated rat genes, orthologous to known mouse genes. Of these new rat genes, 91% (563) could be placed on the rat genome, including Zfp42, an important marker of naïve pluripotency (Figure 4a,b). However, 55 transcripts that matched an annotated mouse gene could not be found on the rat assembly, as was the case for the rat Sall4 ortholog.

Our results show that rat ES cells possess the hallmarks of naïve pluripotency, and that transcriptional differences between mouse and rat reflect species-specific growth properties and behaviour in culture. Intriguingly, rat ES cells overexpress repressive epigenetic modifiers, especially the genes involved in de novo DNA methylation. Currently we are doing more analysis and experiments to determine whether this differential expression of genes is reflected in a difference in the genome-wide methylation state of rat and mouse ES cells.


WP 1.3 Defining transcriptional initiation complexes and epigenetic modifications
This Workpackage aimed to define the genome-wide location of RNA polymerase complexes and the chromatin and DNA methylation state of the rat genome in select tissues and cell types. The cartography of DNA-associated proteins at the genome-wide level has been performed thanks to the development of Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) in rat tissues (Figure 5). The state of DNA methylation as an additional process influencing chromatin structure has been investigated. Novel experimental strategies to capture the dynamic interplay between chromatin and DNA sequence in complex traits have been developed. Alternative chromatin states (epialleles) have been mapped using recombinant inbred lines. Moreover, the data have been integrated with the transcriptome data (WP 1.2) and used for the building of molecular gene regulatory networks (WP 2.1).

In addition, genome-wide mapping of the DNA and histone epigenetic modifications provide precious information concerning the impact of the genome, epigenome and transcriptional activity on (patho) physiological phenotypes in the rat. Comparing the maps with the inventory of transcripts in various rat crosses allowed a detailed description and annotation of the expressed rat genome. Further we addressed the challenging assumption that DNA sequence variants are not the only source of heritable phenotypes.

<Figure 5: Variations of H3K4me1 and H4K20me1 histone marks in BN and SHR rats>

Natural variation of histone modification and its impact on gene expression in the rat genome

Histone modifications are epigenetic marks that play fundamental roles in many biological processes including the control of chromatin-mediated regulation of gene expression. Little is known about interindividual variability of histone modification levels across the genome and to what extent they are influenced by genetic variation. We annotated the rat genome with histone modification maps, identified differences in histone trimethyl-lysine levels among strains, and described their underlying genetic basis at the genome-wide scale using ChIP-seq in heart and liver tissues in a panel of rat recombinant inbred and their progenitor strains. We identified extensive variation of histone methylation levels among individuals and mapped hundreds of underlying cis- and trans-acting loci throughout the genome that regulate histone methylation levels in an allele-specific manner. Interestingly, most histone methylation level variation was trans-linked and the most prominent QTL identified influenced H3K4me3 levels at 899 putative promoters throughout the genome in the heart. Cis-acting variation was enriched in binding sites of distinct transcription factors in heart and liver. The integrated analysis of DNA variation together with histone methylation and gene expression levels showed that histone QTLs are an important predictor of gene expression and that a joint analysis significantly enhanced the prediction of gene expression traits (eQTLs). Our data suggest that genetic variation has a widespread impact on histone trimethylation marks that may help to uncover novel genotype-phenotype relationships.

Genetic analysis of the cardiac methylome at single nucleotide resolution in a model of human cardiovascular disease.

Epigenetic marks such as cytosine methylation are important determinants of cellular and whole-body phenotypes. However, the extent of, and reasons for inter-individual differences in cytosine methylation, and their association with phenotypic variation are poorly characterised. Here we present the first genome-wide study of cytosine methylation at single-nucleotide resolution in an animal model of human disease. We used whole-genome bisulfite sequencing in the spontaneously hypertensive rat (SHR), a model of cardiovascular disease, and the Brown Norway (BN) control strain, to define the genetic architecture of cytosine methylation in the mammalian heart and to test for association between methylation and pathophysiological phenotypes. Analysis of 10.6 million CpG dinucleotides identified 77,088 CpGs that were differentially methylated between the strains. In F1 hybrids we found 38,152 CpGs showing allele-specific methylation and 145 regions with parent-of-origin effects on methylation. Cis-linkage explained almost 60% of inter-strain variation in methylation at a subset of loci tested for linkage in a panel of recombinant inbred (RI) strains. Methylation analysis in isolated cardiomyocytes showed that in the majority of cases methylation differences in cardiomyocytes and non-cardiomyocytes were strain-dependent, confirming a strong genetic component for cytosine methylation. We observed preferential nucleotide usage associated with increased and decreased methylation that is remarkably conserved across species, suggesting a common mechanism for germline control of inter-individual variation in CpG methylation. In the RI strain panel, we found significant correlation of CpG
methylation and levels of serum chromogranin B (CgB), a proposed biomarker of heart failure, which is evidence for a link between germline DNA sequence variation, CpG methylation differences and pathophysiological phenotypes in the SHR strain. Together, these results will stimulate further investigation of the molecular basis of locally regulated variation in CpG methylation and provide a starting point for understanding the relationship between the genetic control of CpG methylation and disease phenotypes.

histoneHMM: Differential analysis of histone modifications with broad genomic footprints
The establishment and comparison of genome wide maps of histone modifications was one of the goals of the project. However, comparative analysis of samples remains challenging for histone modifications with broad domains, such as heterochromatin-associated H3K27me3, as most ChIP-seq algorithms are designed to detect well defined peak-like features. To address this limitation we introduce histoneHMM, a powerful bivariate Hidden Markov Model for the differential analysis of histone modifications with broad genomic footprints. histoneHMM aggregates short-reads over larger regions and takes the resulting bivariate read counts as inputs for an unsupervised classification procedure, requiring no further tuning parameters. histoneHMM outputs probabilistic classifications of genomic regions as being either modified in both samples, unmodified in both samples or differentially modified between samples. We extensively tested histoneHMM in the context of two broad repressive marks, H3K27me3 and H3K9me3, and evaluated region calls with follow up qPCR as well as RNA-seq data. Our results show that histoneHMM outperforms competing methods in detecting functionally relevant differentially modified regions.

Characterization of antibodies against rat nuclear factors
Antibodies against human transcription factors and nuclear receptors produced within the Human Protein Atlas have been validated on Western blot and immunofluorescence microscopy using rat cell lines. Western blot validation was performed for 677 HPA antibodies using nuclear fractions from 9 rat cell lines originating from all 3 germ layers. Based on specificity and signal intensity, 10 antibodies were selected for their potential interest for ChIP-ChIP or ChIP-seq applications. All selected antibodies reveal a nuclear and/or cytosolic staining pattern that supports immunolabeling of transcription factors. In many cases, staining patterns in rat cell-lines is similar to staining patterns observed in human cell lines. The antibodies were further validated by immunocytochemistry.


WP 1.4 Quantitative proteome analysis
In recent years, mass spectrometry-based proteomics has made great technological progress and is increasingly applied in cell culture–based studies. Using the stable-isotope labelling by amino acids in cell culture (SILAC) method, cell lines are isotopically labelled through incorporation of stable ‘heavy’ versions of essential amino acids in the cell populations. This allows for a quantitative comparison of the protein expression levels between investigated cell lines. However, since the SILAC methodology requires complete metabolic labelling of the entire proteome it is usually limited to analysing cell culture, and therefore not fully for quantitative analysis of tissue samples. To address this limitation we have established an in vitro SILAC technology referred to as Super-SILAC, based upon a variety of SILAC-labelled rat cell lines, which we have employed as internal standard for comprehensive rat tissue proteome quantification.
For the eight selected cell lines we assured that the proper SILAC incorporation level was reached (>96%), and that the individual protein expression profiles were properly characterized and broadly represents the entire rat proteome. To this end, we characterized the expression profiles for all cell lines used and hereby identified a total of 8,360 proteins. Demonstrating that the chosen cell lines broadly covers the protein coding genes in the rat proteome, and thus constitute a proper basis for a tissue-wide internal standard (Super SILAC mix).

During the course of the EURATRANS project, we have isolated 10 different tissue types (Left ventricles, liver, adrenals, kidney, spleen, peritoneal fat, aorta, brain, soleus muscle, and brown fat) derived from each progenitor strain (BN-Lx and SHR) as well as the same ten tissue samples across 30 RI strains and from 4 male rats per RI strain, collectively more than 1,200 rat tissue samples.

Using the Super SILAC strategy we have quantitatively analysed the protein expression profiles for all ten tissue types across the progenitor strain, and the specific protein expression profile for liver and left ventrical across 30 RI strains. As validation of the quantification accuracy provided by the Super SILAC approach, we performed replicate analyses of several liver samples derived from the same RI strain but different animals. This is in order to establish whether intra-strain reproducibility is similar to the already confirmed technical reproducibility derived from replicate sample analysis. To this end we compared the protein expression profiles of two BxH2 liver samples from two different animals (same strain but different animals), and two HxB1 liver samples from similarly from two different animals (but same strain). A strong Pearson correlation for these experiments (>0.95) signifies high reproducibility of our established methodology, and confirms that intra-strain comparisons indeed is highly reproducibly and comparable to the obtained technical reproducibility. In contrast, the protein expression profiles between liver samples from BxH2 and BxH1 (different strains) only yields a Pearson correlation of 0.90 while the correlation between RI and parental strains is >0.80 (data not shown).

Having established that the Super SILAC approach is highly reproducible, we set out to quantify rat tissue samples across 30 RI and progenitor strains (Deliverables D40 and D41). Collectively these analysis yielded identification of more than 12,500 rat specific proteins across the different analysed tissues. To obtain such a comprehensive rat proteome profile, we acquired more than 1,000 LC-MS experiments.

Besides tissue-wide quantification, partners in WP1.4 have conducted an in-depth proteomic analysis of liver-specific protein expression in parental strains. Based on whole genome (WP1.1) and liver transcriptome (WP1.2) sequencing of the BN-Lx and SHR strains, we generated a novel rat protein reference database, including strain-specific peptides, gene isoforms and RNA-editing events. This novel rat reference database is publicly available (http://rat.genomes.nl/proteogenomics/). We used this reference to analyse ultra-deep proteomics data for liver samples from the SHR and BN-Lx strain. Each lysate was proteolyzed with five orthogonal proteases, and the resulting 36 SCX fractions per digest were analysed with LC-MS/MS, cumulating in 180 runs per strain, yielding ~12 million tandem MS spectra. We obtained peptide evidence for 26,463 rat liver proteins and validated 1,195 gene predictions, 83 splice events, 126 proteins with nonsynonymous variants, and 20 isoforms with nonsynonymous RNA editing. Integrative quantitative RNA sequencing and proteomics data analysis revealed a very good correlation for either transcriptome or proteome data between strains but a rather poor correlation between data types. Differences are likely to reflect both differences in specificity and sensitivity between methods and biological regulation mechanisms (e.g. transcript or protein stability). Nevertheless, our multilevel analysis identified a genomic variant in the promoter of the most differentially expressed gene Cyp17a1, a previously reported top hit in genome-wide association studies for human hypertension, as a potential contributor to the hypertension phenotype in SHR rats. These results demonstrate the power of and need for integrative analysis for understanding genetic control of molecular dynamics and phenotypic diversity in a system-wide manner.

In a similar data integrated analysis, we employed a principal component analysis for identification of the major contributors to proteomic or transcriptomic differences between the founding and RI strains. The analysis resulted in components that clearly segregated the founding strains from each other, whereas the recombinant strains spanned the continuous space between
BN-Lx and SHR strains. In agreement with the major phenotypic characteristics of the founding strains, enrichment analysis of GO categories and KEGG pathways on these components revealed differences between genes involved in cholesterol and lipid metabolism and oxidative phosphorylation. To incorporate the genetic information, we transformed a genetic map of markers in recombinant strains to a numeric matrix assigning 1 if a marker is present and 0 otherwise. To this end, we performed a Support vector machine (SVR) analysis to fit the above-identified components to the numeric genetic markers map. In order to identify phenotype-related genetic markers, we employed a panel of feature selection strategies and tested the goodness of fit for various combinations of markers. The highest correlation of the fitted model to the prediction variable was obtained with R:0.56 with a small number of markers.

Our results demonstrate that the developed computational approach can be successfully applied to integrate omics platforms in complex biological systems and generates promising findings. Detailed data integration analysis and follow-up studies are still ongoing.


WP 1.5 Mapping the genetic determinants of the metabonome

Work in WP1.5 represents the first successful attempt to profile the metabolome in several tissues and biofluids from large series of animals from segregating populations (recombinant inbred strains and congenic strains) used to map the genetic determinants of the cardiometabolic syndrome. Overall, >2,000 biological samples have been used for metabolomic profiling using several chromatographic methods each designed to detect and quantify different chemical compounds (lipids, polar metabolites). The amount of metabolic data acquired and quantified in a single cohorts (ie. recombinant inbred -RI strains) is unprecedented and provide a comprehensive survey of metabolites present in a biological samples from genetically heterogeneous individuals, that can be used to test their genetic control. We deliberately chose to apply an untargeted strategy, which has the advantage to allow analysis of known metabolites as well as novel chemical compounds.

Initial work in WP1.5 focused on methods designed to process organ samples prior to NMR or mass spectrometry (MS) and to optimise protocols to the preparation of thousands of samples, whilst keeping constant quality control criteria to avoid drifts in MS and NMR data. Using stringent criteria, we have demonstrated that NMR- and MS -based metabolomic profiling of organ extracts can detect unexpectedly high number features (>5,000) likely to correspond to metabolites in each organ analysed. Global unsupervised analytical methods (principal component analysis) showed that each organ investigated in the RI exhibits specific metabolic regulations and that up to 40% of metabolites can be detected in two different organs.
An important achievement of WP1.5 was the design and development of an R-based suite of programs to automatically process metabolomic datasets and to map the genetic control of metabolite abundance in any biological matrices from any species, including human cohorts. We have deposited the software to a public database dedicated to R programs, in order to provide scientific communities interested in genetic analyses of metabolomic variables with a robust and efficient analytical tool for genetic mapping of quantitative features detected by NMR or mass spectrometry.

An original aspect of research in WP1.5 was the use of two different experimental systems, which allowed genome-wide mapping of metabolomic traits (Recombinant inbred strains) or mapping the metabolic consequences of genetic variants present in well-defined genomic regions (congenic strains). These models are highly complementary as the 50% of the genome is shared at the homozygous state in two RI strains, whereas two congenic strains are 90-99% identical with the exception of the targeted homozygous genomic interval of the donor strain. Results from metabolomic profiling in adipose tissue of congenic strains suggested unexpectedly high level of functional redundancy of independent series of genetic polymorphisms on metabolic regulations.

Untargeted metabolomic studies in organs and biofluids the RI represent a significant progress in the application of the concept of systems genetics in cohorts of animal models, with important prospects for application in human GWAS. Methods used in models and humans in a genetic context are exclusively based on targeted metabolomics, which dramatically reduces that range of molecules that can be detected and quantified for genetic analyses. Untargeted metabolomic revealed the unexpected complexity of organ and biofluid metabolic regulations. As a result, the substantial proportion of unknown metabolites detected in organ extracts from the RI was an important result which led to changes in the objectives of WP1.5 in order to elucidate the structure of a selection of chemical compounds. Partners 2 and 17 collaborated closely to work on chemical attribution of MS signals.

QTL mapping of metabolomic variables was a central objective of WP1.5. Results from genetic studies of the metabolome in WP1.5 demonstrate for the first time the possibility to apply untargeted organ and biofluid metabolomic profiling in quantitative genetics. This is a major achievement in EURATRANS, which can have important consequences in human genetics. We were able to determine the heritability of organ metabolomics and map metabolic features to the rat genome. Data analysis in different organs illustrated the existence of concordant genetic control (the same locus [QTL] controls the abundance of a metabolite in different tissues) as well as independent genetic control of metabolite abundance (different QTLs control the abundance of a metabolite in different tissues). As outlined above, active collaboration between Partners 2 and 17 allowed the elucidation of metabolic compounds corresponding to mQTLs.

Multi-omic data integration was an important objective in WP1.5 which was achieved by the integration of organ metabolomic, and transcriptomic in BN.GK and GK.BN congenic strains. Each congenics contains a specific region of the genome from the diabetic GK rat transferred onto the genetic background of the BN strain. In this genetic context, differences in physiological, transcriptome and metabolome regulations are the direct or indirect consequences of genetic polymorphisms present in the congenic intervals. We used interactome data to define functional connections between metabolites and gene transcripts. This approach identified two candidate genes, which were tested for their function in the regulation of glucose and lipid homeostasis.

Several systems were used for metabolomic QTL validation during the EURATRANS program. Expression downregulation of candidate genes was initially preferred and was successfully used to test the function ASNS and GALM in vitro in cell lines. Validation techniques progressed from candidate gene testing to metabolite testing, which allowed investigations both in vitro in primary cell lines and in vivo in animals. Pipelines were set up to systematically test the biological effects of candidate metabolites both in vitro and in vivo.
Within this work package, we defined a submission standard and repository for data generated during the project. We also developed data standards and ontologies that were used by the EURATRANS partners, and defined standard analysis methods for the sequence data. Another aim of this work package was to ensure data release among EURATRANS partners and the wider scientific community.

The repository developed at EMBL-EBI has three different parts:
1. A dropbox, where EURATRANS partners and collaborators can submit generated data.
2. A private repository, only accessible to EURATRANS consortium members where data are held before publication of associated papers.
3. A public repository, globally accessible by http and ftp protocol where data are moved after publication of the associated papers and at the conclusion of the project.

We also implemented the use of ReseqTrack (http://sourceforge.net/projects/reseqtrack/) which is a MySQL database and Perl API originally developed by another group working at EBI (Resequencing Informatics) for tracking files associated with resequencing projects and running analysis pipelines on resequencing data.

We developed the Pinball algorithm (https://github.com/avilella/pinball) as a mechanism for direct data integration that is free of alignment bias and also largely free of artefacts caused by errors or incomplete sections of the reference genome assembly. In short, the Pinball algorithm removes the short read alignment step that is normally used in all analysis of short read data arising from ChIP-seq or DNase-seq allowing regulatory data to be used across species and directly compared between, for example, human, mouse and rat even though their genome assemblies are of different levels of quality and completion. The alignment step is replaced by a direct assembly step that creates clusters of reads by using a string graph overlap technique based on the SGA assembly program that was recently published. The final clusters of reads retain the same peak shape that is characteristic of ChIP-seq and DNase-seq experimental data analysis. The assembled peaks can be aligned to the reference genome or to related genome sequences thus allowing the analysis of functional genomics data from species or strains that are not identical to the reference genome assembly.

We also developed a pipeline to integrate variations discovered from sequence data with networks constructed from multiple tissues in the RI strains, and to prioritize reliable biological candidate genes for the genetic regulation of the networks. The pipeline is illustrated in Figure 6.

<Pillar 2: Building and Validating Models

WP 2.1 Building molecular gene regulatory networks

Our aim was to build molecular gene regulatory networks that are predictive of protein and phenotypic level changes based on integrated analysis of multi-source biological information through coordinated work with WP 1.6 and the Data Integration
Group (DIG). Based on mapping of quantitative trait loci (QTLs) from multiple modalities (physiological, metabonomic, proteomic, expression and epigenetic), transcriptional networks were built using graphical models and cell-type specific modules were identified from analysis of parental strains. In addition we projected transcriptional QTLs (eQTL), networks and modules onto other 'omics phenotypes and finally identified key regulatory genetic loci for the transcriptional modules. These key regulators were experimentally validated by functional genomics studies in WP 2.2.

Mapping of phQTLs, mQTLs, pQTLs, eQTLs, epiQTLs.

Using a newly described Bayesian method for multi-level data integration (HESS), and sparse Bayesian regression models developed by EURATRANS investigators, we mapped eQTLs in the recombinant inbred (RI) rat strains in seven tissues (adrenal, aorta, fat, kidney, liver, left ventricle, and skeletal muscle). From 241 quantitative physiological phenotypes, we mapped 67 new phQTLs and over 25,000 new eQTLs. We also identified hotspots (master regulators) for expression of miRNAs, regulators of more than 70,000 CpG methylation differences between the left ventricles of BN and SHR strains, and more than 15,000 regulators of the histone modifications H3K4me1, H3K4me3, H4K20me1 and H3K27me3. Using metabonomic profiling of the RI strains, we identified 483 distinct mQTLs in 5 tissues.

Building of transcriptional networks using graphical models and identification of cell-type specific modules from analysis of the parental strains.

We have used Weighted Correlation Network Analysis (WGCNA (Langfelder and Horvath, BMC Bioinformatics 2008, 9:559) and Graphical Gaussian Models (GGM; Schäfer and Strimmer. 2005, Statist. Appl. Genet. Mol. Biol. 4: 32) to identify gene co-expression networks using genome-wide expression profiles in 7 tissues form the HXB/BXH RI strains. These studies defined an interferon regulatory factor 7 (IRF7)-driven inflammatory network enriched for viral response genes, representing a molecular biomarker for macrophages and which was regulated in multiple tissues by a locus on rat chromosome 15q25. Using functional genomic approaches, we showed that Epstein–Barr virus induced gene 2 (Ebi2, also known as Gpr183), which lies at this locus and controls B lymphocyte migration, is expressed in macrophages and regulates this inflammatory network. Through comparative genomic studies, we showed that this network and its regulation was conserved in humans and was associated with the risk of type 1 diabetes. These data implicate IRF7 network genes and their regulatory locus in the pathogenesis of T1D. These data were reported in Nature in September 2010.

Project transcriptional networks and modules onto other ‐omics phenotypes.

Integrative analysis of transcriptional networks and phenotypic data

We implemented a permutation-based strategy to test the association of each gene co-expression network identified across 7 tissues with other phenotypes. To show the efficacy of this approach, we first integrated the large set of 241 phenotypic traits measured in the RI strains with the 364 networks and identified 31 significant phenotype- network associations. One application of these approaches was to integrate eQTLs and epiQTLs using quantitative –omics phenotypes that were profiled in the RI strains (histone ChiP-seq: H3K4me3, H3K27me3; RNA-seq data). We followed a two-step procedure. In the first step, we performed separate QTL analyses for each of the three traits representing a gene. When any of the three revealed a genome wide significant association with a locus, we proceeded to a detailed analysis of all three traits and the genotype at the locus in step two. Here we applied a likelihood based model selection technique (Schadt et al 2003, Nature 422:297-302) to identify the best model from a set of competing graphical models. For each gene-locus pair we selected the model that best explains the observed data using Akaike’s information criterion and assessed the robustness of our findings by bootstrapping. We analysed 245 genes in heart tissue that had sufficient coverage for both histone marks and showed at least one QTL. Further, we analysed subsets of the competing models for genes where only data for gene expression and one of two the histone modifications (H3K4me3: 1,173 heart; H3K27me3: 98 heart) yielded sufficient coverage. This approach allowed us to increase the number of gene expression traits that could be attributed significantly to genetically regulated variation in gene expression (eQTLs). Overall we analysed 1,516 genes of which 542 had a genome wide significant eQTL. In contrast, using a model selection approach that included histone trimethylation levels, we were able to link the expression of 774 genes to a
genetic marker. The modelling allowed for two types of links between a locus and gene expression, either direct or indirect through one of the histone modifications. 470 genes were directly linked - i.e. the histone modifications did not contribute to the allelic effect on gene expression. Of the remaining genes, 80 were indirectly linked through one of the histone modifications. This data was published in Genome Research in 2014.

We also made good progress in the systematic genome-wide mapping of loci controlling trans- eQTL clusters and transcriptional networks and completed the development of methods for gene network mapping and Bayesian multi-level data integration. We identified genetic control points for 154 networks across tissues, and 34 of these networks were associated with physiological phenotypes. Some of the identified regulatory ‘hot-spots’ were identified across tissues. For instance, the locus on chromosome 14 (98 Mbp) controlled two large modules of 1,909 and 303 genes in liver and skeletal muscle tissues, respectively. This module was also significantly enriched for several GO terms, including myofibril (P = 3.08 x10-5), contractile fiber (P = 4.04 x 10-5), sarcomere (P = 0.0005) and was associated with physiological phenotypes in liver tissue, including “liver weight corrected to body weight” (P = 0.0007) and “muscle triglycerides (mmol/g)” (P = 0.0006).

Identification of key regulatory genetic loci for the transcriptional modules.

In this Task, we used the data from whole-genome sequence data that we generated in 27 rat strains, reported in WP 1, to map the regulatory loci of transcriptional networks and modules in the RI strains. We reported earlier the successes in mapping eQTLS, epiQTLs and the master regulator of a trans-acting network affecting risk of type 1 diabetes. In this Task, we carried out systematic annotation of the regulatory loci to identify candidate genes and underlying sequence variants for the regulation of the transcriptional modules.

For each co-expression module that we mapped to the genome we integrated the large set 26,990 eQTLs across tissues at FDR <10% and focused on the modules that contained at least 5 trans regulated genes. We then set out to identify the transcription factors (TF) involved in regulating the transcription of these co-expression modules. We used TRAP1 to compute TF affinity to each promoter in the genome and tested for enrichment of predicted targets among the genes of the module using PASTAA2. This analysis identified 93 transcription factors associated at a 0.1% FDR to 19 modules that were mapped to 16 different loci the rat genome (max number of TF per module 26, min 1). For each locus regulating the module’s expression, we retrieved the list of all genes located in the haplotype encompassing the most strongly associated SNP. For each network, we considered all putative regulators at the regulatory locus and sorted them by the sum of their scores across all associated TFs. We further annotated this list by prioritizing genes that presented a non-synonymous mutation or a significant eQTL in the tissue where the network was detected. The results of this analysis point to specific transcription factors underlying the regulation of co-expression modules and the prioritized regulatory genes at the locus.

Using this approach, we identified several putative regulators of transcription factor activity that underlie the detected transcriptional modules. For instance, we highlight here a few candidate genes: Pias3, a protein inhibitor of Stat3 activation, has been found as a likely candidate for the regulation of a Jund/Fos TFs-enriched network in liver; Gfi1b, a transcription factor that is mutated (non synonymous mutation) and is associated with transcription of a Gfi1 TF-enriched transcriptional module in the left ventricle. Those transcriptional modules mapped to the same genetic locus rat chromosome 8, therefore suggesting common underlying regulatory mechanisms. Zbtb16, also known as Plzf, is a polydactyly gene specifically mutated in the BN-Lx strain that EURATRANS investigators also recently proposed as a regulator of cardiac fibrosis in SHR11. These results independently identify Zbtb16 as a primary genetic factor underlying the regulation of a conserved transcriptional module involved in collagen and ECM, and combined with experimental data in the SHR in identify mutant Plzf as a prominent candidate gene for cardiac fibrosis.

WP 2.2 Functional validation of key network discoveries

WP 2.2 aimed to establish various methods for the targeted alteration of the rat genome to functionally validate key components in disease-related networks. For this purpose, technologies for the manipulation of gene expression in rats by loss and gain of function were developed, optimized, and exploited, including classical transgenesis, RNA interference-based targeted gene silencing, homologous recombination in ES cells, transposon-mediated transgenesis and mutagenesis, gene ablation using zinc-finger nucleases, TALENs, and the CRISPR/Cas9 system, ENU mediated mutagenesis, and isolation of single gene polymorphic or mutated congenic rats. The newly generated rat models were characterized by state-of-the-art phenotyping platforms.

shRNA mediated gene knockdown in the rat
In total, six shRNA transgenic rat models targeting Insulin-Receptor, Apolipoprotein E, LDL Receptor, Tryptophan hydroxylase 2, ATP6AP2, and IL22RA2 have been generated and phenotypically at least partially characterized in the EURATRANS consortium. These inducible rat models for different diseases will be instrumental to test novel therapeutic strategies for cardiovascular, metabolic, inflammatory, and psychiatric diseases.

Gene targeting in rat ES cells by homologous recombination.
Several germline-competent rat ES cell lines have been established and characterized by the EURATRANS consortium. A major problem was the karyotype instability of rat ES cells. Nevertheless, at least two novel rat models, a knockout rat for HPRT and a knockin rat for DsRed into the Sox10 locus, have been generated using these cells. For the latter model, already the novel CRISPR/Cas9 technology was used to create the genetic modification with higher efficiency in the rat ES cells.

Transposon-mediated rat genome manipulation
The Sleeping Beauty transposon system (SB100X) was successfully implemented in the EURATRANS consortium to generate transgenic rats with high efficiency. Moreover, Sleeping Beauty transgenesis combined with recombinase-mediated cassette exchange technology (RMCE) has been successfully established in rats.
Rat spermatogonial stem cell lines suitable for genetic manipulation and transplantation into rats were established, transposon-mediated transfection conditions were optimized, and 2,000 clones with gene-trap mutation were isolated. The library has been kept as an oligoclonal library in triplicates of 10 gene trap mutant per clone. The clones are ready for transplantation and for integration site analysis.

Gene ablation using ZFNs and TALENs in the rat:
ZFN, TALENs, and the novel CRISPR/Cas9 technology were widely used in the EURATRANS consortium to induce targeted genetic modifications in rats. Numerous novel rat models have been generated by these methods (Figure 7). EURATRANS partners produced nearly all novel knockout rat models by ZFN and TALENs in the public sector worldwide.

Sperm archive for ENU mutants
In the EURATRANS consortium, male F344/NSlc rats were treated with ENU to generate 10,752 G1 offspring. DNA and sperm of the G1 animals were isolated and stored for further screening. A publicly available homepage can be accessed at http://www.anim.med.kyoto-u.ac.jp/ENU. The average genome wide rate is one mutation per 9.16 MBp. To date, 8 mutant rat strains have already been established using this ENU platform and are in the process of characterization.

Isolation of single gene polymorphic/mutated congenic rats
In the frame of the EURATRANS consortium, a PVG.DA congenic rat strain containing 4 genes has been derived to carry a gene, which underlies a QTL for Experimental Autoimmune Encephalomyelitis and antibody response to MOG. Another PVG.DA congenic rat containing 4 genes has been derived to carry a gene, which underlies a QTL for Experimental Autoimmune Encephalomyelitis. These strains were established by traditional back crossing of congenic strains and are continually backcrossed to identify recombinants that reduce the congenic fragment with the goal of isolating a single gene.


Pillar 3: Comparative Informatics and Translation to Human WP 3.1 Comparative analysis of rodent and human gene regulatory networks

The main achievement of WP3.1 was the incorporation of genetic mapping and sequence based analyses to identify the molecular basis of 160 phenotypes, including many that relate to common human disease. Using outbred rats, we identified 355 quantitative trait loci (QTLs) at high resolution, providing starting points for understanding the biology of the associated phenotypes and diseases. We applied three additional approaches to complement genetic analysis: imputation, expression and complete genomic sequences. Our imputation method efficiently captures missing data and increases our power to detect effects. Expression data, obtained from 200 hearts, provided a resource for integrating with genetic data and constructing networks for identifying genes as drivers of disease state. Sequence of the eight inbred progenitor strains of the outbred rats was combined with the mapping data to provide unparalleled insight into the molecular nature of QTLs. Combining sequence, expression and high resolution mapping data led to the identification of candidate genes, and in some case to the identification of candidate causal variants. Surprisingly though our expectation that there would be overlap between human, mouse and rat QTLs was not upheld. Indeed, formal tests for overlap between at the gene or pathway level yielded little that was statistically significant.

<Figure 8: Summary of genes identified at QTLs and potential functional variants. Shown are the phenotype measured, the chromosome (chr.), the start and stop coordinates of the QTL, gene symbol and description, whether the gene is the only one at a QTL with candidate variants, whether a variant alters an amino acid and, if so, the residue changed and the potential consequences. (Nat Genet 45, 767-75 (2013))>

WP 3.2 Integration of experimental results with human GWAS data

We utilised the Ensembl multispecies alignments, which use the Enrado-Pecan-Ortheous (EPO) pipeline, to find orthologous regions between rat, mouse and human. Within Ensembl we also significantly augmented the annotation, visualisation and underlying data resources for the human GWAS positive regions as shown in Figure 9. The growth of GWAS as a style of analysis and the resulting publications were much larger than was anticipated when EURATRANS was conceived and we responded by updating our resources for GWAS analysis essentially every two months. Features of the GWAS annotation include the location across the genome that is associated with each of now many hundreds of phenotypes (the figure shows the regions of the human genome associated with glaucoma) as well as sortable tables of genomic locations, associated genes, annotation source and strength of the association as measured by p-value.

We also combined clinical Experimental autoimmune encephalomyelitis (EAE) phenotypes with genome-wide expression profiling to enable correlation of transcripts with genotypes, other transcripts and clinical EAE phenotypes to implicate potential genetic causes and pathways in EAE. Sixty out of 599 cis-eQTLs overlapped well-known EAE QTLs and constitute positional candidate genes, including Ifit1 (Eae7), Atg7 (Eae20-22), KlrC3 (eEae22) and Mfsd4 (Eae17). We defined several disease-correlated networks enriched for pathways involved in cell-mediated immunity. The most significant network was enriched for T cell functions, similar to genetic findings in MS, and revealed both established and novel gene interactions. Transcripts in the network have been associated with T cell proliferation and differentiation, the TCR signalling and regulation of regulatory T cells. We compared these genes to the current list of MS risk loci, and a number of network genes and their family members have been associated with MS and/or other autoimmune diseases.

In order to identify the gene networks in rat that are associated with the diseases studied in EURATRANS, we used a data integration strategy. The approach is based on the integration of (i) the haplotype structure from the genome of closely related rat strains associated with disease, (ii) transcription factor data, and finally (iii) Human GWAS SNPs associated with the disease of interest. Ultimately we identified 17 variants from the set of 1,168 mouse and human GWAS/QTL that was overlapped regions identified as being conserved in the mammalian lineage (Table 1). We have created a final list of selected variants by combining these with those variants obtained using methodology (ii) and are preparing a manuscript for publication.

The results of the research and analysis from this work package has been incorporated into major genomics resources such as Ensembl and the Rat Genome Database (RGD) in a way that is complementary to existing data in those resources and other
The objective of Workpackage 3.3 was translation and validation of genetic, metabonomic and/or proteomic findings in the rat to human cardiovascular and inflammatory diseases using tissues/cells obtained from human bio-repositories. Research Ethics approval was obtained for all work involving human tissue.

Human heart biopsies were obtained from the pre-assembled Heart Transplantation bio-repository at Harefield Hospital. These were high quality samples from patients with end stage heart failure and dilated cardiomyopathy.

A bio-repository of human vascular tissues was established during the EURATRANS project which included saphenous veins from patients with established vascular disease obtained during coronary bypass (CABG) procedures carried out at the Glasgow Heart and Lung Centre, and also venous vascular tissue (VV) from varicose vein surgery patients with no evidence of coronary artery disease.

Cerebral spinal fluid cells and mononuclear cells were obtained from a pre-assembled bio-repository at the Karolinska Institute. These cells were obtained from multiple sclerosis patients and patients with other (non-inflammatory) neurological diseases.

Human kidney samples were collected by the NHS Greater Glasgow & Clyde Biorepository and Urology & Surgical Oncology Units, Glasgow. Biopsies were taken from the healthy pole of the kidney during nephrectomy in non-invasive renal cancer patients. The nephrectomy patients were carefully phenotyped and classified according to their hypertension status.

During the course of the EURATRANS project we have generated several examples of successful verification analysis of regulatory networks and key drivers in human disease tissues and cells.

Deep sequencing of the human and rat cardiac transcriptome revealed an RBM20 dependent regulation of alternative splicing in the heart. Our data suggest a model in which reduced activity of RBM20 results in altered isoform expression of proteins that maintain sarcomeric structure and cardiac function – among them titin, CAMKIIδ, LDB3, and CACNA1C. These changes may result in altered biomechanics, electrical activity, and signal transduction that ultimately lead to cardiomyopathy, fibrosis, arrhythmia, and sudden death. This discovery defines a hitherto unappreciated cause of human cardiomyopathies.

Expression analysis of human vascular tissues has been used for the translational validation of networks and pathways identified in our congenic rat studies. We have previously demonstrated impact of a chromosome 2 congenic interval in the SHRSP rat on structure, mechanical properties and vascular reactivity of mesenteric resistance arteries (MRA). We carried out high throughpout proteomic profiling by triple stable-isotope labeling (SILAC) in primary mesenteric VSMCs from SHRSP, WKY and WKY.SPGla2a congenic rats and identified differentially expressed proteins implicated in vascular tone regulation and remodeling. In particular, we identified the RhoA/Rock pathway as highly regulated (-log(p-value)=5.92e-09) and significantly increased in SHRSP. Analysis of human vascular tissue confirmed differential expression of members of the RhoA/Rock pathway including actin (ACTG2), myosin light-chain kinase (MYLK) and ezrin (EZRIN) together with upregulation of collagen alpha-1(III) chain (COL3A1), which is known to be involved in vascular stiffness.

Uromodulin (UMOD) was recently identified as a novel candidate for hypertension when we performed a genome-wide association study (GWAS) on a population cohort consisting of blood pressure extremes. A single-nucleotide polymorphism (rs13333226 at position 16:20365654) in the 5’ region of the uromodulin (UMOD) gene was associated with hypertension and
subsequently validated. Studies in Umod Knockout mice provided further evidence to support a functional role for UMOD on blood pressure regulation. We utilised samples from our human kidney biorepository to investigate the association between human UMOD gene expression, SNP rs13333226 and hypertension status. Our work with Uromodulin is one of the first examples of functional validation of a GWAS hit for human essential hypertension. Translational verification of UMOD as a key driver of hypertensive disease was necessary in order to confirm its potential as a novel therapeutic target.

We have also carried out a comparison of mRNA expression profiling of kidneys from three distinct hypertensive rat models with histologically quantified renal damage (SHR, salt-loaded SHRSP, two-kidney, one clip hypertension) to identify common pathways for development of hypertensive kidney damage. Out of 88 in-common genes, 40 could be connected in a common pathway enriched for genes involved in extracellular matrix metabolism and the complement system. This data was compared to gene-expression profiles from human kidneys with histologically verified fibrosis. We identified a highly significant number of in-common genes that were also represented in the common genetic pathway identified by the hypertensive rat models. The identified genes will be of major interest in future studies as causative genes of renal fibrosis, as well as markers of hypertensive vascular and kidney damage. The identification of a common molecular network that is associated with structural kidney damage is an important finding and a first step to developing new potential treatments against the progression of kidney damage where treatment of the underlying disease has been unsuccessful.

We have combined clinical phenotypes with genome-wide expression profiling in the Multiple sclerosis (MS)-model Experimental Autoimmune Encephalomyelitis (EAE) in rat to detect regulatory networks involved in inflammation. This enabled correlation of transcripts with genotypes, with other transcripts and with clinical EAE phenotypes to implicate potential genetic causes and pathways in inflammatory disease. We defined several disease-correlated networks enriched for pathways involved in cell-mediated immunity. Similar to genetic findings in the human counterpart MS, the most significant network was enriched for T cell functions, and revealed both established and novel gene interactions. Transcripts in the network have been associated with T cell proliferation and differentiation. TCR signaling and regulation of regulatory T cells. We compared the genes in this network to known MS risk loci, and a number of network genes and their family members have been associated with MS and/or other autoimmune diseases. The Cd6 MS susceptibility allele is associated with alterations in T cell proliferation and the transcription factor Ets1 participates in important aspects of early thymocyte development. Several genes, including Lef1, Lck, Crtam and Itk, have previously been linked to functions of the adaptive immune system.

MS risk has been associated with different HLA alleles and more than 100 non-HLA loci. To be able to understand the impact of these associations on the disease it is important to investigate the relationship between MS risk loci and gene expression. Deep sequencing of peripheral blood mononuclear cells (PBMCs) from MS patients and controls revealed that 34 of the known MS risk variants influence the expression levels of 56 genes. By comparing with previously published studies performed in immune cells from healthy individuals, we conclude that the majority of the MS-associated eQTLs identified to date are not acting specifically in the MS disease, but are independent of disease status. Furthermore, some of these effects were active in only one or a few cell types or conditions, while other were active in several or all investigated types. The findings will be helpful in designing future functional genetics studies to pinpoint dysregulated pathways in MS.

We detected transcripts, which may play important roles in regulating key mechanisms in both EAE and MS owing to a large overlap of genes and functions. These can serve to generate novel hypotheses useful in further dissecting pathogenic molecular mechanisms that are dysregulated during chronic autoimmune inflammation. In addition, the gene network analyses denoted genes linked to functional pathways and described their importance in regulating clinical traits of disease. Ultimately, these integrated findings can provide insight to possible diagnostic and prognostic biomarkers or potential therapeutic interventions for autoimmunity.

Training

Several training activities and events for the consortium have been identified at the beginning of the project, and then further developed during the course of the project, in particular to enable early stage researchers to develop their skills and further their careers as independent scientists.

The definition and coordination of these training activities have been possible through the formation of a Training panel, which expressed all relevant partners and multidisciplinary research represented in the consortium. First, the Training panel ensured that all training activities were of scientific relevance to the project and were placed within the remits of the project’s scientific objectives. Second, the Training panel facilitated the implementation of the training activities, including those that emerged as a continuation of valuable activities preceding EURATRANS (see below - symposia of young investigators). Amongst the main training activities, we highlight the travel fellowships and the symposia for young investigators, also called Young EURAT Investigator Symposia (YEIS).

Travel fellowships

A total of 17 short-term travel fellowships (Figure 10) during a time period of 4 years has been awarded by the Training panel with the main aim of supporting the exchange of researchers between the participant laboratories and their international collaborators and instigate skills transfer and cross-disciplinary collaborations. Through these travel fellowships, researchers undertook a research training activity relevant to the EURATRANS project, and worked in a guest lab for a period of time up to 30 days. Preference was given to early stage researchers, i.e. PhD students and researchers with less than 6 years of post-doctoral experience, with the express intent to facilitate skills transfer at an early career stage. Each short-term travel fellowship was awarded on a purely competitive basis, following formal assessment by the Training panel with respect to the proposed EURATRANS research and training activity to be carried out. A summary of all awarded short-term travel fellowships is provided below.

<Figure 10: Short term travel fellowships.>

Fellows and their research training projects:

- Allison Beaty Sarkis, MCW Milwaukee (host: EBI Hinxton): (P. Flicek) Genomic sequence alignment to identify genetic contributors to hypertension.
- Amelie Baud, WTCHG Oxford (host: RUG Groningen): Finish RNSseq data from HS rats and integration of the phenotypic and
expression QTLs.

- Carme Mont Cardona, UAB Barcelona (host: WTCHG Oxford): Creating transgenic rats for genes at QTL loci influencing rat emotionality.
- Christiana Spyrou, IC London (host: MDC Berlin): Genome-wide RNA sequencing analysis to investigate the fine structure of transcriptional variation underlying hypertensive phenotypes.
- Matthias Heinig, MPI Berlin (host: RUG Groningen): Integration of results from statistical methods for the identification of sequence variants affecting histone modification levels with those from tools to identify candidate genomic regions with variation of histone modifications independent of sequence variation.
- Mohammed Dashti, UGL Glasgow (host: ICL London): Utilise SNP information within congenic intervals to highlight functional as well as positional candidate genes and translation of the results from rat to human hypertension.
- Samreen Falak, MDC Berlin (host: WTCHG Oxford): Genetic studies of internal elastic lamina ruptue in the NIH heterogeneous stock, mapping of aortic internal elastic lamina lesions in the HS population.
- Sebastiaan van Heesch, Hubrecht, Utrecht (host: CEA, Paris): Profiling regulatory and epigenetic changes using ChIP-Seq to understand the molecular basis of strain-specific gene expression differences in the rat; generation of a genome-wide “Enhancerome” of SHR and BN rat liver tissue.
- Julien Chilloux, ICL London (host: INSERM Paris): Targeted mechanistic studies of mQTLderived candidate genes and metabolites in cellular systems
- Stephen Meek, UEDIN Edinburgh (host: KU Kyoto): Plasmid-based TALEN production for the enhancement of targeted genetic modifications in rat ES cells
- Judit Menyhert, MDC Berlin (host: UWISC Madison): Deciphering the Genetic Background of Hormone-dependent Mammary Gland Cancer in the Rat Model by Sleeping Beauty Transposon Mutagenesis
- Maxime Rotival, ICL London (host: MDC Berlin): Co-expression Network analysis of RNA-seq data
- Sabrina Ruhrmann, KI Stockholm (host: ICL London): RNA-seq in autoreactive CD4 cells of reciprocal hybrids
- Pernilla Stridh, KI Stockholm (host: UOXF Oxford): Parent-of-origin effects in the Heterogeneous Stock of Rats
- Sasha Prisco, MCW Milwaukee (host: ICL London): Translating Results from the Sorcs1 Mutant Rat Strain to Humans
- Jessica Le Ven, ICL London (host: INSERM Paris and UCPH Copenhagen): Structural validation by high-resolution mass spectrometry of candidate metabolites identified by mQTL mapping
- Ewart Kuijk, KNAW Utrecht (host: MCW Milwaukee): Transplantation of rat liver stem cells to an FAH-/- IL2rg-/- rat model of liver failure

Symposia of young investigators

The EURATRANS consortium is particularly proud of these symposia of young investigators (YEIS meetings), which represent a spontaneous “bottom-up” training activity where early stage researchers came together without senior researchers or Principal Investigators in order to provide an informal setting that is not possible during the more formal and during the official annual consortium meetings. At the YEIS meetings, early stage researchers presented their research project or experimental plan, following a formal yet relaxed scientific meeting-like structure, which gave the young investigators the opportunity to exchange expertise on experimental methods and analytical procedures addressing questions relevant to their specific project within EURATRANS (e.g. “how to analyze RNA-seq data?”, etc.). This kind of meetings has been called YEIS (Young EURAT Investigator Symposia), and three of them were organized during the course of the EURATRANS project between 2012-2014. However, the YEIS meetings were initiated by young investigators of the rat genetics community at the end of the EURATOOLS (2006-2010) EU FP6-funded project, and continued thought the following EURATRANS project (2010-2015). Beyond showing how this community of young investigators has been highly cohesive and pro-active, this continuum of training and education activities well reflects the strong focus on training in both the EURATOOLS and EURATRANS projects. The YEIS meetings that were carried out during the course of the EURATRANS project are briefly described as follows:

YEIS3 (2012), Karolinska Institutet Stockholm, Sweden; organisers: Pernilla Stridh-Forsgren, Ulrika Norin; 26 participants, guest lectures: Lars Klaireskog, Ann-Christine Eklöf.
YEIS4 (2013), Hubrecht Institute, Utrecht, The Netherlands; organisers: Sebastiaan van Heesch, Roel Hermsen; 18 participants; guest lecture: Hans van Maanen.

YEIS5 (2014), Sol Príncipe Hotel, Málaga, Spain; organisers: David Thybert, David Martin-Galvez, Esther Martinez, Carme Mont; 14 participants; guest speakers: Alfonso Valencia, Francisco R. Villatoro.

Potential Impact:
To understand disease and disease susceptibility, and to interpret genomic information a systematic understanding of the functional elements in a cell is required. Moreover, to fully understand human gene function at the molecular level a significant depth and coverage across all molecular components in a cell is required. Such studies may lead to novel insights into disease mechanisms, especially when cross-disciplinary approaches are integrated to better understand the functional aspects of disease in humans.

Genomics
The genomic inventories form an important basis for functional genomics research using the laboratory rat as a model system for human disease. This inventory thus forms a highly valuable resource for the whole rat research community as it facilitates the identification of causal genetic variation driving common diseases, but it may also allow for optimal experimental and control rat strain selection. Finally, sub-strain information might be highly valuable for interpreting differences in experimental results between laboratories.

Transcriptomics
We have produced and analysed small RNA, ncRNA and mRNA data in 30 HXB/ BXH RI strains and the two founder strains, BN-Ix and SHR/Ola, for left ventricle and liver tissue. Being publicly available, these serve as a powerful tool for many applications beyond the project. We have developed statistical tools for the analysis of differential miRNA expression and successfully applied this to small RNA sequencing data from the RI parental strains, BN-Ix and SHR/Ola.

Epigenomics
The analysis of the genome-wide distribution of the histone marks and the comparison between rat strains and their inbred lines required the development of a new method because of the wide distribution of the marks. The method was implemented in histoneHMM that runs in the popular R computing environment and integrates with the extensive bioinformatic tool sets available through Bioconductor making histoneHMM an attractive choice for the differential analysis of ChIP-seq data.

In addition, the human antibodies that have been produced within the framework of the Human Protein Atlas and validated against rat nuclear transcription factors are distributed upon request.

Proteomics
In WP 1.4 we have provided a comprehensive and quantitative measure of proteome expression levels across several tissues and rat strains. These analyses make it feasible to compare protein expression levels across different rat strains, and allows for integrating proteome expression levels with the corresponding transcriptome and genome data. Hereby providing an in depth understanding of the underlying molecular pathways and mechanisms affected by the investigated disease, enabling a better understanding of disease related genes in the context of functional genomic networks. Collectively, the obtained proteome results provide an unprecedented view on posttranslational processes in context of complex genetic variation.

Moreover, the obtained proteome data may lay the foundation for further developments of precision medicine through protein-based diagnosis and mechanisms, especially considering that proteins traditionally are the preferred biomarker in making patient-related decisions.

Data Resources
The EURATRANS project has improved the resources available for rat genomics; these improvements are publically available via RGD and Ensembl, which means that the potential impact for the scientific community is large. For example, there are 20,000 visits per quarter to the Ensembl rat pages alone. Accurate knowledge of the gene structures of an organism is a fundamental requirement for the interpretation of many types of experimental biological datasets and so this research is important to all individuals who carry out research concerning rats. The open availability of the data generated and the software code and tools to access it will ensure its use is maximised. The primary beneficiaries from this proposal will be biologists and bioinformaticians, both in industry and academia, based in Europe and globally. In particular the rat research community will benefit from this proposal, which will impact research in the areas of agriculture, where rodents contribute to loss in crop yields.

Gene networks
The systematic identification and analysis of gene regulatory networks in the rat model (HXB/BXH RI strains in particular) has led to the identification of pathways and key regulators for complex disease with relevance to human disease. A representative example is the implication of pro-inflammatory network genes and their regulatory locus (Ebi2, also known as Gpr183) in the pathogenesis of T1D. These data were reported in Nature in September 2010. In addition, we have generated a large catalogue of gene variants regulating molecular level and endo-phenotypes (e.g. gene expression, epigenetic marks, methylation, metabolites, etc.), which will be of benefit for the larger genetic community. These data provide a resource for comparative genomics and can be directly used to facilitate the functional annotation of the gene variants identified in human studies (e.g. by GWAS).

Functional validation
The numerous novel rat models with targeted genetic alterations generated by the EURATRANS consortium will be instrumental to understand the pathophysiology and to test novel therapeutic strategies for a multitude of diseases. The molecular and cellular tools established by the EURATRANS consortium will enable researchers to generate even more such models in the future.

Comparative analysis
The combined genetic and expression data we have generated, together with the sequence based analysis, provides important clues to the origins of common disease, including psychiatric illnesses. These findings will be the starting point for further functional investigation of the ways that genotypes are related to phenotypes, and thus provide insights into pathophysiology.

However, perhaps the most important finding from our work was the failure to detect a single candidate variant in half of rat QTLs. One major implication of our work is that at many loci contributing to complex traits, no single variant can be held responsible. This suggests a more complex form of genetic action than has hitherto been suspected, making the identification of the molecular basis much harder, and potentially more interesting. It might explain why resequencing studies at individual loci in human studies have not yielded much additional information than that obtained from standard whole genome association analyses.

GWAS integration
Understanding human disease is a key goal of biomedical research and one critical component of research in this area is the use of model organisms such as the rat. This work package set out to explicitly integrate the extensive and phenotypically relevant EURATRANS experimental results with relevant human GWAS data. Using complementary approaches considerable insight into the regions of the rat genome responsible for key human diseases have been both identified and prioritised. By providing these regions to the wider scientific community further insight into the target disease as well as effective strategies for the best use of rat as a model organism may be achieved.

The primary beneficiaries from this proposal will be biologists, medical researchers and bioinformaticians, both in industry and academia, based in Europe and globally. In particular the groups researching cardiovascular and metabolic diseases may see
benefits especially if they are working with rat models. The rat research community will also benefit.

Validation in humans

The data generated as part of EURATRANS WP3.3 provides a unique knowledge base to advance our understanding of regulatory networks and key drivers associated with human cardiovascular and inflammatory diseases. Identification of new pathways for clinical research (such as RBM20 in the pathogenesis of human heart failure, or uromodulin in the development of hypertension) has potential for important socio-economic impact in terms of better disease prevention, improved diagnostics, health promotion and therapy development.

Training

The training and development of your rat investigators has been considered one of the main activities of the EURATRANS project. In this, specific activities tailored to favour young scientists (e.g. travel fellowships to start collaborative activities between your investigators, symposia for young investigators only, etc.) have been put in place. This range of activities facilitated young investigators to develop towards independent career positions, as exemplified by Dr Matthias Heinig (EURATRANS funded post-doc in Hubner’s group, partner 1) appointment as young group leader at the Institute of Computational Biology (Helmholtz Zentrum München, Germany) or Dr Maxime Rotival (EURATRANS funded post-doc in Petretto’s group, partner 2) appointment as Marie Curie Fellow at the Institute Pasteur (Paris, France). Therefore through training of young investigators, the EURATRANS project has contributed in forming the next generation of scientists within the EU.

Dissemination

The major dissemination route for project results was through publications and presentations (as given in the S&T results section and in section (b) below, and in scientific meetings. Data was deposited in publically available repositories.

The most important data dissemination over the length of the project has been associated with the release of the new version of the rat genome assembly (Rnor_5.0) in Ensembl in the January 2013 update of the resource (version 70). These updates, which included data generated by EURATRANS, were a comprehensive update of the rat gene set as well as updates to the extensive comparative genomics and variation resources available within Ensembl. In addition, new SNP tracks were built within RGD in order to display the EURATRANS data. We have also implemented the importation of metadata and files for data generated by EURATRANS accessible from the European Nucleotide Archive (http://www.ebi.ac.uk/ena/).

Internal dissemination activities within this work package have been the development of a repository for EURATRANS partners to share data within the project.

All data on genomes and phenomes of a population of outbred rats and its progenitors have also been made publicly available (Amelie Baud, Victor Guryev, Oliver Hummel, Martina Johannesson, The Rat Genome Sequencing and Mapping Consortium & Jonathan Flint: Scientific Data 1, Article number: 140011 (2014) doi:10.1038/sdata.2014.11 http://www.nature.com/articles/sdata201411).

EURATRANS steering committee members (N. Hübner and E. Cuppen) have been co-organising the most important annual meeting “Rat Genomics and Models” for many years and consortium members have had a significant part in it, as keynote speakers, speakers, poster presenters and participants. It can be said the EURATools and then EURATRANS actually BECAME the European rat research community!

As an expression of the intensive collaboration between project partners, Figure 11 below depicts the joint publication activity within the consortium. It shows that almost 40 % of all publications have co-authors from more than one group and that the
average number of consortium groups involved in those publications is approx. 5.

<Figure 11: Joint publishing of EURATRANS partners. An entry in the matrix means that the groups have a publication with joint authorship. For publications where only one EURATRANS group is involved (or authored it alone), the entry is on the diagonal; this is true for 48 publications and includes those publications with close collaborators that are not beneficiaries to the EURATRANS grant agreement. When a publication is co-authored by more than one EURATRANS group, it produces pairwise entries in the respective fields. In case of 2 groups involved, this is 1 entry, for 3 groups it is 3 entries, for 4 groups 6 entries, for 5 groups 10 entries, and so on. The overall number of entries is 446, an average of 5.7 per publication (meaning almost 4 groups involved on average). The number of entries without the diagonal is 398; therefore the average number of entries for joint publication is 13.3 meaning an average of more than 5 involved groups. Status: June 09, 2015; 78 publications.>

List of Websites:

http://www.euratrans.eu

Coordinator Contact Details

Prof. Norbert Hubner
MAX DELBRUECK CENTRUM FUER MOLEKULARE MEDIZIN in der Helmholtz Gemeinschaft
Robert-Rössle-Str. 10, 13092 Berlin, Germany
Tel: +49 30 9406 3512, Fax: +49 30 9406 3147, E-mail: nhuebner@mdc-berlin.de

List of beneficiaries with corresponding contact names

1. Max Delbrück Center for Molecular Medicine, Berlin (MDC): Norbert Hubner, Michael Bader, Zsuzsanna Izsvak, Nikolaus Rajewsky.
5. European Molecular Biology Laboratory, Heidelberg (EMBL): Paul Flicek (EBI, Hinxton).
6. University of Glasgow (UGL): Anna Dominiczak.
12. Medical College of Wisconsin, Milwaukee (MCW): Howard Jacob.

Related information
Result In Brief

Massive-scale genomics generates disease models

Reported by
MAX DELBRUECK CENTRUM FUER MOLEKULARE MEDIZIN
Germany

Subjects
Life Sciences - Medicine and Health

Last updated on 2015-10-21
Retrieved on 2019-07-27

© European Union, 2019