Nuclear receptor variation in mice

Beratis, Alexander, Iacob, Diana, Boyanova, Dr. Desislava[1], Mewes, Prof. Dr. Hans-Werner[1]

Abstract

Nuclear receptors (NRs) are a large family of ligand-activated transcription factors, that bind directly to DNA to regulate the expression of target genes. They regulate critical functions in cell control, inflammation, fibrosis and tumor formation and are involved in metabolism, development and reproduction. Nuclear receptors influence the metabolism and signalling processes in the cells by changing the expression of target genes and are associated with numerous pathologies such as cancer, cardiovascular disease, and reproductive abnormalities. This paper presents the investigative results of knockout phenotypes and genetic variation for mouse NRs. Based on an assembly of all known mouse SNPs in the mouse NR genes, the phenotype information for genetic knockouts and genetic variation data was compiled from public databases. Knockout phenotypes were extracted from the Mouse Genome Informatics (MGI) database, while the Mouse Phenome Database (MPD) provides SNPs from various mouse strains, which can be correlated with extreme phenotypes measured in these mouse strains. The goal of this analysis is to find NR-associated SNPs in mice that influence changes in biological parameters such as body weight, body fat and other phenotypic traits. Furthermore, these findings will be coupled to phenotypes observed in mice with a targeted or spontaneous mutation of the nuclear receptor and thus provide additional indication for a putative functionality of the investigated SNPs.

Keywords

Nuclear receptors — SNPs — Gene variation

[1] *Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum Mnchen, German Research Center for Environmental Health,*

Introduction

Looking across the evolutionary patterns between mouse and human, numerous research experiments and gene regulation studies have shown striking similarities regarding certain processes and systems in the two organisms. The mouse presents up to 95% genome similarity to humans and is thus often being used as a model organism when investigating anatomical, physiological or genetical markers in humans. Practically, mice are small, have an accelerated life cycle and represent a cost-effective alternative to genetic research and drug development for human diseases. Also, the majority of the genes responsible for complex diseases are shared between mice and humans, enhancing the chances of successfully identifying patterns in mice which would reveal human disease phenotypes [1].

This paper makes use of the publicly available data in the Mouse Genome Informatics database and Mouse Phenome Database, respectively, in order to highlight changes in various biological parameters in mice under the influence of the NR-associated SNPs. Furthermore, these findings can be mapped to genotype - phenotype associations in humans, for studying the human biology and disease.

1. Methods

Both the process of consolidating a genotype - phenotype map, as well as the subsequent analysis of extreme phenotypes observed in mice rely on four biological elements: mouse strains, gene names / symbols, nuclear receptors, SNPs. The correlation between the position of a nuclear receptor and various SNPs on a gene can result in a specific phenotype. Due to increased specificity rate of the phenotype associated with a particular nuclear receptor, different NR and SNPs associations will result in different phenotypes for the same gene. Similarly, diverging mouse

strains present non-identical phenotypes for the genetic parameters.

1.1 Mouse nuclear receptors

The analysis presented in this paper was based upon a dataset consisting of the 49 nuclear receptors of mouse [2] and their associated genes (242 unique gene Ids, 49 gene names / symbols); for a full list of gene names, see Appendix 6). The gene names were further on used to compile additional information regarding the position, associated SNPs and phenotypes for the nuclear receptors in mice.

1.2 MGI - Mouse Genome Informatics

Mouse Genome Informatics[1] [3] is a free, online database for the laboratory mouse and provides access to information about integrated genetics and associations between specific phenotypes and their corresponding alleles. It contains over 24000 genes and their protein sequences and approximately 48000 genotypes and phenotype annotations. For this research, the MGI database was solely used for building a connection between the nuclear receptor genes in the mouse and the associated phenotypes dependent on miscellaneous strains.

1.3 MPD - Mouse Phenome Database

Mouse Phenome Database[2] [4] includes annotations of measured data on the laboratory mouse strains and populations, as well as SNPs and phenotypes of the examined strains. More than 1330 strains were examined, providing annotations for over 3500 phenotype and 1.8 billion genotypes.

The MPD database is more detailed and comprehensive than the MGI, so that the phenotypes found in MGI can be traced back to MPD. However, the MPD strictly associates individual phenotypes to their corresponding strain, posing difficulties in the mapping process between the mouse nuclear receptors and the MGI data.

2. Database

Since annotation from multiple public databases and internal data were used, there was a need to design and implement a database supporting the research. The database structure thereby presents 7 tables and aims to facilitate the processes of gathering, parsing and statistical analysis of the nuclear receptors data, as well as to allow for a clear mapping between the nuclear receptor genes and the phenotypes and SNPs associated with them. The main challenge was to establish a connection between the MGI, MPD and UCSC tables. Therefore, there are 2 tables containing information from the MGI, 3 tables for MPD, 1 USCS table and the core table:

- **nr_mapping** - core table, which contains the associations between the 49 nuclear receptors in the mouse and their corresponding gene names.
- **mgi** - contains the MGI gene annotations (MGI gene id, type, attributes, transmission etc.).
- **mgi_phenotypes** - stores phenotypic information in association with MGI genes; more than one phenotype can be associated with a specific gene.
- **mpd_snp** - contains the MPD SNP-annotations in correlation with the nuclear receptor genes.
- **ucsc** - contains a more detailed overview on the MPD mutations from *mpd_snp*
- **mpd_strains** - contains the MPD strain annotations (MPD strain id, sex, number of mice, Z-score etc.)
- **mpd_phenotypes** - stores phenotypic information in association with MPD strains; more than one phenotype can be associated with a specific strain.

[1] `http://www.informatics.jax.org/`, March 13, 2015
[2] `http://phenome.jax.org/`, March 13, 2015

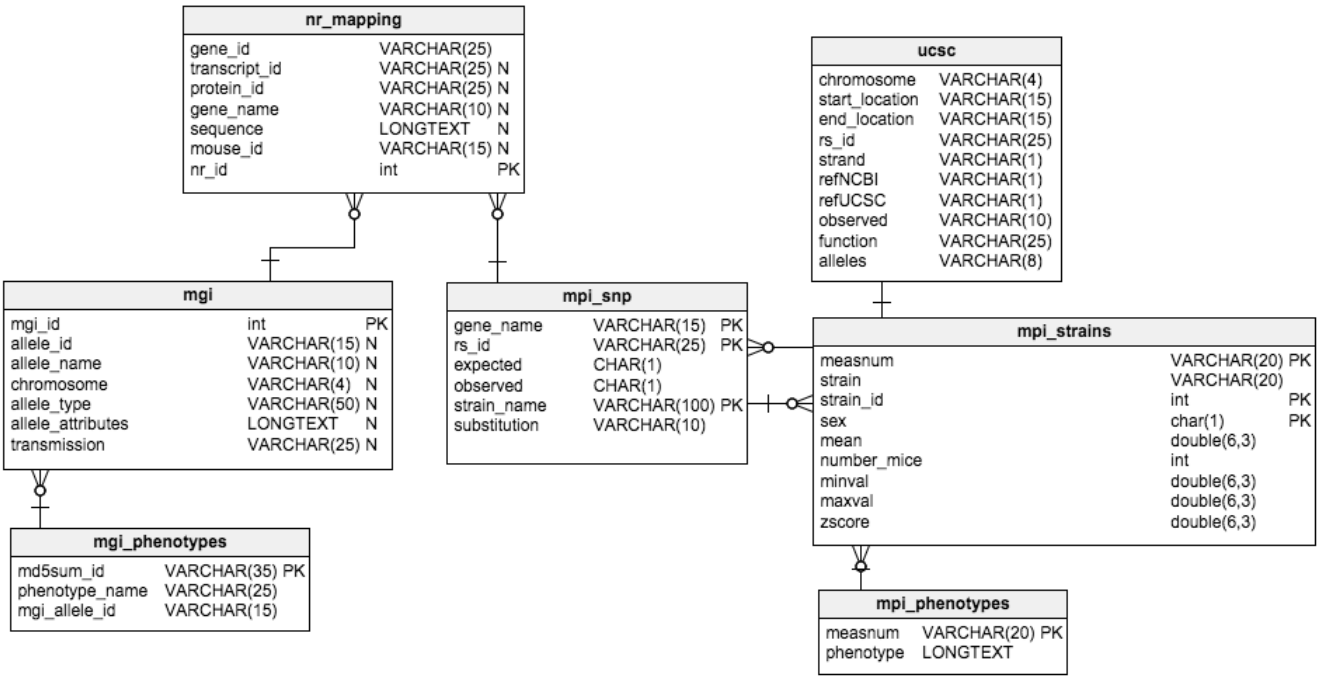

Figure 1. Database.
Database scheme for the 49 nuclear receptors with information about MGI, MPI and UCSC.

3. Results

The goal hereby was to visualise the most important phenotypes associated with the 49 nuclear receptors in the mouse. In this regard, the phenotypic and genomic data from the MGI and MPD was compared and analysed. All graphics were generated using the R package[3].

3.1 MGI Statistics

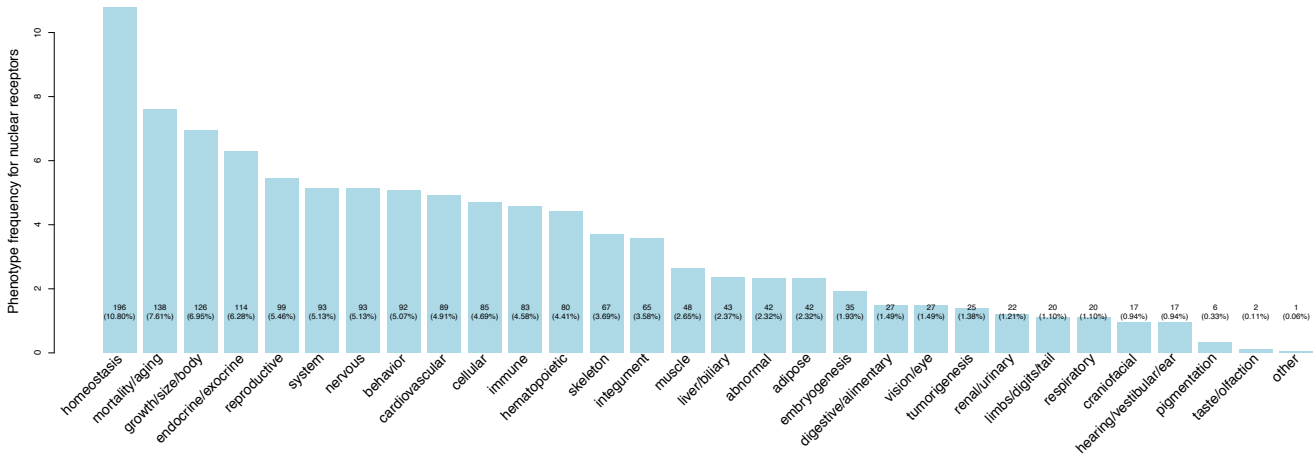

Figure 2. MGI phenotypes.
Mouse Genome Informatics phenotype distribution over the genes associated with the 49 nuclear receptors in the mouse.

[3] `www.r-project.org/`, March 13, 2015

The MGI database consists of generalised definitions of the phenotypes associated with the nuclear receptors found in the mouse. Figure 2 illustrates the most significant phenotypes and their occurrence frequency across the 242 genes corresponding to the nuclear receptors in the mouse. With a count of 196 matches across the dataset, the most prominent phenotype correlates with the homeostatic metabolic processes, such as temperature regulation and pH-balance. Moreover, Figure 3 provides an insight into the exact association of each phenotype with the corresponding nuclear receptor genes, such that homeostatis, for instance, is prominently found in the following genes: *Esr1*, *Pparg*, *Thrb*, *Thra*, *Vdr* etc. Other phenotypes describe body size and growth features and are representative for the *Pparg* genes.

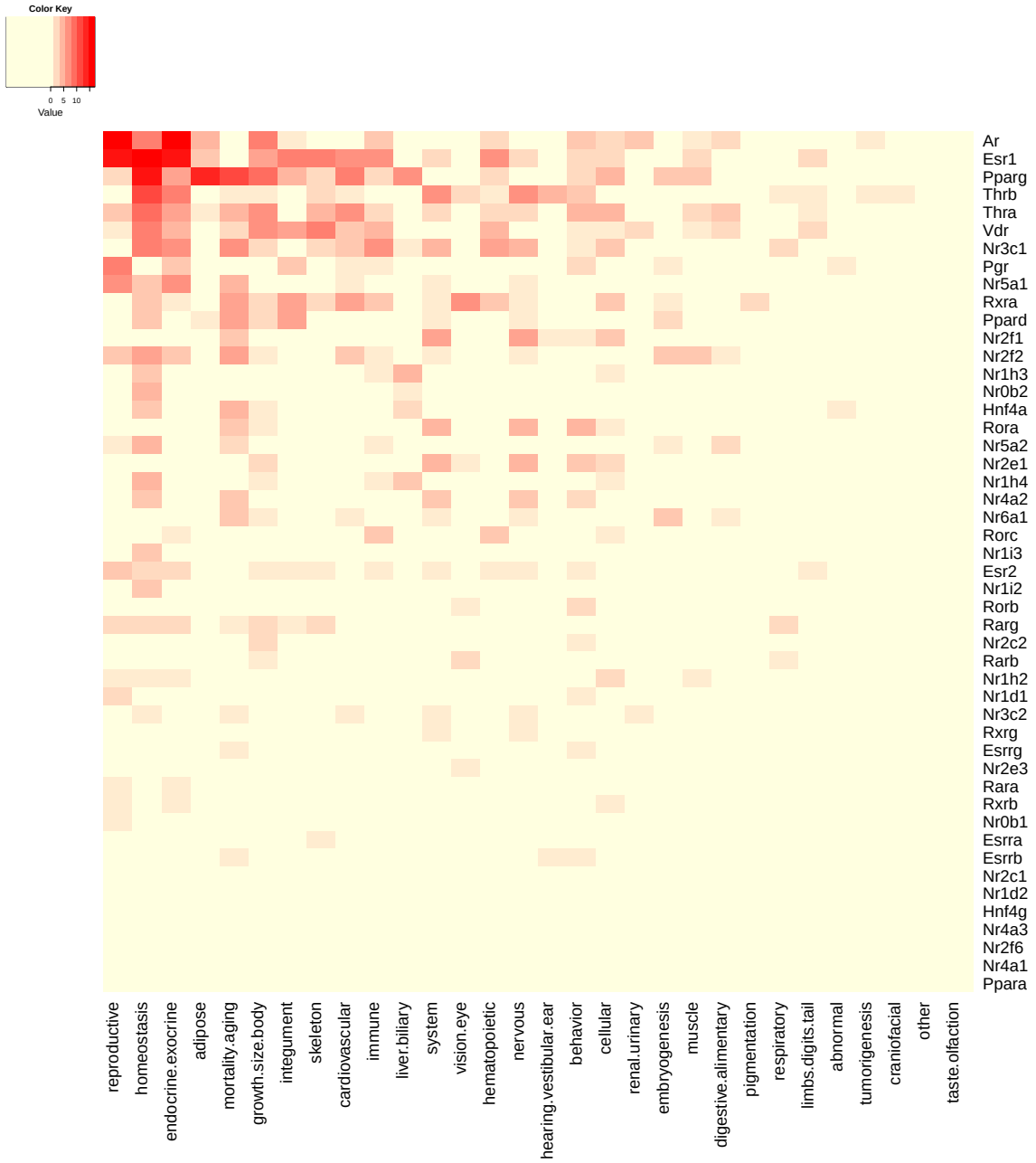

Figure 3. MGI phenotype - nuclear receptor gene associations.
Mouse Genome Informatics phenotype occurrence frequency among the nuclear receptor genes.

3.2 MPD Statistics

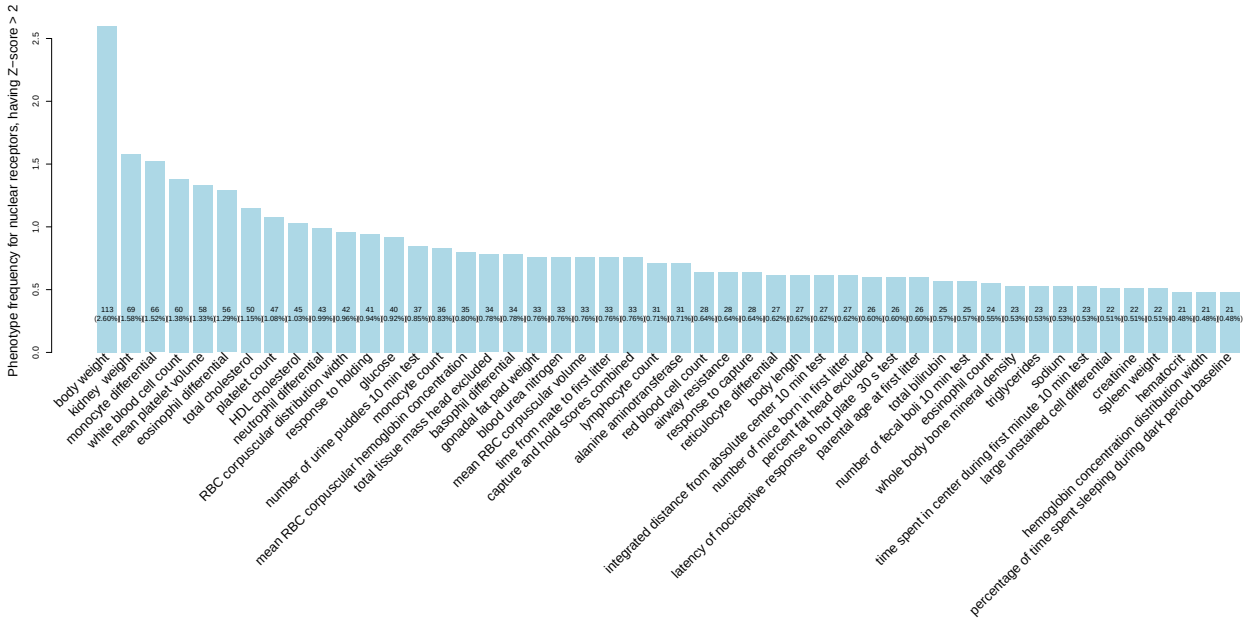

Figure 4. MPD extreme high phenotypes, having Z-score $>$ 2.
Mouse Phenotype Database extreme high phenotype distribution over the genes associated with the 49 nuclear receptors in the mouse, having Z-score $>$ 2

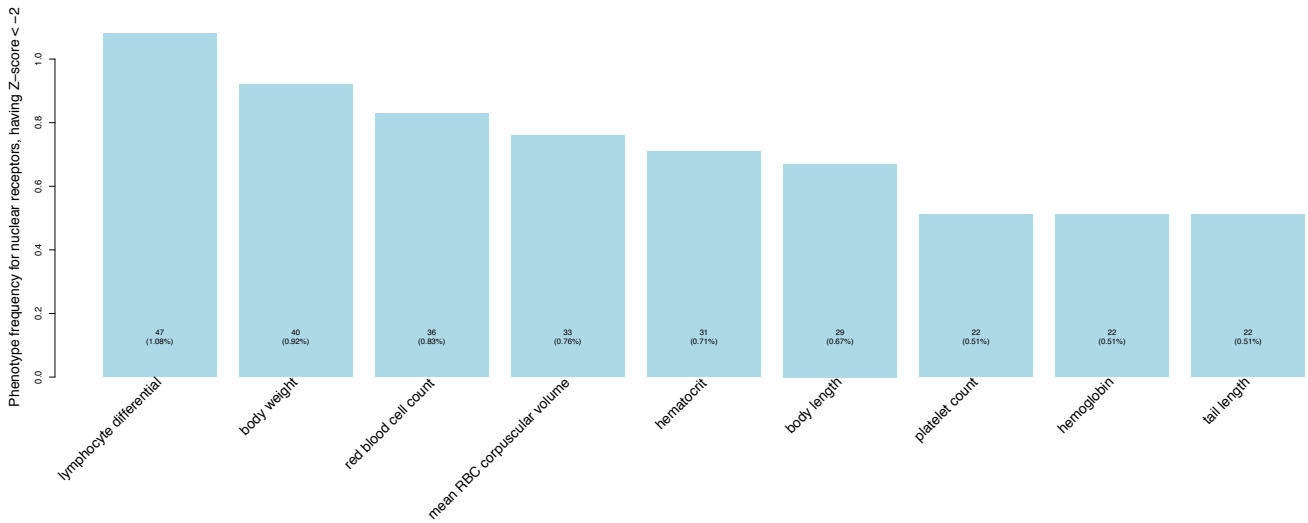

Figure 5. MPD phenotypes, having Z-score $<$ -2.
Mouse Phenotype Database extreme low phenotype distribution over the genes associated with the 49 nuclear receptors in the mouse, having Z-score $<$ -2

The MPDatabase contains annotations for extreme phenotypes associated with different mouse strains. An extreme phenotype is described by a Z-score above 2 (very high) or below -2 (very low), relative to other measurement means. Based on these specifications, Figure 4 illustrates the most significant phenotypes and their occurrence frequency across the 242 genes corresponding to the nuclear receptors in the mouse, with a Z-score > 2 (extreme high phenotypes). Here, the most significant phenotypes include *body weight*, *kidney weight* and *cholesterol*. Similarly, Figure 5 presents the most significant phenotypes and their occurrence frequency across the 242 genes corresponding to the nuclear receptors in the mouse, having a Z-score $<$ -2 (extreme low phenotypes). In this case, the most significant phenotypes include *body weight* as well, but the focus lies on several blood phenotypes (e.g. lymphocyte differential, red blood cell count, hemoglobin).

Table 1 shows the strains and gene names associated with the highest ranking *MPD* phenotypes relative to their Z-scores. As seen in the table, the phenotypes are not directly associated with the expressed proteins, but rather with the observed abnormal condition in the mouse strains.

Strain	Gene	Phenotype	Z-score
CAST/EiJ	Rora (RAR-related orphan receptor alpha)	trigonelline relative abundance	8.8
C58/J	Rorc (RAR-related orphan receptor gamma)	N-acetylglutamate relative abundance	8.4
CAST/EiJ	Rara (retinoic acid receptor, alpha)	N2-acetyllysine (16h fast), relative abundance	4.6
NZB/BlNJ	Nr1i3 (nuclear receptor subfamily 1, group I, member 3)	ECG parameters interval between peak of P-wave to R-wave (PR)	4.6
CE/J	Nr0b1 (nuclear receptor subfamily 0, group B, member 1)	kidney total weight	4.4
NZB/BlNJ	Nr5a1 (nuclear receptor subfamily 5, group A, member 1)	succinylcarnitine relative abundance	3.8
C57L/J	Esr2 (estrogen receptor 2 beta)	relative size of perivascular immune cell clusters	3.5
CE/J	Rorb (RAR-related orphan receptor beta)	tigloylglycine (16h fast), relative abundance	3.3
NOD/ShiLtJ	Ppard (peroxisome proliferator activator receptor delta)	percentage of parasites in brain relative to all organs tested	3.1
BUB/BnJ	Esr1 (estrogen receptor 1 alpha)	3-dehydrocarnitine relative abundance	2.7
tdb	Esrrg (...)	tbd	tbd

Table 1. Top 10 extreme phenotypes.
Mouse strains associated with MPD extreme phenotypes, based on their Z-score.

Moreover, Figures 6 and 7 provide an insight into the exact association of each phenotype with the corresponding nuclear receptor genes. Therefore, there are 4 genes associated with the extreme high and low phenotypes respectively, as following: *Rora - RAR-related orphan receptor alpha*, *Esr1 - estrogen receptor 1 alpha*, *Esrrg - estrogen-related receptor gamma*, *Thrb - thyroid hormone receptor, beta* and a fifth gene associated only with extreme low phenotypes - *Nr3c2*.

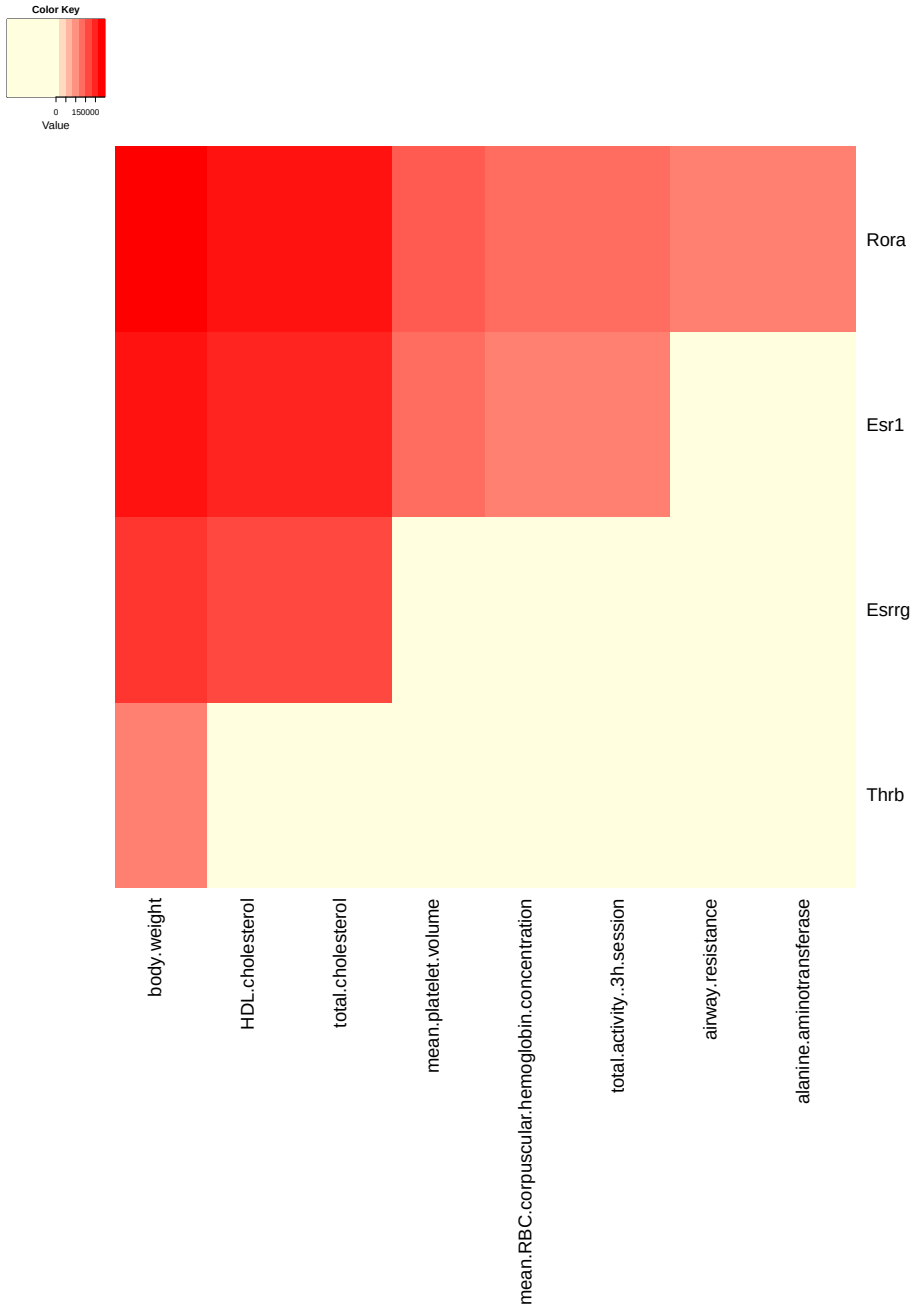

Figure 6. MPD extreme phenotypes, having Z-score > 2.
Mouse Phenome Database extreme phenotype occurrence frequency among the 49 nuclear receptors in the mouse, having Z-score > 2

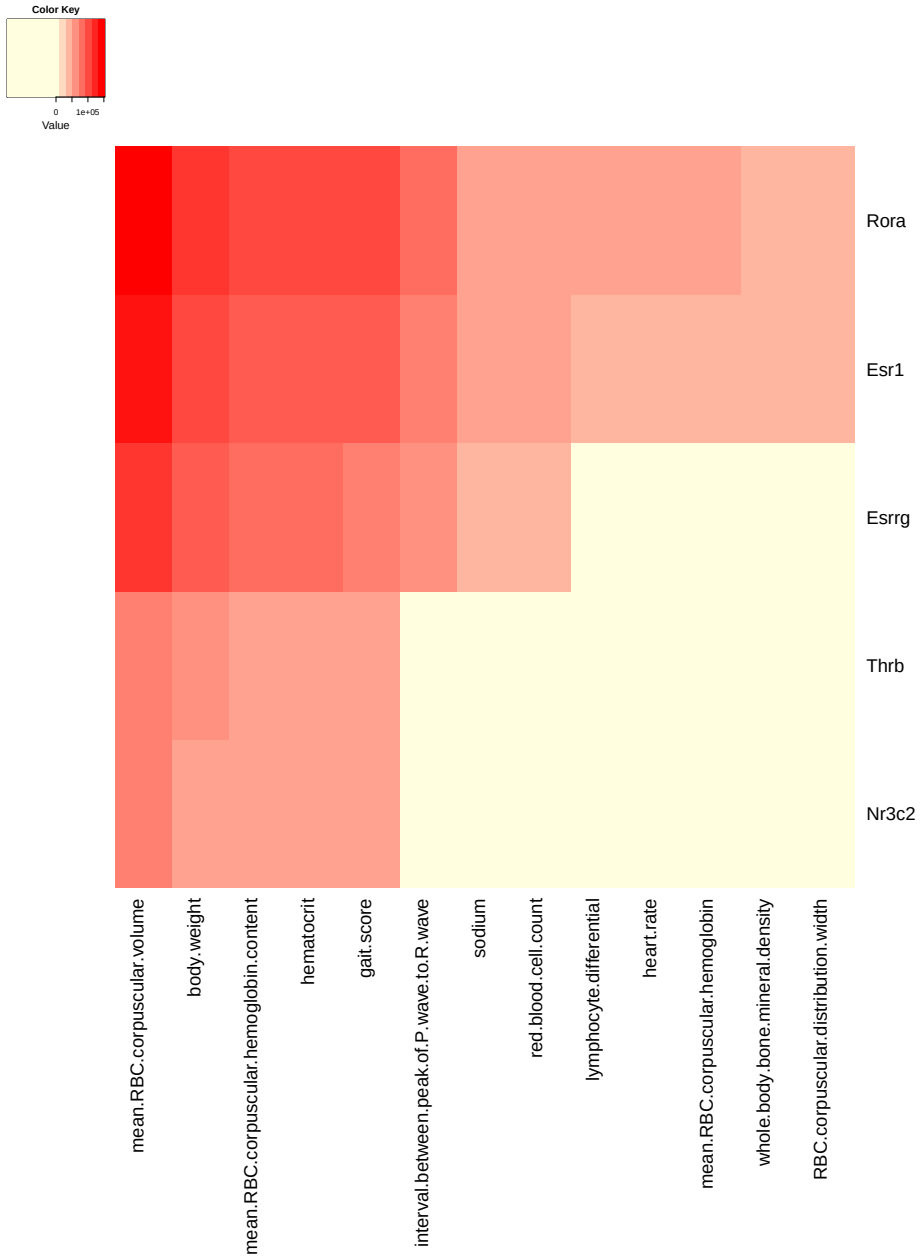

Figure 7. MPD extreme phenotypes, having Z-score $<$ -2.
Mouse Phenome Database extreme phenotype occurrence frequency among the 49 nuclear receptors in the mouse, having Z-score $<$ -2

3.3 Nuclear receptors in the human genome

The nuclear receptor variation has been widely studied not only in the mouse, but also in the human genome. The research hereby refers to two published genome-wide association studies in this field of interest, namely *Kora*[4] - a GWAS study initiated at the Helmholtz Zentrum in Munich involving approximately 18,000 adults from Southern Germany - and *TwinsUK*[5], being the biggest adult twin genetic registry in the UK.

Using the information provided by these studies regarding the correlations between different SNPs and metabolites

[4] `http://epi.helmholtz-muenchen.de/kora-gen`, March 13, 2015
[5] `http://www.twinsuk.ac.uk/`, March 13, 2015

to metabolic pathways, this research aimed to identify the similarities between the nuclear receptors and their associated phenotypes in human and in the mouse, respectively. Due to missing or incomplete annotations for the nuclear receptors in the mouse, comparisons could only be made between the phenotypes extracted from the MGI and MPD databases (see tables *mgi_phenotypes* and *mpg_phenotypes* from the internal database 1). Table 2 shows a list of nuclear receptors found in both the mouse and human and which present a similar phenotype or share the same biological function. As seen there are only a few phenotypes shared between these two species. The reason therefore is that the phenotypes for the nuclear receptors in both, human and mouse do not only have influence on one metabolic pathway. For each nuclear receptor there were roughly about 8 to 10 phenotype annotations available, but only the ones which share the same function were written into the table. Nevertheless as an example the gene *Nr2e1* without limitation regulates the expression for the same kind of molecule, a steroid hormone. This leads to the assumption that the nuclear receptors have somehow a similar influence on the individual in each species, albeit there are only few annotated by this time.

Nuclear Receptor Gene	Mouse phenotype	Human phenotype
Nr2e1	corticosterone, *steroid hormone*	lathosterol, *steroid hormone*
Essrg	carnitine, *lysine metabolism*	glutaroyl carnitine, *lysine metabolism*
Thrb	*thyroid hormone resistance*	*thyroid hormone resistance*
Nr0b1	*adrenal hypoplasia*	*adrenal hypoplasia*
Vdr	*osteoporosis*	*osteoporosis*
Ar	*spinal and bulbar muscular atrophy*	*spinal and bulbar muscular atrophy*

Table 2. Nuclear receptors with their coding protein or phenotype
Nuclear receptors in human and in the mouse which share a similar phenotype or biological function.

4. Discussion and Outlook

Over the last year, the number of studies and publications involving the nuclear receptor variation has increased and scientist have been focusing on exploring the effects of nuclear receptors in different organisms, on various levels: in the cells, tissues, and even full-body impact. The goal hereby is to link their metabolic activity to possible human diseases and facilitate the drug development processes. In this regard, extensive experiments and studies have been performed on humans, as well as on laboratory mice, which are known to present a striking genomic resemblance to humans (up to 95%) and and time and money-wise more advantageous. As a result, it has been proved that nuclear receptors have a great influence over the extreme phenotypes showcased in the mouse, as well as in human.

In spite of the ascending progress in this field of research, there annotation data provided in the mouse public databases - *Mouse Genome Informatics* and *Mouse Phenome Database* on nuclear receptors is still rather limited. However, the two databases provide complementary information: the phenotypes annotated in *MGI* describe roughly the biological systems in which they occur rather than the description of differences between non-variants, as of *MPD*. On the other hand, the UCSC database for the human genome is far more detailed, comprising also information regarding the metabolic pathways associated with various phenotypes.

This research presents the most significant phenotypes found in the mouse, based on the 49 nuclear receptor variation, as well as a mapping of these phenotypes on the human phenotypic expression, providing insight into possible human disease conditions. Nevertheless, the mouse public databases are far less comprehensive than the UCSC, for instance, allowing for only a generalised comparison between the two species, relative to the phenotypic expression in case of a variation in one of the same nuclear receptors.

The mapping of phenotype information between mouse and human showed that regardless if the genes are thomolog,

their function can differ between these two species, at least with the information we have by this time on the human nuclear receptors. As long as the human nuclear receptors are not enough researched, mapping the nuclear receptor variatons from mouse on the human can only result in an indicator to get information about what protein the gene could code for or in which biological process is it likely to be involved.

5. Acknowledgments

Our gratitude to Prof. Dr. H. W. Mewes for offering us the topic and the opportunity to work with the Helmholz Zentrum research centre in Munich. Many thanks to our supervisor, Dr. Desislava Boyanova for all the input and ideas, discussions, advices and foremost her useful and critical suggestions which motivated us a lot. Last but not least, we would like to thank our fellow colleagues for their support, as well as the entire *Helmholz Zentrum* group.

6. Appendix

List of gene names associated with the 49 mouse nuclear receptors:

Esrrg, Rorc, Nr1i2, Hnf4g, Nr0b1, Nr3c1, Hnf4a, Rxrg, Esrra, Thrb, Nr5a1, Pparg, Nr4a1, Nr5a2, Nr1h3, Nr4a3, Nr1i3, Nr2f2, Nr4a2, Nr2f6, Esrrb, Ppard, Nr2e1, Vdr, Nr6a1, Nr1h5, Rarb, Nr1h2, Esr1, Rora, Rarg, Nr1h4, Rara, Nr3c2, Pgr, Ar, Rxrb, Thra, Esr2, Rorb, Ppara, Nr1d1, Nr0b2, Nr2e3, Nr2f1, Nr1d2, Nr2c1, Rxra, Nr2c2.

References

[1] Ali J. Marian. Modeling human disease phenotype in model organisms. 2011.

[2] Qiongming Liu, Chen Ding, Wanlin Liu, Lei Song, Mingwei Liu, Liang Qi, Tianyi Fu, Anna Malovannaya, Yi Wang, Jun Qin, and Bei Zhen. In-depth proteomic characterization of endogenous nuclear receptors in mouse liver. 2013.

[3] Judith A. Blake, Carol J. Bult, Janan T. Eppig, James A. Kadin, and Joel E. Richardson. The mouse genome database: integration of and access to knowledge about the laboratory mouse. 2013.

[4] Terry P. Maddatu, Stephen C. Grubb, Carol J. Bult, and Molly A. Bogue. Mouse phenome database (mpd). 2011.