

# Nuclear receptor variation in mice

Beratis, Alexander, Iacob, Diana, Boyanova, Dr. Desislava<sup>1</sup>, Mewes, Prof. Dr. Hans-Werner<sup>1</sup>

## 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

<sup>1</sup> Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, 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 (for a full list of gene names, please 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<sup>1</sup> [3] is free, online database for the laboratory mouse, which provides access to information about integrated genetics and associations between specific phenotypes and their corresponding alleles.

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. Information about the SNPs and the variation of the nuclear receptors were not taken from MGI, because the annotation regarding this matter was far better in the MPI and UCSC database.

### 1.3 MPD - Mouse Phenome Database

Mouse Phenome Database<sup>2</sup> [4] includes information about measured data on laboratory mouse strains and populations as well as SNPs and phenotypes of the examined strains. More than 1330 were examined in several ways, over 3500 phenotype and 1.8 billion genotype annotations were made [4]. Given that the MGI phenotypes can be compared to the ones found in MPI which are by far more detailed. The drawback of the MPI database is that every sort of data is always referred on the regarding strain, making the mapping between the nuclear receptors and the MGI data more difficult. Mouse Genome Informatics<sup>3</sup> is a database for the laboratory mouse, which makes information about integrated genetics and associated phenotypes with their alleles available. The Mouse Genome Database contains over 24000 genes with protein sequences and approximately 48000 genotypes with phenotype annotation [?]. For this paper the MGD was solely used for building a connection between the nuclear receptor genes and the associated phenotypes always dependent on miscellaneous strains. Information about SNPs and variation of the nuclear receptors were not taken from MGD, because the annotation regarding this matter was far better in the MPI and UCSC database.

### 1.4 MPD

Mouse Phenome Database<sup>4</sup> [4] includes information about measured data on laboratory mouse strains and populations as well as SNPs and phenotypes of the examined strains. More than 1330 strains were examined in several ways, over 3500 phenotype and 1.8 billion genotype annotations were made [4]. Given that the MGI phenotypes can be compared to the ones found in MPI which are by far more detailed. The drawback of the MPI database is that every sort of data is always referred on the regarding strain, making the mapping between the nuclear receptors and the MGI data more difficult.

## 2. Database

The goal of the database was to create several tables which could be mapped with the reference table of the 49 nuclear receptors, to create statistics about the relationship between the nuclear receptors and the phenotypes and SNPs associated with them. As mentioned above, the main problem was the connection between the MGI and the MPI tables, which was gone around by using an extra table which contained only the strains of MGI which are the same as the ones of the MPI table. As additional information regarding the SNPs another database was

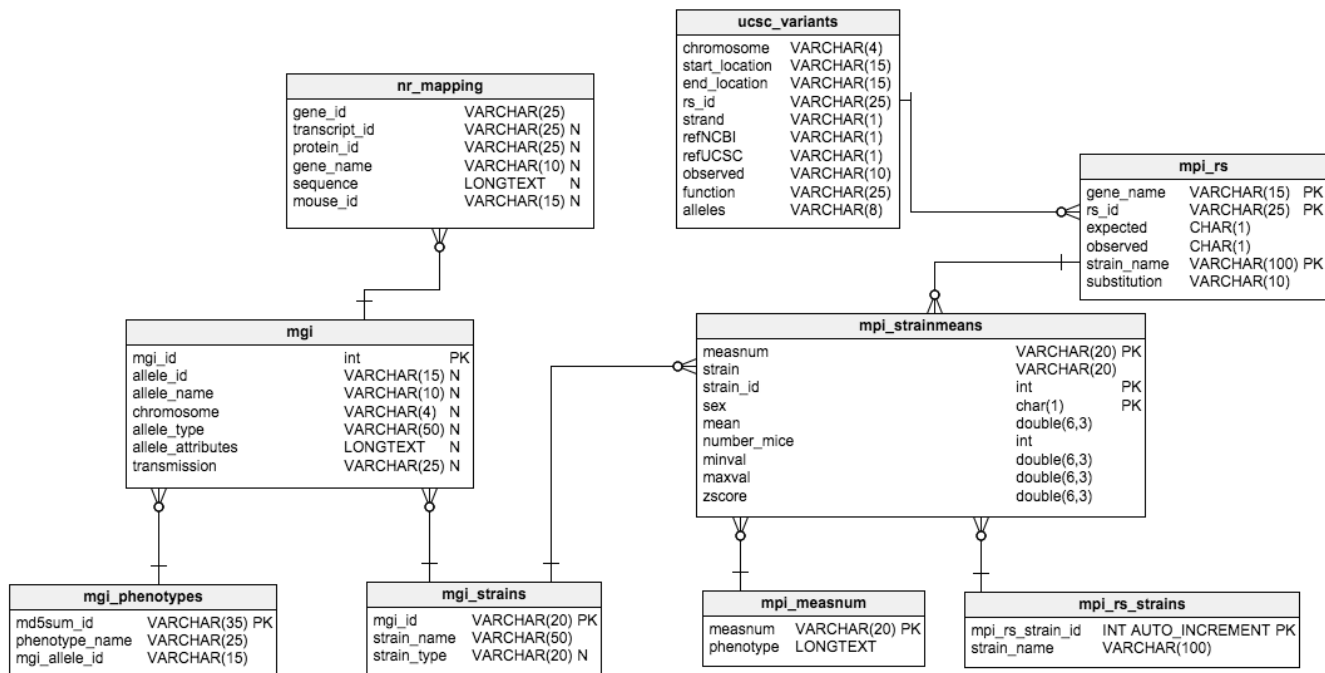
<sup>1</sup><http://www.informatics.jax.org/>, March 10, 2015

<sup>2</sup><http://phenome.jax.org/>, March 10, 2015

<sup>3</sup><http://www.informatics.jax.org/>, March 10, 2015

<sup>4</sup><http://phenome.jax.org/>, March 10, 2015

consulted, the UCSC database. This table contains every SNP of the MPI table, but with information about the reference genome and the locations of the SNPs. Having that information, it is possible to create statistics regarding the substitutions of the nucleotides, locations and the phenotypic results arising from those SNPs. The goal of the database was to create several tables which could be mapped with the reference table of the 49 nuclear receptors, to create statistics about the relationship between the nuclear receptors and the phenotypes and SNPs associated with them. As mentioned above, the main problem was the connection between the MGD and the MPD tables, which was gone around by using an extra table which contained only the strains of MGI which are the same as the ones of the MPD table. As additional information regarding the SNPs another database was consulted, the UCSC database. This table contains every SNP of the MPD table, but with information about the reference genome and the locations of the SNPs. Having that information, it is possible to create statistics regarding the substitutions of the nucleotides, locations and the phenotypic results arising from those SNPs.

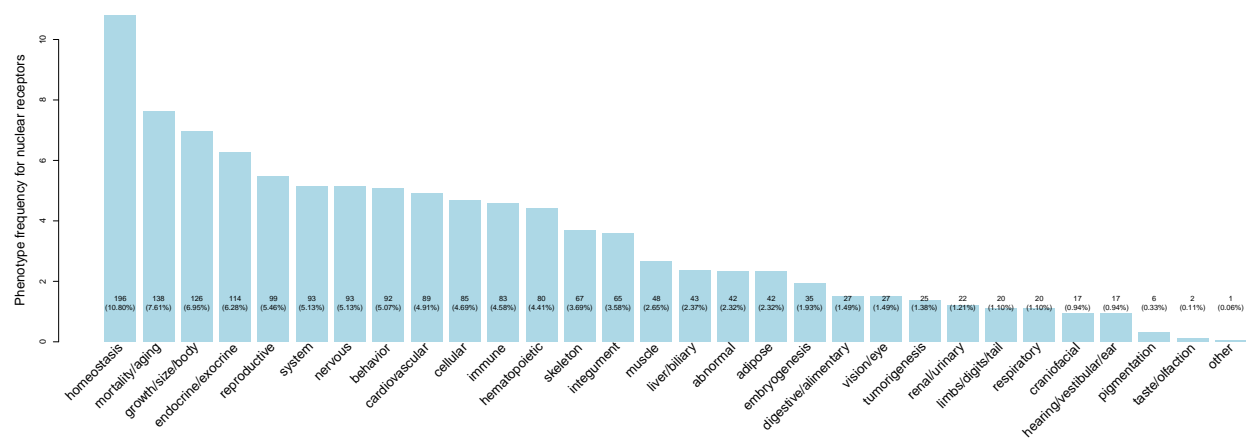


**Figure 1. Database.**

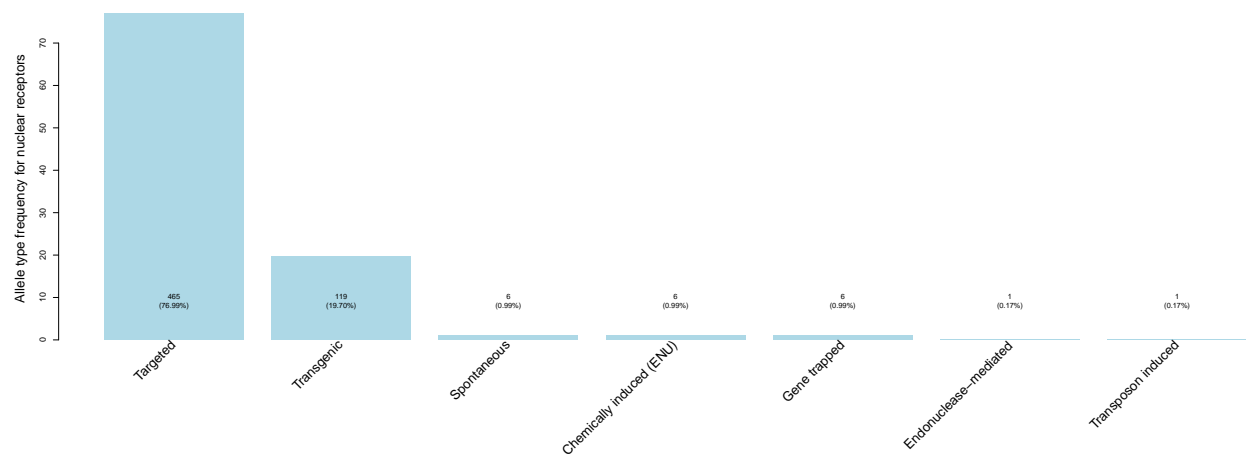
Database scheme for the 49 nuclear receptors with information about MGI, MPI and UCSC.

All databases were build upon the *nr\_mapping* table, which contains the important association between the nuclear receptors and the gene names. Given that the table for MGI was created by searching the gene names in the database and extracting all necessary information, analogous for the MPI database. All databases were build upon the *nr\_mapping* table, which contains the important association between the nuclear receptors and the gene names. The *mgi* table with the information regarding the genes itself like allele attributes and types was connected with the *mgi\_phenotypes* table which stored the phenotypic information in association with the internal MGD ID. To map the nuclear receptors and the data of MGD on MPD we used the *mgi\_strains* table which contained every available strain in mgi and built with that and the gene names of *mgi* the *mpi\_strainmeans* table. This table stores information about experimental results from strains like means, maximum values, minimum value and the zscore which will be an important indicator in the results chapter. Further the table includes an *measurementsnumber* which was the key identifier for an element in the *mpi\_strainmeans* and the *mpi\_measnum* table which stores information about phenotypes associated with the regarded genes. To get a more view on the genotypic aspect besides the phenotypic, the *mpi\_rs* and the *ucsc\_variants* tables were built. They store the information about SNPs in the genes which are affected by the nuclear receptors. Given that the table for MGD was created by searching the gene names in the database and extracting all necessary information, analogous for the MPD database.

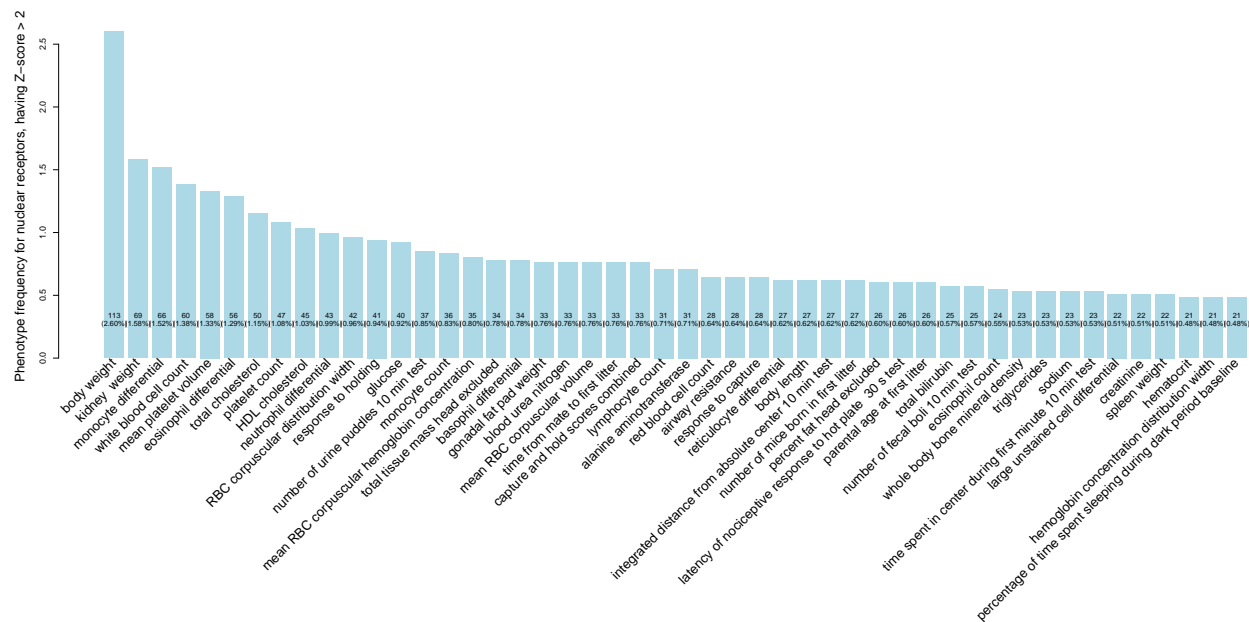
3. Results



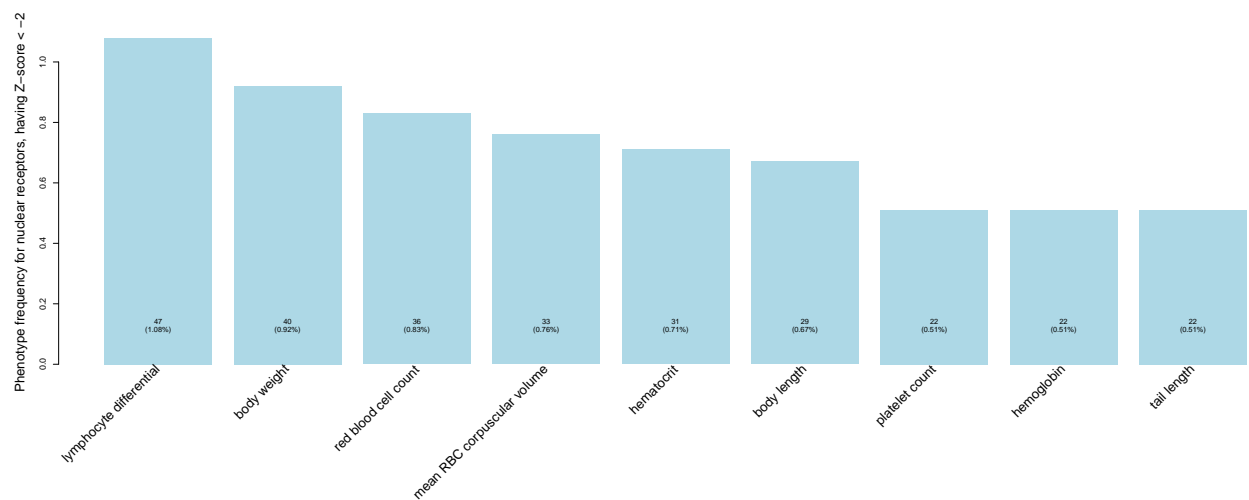
**Figure 2. MGI phenotypes.**  
Mouse Genome Informatics phenotype distribution over the nuclear receptors.



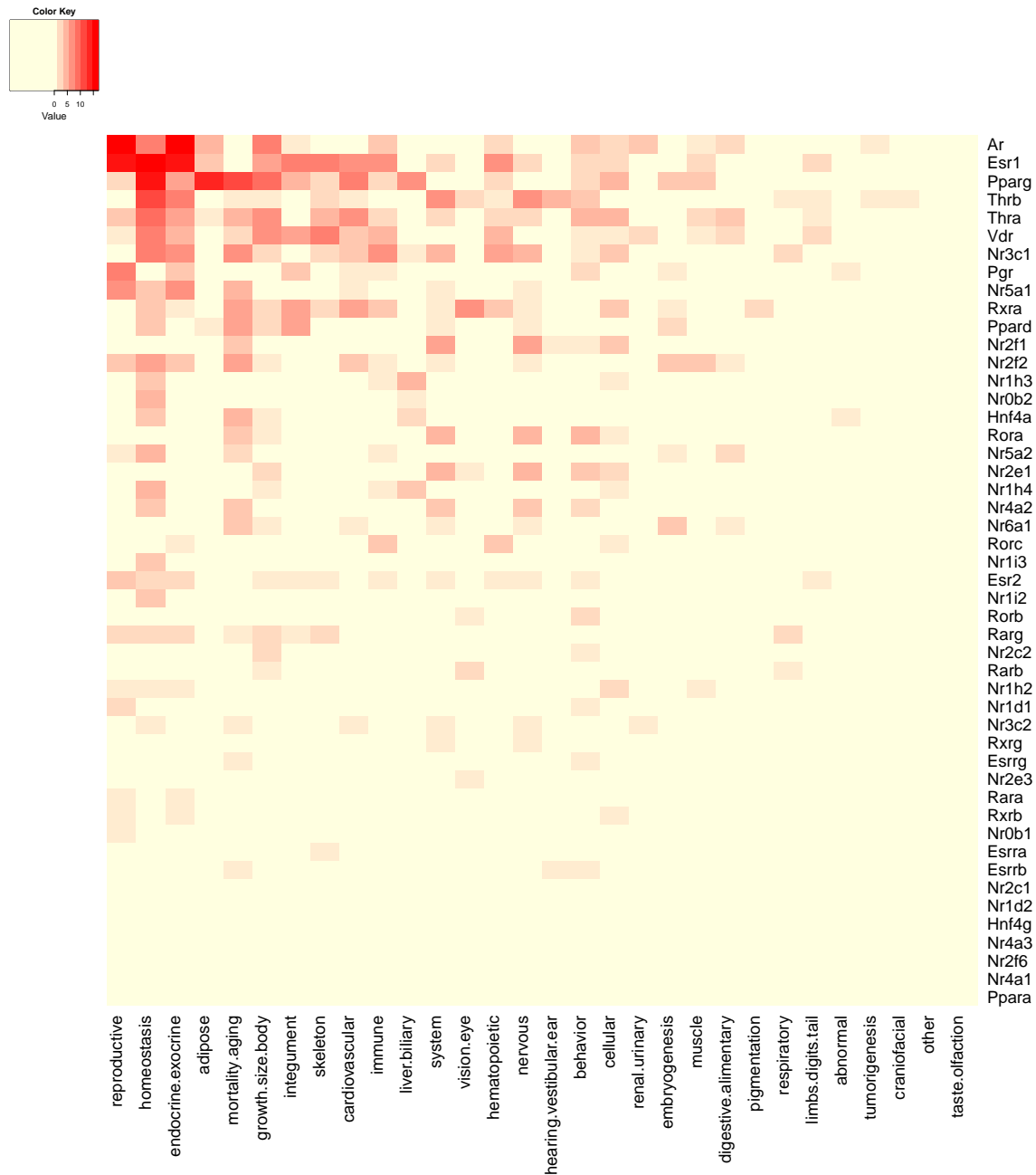
**Figure 3. MGI allele types.**  
Mouse Genome Informatics allele type distribution over the nuclear receptors.



**Figure 4. MPD phenotypes, having Z-score > 2.**  
Mouse Phenotype Database phenotype distribution over the nuclear receptors, having Z-score > 2

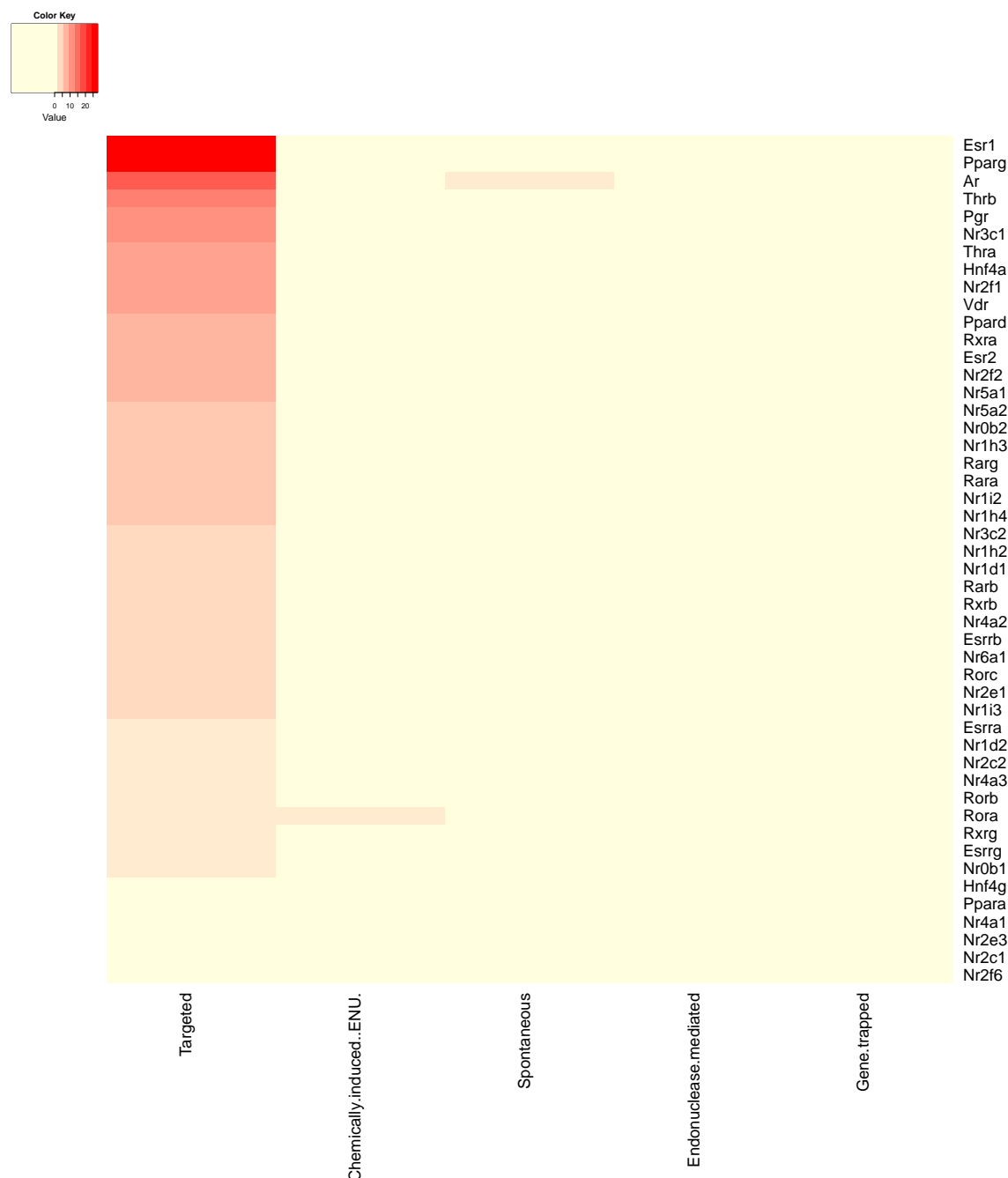


**Figure 5. MPD phenotypes, having Z-score < -2.**  
Mouse Phenotype Database phenotype distribution over the nuclear receptors, having Z-score < -2



**Figure 6. MGI phenotype - nuclear receptor associations.**

Mouse Genome Informatics phenotype occurrence frequency among the nuclear receptors.



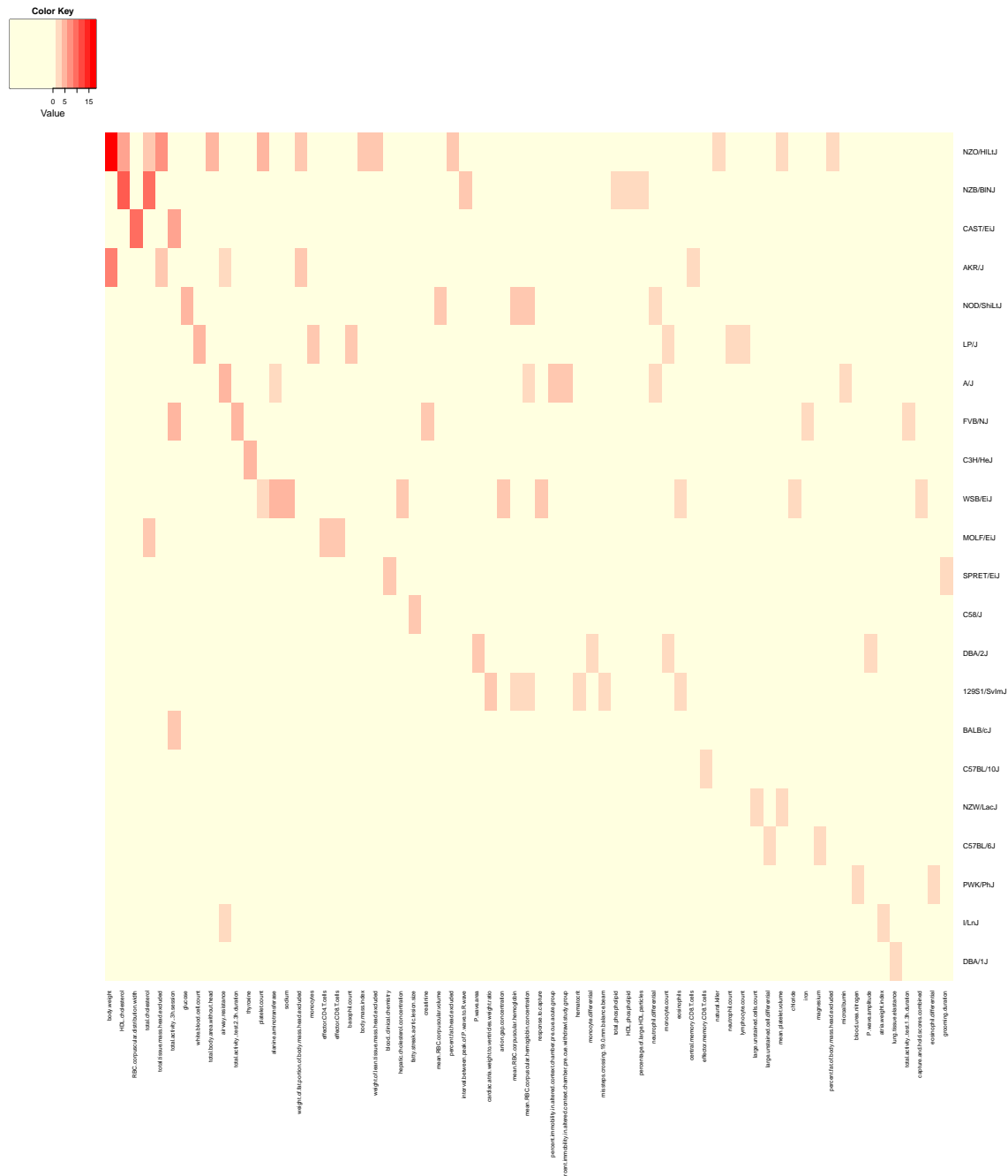
**Figure 7. MGI allele type - nuclear receptor associations.**  
Mouse Genome Informatics allele type occurrence frequency among the nuclear receptors.

Strain	Gene	Feature	Phenotype	Z-score	Occurrence
CAST/EiJ	Rora (RAR-related orphan receptor alpha)	protein coding gene	trigonelline relative abundance	8.870	2440
C58/J	Rorc (RAR-related orphan receptor gamma)	protein coding gene	N-acetylglutamate relative abundance	8.450	1398
BUB/BnJ		3-dehydrocarnitine relative abundance	8.420	1716	
NOD/ShiLtJ		glucose	7.840	2599	
NZW/LacJ		glycerate relative abundance	7.410	2230	
FVB/NJ		succinylcarnitine relative abundance	6.450	3335	
NZB/BINJ		cornea-lens synechia	6.250	1342	
SEA/GnJ		optic cup changes	6.170	912	
NZO/HiLtJ		body weight	6.130	1007	
129S1/SvImJ		latency to reach periphery 10 min test	6.080	3360	

**Table 1. Top 10 extreme phenotypes.**

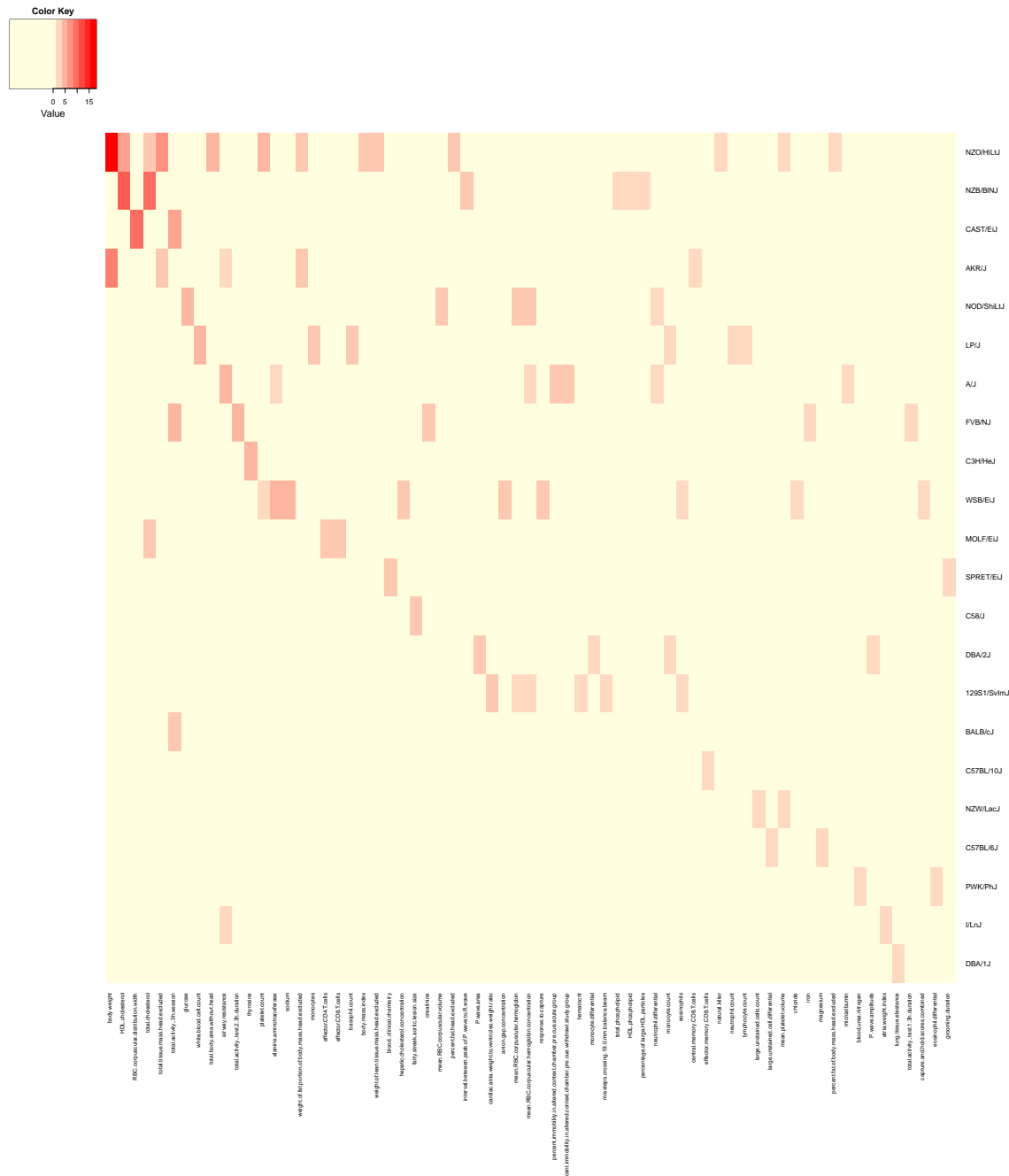
Mouse strains associated with extreme phenotypes, based on their Z-score.





**Figure 8. MPD phenotype - strain associations, having Z-score > 2.**

Mouse Phenome Database phenotype occurrence frequency among the strains associated with the nuclear receptors, having Z-score >



**Figure 9. MPD phenotype - strain associations, having Z-score  $< -2$ .**

Mouse Phenome Database phenotype occurrence frequency among the strains associated with the nuclear receptors, having Z-score <

-2

## 4. Discussion and Outlook

Blabla

## 5. Acknowledgments

Our gratitude to Prof. Dr. H. W. Mewes for offering us the topic and the opportunity to work with the Helmholtz 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 *Helmholtz 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.