Chromosomal clustering of tissue restricted antigens

Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Diplom - Biologist Maria Agnes Dinkelacker

born in: Göttingen, Germany

Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Diplom-Biologist Maria Agnes Dinkelacker

born in: Göttingen, Germany

Oral-examination:

Chromosomal clustering of tissue

restricted antigens

Referees:

Prof. Dr. Benedikt Brors

Prof. Dr. Ludger Klein

Acknowledgment

I hereby want to acknowledge all people who have helped me in all these years to do this research and work on my PhD thesis.

Most of my thanks go to my suporvisors Prof. Dr. Benedikt Brors of the German Cancer Research Center (DKFZ) in Heidelberg, Prof. Dr. Bruno Kyewski of the German Cancer Research Center (DKFZ) in Heidelberg as well as Prof. Dr. Roland Eils of the German Cancer Research Center (DKFZ) in Heidelberg for giving me the chance and the opportunity to work on this most exciting field of research and financing me through all these years.

Furthermore I want to thank HBIGS, the Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology of the Excellence Initiative of the University of Heidelberg for their financing of my childcare and their carreer support. In this respect I also want to thank to the womens politics of the German Research Foundation (DFG - Deutsche Forschungsgemeinschaft) for their change in politics towards a family friendly research to young scientific mothers.

I would like to acknowledge the German Cancer Research Center (DKFZ) for their wonderful family friendly research environment and thank the Helmholtz International Graduate School for Cancer Research from the DKFZ for their suport.

In the end my special thanks go to all immunologists helping me to substitute my suporvisor Prof. Dr. Bruno Kyewski after his leave of this world, especially Prof. Dr. Ludger Klein (LMU Munich), Prof. Dr. Mark Anderson (UC San Francisco) and Prof. Dr. Rienk Offringa (DKFZ).

I dedicate this work to my three little children Lea Sophie, Tim Jonas and Vincent Anton and I also want to thank my brother Prof. Dr. Friedrich Dinkelacker and my father Dr. Albrecht Dinkelacker, who gave me their heart and the critical way of scientific thinking and this way their support.

Summary

The every day response of our immune system is a delicately balanced system between protecting the body of foreign pathogens, such as worms, viruses, infections and other intruding particles on the one side and tolerating the cells of our body on the other side. This balance is conducted by educating the T cells of the immune system during their development in the thymus upon their specific T cell receptors (TCRs) on MHC molecules of the thymic stroma cells. During their development T cells undergo two checkpoints, the positive selection in the cortical part of the thymus on ctecs (corticular thymic epithelial cells) and the negative selection in the medullary part of the thymus on mtecs (medullary thymic epithelial cells).

While in the first step the TCR is tested to react at all to self MHC molecules on ctecs, in the second step the TCR is outselected if it reacts too strong to self MCH molecules loaded with self antigens on mtecs. This is an important step in order to make sure that T cells and with them the immune system tolerates self antigens and does not attack the body itself. This way it learns how to discriminate between self and foreign in order to ensure a healthy life of the individual. The peptides loaded to the self MHC molecules on mtecs are called tissue-restricted antigens (TRAs) and represent all tissues of our body.

If this presentation and selection procedure is disturbed, the individual suffers of severe autoimmune diseases, in which the immune system attacks the body. Therefore the study of TRAs is most important for the medical study of autoimmunity. In former times it was thought that TRAs are located on chromosomes in a random fashion and are also randomly presented upon surface molecules on mtecs such as MHC molecules. In this thesis we could prove that the localization on a chromosomal level of TRAs is not at all random, but rather organized in so called TRA clusters and furthermore also other genes of a common function are not randomly scattered but ordered in a chromosomal context. We could also show that the gene order is not only due to gene duplication, which was often argued in the past and might give insight into gene regulation in general.

One of the regulatory elements of TRA expression is known as the autoimmune regulator (AIRE). We calculated overlaps of AIRE regulated genes with our TRAs and found out that some, but not all TRAs are AIRE regulated, which highly suggests that there might be more transcription factors besides AIRE involved in the formation of central self tolerance.

With this thesis I could contribute to the elucidation of more knowledge about central self tolerance, tissue-restricted antigens, gene order and gene

expression in terms of chromosomal clustering of functionally related genes in general and TRA-DB is a useful tool which can also be used in the development of an immune therapy against cancer, since the problem of autoimmunity is a severe danger in the immunotherapy treatment in cancer patients.

Zusammenfassung

Jeden Tag in unserem Leben hat unser Immunsystem die Aufgabe, das fein ausbalancierte System zwischen dem Schutz unseres eigenen Körpers gegen Pathogene, wie Würmer, Viren, Infektionen und anderen Fremdpartikeln auf der einen Seite und der Toleranz unseres eigenen Körpers, unserer eigenen Zellen und unserer eigenen Gene auf der anderen Seite zu führen. Diese Balance zu halten wird dadurch erreicht, dass T-Zellen unseres Immunsystems während ihrer Entwicklung im Thymus aufgrund ihres spezifischen T-Zell Rezeptors (TCR) anhand der umliegenden Thymus-Stroma-Zellen selektiert werden. Während ihrer Entwicklung durchlaufen sie zwei Checkpoints, den der positiven Selektion auf kortikalen Thymus-Epithel-Zellen (cTECs) und den der negativen Selektion auf medullären Thymus-Epithel-Zellen (mTECs), anhand derer sie sowohl auf Reaktivität (positive Selektion) als auch auf Selbst Toleranz (negative Selektion) aussortiert werden.

Bei der positiven Selektion auf cTECs wird der spezifische T-Zell Rezeptor auf selbst MHC Molekülen getestet, ob er überhaupt reagiert und bei der negativen Selektion wird er auf selbst-MHC Molekülen getestet, die mit Selbstantigenen beladen sind. Dieser zweite Schritt der negativen Selektion ist essentiell, um sicher zu gehen, dass unsere eigenen T-Zellen und damit unser gesamtes Immunsystem unsere eigenen Gewebe toleriert und nicht angreift. Damit wird auch sichergestellt, dass unser Immunsystem den Körper beschützt, aber selbst nicht als fremd erkennt und hierzwischen unterscheiden lernt. Die Gene, die hierzu nötig sind, heissen gewebe-spezifische Antigene (TRAs). Sie repräsentieren sämtliche Gewebe des eigenen Körpers und werden auf medullären Thymus-Epithelzellen auf selbst-MHC geladen und den sich entwickelnden T-Zellen präsentiert.

Wenn in diesem Selektionsprozess irgendetwas falsch läuft, können schwere Autoimmunerkrankungen die Folge sein. Während man früher davon ausgegangen ist, dass Gene generell und insofern auch TRAs eher zufällig im Genom verteilt sind, so geht man heutzutage davon aus, dass Gene, die funktionell gekoppelt sind gruppiert vorliegen und daher gemeinsam reguliert werden können. In dieser Arbeit konnten wir zeigen, dass gewebe-spezifische Gene chromosomal gruppiert vorliegen, in medullären Thymus-Epithelzellen insofern gemeinsam reguliert werden können und die Natur einzelner dieser TRA-cluster im Detail dargestellt. Ferner konnten wir zeigen, dass auch andere funktionell gekoppelte Gene gruppiert vorliegen und dies eine allgemeine Genorganisation darstellen könnte. Während einer der Transkriptionsfaktoren, AIRE, bekannt ist, so scheint er aber nur einen Teil der hier gefundenen gewebe-spezifische Antigene zu regulieren. Es wäre interessant zu untersuchen, ob noch weitere Faktoren hier involviert sein könnten.

Die systematische Untersuchung gewebe-spezifischer Antigene hat es uns in dieser Arbeit möglich gemacht, weitere für Autoimmunerkrankungen wichtige Gene herauszufinden, ebenso wie das Auffinden weiterer wichtiger Gene, wie der Cancer-Testis Antigens (CTAs), die für eine Immuntherapie in Krebs wichtig sein können. Das Verständnis der zentralen Selbsttoleranz ist auch in diesem Kontext von grosser Wichtigkeit, da man nicht gegen ein Gen impfen kann oder das Immunsystem stimulieren, das bereits eine ausgeprägte Selbsttoleranz aufweist. Dies hat in der Vergangenheit zu einigen Fehlschlägen in der klinischen Anwendung bekannter CTAs geführt, wie im Falle von MUC1, ebenso wie in CEACAM6.

Ich konnte mit dieser Arbeit einen weiteren Schritt aufklären, wie die zentrale Selbsttoleranz funktioniert, welche Gene hier involviert sind, wie diese reguliert sein könnten und somit zur weiteren Aufklärung von Autoimmunerkrankungen beitragen, und hoffentlich später auch klinisch relevante neue Faktoren (Gene) finden, um Autoimmunerkrankungen in der Zukunft zu heilen.

List of Abbreviations and Glossary

ACKR4	atypical chemokine receptor 4
adaptive immune system	aquired immune system after the birth, including lymphocytes, B cells, T cells and NK cells.
Affymetrix chips	Affymetrix microarrays by Affymetrix, Santa Clara, CA.
AIRE	autoimmune regulator
antibody	secreted form of B cell specific BCR, immunglobulin, in 5 different allotypes, IgG, IgM, IgE with different effector functions.
APC	antigen presenting cell
APECED	autoimmune polyendocrinopathy candidiasis ectordermal dystrophy syndrom
autoimmunity	phenomenon, where the immune system attacks the body, illness.
autoimmune disease	Illness, where the immune system attacks the body.
B cells	immune cells, which derive in the bone marrow with their specific B cell receptor, and diferentiate upon stimulation to antibody producing plasma cells, B memory cells or regulatory B cells.
BCR	B cell receptor
Bioconductor	Bioconductor is an open software project for biologists, statisticians and bioinformaticians, supplying the CRAN network with biological applications.
BM	bone marrow is the place, where all precursors of the immune system derive, develope and outmigrate at one point in their life.
Brainarray	platform for microarray annotations.
CBP	creb binding protein
CCL17, 19, 22	chemokine C-C ligand 17, 19, 22
CCR4	C-C chemokine receptor type 4, 7
central self tolerance	self tolerance of T cells established through the negative selection of T cells in the thymus.
cell cycle genes	genes of the cell cycle, used as a control group for chromosomal clustering of TRAs.

chromosomal clustering	chromosomal clustering is here refered to as a gene order in direct neighborhoods of genes, which is organized in a higher manner, than random gene lists of the same length.
CLP	common lymphoid progenitor cell
CMP	common myeloid progenitor cell
conservation of TRAs	TRAs and clustering of TRAs is evolutionary conserved among different species.
cTECs	corticular thymic epithelial cells
DC	dendritic cells, immune cells specialized for antigen uptake
DLL4	delta ligand 4
DN	double negative state (T cells)
DNA	desoxy ribonucleic acid, the four letter code of the genome, consisting of the four bases A, C, G and T
DP	double positive state (T cells)
Ensembl Biomart	Ensembl Biomart - annotation server for gene annotation by the Ensembl database.
ETP	early thymic progenitor cell
evolutionary tree	TRA clustering was also tested down the evolutionary tree
FasL	fas ligand
FOXP3	forkhead box protein P3
gene duplication	chromosomal clustering of TRAs is not only due to gene duplication
GEO, gene annotation omnibus	database to store microarray data.
GMP	granulocyte megacaryocyte progenitor cell
granulocytes	granulocytes belong to the native immune system. They stimulate the adaptive immune system and react to infected cells, inflammations and intruding pathogens.
GTEX	The GTEX dataset is a human NGS dataset provided by the GTEX consortium, here used as the preprocessed RPKM values
hematopoesis	hematopoesis is the blood forming system in which for example the immune cells develope.

HSC	The hematopoetic stem cell, is the pluripotent precursor cell of all blood cells and all cells of the immune system.
housekeeping genes	housekeeping genes is the opposite of tissue specific genes, genes which are expressed basically in all tissues used for housekeeping functions.
intrahuman variability	The intrahuman variability reflects the variation in gene expression between human individuals, very nicely shown in the GTEX dataset.
lymphocytes	white blood cells, T cells, B cells and NK cells
MEP	megakaryocyte erythrocyte progenitor cell
microarrays	gene chips, where extracted RNA is transverted into cDNA, labeled with flourescent dyes and measured upon gene expression
MHC	major histo compatability complex, surface protein on cTECs, mTECs and other cells of the body
MHC I	MHC I proteins
MHC II	MHC II proteins
MHC loading prediction	netMHC and netMHCpan are MHC loading prediction tools, programmed by binding affinities on the basis of neuronal networks
MHC locus	gene locus with a highly diverse type of MHC accociated genes
MPP	myeloid pluripotent progenitor cell
mTECs	medullary thymic epithelial cells
myeolocites	immune cells of the myeloid lineage, granulocytes, erythrocytes
negative selection	selection procedure where T cells are selected upon self MHC plus self antigen, which in case of too strong binding are induced to undergo apoptosis.
NIK	NFκB inducing kinase
NK cells	natural killer cells, belong to the adaptive immune system
Novartis dataset	The Novartis foundation provided the here used human, mouse and rat microarray dataset
macrophages	macrophages are cells of the innate immune system, which phagocyte pathogens as well as cells, which are out of function, in an healthy individual for example cancer cells as well as infected cells.

Illumina chips	Illumina microarrays, bead arrays with labeled cDNA on beads, here preprocessed by the bead studio, an Illumina internal program for preprocessing
immunglobulin	Ig, protein of the antibodys, secreted form of the B cell receptor
Lattin dataset	The Lattin dataset is a murine microarray dataset, published by Lattin et al. 2008
LPS	lipopolysaccharide
pathogens	pathogens are worms, bacteria, viruses, funghi and parasites.
peripheral tolerance	peripheral tolerance is induced through stimulation of T cells to Treg cells, suppressing the immune answer to the specific antigen.
Perl	Perl is a free and platform independent scripting language, which can be easily used for biological applications, for example with packages, like bioperl.
positive selection	selection procedure of T cells upon self MHC molecules in the corticular part of the thymus, which in the positive effect leads to a survival signal and further development of T cells.
R, CRAN	R is an open source programming language under the GNU licence for statistical data analysis, which can be downloaded from the CRAN home page, CRAN stands for the Comprehensive R Archive Network, R packages can be sustained by Bioconductor packages, see above
RNA	ribonucleic acid, four letter code, A, C, G, U
RMA normalization	rma normalization, first normalization method developed for the normalization of microarrays
Roth dataset	The human Roth dataset is a microarray dataset used in this study
SCF	stem cell factor
self tolerance	self tolerance is the system where the immune system protects the body and does not attack its tissues or cells.
shell script	shell scripts are little programs, which can be run on the unix shell and direct other programs, such as R, Perl and Python
SIR	sphingosine-1-phostphate receptor

SP	single positive state (T cells)
T cells	immune cells, which derive from the bone marrow and further develope in the thymus
TCR	T cell receptor
thymus	organ behind the heart, where T cells are selected
TLR	toll like receptor
TRA	tissue-restricted antigens
TRA-DB	TRA-DB is database with tissue-restricted antigens
TRAF6	TNF receptor associated factor 6
Treg	regulatory T cells, which suppress the immune answer in the peripheral self tolerance [203].
vsn normalization	variance stabilization normalization, developed by Huber et al. for microarrays

List of Figures

1.1 Hematopoiesis 5
1.2 Developing T cells in the thymus 7
1.3 The autoimmune regulator AIRE and promiscuous gene expression 10
2.1 Quality control - single chip control in the mouse Novartis dataset 24
2.2 Scatterplot - quality control of the cerebral cortex 25
2.3 Quality control - single chip control of the human Novartis dataset 1 27
2.4 Quality control - single chip control of the human Novartis dataset 2 28
2.5 RNA degradation plot - quality control of the mouse Novartis dataset 31
2.6 Boxplot - quality control of the mouse Novartis dataset before normalization 33
2.7 Boxplot - quality control of the mouse Novartis dataset after vsnrma normalization 34
2.8 meanSdPlot - quality control of the mouse Novartis dataset . 35
2.9 Density plot - quality control of the mouse Novartis dataset before and after vsnrma normalization 38
2.10 Scatterplots - quality control of the human Novartis dataset . 39
2.11 Boxplot - after averaging double measurements in the mouse Novartis dataset 40
2.12 Distribution function - of multiple measurements per tissue type in the GTEX dataset 41
2.13 Definition of tissue-restricted antigens (TRAs) in the mouse Novartis dataset 42
2.14 Saturationplot - of defining and finding tissue-restricted antigens (TRAs) in the mouse Novartis dataset 45
2.15 Chromosomal clustering - The sliding 10-gene window method 47
2.16 Chromosomal clustering - The sliding-gene window method of fixed size 48
3.1 Tissue types and tissue grouping - from adipocytes to the hippocampus in all datasets 53
3.2 Tissue types and tissue grouping - from hypothalamus to the B lymphoblasts in all datasets 54
3.3 Tissue types and tissue grouping - from the b220 bcells to the spleen in all datasets 55
3.4 Tissue types and tissue grouping - from the stomach to the whole blood in all datasets 56
3.5 Frequency of tissue types in the human GTEX dataset 56
3.6 Overlap of TRAs in the human and mouse datasets 57

3.7 Overlap of housekeeping genes in the human and mouse datasets 59
3.8 Cellular functions of housekeeping genes in the mouse Novartis dataset 60
3.9 TRA examples for one tissue over the cutoff in the mouse Novartis dataset 61
3.10 TRA examples for two and three tissues over the cutoff in the mouse Novartis dataset 62
3.11 TRA examples for four and five tissues over the cutoff in the mouse Novartis dataset 63
3.12 Examples of non TRAs in the mouse Novartis dataset 64
3.13 TRA examples with one and two tissues over the cutoff line in the human GTEX dataset 66
3.14 TRA examples for three tissues in the human GTEX dataset 67
3.15 Examples of non TRAs in the human GTEX dataset 68
3.16 Insulin 1 and 2 involved in autoimmune diabetes type 1 . . . 75
3.17 CTRB2 and GCG being involved in autoimmune diabetes type 1 76
3.18 IAPP and PCSK1 being involved in autoimmune diabetes type 1 77
3.19 REG3A and SPINK1 being involved in autoimmune diabetes type 1 78
3.20 ACHE and BIN1 involved in autoimmune myasthenia gravis (MG) 85
3.21 ACYP2 and ACDY2 involved in autoimmune myasthenia gravis 86
3.22 MBP and PLP1 involved in autoimmune multiple sclerosis (MS) 104
3.23 MOG and MOBP involved in autoimmune multiple sclerosis (MS) 105
3.24 CYP11A1 and MC2R being involved in autoimmune addison's disease 123
3.25 MRAP and STAR involved in autoimmune addison's disease 124
3.26 S100a8 and S100a9 involved in juvenile idiopathic arthritis (SJIA) 125
3.27 TG and TPO involved in autoimmune hashimoto thyroiditis . 126
3.28 TSHR and FOXE1 involved in autoimmune hashimoto thyroiditis 127
3.29 ADAM2 and PIWIL1 testis-specific antigens (CTAs) 153
3.30 SPATA3 and ZNF546 testis-specific antigens (CTAs) 154
3.31 DIO2 and TSHR involved in autoimmune grave's disease . . . 155
3.32 NEK7 and IFI44 involved in systemic lupus erythematosus . 156
3.33 Chemokines with its corresponding tissue-specificity 1 166
3.34 Chemokines with its corresponding tissue-specificity 2 167
3.35 Number of TRAs in different tissue-types in the mouse and human Novartis dataset 169

3.36 Number of TRAs in different tissue-types in the mouse Lattin and human Roth dataset . . . 170
3.37 Number of TRAs in different tissue-types in the human GTEX dataset . . . 171
3.38 Chromosomal clustering of TRAs in the mouse and human Novartis dataset . . . 172
3.39 Chromosomal clustering of TRAs in the mouse Lattin and human Roth dataset . . . 173
3.40 Neighbourhood analysis of the mouse Novartis dataset . . . 175
3.41 Neighbourhood analysis of the human Novartis dataset . . . 176
3.42 Size and distribution of TRA clusters in the mouse Novartis dataset . . . 177
3.43 Tissue-types in TRA clusters on two examples of chromosome two and X in the mouse Novartis dataset . . . 179
3.44 Synteny plot between human, mouse and rat 1 . . . 180
3.45 Synteny plot between human, mouse and rat 2 . . . 181
3.46 Synteny plot of TRA clusters between human and mouse with tissue types . . . 182
3.47 TRAs in the evolutionary tree 1, chromosome 1, cluster 7 . . . 183
3.48 TRAs in the evolutionary tree 2, chromosome 2, cluster 8 . . . 184
3.49 Examples of TRA clusters with great diversity in terms of tissue-specificity . . . 185
3.50 Examples of TRA clusters with mono-clusters in terms of tissue-specificity . . . 186
3.51 Example of a TRA cluster tissue-specific for only one tissue-type 187
3.52 Homology plot in the mouse and human Novartis dataset . . . 189
3.53 Homology plot of random clusters in the mouse and human Novartis dataset . . . 190
3.54 The TRA database (TRA-DB) . . . 191
4.1 TRA numbers and percentages in the different datasets . . . 195
4.2 Overlap of TRAs in four different human datasets . . . 196
4.3 Commonly regulated gene families in TRA clusters . . . 201
4.4 Gene families within TRA clusters . . . 202
4.5 The KLK cluster in the human Novartis dataset . . . 203
4.6 The S100 cluster in the human Novartis data . . . 204
4.7 Tissue types represented in the TRA data in the mouse Novartis dataset . . . 209

List of Tables

2.1 Overview of datasets used for TRA detection 19
2.2 Brainarray packages used in this study 20
2.3 R and Bioconductor packages used in this study 21
2.4 Single chip quality control (QC) of microarrays - overview . . 30
2.5 Overview of the numbers of tissue grouping in all different datasets 37
3.1 Results - TRA numbers and percentages in all different datasets 52
3.2 TRAs tissue-specific for the pancreatic islet cells, potentially involved in autoimme diabetes type 1 74
3.3 TRAs tissue-specific for skeletal muscle, potentially involved in myasthenia gravis 84
3.4 TRAs tissue-specific for the central nervous system (CNS), potentially involved in multiple sclerosis (MS) 103
3.5 TRAs tissue-specific for the adrenal gland, potentially involved in autoimmune addison's disease 113
3.6 S100 gene family involved in autoimmune juvenile arthritis . 115
3.7 TRAs tissue-specific for the thyroid, potentially involved in hashimoto thyroiditis 122
3.8 TRAs tissue-specific for the testis, potential drug targets for cancer immunotherapy cancer-testis antigens (CTAs) 152
3.9 The roadmap of chemokines 165
4.1 The KLK cluster in the human Novartis dataset 199
4.2 The S100 cluster in the human Novartis dataset 200
4.3 The CSN cluster in the mouse Novartis dataset 200

Contents

1 Introduction **1**
1.1 The immune system and autoimmunity 1
1.1.1 Hematopoiesis 2
1.1.2 T cell development in the thymus 4
1.1.3 Central self tolerance 6
1.1.4 The autoimmune regulator (AIRE) 8
1.2 Tissue-restricted antigens (TRAs) 9
1.3 Housekeeping genes 12
1.4 Chromosomal clustering and gene organization 13
1.5 Aim of this thesis 15

2 Methods **18**
2.1 Datasets and pre-processing 18
2.1.1 The human and mouse Novartis datasets 18
2.1.2 The human Roth dataset 18
2.1.3 The mouse Lattin dataset 18
2.1.4 The human GTEX RNAseq dataset 19
2.2 Bioinformatical tools and programs 19
2.2.1 Gene expression omnibus (GEO) 19
2.2.2 Brainarray packages 20
2.2.3 The programming language R and Bioconductor . . . 20
2.2.4 Annotation of microarray data 20
2.2.5 Perl, mySQL and PHP 20
2.3 Technical background: microarrays and RNAseq data 22
2.3.1 Microarray chip technology: Affymetrix chips 22
2.4 Quality control (QC) of microarray data 23
2.4.1 QC in the mouse Novartis dataset 23
2.4.2 QC in the human Novartis dataset 23
2.4.3 QC in the rat Novartis dataset 26
2.4.4 QC in the human Roth dataset 26
2.4.5 QC in the mouse Lattin dataset 26
2.4.6 QC of the GTEX RNAseq data 26
2.4.7 QC of Microarrays 29
2.4.8 The RNA degradation plot (QC) 30
2.4.9 Variance stabilization (vsnrma) normalization 30
2.4.10 Mean versus standard deviation (meanSdPlot) 32
2.4.11 The density plot (QC) 32
2.4.12 The scatter plot (QC) 36
2.5 Averaging over multiple measurements per tissue 36
2.6 Tissue grouping for TRA detection 36
2.7 The definition of tissue-restricted antigens (TRAs) 37
2.8 The definition of housekeeping genes 44

2.9 The database of tissue-restricted antigens (TRA-DB) 44
2.10 Gene annotation 44
2.10.1 Annotation of microarrays with Biomart and Brainarray 44
2.11 Chromosomal clustering of tissue-restricted antigens 46
2.11.1 The 10-gene window method 46
2.11.2 The sliding-gene window method of fixed size 46
2.12 Interspecies comparison of human, mouse and rat 48
2.13 Interspecies comparison of TRA clusters 49
2.14 Homology plot within TRA clusters 49

3 Results **50**
3.1 Data analysis in order to detect tissue-restricted antigens . . 50
3.2 TRA definition 51
3.3 TRA numbers and percentages 51
3.4 Housekeeping genes numbers and percentages 52
3.5 Housekeeping genes in more detail 58
3.5.1 TRAs in the mouse Novartis dataset 58
3.5.2 TRAs in the human GTEX dataset 61
3.6 TRAs associated to autoimmune diseases 65
3.6.1 Diabetes type 1 65
3.6.2 Known autoantibodies in diabetes type 1 67
3.6.3 Myasthenia gravis 74
3.6.4 Multiple sclerosis 84
3.6.5 Addison's disease 103
3.6.6 Autoimmune gastritis 113
3.6.7 Juvenile idiopathic arthritis 113
3.6.8 Hashimoto thyroiditis 115
3.6.9 Testis-specific antigens 122
3.6.10 Grave's disease 152
3.6.11 Systemic lupus erythematosus 152
3.7 Summary of finding TRAs 152
3.8 The roadmap of chemokines in different tissue types 157
3.8.1 TRAs per tissue type 165
3.9 Chromosomal clustering of tissue-restricted antigens 168
3.9.1 The 10-gene window method 168
3.9.2 The neighbourhood analysis 174
3.9.3 Chromosomal clustering of housekeeping genes and other functionally related genes 174
3.9.4 Distribution of TRA clusters per chromosome 174
3.9.5 TRA clusters are an intermingle of different tissue types 174
3.9.6 Conservation of chromosomal clustering of TRAs . . . 178
3.9.7 Tissue specificity in TRA clusters 178
3.10 Intra-cluster homology analysis 188
3.11 TRA-DB: database of tissue-restricted antigens 188

4 Discussion **192**
4.1 Chromosomal clustering of tissue-restricted antigens 192
4.2 The selection of datasets and databases 193
4.3 Number of TRAs and proportion 194
4.4 Tissue-grouping, TRAs and housekeeping genes 197
4.5 TRA clustering 197
4.5.1 The kallikrein gene cluster 198
4.5.2 The S100 cluster 199
4.5.3 The casein cluster 200
4.6 Housekeeping genes numbers and percentages 205
4.7 Former attempts to find tissue-restricted antigens and define housekeeping genes 205
4.8 A TRA database (TRA-DB) 206
4.9 TRAs in autoimmune diseases 207
4.10 TRAs and cancer testis antigens 209
4.11 TRAs and systemic lupus erythematosus 210
4.12 Evolutionary conservation of TRA clusters 211
4.13 Alternative splicing 212
4.14 Conclusion 212

5 Technical appendix **272**
5.1 Part A: programming code 272
5.1.1 Data retrieval Microarray data 272
5.1.2 Data retrieval GTEX data 273
5.1.3 Calculating tissue-restricted antigens (TRAs) 273
5.1.4 Chromosomal clustering of TRAs 274
5.1.5 TRA plots and plotting chromosomal clustering of TRAs 275
5.1.6 Gene annotation and TRA tables 275
5.1.7 Plot TRA clusters 275
5.1.8 Aire genes in TRAs 276
5.1.9 Homology plots 276
5.1.10 Synteny maps 276
5.1.11 Calculate paired t-tests 276
5.2 Part B: Additional figures 277
5.2.1 Quality control 277
5.2.2 Quality control - RNA degradation plots 277
5.2.3 Boxplots of all datasets 277
5.2.4 Density plot before and after vsnrma normalization . 278
5.2.5 Tissue types in the different datasets 278
5.2.6 Saturationplots in the different datasets 278
5.2.7 Some example TRAs in all different datasets 278
5.2.8 TRA clusters per chromosome in the human and mouse Novartis dataset 280

5.2.9 Synteny plots of TRA clusters between human, mouse and rat 282
5.2.10 Synteny maps on tissue-specificity in TRA clusters between human and mouse 282
5.2.11 conservation of TRAs in the evolutionary tree 283
5.2.12 TRA clusters in the mouse 283
5.2.13 TRA clustered genes in the mouse 284
5.2.14 TRA clustered tissues in the mouse 284

1 Introduction

1.1 The immune system and autoimmunity

The subject of immunology, the science of the human and vertebrate immune system, came into focus when Edward Jenner performed in 1796 the first vaccination against smallpox. He vaccinated a human with the cowpox virus and could this way start a process that finally proved to erase the illness of smallpox [176]. In 1876 Robert Koch discovered that microbes can cause diseases. In 1876 he found *Bacillus anthracis*, and in 1882 *Mycobacterium tuberculosis*. In 1878 Paul Ehrlich described the first immune cells, the mast cells, in his doctoral thesis [332, 42]. Over the years, immunology became widely accepted.

The human immune system is composed of the innate immune system, which is fully established at birth, and the adaptive immune system, which has to be acquired after birth. While cells of the innate immune system recognize conserved structures, for example components of bacterial cell walls, the adaptive immune system is regulated through stimulation of its cell surface receptors, such as T cell and B cell receptors. In 1890 Emil von Behring could demonstrate the first antibody activity opening up the field of humoral immunity [192]. In 1961 Mac Farlane Burnet found out that antibodies are the secreted B cell receptors which resolved a long standing discussion between two groups of scientists over many years [79].

The cells that make up the adaptive immune system are lymphocytes, which derive from a common hematopoietic stem cell (HSC) in the bone marrow (BM), to differentiate into T cells, B cells, and natural killer cells (NK cells). Cells of the innate immune system are dendritic cells (DCs), macrophages, granulocytes and also a subgroup of natural killer cells, which can be counted to both systems [201]. While B cells remain for their final development in the bone marrow, T cells migrate to the thymus and further differentiate there. Both T cells and B cells have their individual surface receptors which are highly variable and can thus recognize foreign antigens.

While the immune system has to protect our body from intruding pathogens, such as worms, bacteria, viruses, fungi and parasites, it has to tolerate on the other side our body and its own tissues. This requires that the adaptive immune system learns and understands how to discriminate between foreign and self [214]. If this learning of so-called self-tolerance is not well functioning, the individual will suffer from multiple autoimmune diseases, which are in many cases very harmful. There are two main tolerance mechanisms, the central self-tolerance, where potentially autoreactive T cells are eliminated in the thymus, if they react and bind too strongly to self-antigens, and the

peripheral tolerance, where autoreactive T cells are induced to become inert or develop into regulatory T cells. These two main mechanisms stabilize the human immune system in order to protect our self [408].

In order to detect and destroy auto-reactive T cells, auto-antigens need to be presented in the thymus. This is mediated by key mechanisms. The auto immune regulator AIRE promotes promiscuous gene expression in medullary thymic epithelial cells (mTECs). Tissue-restricted antigens (TRAs) can also be cross-presented in the medullary part of the thymus on thymic dendritic cells (DCs). Both result in presentation of self-antigens to developing T cells for negative selection of potentially autoreactive T cells [214, 102].

Promiscuous gene expression of tissue-restricted antigens means that otherwise tissue-specifically expressed genes are commonly expressed in mTECs in the thymus and are presented to developing T cells. In case of a too strong binding of the antigen-specific T cell receptor (TCR), T cells are co-stimulated by an apoptosis signal and are selected out before being released into the peripheral blood [214].

T cells are also positively selected in the cortical part of the thymus, where they are presented to self antigens bound to MHC ligands and receive a co stimulatory survival signal in case of strong enough binding [204]. Only 5% of all developing T cells survive both selection procedures [214]. The common hematopoietic stem cell (HSC) in the bone marrow gives not only rise to T cells, B cells or other immune cells, but also to erythrocytes as well as platelets. While erythrocytes are important for oxygen transport in the blood, platelets are needed for blood clotting and wound healing. For an overview of the hematopoietic system please refer to Fig. 1.1.

1.1.1 Hematopoiesis

Hematopoiesis is the formation of the blood cells which constitute a major part of the immune system. It occurs during embryonic development as well as during the adult life [175]. It starts with the development of the totipotent and self-renewing hematopoietic stem cell (HSC) in the bone marrow, which differentiates into all immune cells as well as into erythrocytes and platelets [201]. In all vertebrate species the hematopoiesis starts already in early embryogenesis. It involves two waves of developing blood cells, the primitive wave and the definitive wave [124]. The primitive wave involves erythroid progenitor cells, which give rise to erythrocytes and macrophages whose purpose is to produce red blood cells in order to oxygenate the embryo [279]. The primitive wave is only transitory, and soon replaced by the definitive wave, its progenitor cells are not pluripotent and do not renew themselves in contrast to the hematopoietic stem cell of the adult organism.

In humans hematopoiesis is starting in the yolk sac, continuing in the liver and finishing up in the bone marrow and thymus. The hematopoietic stem cell (HSC) is, in contrast to its offspring, the only self-renewing immune cell and is present in close proximity to endothelial cells [379].

In primitive hematopoiesis the main transcription factor is GATA1, which drives the development of the HSC into the lymphoid lineage, while the transcription factor PU.1 propagates the development into the myeloid lineage. Both indirectly supress the other line of development [55, 345]. In the adult hematopoiesis Runx1 is a transcription factor which is needed for the development of both the myeloid as well as the lymphoid lineages. While the hematopoietic stem cell is the last self-renewing cell, its offsprings are further differentiated and lineage-commited cell lines. Both Wnt as well as Notch signalling may play a role in the self renewal of HSCs. While the influence of Notch signalling is well perceived, the role of Wnt signalling is still under debate. Besides these two signalling pathways also the microenvironment plays a major role in hematopoiesis of HSCs, and the bone marrow seems to be a suitable niche for the development of the blood forming system [175].

The first offspring of the hematopoietic stem cell is the multipotent progenitor cell (MPP) (Fig. 1.1). It has, in contrast to the hematopoietic stem cell, only a limited repopulating capacity as well as a finite self-renewing potential. The MPP can commit either to the lymphoid or the myeloid line [201]. First MPPs loose their commitment to the megakaryocyte/erythrocyte lineage (MEP) megakaryocyte-erythrocyte progenitor cell, next they loose their commitment to the myeloid line (CMP and GMP), the common myeloid progenitor cell, as well as the granulocyte-macrophage progenitor cell and last it differentiates into the common lymphoid progenitor cell (CLP), which gives then rise to T cells, B cells and NK cells. The differentiation step between B cells and T cells is mainly influenced by transcription factors. The feed-forward regulatory cascade of the B cell development is mainly driven by PU.1, E2A, EBF1 and Pax5. EBF1 is the main transcription factor of the B cell line, regulated through IKAROS, E2A and PU.1 and extrinsic IL-7 signaling, and represses the myeloid genes while ensuring a stable B cell commitment [315].

While the determination of the B cell lineage is defined early in their development, the differentiation of the T cell line is more dependent on environmental factors of the thymic microenvironment. In the case of T cells the main lineage commitment is driven by Notch receptor signaling which is stimulated by Delta-like 1 and 4 expressing cortical thymic epithelial cells (cTECs). Through its signaling, lymphocyte progenitor cells (LMPPs, lymphocyte multipotent progenitor cells) are turned into the pro T cell developmental program, which activate the typical T cell genes GATA3, TCF7,

while suppressing EBF1, E2A and PU.1 of the B cell line and ID2 of the NK cell line [315]. Whether or not lineage commitment is unidirectional is still an ongoing question [340].

1.1.2 T cell development in the thymus

The thymus is the final site of T cell differentiation. It is located right behind the heart. Lymphoid progenitor cells are attracted to the thymus via chemokine signalling. The seeding of T cells both in the fetal thymus as well as in the adult thymus is occuring in waves and conducted mainly by the expression of CC-chemokine ligands CCL21, CCL25 and its equivalant receptors CCR7 and CCR9 on lymphoid progenitor cells [377]. In the adult thymus the thymic seeding of T cells is also guided through the expression of P-selectin and its PSGL1 ligand on thymic endothelial cells [326].

After entering the thymus, developing T cells go through different stages, starting with the double negative (DN) CD4– CD8– state. They show, besides CD4– and CD8– during this time, a CD25+ CD44– phenotype, the DN3 state, and move attracted by chemokine signalling from the corticomedullary junction to the subcapsular zone [377]. This movement is guided by Notch-mediated signaling, binding to Delta ligands, as well as by IL-7. During this developmental step they rearrange their TCR *beta* chain, and only T cells with an in-frame rearrangement can go on in their development. It is a two-sided interaction of signalling and development between thymic epithelial cells and developing thymocytes which helps them to differentiate and to interact. Also the immune cells help the cortical thymic epithelial cells (cTECs) to differentiate by expressing keratin 5 and keratin 8, which are important for the maturation of cTECs [377].

During the outward movement of T cells to the subcapsular zone the chemokine receptors CXCR4, CCR7 and CCR9 play a role, and the assembly of the TCR *beta* chain together with the pre-TCR *alpha* chain forms the pre-TCR complex. Double negative T cells go over into the double-positive state expressing a TCR *alpha beta* antigen receptor [377]. Through a low-avidity interaction between DP T-cells and cortical thymic epithelial cells (cTECs), T cells get positively selected and can move on into the medullary part of the thymus [43]. About 3-5% of all cells survive this positive selection procedure [110, 138].

After the positive selection T cells develop into CD4+ CD8– or CD4– CD8+, single-positive (SP) T cells. Through the expression of the chemokine receptor 7 (CCR7) they are attracted to the medullary part of the thymus, completing their journey and final development. Besides the interaction with thymic epithelial cells, mostly mTECs, they interact also with accom-

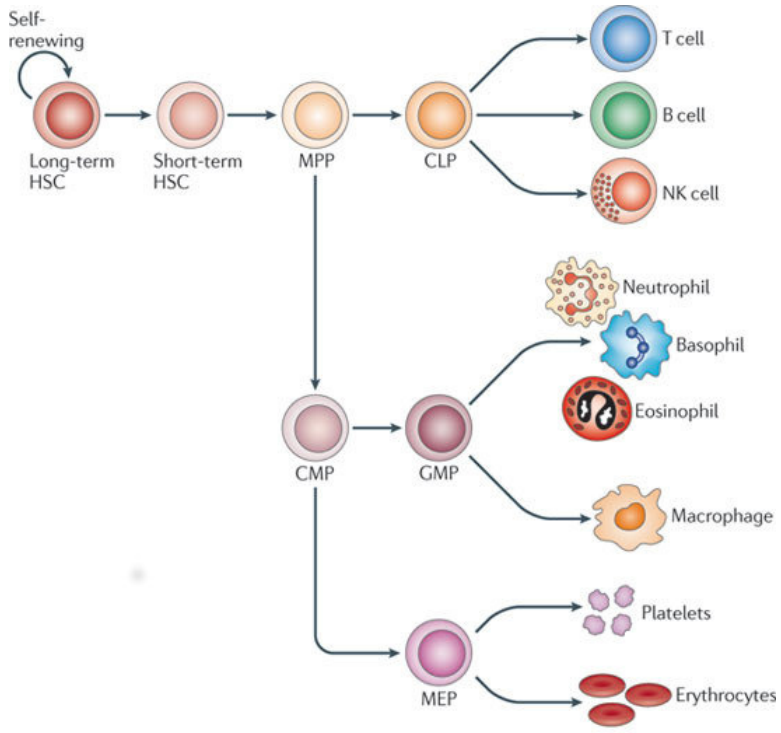


Figure 1.1: **Hematopoiesis:** All immune cells derive from a *common hematopoietic stem cell (HSC)* in the bone marrow, which then further differentiate into a *common myeloid progenitor cell (CMP)* and a *common lymphoid progenitor cell (CLP)*. While the CLP gives rise to cells of the adaptive immune system, such as *T cells (TCs)*, *B cells (BCs)* and *NK cells*, the common myeloid progenitor cell differentiates into a *granulocyte macrophage progenitor cell (GMP)* and a *megakaryocyte erythrocyte progenitor cell (MEP)*. The GMP gives rise to the cells of the innate immune system, such as *granulocytes* and *macrophages*, the MEP forms into platelets for blood clotting as well as wound healing and into *erythrocytes*, which are important for the oxygen transport in the body. (Figure taken from King et al. 2011 [201], reprinted with permission by Springer Nature).

panying stromal cells, such as dendritic cells (DCs) and macrophages. In addition to the above cited chemokines other factors, such as NFκB, the lymphotoxin-β-receptor (TRAF) and NFκB inducing kinase (NIK) play a critical role [377, 37, 9].

During this journey developing thymocytes spend about twelve days only in the medulla, highlighting the fact that the interaction with thymic epithelial cells is very critical. Besides CD4+ and CD8+, developing thymocytes also express the L-selectin CD62L as well as the P-selectin CD69. While CD4+ T cells later on in their development develop into T helper cells, CD8+ cells develop into cytotoxic T cells. Besides these two subgroups the third major subgroup of T cells are the Foxp3+ regulatory T cells. These are also attracted to medullary stroma cells such as DCs and Hassals's corpuscles via chemokine signalling, mainly by CCL17 and CCL22 binding to the surface receptor CCR4 [119, 10, 249]. The role of regulatory T cells is in contrast to T effector cells such as CD4+ and CD8+ T cells is to down-regulate the immune system, rather than to stimulate it (Fig. 1.2).

1.1.3 Central self tolerance

During their journey through the thymus, T cells reshuffle their T cell receptor. Through a recombination of different genes, the potential T cell specificity is very high. This has a drawback to create also T cell receptors without the right specificity for self-MHC as well as to the potentially overreact towards self-antigens. In order to ensure a sufficiently diverse T cell pool, but avoid autoimmune diseases, T cells undergo two checkpoints, the positive selection in the cortical part of the thymus and the negative selection in the medullary part of the thymus. Only 5% of all T cells survive this selection procedure [377].

While T cells are tested upon self MHC in the positive selection, they are selected upon self-MHC-self antigen ligand in the negative selection. In case of not binding strongly enough to self MHC on cTECs, T cells do not get a survival signal and undergo a death by neglect. In case of too strong binding to self-MHC-self-antigen on medullary thymic epithelial cells, T cells get an apoptosis signal and are induced to undergo apoptosis or are turned into regulatory T cells. In addition to the central-self tolerance there is also peripheral tolerance outside the thymus, where T cells can be induced to become inert.

In the process of negative selection tissue-restricted antigens are expressed *"ectopically"* in the thymus representing otherwise externally expressed tissue-restricted antigens. This *"promiscuous gene expression"* was first detected in 1994, where the first self-antigen was found to be expressed in the thy-

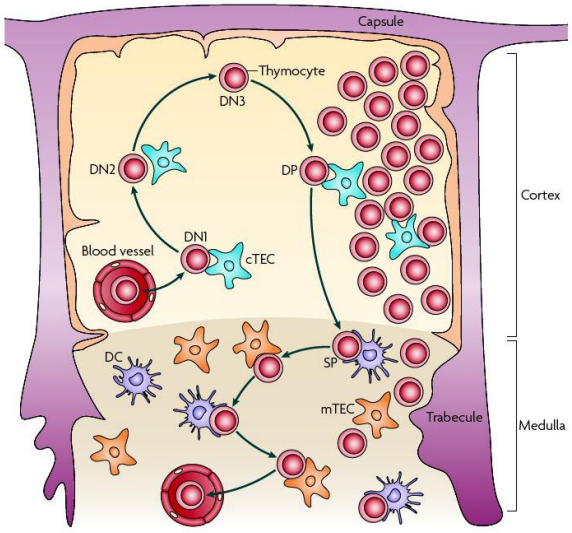


Figure 1.2: **Developing T cells in the thymus:** After entering through blood vessels, the developing T cells are migrating through the thymus. They undergo two main checkpoints, the *positive selection* in the cortical part of the thymus, and the *negative selection* in the medullary part of the thymus. During this time the T cells are reshuffling their T cell specific *T cell receptors (TCRs)*, passing by *corticular thymic epithelial cells (cTECs)*, *dendritic cells (DCs)* as well as *medullary thymic epithelial cells (mTECs)*. During this time the T cells pass different stages by the presence or absense of their surface receptors, the *double-negative stage (DN, CD4– CD8–)*, the *double-positive stage (DP, CD4+ CD8+)*, and the single positive stage (SP) depending on their specificity *CD4+* for the *T helper cells*, or *CD8+* for the *cytotoxic T lymphocytes*. After passing the journey through the thymus, the differentiated T cells are released into the peripheral blood. (Figure taken from Klein et al. 2009 [204], reprinted with permission from Springer Nature).

mus (Pugliese et al. 1994). This finding of the first rat insulin important in autoimmune diabetes type 1 was followed by the finding of more tissue-restricted antigens in the following years. Among them are the insulin gene (INS) in the mouse, glucagon (GLC), the pancreatic polypeptide (PP), somatostatin, trypsin as well as the elastase tissue-specific for the pancreatic islet cells [184], the myelin basic protein (MBP), albumin, GABDH, the thyreoglobulin (TG), the thyroid peroxidase (TPO) [366], the acetylcholine receptor (ACHR), IRBP, CRP and the SAP gene, the myelin oligodendrocyte glycoprotein (MOG), and IA2 gene [333, 202, 98].

Through finding of the proteolipid protein PLP expressed in the thymus, which is tissue-specific for the central nervous system and involved in multiple sclerosis, alternative splicing came into the focus of tissue-specific gene expression [203]. The finding of the autoimmune regulator AIRE finally explained how this gene regulation of tissue-restricted antigens in the thymus might be regulated, since AIRE regulates many TRAs in the thymus [265, 4, 99, 16].

In 2004 Gotter et al. found that tissue-restricted antigens are clustered on a chromosomal level. They described that most TRAs represent different tissue-types of the body. In 2005, Derbinski et al. found that in the differentially expressed genes between mTECs and cTECs most genes are colocalized in chromosomal clusters. Three of these gene clusters of size up to sixteen genes in a cluster were studied in further detail. This includes the kallikrein cluster on chromosome 7 in the mouse, the S100 cluster on chromsome 3 in the mouse and the casein locus on chromosome 5 in the mouse. Many of these genes they found to be AIRE -regulated. Also Johnnidis et al. 2005 found chromosomal clustering of genes controlled by AIRE [144, 99, 182].

1.1.4 The autoimmune regulator (AIRE)

The autoimmune regulator AIRE has been first discovered by the Finnish-German APECED Consortium in 1997 [1]. AIRE has been linked to the failure of central self-tolerance and thus the establishment of multiple autoimmune diseases. Mutations in AIRE have been reported to be linked to APECED (autoimmune polyendocrinopathy-candidiasis-ectodermal- dystrophy) as well as to APS-1 (autoimmune polyendocrine syndrome type 1), an autoimmune failure in many different tissues of the body [265]. The murine equivalent AIRE has been found in 1999 by positional cloning with three different splice variants [409], and it was found to be expressed in the thymus [33]. Also the authors noticed its gene expression to be an important factor in hematopoiesis [260]. In addition, the gene expression of AIRE has also been found in other immunological relevant tissues, such as the lymph

nodes, spleen as well as in the fetal liver. A role of AIRE in the induction of immune tolerance has been suggested [155, 205].

AIRE is a transciption factor, whose encoding gene is 500kb long [265], including two zinc finger domains and one SAND domain [131]. It is a chromatin-associated protein, which remodels chromatin folding [131]. AIRE interacts with many other associated proteins, such as the common transcriptional coactivator CREB-binding protein (CBP). It recruits further transciption factors such as Jun, Fos, NFκB and STAT and this way enhances transcriptional regulation [299]. Heino et al. found 42 different mutations in the AIRE protein in 200 APECED patients, which caused addison's disease, hypoparathryroidism and type-1 diabetes [156]. In 2001 Kumar et al. suggested that recombinant AIRE oligomerizes spontaneously upon phosphorylation with the cAMP dependent protein kinase A or C and forms dimers [213].

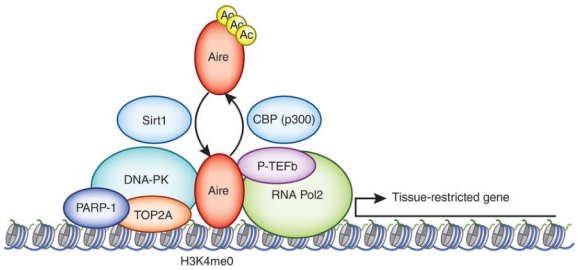
Through systematic studies on AIRE knock-out mice both Anderson et al. 2002 as well as Derbinski et al. 2001 found that the transcription factor AIRE regulates a battery of tissue-restricted antigens, whose failure results in multiple autoimmune diseases. They suggested for this the role of AIRE in the negative selection of T cells in the thymus [16].

In the last couple of years AIRE has been intensively studied, it acts in concert with several other proteins, such as CBP, the poly ADP ribose polymerase 1 (PARP-1), the topoisomerase 2a (TOP2A), the positive transcription elongeation factor (p-TEFb) and many others (Fig. 1.3).

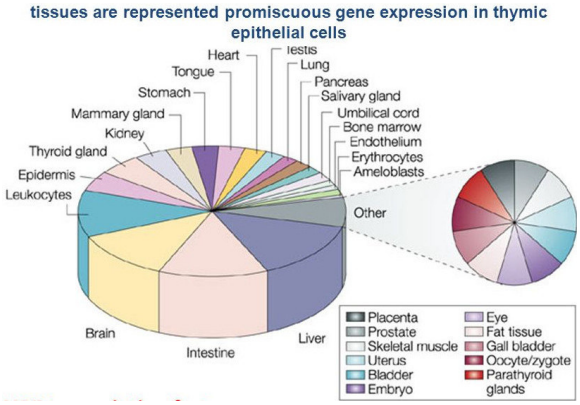
1.2 Tissue-restricted antigens (TRAs)

Tissue-restricted antigens (TRAs) are genes which are highly expressed in a few tissues of the body in comparison to other tissues. They stand in contrast to housekeeping genes, which are expressed in many tissues of the body. Tissue-restricted antigens are important to characterize the protein content of certain tissues, but also in understanding autoimmune diseases. TRAs are *"promiscuously"* expressed in the thymus and presented to T cells on MHC II molecules by medullary thymic epithelial cells (mTECs). T cell depletion is taking place under the influence of apoptotic signals and will remove all autoreactive T cells [214, 102].

In case of AIRE deficiency, or other regulatory problems of the negative selection, e.g. a destructive disorder of the thymic microenvironment or deficiencies in other transcription factors affecting the TRA expression, such as Sirt1, autoimmunity is the result. The first to detect the expression of tissue-restricted antigens in the thymus was the group of Jolicoeur et al. in



(a)



(b)

Figure 1.3: **The autoimmune regulator AIRE:** (a) The autoimmune regulator AIRE upregulates the transcription of tissue-restricted antigens in medullary thymic epithelial cells (mTECs). It binds to ummethylated Lys4 of histone H3 (H3K4me0) on the chromatin and interacts with a group of chromatin-bound proteins which promote transcriptional elongeation (P-TEFb) and DNA double strand breaks (DNA-PK). The Protein Sirt1 deacetylates lysine in AIRE and this way promotes TRA expression, while CBP, the CREB binding protein acetylates AIRE and opposes the TRA transcription. AIRE works together with the DNA dependent protein kinase (DNA-PK), the RNA polymerase, RNA Pol2, the topisomerase 2a (TOP2A). (Figure taken from Peterson et al. 2015 [294], reprinted with kind permission from Springer Nature). (b) The variety of tissue-specificity of TRAs that are regulated by AIRE. (Figure taken from Kyewski et al. 2004 [214], reprinted with permission from Springer Nature).

1994. They found the expression of a Tag antigen in pancreatic β cells in the thymus and could also show that the insulin gene involved in diabetes type I is virtually expressed in thymic epithelial cells. From this they conclude that gene expression of tissue-restricted antigens in the thymus could generally play a role in the selection of autoreactive T cells and prevent autoimmune diseases [184].

Further tissue-restricted antigens have been found to be expressed in the thymus in the following years, for example the human C reactive protein (hCRP), as an acute phase protein of the hepatocytes. For this they also discovered that a certain threshold in order to eliminate autoreactive T cells is necessary (avidity model) [202, 203]. This finding was providing the importance of central-self tolerance versus peripheral tolerance for the deletion of autoreactive T cells. Gene expression of the C reactive protein in the thymus in their study led to the physical elimination of the autoreactive T cell. This clonal deletion of T cells is of major importance for the establishment of central self tolerance. Furthermore they observed a receptor down-regulation as well as a functionally inactivation of the responding T cells. Tissue-restricted antigens were presented to T cells on MHC II molecules on medullary thymic epithelial cells as well as on dendritic cells through cross-presentation [202, 203].

Visan et al. could show this phenomenon for another tissue-restricted antigen in 2004, the myelin protein zero, which is over-expressed in the thymic periphery, and its gene expression is coupled to the autoimmune disease of motor sensory neuropathy, a disorder of the peripheral nervous system leading to severe neurological disabilities [404]. In the meantime Anderson et al. as well as Klein et al. found that the myelin proteolipid protein (PLP) involved in multiple sclerosis (MS) is also expressed ectopically in the thymus [203, 14].

All these observations taken together led to the conclusion that the expression of tissue-restricted antigens (TRAs) in the thymus is a physiological property of the thymus, especially in the medullary part of the thymus and particularly in medullary thymic epithelial cells (mTECs) [214]. Derbinski et al. found more self antigens, such as GAD65, S100, Insulin and other molecules to be expressed *"ectopically"* in the thymus as well as being involved in autoimmune diseases [98].

Knock-out of the transcription factor AIRE in mice led to the detection of all genes up- and down- regulated in mTECs [16, 99]. Furthermore the *"terminal differentiation model"* proposed by Derbinski et al. [98, 214] has been shown to be true in terms of tissue-restricted gene expression by MHC II hi, CD 80 hi mTECs versus MHC II lo, CD 80 lo mTECs [297]. The

number of genes upregulated as well as the number of AIRE upregulated genes raises from cTECs to mTECs and within the mTEC pool from MHC II lo CD 80 lo towards MHC II hi CD 80 hi most differentiated mTECs [297].

Velculescu et al. found the first tissue-restricted antigens. Some of them have been annotated already during this time to be tissue-specific for the colon, such as the guanylate cyclase activator (GUCA2B) [401], the pregnancy-specific *beta* 1 glycoprotein 6 (PSG6), tissue-specific for the placenta, the aquaporine 8 (AQP8), tissue-specific for the large intestine, and placenta, the alkohol dehydrogenase 1 (ADH1A), tissue-specific for the liver, LY6H tissue-specific for the central nervous system, the myelin-associated oligodendrocytic basic protein (MOPB) tissue- specific for the central nervous system, the Ll-cadherin (CDH17) tissue-specific for the large intestine as well as the calmodulin-stimulated phosphodiesterase (PDE1B1), tissue-specific for the central nervous system [401].

1.3 Housekeeping genes

Housekeeping genes are genes which are expressed in many if not all tissues and thereby stand in contrast to tissue-restricted genes. Housekeeping genes are thought to fullfill housekeeping functions [52]. Housekeeping genes have several times been tried to be defined. She et al. found 1,522 housekeeping genes and defined them as genes with a low variance [348]. Lercher et al. declared housekeeping genes to be strongly clustered in human and showed that housekeeping genes evolve much slower than tissue-restricted genes due to evolutionary pressure [222, 455]. Vinogradov et al. stated in 2004 that housekeeping genes are much shorter than tissue-restricted genes [402].

Several groups have made an attempt to characterize housekeeping genes. They define housekeeping genes in different ways. Weber et al. 2011 define a housekeeping gene as a gene, which is expressed in fourteen out of fourteen tissues [418]. What they mean by *"is expressed"* is not further explained. In 1998 Werdelin et al. stated that the number of tissue-restricted genes expected in the genome was probably higher than the number of housekeeping proteins [424]. But what they exactly meant by either of them, was not clarified. Eisenberg et al. found in 2003 575 housekeeping genes by identifying housekeeping genes as genes, which are expressed constitutively [111]. They used for their calculations however the same dataset by Su et al. 2002 and Su et al. 2004 [371, 372] as we have used in this work. Zhang et al. defined in 2004 a housekeeping gene as a gene, which is always expressed in any tissue and is there in order to maintain cellular functions [455]. Watson et al. wrote in 1965 about housekeeping genes, but did not further specify a definition. Lercher et al. 2002 defined a housekeeping gene as a gene that

is expressed in 9 out of 14 different tissues [222], without saying what the term *"is expressed"* means, and Zhang et al. specified this definition in 2004 into the definition that a housekeeping gene is a gene which is expressed in at least 19 out of 60 tissues [455, 222].

Butte et al. state that a housekeeping gene is a gene, which is *"constitutively expressed"* and is there to *"maintain cellular function"*[52]. Veculescu et al. identified a first *"starter set"* of housekeeping genes in 1999 [52, 401]. Warrington et al. find 535 *"maintenance genes"* as *"likely candidates for housekeeping genes"* [52, 415]. They state, that housekeeping genes are *"consitutively expressed"* [52]. Hsiao et al. identified any gene expressed in 19 tissues as *"housekeeping"* or *"maintenance"* genes [52, 160]. They found 451 genes to be maintenance genes, coding for proteins mediating cellular functions, such as transcription, translation as well as signaling [52]. A precise definition of a housekeeping gene was not given and differentiated to a tissue-specific gene, which would be a housekeeping gene of tissue-specific function.

Velculescu et al. posed the question *"how many human genes are expressed ubiquitously, in all human tissues, and how many are expressed in a tissue-specific pattern"* [401]. they analyzed 3.5 million transcripts from 19 normal and diseased tissue types and found 43,500 genes to be expressed in only a single cell type and 1,000 genes in all cell types. As tissues they studied the gene expression of colon epithelium, breast epithelium, lung epithelium, melanocytes, prostate, kidney epithelium, cardiomyocytes as well as the brain. Many of these genes had not been fully annotated at this time [401].

1.4 Chromosomal clustering and gene organization

Chromosomal clustering of functionally related genes and a non random distribution of genes has been shown by many groups [167]. The common regulation of functionally related genes can be organized through common open chromatin domains, the same transcriptional regulator elements, or by methylation. The distribution of genes plays a major role, as well as the arrangement of common genes on the chromatin level. Also in higher dimensions after chromatin folding, condensation and decondensation play a role [343, 137].

According to Lercher et al. the human genome is a mosaic on many structural levels, which comprise cytogenetic bands, GC-rich areas, isochores, as well as gene clusters. For an optimal gene regulation, housekeeping genes should be concentrated on transcriptionally active chromosomal domains and associate these with GC- rich regions and Giemsa bands [221]. Lercher et al. stated that genes that are broadly expressed tend to cluster. Earlier

they had found that procaryotic genes were organized in groups, called operons, while eukaryotic genes first did not seem to show this effect [222].

Some people argue that the GC rich sequence is an actual driving force to select certain gene patterns, and local chromatin characteristics might have an affect on the accessibility by components of the transcriptional machinery [84, 427]. Lercher et al. state that housekeeping genes, in contrast to tissue-specific genes, are frequently associated with CpG islands [18]. In 2007 Meaburn brought up the idea of chromosome territories. They describe a looped chromatin structure in higher order eukaryotes. Chromosomal arrangement however can also be cell- and tissue-specific and can be rearranged during differentiation as well as development. One example is the outward movement of chromosome number 6 in the differentiation of T cells [255].

The higher chromatin folding has been shown using methods such as 3-C, 4-C, 5-C, Hi-C as well as FISH (fluorescent in-situ hybridization). If promotors and enhancers are assembled in common spatial areas, genes can be more easily transcribed [92, 12, 141, 391]. Also insulator proteins may play a role here [246]. Clustering of CpG islands containing promoters of housekeeping genes has been reported to be an important factor of the spatial organization of interphase chromosomes [149].

Chromosomal clustering of many gene groups has been shown in the past. Shoguche et al. found clustering of housekeeping genes in human and *Caenorhabditis elegans*. While studying 158 genes in 11 different tissues in *Ciona intestinalis*, they could not detect any chromosomal clustering. Looking for chromosomal clustering of tissue-specific genes in the same species they could not find any clustering [352]. Lercher et al. found clustering of housekeeping genes in the human genome [222]. Roy et al. found clustering of muscle specific genes in the nematode worm *Caenorhabditis elegans* [328]. Miller et al. could show clustering of spermatogenesis genes and oogenesis genes [258] in the same species. Pauli et al. found clustering of intestine-specific genes also in *Caenorhabditis elegans* [292]. Blanco et al. detected clustering of testis-specific genes in the fruit fly *Dro-sophila melanogaster*, and Butanaev et al. showed that many tissue-specific genes in *Arabidopsis thaliana* are clustered as well, e.g. root genes, genes related to seedlings, ovules, sliques, flowers, seeds, and genes related to biotic stress [44]. Cohen et al. showed chromosome correlation maps of functionally related genes in cell cycle, sporulation and pheromone response in yeast [78]. This means that chromosomal clustering of functionally related genes seems to be a general organization principle in many species [50, 435, 368, 35, 32].

1.5 Aim of this thesis

The aim of this thesis is to define and identify tissue-restricted antigens from gene expression microarray data as well as from next generation sequencing datasets. A tissue-restricted antigen (TRA) is a gene which is highly expressed in one or only a few tissues in the body compared to its gene expression in all other tissues in the body. We will provide here an operational definition of tissue-restricted antigens (TRAs). TRAs are thus outliers with respect to tissue-specific gene expression for which we will search and find cutoffs as well as rules in order to define the significance of an outlier. Since the main focus of this work will be how to solve the question of tissue-restricted antigens in the context of promiscuous gene expression in the thymus in order to eliminate autoreactive T cells from the T cell repertoire, we were led by the idea to first search for all known antigens being involved in autoimmune diseases such as insulin-1 and 2 [102].

In this work we will define a criteria for tissue-restricted antigens (TRAs) and calculate them in all given datasets. These TRAs will be compared to already known tissue-restricted antigens (TRAs), especially those, which are involved in autoimmune diseases as well as in the negative selection of T cells in the thymus. After this we will compare all found TRAs in all different datasets and calculate their differences as well as overlaps in order to get a good common overview over the given criteria [102].

In a next step, we will annotate all TRAs with the given identifiers, identify the tissue, for which they are specific, and plot their gene expression profile over all tissue types available in each dataset. All of this data will then be stored in a user-based interactive database, which can be queried for over the internet for different criteria. This database will be available to the community and called TRA-DB. The aim is to provide a comprehensive database of all tissue-restricted antigens according to different datasets and species based on microarray data as well as RNAseq data [102].

In this database it will be possible to limit the criteria to one tissue only, giving the user the oportunity to clarify the question of research of interest. In the case of the RNAseq dataset of the human GTEX data, the intrahuman variability will be shown in boxplots, since this dataset provides the necessary data for this question in contrast to the microarray datasets which will be used in this work, which only have a sample size of n=2 datasets per tissue type. In the case of mouse, data will not show this vast difference in gene expression due to pooling of different mice and using inbread mouse strains, in contrast to the variability within the human population, as represented by donors in the GTEX dataset.

After finding and defining tissue-restricted antigens (TRAs) we will follow the question of chromosomal clustering of tissue-restricted antigens (TRAs) compared to randomly picked genes of the same length. For this cluster analysis, two different methods of chromosomal clustering will be applied to the previously calculated data. This will be done for all datasets and species, and an interspecies comparison between TRA clusters will be done. One of these methods will be the sliding 10-gene window method developed by Roy et al. [328], the other will be the sliding gene window method of fixed size developed by Gotter et al. [144]. Both methods have proven to be effective to study chromosomal clustering of the genes of interest. One will take regional differences in terms of gene density into account, the other method shows how many genes of interest are in direct neighborhood also in terms of base pairs with a fixed distance. The combination of both will ensure that chromosomal clustering of tissue-restricted antigens is robust with respect to the method used [102].

In case of chromosomal clustering of tissue-restricted antigens (TRAs) we will go into further questions such as showing that not only TRAs, but also housekeeping genes, as well as other functionally related genes might be clustered on a chromosomal level. For this we will use the GO annotation of genes in order to find functionally related gene groups, such as cell cycle genes, cytoskeleton genes, genes of the glycolysis as well as other common gene groups of interest [102].

In case the chromosomal clustering of tissue-restricted antigens (TRAs) can be shown in the different datasets, we will further search for a common evolutionary driving force, with a possible common regulatory mechanism in gene expression for example in medullary thymic epithelial cells (mTECs) and prove if chromosomal clustering of TRAs might be only due to gene duplication as was previously argued, or is the result of a common regulatory mechanism of gene expression in mTECs, for example triggered by transciption factors such as the autoimmune regulator (AIRE). For this, our TRA list will be compared to AIRE knock-out versus wild-type mice, for which we have microarray gene expression data.

Defining a whole set of tissue-restricted antigens, by studying the gene expression of TRAs in medullary thymic epithelial cells as well as looking for its regulation by the autoimmune regulator AIRE will give us a deep inside into the molecular machinery of the development of autoimmune diseases and into promiscuous gene expression and might help us to illucidate fundamental features in order to solve, understand and maybe cure autoimmune diseases in the future [102].

The idea of this work has been based on my diploma thesis with the ti-

tle *"A database of genes that are expressed in a tissue-restricted manner to analyse promiscuous gene expression in medullary thymic epithelial cells"*, they are as such indicated in the text of the work and have been significantly enlarged and changed during this PhD thesis [102].

2 Methods

Some parts of the methods have been adapted from my previous work in my diploma thesis with the title *"A database of genes that are expressed in a tissue-restricted manner to analyse promiscuous gene expression in medullary thymic epithelial cells"* [102]. They have been substantially extended and updated, for example by including sequencing-based expression data, new datasets, updated annotations and the whole work on chromosomal clustering has been added. TRA detection has been done, interspecies comparison has been added.

2.1 Datasets and pre-processing

In order to study chromosomal clustering of tissue-restricted antigens (TRAs) we analyzed four different microarray datasets, two in mouse, two in human and one RNAseq dataset in human [371, 372, 327, 218, 3, 2]. The rat dataset was excluded from the study, because it only contained one tissue type and was thus not suited for our study (Table 2.1.4). The calculations were mostly done with the statistical open source programming language R, using bioconductor packages, ensembl biomart annotation files, annotation packages from brainarray as well as perl and shell scripts (see technical appendix, part A on CD) [413, 130, 88, 337, 102, 164].

2.1.1 The human and mouse Novartis datasets

The Novartis microarray datasets by Su et al. [371, 372] contain two datasets one in human and one in the mouse. It is gene expression data of 61 murine and 79 human tissue-types. The rat dataset only represented tissue of the central nervous system (CNS) and was therefore excluded from the study. In our study we used the microarray raw data with gene expression data of 44,775 human and 36,182 mouse genes [371, 372, 102].

2.1.2 The human Roth dataset

The human Roth dataset from 2006 [327] contains gene expression data of 65 different tissue-types in human. On the chips there are 20,774 genes represented which have been measured on 353 chips. The sample size per tissue-type varies between two an nine measurements per tissue-type.

2.1.3 The mouse Lattin dataset

The mouse Lattin dataset from 2008 [218] shows gene expression data in mice for 91 different tissue-types. The C57Bl/6 mice have been 3-10 weeks of age, the RNA was pooled from these mice. There have been double measurements per tissue-type for 17,079 genes in total.

dataset	number of tissues studied	data type	species	source
mouse Novartis	61	Affymetrix, gngnf1mus custom array	mouse	GSE1133 [371, 372]
human Novartis	79	Affymetrix, hg-u144a microarray	human	GSE1133 [371, 372]
rat Novartis	12	Affymetrix	rat	GSE1133 [371, 372]
human Roth	65	Affymetrix, u133 plus 2.0 microarray	human	GSE3526 [327]
mouse Lattin	91	Affymetrix, mouse 4302 micrarray	mouse	GSE10246 [218]
GTEX	54	human	RNASeq	GTEX [3, 2]

Table 2.1: Overview of datasets used for TRA detection

2.1.4 The human GTEX RNAseq dataset

The human RNAseq dataset from the GTEX Consortium covers the measurement of 190 different donors. It contains 1,814 samples with 47 different tissue-types in the human GTEX dataset from 2013 and 1,641 samples from 175 different patients with 43 different tissue-types in the dataset from 2015 [3, 2]. Combining both datasets, we have the measurements of 54 different human tissue-types (Table 2.1.4).

2.2 Bioinformatical tools and programs

The microarray data [371, 372, 327, 218] was analyzed on the basis of the transcript level by reading in the CEL files as raw data (Table 2.1.4). The data was downloaded from the GEO database and read in using the brainarray package (version 18.0.0). The annotation was done with ensembl biomart with the actual version (see technical appendix on CD).

2.2.1 Gene expression omnibus (GEO)

The Gene Expression Omnibus (GEO) is an open repository for microarray data, provided by the National Center for Biotechnology Information (NCBI) of the National Library of Medicine in the National Institute of Health in Bethesda, MD, USA. It held already in 2012 more than 20,000 published datasets [109, 25].

package	version	title
gngnf1musamm enstcdf	Version 18.0.0	Annotation for the mouse Novartis dataset [371, 372]
hgu133ahs enstcdf	Version 18.0.0	Annotation for the human Novartis dataset [371, 372]
mouse4302mm enstcdf	Version 18.0.0	Annotation for the mouse Lattin dataset [218]
hgu133plus2h enstcdf	Version 18.0.0	Annotation for the human Roth dataset [327]

Table 2.2: Brainarray packages used in this study

2.2.2 Brainarray packages

Brainarray is a microarray re-annotation project from the Micro Array Lab at the University of Michigan. It offers custom cdf files, with which the import and annotation of published microarrays yields more recently annotated data for virtually all available microarrays. Dai et al. as well as Sandberg et al. showed that gene annotation and calculation with the custom cdf files from Brainarray gave better results than previously offered annotations by the microarray manufacturers themselves [413]. They are regularly updated to the most recent standard of gene databases and annotations [88, 337]. In our study we used the Brainarray version 18.0.0 on the basis of Ensembl transcript levels.

2.2.3 The programming language R and Bioconductor

For the analysis of microarray data and statistical computing the open source statistical programming language R was used. Together with Bioconductor packages it suits well for statistical testing, graphical display, genomic annotation and computing. R can be downloaded for different platforms from the CRAN network and is realeased in regular updates. For this work we used R version 3.1.1. and the Bioconductor version 3.4. (technical appendix, part A: programming code).

2.2.4 Annotation of microarray data

For the annotation of the microarray data, we used Ensembl Biomart, Version 87 of the patch GRCh38.p7 and GRCm38.p5.

2.2.5 Perl, mySQL and PHP

Perl is a scripting language with many biological applications. In this work it was used for the analysis of chromosomal clustering of TRAs. For the

package	platform	version	title	author
affy	Biocond.	1.52.0	Methods for Affymetrix oligonucleotide arrays	Gautier et al. [127]
affyio	Biocond.	1.46.0	Tools for parsing Affymetrix data files	Bolstad [36]
AnnotationDbi	Biocond.	1.38.2	annotation database interface	Pages et al. [282]
Biobase	Biocond.	2.36.2	base functions for Bioconductor	Huber et al. [164]
Cairo	R	1.5-9	graphical device	Urbanek 2015 [395]
geneplotter	Biocond.	1.54.0	graphics related functions for Bioconductor	Gentleman et al. [308]
GO.db	Biocond.	3.4.1	a set of annotation maps describing the entire gene ontology	Carlson [239]
limma	Biocond.	3.32.6	linear models for microarray data	Ritchie et al. [317]
VennDiagram	R	1.6.17	a set of functions to generate high-resolution venn and euler plots	Chen et al. [64]
vsn	Biocond.	3.42.3	variance stabilization and calibration for microarray data	Huber et al. [165]

Table 2.3: R and Bioconductor packages used in this study

establishment of the database of TRAs we used MySQL as well as PHP. MySQL is a database management tool and PHP is a scripting language, which has been used here in order to construct a webinterface for the TRA database.

2.3 Technical background: microarrays and RNAseq data

As gene expression data we used Affymtrix microarrays, Illumina microarrays as well as pre-processed RNAseq data by the GTEX consortium. The Affymetrix arrays were read in as raw .CEL files and annotated with predefined Brainarray packages. As identifiers we used the ensembl transcript IDs on the basis of the Brainarray packages. The illumina microarrays were pre-analyzed with bead studio and then imported into R. The annotation was done on the basis of the DNA sequence by the nuID, developed by Pan Du et al. [105]. The matching of probes from one system to the other was done on the transcript level by the ensembl transcipt IDs. For normalization of Affymetrix microarrays we used vsnrma, variance stabilization normalization, developed by Huber et al. [165] and for the Illumina microarrays quantile normalization, since the variation otherwise was too high and made comparisons difficult. The GTEX RNAseq data was pre-annotated by the GTEX consortium and used as ready to use RPKM values [3, 2].

2.3.1 Microarray chip technology: Affymetrix chips

The microarray chip technology by Affymetrix, Santa Clara, CA, USA synthesizes oligonucleotides up to a length of 25 base pairs directly on a small glass slide, hybridizing to analyse by complementary base pairing of the applied cRNA. The cRNA is labelled with flourescent dye. Here it is assumed that the intensity of the fluorescent dye and the expression level per gene are proportional to each other. The intensity level can be read in from a picture file, which is then further analyzed for its gene expression level per gene or transcript. For each gene or transcript, there is a set of probe pairs, consisting of 11 up to 20 probes spread out over the whole microarray. Each prope pair is consisting of a perfect match (PM) and a mismatch (MM), which makes the correction of unspecific binding possible. The distribution of probe pairs on the chip helps to account for regional differences of dye distribution and a more stable statistic of gene expression levels on the chip.

Microarrays are read in as .CEL files and annotated with Brainarray packages [413]. For the normalization of Affymetrix chips we used the variance stabilization normalization (vsnrma) developed by Huber et al. [165]. The vsn normalization integrates background substraction and normalization in a non-linear model. As a summarization method we used rma in the vsnrma method. Both vsn normalization as well as rma summarization are

integrated into one step. The purpose of the summarization is to combine the multiple probe intensities for each probeset to produce an expression value. The rma method provides a background correction and the vsn the normalization, a summarization based on a multi-array model fit by using the median polish algorithm [171].

2.4 Quality control (QC) of microarray data

After reading in .CEL files into R, microarrays have to be examined for quality control with a single-chip analysis. This way the four major quality problems can be eliminated [130]. These are low quality chips, artifacts such as fingerprints, imprints of pipette tips on the chip, extremes in terms of light intensities as well as local irregularities in the dye distribution (Fig. 2.1, 2.3 and Fig. 2.4).

Depending on the experimental setup, low quality chips should be excluded from the study, in case they can be substituted by enough biological replicates, which is not always the case. For instance in the human as well as in the mouse Novartis dataset, there are only double measurements per tissue, so that sometimes due to calculation reasons we had to accept some variability in the quality of the chips.

2.4.1 QC in the mouse Novartis dataset

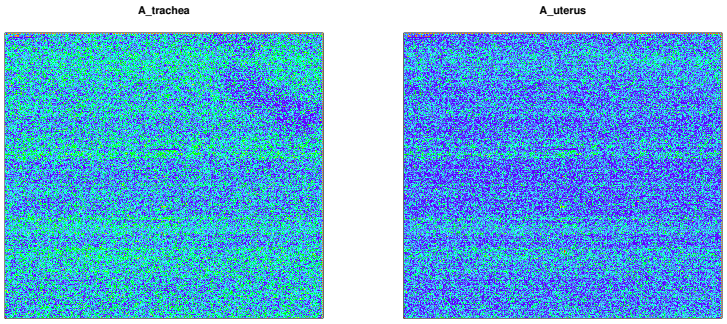
Quality control in the mouse Novartis dataset gave the result of two chips with quality problems. In the case of the A trachea there was a fingerprint on the chip (Fig. 2.1 (a)) and in the case of A uterus there were light stripes on the chip (Fig. 2.1 (b)). In both cases the chips were kept in the study, due to the lack of more replicates. Also in the case of Affymetrix microarrays small irregularieties can be equaled out by the distribution of twelve measurements per gene or transcript on the chip.

In the case of the cerebral cortex there was an irritation with the naming of the chips, but looking at the scatterplot of the gene expression data of one chip versus the other it became evident, that this was only a spelling mistake in the data and both chips refered to the same tissue-type (Fig. 2.2).

So all in all we could keep all n=122 microarrays refering to 61 different tissues in the mouse in the study concerning the mouse Novartis dataset (Table 2.4).

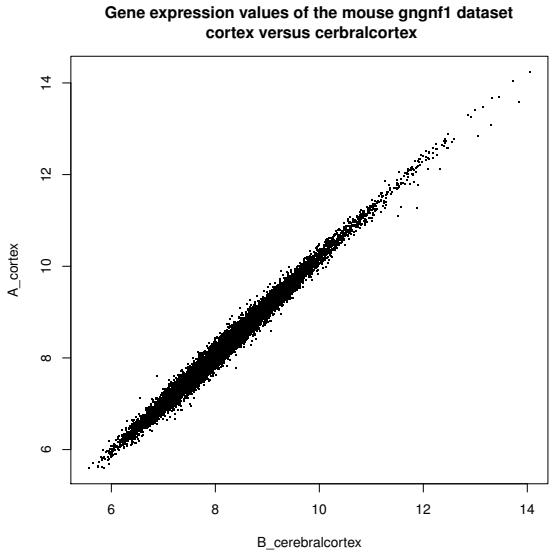
2.4.2 QC in the human Novartis dataset

In the human Novartis dataset there were eight chips with quality problems. One chip for the hypothalamus 2 had light stripes (Fig. 2.3 (a)), one for



(a) fingerprint (b) light stripes

Figure 2.1: **Quality control: single chip inspection of the Novartis mouse data, trachea A (a), uterus A (b)** In the visual inspection of the CEL files regional artifacts and low quality microarrays can be seen, in this step each chip was analyzed seperately and low quality chips were sorted out. In Fig. (a) of the mouse Novartis dataset for the measurement of trachea A a fingerprint can be seen in the upper right corner in Fig. (b) of the Novartis dataset of the uterus A light stripes can be seen as irregularities. If these chips are sorted out or kept in the study depends on the quality problem they have, as well as further quality control and number of biological replicates to choose from, little quality problems can be overcome in the case of Affymetrix microarrays, since each gene is measured in several different spots dispersed over the whole chip.



(a) cerebral cortex

Figure 2.2: **Scatterplot of the cerebral cortex versus the cortex in the mouse Novartis dataset:** Since two of the chip names in the mouse Novartis dataset did not have the exact same name, we plotted the gene expression of both chips versus each other in a scaterplot. According to the gene expression data, both chips seem to refe to the same tissue type, thus names can be changed to the same tissue type, furhtermore both tissues were the only single measurements in the study, so it seems that there has been a spelling mistake in the labels of the microarrays.

cardiac myocytes 2 was a very light intensive chip with dye irregularieties (Fig. 2.3 (b)), one for the testis leydig cell 1 had a pipette tip imprint on the chip (Fig. 2.3 (c)), one for the testis germ cells 2 had irregular dye as well as stripes (Fig. 2.3 (d)), one for the adrenal cortex had irregular dye and a small pipette tip imprint (Fig. 2.4 (a)), one for the trigeminal ganglion had low light intensity, some irregular dye (Fig. 2.4 (b)), one for the uterus corpus 1 had low light intensity with a light edge (Fig. 2.4 (c)) and one for the uterus corpus 2 had a very light intensity with a fingerprint on the edge (Fig. 2.4 (c)). All chips were further tested upon their quality in the further steps of quality control, for an overview please refer to Table 2.4.

2.4.3 QC in the rat Novartis dataset

In the wistar and sprague rat dataset there was one chip with quality problems. In the data of the nucleus accumbens of the sprague rat there has been a fingerprint (Fig. 1, technical appendix, part B: additional figures and Table 2.4). But since this dataset has been excluded from the study due to the lack of different tissue-types apart from the central nervous system the quality control does not fall further into account.

2.4.4 QC in the human Roth dataset

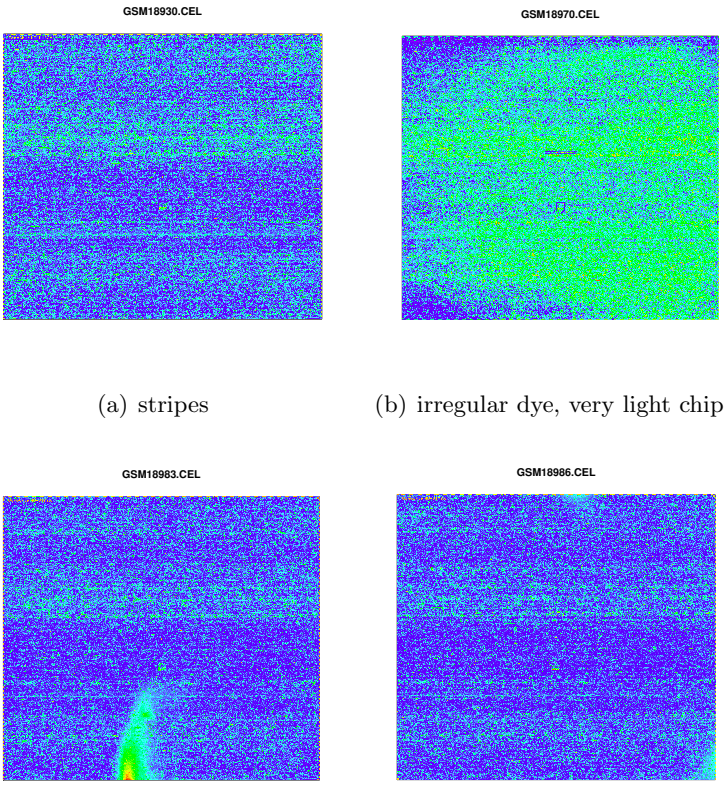
From the human Roth dataset thirteen of the 252 chips have quality problems, covering 65 different tissue-types in human. Since there have been between four to nine measurements per tissue-type seven low quality chips have been eliminated from the study. From all problematic chips of adipose tissue omental 1 (n=4), bronchus 4 (n=4), corpus callosum 7 (n=10), kidney cortex 2 (n=4), midbrain 9, 10 (n=10), ovary 8 (n=10), oral mucosa 1, 2 (n=4) and ventral tegmental area 4 (n=4), the following chips were sorted out: ovary 8 (n=10), midbrain 5, 9 and 10 (n=10), ventral tegmental area 4, 6 and 9 (n=10), (Fig. 2-3, technical appendix, part A: additional figures, Table 2.4).

2.4.5 QC in the mouse Lattin dataset

In the mouse Lattin dataset there have been quality problems in five tissues, these tissues were: adipose white B, amygdala A, iris B, spinal cord B, 3T3-L1 A and B, none of these chips had to be excluded from the study (Fig. 4, technical appendix part A: additional figures, Table 2.4).

2.4.6 QC of the GTEX RNAseq data

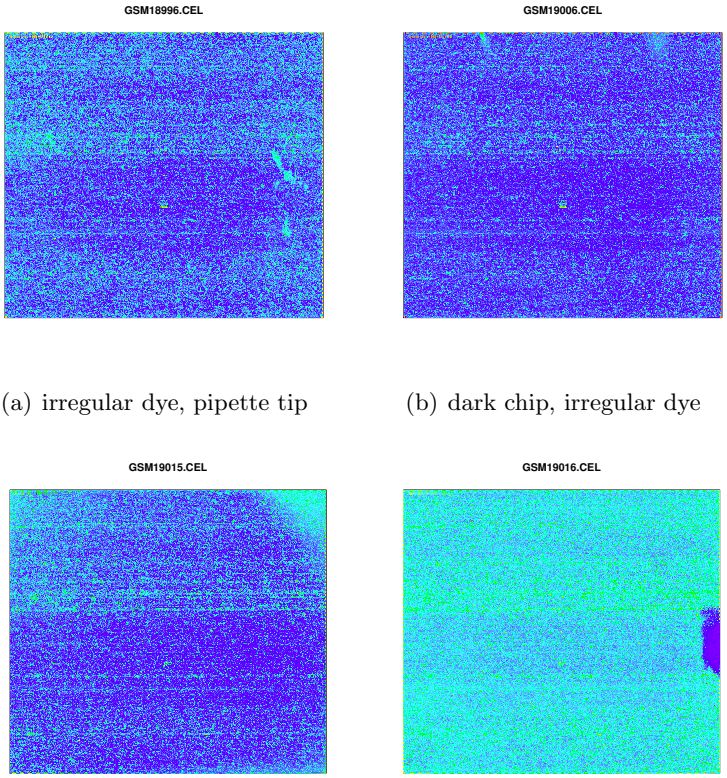
The quality control of the GTEX dataset was done by the GTEX consortium, since we used the ready to use RPKM values, there was no further quality control done.



(a) stripes (b) irregular dye, very light chip

(c) finger tip, pipette tip (d) irregular dye, stripes

Figure 2.3: **Quality control: single chip control in the human Novartis dataset, hypothalamus (a), cardiac myocytes (b), testis leydig cell (c), testis germ cells (d)** The same holds true for the quality control in the human Novartis dataset, which we have already seen in the mouse Novartis dataset. Here we can clearly see different types of quality problems in the chips. In Fig. (a) of the hypothalamus, we see stripes from moving the microarray during the application of the dye from right to the left, in Fig. (b) in the measurement of the cardiac myocytes we can see a big difference in the application of the dye, which will clearly lead to quality problems, in Fig. (c) of the testis leydig cell we can see a pipette or finger tip, which was probably accidently inserted during the application of the dye, in Fig. (d) testis germ cells there is some irregular dye and light stripes



(a) irregular dye, pipette tip

(b) dark chip, irregular dye

(c) dark chip with light edge

(d) light chip with dark fingerprint

Figure 2.4: **Quality control: single chip control in the human Novartis dataset, adrenal cortex (a), trigeminal ganglion (b), uterus corpus 1 (c), uterus corpus 2 (d).** Also in these chips of the human Novartis dataset we can see further quality problems. In Fig. (a) of the adrenal cortex, we can see irregular dye and a pipette tip, in Fig. (b) of the trigeminal ganglion the dye is very dark and irregular, in Fig. (c) of the uterus corpus 1, there is a light edge with irregularities and in Fig. (d) of the uterus corpus 2, the chip is so light, that it most probably has to be excluded from the study.

2.4.7 QC of Microarrays

Microarrays with obvious defects or chips with irregular dye distribution and a big difference in gene expression compared to other chips of the same tissue-type were sorted out (Table 2.4).

dataset	**tissue type**	**no. of chips**		**problem**
Novartis mouse	A trachea	n=2		fingerprint
	A uterus	n=2		light stripes
Novartis human	hypothalamus 2	n=2		stripes
	cardiac myocytes 2	n=2		irregular dye, light chip
	testis leydig cells 1	n=2		pipette tip
	testis germ cells 2	n=2		irregular dye, stripes
	adrenal cortex 2	n=2		irregular dye, pipette tip
	trigeminal ganglion 2	n=2		dark chip, irregular dye
	uterus corpus 1,2	n=2		dark chip with light edge, light chip with dark fingerprint
Novartis rat	nucleus accumbens core (sprague rat)			fingerprint
human Roth	adipose tissue omental 1	n=4		stripes, irregular dye
	bronchus 4	n=4		dark chip
	corpus callosum 7	n=10		stripes
	kidney cortex 2	n=4		stripes
	midbrain 5, 9, 10	n=10	excluded	dark chip, irregular dye
	oral mucosa 1,2	n=4		tip, stripes
	ovary 8	n=10	excluded	very light chip
	ventral tegmental area 4, 6, 9	n=4	excluded	tip

dataset	**tissue type**	**no. of chips**	**problem**
mouse Lattin	adipose white B		stripes with a scratch
	amygdala A		stripes
	iris B		small fingerprint
	spinal cord B		dark chip
	3T3-L1 A, B		stripes
GTEX	RPKM values		no quality control done

Table 2.4: Quality control of microarrays

2.4.8 The RNA degradation plot (QC)

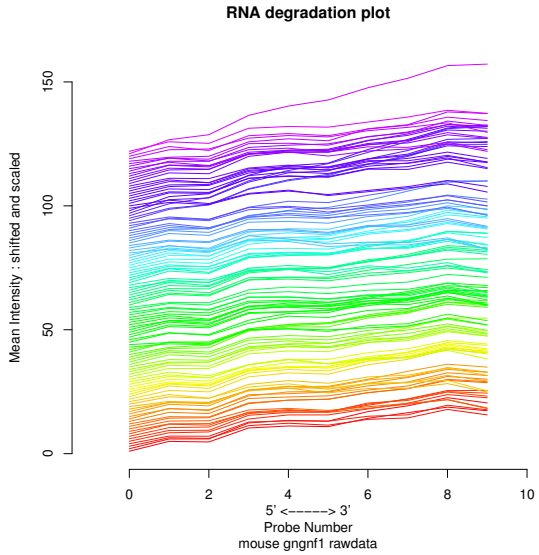
The second step of quality control (QC) is the analysis of RNA degradation [130]. The RNA degrades from the 5' end to the 3' end, strong degradation leads to irregularieties in these plots and crossing lines, which can be visualized in two pre-defined plots both shifted and scaled or shifted only. In the RNA degradation plot probesets are ordered by their location from the 5' to the 3' end. Over each point the expression values are averaged. RNA of one probeset should not differ the *"groups"* pattern and crossing lines should be considered to possibly represent bad RNA quality. The exclusion of single chips might be necessary. In the RNA degradation plot of the mouse Novartis data, we can see some irregularities in the shifted and scaled plot Fig. 2.5 (a), the RNA degradation plot scaled only does not show any effect Fig. 2.5 (b). Due to the lack of biological replicates we still left the chips in and went on for further quality control.

Similar effects can be seen in the other three datasets. Until here, no extra chips had to be excluded (Fig. 5-8, technical appendix, part B: additional figures).

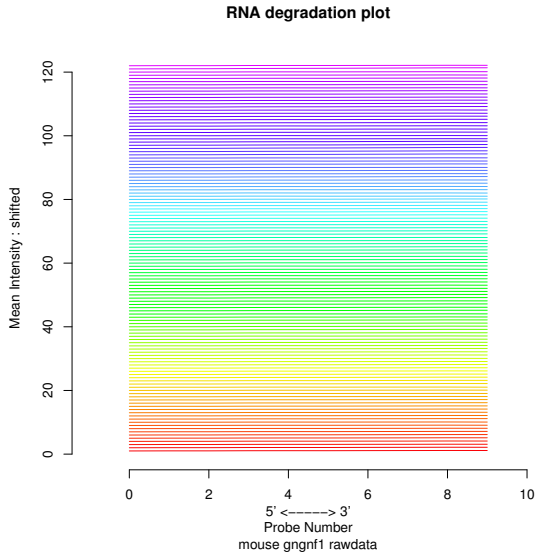
2.4.9 Variance stabilization (vsnrma) normalization

After quality control all microarrays had to be normalized and a background correction of all chips had to be done. We have used here the vsn normalization developed by Huber et al. [165] and a background correction and summarization rma, which had been existent earlier [171]. The combination of both methods seemed to be the perfect combination of both methods.

The vsnrma normalization is a per-probe normalization of transformed val-



(a) shifted and scaled



(b) scaled

Figure 2.5: **Quality control and RNA degradation:** In the RNA degradation plots, the degradation of the RNA can be seen, in case of crossing lines, as can be seen above in Fig. (a) shifted and scaled, there might be some chips, which have to be excluded, in the scaled version only there are no irregularieties (Fig. (b)). This figure has been adapted from Dinkelacker 2007 [102], the method adapted from Gentleman et al. 2005 [130].

ues (ashinh/glog) by robust linear regression. The variance stabilization normalization fits scaling factors per chip according to the distribution of their variance. This stands in contrast to other normalization methods, such as the quantile normalization, where the quantile values of the distribution of each microarray are adjusted. While the quantile normalization method sometimes has to be used for example in the case of Illumina microarrays, vsnrma normalization does not overfit the distribution and seems to be the better method in this case.

Through the vsnrma normalization, dye effects, local irregularieties as well as other artifacts can be smoothed out and gene expression values per chip can be compared afterwards. Here both distributions of microarrays before and after vsnrma normalization are shown (Fig.2.6, 2.7). The y-axis in the normalized dataset is refering to the log2 transformation of the fluorescence intensity values (gene expression).

The vsnrma normalization was sufficient to normalize all Affymetrix based datasets, so that afterwards the data could be used for further calculations. In Fig. 2.6 and Fig. 2.7 the black horizontal lines refer to the median expression value of all probes on the chips, the boxplots further show the 25% as well as 75% quantile ranges of gene expression of all genes on the microarray. The black dots after normalization refer to outlier values, which are due to the log2 transformation further scattered for higher intensity values, than for lower intensity values. For all other datasets, please refer to the technical appendix: part B additional figures, Fig. 9-17.

2.4.10 Mean versus standard deviation (meanSdPlot)

After vsnrma normalization the quality of the variance stabilization can be measured and shown in the meanSd plot. The meanSd plot shows the standard deviation (sd) versus the row means of each dataset. If the stabilization procedure has worked out well, there should not be a strong dependency of the standard deviation or variance on the mean. The red lines depict the running median estimator. If there is no variance-mean dependency, the line will be approximately horizontal [165]. For all available datasets the meanSd plot is quite stable, the little upward movement of the curve at the end remains in the range of our expectations (Fig. 2.8).

2.4.11 The density plot (QC)

The density plot gives an overview of the quality of vsnrma normalization. In the density plot the density function is plotted for the log intensity of gene expression on each chip. In the plot each colored line is referring to the density function of each microarray. While the curves are distributed

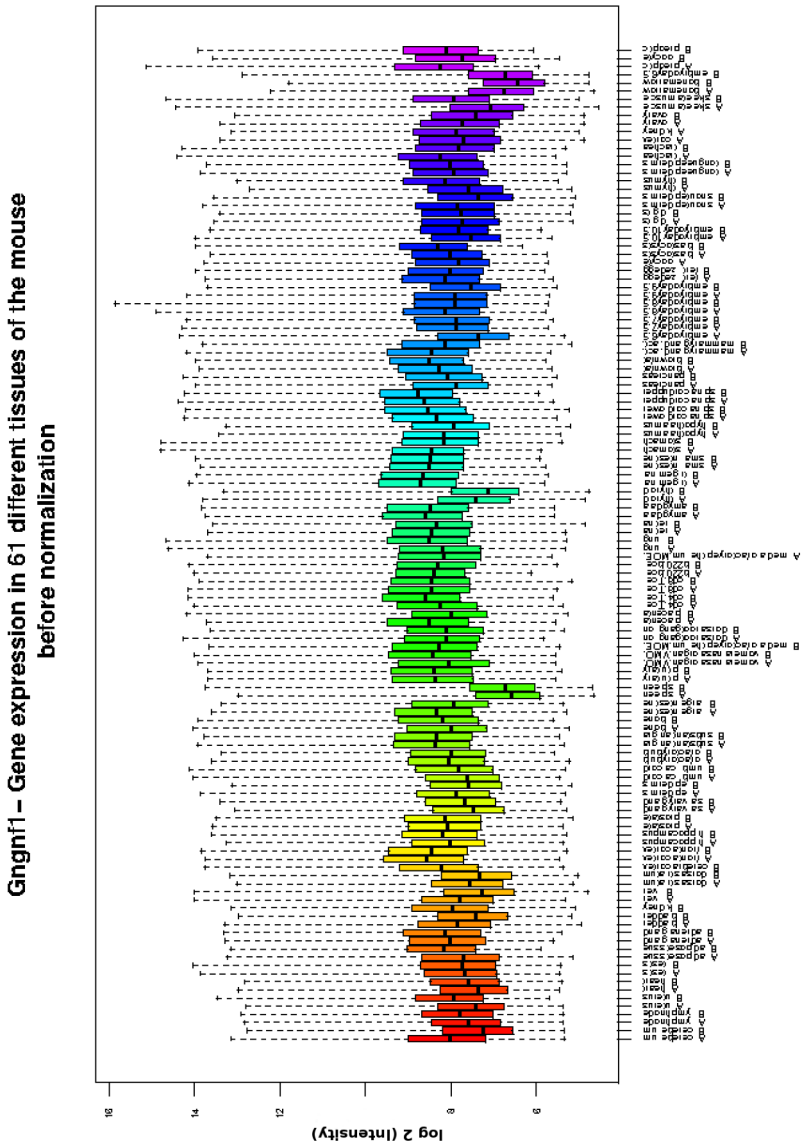


Figure 2.6: **Affymetrix microarrays of the mouse Novartis raw data:** In the microarrays of the mouse Novartis raw data, the intensity which reflects the gene expression per chip varies depending on the dye intensity applied on the chip. The boxplot per chip shows the median gene expression (black line), as well as the 25% and the 75% quantile range. This figure has been re-calculated from Dinkelacker 2007 [102], the method is taken from Huber et al. 2002, Irizarry et al. 2003 [165, 171].

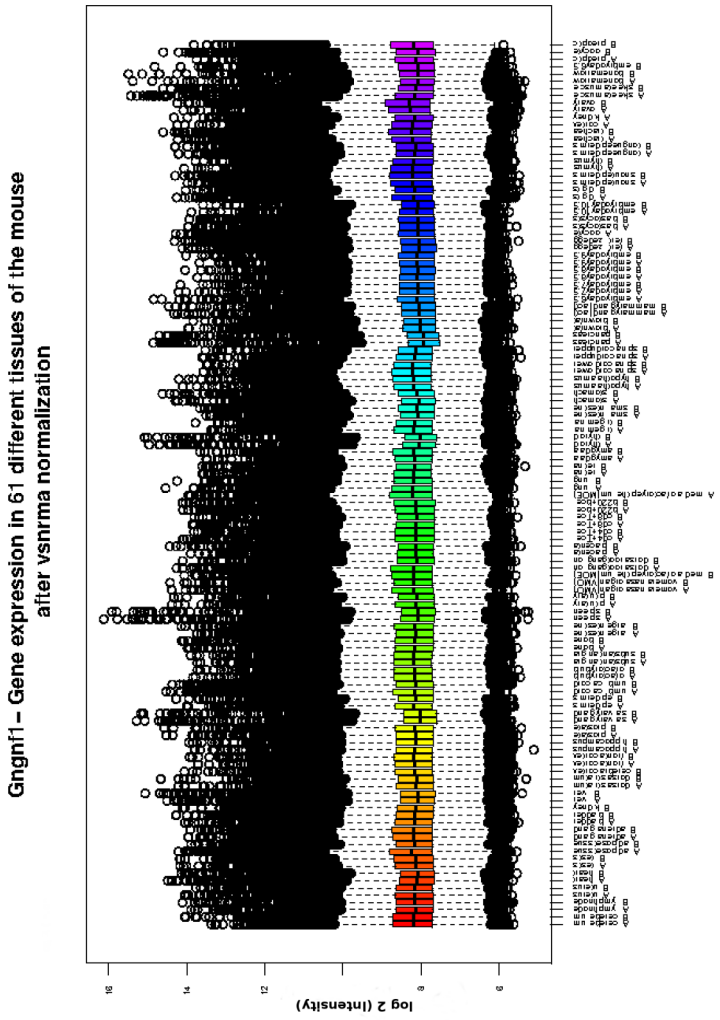
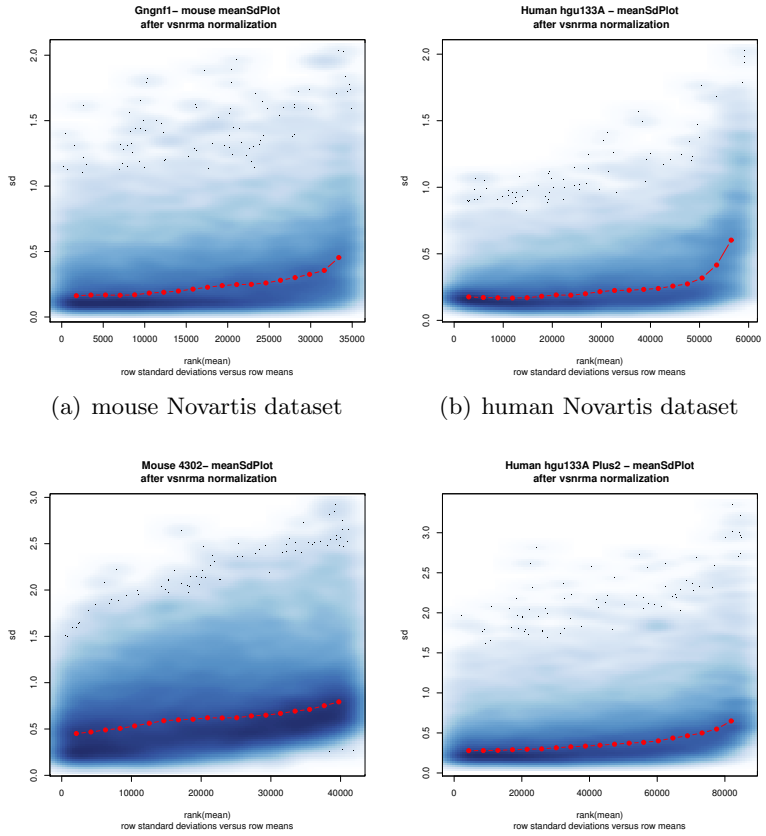


Figure 2.7: **Affymetrix microarrays of the mouse Novartis data after vsnrma normalization:** After vsnrma normalization the distribution of gene expression shown in boxplots is adjusted. through the vsnrma normalization outlier chips are equalized in the same range of all other microarrays. Although the 25% quantile ranges as well as the 75% quantiles are not exactly adjusted, gene expression is now among the chips comparable and the gene expression data is log2 transformed. The vsnrma normalization also includes background correction as well as summarization. The y-axis refers to the log2 transformed intensity values. The black dots refer to outlier values which are higher above than below, due to the log2 transformation. The figure has been re-calculated from Dinkelacker 2007 [102], the method has been developed by Huber et al. 2002, Irizarry et al. 2003 [165, 171].



(a) mouse Novartis dataset (b) human Novartis dataset

(c) mouse Lattin dataset (d) human Roth dataset

Figure 2.8: **meanSdPlot after vsn normalization of all four datasets:** The meanSd plot gives the standard deviation (sd) versus row means of each dataset. If there is no variance-mean dependence the median estimator (red line) should be approximately horizontal. Fig. (a) is showing the meanSd-Plot of the mouse Novartis datatset, (b) human Novartis dataset, (c) mouse Lattin dataset and (d) human Roth dataset

differently across the different chips before normalization (Fig. 2.9 (a)), they are well adjusted after vsnrma normalization (Fig. 2.9 (b)). For the other datasets, please refer to the technical appendix, part B: additional figures, Fig. 18-20.

2.4.12 The scatter plot (QC)

The scatter plot is a means of quality control between the different chips. They show the distribution of gene expression of two chips of the same tissue against each other. The distribution of dots after normalization should roughly follow a straight line (Fig. 2.10 A). If single chips deviate from this distribution, e.g. resulting in a banana-like shaped form (Fig. 2.10 B), the scatter plot is a good help to decide which chip should be excluded (Fig. 2.10 A-D). In this example the two measurements of the adrenal cortex in the human Novartis dataset fit well (Fig. 2.10 A), the two of cardiac myocytes show a banana-like form, one has to be excluded (Fig. 2.10 B), the two of testis germ cells scatter quite a bit, so one could be excluded (Fig. 2.10 C) and the two of the testis leydig cells fit very well (Fig. 2.10 D).

2.5 Averaging over multiple measurements per tissue

In the case of double measurements per tissue-type, we calculated the mean over double measurements as an expression value for each gene (Fig. 2.11). If there were more than n=4 values per tissue, as for example in the human RNAseq dataset, we also calculated the range of gene expression of each gene. Especially in human the intrahuman variability is very high in terms of gene expression per tissue-type (Fig.2.12). In this case we plotted the mean as well as the interquartile ranges of the gene expression in TRA plots (TRA-DB) [102].

2.6 Tissue grouping for TRA detection

From all datasets we excluded all embryonic tissues, cancer cell lines, as well as moving cells, such as the main immune cell types from TRA calculation. Functionally related tissues were grouped together according to their main tissue-type (Table 2.5) [102].

datasets	**Novartis mouse**	**Novartis human**	**mouse Lattin**	**human Roth**
tissue no. per dataset	61	79	91	65

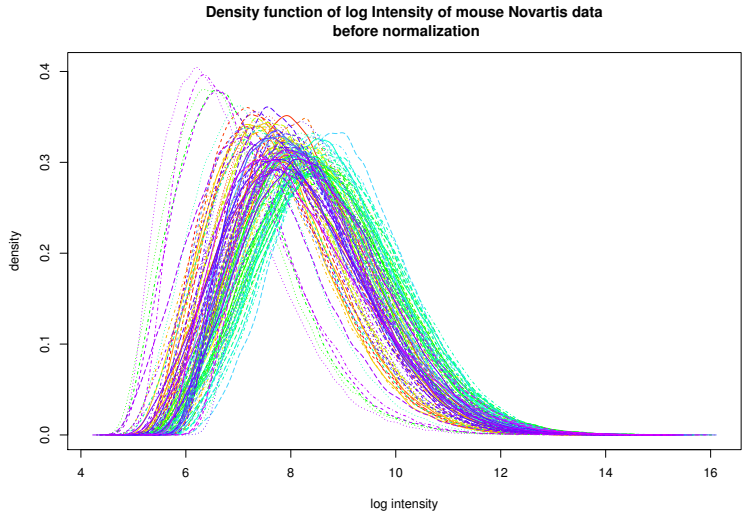
datasets	**Novartis mouse**	**Novartis human**	**mouse Lattin**	**human Roth**
excluded tissues per dataset	embryos(5), immune cells(3)	carcinoma (1), leukemia(3), lymphoma(2), cell lines(2), embryos(4), immune cells(11)	cell lines(9), embryos(2), immune cells(28), osteoblasts(3), osteoclasts(1)	-
tissue no. remaining per dataset	53	56	49	65
tissue no. per group per dataset	CNS(13), epidermis(4), intestine(2), ovary(2), PNS(4)	adrenal gland(2), CNS(19), epidermis(2), lymphoide structure(3), muscle(2), pancreas(2), PNS(4), testis(5), uterus(2)	adipose tissue(2), CNS(12), eyes(8), intestine(2), mammary gland(2)	adipose tissue(3), CNS(22), epidermis(2), heart(3), kidney(2), lymphoide structure(2), mammary gland(2), PNS(2), stomach(3), uterus(6)
tissue no. remaining per dataset	33	24	28	28

Table 2.5: Tissue grouping in all different datasets

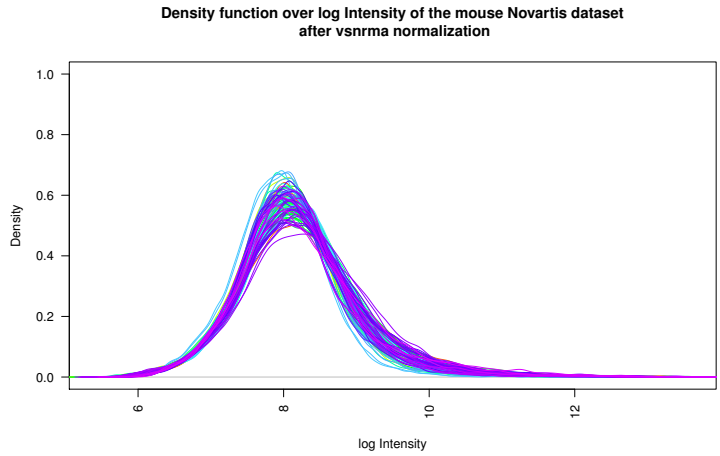
2.7 The definition of tissue-restricted antigens (TRAs)

In order to find tissue-restricted antigens we plotted the distribution of expression values of each gene/transcript over the studied tissue-types (Fig. 2.13). After calculating the median gene expression level of this gene or transcript over all tissues (horizontal black line), we defined tissue-specificity by giving a cutoff value of 5 times the median gene expression over all tissues (horizontal red line). A tissue-restricted antigen (TRA) is then a gene/transcript, which has higher expression in 5 times the median gene expression in at least one and not more than five tissues (Fig. 2.13). This definition has been adapted from Dinkelacker 2007 [102].

This definition is effective for two reasons. First, calculating 5 times the

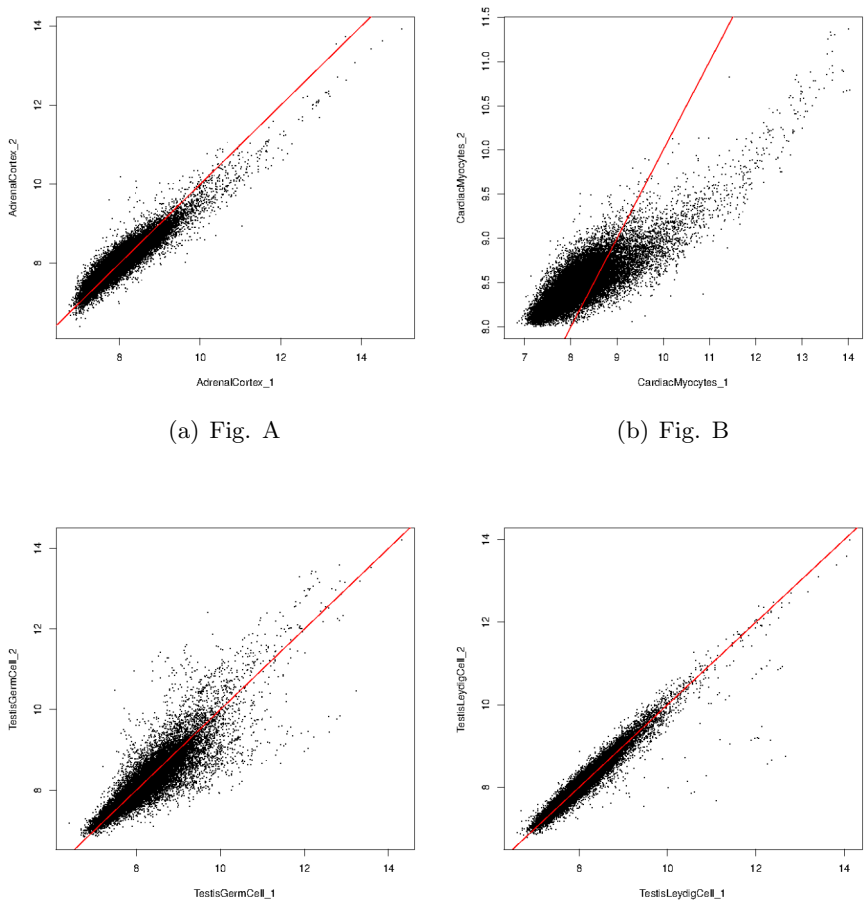


(a) Fig. A



(b) Fig. B

Figure 2.9: **Density plot of the mouse Novartis data before and after vsnrma normalization:** The density plot represents the density function of each microarray over their log intensities. In Fig. (a) many curves are widely dispersed over the log intensity, curves overlay in Fig. (b) and lie in close proximity. This figure has been recalculated according to Dinkelacker 2007 [102], the method taken from Huber et al. 2002 [165].



(a) Fig. A

(b) Fig. B

(c) Fig. C

(d) Fig. D

Figure 2.10: **Scatterplot:** These scatterplots show the distribution of one microarray versus the other being measured in the same tissue. The dots should be scattered around the red line if the samples show similar expression patterns for each gene expression within the samples, in Fig. A we can see the adrenal cortex 2 versus the adrenal cortex 1 in one dataset, it shows a small slope, but the quality is still good enough to keep the data. In Fig. B we can see a big difference in the gene expression pattern of cardiac myocytes 2 versus cardiac myocytes 1, this clearly shows, that there is a big difference in gene expression between the two tissues, which might give a hint to quality problems in one or the other probe. In Fig. D we can see a clear line scattered very narrowly around the red line which shows a very close proximity of expression patterns and a very good quality. Scatterplots are a good way of quality control from one probe versus the other.

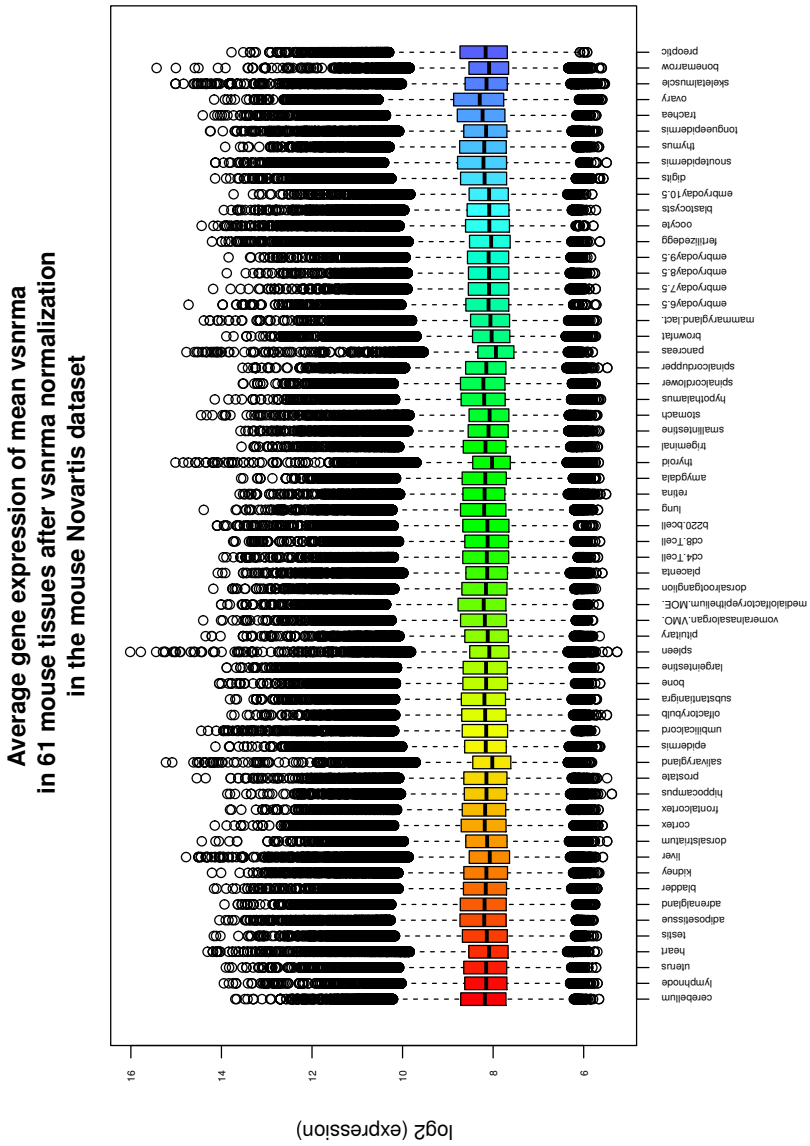


Figure 2.11: **Averaging over multiple measurements per tissue type in the mouse Novartis dataset:** Distribution of log2 (intensity) values, gene expression values per tissue type averaged over the double measurements in the microarray datasets, here the mouse Novartis dataset. This figure has been recalculated from Dinkelacker 2007 [102]. The method has been taken from Huber et al. 2002, Irizarry et al. 2003 and Dinkelacker 2007 [165, 171, 102].

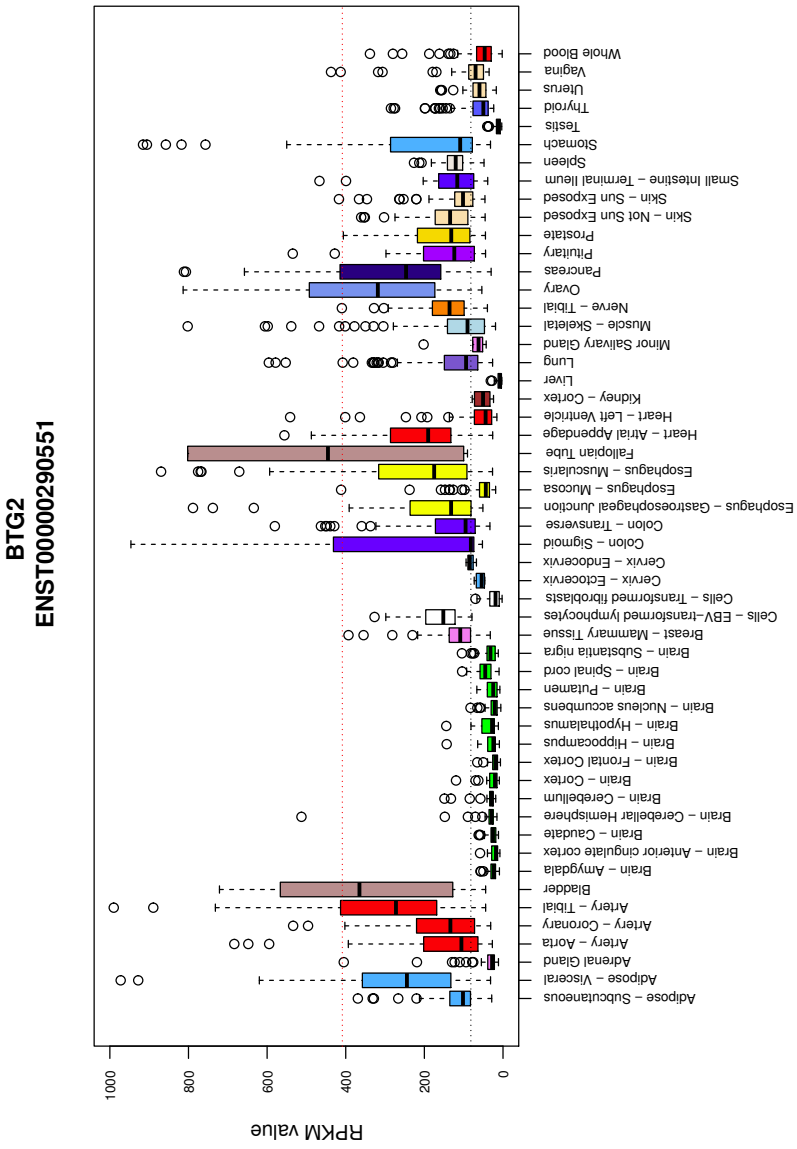


Figure 2.12: **Distribution function of multiple measurements per tissue type in the GTEX dataset:** Due to the high intrahuman variability in gene expression per tissue type, the distribution function per gene per tissue type was plotted in form of a boxplot.

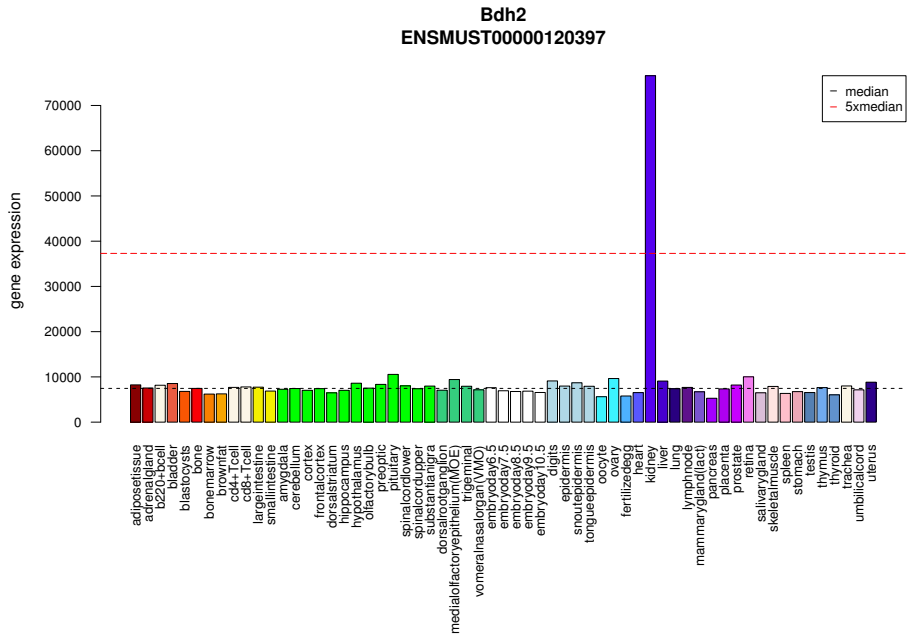


Figure 2.13: **Definition of tissue-restricted antigen (TRAs):** A tissue-restricted antigen is defined as a gene, which has higher expression in 5x the median gene expression of this gene over all tissues in at least one and not more than five tissues out of the 61 tissues in the mouse. For tissues belonging to one tissue entity we applied tissue grouping, such as for example for tissues of the central nervous system (CNS), light green bars, if in more than one tissue out of this group exceed 5x the median line, they are only counted as one tissue type. Embryonal tissues, as well as cell lines, including immune cells are plotted but not taken into account in the calculation. This method has been adapted from Dinkelacker 2007 [102].

median over all tissues is very easy and fast. Second, the median is a robust estimate in case of outliers. The cutoff of 5 times the median has proven to be a good estimate of tissue-specificity. For comparison reasons, we also tested other cutoff values, such as 3x, 10x, and 20x the median. According to already known tissue-restricted antigens (TRAs) from the literature, we had the best hit rate with 5 times the median line in the tested datasets. Of course this is an operational definition, giving a reasonable good balance between sensitivity and specificity. Pre-studies of this method have been adapted from Dinkelacker 2007 [102].

In order to not have any bias in our data towards certain tissue-types, such as the central nervous system (CNS) in the mouse Novartis dataset, we applied an additional tissue grouping in case of tissues, which belong biologically to the same organ. In the calculation of TRAs, we consider them to be one tissue only. The tissue groups can be seen in Fig. 3.1-3.4. Tissues of the same tissue group are plotted in similar colours (Table 2.5) [102].

From the tissues available, all embryonic tissues, as well as cancer cell lines and immune cells have been excluded from the calculation [102]. The embryonic tissue-types because the tissue groups in the embryo develop so fast, that the differentiation between different tissues is very difficult, the cancer cell lines, because they were not in the matter of our scientific question and the immune cells, because we have aimed to measure tissue-restricted antigens on the basis of tissue-types rather than cell lines. It remains clear however, that immune cells are important on one hand side, and included in the measurement of all secondary immunological relevant tissue-types, such as the thymus, the lymph nodes, and other secondary lymphoid organs. For this reason we did not consider them in order to calculate TRAs but still plotted them in our TRA plots, so that the gene expression if available could be still illustrated in the outcoming plots. In this case these cell lines or tissue-types have been depicted in white (Fig. 2.13).

In order to check the cutoff, such as here the 5x median line, we plotted a *"saturation curve"* for each dataset in terms of tissue number and cutoff criteria. It becomes obvious that if we set the criteria too stringent, we lose too many true positives, while if we set the criteria to loose, we have too many false negatives. The goal is however to hit the curve at the point of their biggest slope, before reaching the saturation of too many tissues and after being in the steep part (too many changes with little changes in the criteria) of the curve (Fig. 2.14) [102].

Applying our criteria in this example of the mouse Novartis dataset we find about 8,000 transcripts, referring to about 5,800 genes, to be tissue-restricted with 5x the median line in at least one and not more than 5 out

of 61 murine tissues (black vertical line).

2.8 The definition of housekeeping genes

In addition to the definition of tissue-restricted antigens (TRAs), we also calculated all housekeeping genes in a similar approach. A housekeeping gene is defined as a gene which is not higher expressed than 3x the median gene expression of this gene over all tissues in any of the tissues.

2.9 The database of tissue-restricted antigens (TRA-DB)

After defining and finding all tissue-restricted antigens (TRAs) as well as housekeeping genes for any of the given datasets and plotting the gene expression profiles with the given tissue colours, we established a table including gene annotation with different identifiers, start site as well as tissues in which they are tissue-restricted and imported all this data into a mySQL based database which can be queried from the internet via a PHP-based web interface. Users can export query results and demand both lists and plots as pdf output or lists as an Excel based csv file. This database has been re-calculated and updated from a previous version from Dinkelacker 2007 [102].

2.10 Gene annotation

TRAs were determined on the basis of transcipt level, in contrast to other databases of TRA genes. Microarrays are read into R using the Brainarray packages and parsed on the Ensembl transcript IDs and in this way the whole workflow is calculated. TRA tables and plots were than further annotated using biomart annotation files based on these Ensembl transcript identifiers. For the annotation we used gene Identifier, start side, transcript Id, chromosome, gene symbol as well as other Ids.

2.10.1 Annotation of microarrays with Biomart and Brainarray

For the annotation of microarrays we used custom CDF environments and probeset data provided by Brainarray, version 18.0.0. As an annotation package of R we used the R package, Annotation Dbi, version 1.24.0. For the ensembl Version with which we annotated the files, we always used the most up-to-date Ensembl version at that time. In the mouse Novartis data, we could annotate 17,121 genes (34,589 transcripts) on the chip, in the human Novartis data we could annotate 13,663 genes (49,028 transcripts) on the chip, in the human Roth data we could annotate 21,159 genes (77,834 transcripts) on the chip, and in the mouse Lattin data we could annotate 16,864 genes (29,590 transcripts) on the chip. The GTEX data was pre-annotated by the GTEX consortium. Gene numbers may vary depending

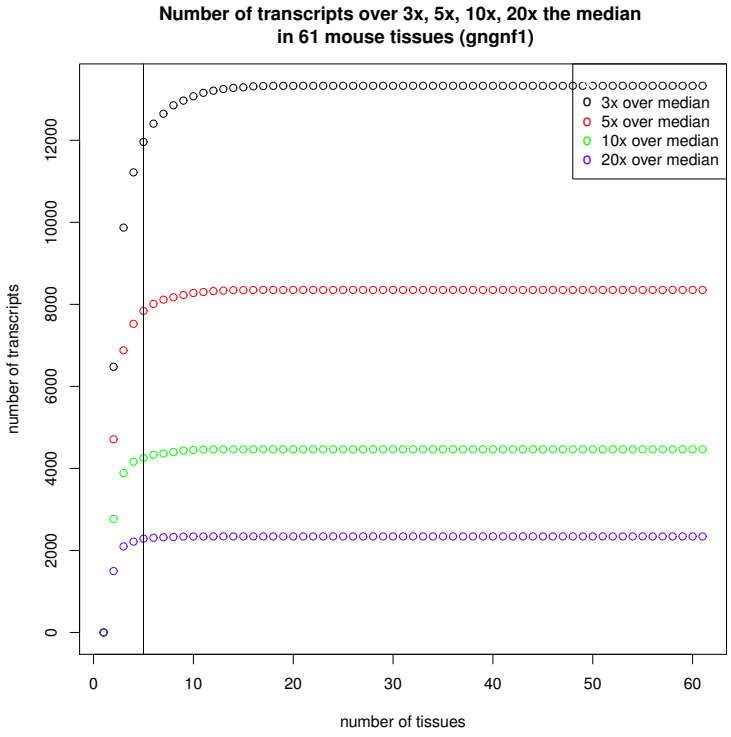


Figure 2.14: **Saturationplot:** The saturationplot shows the number of TRAs detected per cutoff criteria of 3x (black), 5x (red), 10x (green) and 20x (blue) the median gene expression per gene over all tissues over the number of tissues, which have to exceed the cutoff line. The black vertical line depicts the cutoff of up to five tissues as being regarded to be tissue-specific exceeding over the cutoff of 5x the median gene expression (red dots) as an interplay of not too many false positives and not too many false negative TRAs. This figure and method has been adapted from Dinkelacker 2007 [102].

on the annotation version. The exact versions can be found in the technical appendix: part A, programming code on CD [102].

2.11 Chromosomal clustering of tissue-restricted antigens

For the calculation of chromosomal clustering of TRAs we used two different approaches. The first one has been reported by Roy et al. [328] for testing the clustering of muscle-specific genes in the worm *C. elegans* with a sliding 10-gene window method over the chromosomes. The second has been first applied by Gotter et al. for the calculation of chromosomal clustering of a first set of tissue-restricted antigens (TRAs) [144]. Both methods are applied here for all TRA groups as well as for housekeeping genes and have also been tested on other gene groups of functionally related genes, such as TCA genes, cell cycle genes, cytoskeleton genes, glycolysis genes according to their GO annotation. While the 10-gene window method takes into account that genes are sometimes more densely distributed and in other times more dispersed on the chromosomes, the sliding gene window method by Gotter et al. rather counts duplicates on the basis of fixed window sized at different physical distances. This clustering method has been adapted and re-calculated as well as applied to new datasets according to Dinkelacker 2007 [102].

2.11.1 The 10-gene window method

The 10-gene window method, developed by Roy et al. [328] calculates the number of TRAs within a moving gene window of 10 adjacent genes of all genes being present in each dataset. As long as at least one gene in the moving 10-gene window is a TRA, the gene window is moved on and the number of TRAs are summed up into a total number of TRAs and further considered as a TRA gene cluster. Only if a gap of at least 10 adjacent genes includes no TRA, the window is closed and the calculation of TRAs is started again. This calculation is applied to all chromosomes, and the number and sizes of all TRA clusters are calculated and inserted into a list. The same is done for statistical significance with 1,000 randomly drawn lists (Fig. 2.15) [144, 102].

2.11.2 The sliding-gene window method of fixed size

The sliding-gene window method of fixed size developed by Gotter et al. [144] calculates the number of neighbors of TRAs within a sliding gene window of fixed size. For window sizes here we used windows ranging from 50kb, 100kb, 200kb, 500kb, 800kb, 2Mp and 5Mp. The sliding gene window of fixed size is moved over the chromsome and calculates the number of TRA neighbors, and calculates the number of TRA neighbors within this window. A window does not drop off until there is at least one TRA in the directly adjacent gene window. The number of gene pairs, triplets, quadruplets of TRAs are

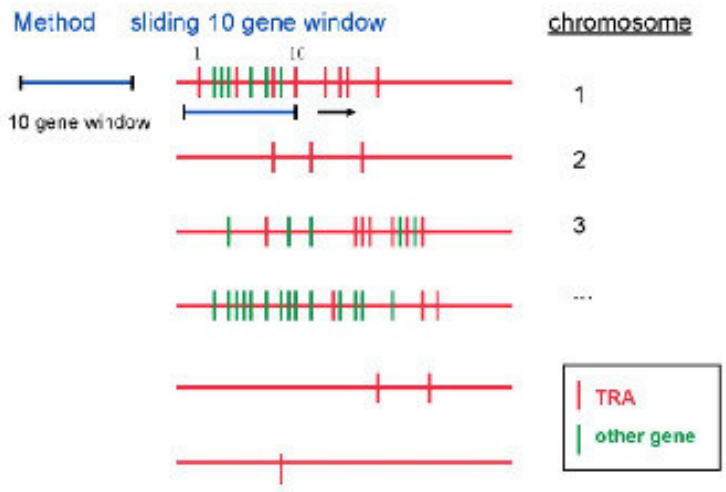


Figure 2.15: **Chromosomal clustering of TRAs, the 10-gene window method:** In the 10-gene window method chromosomal clustering of tissue-restricted antigens (TRAs) is calculated while adding up TRAs which appear within the range of a sliding 10-gene window, which drops of if it encounters a region of 10 adjacent genes including no TRA. TRAs in this figure are depicted by the red vertical lines and non TRAs with green vertical lines. This sliding 10-gene window is applied for each chromosome and compared to a list of 1,000 randomly drawn genes with the same length than our TRA list. This figure has been adapted from Dinkelacker 2007 [102], the method taken from Roy et al. 2002 [328].

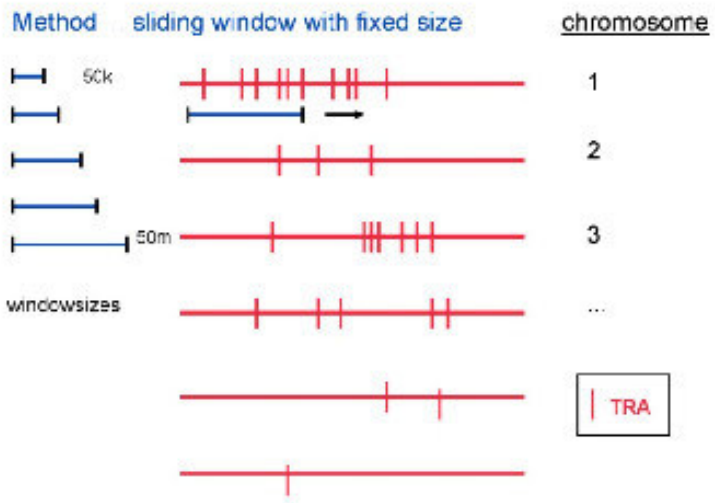


Figure 2.16: **Chromosomal clustering of TRAs, the sliding gene-window method of fixed size:** In the sliding gene-window method of fixed size chromosomal clustering of tissue-restricted antigens (TRAs) is calculated while counting the number of TRAs within a sliding gene-window of fixed size. As long as the gene-window does not encounter a part in which it does not find any TRA the number of duplets, triplets, quadruplets, etc. pp. are counted up. The same is done for a gene list of the same length as the TRA gene list, which are randomly drawn. This experiment is repeated 1,000 times. This figure has been adapted from Dinkelacker 2007 [102], the method taken from Gotter et al. 2004 [144].

counted and the same is done for 1,000 randomly drawn genes from the same gene list with the same length as the TRA gene list. In contrast to the 10-gene window this method does not account for gene density but captures physically close gene neighborhoods [102].

2.12 Interspecies comparison of human, mouse and rat

In this work we claim a biological meaning of TRA clustering for the purpose of the immune system. We expect hence that this gene clustering should be conserved among species that have an adaptive immune system. This includes mammals from the top of the tree down to jawed fishes, which were the first to develop an adaptive immune system [214]. We had data for three species within the mammalian clade. We compared human, mouse, and rat TRAs in terms of chromosomal location, gene order and TRA clustering. For this, TRA clusters from mouse were plotted and mapped through homology mappings to the other two species. The same was done for non-TRAs within the TRA clusters. As a mapping result, we took always the best

match for each TRA cluster with respect to the highest number of matching TRAs to the equivalent TRA cluster in the other species. For the result, please refer to the results section of interspecies comparison of this work.

The rat dataset only consisted of tissues belonging all to the neuronal system and therefor we considered a gene to be regarded as a TRA in the rat, if it is homologous to a TRA in the mouse, and as well a gene not to be a TRA in the rat, if it was a non TRA also in the mouse and vice versa. Since gene order and especially TRA gene order was the focus of this step, we considered this to be an adequate and allowed step.

2.13 Interspecies comparison of TRA clusters

In the interspecies comparison between mouse TRA clusters and other model organisms, such as kangaroo, platypus, frog, tetraodon, *Danio rerio*, sticklebacks, *Ciona intestinalis*, *Drosophila* as well as *Caenorhabditis elegans*, we defined TRAs in the other species, as being homologous genes to the mouse TRAs. For each TRA cluster, we searched for the TRA cluster or region of the best possible hit in the other species and compared the gene order of the mouse TRA with the one in the other species, while drawing TRAs in red and non-TRAs in grey lines. We further calculated the percent identity between clusters of both species as well as the total size of the TRA cluster and the number of TRAs and found homologous genes in the other species (Fig. 3.47 and 3.48).

2.14 Homology plot within TRA clusters

For the homology plot within TRA clusters, we calculated the % identity of all TRAs within the cluster and ordered them according to their median line of % identity. As can bee seen in Fig. 3.52, the outliers of gene families, with high % identity have been plotted as outliers to the boxplots. The same has been done to a number of random clusters with gene sizes of 5, 10, 20, 50 and 100 genes (Fig. 3.53). This method has been adapted from scripts from Prof. Dr. Benedikt Brors.

3 Results

Chromosomal clustering of tissue-restricted antigens (TRAs) can help to elucidate the background of central self tolerance, the expression of self antigens to potentially autoreactive T cells and explain the ectopic gene expression of otherwise tissue specifically expressed genes outside the thymus [102]. According to Hurst et al. the gene order in the human chromosome is non random and higher order gene organization can explain a common regulatory mechanism in gene expression [167].

We analyzed four different microarray datasets as well as one RNAseq dataset, both for human and mouse and determined tissue-restricted antigens (TRAs) in each dataset according to an operational definition. These TRAs were put into a database, where they can be queried upon different approaches, such as identifiers, gene names, symbols as well as tissues they are restricted for. These lists can then be extracted as Excel files as well as downloaded as plots.

We analyzed TRAs for chromosomal clustering according to gene density as well as gene neighbourhoods and tested chromosomal clustering also of housekeeping genes and other functionally related gene groups, such as cell cycle genes, cytoskeleton genes, glycolysis genes, TCA genes, caspase genes, muscle genes, actin cytoskeleton genes and apoptosis genes. We also analyzed the gene order conservation across the three species: human, mouse and rat. Furthermore we studied tissue specificity in TRA clusters and tested whether or not gene duplication may be the main cause for chromosomal clustering of tissue-restricted antigens.

Finally we determined the conservation of TRA clusters across more species further down the evolutionary tree and provided insight into the gene expression in medullary thymic epithelial cells (mTECs) as a basis of the negative selection of potentially autoreactive T cells in the thymus.

3.1 Data analysis in order to detect tissue-restricted antigens

In order to detect tissue-restricted antigens, tissues have been grouped according to their biological function and structure. Similar tissue types were put together so that there was no bias towards tissues of certain biological instances, such as the brain with many different regions in comparison to other tissues, which were presented only a few times (Table 2.5, Fig. 3.1 - 3.4) citeDinkelacker2007. Some tissue types have been excluded from TRA calculation, among them all cell lines, cancer cells as well as all embryonic tissues. Also immune cells have been taken out of the calculation. These

tissues or cells were however still plotted in the TRA plots [102]. Not all tissues are represented in each data set, but most datasets have examples of each tissue group (Fig. 3.1 - 3.4). The most represented tissues in most datasets are tissues of the central nervous system (CNS).

3.2 TRA definition

For the calculation of tissue-restricted antigens (TRAs) we first calculated the mean expression in all microarray datasets and the median expression in the RNAseq dataset of the gene expression per transcript or gene over double measurements. For the RNAseq dataset we used the median, rather than the mean due to the big variability in terms of number of replicates per tissue type as well as of intrahuman variability in gene expression (Fig. 3.5) [102].

For finding the right criteria in order to define tissue-restricted antigens (TRAs), we used an operational definition in the context of a 5x median gene expression cutoff line, after testing several different cutoffs (3x, 5x, 10x, 20x the median) and the number of tissues, which have to be over this cutoff line in at least one and not more than five tissues. This decision has been drawn from previous knowledge of already known TRAs, as well as the interplay of sensitivity and specificity of the saturation plot (Fig. 2.14) [102].

3.3 TRA numbers and percentages

In all datasets we could identify tissue-restricted antigens, varying between 4,172 in the mouse Novartis dataset from 2004 and 27,339 in the human GTEX dataset from 2013 and 2015. This accounts for about 24.6% of all genes to be tissue-restricted in the mouse Novartis dataset and 47.60% of all genes to be tissue-restricted in the human GTEX dataset. As the knowledge about gene and its gene expression has increased over the years, this increase in number can mainly be explained by this phenomenon (Table 3.1). For all datasets within the same species we calculated the overlap of tissue-restricted antigens across datasets (Fig. 3.6).

dataset	**mouse Novartis**	**human Novartis**	**mouse Lattin**	**human Roth**	**human GTEX**
ensembl.75	94,929	215,647	94,929	215,647	ensembl
No. of transcripts (genes)	(39,179)	(64,102)	(39,179)	(64,102)	Version 87 CRCh38.p7

dataset	**mouse Novartis**	**human Novartis**	**mouse Lattin**	**human Roth**	**human GTEX**
genes per chip	35,076 (16,960)	59,344 (14,522)	41,770 (17,079)	86,103 (20,774)	194,844 (57,431)
TRAs per chip	-		-	-	
3x median	12,377 (6,285) 37.06%	13,415 (3,448) 23.74%	23,531 (10,398) 60.88%	35,392 (9,837) 47.35%	71,297 (30,899) 53.80%
5x median	7,986 (4,172) 24.6%	7,726 (2,055) 14.15%	19,761 (8,924) 52.25%	22,289 (6,515) 31.36%	60,131 (27,339) 47.60%
10x median	4,329 (2,340) 13.8%	4,003 (1,042) 7.18%	14,289 (6,733) 39.42%	12,458 (3,775) 18.17%	51,352 (25,145) 43.78%
20x median	2,307 (1,307) 7.71%	2,115 (579) 3.99%	9,917 (4,833) 28.3%	6,897 (2,165) 10.42%	46,776 (23,808) 41.45%
housekeeping	(8,692) 51.25%	(9,938) 68.43%	(2,863) 17.08%	(9,148) 20.78%	-

Table 3.1: TRAs numbers and percentages

3.4 Housekeeping genes numbers and percentages

Besides tissue-restricted antigens, we also defound housekeeping genes as genes which are not higher expressed in any tissue than 2x the median gene expression line. With this definition we found between 2,863 housekeeping genes in the mouse Lattin dataset and 9,938 housekeeping genes in the human Novartis dataset. Housekeeping genes in the GTEX dataset have not been calculated in this study (Table 3.1). This accounts for about 17.08% of all genes in the mouse Lattin dataset and up to 68.43% in the human

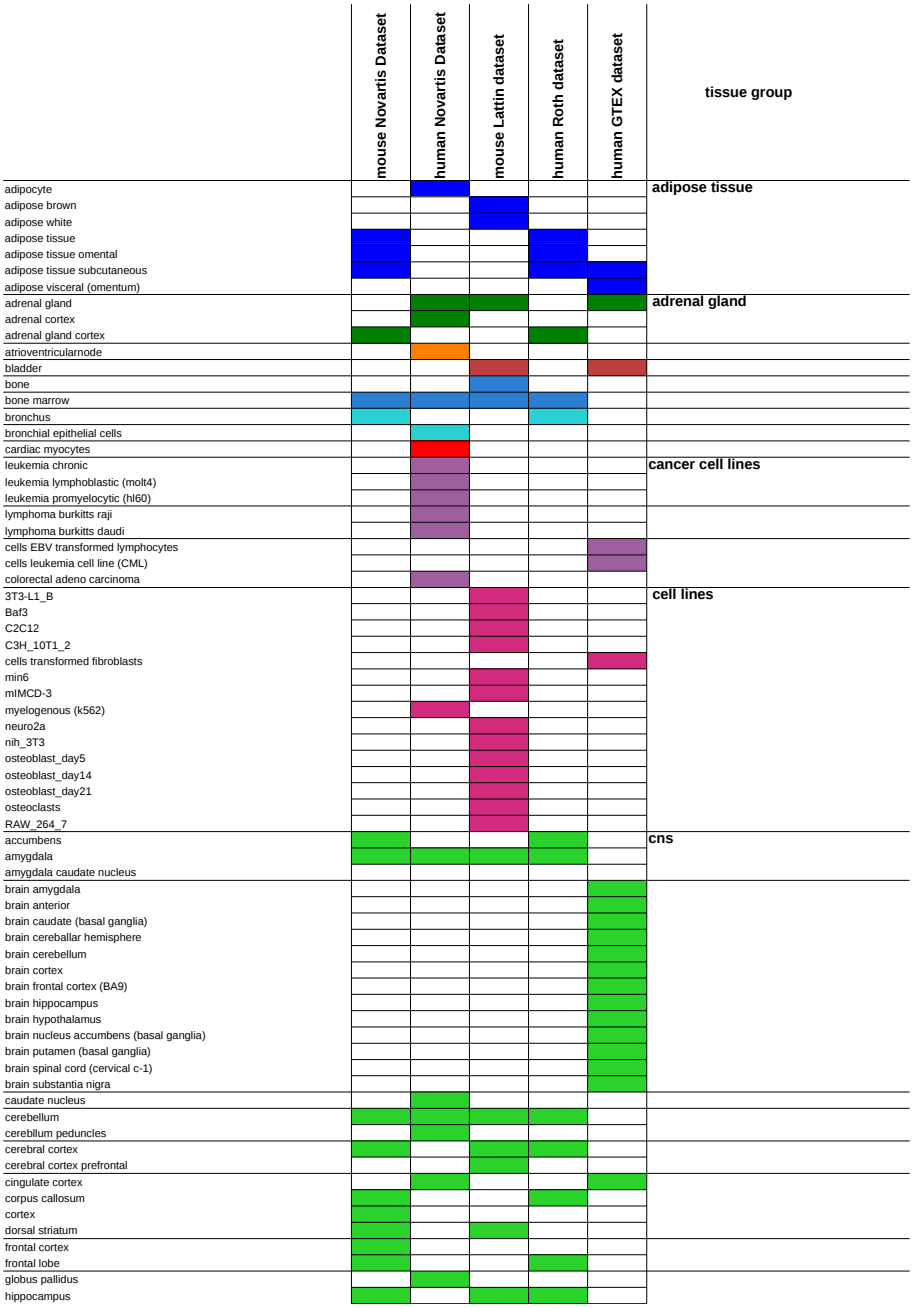


Figure 3.1: **Tissue types in all five datasets.** All represented tissue types are shown here in the various datasets. All tissue types belonging to the same biological instance, such as the central nervous system (CNS), the peripheral nervous system (PNS) and some other subgroups are grouped together and regarded in the TRA calculation as one tissue. Cells lines, such as cancer cell, motile cell lines, embryonic tissue, as well as all immune cells have been plotted in the TRA plots, but not regarded in TRA calculation.

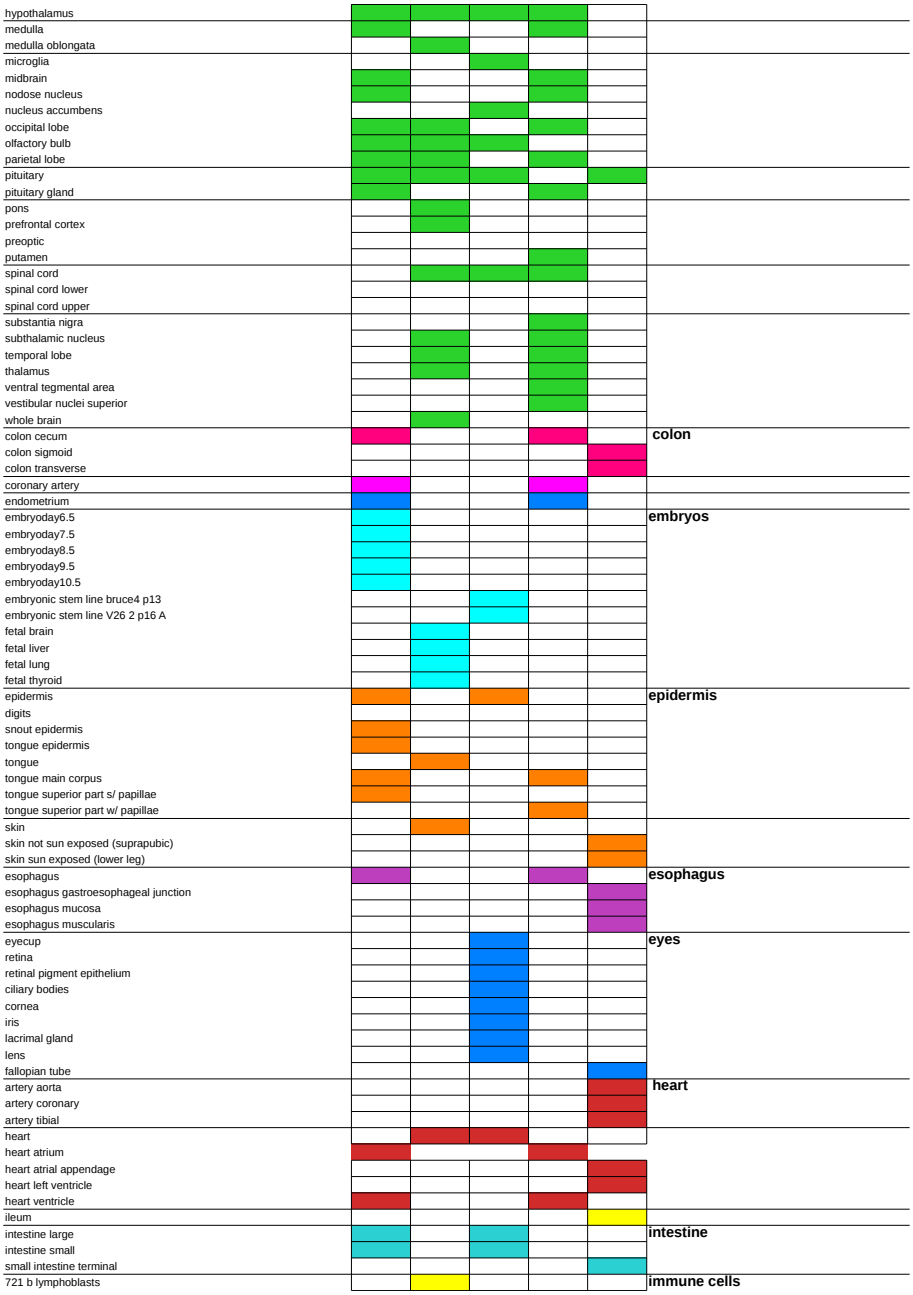


Figure 3.2: **Tissue types in all five datasets.** Most tissues are represented by at least one subtype in each dataset, some are represented by more than one tissue subtype per dataset.

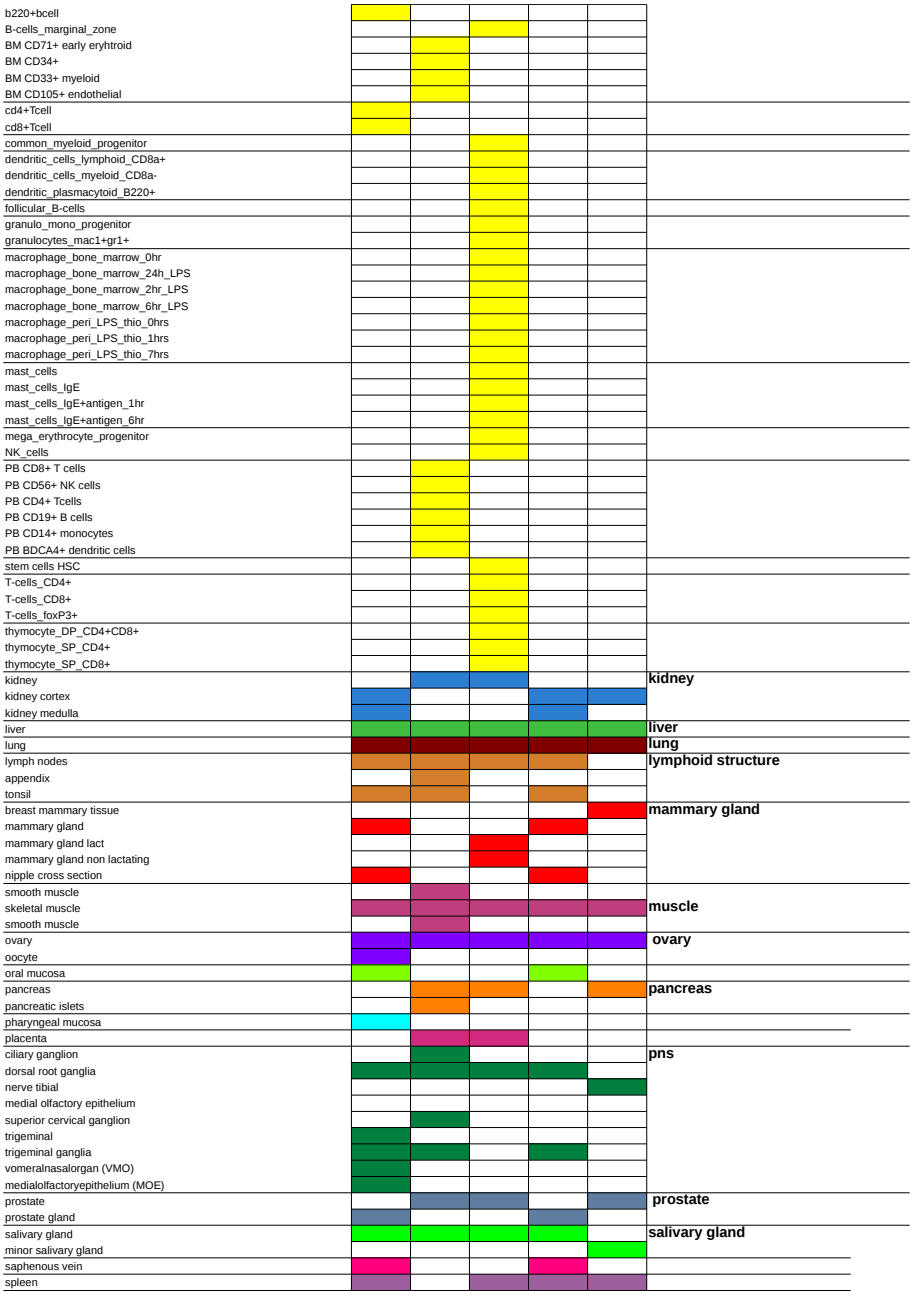


Figure 3.3: **Tissue types in all five datasets.** The yellow tissues are related to the different subtypes of immune cells, which have been plotted in the TRA plots but not been taken into account in the TRA definition, since they are motile cells.

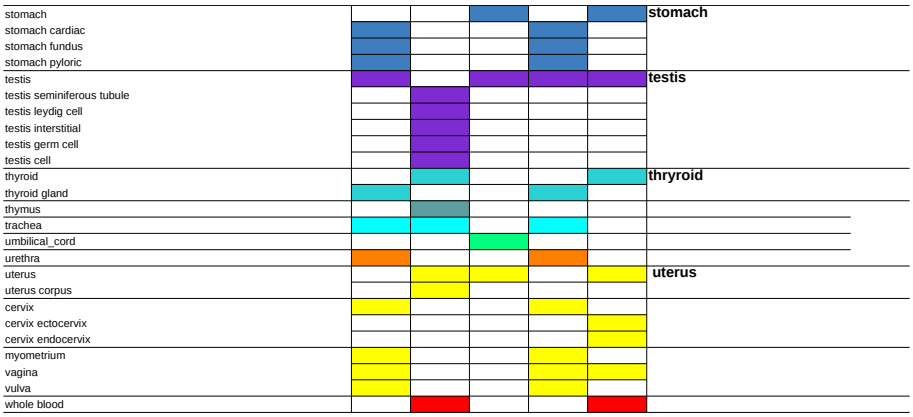


Figure 3.4: **Tissue types in all five datasets.** Altogether there are twenty-eight tissue groups in all datasets. Tissue grouping was discussed with Dr. Sheena Pinto, personal communication and adapted from previous work Dinkelacker 2007 [102].

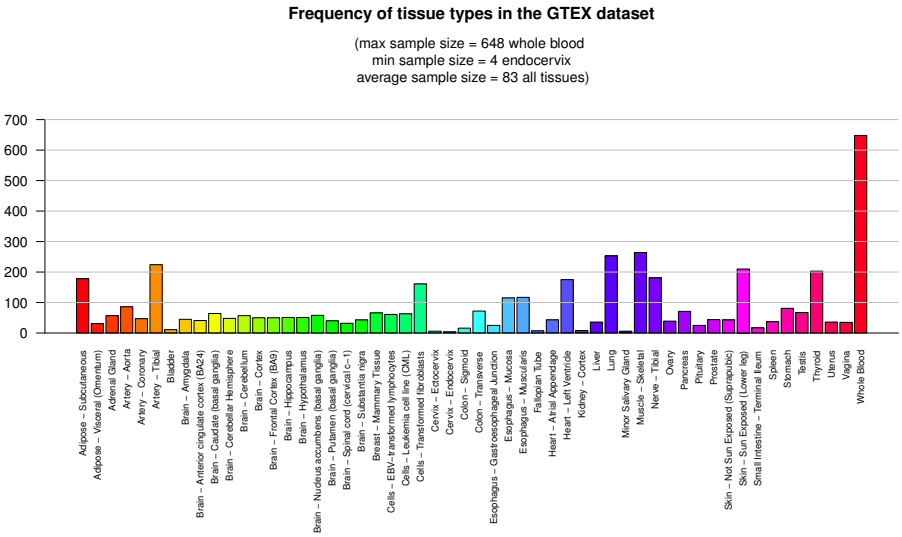


Figure 3.5: **Frequency of the tissue types in the human GTEX dataset.** The frequency of the tissue type within the human GTEX RNAseq dataset varies, depending on the type of tissue and donation. While there have only been a few samples for tissues such as the cervix, kidney and bladder (post mortem), there have been over 600 measurements in tissues such as the whole blood.

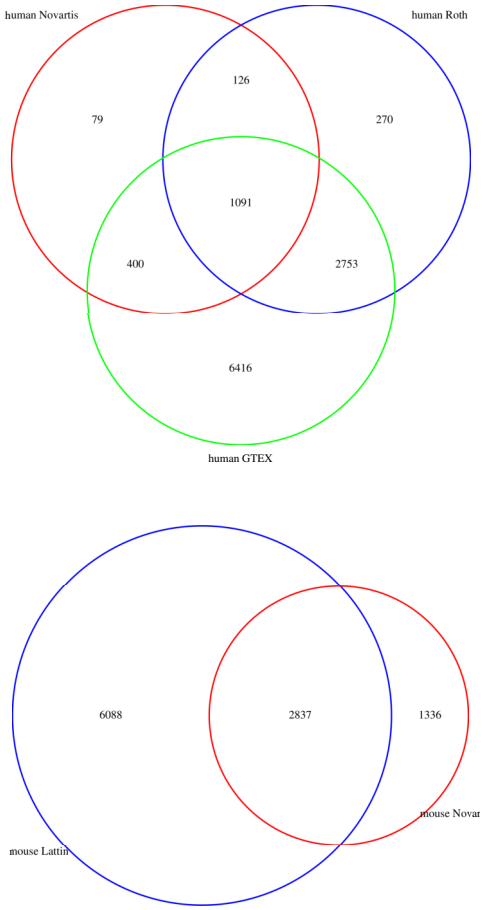


Figure 3.6: **Overlap of the TRAs in all datasets.** The overlap of common genes, which can be found to be tissue-restricted in each dataset is 1,091 common genes in the three human datasets and 2,837 genes in the two mouse datasets. Since the human Novartis dataset has been from 2002 already and resembles the oldest dataset here, we can only find 79 TRAs which are tissue-restricted in this dataset but not in the others. In the human Roth dataset, which is from 2006 we can find 270 extra TRAs and in the newest GTEX dataset from 2013 and 2015 we can find 6,416 TRAs more than in the other datasets. The same holds true for the mouse datasets, the oldest Novartis dataset in the mouse is from 2004 and contains 1336 TRAs more than the mouse Lattin dataset from 2008, which has 6,088 genes to be TRAs, which can not be found in the other dataset. Alltogether there is quite a high number of TRAs identified in all different datasets and thus builds a stable basis of TRA calculation.

Novartis dataset (Table 3.1). For all housekeeping genes we have calculated the overlap of genes found within the same species (Fig. 3.7).

Within different datasets there is an obvious relationship between the numbers of tissue-restricted genes and the numbers and percentages of housekeeping genes found per dataset. Besides tissue-restricted genes and housekeeping genes there are also genes which are not falling into either of the two groups, since the definition is not complementary. This method is limited in the aspect, that if genes are not highly expressed in any of the tissues, the cutoff line is so low, that 5x the median in the case of TRAs or 2x the median in the context of housekeeping genes is also very low, which gives in principle wrong results. A lower additional cutoff line would be helpful in order to solve this problem in the future.

3.5 Housekeeping genes in more detail

Analyzing the nature of housekeeping genes in further detail, we found many gene groups of cellular function. Among these we could detect 18 genes coding for actin filaments, 8 genes for apoptotic processes, 313 genes for ATP binding, 6 genes for autophagy, 149 genes for calcium ion binding, 348 genes for DNA binding and many other gene groups of the usual cellular biology needed in any tissue-type. Among the 350 genes specific for DNA binding, we could also detect the autoimmune regulator AIRE. This means that AIRE is actually expressed in basically all tissues. Varyfying this finding in bioGPS seems to give the same picture [434, 433, 432]. As AIRE is binding to histones, we also looked at its interaction partners in the STRING database [374] where this finding can also be seen. Besides AIRE we find many other DNA binding proteins in our housekeeping gene lists, such as different topoisomerases as well as polymerases (see housekeeping gene lists, technical appendix on CD). The distribution of the main functional gene groups of housekeeping genes are shown in Fig. 3.8.

3.5.1 TRAs in the mouse Novartis dataset

With the definition of TRAs to be genes, which are higher expressed than the cutoff of 5x the median gene expression in at least one and not more than five tissues of the 5x median, we can find 4,172 TRAs to be tissue-restricted in the mouse Novartis dataset. As example plots, we can thus find genes which are tissue-specific for one, two, three, four as well as five tissues over the 5x median gene expression line. Example plots are given in Fig. 3.9 - 3.11. Many of these genes have already been known to be involved in autoimmune diseases [102].

Similar to tissue-specific genes we can also find genes, which are too un-

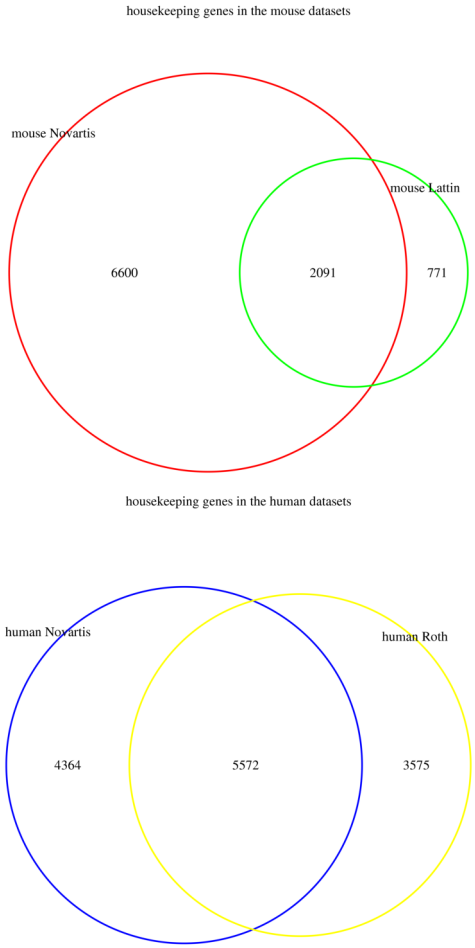


Figure 3.7: **Overlap of housekeeping genes in the human and mouse datasets.** A housekeeping gene was defined as a gene, which is not higher expressed in any tissue than 2x the median gene expression. In the mouse Novartis dataset, we found 8,692 housekeeping genes, in the mouse Lattin dataset, we found 2,863 housekeeping genes, 2091 of these genes overlap in both mouse datasets. In the human Novartis dataset, we found 9,938 housekeeping genes, in the human Roth dataset, we found 9,148 housekeeping genes, 5,572 of these overlap in both datasets.

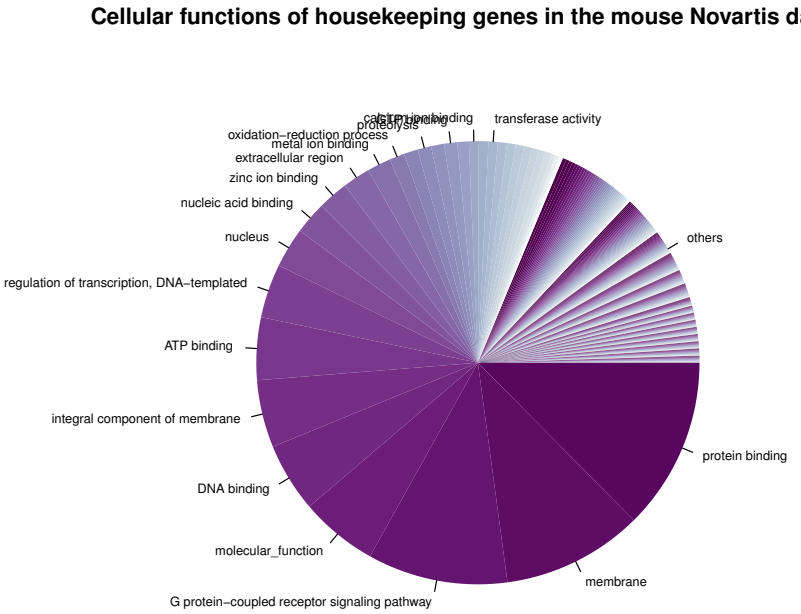


Figure 3.8: **Cellular functions of housekeeping genes in the mouse Novartis dataset.** Most of the 8,692 housekeeping genes in the mouse Novartis dataset are genes important for protein binding 871 genes (18.01%), directly followed by genes specific for the membrane 712 genes (10.02%), G protein coupled recptor signaling pathways 708 genes (8.19%), molecular function 392 genes (8.15%), DNA binding 350 genes (4.51%), integral component of the membrane 341 genes (4.03%), ATP binding 316 genes (3.92%), regulation of transcription 272 genes (3.64%), the nucleus 202 genes (3.13%), nucleic acid binding 165 genes (2.32%), zinc ion binding and extracellular region 154 genes (1.99%). Smaller groups are collected together to other functions, depicted here as "others". Genes which could not be annotated with GO terms were not plotted here.

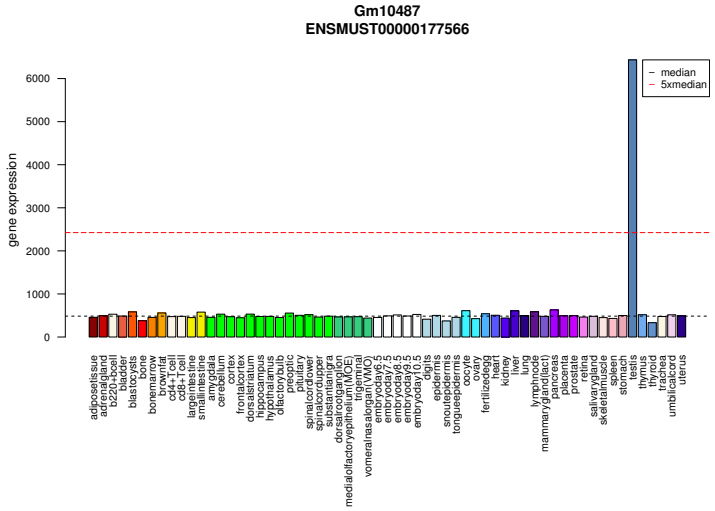


Figure 3.9: **TRA example for one tissue over the cutoff in the mouse Novartis dataset** A TRA is a gene, which is higher expressed in at least one and not more than five tissues over the 5x median gene expression line (red horizontal line) over all tissues. This example shows a TRA, which fulfills this criteria for one tissue over the cutoff. The Gm10487 gene is tissue specific for the testis and thus resembles a typical testis-specific antigen. Other examples of TRAs for one tissue are Insulin 1 and Insulin 2, shown in Fig. 3.16. In the TRA database (TRA-DB) the criteria of how many tissues to represent can be choosen upon the users wish. These figures has been adapted by and recalculated from previous work in Dinkelacker 2007 [102].

specific and expressed in more than five tissues over the 5x median gene expression line or genes, which are non TRAs and for example fall into the group of housekeeping genes, if they are not expressed higher than 5x the median gene expression line in any tissue of the dataset. Examples of these genes can be seen in Fig. 3.12. All TRAs can be found in the TRA database (TRA-DB), https://ibios.dkfz.de/tra/ [102].

3.5.2 TRAs in the human GTEX dataset

Identifying TRAs in the human GTEX dataset has been more difficult than in the microarray datasets, due to the high intrahuman variability as well as the big range of sample sizes per tissue-type. We accounted for this by calculating TRAs upon the median gene expression per gene/transcript per tissue-type over all samples and considered the 50% range to be the value used for TRA calculation. The intrahuman variability has been shown by boxplots in the TRA plots. Examples of TRAs in the GTEX dataset can be seen in Fig. 3.13. Examples of non TRAs can be seen in Fig. 3.15. Altogether

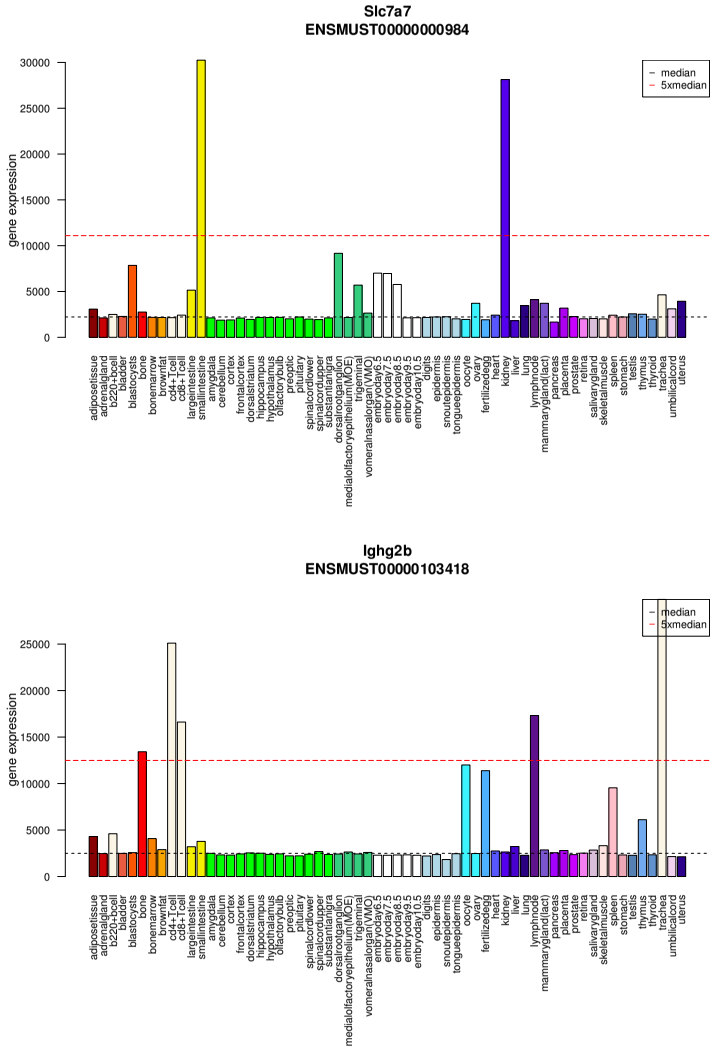


Figure 3.10: **TRA example with two and three tissues over the cutoff in the mouse Novartis dataset** SLC7A7 is an example of a TRA, which exceeds 5x the median gene expression line in two tissues over the cutoff. It is an amino acid transporter, which is tissue-specific for the kidney as well as the small intestine [40]. SLC7A7 is known to be involved in Lupus erythematosus [383, 247]. The gene IGHG2B is a TRA example for three tissues over the cutoff line it is tissue-specific for the bone, lymphnode as well as the trachea. Also IGHG2B is known to be connected to Lupus erythematosus [117, 306]

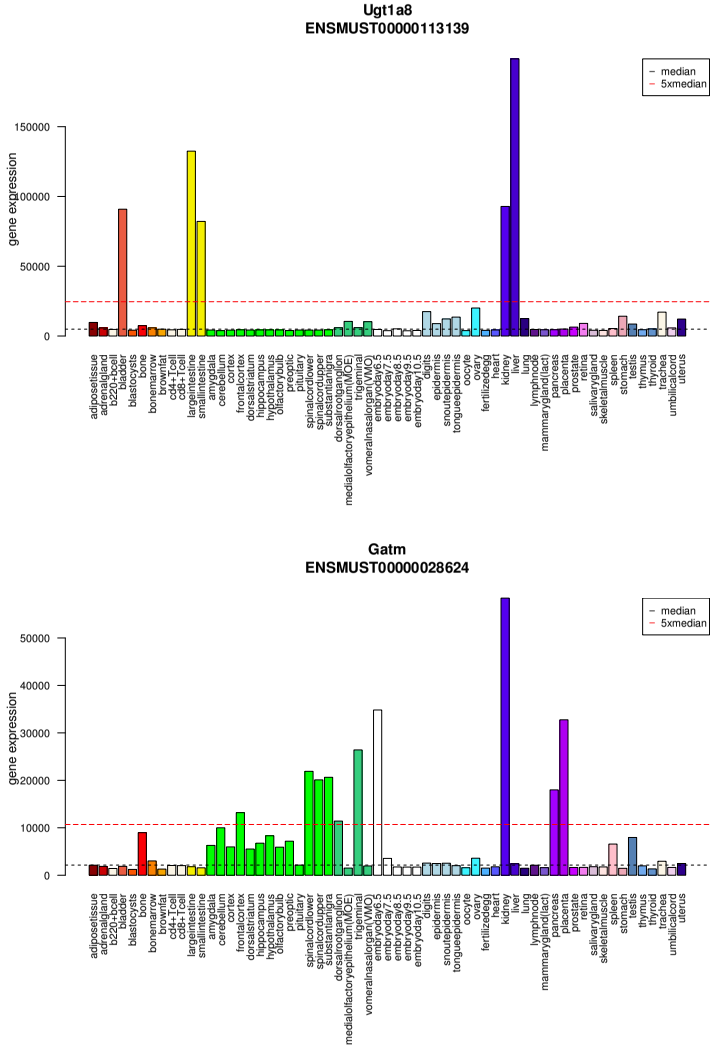
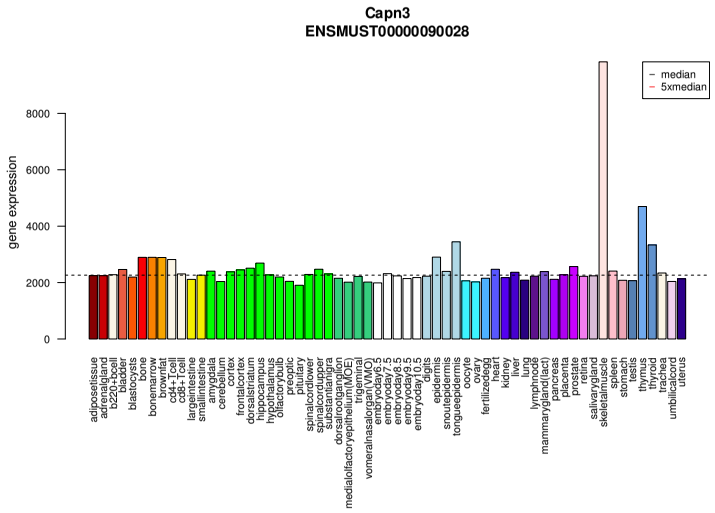
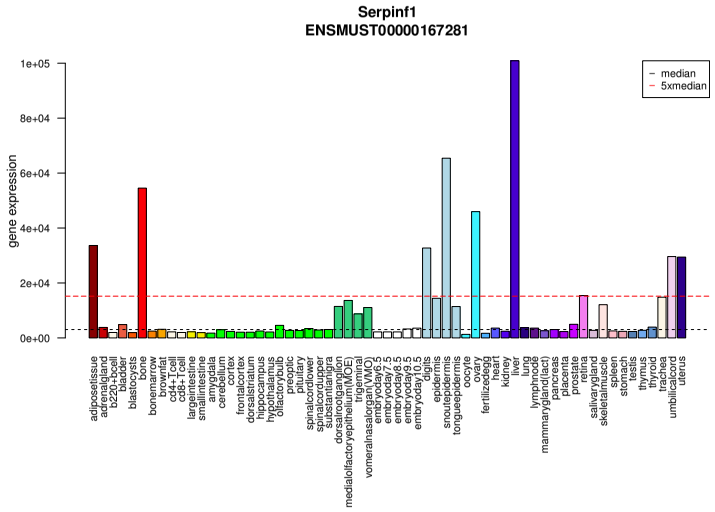


Figure 3.11: **Some examples of TRAs with four and five tissues over the cutoff in the mouse Novartis dataset.** The gene UGT1A8 is a TRA which exceeds the cutoff line in four tissues. It is tissue specific for the kidney, the liver, the intestine and the bladder. The genes UGT1A1, UGT1A4 [144], UGT1A6, UGT1A7, Ugt1A9, UGT1A10 and UGT2B7UGT1A7 have been found to be connected to inflammatory liver disease [21], In our data we can find UGT1A2, UGT1A3, UGT1A5, UGT1A8, UGT2B1, UBT2B15, UBGT2B17, UGT2B34, UGT2B36, UGT2B38 and UGT2B4 most of them tissue specific for the liver and the kidney. GATM is an example of a TRA which is tissue-specific for five tissues over the cutoff, the tissues of the central nervous system (CNS) are grouped together, light green bars, so are the tissues of the peripheral nervous system (PNS), dark green bars. GATM is known to be associated with kidney failure [57].



(a) 0 tissues over cutoff



(b) more than 5 tissues over cutoff

Figure 3.12: **Some examples of non TRAs in the mouse Novartis dataset.** In order to give some examples of non TRAs, there are genes which are expressed in virtually all tissues over the median gene expression line (CAPN3), such as the example here, as also a gene which is higher expressed in more than five tissues over the cutoff as the example of SERPINF1.

we could identify 1,091 common human TRAs and 2,837 common mouse TRAs in all datasets used in this study (Fig. 3.6). All TRA lists are in the technical appendix on CD.

3.6 TRAs associated to autoimmune diseases

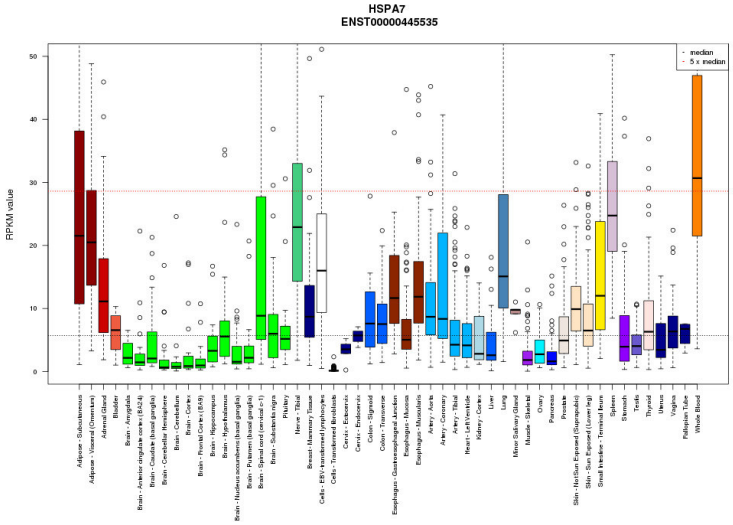
Since TRAs are expressed by medullary thymic epithelial cells (mTECs) in the thymus and are important for the negative selection of autoreactive T cells [214], we tested the association between TRAs and already known genes involved in autoimmune diseases. Doing a literature research by an automized pubmed search in order to look for the combination of the TRA gene symbols and certain keywords, we found about 1,129 genes in the human Novartis data to be connected within the title or abstract to the keyword "*autoimmune*", 40 TRAs to be involved in "*autoimmune gastritis*", 106 being related to "*hashimoto thyroiditis*", 189 TRAs being linked to "*juvenile idiopathic arthritis*", 587 to "*lupus erythematosus*", 311 to "*male infertility*" and 662 to "*multiple sclerosis*".

Going further into detail of these gene lists, we can find, that from the 40 TRAs related to autoimmune gastritis three are tissue-specific for the liver (the stomach was not measured in this dataset). In the case of the 54 genes related to hashimoto, only one matched to the thyroid. Looking at genes related to lupus erythematosus, we can find many different tissue-specificities. Since lupus erythematosus is not linked to a specific tissue-type this has been expected. In the case of male infertility at least fifteen of the genes are related to the testis. In the case of multiple sclerosis (MS) we can find eighty of the TRAs to be tissue-specific for the CNS.

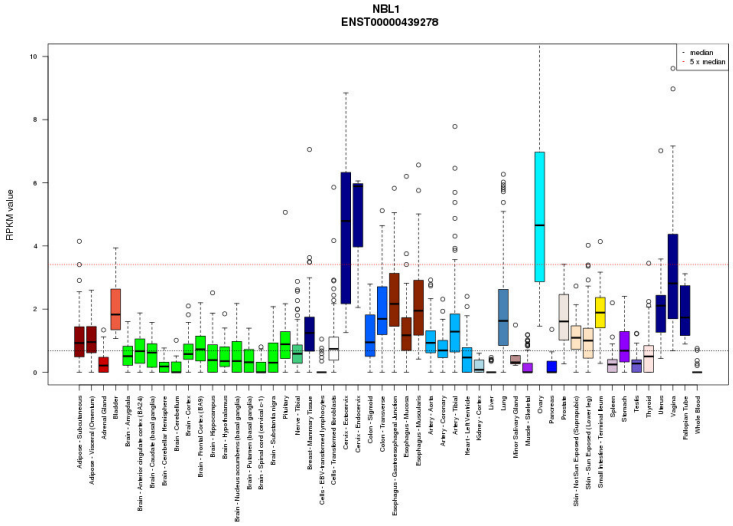
For a manual and thus more specific search, we extracted all genes being tissue-specific for only one tissue-type, for example the pancreatic islet cells and analyzed the data genewise upon previous publications, known to be involved in diabetes type 1, including these genes of interest. Besides diabetes type 1 in the case of pancreatic-specific genes, we also searched for the connection of diabetes type 2, pancreatitis as well as pancreatic cancer (Table 3.2).

3.6.1 Diabetes type 1

Diabetes type 1 is an autoimmune disease, which is characterized by a chronic insulin deficiency due to the loss of beta-pancreatic islet cells, which leads to hyperglycaemia. Patients usually get autoimmune diabetes type 1 during infancy, but symptoms can also develop much later [191]. Diabetes type 1 goes along with a chronic inflammation, which is characterized by autoimmune destruction of insulin-producing pancreatic beta cells [314].

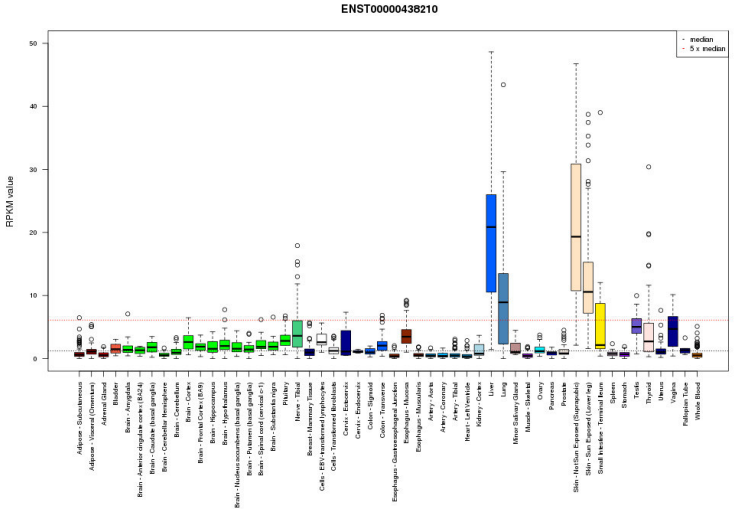


(a) 1 tissue over the cutoff



(b) 2 tissues over the cutoff

Figure 3.13: **Examples of TRAs with one and two tissues over the cutoff line in the GTEX dataset.** To define TRAs in the human GTEX dataset we have to account for intrahuman variability as well. Therefore we plotted the distribution of gene expression per gene over all samples in boxplots. Some genes have a high variability, others don't this variance is also dependent on the different tissue-types. The first example HSPA7 is a TRA which is tissue-specific for the whole blood, the second is tissue specific for the ovary, as well as the cervix.



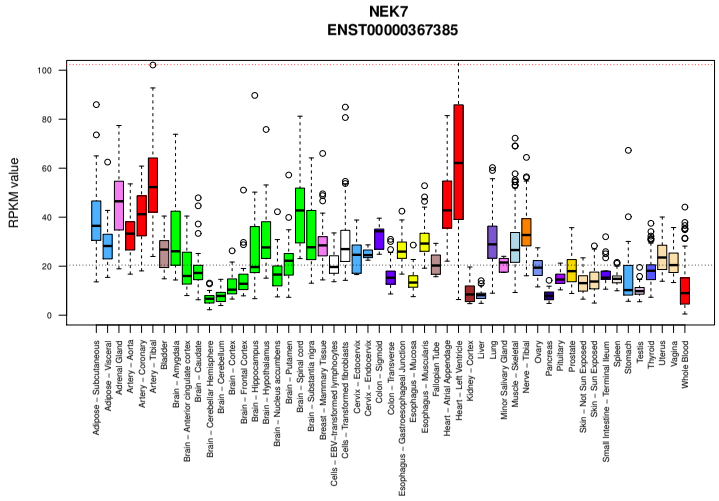
(a) 3 tissues over the cutoff

Figure 3.14: **Example of a TRA with three tissues over the cutoff in the human GTEX dataset.** As an example for three tissues over the cutoff line we have the Transcript ENST00000438210, which is tissue-specific for the liver, lung and skin, the two skin tissues are grouped together and regarded as one tissue.

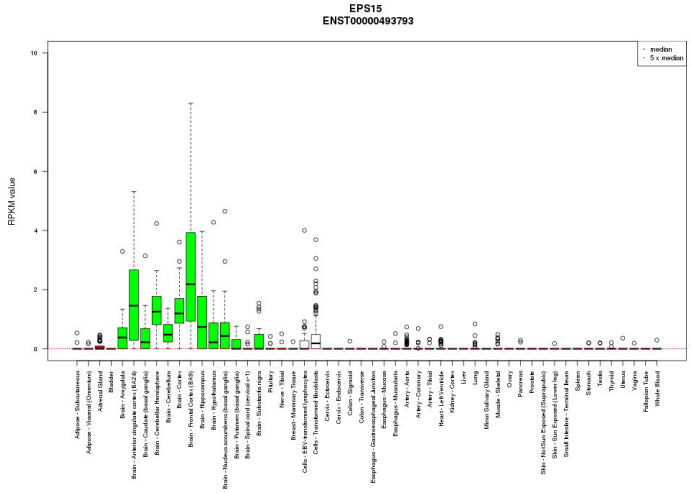
Through the inflammation antigen presenting cells make tissue-specific genes visible to the immune system and lead to autoimmune reactions. This distruction of the pancreatic beta cells leads to the failure of insulin production and this to diabetes type 1 typical symptoms. By the time that diabetes type 1 is diagnosed most of the beta pancreatic islet cells are usually already destroyed [381].

3.6.2 Known autoantibodies in diabetes type 1

According to the literature there is a list of already known autoantibodies, mostly TRAs tissue-specific for the pancreatic islet cells and involved in autoimmune diabetes type 1. The first autoantibody discovered in this context was the rat insulin gene [184]. At the same time they also found glucagon (GCG) as well as thy pancreatic polypeptide, somatostatin (SST), trypsin, amylase, caboxypeptidase A as well as GAD65 [184]. GAD65 was not expressed in the pancreas [184]. The amylase, carboxypeptidase A as well as GAD65 were not expressed in the thymus [184]. Derbinski et al. 2001 could show most of these TRAs to be mainly expressed in mTECs [98]. Gotter et al. 2004 increased this list by the elastase [144]. In our data, we could find most of these TRAs (Table 3.2).



(a) 0 tissues over the cutoff



(b) many tissues over the cutoff

Figure 3.15: **Examples of non TRAs in the GTEX dataset.** As an example of a non TRA in the human GTEX dataset, we found the NEK7 gene, which is not expressed in any of the tissues over the cutoff line, and the EPS15 gene, which is expressed higher than the cutoff line in many tissues. Most of them belong to the central nervous system, but since the median gene expression line is very low in most of the tissues, the five times median gene expression line is also very low. This gene is actually tissue-specific for the central nervous system (CNS), but not regarded as a TRA in our data, due to the criteria.

The insulin 1 and 2 genes [223] can be found in all of our datasets (Fig. 3.16), it is tissue-specific for the pancreatic islet cells. Other TRAs formerly associated with diabetes type 1 or genes, which are ought to be tissue-specific for the pancreas, such as GAD65 [366], IA-2 [98], GADA, ICA as well as ZNT98 [144, 194, 365, 235] can not be found in our data. But instead we can find other genes, such as CTRB2, GCG, IAPP, PCSK1, REG3A as well as SPINK1 [274] to be tissue-specific for the pancreatic islet cells (Fig. 3.17, 3.18). They have formerly been connected to diabetes type 1 [312, 85, 97, 416, 61].

For an overview of all TRAs tissue-specific for the pancreatic islet cells and potentially involved in autoimmune diabetes type 1, please refer to Table 3.2.

gene name	**gene description**	**associated with**	**autoimmune**	**reference**
C9[3]	-	diabetes type 1	autoimmune, tissue-specific for the liver	Nyalwidhe et al. 2017 [274]
CEL[3]	Carboxyl Ester Lipase	grave's disease	autoimmune	Strzelczyk et al. 2016 [370]
	-	pancreatic disease	-	Johansson et al. 2018 [181]
CELA2A	**Elastase-1**	**diabetes type 1**	**upregulated in MHC II hi mTECs, Pinto et al. 2008**	**Jolicoeur et al. 1994** [184]
CELA2B	-	**diabetes type 1**	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
CELA3B[3]**		**diabetes type 1** [274]	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
CEL2B	Chymotrypsin Like Elastase	-	-	
CEL3A	-	chronic pancreatitis	-	Párniczky et al. [307]
CEL3B	-	chronic pancreatitis	-	Párniczky et al. [307]
	-	diabetes type-2	-	Han et al. 2011 [150]
CELP	Carboxyl Ester Lipase Pseudogene	pancreatic diseases	-	Johansson et al. 2018[181]
CLPS	**Colipase**	diabetes type 2 [230]	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**

Table 3.2 – TRAs tissue-specific for pancreatic islet cells

gene name	gene description	associated with	autoimmune	reference
CPA1	Carboxypeptidase A1	pancreatitis	-	Masamune et al. 2014 [248]
	-	pancreatic cancer	-	Nagaraja et al. 2013 [267]
	-	diabetes type 2	not autoimmune	Han et al. 2011 [150]
CD69[1]	-	diabetes type 1, tissue-specific for adipose tissue, lymphnode, lung and spleen	autoimmune	Reddy et al. 2011 [312]
CHGA*[2]**	-	pancreatic cancer [106]	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
CHGB*[2]**	-	pancreatic cancer [106]	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012** [144]
CLEC16A	-	diabetes type 1, no TRA	autoimmune	Reddy et al. 2011 [312]
CPA2	Carboxypeptidase A2	diabetes type 2	not autoimmune	Han et al. 2011 [150]
CPB1[3]	Carboxypeptidase B1	diabetes type 1	autoimmune	Nyalwidhe et al. 2017 [274]
CRP	C-Reactive Protein	autoimmune diseases	autoimmune	Klein et al. 1998 [202]
CTLA4[1]	-	diabetes type 1	autoimmune	Reddy et al. 2011 [312]
CTRB2	Chymotrypsinogen B2	diabetes type 1	autoimmune	Reddy et al. 2011 [312]
CTRC	Chymotrypsin C	chronic pancreatitis	-	Giefer et al. 2017 [132]
CTRL[3]**	-	diabetes type 1	-	Nyalwidhe et al. 2017 [274]
CTSH[1]	-	**diabetes type 1** [312]	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
CUZD1	**CUB and zona pellucida-like domain-containing protein 1**	inflammatory bowel diseases[113]	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**

Table 3.2 – TRAs tissue-specific for pancreatic islet cells

gene name	gene description	associated with	autoimmune	reference
DNAJC12	DnaJ heat shock protein family (Hsp40) member C12	-	-	-
EPB41L4B	Erythrocyte membrane protein band 4.1 like 4B	-	-	-
ERBB3[1]	-	**diabetes type 1** [312]	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
GCG	**glucagon**	**diabetes type 1**	**expressed in MHC II hi mTECs, Pinto et al. 2008**	**Jolicoeur et al. 1994** [184, 144]
GGT1	glutathione hydrolase 1 proenzyme precursor	pancreatitis	-	Shelton et al. 2014 [349]
GGT3P	putative glutathione hydrolase 3 proenzyme	-	-	-
GGT8P	gamma-glutamyltransferase 8 pseudogene	-	-	-
GGTLC2	Glutathione hydrolase light chain 2	-	-	-
GGTLC3	gamma-glutamyltransferase light chain family member 3	-	-	-
GLIS3[1]	-	**diabetes type 1** [312]	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
CP2	cystein protease	-	-	-
IAPP	**Islet amyloid polypeptide precursor**	**diabetes type 1**[97]	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
IFIH1[1]	-	diabetes type 1	autoimmune, no TRA	Reddy et al. 2011 [312]
IL27[1]	-	diabetes type 1	autoimmune, no TRA	Reddy et al. 2011 [312]

Table 3.2 – TRAs tissue-specific for pancreatic islet cells

gene name	gene description	associated with	autoimmune	reference
INS[1]	**insulin**	**diabetes type 1**	**upregulated in MHC II hi mTECs, Pinto et al. 2008**	**Jolicoeur et al. 1994** [184]
ITPR3[1]	-	**diabetes type 1** [312]	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
MMP7	**matrix metallopeptidase 7**	systemic lupus erythematosus [403]	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
PCSK1[3]	proprotein convertase subtilisin/kexin type 1	diabetes type 1	autoimmune	Wasserfall et al. 2017 [416]
PCSK1N[3]**	-	diabetes type 1	autoimmune	Nyalwidhe et al. 2017 [274]
PCA2G1B	calcium-transporting ATPase G1B	-	-	-
PNLIP[3]**	pancreatic lipase	diabetes type 1	autoimmune, tissue-specific for the pancreas and the spleen	Nyalwidhe et al. 2017 [274]
PNLIPRP1	**pancreatic lipase related protein 1**	pancreatic cancer [453]	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
PNLIPRP2	pancreatic lipase related protein 2	pancreatic cancer	-	Zhang et al. 2013 [453]
PKD2[1]	-	diabetes type 1, tissue-specific for the uterus, umbilicalcord and kidney	autoimmune	Reddy et al. 2011 [312]
PP, PPY	**pancreatic polypeptide**	**diabetes type 1**	**autoimmune**	**Jolicoeur et al. 1994** [184]
PRKCQ[1]	-	diabetes type 1, tissue-specific for the skeletal muscle	autoimmune	Reddy et al. 2011 [312]
PRSS2	**serine protease 2**	hashimoto [140]	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012** [144]

Table 3.2 – TRAs tissue-specific for pancreatic islet cells

gene name	gene description	associated with	autoimmune	reference
PRSS3	serine protease 3	autoimmune pancreatitis	autoimmune	Loehr et al. 2010 [238]
PTPN2[1]	-	diabetes type 1, no TRA	autoimmune	Reddy et al. 2011 [312]
PTPN22[1]	-	**autoimmune diabetes type 1** [312]	**downregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
REG1	-		**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
REG1A**	regeneration Family Member 1	Sjögren's syndrome	autoimmune	Yoshimoto et al. 2013 [449]
REG1B[3]	regeneration Family Member 1 beta	diabetogenesis	autoimmune	Nyalwidhe et al. 2017 [274]
REG1P	regeneration Family Member 1	colorectal cancer	not autoimmune	Lennard et al. 2016 [220]
REG2**	-		pancreas, spleen	-
REG3A[3]	**regeneration Family Member 3**	**type 1 diabetes** [274]	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
RNLS[1]	-	diabetes type 1, tissue-specific for the testis	autoimmune	Reddy et al. 2011 [312]
S100A8[3]	-	**diabetes type 1** [274]	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
S100A9[3]	-	**diabetes type 1** [274]	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
SH2B3[1]	-	type 1 diabetes	autoimmune, TRA tissue specific for whole blood and the spleen	Reddy et al. 2011 [312]
SIRPG[1]	-	diabetes type 1, no TRA	autoimmune	Reddy et al. 2011 [312]
SKAP2[1]	-	diabetes type 1, tissue specific for the testis, CNS and oocyte	autoimmune	Reddy et al. 2011 [312]
SPINK1[3]	serine protease inhibitor Kazal-type 1	type 1 diabetes	autoimmune	Chang et al. 2014 [61]

Table 3.2 – TRAs tissue-specific for pancreatic islet cells

gene name	gene description	associated with	autoimmune	reference
SSP[3]	**somatostatin**	**diabetes type 1**	**autoimmune**	**Jolicoeur et al. 1994** [184]
SYCN[3]	-	**diabetes type 1** [274]	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**

Table 3.2: **TRAs tissue-specific for pancreatic islet cells**, ** tissue-specific for more than one tissue, *** tissue-specific for more than two tissues, [2] pancreatic cancer, Dugnani et al. 2018 [106], [3] diabetes type 1, Nyalwidhe et al. 2017 [274], [4] Pinto et al. 2008 (internal data), Gaertner et al. 2012 (internal data)

Table 3.2 – TRAs tissue-specific for pancreatic islet cells

3.6.3 Myasthenia gravis

Myasthenia gravis (MG) is an autoimmune disorder of the neuromuscular junction. Known autoantibodies involved in MG are the acetylcholin receptor (ACHR) [333], the muscle-specific kinase (MUSK) as well as the lipoprotein-related protein 4 (LPR4) [185]. Neither of them can be found in our TRA database. According to Petrov et al. [295] also the inhibition of the acetylcholin esterase might help in order to cure myasthenia gravis. ACHE is a TRA, which is tissue-specific for the central nervous system (CNS) as well as for the skeletal muscle (Fig. 3.20). Genes with a similar expression pattern as the acetylcholinesterase are shown in Fig. 3.21 and Table 3.3.

In order to find more possible candidates which might be involved in myasthenia gravis (MG) we are looking for genes, which are tissue-specific for the skeletal muscle as well as the central nervous system (CNS). Since the search for the central nervous system is difficult due to the high amount of different central nervous system tissues, we started the search for TRAs tissue-specific for the skeletal muscle and apointed double tissue-specificities if available (Table 3.3). Several of these genes have already been linked according to the literature to myasthenia gravis, among them, the ACHE [295], ACHR (no TRA) [333], BIN1 [227], CAP2 [227], CAV3 [341], CRMP5 [261] and many others (Table 3.3). This means that finding new autoantibodies as well as diagnostic markers for MG should be possible this way.

Futhermore also other muscle related illnesses or biological functions not yet known can be found with the list of muscle-specific genes, according to the literature many of these genes have already previously been linked to muscle specificity (Table 3.3). A systematic search with these genes in myasthenia gravis (MG) patients will highlight the real impact of this gene

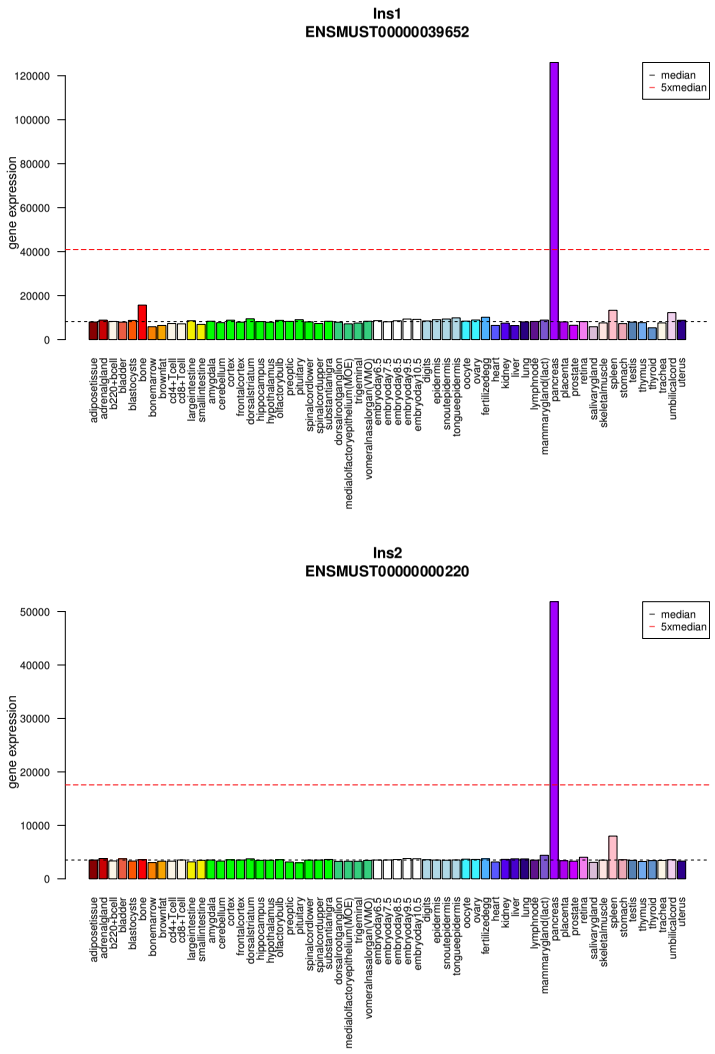


Figure 3.16: **Insulin 1 and 2 being involved in autoimmune diabetes type 1.** Both Insulin 1 and 1 have been known to be involved in autoimmune diabetes type 1 [223]. Both are tissue restricted for the pancreatic islet cells and can be found as a TRA in any dataset. These figures have been re-calculated and reproduced from previous work in Dinkelacker 2007 [102].

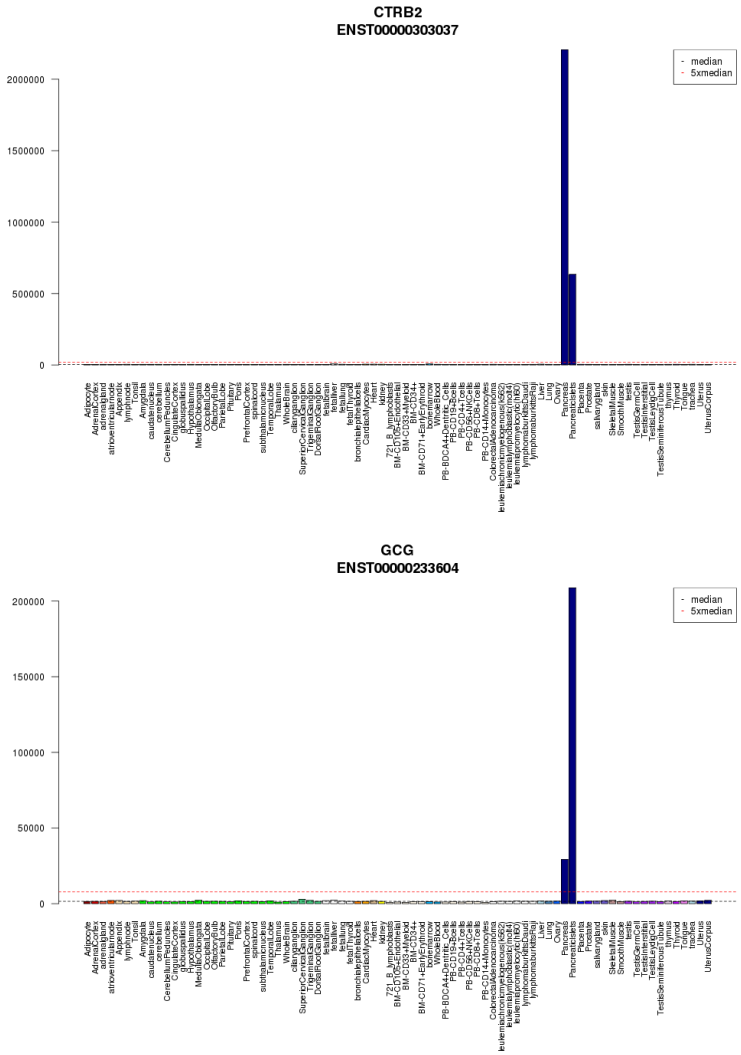


Figure 3.17: **CTRB2 and GCG being involved in autoimmune diabetes type 1.** Chymotrypsinogen B2 [312], CTRB2 [85] as well as glucagon, GCG [144] are known to be involved in autoimmune diabetes type 1. While CTRB2 is also associated to pancreatic cancer [431] as well as to chronic pancreatitis [421], glucagon (GCG) has already been related to autoimmune diabetes type 1 [184].

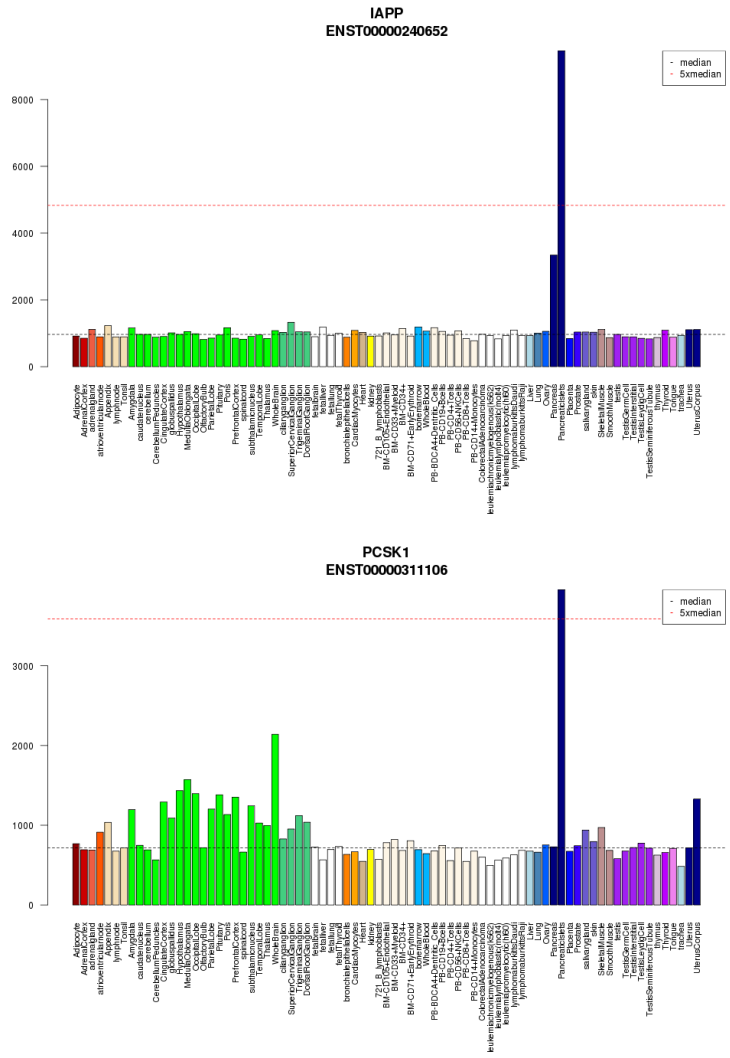


Figure 3.18: **IAPP and PCSK1 being involved in autoimmune diabetes type 1.** Both IAPP (Islet amyloid polypeptide precursor) as well as PCSK1 (Proprotein convertase subtillisin/kexin type 1) are known to be involved in autoimmune diabetes type 1 [97, 416]. Both of these genes are tissue-specific for the pancreatic islet cells. IAPP is the islet amyloid polypeptide (amylin), a circulating peptide, which is produced in beta cells by a precursor pro IAPP [82]. PCSK1 as well as PCSK2 are prohormone convertases, which process proinsulin into insulin. PCSK2 processes also glucagon. The gene expression of PCSK1 is reduced in diabetes type 1 patients [416].

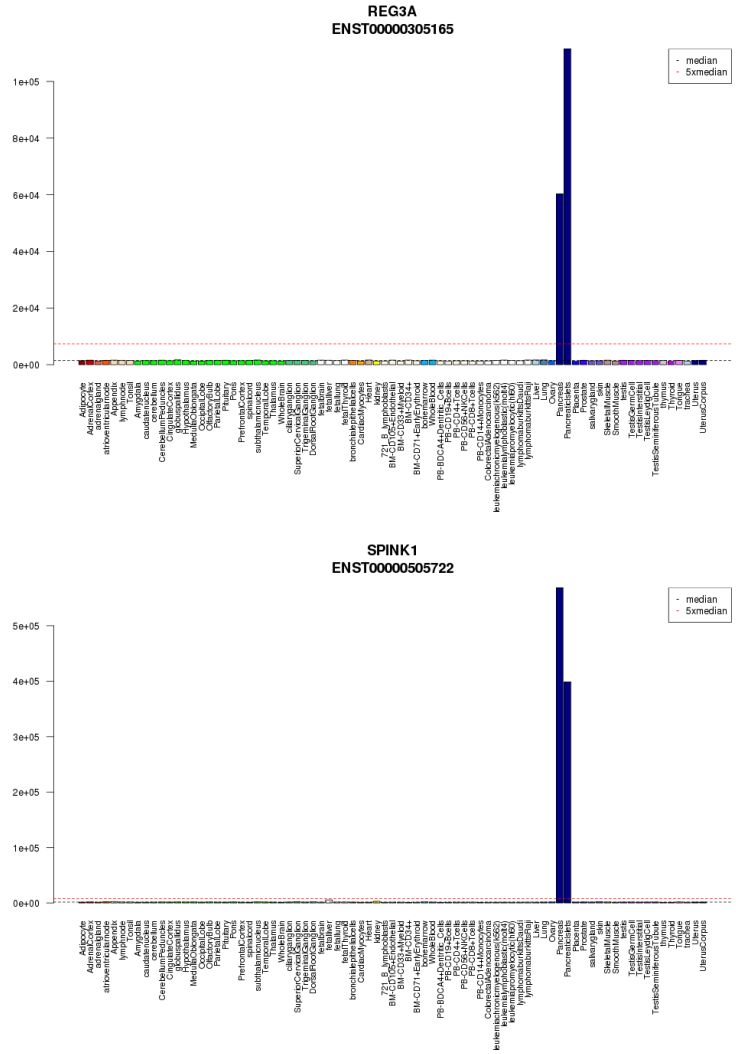


Figure 3.19: **REG3A and SPINK1 being involved in autoimmune diabetes type 1.** REG3A as well as REG1, REG1A, REG1B, REG1P, REG2 and REG3D are tissue-specific for the pancreatic islet cells, and known to be highly expressed in diabetes type 1 [274]. While REG1B is associated with the pancreas regeneration (Yamauchi et al. 2015), the upregulation of REG3A is known to be associated in acute phase pancreatitis [278]. The serine protease inhibitor SPINK1, a pancreatic secretory trypsin inhibitor is an inactivation factor of intra-pancreatic trypsin activity [184]. Its mutation is associated with idiopathic chronic pancreatitis [30, 428]

list to actual patients.

gene name	gene description	associated with	autoimmune	reference
AARSD1	skeletal muscle	-	muscle gene	Echeverría et al. 2016 [108]
ACHE	**CNS, skeletal muscle**	-	myasthenia gravis	Petrov et al. 2018 [295]
	-		**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
ACHR	-	no TRA	myasthenia gravis	Salmon et al. 1998[333]
ACTN3	skeletal muscle	-	muscle gene	Garton et al. 2018 [126]
ACYP2	skeletal muscle	-	-	Kim et al. 2007 [198]
ADCK3	skeletal muscle	-	muscle gene	Rooney et al. 2017 [323]
ADCY2	skeletal muscle	-	-	Silver et al. 2012 [354]
ADSL	skeletal muscle	-	-	-
ANKRD2	skeletal muscle	-	muscle gene	Koskinen et al. 2017 [207]
ANO5	skeletal muscle, PNS	-	-	-
AP2M1	smooth muscle, CNS	-	-	-
AQP4	skeletal muscle, CNS	-	-	-
ARPP21	skeletal muscle, CNS	-	muscle gene	Davegårdh et al. 2017 [89]
ASB8	skeletal muscle	-	-	-
ASPH	skeletal muscle	-	-	-
ATP5D	skeletal muscle	-	-	Chang et al. 2018 [60]
ATP1A2	skeletal muscle, CNS	-	-	-
BIN1	skeletal muscle, CNS	-	myasthenia gravis	Liewluck et al. 2011 [227]
BNIP3	-		-	
C11orf67	skeletal muscle	-	myasthenia gravis	Zhang et al. 2019
C20orf166	skeletal muscle	-	-	-
C22orf15	skeletal muscle	-	myasthenia gravis	Zhang et al. 2019
CACNA2D1	skeletal muscle	-	-	-

Table 3.3 – TRAs tissue-specific for skeletal muscle

gene name	gene description	associated with	autoimmune	reference
CACNG1	**skeletal muscle**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
CAMK2B	skeletal muscle, CNS	-	-	-
CAP2	skeletal muscle, CNS	-	myasthenia gravis	Liewluck et al. 2011 [227]
CAPN3	skeletal muscle, CNS	-	-	-
CAV3	skeletal muscle	-	myasthenia gravis	Schoser et al. 2009 [341]
CDK2AP1	-	-	-	-
CLIP1	skeletal muscle	-	coronary artery	Cho et al. 2019 [71]
CNBP	skeletal muscle	-	Il-12 transcription, Th1 immunity	Chen et al. 2018 [68]
COQ9	skeletal muscle	-	cardiomyopathy	Sondheimer et al. 2017 [362]
CRMP1	CNS	-	inflammation on the neurotransmission of vascular smooth muscle	Gan et al. 2017 [125]
CRMP5	-	no TRA	myasthenia gravis, autoantibodies	Monstad et al. 2009 [261]
CUTC	skeletal muscle	-	cardiac troponin	Mahmud et al. 2019 [244]
DCAF6	skeletal muscle	-	-	Chen et al. 2015 [65]
DNAJB5	**skeletal muscle**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
DNM1	CNS	-	-	-
DNM2	-	no TRA	myasthenia gravis	Liewluck et al. 2011 [227]
DOK7	skeletal muscle	-	-	-
DPYSL2	**skeletal muscle, CNS**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012** [144]
DPYSL5	CNS	-	myasthenia gravis, autoantibodies	Monstad et al. 2009 [261]

Table 3.3 – TRAs tissue-specific for skeletal muscle

gene name	gene description	associated with	autoimmune	reference
EEF1A2	skeletal muscle, CNS	-	-	-
EHBP1L1	skeletal muscle	-	-	-
ENDOG	skeletal muscle	-	-	-
ENO3	skeletal muscle	-	human mTECs	Gotter et al. 2004 [144]
EPM2A	skeletal muscle	-	muscle glycogen	Irimia et al. 2015 [170]
EYA4	skeletal muscle	-	-	-
FBXO3	skeletal muscle	-	-	-
FBXO32	skeletal muscle	-	mysthenia gravis	Chauhan et al. 2013 [63]
FEZ2	skeletal muscle	-	-	-
FNDC5	**skeletal muscle**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
FXR1	skeletal muscle	-	-	-
GAMT	skeletal muscle	-	sceletal muscle	Chen et al. 2019 [67]
IDI2	skeletal muscle	-	-	-
IGFN1	skeletal muscle	-	-	-
IP6K3	skeletal muscle	-	-	-
IL6	smooth muscle	-	myasthenia gravis	Huang et al. 2018 [161]
IL17D	-	-	-	-
JPH2	skeletal muscle	-	-	-
KPNA4	skeletal muscle	-	-	-
LRRC20	skeletal muscle	-	-	-
LPR4	-	no TRA	myasthenia gravis	Jordan et al. 2018[185]
MUSK	-	no TRA	myasthenia gravis	Jordan et al. 2018[185]
MTM1	-	no TRA	myasthenia gravis	Liewluck et al. 2011 [227]
MACROD1	skeletalmuscle	-	-	-
MAPK12	skeletal muscle	-	-	-
MDH2	**skeletal muscle**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
MEF2C	skeletal muscle, CNS	-	-	-
METTL21C	skeletal muscle	-	skeletal muscle	Wiederstein et al. 2018 [426]
MRPL15	skeletal muscle	-	-	-
MYBPC2	**skeletal muscle**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**

Table 3.3 – TRAs tissue-specific for skeletal muscle

gene name	gene description	associated with	autoimmune	reference
MYH4	skeletal muscle	-	-	-
MYLK2	skeletal muscle	-	-	-
MYOZ1	skeletal muscle	-		Roberts et al. 2018 [319]
MYPN	skeletal muscle	-	-	-
NANOS1	skeletal muscle	-	-	-
NDUFS7	skeletal muscle	-	-	-
NEDD4	skeletal muscle	-	-	-
NFE2L1	skeletal muscle	-	muscle atrophy	Furuya et al. 2014 [123]
NRD1	skeletal muscle	-	-	-
NDRG2	skeletal muscle, CNS	-	-	-
OPTN	skeletal muscle	-	-	-
OTUD1	skeletal muscle	-	-	-
PABPC4	skeletal muscle	-	-	-
PARVB	skeletal muscle	-	-	Matsuda et al. 2008 [250]
PDE4DIP	skeletal muscle	-	-	-
PDLIM5	skeletal muscle	-	-	-
PDLIM7	skeletal muscle	-	-	-
PHKA1	skeletal muscle	-	-	-
PLCD4	skeletal muscle	-	-	-
PLEC	skeletal muscle	-	-	Selcen et al. 2011 [346]
PPP3CB	skeletal muscle	-	-	-
PRKAG3	skeletal muscle	-	-	-
PRKCQ	skeletal muscle	-	-	-
RAD23A	skeletal muscle	-	-	-
RBFOX1	-	-	-	-
RBFOX3	CNS	-	-	-
RIF1	skeletal muscle	-	-	-
RNF123	skeletal muscle	-	-	-
RPUSD4	skeletal muscle	-	-	-
RYR1	**skeletal muscle**	-	**myasthenia gravis**	Stefanou et al. 2016 [367]
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
RYR3	**CNS**	-	myasthenia gravis	Hong et al. 2016 [159]
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
RTN1	**CNS**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**

Table 3.3 – TRAs tissue-specific for skeletal muscle

gene name	gene description	associated with	autoimmune	reference
RTN2	muscle	-	-	-
RTN3	CNS	-	-	-
RTN4R	CNS	-	-	-
SCN4A	skeletal muscle	-	-	Kao et al. 2018 [189]
SGCA	skeletal muscle	-	-	-
SHISA4	skeletal muscle	-	-	Rodrigues et al. 2019 [320]
SNTA1	skeletal muscle	-	-	-
SRPK3	skeletal muscle	-	-	-
SVIL	skeletal muscle	-	-	-
S100A1	-	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
S100b	CNS	-	myasthenia gravis	Mu et al. 2011 [262]
SLN	**muscle**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
SLN1B	-	no TRA	myasthenia gravis	Feng et al. 2019 [114]
TNIP1	immune cell specific	-	myasthenia gravis	Geng et al. 2016 [129]
TACC2	skeletal muscle	-	-	-
TBX15	skeletal muscle	-	-	-
TEAD4	skeletal muscle	-	-	-
TIAF1	skeletal muscle	-	-	-
TMEM38B	skeletal muscle	-	-	Webb et al. 2017 [417]
TMEM70	skeletal muscle	-	-	-
TNNI1	skeletal muscle	-	human mTECs	Gotter et al. 2004 [144]
TNNT3	**skeletal muscle**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
TPI1	skeletal muscle	-	-	-
TPM2	**skeletal muscle**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
TPM3	**skeletal muscle**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**

Table 3.3 – TRAs tissue-specific for skeletal muscle

gene name	gene description	associated with	autoimmune	reference
TPM4	**skeletal muscle**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
TRIM7	skeletal muscle	-	-	-
UBAC1	**skeletal muscle**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
UBE2B	skeletal muscle	-	-	-
UBE2D1	skeletal muscle	-	-	-
UBE2G1	skeletal muscle	-	-	-
UBELQL1	-	no TRA	-	Topakian et al. 2019[382]
UBR3	skeletal muscle	-	-	-
UCHL1	skeletal muscle, CNS	-	-	-
UQCRC1	skeletal muscle	-	-	-
UTP11L	skeletal muscle	-	-	-
VDAC3	skeletal muscle	-	-	Poleti et al. 2018 [301]
VGLL2	skeletal muscle	-	-	-
VPS72	skeletal muscle	-	-	-
ZMYND17	skeletal muscle	-	-	-

Table 3.3: **TRAs tissue-specific for muscle specific genes potentially involved in myasthenia gravis (MG)**, not all genes are annotated due to time reason, ** more than one tissue, Pinto et al. 2008 (internal data), Gaertner et al. 2012 (internal data)

Table 3.3 – TRAs tissue-specific for skeletal muscle

3.6.4 Multiple sclerosis

Multiple sclerosis (MS) is an autoimmune illness, where autoreactive lymphocytes cross the blood-brain barrier (BBB) and lead to demyelination, gliosis and neuroaxonal degeneration of the central nervous system, disrupting the neuronal signalling [96]. Known autoantibodies in multiple sclerosis patients are the myelin basic protein (MBP) (Fig. 3.22), being highly tissue-specific for the central nervous system (CNS), the proteolipid protein (PLP) [203] (Fig. 3.22), the myelin oligodendrocyte glycoprotein (MOG) [98] (Fig. 3.23), tissue-specific for the central nervous system (CNS) and the oligodendrocyte basic protein (MOBP) [179] (Fig. 3.23).

This leads us to the assumption that all CNS tissue-specific TRAs might be potentially important for the development of multiple sclerosis (MS). A complete list of these can be seen in Table 3.4. Already known autoantibod-

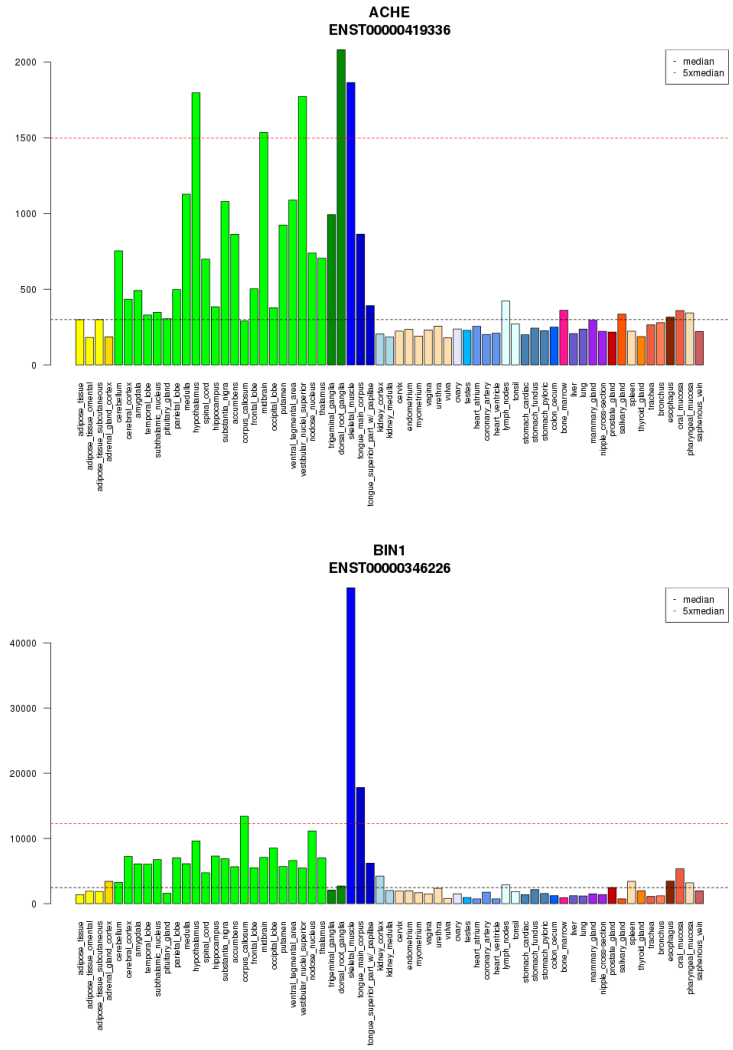


Figure 3.20: **ACHE and BIN1 involved in autoimmune myasthenia gravis.** The acetylcholinesterase (ACHE) is tissue-specific for the skeletal muscle and the central nervous system (CNS). ACHE is involved in the autoimmune myasthenia gravis [295]. BIN1 is also tissue-specific for skeletal muscle as well as the central nervous system (CNS) and also known to be involved in autoimmune myasthenia gravis [227].

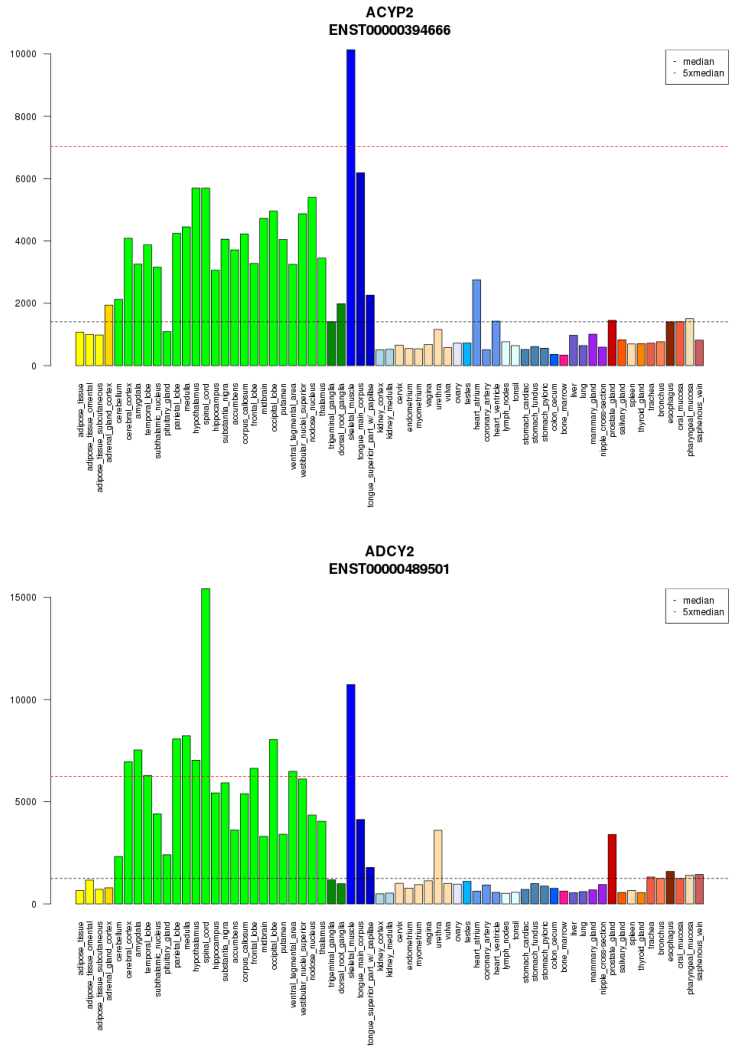


Figure 3.21: **ACYP2 and ACDY2 involved in autoimmune myasthenia gravis.** ACYP2 as well as ACDY2 are tissue-specific for the skeletal muscle as well as the central nervous system (CNS). Both are known genes [198, 354] but have not been known yet to be involved in autoimmune myasthenia gravis. More studies upon patients and diagnostic immune chips will highlight this finding.

ies important in multiple sclerosis are ADEM, AQP4, NMOSD [177], some of them are in our TRA list (Table 3.4). TRAs tissue-specific for the central nervous system might also be important to other brain related dieseases such als Alzheimers diesease.

gene name	gene description	associated with	autoimmune	reference
AACS	Acetoacetyl-CoA Synthetase	corpus callosum	-	Maccaferri et al. 2000 [242]
ABCA2	corpus callosum	-	-	Zhou et al. 2002 [457]
ACSBG1	spinal cord	-	-	-
AGPAT4	nodose nucleus	-	-	-
AGXT2L1	amygdala	-	-	McQuillin et al. 20017 [254]
AHCYL1	spinal cord	-	-	-
AK5	amygdala, hippocampus	-	-	Ansoleaga et al. 2015 [17]
AKAP11	amygdala	-	-	Blotta et al. 2009 [34]
ALDOC	**cerebellum**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
AMPH	occipital lobe	-	-	Anderson et al. 2006 [15]
ANKRD43	putamen	-	ganglion	Tucker et al. 2008 [385]
ANKS1B	**accumbens, corpus callosum**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
ANLN	**corpus callosum**	-	**down regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
ANO3	putamen	-	-	Yoo et al. 2018 [448]
AP3B2	amygdala	-	-	-
APBA2	parietal lobe	-	-	-
APBB1	parietal lobe	-	-	-
APLP1	nodose nucleus, spinalcord	-	human brain diseases	Preciados et al. 2016 [305]
AQP4	**amygdala**	-	**multiple sclerosis**	**Berger et al. 2017**
	-		**upregulated by aire**	**Pinto et al. 2008**
ARF3	parietal lobe	-	-	-
ARHGAP21	corpus callosum, nodose nucleus	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
ARHGAP22	nodose nucleus	-	-	-
ARHGEF26-AS1	spinal cord	-	-	-
ARHGEF4	parietal lobe, whole brain	-	-	-
ARNT2	**parietal lobe, whole brain**	-	multiple sclerosis,**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
ARPP21	accumbens	-	-	-
ASPHD1	putamen	-	-	-
ASTN1	amygdala, occipital lobe	-	-	-
ATCAY	amygdala	-	-	-
ATP1A2	prefrontal cortex	-	-	-
ATP1A3	amygdala	-	-	-
ATP1B2	hypothalamus	-	-	-
ATP6V1G2	**whole brain**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
ATP6V1H	amygdala	-	-	-
ATP9A	whole brain	-	-	-
B3GAT1	whole brain	-	-	-
B3GAT2	spinal cord	-	-	-
B3GNT1	parietal lobe	-	-	-
BAALC	spinal cord	-	-	-
BACE1	nodose nucleus	-	-	Mattsson et al. 2009 [252]
BAI2	amygdala, hippocampus, whole brain	-	-	-
BAI3	amygdala	-	-	-
BAIAP2	accumbens	-	-	-
BEX1	amygdala	-	-	-
BEX4	whole brain	-	-	-
BEX5	parietal lobe	-	-	-
BOK	corpus callosum	-	-	-
BSN	parietal lobe	-	multiple sclerosis	Marquez et al. 2009 [264]
BTBD3	amygdala	-	-	-
CA10	cerebellum	-	-	-
CA11	putamen, whole brain	-	-	-
CACNA1A	cerebellum	-	-	-
CACNG3	amygdala	-	-	-
CADM2	occipital lobe	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
CADPS	amygdala	-	human mTECs	Gotter et al. 2004 [144]
CALM1	occipital lobe, whole brain	-	-	Preciados et al. 2016 [305]
CALM3	parietal lobe	-	-	-
CALN1	cerebellum	-	-	-
CALY	amygdala, hypothalamus, pituitary gland, Prefrontal Cortex	-	-	-
CAMK2B	accumbens, CerebellumPeduncles, whole brain	-	-	-
CAMKV	accumbens	-	-	-
CAP2	prefrontal cortex	-	neuronal	Kumar et al. 2060 [212]
CARNS1	nodose nucleus	-	-	-
CBLN4	amygdala	-	-	-
CCDC88A	nodose nucleus	-	-	-
CCDC92	amygdala	-	-	-
CCK	**frontal lobe, whole brain**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008** [144]
CDH10	cerebellum	-	-	-
CDH18	cerebellum, nodose nucleus, spinalcord	-	-	-
CDK5	amygdala	-	-	-
CDR1	cerebellum	-	cerebellum	Totland et al. 2018 [384]
CELF4	parietal lobe	-	-	-
CERCAM	corpus callosum	-	-	-
CHD5	parietal lobe, Prefrontal Cortex	-	-	-
CHL1	fetalbrain, parietal lobe, Prefrontal Cortex	-	-	-
CHST1	amygdala	-	-	-
CLASP2	corpus callosum, occipital lobe	-	-	-
CLDN11	**corpus callosum**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
CLDND1	hypothalamus, nodose nucleus	-	-	-
CLIP2	nodose nucleus	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
CLIP3	**prefrontal cortex**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
CLSTN1	prefrontal cortex	-	multiple sclerosis	Yin et al. 2009 [447]
CNP	occipital lobe	-	-	-
CNTN2	corpus callosum	-	-	-
CNTNAP4	spinal cord	-	-	-
COL9A2	nodose nucleus	-	-	-
CPLX2	hippocampus	-	-	-
CREG2	parietal lobe	-	-	-
CRMP1	amygdala	-	-	-
CSPG5	amygdala	-	-	-
CST3	PB-CD14+Monocytes	-	-	-
CTNNA2	amygdala	-	-	-
CTNND2	spinal cord	-	-	-
CTTNBP2	corpus callosum, nodose nucleus	-	-	-
CTXN1	accumbens	-	-	-
CYFIP2	**PB-CD8+Tcells**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
CYP46A1	caudatenucleus, putamen	-	-	-
DAAM2	**corpus callosum, spinalcord**	-	**upregulated in MHC IIlo mTECs**	**Pinto et al. 2008**
DBC1	occipital lobe	-	-	-
DBNDD2	nodose nucleus	-	-	-
DCLK1	fetalbrain, occipital lobe, parietal lobe	-	-	-
DDAH1	**prefrontal cortex**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
DDN	amygdala	-	-	-
DGKZ	hippocampus	-	-	-
DIP2B	nodose nucleus	-	-	-
DIRAS2	whole brain	-	-	-
DKK3	whole brain	-	-	-
DLG2	hippocampus, occipital lobe	-	-	-
DLG4	occipital lobe	-	-	-
DNAJC6	amygdala, parietal lobe	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
DNM1	whole brain	-	-	Ou et al. 2017 [280]
DNM3	hypothalamus	-	-	-
DOCK3	occipital lobe	-	-	-
DOCK4	amygdala	-	-	-
DOK6	amygdala	-	-	-
DPF1	parietal lobe	-	-	-
DPP6	amygdala, putamen	-	-	-
DPYSL5	spinal cord	-	-	-
DTNA	spinal cord	-	-	-
EDIL3	corpus callosum, nodose nucleus	-	-	-
EDNRB	spinal cord	-	-	-
EHD3	whole brain	-	-	-
ELAVL3	cerebellum	-	-	-
ELMO1	occipital lobe	-	-	-
ELOVL1	nodose nucleus	-	-	-
ENC1	parietal lobe, whole brain	-	-	-
ENHO	amygdala, thalamus	-	-	-
ENO2	cerebellum	-	-	-
ENPP2	**hypothalamus, nodose nucleus, corpus callosum**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
ERMN	nodose nucleus	-	-	-
FA2H	nodose nucleus	-	-	-
FABP7	**fetalbrain**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
FAM107A	prefrontal cortex	-	-	-
FAM120B	putamen	-	-	-
FAM123A	corpus callosum	-	-	-
FAM125B	nodose nucleus	-	-	-
FAM131B	occipital lobe	-	-	-
FAM13C	nodose nucleus	-	-	-
FAM169A	occipital lobe	-	-	-
FAM171B	amygdala	-	-	-
FAM19A2	occipital lobe	-	-	-
FAM5B	amygdala	-	-	-
FAM5C	amygdala	-	-	-
FEZ1	spinalcord	-	-	-
FGF13	**amygdala**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
FMNL2	**corpus callosum**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
FOXG1	**fetalbrain**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
FRMD5	corpus callosum	-	-	-
FXYD6	parietal lobe, whole brain	-	-	-
FXYD7	whole brain	-	-	-
GABBR1	amygdala, cerebellum	-	human mTECs	Gotter et al. 2004 [144]
GABBR2	parietal lobe, whole brain	-	-	-
GABRA1	occipital lobe	-	-	-
GABRA2	occipital lobe	-	-	-
GABRA5	accumbens, amygdala	-	human mTECs	Gotter et al. 2004 [144]
GABRB1	amygdala	-	-	-
GABRB2	occipital lobe	-	-	-
GABRB3	**amygdala, temporal lobe**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
GABRG1	vestibular nuclei superior	-	-	-
GABRG2	occipital lobe	-	-	-
GAD1	**hypothalamus**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012** [144]
GALNTL2	spinal cord	-	-	-
GAP43	whole brain	-	-	-
GDF1	spinal cord	-	-	-
GFAP	spinal cord	-	-	-
GLT25D2	nodose nucleus	-	-	-
GNAI1	corpus callosum	-	-	-
GNAL	putamen	-	-	-
GNAO1	amygdala	-	-	-
GNG3	whole brain	-	-	-
GP1BB	accumbens	-	-	-
GPC5	amygdala	-	-	-
GPM6A	amygdala, temporal lobe	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
GPM6B	**spinal cord**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
GPR37	whole brain	-	human mTECs	Gotter et al. 2004 [144]
GPRC5B	corpus callosum, whole brain	-	-	-
GRIA1	**hippocampus**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
GRIA2	hippocampus, prefrontal cortex	-	-	-
GRIA3	hippocampus, occipital lobe	-	-	-
GRIA4	parietal lobe	-	-	-
GRIN2A	occipital lobe	-	-	-
GRM3	putamen	-	-	-
HAPLN2	nodose nucleus	-	-	-
HEPACAM	nodose nucleus	-	-	-
HHIP	nodose nucleus	-	-	-
HIPK2	corpus callosum	-	-	-
HPCA	putamen	-	-	-
HPCAL4	amygdala	-	-	-
INA	midbrain	-	-	-
INPP5F	**hypothalamus**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
IPO13	nodose nucleus	-	-	-
IQCJ-SCHIP1	corpus callosum	-	-	-
ITM2C	spinal cord	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
JPH3	hippocampus	-	-	-
JPH4	putamen	-	-	-
KALRN	parietal lobe	-	-	-
KBTBD11	occipital lobe	-	-	-
KCNC1	cerebellum	-	-	-
KCNH8	corpus callosum	-	-	-
KCNJ10	nodose nucleus	-	-	-
KIAA0284	whole brain	-	-	-
KIAA1107	parietal lobe	-	-	-
KIAA1598	hypothalamus, nodose nucleus	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
KIF1B	**prefrontal cortex**	-	**upregulated in MHC IIlo mTECs**	**Pinto et al. 2008**
KIF5C	prefrontal cortex	-	-	-
KLC1	whole brain	-	-	-
KLHL2	amygdala	-	-	-
KLHL32	corpus callosum	-	-	-
KLK6	**spinalcord**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
KNDC1	amygdala	-	-	-
LAMP2	**corpus callosum**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
LANCL1	nodose nucleus	-	-	-
LARP6	nodose nucleus	-	-	-
LGI3	nodose nucleus	-	-	-
LHPP	nodose nucleus	-	-	-
LHX2	amygdala, temporal lobe	-	-	-
LINC00320	corpus callosum	-	-	-
LINC00323	spinal cord	-	-	-
LINGO1	parietal lobe	-	-	-
LMO3	prefrontal cortex	-	-	-
LPAR1	**corpus callosum**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
LRRC3B	amygdala	-	-	-
LRRC4C	amygdala	-	-	-
LY6H	amygdala	-	-	-
LZTR1	occipital lobe	-	-	-
MAG	nodose nucleus, spinalcord	-	-	-
MAN2A1	nodose nucleus	-	-	-
MAP1A	cerebral cortex	-	-	-
MAP2	parietal lobe	-	-	-
MAP4K4	corpus callosum	-	-	-
MAP6D1	nodose nucleus	-	-	-
MAPK9	parietal lobe	-	-	-
MARCKSL1	**fetalbrain**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
MAST3	parietal lobe	-	-	-
MBP	**myelin basic protein**	-	**multiple sclerosis**	**Jensen et al. 1994**

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
	-		**upregulated by aire**	**Pinto et al. 2008**
MC1R	fetalbrain	-	-	-
MEGF10	corpus callosum	-	-	-
METRN	spinal cord	-	-	-
MIAT	occipital lobe	-	-	-
MICAL2	cerebellum	-	-	-
MKL2	Prefrontal Cortex	-	-	-
MLC1	spinal cord	-	-	-
MLLT11	fetalbrain	-	-	-
MMD	**amygdala**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
MOBP	**CNS, spinalcord**	-	**multiple sclerosis**	-
	-		**upregulated by aire**	**Pinto et al. 2008**
MOG	**CNS, spinalcord upper, spinalcord lower, substantia nigra**	-	**multiple sclerosis**	**Derbinski et al. 2001** [98]
	-		**upregulated by aire**	**Pinto et al. 2008**
MT3	whole brain	-	-	-
MTUS1	nodose nucleus	-	-	-
MYT1L	fetalbrain, occipital lobe	-	-	-
NACC2	**spinalcord**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
NAP1L2	amygdala	-	-	-
NAP1L3	amygdala	-	-	-
NAT8L	occipital lobe, parietal lobe	-	-	-
NCAM1	amygdala, corpus callosum	-	-	-
NCAN	amygdala, fetalbrain	-	-	-
NCDN	accumbens	-	-	-
NDFIP1	amygdala	-	-	-
NDRG2	cerebellum	-	-	-
NDUFA5	**amygdala**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
NECAB1	parietal lobe	-	-	-
NEFL	whole brain	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
NELF	colorectal adenocarcinoma, hippocampus	-	-	-
NFASC	prefrontal cortex	-	-	-
NGFRAP1	whole brain	-	-	-
NINJ2	corpus callosum	-	-	***
NLRP1	pituitary	-	-	-
NNAT	**pituitary gland**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
NPAS3	spinal cord	-	-	-
NPTN	whole brain	-	-	-
NPTX1	amygdala	-	-	-
NPTXR	amygdala, hippocampus	-	-	-
NR2E1	amygdala	-	-	-
NRCAM	parietal lobe, temporal lobe	-	-	-
NRGN	parietal lobe	-	-	-
NRN1	whole brain	-	-	-
NRSN1	occipital lobe, parietal lobe	-	-	-
NRXN2	cerebellum	-	-	-
NSF	hypothalamus	-	-	-
NTM	occipital lobe	-	-	-
NTRK2	**occipital lobe, prefrontal cortex**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
OLFM1	whole brain	-	-	-
OLIG1	**nodose nucleus**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
OLIG2	corpus callosum	-	-	-
OMG	Hypothalamus, nodose nucleus	-	-	-
OPALIN	corpus callosum	-	-	-
OPCML	parietal lobe	-	-	-
OSBPL1A	occipital lobe	-	-	-
P2RX7	nodose nucleus	-	-	-
P2RY12	subthalamic nucleus	-	-	-
PADI2	spinalcord	-	human mTECs	Gotter et al. 2004 [144]
PAK3	amygdala	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
PALM	**amygdala**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
PAQR6	Hypothalamus	-	-	-
PAQR8	nodose nucleus	-	-	-
PAX6	cerebellum	-	-	-
PCDH10	accumbens	-	-	-
PCDH11X	accumbens	-	-	-
PCDH11Y	accumbens	-	-	-
PCDH8	amygdala	-	human mTECs	Gotter et al. 2004 [144]
PCDHGA3	amygdala	-	-	-
PCDHGC4	amygdala	-	-	-
PCDHGC5	amygdala	-	-	-
PCLO	cerebellum	-	-	-
PCMT1	whole brain	-	-	-
PDE2A	**amygdala**	-	**down regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
PDXP	parietal lobe	-	-	-
PDZD4	parietal lobe	-	-	-
PEA15	spinal cord, whole brain	-	-	-
PEX5L	nodose nucleus	-	-	-
PFN2	whole brain	-	-	-
PGBD5	hippocampus	-	-	-
PGM2L1	amygdala	-	-	-
PHACTR3	corpus callosum	-	-	-
PHLPP1	corpus callosum	-	-	-
PHYHIP	putamen, Whole Brain	-	-	-
PIP4K2A	nodose nucleus	-	-	-
PJA1	whole brain	-	-	-
PJA2	amygdala	-	-	-
PKP4	nodose nucleus, spinalcord	-	-	-
PLCL1	corpus callosum	-	-	-
PLD3	amygdala, pituitary	-	-	-
PLEKHB1	**spinalcord**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
PLP1	**CNS**	-	**multiple sclerosis (MS)**	**Klein et al. 2000** [203]
PLXNB3	corpus callosum	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
PMP2	olfactory bulb	-	-	-
PNMA1	whole brain	-	-	-
PNMA2	frontal lobe	-	-	-
PNMAL1	amygdala, parietal lobe	-	-	-
PON2	amygdala, fetal-lung	-	-	-
POU3F2	spinal cord	-	-	-
PPP2R2C	**occipital lobe, parietal lobe**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
PPP3CA	caudate nucleus, Prefrontal Cortex	-	-	-
PREX1	nodose nucleus	-	-	-
PRKACB	occipital lobe	-	-	-
PRKCZ	whole brain	-	-	-
PRNP	**amygdala**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
PRR18	corpus callosum	-	-	-
PRRG1	corpus callosum	-	-	-
PRTFDC1	corpus callosum	-	-	-
PSAT1	corpus callosum	-	-	-
PSD2	**hypothalamus**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
PSD3	prefrontal cortex	-	-	-
PSEN1	nodose nucleus	-	-	-
PSRC1	corpus callosum	-	-	-
PTPN5	putamen	-	-	-
PTPRD	corpus callosum	-	-	-
PTPRN	amygdala, pituitary gland	-	-	-
PTPRR	cerebellum	-	-	-
PTPRZ1	amygdala	-	-	-
PXK	nodose nucleus	-	-	-
QKI	corpus callosum	-	-	-
RAB33A	corpus callosum	-	-	-
RAB3A	parietal lobe	-	-	-
RAB40B	accumbens, amygdala	-	-	-
RAPGEF4	amygdala	-	-	-
RAPGEF5	nodose nucleus	-	-	-
RASGRF2	parietal lobe	-	-	-
RASGRP3	nodose nucleus	-	-	-
RASL10A	amygdala	-	-	-
RASSF2	nodose nucleus	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
RBFOX3	cerebellum	-	-	-
RFPL1-AS1	occipital lobe	-	-	-
RGS20	accumbens	-	-	-
RGS7	occipital lobe	-	-	-
RHOU	corpus callosum	-	-	-
RIMBP2	**amygdala**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
RNF11	amygdala	-	-	-
RNF13	nodose nucleus	-	-	-
RNF182	thalamus	-	-	-
ROGDI	nodose nucleus	-	-	-
RTKN	nodose nucleus	-	-	-
RTN1	**fetalbrain**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
RUFY3	prefrontal cortex	-	-	-
RUNDC3A	parietal lobe, whole brain	-	-	-
RYR3	putamen	-	-	-
SCAMP5	parietal lobe	-	-	-
SCD5	**spinal cord**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
SCHIP1	**amygdala, corpus callosum, spinal cord**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
SCN2A	cerebellum	-	-	-
SCN3A	accumbens	-	-	-
SCN3B	amygdala, hippocampus	-	-	-
SCRN1	olfactory bulb	-	-	-
SEC14L5	nodose nucleus	-	-	-
SEMA4D	spinal cord	-	-	-
SEPT11	parietal lobe	-	-	-
SEPT3	**amygdala**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
SEPT4	corpus callosum, hypothalamus	-	-	-
SEPT5	accumbens, parietal lobe	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
SEPT8	corpus callosum, nodose nucleus, occipital lobe, spinalcord	-	-	-
SERINC1	amygdala	-	-	-
SERPINI1	**cccipital lobe, parietal lobe, prefrontal cortex**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
SGK3	corpus callosum	-	-	-
SH3BP1	parietal lobe	-	-	-
SH3GL2	Prefrontal Cortex	-	-	-
SHANK2	amygdala	-	-	-
SHC3	temporal lobe	-	-	-
SKAP2	spinal cord	-	-	-
SLAIN1	nodose nucleus	-	-	-
SLC12A5	occipital lobe, whole brain	-	-	-
SLC17A7	amygdala, hippocampus	-	-	-
SLC1A2	amygdala, temporal lobe	-	-	-
SLC1A3	**amygdala, prefrontal cortex**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
SLC22A17	**parietal lobe**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
SLC39A12	temporal lobe	-	-	-
SLC44A1	corpus callosum, nodose nucleus	-	-	-
SLC5A11	nodose nucleus	-	-	-
SLC6A1	amygdala	-	-	-
SLC7A11	spinal cord	-	-	-
SLC9A6	Hypothalamus	-	-	-
SLCO1A2	corpus callosum	-	-	-
SLCO1C1	amygdala	-	-	-
SLIT1	amygdala, fetalbrain	-	-	-
SLITRK1	temporal lobe	-	-	-
SNAP25	whole brain	-	-	-
SNAP91	amygdala	-	-	-
SNCB	parietal lobe	-	-	-
SNCB	whole brain	-	-	-
SNRPN	whole brain	-	-	-
SOBP	fetalbrain	-	-	-
SORL1	nodose nucleus	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
SOX2	spinal cord	-	-	-
SPOCK3	corpus callosum, nodose nucleus	-	-	-
SRCIN1	cerebellum	-	-	-
SRP9	-	-	-	-
ST18	corpus callosum	-	-	-
STMN4	nodose nucleus, whole brain	-	-	-
STX1A	parietal lobe	-	-	-
STXBP1	whole brain	-	-	-
SULT4A1	whole brain	-	-	-
SV2A	**parietal lobe, whole brain**	-	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
SV2B	**parietal lobe, prefrontal cortex**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012** [144]
SVOP	parietal lobe	-	-	-
SYBU	fetalbrain	-	-	-
SYN1	amygdala, parietal lobe	-	-	-
SYN2	amygdala	-	-	-
SYNDIG1	spinal cord	-	-	-
SYNGR1	**parietal lobe**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012** [144]
SYNGR3	whole brain	-	-	-
SYNJ2	**nodose nucleus**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
SYNPR	accumbens	-	-	-
SYP	parietal lobe	-	-	-
SYS1-DBNDD2	nodose nucleus	-	-	-
SYT1	amygdala	-	human mTECs	Gotter et al. 2004 [144]
SYT17	hippocampus	-	-	-
TAC1	**caudatenucleus**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008** [144]
TAGLN3	**whole brain**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
TBC1D12	spinal cord	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
TBCB	nodose nucleus	-	-	-
TCEAL2	**occipital lobe, prefrontal cortex**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
THY1	whole brain	-	-	-
TJP2	**spinal cord**	-	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
TMEFF1	temporal lobe	-	-	-
TMEM132B	parietal lobe	-	-	-
TMEM144	corpus callosum	-	-	-
TMEM151B	cerebellum	-	-	-
TMEM235	corpus callosum	-	-	-
TMEM35	amygdala	-	-	-
TMEM59L	amygdala	-	-	-
TMEM74	spinal cord	-	-	-
TMOD2	parietal lobe	-	-	-
TNIK	occipital lobe	-	-	-
TP53INP2	nodose nucleus	-	-	-
TPPP	corpus callosum	-	-	-
TRAK2	nodose nucleus	-	-	-
TRIL	spinal cord	-	-	-
TRIM2	corpus callosum	-	-	-
TRIM37	prefrontal cortex	-	-	-
TRIM59	nodose nucleus	-	-	-
TRIM9	**cerebellum, occipital lobe**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
TSPAN7	occipital lobe, Whole Brain	-	-	-
TSPYL1	whole brain	-	-	-
TSPYL4	whole brain	-	-	-
TTC9B	amygdala	-	-	-
TTLL7	corpus callosum	-	-	-
TTYH1	amygdala, whole brain	-	-	-
TTYH2	corpus callosum	-	-	-
TUBB2B	**fetalbrain**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
TUBB4	amygdala	-	-	-
UNC5A	**parietal lobe**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
UNC80	cerebellum	-	-	-

Table 3.4 – TRAs tissue-specific for the central nervous system

gene name	gene description	associated with	autoimmune	reference
USP54	nodose nucleus	-	-	-
VAMP2	amygdala	-	-	-
VSNL1	OccipitalLobe	-	-	-
VSTM2A	occipital lobe	-	-	-
WASF1	fetalbrain	-	-	-
WASF3	amygdala	-	-	-
WDR17	temporal lobe	-	-	-
WDR47	amygdala	-	-	-
WIF1	amygdala	-	-	-
WSB2	whole brain	-	-	-
YWHAH	whole brain	-	-	-
ZCCHC12	**hypothalamus**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
ZCCHC24	**nodose nucleus**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
ZDHHC9	corpus callosum	-	-	-
ZEB2	nodose nucleus	-	-	-
ZNF365	occipital lobe	-	-	-
ZNF488	nodose nucleus	-	-	-
ZNF536	corpus callosum	-	-	-

Table 3.4: **TRAs tissue-specific for the central nervous system (CNS) potentially involved in multiple sclerosis (MS)** not all genes are annotated, due to time limiting reasons, Pinto et al. 2008 (internal data), Gaertner et al. 2012 (internal data)

Table 3.4 – TRAs tissue-specific for the central nervous system

3.6.5 Addison's disease

Autoimmune Addison's disease (AAD) is an autoimmune illness related to the adrenal gland. Patients with Addison's disease often suffer from a combination of fatigue, loss of weight, hyperpigmentation of the skin as well as pathological changes as well as failure of the adrenal gland [259].

Known autoantibodies related to addison's disease are CYP11A1 [118], which is tissue-specific for the adrenal gland (Fig. 3.24), MC2R [118] (Fig. 3.24), MRAP [118] (Fig. 3.25), STAR [118] (Fig. 3.25). The TRA NNT [118] in our data is tissue-specific for the skeletal muscle not the adrenal gland. The TRA CDKN1C [118] is tissue-specific for the placenta, MCM4 [118] is tissue-specific for the testis and the TRA SAMD9 [118] is tissue-specific for the spleen. The TRAs TRXR2, MIRAGE and SGLPL1 [118] can not be found

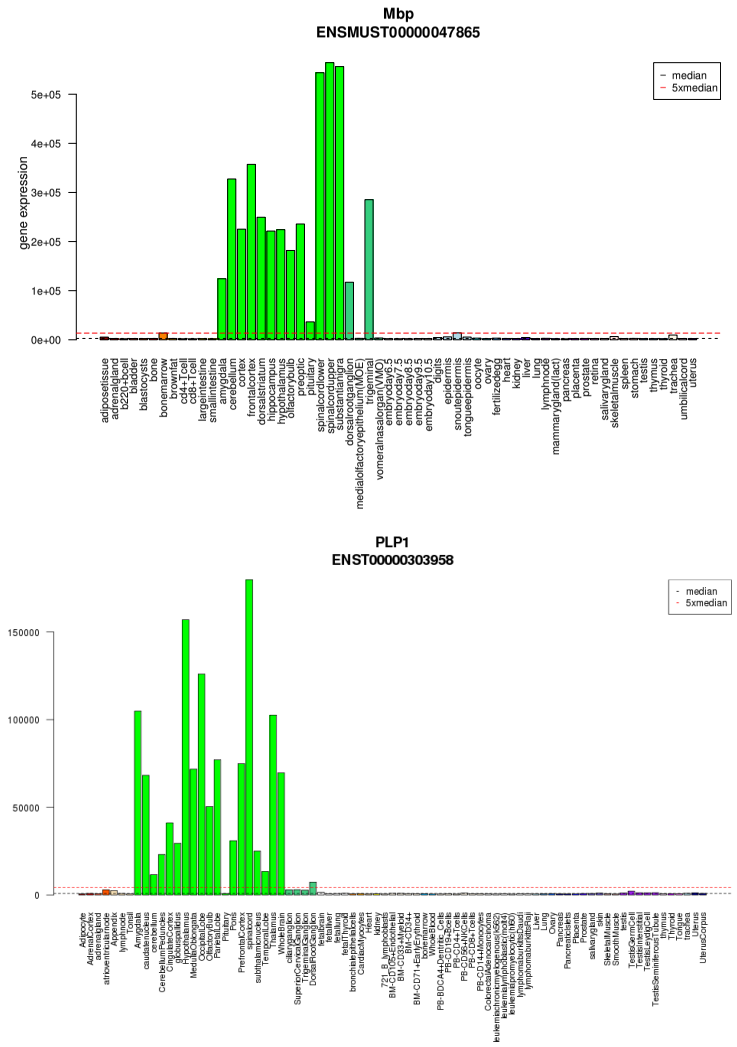


Figure 3.22: **MBP and PLP1 known to be involved in autoimmune multiple sclerosis (MS).** The myelin basic protein (MBP) has been known to be involved in autoimmune multiple sclerosis (MS). As well MBP as also the proteolipid protein (PLP) have been known to exist as autoantibodies in MS patients [203]. Both of these genes are tissue-specific for the central nervous system (CNS).

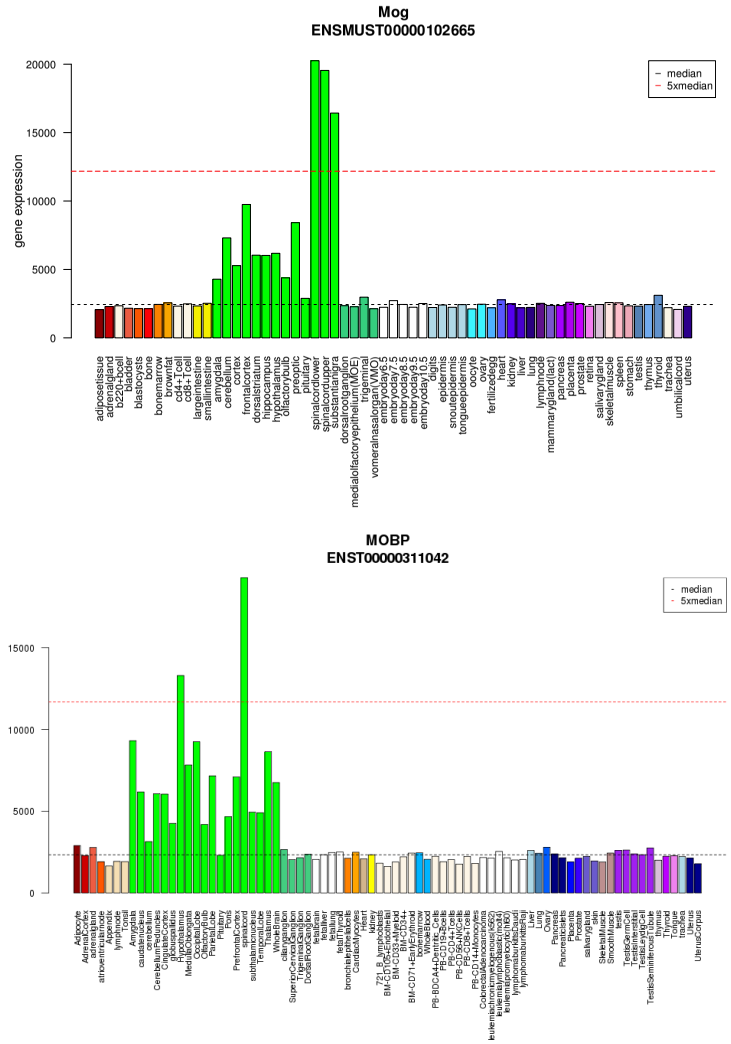


Figure 3.23: **MOG and MOBP involved in autoimmune multiple sclerosis (MS).** The myelin oligodendrocyte glycoprotein (MOG) [98], as well as the myelin associated oligodendrocyte basic protein (MOBP) are known to be involved in autoimmune multiple sclerosis (MS), both TRAs are tissue-specific for the central nervous system (CNS).

in our data (Table 3.5). Also in cancer the adrenal gland genes are important, as Pan et al. found in 2011 [283] there are several adrenal cortical genes involved in liver cancer, such as CYP11B2, CYP21A1, HSD3B1 as well as FDX1. All of these genes are tissue-specific for the adrenal gland and might be important for Addison's diesease (Table 3.5).

Also the development of the adrenal gland is dependent on different factors, which also reflect in our TRA table tissue-specific for the adrenal gland. Some of these are important for different illnesses, such as for example adrenal hypoplasia. Some of these genes seem to be TRAs and others are not. Adrenal hypoplasiea is a defect in the synthesis of adrenocorticotropin (ACTH), which is no TRA. In the secondary hypoplasia, some important factors are HESX1, LHX4, SOX3, TPIT, which are no TRAs. The same accounts true for the genes POMC, which is a TRA, but tissue-specific for the pituitary and not the adrenal gland and PC1, which is no TRA [116]. Some other TRAs can be found for example the resistance against ACTH through the gene MC2R (Fig. 3.24), the ACTHR as well as MRAP (Table 3.5) [116]. For an overview of TRAs tissue-specific for the adrenal gland, please refer to Table 3.5. The gene expression within the thymus, if known is appointed in the table.

gene name	**chromosome**	**tissue**	**autoimmune**	**reference**
ALAS1	3	adrenal gland,adrenal cortex	adrenal gland,	Okano et al. 2010 [276]
ATP1B3	3	Adrenal Cortex	-	-
ABCA1	9	adrenal gland cortex, liver	-	-
ABCC3	**17**	**adrenal glandcortex, liver**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
ACADVL	17	adrenal glandcortex, liver	-	-
AGTR1	3	adipose tissue subcutaneous,adrenal gland cortex, liver	-	-
AKR1B1	7	adrenal gland cortex, dorsal root ganglia, kidney medulla	-	-
ALAS1	3	adrenal gland cortex, liver	-	-
ALDH3A2	17	adrenal gland cortex, kidney cortex	-	-
ALDH3A2**	17	adrenal gland cortex, kidney cortex	-	-

Table 3.5 – TRAs tissue-specific for the adrenal gland

gene name	chromosome	tissue	autoimmune	reference
ARHGEF40	14	adrenal gland cortex, spleen	-	-
AS3MT	10	adrenal gland cortex	-	-
ATP1B3	3	adrenal gland cortex	-	-
C2CD2	21	adrenal gland	-	-
C4A**	-	**adrenal gland, liver**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
C4B	-	adrenal gland, liver	-	-
C4B 2	-	adrenal gland, liver	-	-
C4B**	-	adrenal gland, liver	-	-
CYP11A1**	15	adrenal gland, adrenal cortex, placenta	-	-
CYP11B1	8	adrenal gland, adrenal cortex	-	-
CYP17A1	10	adrenal gland, adrenal cortex	-	-
CYP21A2	-	adrenal gland, adrenal cortex	-	-
C10orf11	10	adrenal gland cortex	-	-
C10orf32-ASMT	10	adrenal gland cortex	-	-
C2CD2	21	adrenal gland cortex, spleen	-	-
C7	5	adrenal gland cortex, lymph nodes	-	-
CCDC69	5	adrenal gland cortex, heart atrium, heart ventricle, lymph nodes, skeletal muscle, spleen	-	-
COLEC11***	2	adrenal gland cortex, kidney cortex, liver, ovary	-	-
CPB1	3	adrenal gland cortex	-	-
CREM	10	adrenal gland cortex	-	Zwermann et al. 2007 [466]
CSDC2	22	adrenal gland cortex, heart atrium	-	-
CTNNAL1	9	adrenal gland cortex, dorsal root ganglia, heart atrium, ovary, thyroid gland, trigeminal ganglia	-	-
CYB561A3	11	adrenal gland cortex	-	-
CYB5B	16	adrenal gland cortex	-	Pan et al. 2011 [283]

Table 3.5 – TRAs tissue-specific for the adrenal gland

gene name	chromosome	tissue	autoimmune	reference
CYP11A1**	15	adrenal gland cortex, ovary	addison's disease	Flueck et al. 2017 [118]
CYP11B1	8	adrenal gland cortex, lymph nodes	-	-
CYP11B2	8	adrenal gland cortex	-	Pan et al. 2011 [283]
CYP17A1	10	adrenal gland cortex, kidney cortex, lymph nodes, ovary	-	-
CYP21A2	-	adrenal gland cortex	-	Pan et al. 2011 [283]
CYYR1	21	adrenal gland cortex, spleen	-	-
DDR2	1	adipose tissue omental, adrenal gland cortex, urethra	-	-
DFNB31	9	adrenal gland cortex	-	-
DHCR24	**1**	**adrenal gland cortex, liver, spinal cord**	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
DHCR7	11	adrenal gland cortex	-	-
DLG5	10	adrenal gland cortex	-	-
DNAJC12**	10	adrenal gland cortex, cerebellum	-	-
EBP	X	adrenal gland cortex, liver	-	-
EPB41L1	20	adrenal gland cortex	-	-
ERN1	17	adrenal gland cortex	-	-
FDX1**	11	adrenal gland, adrenal cortex, placenta	-	-
FDXR	17	adrenal gland	-	-
FADS1	11	adrenal gland cortex	-	-
FAM114A1	4	adrenal gland cortex	-	-
FAM150B	2	adrenal gland cortex, myometrium, ovary, testes	-	-
FAM166B	9	adrenal gland cortex	-	-
FBN2	**5**	**adrenal gland cortex, testes**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
FDX1	11	adrenal gland cortex	-	Pan et al. 2011 [283]
FDXR	17	adrenal gland cortex	-	-
GPR98	5	adrenal gland cortex	-	-

Table 3.5 – TRAs tissue-specific for the adrenal gland

gene name	chromosome	tissue	autoimmune	reference
GRAMD1B	11	adrenal gland cortex, dorsal root ganglia, cerebellum, trigeminal ganglia	-	-
GSTA3	6	adrenal gland cortex	-	-
GSTA4	**6**	**adrenal gland cortex, vulva**	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
HSD3B2	1	adrenal gland, adrenal cortex	-	-
HDHD3	9	adrenal gland cortex	-	-
HOXA5	7	adipose tissue omental, adrenal gland cortex, kidney cortex, kidney medulla, stomach fundus	-	-
HSD3B2	1	adrenal gland cortex, lymph nodes, ovary	-	-
HSPE1	2	adrenal gland cortex	-	-
IKBKAP	9	adrenal gland cortex	-	-
ING2	4	adrenal gland cortex	-	-
IDH1***	2	adrenal gland, adipocyte, prostate	-	-
KCNK3	2	adrenal gland cortex	-	-
KIAA1024	15	adrenal gland cortex	-	-
KCNK3	2	adrenal gland	-	-
KLHDC8B	3	adrenal gland cortex	-	-
LDLR	**19**	**adrenal gland cortex, lung, ovary**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
LONP1	19	adrenal gland cortex	-	-
LONRF2	2	amygdala, putamen, accumbens, adrenal gland cortex, cerebellum, cerebral cortex, frontal lobe, parietal lobe, occipital lobe, temporal lobe	-	-
LRRN3	7	amygdala, accumbens, adrenal gland cortex, cerebral cortex, heart atrium, frontal lobe, parietal lobe, occipital lobe	-	-
MAP3K5	6	adrenal gland cortex	-	-
MC2R	-	adrenal gland	disorder of adrenal development	Ferraz-de-Souza et al. 2008 [116]

Table 3.5 – TRAs tissue-specific for the adrenal gland

gene name	chromosome	tissue	autoimmune	reference
MCFD2	2	adrenal gland cortex, salivary gland	-	-
MCOLN3	1	adrenal gland cortex	-	-
MGARP	4	adipose tissue omental, adrenal gland cortex, ovary	-	-
MRAP	21	adipose tissue omental, adipose tissue subcutaneous, adrenal gland cortex	addison's disease	Flueck et al. 2017 [118]
MRPL33	2	adrenal gland cortex	-	-
MSI2	**17**	**adrenal gland cortex, ovary**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
MT2A	16	adrenal gland cortex	-	-
MYO7A	11	adrenal gland cortex	-	-
NOV	8	adrenalgland, adrenal cortex	-	-
NR4A1	12	adrenal cortex	-	-
NR4A2	2	adrenal cortex	-	-
NOV* **	**8**	**adrenal gland cortex, coronary artery, saphenous vein, trigeminal ganglia, urethra**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
NPC1	**18**	**adrenal gland cortex, corpus callosum**	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
NXPH1	7	amygdala, adrenal gland cortex, cerebral cortex, hypothalamus, parietal lobe, occipital lobe, temporal lobe, vestibular nuclei superior	-	-
PEBP1	12	adrenalgland	-	-
PAPSS2	**10**	**adrenal gland cortex, lung, liver, ovary**	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
PDGFD	11	adrenal gland cortex, ovary	-	-
PEBP1	12	adrenal gland cortex, liver, kidney medulla	-	-

Table 3.5 – TRAs tissue-specific for the adrenal gland

gene name	chromosome	tissue	autoimmune	reference
PEG10	7	amygdala, putamen, accumbens, adrenal gland cortex, cerebellum, hypothalamus, midbrain, medulla, ovary, pituitary gland, substantia nigra, testes, ventral tegmental area, vestibular nuclei superior	-	-
PHACTR2	6	adrenal gland cortex	-	-
QPCT	2	bone marrow, adrenal gland cortex	-	-
RARRES2*	**7***	**adipose tissue omental, adipose tissue subcutaneous, adrenal gland cortex, lung, liver, ovary**	**upregulated in MHC II hi mTECs**	**Pinto et al. 2008**
RBBP7	**X**	**adrenal gland cortex**	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
RIMS2	8	adrenal gland cortex, parietal lobe, occipital lobe	-	-
RIMS2**	8	adrenal gland cortex, parietal lobe, occipital lobe	-	-
RMDN2	2	adrenal gland cortex	-	-
RORA	15	adrenal gland cortex, nipple cross-section	-	-
RUNDC3B	7	putamen, accumbens, adrenal gland cortex, cerebellum, midbrain, occipital lobe	-	-
RHOB***	2	adrenalgland, placenta, Lung	-	-
SCARB1**	12	adrenalgland, adrenal cortex, placenta	-	-
SLC47A1	17	adrenal cortex	-	-
SORBS2**	4	adrenalgland, thyroid	-	-
STAR**	8	adrenal gland, adrenal cortex, ovary	addison's disease	Flueck et al. 2017 [118]
SCAP	3	adrenal gland cortex	-	-

Table 3.5 – TRAs tissue-specific for the adrenal gland

gene name	chromosome	tissue	autoimmune	reference
SCARB1	12	adrenal gland cortex, liver, ovary	-	-
SEMA3B	-	adrenal gland cortex, trigeminal ganglia	human mTECs	Gotter et al. 2004 [144]
SEMA6A	**5**	**adrenal gland cortex**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
SH3BP5	3	adrenal gland cortex	-	-
SIAH2	3	adrenal gland cortex	-	-
SLC23A2	20	adrenal gland cortex	-	-
SLC47A1	17	adrenal gland cortex, endometrium, kidney cortex, liver, kidney medulla, trigeminal ganglia	-	-
SMIM4	3	adrenal gland cortex	-	-
SOAT1	1	adrenal gland cortex, prostate gland, salivary gland	-	-
SPTSSA	14	adrenal gland cortex	-	-
ST3GAL5	2	adrenal gland cortex	-	-
STAR	8	adrenal gland cortex, lymph nodes, ovary, testes	-	-
STK19	6	adrenal gland cortex	-	-
STK19P	-	adrenal gland cortex	-	-
TBC1D4	**13**	**adrenal gland cortex**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
TBC1D8B	X	adrenal gland cortex	-	-
TBX3	12	adrenalgland	-	-
TM7SF2	11	adrenalgland	-	-
TBX3	12	adrenal gland cortex, prostate gland, thyroid gland, urethra	-	-
TBX3****	12	adrenal gland cortex, prostate gland, thyroid gland, urethra	-	-
TM7SF2	11	adrenal gland cortex	-	-
TOB1	17	adrenal gland cortex, liver, skeletal muscle, trachea, urethra	-	-
UGCG	**9**	**adrenal gland cortex, lymph nodes**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**

Table 3.5 – TRAs tissue-specific for the adrenal gland

gene name	chromosome	tissue	autoimmune	reference
ZNF275	**X**	**adrenal gland cortex, ovary**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
ZNF331	19	adrenal gland cortex	-	-

Table 3.5: **TRAs tissue-specific for the adrenal gland potentially involved in autoimmune addison's disease**, not all genes are annotated due to time reason, ** more than one tissue, Pinto et al. 2008 (internal data), Gaertner et al. 2012 (internal data)

Table 3.5 – TRAs tissue-specific for the adrenal gland

3.6.6 Autoimmune gastritis

The autoimmune atrophic gastritis (AIG) is an autoimmune disease against parietal cells, resulting in the mucosal distruction mainly of the stomach. AIG often occurs in association with other autoimmune diseases, such as hashimoto thyroiditis, autoimmune diabetes type 1, addison's disease as well as other autoimmune diseases [321]. Autoimmune atrophic gastritis can lead to gastric cancer [225]. Autoantibodies in AIG are mostly related to be anti-parietal cell antibodies (APCA), intrinsic factor antibodies (IFA), anti-mitochondrial antibodies (AMA), as well as anti-smooth muscle antibodies (ASMA) [397]. None of these autoantibodies can be found in our TRA data. Tissue-specific genes for the gasterointestinal tract might be important in autoimmune gastritis but have not been futher studied here (TRA-DB).

3.6.7 Juvenile idiopathic arthritis

Juvenile idiopathic arthritis (JIA) is a chronical arthritis starting in childhood. SJIA has been related to the S100 gene family [195, 158], which can be found in our data. On chromosome three in the mouse there is a S100 TRA cluster including nine different S100 genes (Fig. 3.26).

gene name	chromosome	tissue	autoimmune	reference
S100a1	-	**bladder, heart, thyroid**	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**

Table 3.6 – TRAs of the S100 gene family

gene name	chromosome	tissue	autoimmune	reference
S100a2	-	**bronchial epithelial cells**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
S100a3	-	-	-	-
S100a4	-	-	-	-
S100a5	-	**spleen, pancreas**	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
S100a6	-	**bronchial epithelial cells, whole blood, lung, smooth muscle**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
S100a7	-	-	-	-
S100a7a	-	-	-	-
S100a8	-	**bone, bone marrow**	JIA	Holzinger et al. 2018 [158]
	-		**up regulated in MHC II hi mTECs**	**Pinto et al. 2008**
S100a9	-	**bone, bone marrow**	JIA	Holzinger et al. 2018 [158]
	-		**up regulated in MHC II hi mTECs**	**Pinto et al. 2008**
S100a10	-	**bronchial epithelial cells**	**up regulated in MHC II hi mTECs**	**Pinto et al. 2008**
S100a11	-	**oral mucosa, pharyngeal mucosa**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
S100a12	-	-	-	-
S100a14	-	digits	-	-
S100a14P1	-	oral mucosa	-	-
S100a16	-	**epidermis, digits, tongue epidermis, trachea**	**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
S100b	-	-	-	-
S100g	-	-	-	-
S100p	-	-	-	-

Table 3.6 – TRAs of the S100 gene family

gene name	chromosome	tissue	autoimmune	reference

Table 3.6: **S100 gene family involved in autoimmune juvenile idiopathic arthritis.** The S100 gene family is involved in autoimmune juvenile arthritis (JIA) [158]. many of them are clustered on chromosome three in the mouse. Pinto et al. 2008 (internal data), Gaertner et al. 2012 (internal data)

Table 3.6 – TRAs of the S100 gene family

3.6.8 Hashimoto thyroiditis

Hashimoto thyroiditis is an autoimmune disease against the thyroid gland leading to hypothyroiditis. Thyroid specific autoantigens have been found to be related to hashimoto thyroiditis, such as thyroglobulin (TG) [366], thyroid peroxidase (TPO)[366] (Fig. 3.27, 3.28), the TSH receptor (TSHR) [8], PTPN22 [312], FCRL3, FOXE1, ITGAM, PRICKLE1, LPP as well as TRIB2 [355]. According to our data PTPN22 as well as ITGAM are tissue-specific for the bone marrow and not the thyroid, FCRL3 is tissue-specific for the spleen, PRICKLE1 is tissue-specific for the oocyte and and LPP is tissue-specific for the bladder as well as the umbilicalcord, it hence remains unclear if they are really related to hashimoto thyroiditis. TSHR, FOXE1 and TRIB2 are however tissue-specific for the thyroid. Table 3.7 is showing all TRAs, which are tissue-restricted for the thyroid and thus might have an impact on hashimoto thryoiditis or other thyroid related illnesses. Already known autoantibodies in hashimoto thyroiditis are pointed out in the table.

gene name	gene description	associated with	autoimmune	reference
AFAP1L2	Actin Filament Associated Protein 1 Like 2	thyroid Cancers	-	Iyama et al. 2017 [173]
ATOH8	Atonal BHLH Transcription Factor 8	colorectal cancer	-	Ye et al. 2017 [445]
ATP6AP2	ATPase H+ Transporting Accessory Protein 2	granular cell tumors	-	Pareja et al. 2018 [286]
BEND7	**BEN Domain Containing 7**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**

Table 3.7 – TRAs tissue-specific for the thyroid

gene name	gene description	associated with	autoimmune	reference
C5orf28	Transmembrane Protein 267(TMEM267)	-	-	-
C7orf23	transmembrane protein 243 (TMEM243)	-	-	-
C16orf89	Chromosome 16 Open Reading Frame 89	thyroid specific	-	Afink et al. 2010 [6]
C20orf3	adipocyte plasma membrane associated protein	diabetes type 2	not autoimmune	Ma et al. 2016 [240]
CALR	Calreticulin	thyroid tissue, grave's diesease	autoimmune	Meng et al. 2017 [257]
CAV2	Caveolin 2	thyroid cancer	-	Grosse et al. 2012 [146]
CAV2	Caveolin 2	diabetes type 1	autoimmune	Bhandage et al. 2018 [29]
CCL21	**CC-chemokine ligand 21**	**thyroid cancer**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
CCL21	CC-chemokine ligand 21	central self tolerance	autoimmune	Kozai et al. 2017 [208]
CD24	**BCs,TCs**	**hashimoto (Breg cells)**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
CD24	BCs,TCs	thyroid carcinoma	-	Han et al. 2011 [150]
CITED2	Cbp/P300 Interacting Transactivator	thyroid cancer	-	Ye et al. 2018 [446]
COL9A3	Collagen Type IX Alpha 3 Chain	thyroid cancer	-	Ye et al. 2018 [446]
CRABP1	**Cellular Retinoic Acid Binding Protein 1**	**thyroid carcinoma**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
CREB3L2	CAMP Responsive Element Binding Protein 3 Like 2	thyroid cancer	-	Chang et al. 2018 [60]
CRELD2	Cysteine Rich With EGF Like Domains 2	thyroid target	-	Fernández et al. 2013 [115]

Table 3.7 – TRAs tissue-specific for the thyroid

gene name	gene description	associated with	autoimmune	reference
CSGALNACT1	Chondroitin Sulfate N-Acetylgalactosaminyltransferase 1	thyroid carcinoma	-	Schulten et al. 2015 [342]
CTBP2	C-terminal-binding protein 2	thyroid carcinoma	-	Lui et al. 2005 [236]
CTSB	Cathepsin B	thyroid carcinoma	-	Tedelind et al. 2010 [380]
DIO2	Iodothyronine Deiodinase 2	thyroid hormone	-	Park et al. 2018 [287]
DIO2	Iodothyronine Deiodinase 2	grave's disease	autoimmune	Shahida et al. 2018 [347]
DLG5	Discs Large MAGUK Scaffold Protein 5	thyroid cancer	-	Ibrahimpasic et al. 2017 [168]
EPB41L4B	Erythrocyte Membrane Protein Band 4.1 Like 4B	thyroid cancer	-	-
FAM167A	Family with sequence similarity 167	Autoimmune Thyroid Disease	autoimmune	Song et al. 2018 [363]
FAM189A2	Family With Sequence Similarity 189 Member A2	thyroid tumours	-	Wojtas et al. 2017 [429]
FOXE1	**Forkhead Box E1**	**hypothyroidism**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
GLCCI1	Glucocorticoid Induced 1	-	-	-
GLIS3	GLIS Family Zinc Finger 3	hypothyroidism	-	Rurale et al. 2018 [330]
GLIS3	GLIS Family Zinc Finger 3	type 1 diabetes	autoimmune	Wen et al. 2017 [423]
GNAS***	Guanine Nucleotide binding protein, Alpha Stimulating	hypothyroidism, 6% altered in breast cancer [406]	-	Long et al. 2018 [234]
GNAS	Guanine Nucleotide binding protein, Alpha Stimulating	thyroid cancer	-	Untch et al. 2018 [393]
GOLGA8A	Golgin A8 Family Member A	-	-	-

Table 3.7 – TRAs tissue-specific for the thyroid

gene name	gene description	associated with	autoimmune	reference
GOLGA8B	**Golgin A8 Family Member B**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
HEBP1	**Heme Binding Protein 1**	**Crohn's disease**	autoimmune	Cagliani et al. 2013 [53]
	-		**down regulated in MHC II hi mTECs**	**Pinto et al. 2008**
HHEX	Hematopoietically-expressed homeobox protein	thyroid cancer	-	Zhu et al. 2014 [463]
HIRA	Histone Cell Cycle Regulator	-	not autoimmune	Goroshi et al. 2017 [142]
HIRA	Histone Cell Cycle Regulator	grave's disease	autoimmune	Yuk et al. 2016 [451]
HSP90B1**	High Purity HSP 90kDa ß1	thyroid papillary carcinoma, drugable target *1	-	Cong et al. 2015 [80]
HSPA5	**High Purity GRP78/BIP Antigen**	**thyroid cancer**	-	Lee et al. 2018 [219]
	-	**grave's disease**	-	Meng et al. 2017 [257]
	-		**up regulated in MHC II hi mTECs**	**Pinto et al. 2008**
HSPB11	Heat Shock Protein B Member 11	multiple sclerosis	-	Gorter et al. 2018 [143]
ID3	**Inhibitor Of DNA Binding 3**	**thyroid cancer**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
ID4	**Inhibitor Of DNA Binding 4**	**thyroid cancer**	-	Amaral et al. 2018 [13]
	-		**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
INPP5J	Inositol Polyphosphate-5-Phosphatase J	ovarian cancer	-	Zhu et al. 2015 [464]

Table 3.7 – TRAs tissue-specific for the thyroid

gene name	gene description	associated with	autoimmune	reference
IQCA1	IQ Motif Containing With AAA Domain 1	-	-	-
KCNJ16	Potassium Voltage-Gated Channel Subfamily J Member 16	thyroid carcinoma	-	Liu et al. 2016 [231]
KLHL14	Kelch Like Family Member 14	thyroid gene	-	Credendino et al. 2017 [83]
LCN12	Members of the lipocalin family	-	-	-
LPCAT2	-	thyroid function	-	Porcu et al. 2013 [304]
LRIG3	Leucine Rich Repeats And Immunoglobulin Like Domains 3	prostate cancer	-	Chen et al. 2018 [68]
LRP1B	LDL Receptor Related Protein 1B	thyroid cancer	-	Gomez-Rueda et al. 2016 [139]
LRP8	Cell surface receptor for Reelin	breast cancer	-	Maire et al. 2018 [245]
METTL7A	Methyltransferase Like 7A	thyroid cancer	-	Zhou et al. 2017 [461]
MGAT4C	MGAT4 Family Member C	prostate cancer	-	Demichelis et al. 2012 [94]
MT1E	Metallothionein 1E	malignant thyroid lesions	-	Wojtczak et al. 2017 [430]
MTCH1	Mitochondrial Carrier 1	neuro-Behcet's disease	autoimmune	Vural et al. 2013 [407]
NKX2-1	NK2 Homeobox 1	thyroid autoimmunity	autoimmune	Giuliani et al. 2018 [135]
NUPR1	Nuclear Protein 1, Transcriptional Regulator	cancer	-	Chowdhury et al. 2009 [74]
OBSL1	Obscurin Like 1	3-M syndrome	-	Demir et al. 2013 [95]
PAX8	Paired-Box-Protein 8	thyroid cancer	-	Suzuki et al. 2018 [373]
PBX4	PBX Homeobox 4	acute lymphoblastic leukemia	-	Rosales-Avina et al. 2011 [324]
PDE8B	**Phosphodiesterase 8B**	**hashimoto thyroiditis**	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**

Table 3.7 – TRAs tissue-specific for the thyroid

gene name	gene description	associated with	autoimmune	reference
PDIA4	Protein Disulfide Isomerase Family A Member 4	adenocarcinoma.	-	Tufo et al. 2014 [386]
PDIA6	Protein disulfide isomerase A6	adenocarcinoma.	-	Tufo et al. 2014 [386]
PGF	Placental Growth Factor	thyroid carcinoma	-	He et al. 2015 [153]
PKHD1L1	PKHD1 Like 1	proliferative diabetic retinopathy	-	Ung et al. 2017 [392]
PLVAP	Plasmalemma Vesicle Associated Protein	thyroid tumors	-	Wojtas et al. 2017 [429]
PRKX	Protein Kinase X-Linked	hematopoietic neoplasm	-	Saloustros et al. 2015 [334]
PRSS16	Serine Protease 16	antitumoral immunity	-	Brisson et al. 2015 [47]
PTH	Parathormon	thyroid cancer	-	Vargas-Ortega et al. 2018 [398]
RCBTB1	RCC1 And BTB Domain Containing Protein 1	chronic lymphocytic leukemia	-	Parker et al. 2011 [288]
RGL3	Ral Guanine Nucleotide Dissociation Stimulator Like 3	-	-	-
SCPEP1	**Serine Carboxypeptidase 1**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
SDC2	Syndecan 2	-	-	-
SHISA2	Shisa Family Member 2	-	-	-
SLC12A8	**Solute Carrier Family 12 Member 8**	**psoriasis**	-	Cabaleiro et al. 2016
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
SLC25A29	solute carrier family 25	cancer	-	Zhang et al. 2018 [454]
SLC26A4	solute carrier SLC26A4	thyroid carcinoma	-	Poma et al. 2018 [302]
SLC26A7	Solute Carrier Family 26 Member 7	thyroid carcinoma	-	Weinberger et al. 2017 [420]

Table 3.7 – TRAs tissue-specific for the thyroid

gene name	gene description	associated with	autoimmune	reference
SNHG5	Small Nucleolar RNA Host Gene 5	acute myeloid leukemia	-	Li et al. 2018 [224]
SORBS2	Sorbin And SH3 Domain Containing 2	thyroid carcinoma	-	Stein et al. 2010 [369]
ST6GAL2	ST6 Beta-Galactoside Alpha-2,6-Sialyltransferase 2	thyroid carcinoma	-	Liang et al. 2018 [226]
TCERG1L	**Transcription Elongation Regulator 1 Like**	**Crohn's disease**	-	Bae et al. 2014 [22]
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
TG	thyroglobulin	thyroid antibodies	autoimmune	Sospedra et al. 1998 [366]
TMEM30B	Transmembrane Protein 30B	-	-	-
TPO	thyroid peroxidase	hashimoto thyroiditis	autoimmune	Sospedra et al. 1998 [366]
TPST2	Tyrosylprotein Sulfotransferase 2	hypothyroidism	-	Sasaki et al. 2007 [338]
TSHR	Thyroid Stimulating Hormone Receptor	thyroid disease	autoimmune	Patel et al. 2018 [290]
VEGFA**	**Vascular endothelial growth factor A**	**thyroid carcinoma, drugable target *1**	-	Wang et al. 2018 [414]
	-		**up regulated in MHC II hi mTECs**	**Pinto et al. 2008**
VEGFC	Vascular Endothelial Growth Factor C	thyroid cancer	-	Gao et al. 2018
ZFP36L2	ZFP36 Ring Finger Protein Like 2	-	-	-
ZMAT1	Zinc Finger Matrin-Type 1	-	-	-

Table 3.7 – TRAs tissue-specific for the thyroid

gene name	gene description	associated with	autoimmune reference

Table 3.7: **TRAs tissue specific for the thyroid**, *1 - drugable targets, ** tissue-specific for more than one tissue-type, *** expressed in more than in one tissue-type, Pinto et al. 2008 (internal data), Gaertner et al. 2012 (internal data)

Table 3.7 – TRAs tissue-specific for the thyroid

3.6.9 Testis-specific antigens

Testis-specific antigens have two major roles in immunology as well as in cancer treatment. First it has been observed as for most, if not all of the TRAs, that TRAs are in general involved in autoimmune diseases. This is also true for testis-specific antigens [144]. Autoimmune responses to testis-specific antigens have been observed mainly in male infertility [387]. Due to its immune-priviliged status, the testis is a site in which immune cells have no or just restricted access, coming from the blood-testis barrierer formed by Sertoli cells, which physically prevent immune cell infiltration into the site of sperm cell maturation [425]. Also Sertoli cells secrete TGF-*beta* (tumor growth factor-*beta*), as also activin A, granzyme B and FAS ligand in order to inhibit growth and survival of immune cells in the interstitial epithelium of the testis [425].

The immune-priviliged status of the testis, as well as the germline expression of testis-specific antigens make them such an interesting target of immunotherapy, because it has been observed in many tumor types, that especially the germline encoded testis-specific antigens are reactivated in many cancers. In 1991 Van den Bruggen et al. found the first cancer testis antigen (CTA) in melanoma patients, called MAGE [396]. And De Plaen et al. discovered in 1994 twelve genes of the MAGE family [91]. This discovery was followed by the BAGE and GAGE genes [45, 90] and many others in the following years. In 2005 Simpson et al. cite about 40 different cancer testis antigens [356]. And 2014 Whiterhurst et al. report about 225 genes coding for testis-specific antigens, which might be possible drug targets [425]. In 2017 Tung et al. state, that many of these genes are known, but their special expression pattern is still not well understood yet [387]. In 2018 Peer et al. again came up with a list of testis-specific antigens, as potentially new cancer immunology drug targets.

Since we have good access here through the TRA-DB to the real gene expression profile of CTAs as well as potential side effects due to gene expression in other tissues as well then the testis, our TRA data might actually be

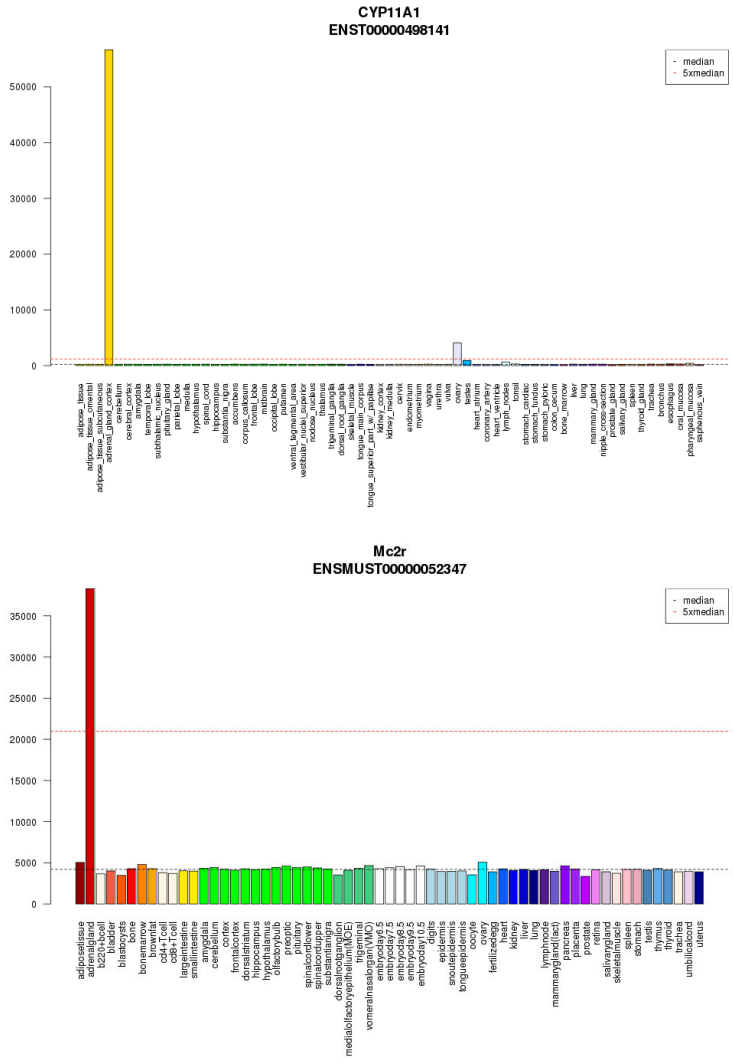


Figure 3.24: **CYP11A1 and MC2R being involved in autoimmune addison's disease.** The cytochrome P450 side chain cleavage enzyme (CYP11A) is known to be involved in the autoimmune addison's disease, it is tissue-specific for the placenta, the adrenal gland and the adrenal cortex. The melanocortin receptor 2 (MC2R) is an ACTH receptor which leads to adrenocorticotropin resistance in case of MC2R defect [116]. Also this is tissue-specific for the adrenal gland.

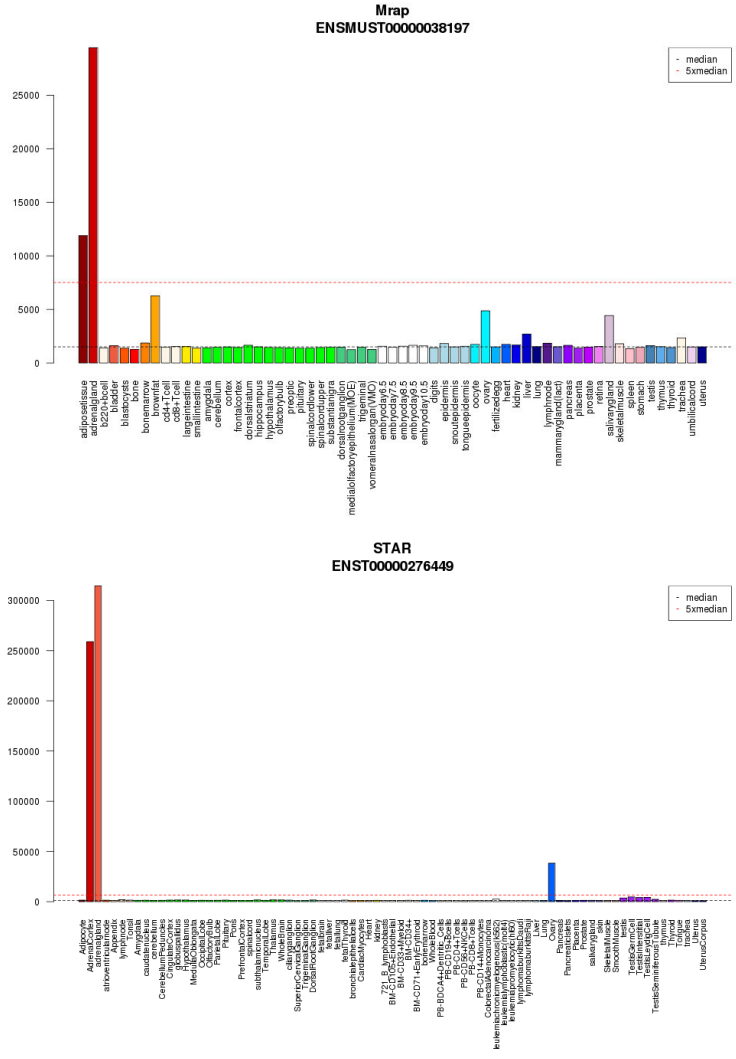


Figure 3.25: **MRAP and STAR being involved in autoimmune addison's disease.** The melanocortin receptor 2 accessory protein (MRAP) is involved in adrenocorticotropin resistance (ACTH) and can thus lead to addison's disease. The same hold true for the steroidogenic acute regulatory protein (STAR).

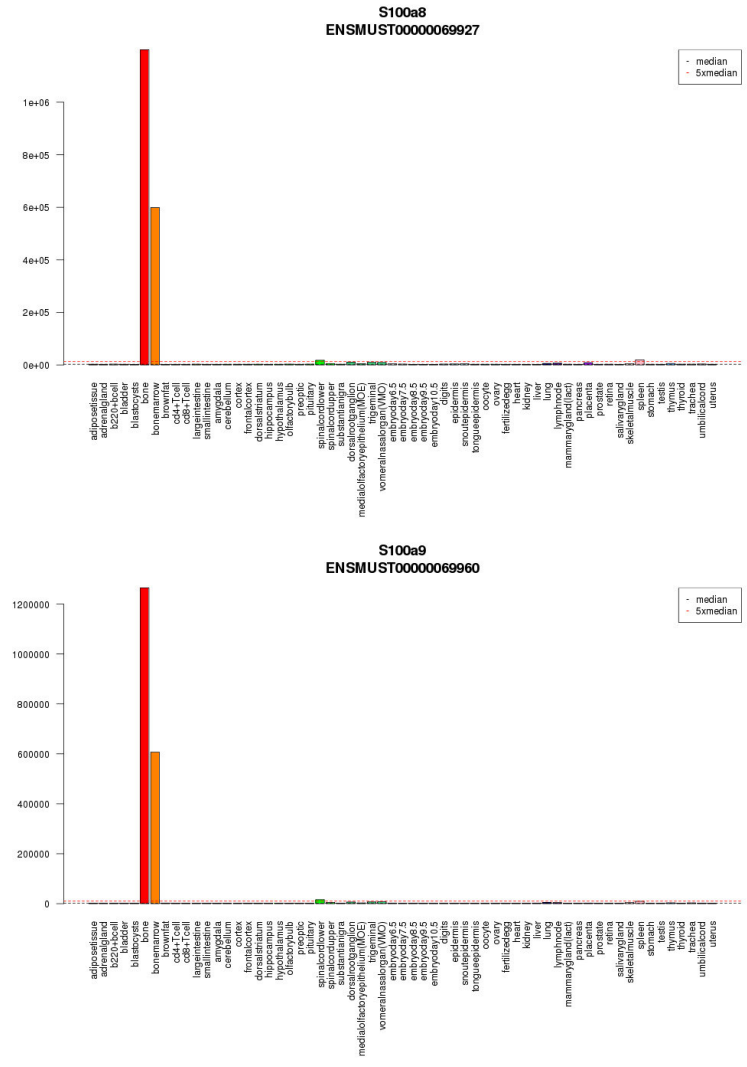


Figure 3.26: **S100a8 and S100a9 being involved in the autoimmune disease of juvenile idiopathic arthritis (SJIA)**. Both S100a8 as well as S100a9 are involved in autoimmune juvenile idiopathic arthritis (SJIA) [195, 158]. Both are tissue-specific for the bone and the bone marrow.

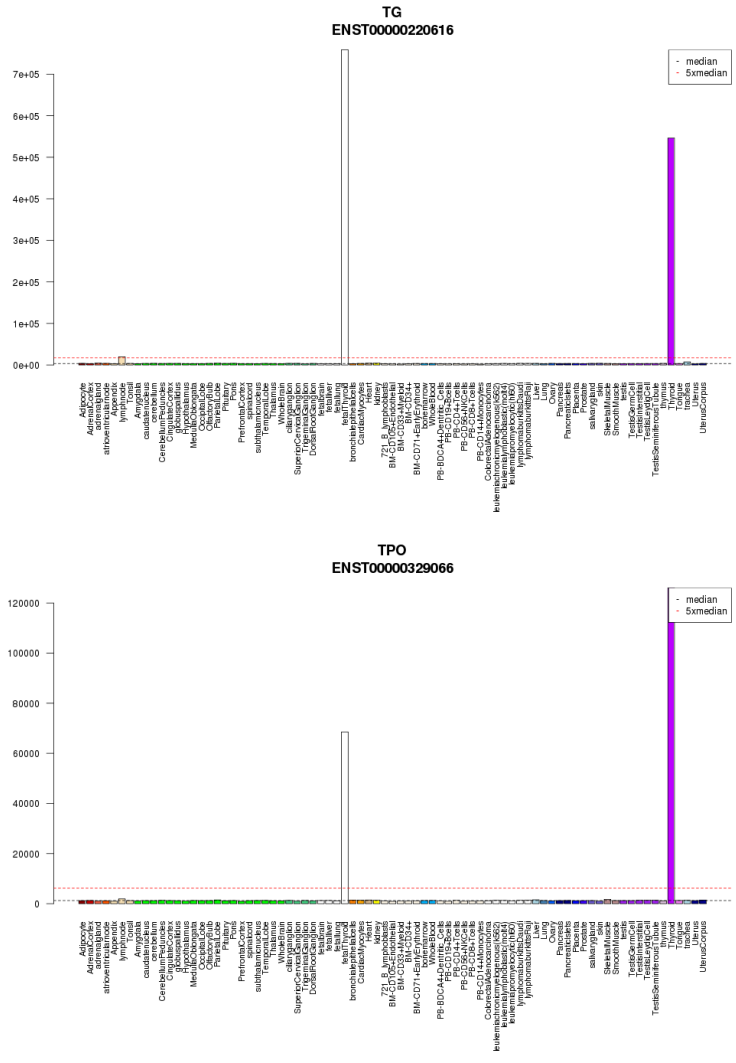


Figure 3.27: **TG and TPO involved in autoimmune diseases.** The thyroglobulin (TG) as well as the thyroid peroxidase (TPO) are involved in autoimmune hashimoto thyroiditis. Both of them are tissue-specific for the thyroid [366].

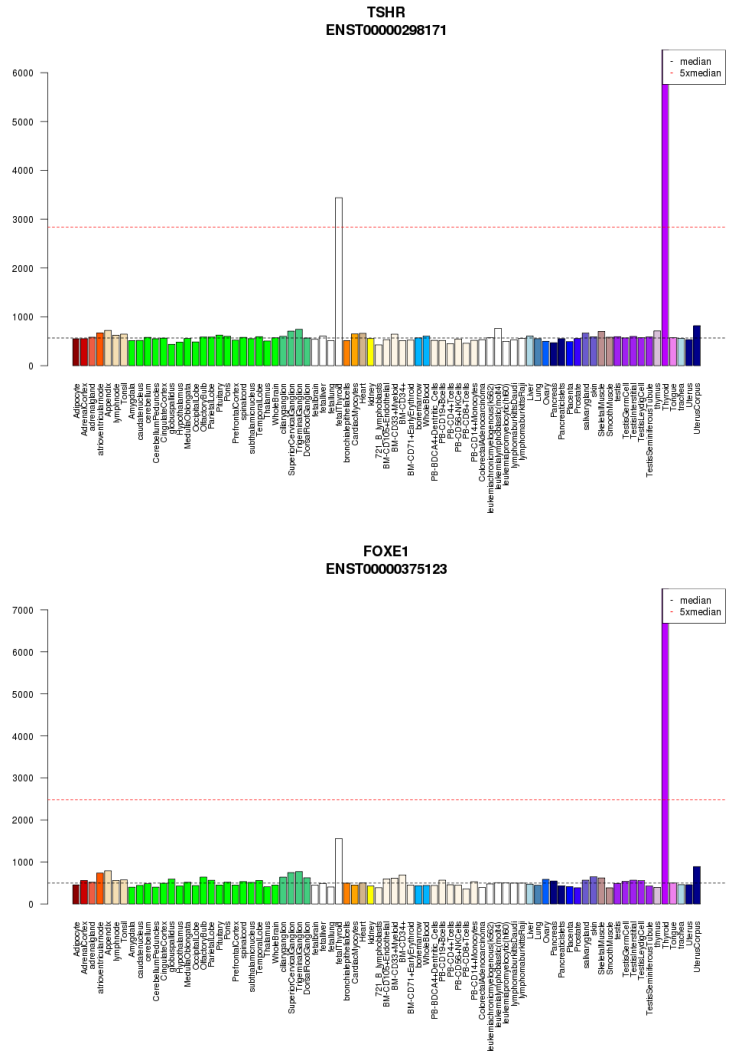


Figure 3.28: **TSHR and FOXE1 involved in autoimmune hashimoto thyroiditis.** The thyroid stimulating hormone receptor (TSHR) as well as FOXE1 are highly tissue-specific for the thyroid. Both are known to be involved in thyroid specific illnesses [290, 410] and might be involved as all other TRAs, which are tissue specific for the thyroid in autoimmune hashimoto thyroiditis.

important for the study of cancer testis antigens. For this purpose we have gone through the list of all CTAs found in our database and connect them to previous knowledge if possible. Through this approach we could identify many new cancer testis antigens, which might be important for the cancer immunotherapy. We have further used the function of our TRA-DB in order to enlarge the knowledge of gene families in the context of cancer testis antigens.

Since CTAs have been formerly subgrouped into X linked and non X linked CTAs we copied this way of representation in our Table 3.8. This way it becomes obvious that most of the cancer testis antigens, we found were not linked to the X chromosome in human. Furthermore we noticed that not all formerly predicted CTAs are really tissue-specific for the testis only, which might cause problems in immunotherapy treatment, especially man of the CTAs found in Peer et al. 2018 are highly expressed in the bone marrow as well, which might be actually very dangerous to use in clinical trials on patients. Examples of these are ASPM, DLGAP5, NCAPG2, UBE2C, AURKA, KIF4a, KIF15, KIF23, TPX2, CDC20 RACGAP1, FOXM1, PRC1 as well as HJURP (Peer et al. 2018).

In the CTA list, we represent here, we only focus on those CTAs, which have either been previously known or also only, if not differently described expressed in the testis, but no other tissue, measured in our datasets. Since we have different subgroups of testis associated tissues, we could maybe even go into further detail of this study.

Already Gotter et al. 2004 [144] pointed out that the immunogenicity of a gene is highly dependent on the central self tolerance of T cells, it might be very important for future research to keep the actual gene expression level of CTAs in the medullary part of the thymus in mind, before starting any clinical trial on tumor vaccination against CTAs.

Most of the already known CTAs can be found in our data. As can be seen in Table 3.8 we can identify twelve representatives of the GAGE family, one member of the MAGE family (MAGEb), one member of the PAGE family (PAGE4) [339], which is however not tissue-specific for the testis in our data, but instead for the placenta (TRA-DB). We could find the CTAs SSX2IP [331] in our TRA-DB, as well as IL3, but not the IL31RA1 receptor, as had been proposed by Simpson et al. 2005 [356]. Other examples of previously known CTAs we could not find are the SCP1 gene [389] as well as CSAGE or CAGE [356]. The supposively CTA E2F1 is in our data tissue-specific for the oocyte and not the testis. Sometimes we find different subtypes as CTAs, as for example in the case of NXF2 [356, 412], which we can not find, but instead detected the subtype NXF3. Sometimes also

the tissue-specificity of a gene measured in our data, varies according to the dataset used, as can be seen in the case of TAF7L [356], which is tissue-specific for the testis in the human Roth dataset, but tissue-specific for the placenta in the mouse Novartis dataset. If this is due to the different species or just the different dataset, remains unclear and would need further examination.

Summing up the results we found for cancer testis antigens, the TRA-DB seems to be good tool in order to detect new TRAs, in order to claim these as potential drug targets, its gene expression within the thymus should clearly be clearified first, before going into any clinical trial. Under www.cta.lncc.br there is a database of cancer-testis antigens [11]. For further CTAs please refer to Table 3.8 as well as Fig. 3.29 and 3.30.

gene name	**gene description**	**chromosome**	**autoimmune**	**reference**
ACE2*	Angiotensin converting enzyme 2	X - germ cell	expressed by testes leydig cells	Douglas et al. 2004 [104]
AKAP4*	A-Kinase Anchoring Protein 4)	X	testis antigen	Jagadish et al. 2016 [174]
C9orf9	Sperm Acrosome Associated 9	X	-	-
CITED1*	Glu/Asp Rich Carboxy-Terminal Domain 1	X - germ cell	gonad development	Del Valle et al. 2017 [93]
GAGE2A*	G Antigen 2A	X - k562 cells	testis-specific	Chao et al. 2018 [62]
GAGE2B*	G Antigen 2B	X	testis antigen family	Gjerstorff et al. 2008 [136]
GAGE2C*	G Antigen 2C	X	testis antigen family	Gjerstorff et al. 2008 [136]
GAGE3*	G Antigen 3	X	testis antigen family	Vodolazhsky et al. 2018 [405]
GAGE8*	G Antigen 8	X	testis antigen family	Almeida et al. [11]
GAGE10*	G Antigen 10	X	testis antigen family	Gjerstorff et al. 2008 [136]
GAGE12C*	G Antigen 12C	X	testis antigen family	Kulkarni et al. 2012 [211]
GAGE12D*	G Antigen 12D	X	testis antigen family	Gjerstorff et al. 2008 [136]
GAGE12E*	G Antigen 12E	X	testis antigen family	Gjerstorff et al. 2008 [136]
GAGE12H*	G Antigen 12H	X	testis antigen family	Gjerstorff et al. 2008 [136]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
GAGE12J*	G Antigen 12J	X	testis antigen family	Gjerstorff et al. 2008 [136]
GAGE13*	G Antigen 13	X	testis antigen family	Gjerstorff et al. 2008 [136]
GPR64	G protein-coupled receptor 64	X	spermatogenesis	Yap et al. 2011 [444]
NXF3	Nuclear RNA Export Factor 3	X	spermatogenesis	Zhou et al. 2011 [459]
TKTL1	Transketolase-like-1	X	cancer testis antigens	Djureinovic et al. 2016 [103]
ACSBG2*	Acyl-CoA Synthetase Bubblegum Family Member 2	-	testis gene	Pei et al. 2006 [293]
ACRBP	Acrosin Binding Protein	-	CT database	Whitehurst et al. 2014 [425]
ACTL6A	Actin Like 6A	3	-	-
ACTL7A*	Actin Like 7A	9	testis antigen	Afsharpad et al. 2019 [7]
ACTL7B*	Actin Like 7B	9	testicular germ cell specific	Hisano et al. 2003 [157]
ACTL9	Actin Like 9	19	-	-
AGBL5*	ATP/GTP Binding Protein Like 5	-	-	-
ADAM23***	CNS, PNS, umbilicalcord	no CTA	human mTECs	Gotter et al. 2004 [144]
ADAM2	ADAM Metallopeptidase Domain 2	altered in 7% in breast cancer [243]	CT database	Kulkarni et al. 2012 [211]
ADAM3	ADAM Metallopeptidase Domain 3	-	sperm membrane	Fujihara et al. 2018 [122]
ADAM3A	ADAM Metallopeptidase Domain 3A	altered 6% in breast cancer	cancer testis antigen	Kim et al. 2009 [197]
ADAM5	ADAM Metallopeptidase Domain 5	altered 7% in breast cancer [400]	testis gene	Cho et al. 1996 [70]
ADAM18	ADAM Metallopeptidase Domain 18	altered 6% in breast cancer [462]	testis gene	Frayne et al. 1997 [121]
ADAM29	ADAM Metallopeptidase Domain 29	altered 6% in breast cancer	CT database	Kulkarni et al. 2012 [211]
ADAM30	ADAM Metallopeptidase Domain 30	-	testis specific gene	Cerretti et al. [59]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
ADAM32	ADAM Metallopeptidase Domain 32	altered 7% in breast cancer [400]	testis gene	Choi et al. 2003 [72]
ASRGL1	Asparaginase Like 1	11	sperm autoantigen	Bush et al. 2002 [51]
ACSBG2	Acyl-CoA Synthetase Bubblegum Family Member 2	19	testis gene	Pei et al. 2006 [293]
AKAP3*	A-Kinase Anchoring Protein 3	12	cancer testis antigen	Kulkarni et al. 2012 [211]
AKAP4*	A-Kinase Anchoring Protein 4	-	testis antigen	Jagadish et al. [174]
AKAP14	A-Kinase Anchoring Protein 14	-	testis antigen	Jagadish et al. [174]
ALDH1A1***	Aldehyde Dehydrogenase 1 Family Member A1	-	testis gene	Nourashrafeddin et al. [273]
APH1B	Aph-1 Homolog B	15	breast cancer	Peltonen et al. 2013
AGBL5	ATP/GTP Binding Protein Like 5	2	breast cancer	Peltonen et al. 2013
ANKRD5	Ankyrin Repeat Domain-Containing Protein 5	-	prostate cancer	Jin et al. 2016 [180]
ANKRD7*	Ankyrin Repeat Domain 7	7	testis specific	Ozaki et al. 1996 [281]
ANKRD13A	-		cell migration	Avellino et al. 2013 [19]
ANKRD20A1	-		germ cell arrest	Catford et al. 2019 [58]
ANKRD20A2	-		lung cancer	Kanwal et al. 2018 [188]
ANKRD20A3	-		male infertility	Zhou et al. 2019 [460]
ANKRD20A4	-		male infertility	Zhou et al. 2019 [460]
ANKRD20A7	P		male infertility	Zhou et al. 2019 [460]
ANKRD20A8	P		male infertility	Zhou et al. 2019 [460]
APH1B*	Aph-1 Homolog B, Gamma-Secretase Subunit	-	breast cancer	Peltonen et al. 2013

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
APOA1**	**Apolipoprotein A1**	human mTECs	**infertility**[336]	Gotter et al. 2004 [144]
	-		**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
ARMC3	Armadillo Repeat Containing 3	-	CT database	Almeida et al. 2009 [11]
ASRGL1*	Asparaginase Like 1	-	sperm autoantigen	Bush et al. 2002 [51]
	-	overexpressed in breast cancers	sperm autoantigen	Weidle et al. 2009 [419]
ATAD2	AAA Domain Containing 2	altered 11% in breast cancer [376]	CT database	Whitehurst et al. 2014 [425]
ACTG2***	Actin, Gamma 2	-	prostate gene	Untergasser et al. 2005 [394]
BRD8*	Bromodomain Containing 8	5	colorectal cancer	Yamada et al. 2009 [441]
BRDT*	Bromodomain testis-specific protein	1	testis-specific	Kulkarni et al. 2012 [211]
BRP44***	Brain Protein 44	-	immune cells	-
BSCL2**	Seipin Lipid Droplet Biogenesis Associated	-	spermatogenesis	Ebihara et al. 2015 [107]
CABYR	Calcium Binding Tyrosine Phosphorylation Regulated	-	CT database	Kulkarni et al. 2012 [211]
CALR3**	Calreticulin 3	-	CT database	Ikawa et al. 2011 [169]
CASC1	Cancer Susceptibility 1	-	-	Sinnott et al. 2014 [357]
CASC5	cancer susceptibility candidate 5	-	CT database	Kulkarni et al. 2012 [211]
CATSPERZ*	-		testis derter-mining SRY gene	Brown et al. 2018 [48]
CCDC7	Coiled-Coil Domain Containing 7	-	testis development	Wang et al. [411]
CCDC11	Coiled-Coil Domain Containing 11	-	prostate cancer	Yamamoto et al. 2007 [442]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
CCDC15	Coiled-Coil Domain Containing 15	-	lymphoma	Yang et al. 2018 [443]
CCDC19	Coiled-Coil Domain Containing 19	-	nasopharyngeal carcinoma	Fang et al. 2012[112]
CCDC30	Coiled-Coil Domain Containing 30	-	thyroid cancer	Ritterhouse et al. 2016 [318]
CCDC34	Coiled-Coil Domain Containing 34	-	cervical cancer	Liu et al. 2018 [232]
CCDC38	Coiled-Coil Domain Containing 38	-	testis specific gene	Lin et al. 2016 [229]
CCDC42	Coiled-Coil Domain Containing 42	-	male fertility	Pasek et al. 2016 [289]
CCDC46	Coiled-Coil Domain Containing 46	-	cancer testis antigen	Xie et al. 2019 [437]
CCDC53	Coiled-Coil Domain Containing 53	-	-	-
CCDC54	Coiled-Coil Domain Containing 54	-	spermiogenesis	Bai et al. 2018 [24]
CCNA1	Cyclin A1	13	testis gene	Zhang et al. 2018 [454]
CD52***	CD52 Molecule	-immune cells	Sertoli cells, diabetes type 1	Skurikhin et al. 2017 [359]
CDC20***	Cell-division cycle protein 20	-immune cells	gonad differentiation	Groh et al. 2013 [145]
CENPH	Centromere protein H	5	testis specific	Peer et al. 2018
CENPL	Centromere Protein L	1	testis specific histone	Tachiwana et al. 2008 [375]
Cenpv	Centromere Protein V	11 mouse	testis specific histone	Tachiwana et al. 2008 [375]
CENPW	Centromere Protein W	6	-	Geister et al. 2015 [128]
CEP55	Centrosomal Protein 55	-	CT database	Kulkarni et al. 2012 [211]
CEP63	Centrosomal Protein 63	-	bladder carcinoma	Buim et al. 2005 [49]
CEP70	Centrosomal Protein 70	-	breast cancer	Sirkisoon et al. 2018 [358]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
CEP112	Centrosomal Protein 112	altered 5% in breast cancer [284]	breast cancer	Panda et al. 2018[284]
CKM* **	**Creatine Kinase, M-Type**	-	colorectal cancer	Francis et al. 2016 [120]
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
CKS2***	cyclin dependent kinases	-immune cells	spermatogenesis	Smirnova et al. 2006 [360]
C11orf20	-	11	ovarian carcinoma	Salzman et al. 2011 [335]
CYB5R2	Cytochrome B5 Reductase 2	11	prostate cancer,**mTECs**	Devaney et al. 2013 [101]
CLPB	**Caseinolytic peptidase B protein homolog**	**11**	**testis gene**	Guan et al. 2013 [147]
	-		**up regulated in MHC II lo mTECs**	**Pinto et al. 2008**
CLU***	Clusterin	-	sperm associated biomarker	Dere et al. 2017 [100]
CNN1***	Calponin 1	-	spermatozoa	López-Cardona et al. 2018 [237]
COL1A2***	Collagen Type I Alpha 2 Chain	-	testis gene	Baert et al. 2015 [23]
CABYR*	Calcium Binding Tyrosine Phosphorylation Regulated	18	testis specific	Shen et al. 2015 [350]
CCNA1*	Cyclin A1	13	testis specific	Panigrahi et al. 2012 [285]
CCT6B*	Chaperonin Containing TCP1 Subunit 6B	17	testis specific	Kubota et al. 1997 [209]
CDKN3*	Cyclin Dependent Kinase Inhibitor 3	14	testis genes	Peer et al. 2018
CEP55	-	10	testis genes	Peer et al. 2018
CETN3*	Centrin 3	5	cancer testis antigen	Kim et al. 2013 [199]
CHIC2*	Cysteine Rich Hydrophobic Domain 2	4	leukemia	Kuchenbauer et al. 2005 [210]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
CLPB*	Caseinolytic peptidase B protein homolog	-	testis gene	Guan et al. 2013 [147]
COX6B2	Cytochrome c oxidase subunit VIb polypeptide 2	-	CT database	Whitehurst et al. 2014 [425]
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
COX7A2*	Cytochrome c oxidase (COX)	6	testis gene	Chen et al. 2012 [66]
CPXCR1	CPX Chromosome Region	-	CT database	Almeida et al. [11]
CREM*	CAMP Responsive Element Modulator	-	testis	Wang et al. 2018 [414]
CRISP2*	Cysteine-rich secretory protein 2	6	cancer testis antigen	Zamuner et al. 2015 [452]
CRISP3	Cysteine-rich secretory protein 3	-	testis specific gene	Giese et al. 2002 [133]
CUL3*	Cullin 3	2	spermatozoa gene	Nguyen et al. 2009 [269]
CSDAP1**	-	immune cells	lung cancer	Xu et al. 2018 [439]
CT62	-		cancer testis antigen	Kulkarni et al. 2012 [211]
CTAG1A	-	vaccine available	CT database	Kulkarni et al. 2012 [211]
CTAG1B*	-	drugable target, *1	CT database	Whitehurst et al. 2014 [425]
CTCFL	-		CT database	Kulkarni et al. 2012 [211]
CYB5R2*	-		prostate cancer	Devaney et al. 2013 [101]
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
CYLC2*	-		sperm-specific cylicin II	Xie et al. 2001 [438]
DAZL*	Deleted In Azoospermia Like	3	testis specific gene	Hashemi et al. 2018 [152, 412]
DBF4**	-	immune cells	anti cancer target	Cheng et al. 2018 [69]
DCAF12L1	-		azoospermia	Ramasamy et al. 2014 [309]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
DDX3X	-		oncogene	He et al. 2018 [154]
	-		**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
DDX4*	DEAD-Box Helicase 4	5	germ cells	Aduma et al. 2018 [5]
DDX20	-		tumorigenesis	Chen et al. 2016
DDX25	-		testicular germ cell	Kavarthapu et al. 2015 [193]
DDX39A	-		testis gene	Soboleva et al. 2017 [361]
DDX39***	-	immune cells	cancer cells	Yuan et al. 2014 [450]
DDX43	-		CT database	Kulkarni et al. 2012 [211]
DDX53	-		testis gene	Liggins et al. 2010 [228]
DNALI1	**Dynein Axonemal Light Intermediate Chain 1**	**1**	flagellar protein	Rashid et al. 2006 [310]
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
DKKL1*	Dickkopf Like Acrosomal Protein 1	19	cancer testis antigen	Tarnowski et al. 2016 [378]
DNAJB3	-		testis gene	Berruti et al. 2002 [28]
DNAJB6	-		**male reproduction**	Meccariello et al. 2008 [256]
	-		**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
DNAJB7	-		-	-
DNAJB8	-		CT database	Almeida et al. [11]
DNALI1*	-		testis gene	Rashid et al. 2006 [310]
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
DMRT1	-		CT database	Almeida et al. [11]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
DMRT1B1	-		testis gene	Ruan et al. 2019 [329]
DMRTC2	-		cell cycle	Odajima et al. 2016 [275]
DRG1*	-		-	Ishikawa et al. 2003 [172]
DSCR8	-		cancer testis gene	Risinger et al. 2007 [316]
ELP5*	Elongator Acetyltransferase Complex Subunit 5	17	melanoma	Close et al. 2012 [76]
FAM24A	-		male fertility	Niu et al. 2019 [272]
FAM46C*	**family with sequence similarity 46, member C**	1-immune cells	lung cancer	Xia et al. 2018 [436]
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
FAM47B	-		testis gene	Ruan et al. 2019 [329]
FAM48A	-		cervical cancer	Lando et al. 2009 [217]
FATE1	-		CT database	Almeida et al. [11]
FBXO15	-		**germline stem cell**	Okita et al. 2007 [277]
	-		**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
FBXO24	-		gastric cancer	Petrovchich et al. 2016 [296]
FBXO25	-		cancer	Jiang et al. 2016 [178]
FMR1NB	-		CT database	Almeida et al. [11]
FNDC8*	-		testis gene	Ruan et al. 2019 [329]
FNDC11*	Fibronectin Type III Domain Containing 11	20	testis gene	Ruan et al. 2019 [329]
FXR1*	FMR1 Autosomal Homolog 1	3-immune cells	mouse testis	Huot et al. 2001 [166]
GAPDHS*	-		sperm function	Huang et al. 2017 [163]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
GGA1*	-		-	-
GK2*	Glycerol Kinase 2	4	-	Zuo et al. 2016 [465]
GSTM3*	**Glutathione S-Transferase Mu 3**[87]	**1**	human mTECs	Gotter et al. 2004 [144]
	-		**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
GTF2A1L	General Transcription Factor IIA Subunit 1 Like	2	-	Huang et al. 2006 [162]
GYG1	Glycogenin 1	3	-	Morimoto et al. 2017
GAPDHS	Glyceraldehyde-3-Phosphate Dehydrogenase	19	-	Huang et al. 2017 [163]
GNAS***	-	drugable target, *1	altered 5% in breast cancer	Bhattacharya et al. 2019 [31]
GPR64*	-		-	Yap et al. 2011 [444]
GPX4*	Phospholipid-Hydroxyperoxid-Glutathion-Peroxidase	19	-	Guerriero et al. 2014 [148]
GSG1*	Germ Cell Associated 1	12	-	Zheng et al. 2018 [456]
GSTA***	-		-	Paul et al. 2009 [291]
GTF2A2*	General Transcription Factor IIA Subunit 2	15	-	Han et al. 2001 [151]
GTF2A1L*	-		-	Huang et al. 2006 [162]
GYG1*	-		-	Marimoto et al. 2017
HN1*	-	-immune cells	-	Zhou et al. 2004 [458]
HRASLS*	HRAS Like Suppressor	3	-	Shyu et al. 2013 [353]
HSPA1L	-		-	Rogon et al. 2014 [322]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
HSPA2***	-		-	Samanta et al. 2018 [336]
	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
HSPA2L	-		-	Ruan et al. 2019 [329]
HSPA4L*	Heat Shock Protein Family A (Hsp70) Member 4 Like	4	testis specific	Liu et al. 2016 [231]
HSPB9	-		CT database	Almeida et al. [11]
HN1	Hematological and neurological expressed 1 protein	17	-	-
HORMAD1	-	altered 8% in breast cancer [414]	CT database	Shin et al. 2010 [351]
HORMAD2	-		CT database	Almeida et al. 2009 [11]
IL13RA2	-		CT database	Kulkarni et al. 2012 [211]
IFT57*	Intraflagellar Transport 57	3-immune cells	-	-
INSL3*	Insulin-like 3	19	-	-
IPO5*	Importin 5	13-immune cells	-	-
ISYNA1*	Inositol-3-Phosphate Synthase 1	19 - germ cell	-	-
IZUMO1 IZUMO2 IZUMO4*	IZUMO Family Member 4	19	-	-
KHDRBS3*	KH RNA Binding Domain Containing	8, altered 10% in breast cancer [251]	-	-
KDM5B*	-		-	Kulkarni et al. 2012 [211]
KIAA0895	-		-	-
KIAA0100	-		CT database	Kulkarni et al. 2012 [211]
KIAA1210	-		-	-
KIAA1257	-		-	-
KIF2A	-	5	-	-
KIF2B	-	17	-	-
KIF2C	-	1	testis genes	Peer et al. 2018

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
KPNA2**	-	-immune cells, altered 5% in breast cancer [241]	-	-
LAPTM4A***	-		-	-
LEMD1	-		CT database	Almeida et al. [11]
LDHC*	-		CT database	Kulkarni et al. 2012 [211]
LPIN1*	Lipin 1	2	-	-
LDHC	L-lactate dehydrogenase C	11	testis-specific	Simpson et al. 2005 [356]
LY6K	-	altered in 10% in breast cancer [206]	CT database	Whitehurst et al. 2014 [425]
LYAR	-		-	-
MAEL	-		CT database	Soper et al. 2008 [364]
MAGEA1	no TRA	-	human mTECs	Gotter et al. 2004 [144]
MAGEA3	no TRA	-	human mTECs	Gotter et al. 2004 [144]
MAGEA4	no TRA	-	human mTECs	Gotter et al. 2004 [144]
MAGEB2	-		-	-
MART	no TRA	-	human mTECs	Gotter et al. 2004 [144]
MEA1*	Male-Enhanced Antigen 1	6	-	-
MLF1*	Myeloid Leukemia Factor 1	3	-	-
MLF1IP**	-	-immune cells	-	-
MLLT10	-	10	-	-
MORC2-AS1	-		-	-
MORC2B	-		-	-
MRGBP*	MRG Domain Binding Protein	-	-	-
MYL2* **	-		**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008** [144]
MYLK***	-		-	-
MYL6B*	Myosin Light Chain 6B	12 - germ cell	-	-
NEK2	-	1	testis genes	Peer et al. 2018

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
NDUFAF3*	NADH:Ubiquinone Oxidoreductase Complex Assembly Factor 3	3	-	-
NME5*	**NME/NM23 Family Member 5**	**5**	**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
NPTX2***	-		-	-
NUF2*	-		-	Whitehurst et al. 2014 [425]
NUP155*	Nucleoporin 155	5	-	-
NCRNA00081*		10	-	-
NXF3	-	-immune cells	-	-
NY-ESO-1	no TRA	-	human mTECs	Gotter et al. 2004 [144]
OAZ3*	Ornithine Decarboxylase Antizyme 3	1, altered 8% in breast cancer	-	-
OBSL1*	Obscurin Like 1	2	-	-
ODF1*	Outer Dense Fiber Of Sperm Tails 1	8, altered 10% in breast cancer [20]	-	Kulkarni et al. 2012 [211]
ODF2*	Outer Dense Fiber Of Sperm Tails 2	9	-	Kulkarni et al. 2012 [211]
ODF3	Outer Dense Fiber Of Sperm Tails 3	7	-	Almeida et al. [11]
ODF3L1	Outer Dense Fiber Of Sperm Tails 3 L1	15	-	Almeida et al. [11]
Odf4	Outer Dense Fiber Of Sperm Tails 3 L1	11 mouse	testis antigen	Afsharpad et al. 2019 [7]
OIP5	-	15	testis antigen	Afsharpad et al. 2019 [7]
PAOX*	-		-	-
PASD1	-		-	-
PBK**	-	-immune cells	CT database	Kulkarni et al. 2012 [211]
PDHA2*	Pyruvate Dehydrogenase E1 Alpha 2 Subunit	4	-	-
PDGDS***	-		-	-
PENK**	-		-	-
PIWIL1	-		-	Whitehurst et al. 2014 [425]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
PIWIL2	-		CT database	Kulkarni et al. 2012 [211, 412]
PGK2*	Phosphoglycerate Kinase 2	6	-	-
PHF7*	PHD Finger Protein 7	3	-	-
PLAC1L	-		-	Whitehurst et al. 2014 [425]
PRKAA1*	Protein Kinase AMP-Activated Catalytic Subunit Alpha 1	5	-	-
PAFAH1B1*	Platelet Activating Factor Acetylhydrolase 1b Regulatory Subunit 1	17	-	-
PDXK*	Pyridoxal Kinase	21	-	-
PIAS2*	Protein Inhibitor Of Activated STAT 2	18	-	-
PIWIL1	-	12	-	-
PIWIL2	-	8	testis antigen	Afsharpad et al. 2019 [7]
PRAME*	-	22 - vaccine available	(PRAME1) testis-specific antigen, drugable target *1	Kulkarni et al. 2012 [211]
PRM1*	**Protamine 1**	**16**	CT database	Kulkarni et al. 2012 [211]
	-		**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
PRM2*	Protamine 2	16	CT database	Kulkarni et al. 2012 [211]
PRM3	Protamine 3	16	CT database	Almeida et al. [11]
PRND*	Prion Like Protein Doppel	20 - germ cell	-	-
PRPS2**	-		**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
PRSS21*	Serine Protease 21	16	-	-
PRSS37	-		male infertility	Liu et al. 2016 [231]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
PSMG1**	-	-immune cells	-	-
PTTG1**	-		**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
PAOX	Polyamine Oxidase	10	-	-
RACGAP1P**	-	-immune cells	-	-
REC114	Meiotic Recombination Protein REC114-Like	-	CT database	Almeida et al. [11]
REEP1*	Receptor Accessory Protein 1	2	-	-
ROPN1B*	Rhophilin Associated Tail Protein 1B	3	CT database	Whitehurst et al. 2014 [425]
RGS22	-	altered 10% in breast cancer [303]	CT database	Almeida et al. [11]
RPL39L*	Ribosomal Protein L39 Like	3	-	-
RNF114*	Ring Finger Protein 114	20	-	-
RNASE1***	-	- germ cell	human mTECs	Gotter et al. 2004 [144]
RUVBL2*	RuvB Like AAA ATPase 2	19	-	-
RBM46	-		CT database	Almeida et al. [11]
RBMXL2*	RBMX Like 2	11	-	-
SCRG1**	Stimulator Of Chondrogenesis 1	-	-	-
SERPINA5**	-	- germ cell	human mTECs	Gotter et al. 2004 [144]
SERPINF1***	-	-	-	-
SLC6A16*	Solute Carrier Family 6 Member 16	19	testis specific	Nishimura et al. 2008 [271]
SLC9B1	Solute Carrier Family 9 Member B1	-	-	-
SLC12A6	Solute Carrier Family 12 Member 6	-	-	-
SLC25A31*	Solute Carrier Family 25 Member 31	4	-	-

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
SLC25A33	Solute Carrier Family 25 Member 33	-	-	-
SLC26A8	Solute Carrier Family 26 Member 8	-	-	-
SLC39A13	Solute Carrier Family 39 Member 13	-	-	-
SLC39A3	Solute Carrier Family 39 Member 3	-	-	-
SLCO6A1*	Solute Carrier Organic Anion Transporter Family Member 6A1	-	CT database	Almeida et al. [11]
SLCO6C1	solute carrier organic anion transporter family, member 6c1	-	-	-
SLCO6D1	solute carrier organic anion transporter family, member 6d1	-	-	-
SAMD4A*	Sterile alpha motifs (SAMs)	14	-	-
SIK3*	SIK Family Kinase 3	11	-	-
SMCP*	Sperm Mitochondria Associated Cysteine Rich Protein	1	-	-
SPINLW1*	Serine protease inhibitor-like protein	20	-	-
SPACA1*	Sperm Acrosome Associated 1	-	-	-
SPACA3*	Sperm Acrosome Associated 3	-	CT database	Almeida et al. [11]
SPACA7*	Sperm Acrosome Associated 7	-	-	-
C9orf9*	Aliases for SPACA9	-	-	-
SPAG4*	Sperm Associated Antigen 4	-	CT database	Kulkarni et al. 2012 [211]

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
SPAG5**	Sperm Associated Antigen 5	-	-	-
SPAG6*	Sperm Associated Antigen 6	CT database [11]	human mTECs	Gotter et al. 2004 [144]
SPAG11A*	Sperm Associated Antigen 11A	8	-	-
SPAG11B*	Sperm Associated Antigen 11B	8	-	-
SPAG16*	Sperm Associated Antigen 16	-	-	-
SPA17*	Sperm Autoantigenic Protein 17	11	testis-specific	Kulkarni et al. 2012 [211]
SPA17P1*	Sperm Autoantigenic Protein 17 Pseudogene 1	-	-	-
SPATA1*	Spermatogenesis Associated 1	-	fertility	Giesecke et al. 2009 [134]
SPATA3*	Spermatogenesis Associated 3	-	spermatogenesis	-
SPATA4*	Spermatogenesis Associated 4	-	spermatogenesis	-
SPATA6*	Spermatogenesis Associated 6	-	spermatogenesis	-
SPATA7*	Spermatogenesis Associated 7	-	spermatogenesis	-
SPATA8*	Spermatogenesis Associated 8	-	spermatogenesis	-
SPATA9*	Spermatogenesis Associated 9	-	spermatogenesis	-
SPATA12*	Spermatogenesis Associated 12	-	spermatogenesis	-
SPATA16*	Spermatogenesis Associated 16	-	spermatogenesis	-
SPATA17*	Spermatogenesis Associated 17	-	spermatogenesis	-
SPATA18*	Spermatogenesis Associated 18	-	spermatogenesis	-
SPINLW1*	-	- germ cell	CT database	Kulkarni et al. 2012 [211]
SPINK2*	**Serine Peptidase Inhibitor, Kazal Type 2**	**4**	**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008**

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
SSX2IP*	SSX Family Member 2 Interacting Protein	-	-	-
STAG3	Stromal Antigen 3	7-immune cells	-	-
STMN1***	Stathmin 1	-	-	-
STON1-GTF2A1L*	read-through products of the neighboring STON1 and GTF2A1L genes	2	-	-
SYCE1*	Synaptonemal Complex Central Element Protein 1	-	CT database	Bolcun-Filas et al. 2009 [38]
SYCE2*	Synaptonemal Complex Central Element Protein 2	-	-	-
SYCE3*	Synaptonemal Complex Central Element Protein 3	-	-	-
SYCP1*	Synaptonemal Complex Protein 1	-	CT database	Kulkarni et al. 2012 [211, 412]
SYCP2*	Synaptonemal Complex Protein 2	-	-	Wang et al. 2001 [412]
SYCP3*	Synaptonemal Complex Protein 3	-	-	Wang et al. 2001 [412]
SYPL1*	Synaptophysin Like 1	7	testis specific antigen	Vodolazhsky et al. 2018 [405]
TAF7L*	TATA-Box Binding Protein Associated Factor 7 Like	-	CT database	Kulkarni et al. 2012 [211]
TAF9*	**TATA-Box Binding Protein Associated Factor 9**	-	**up regulated in MHC IIlo mTECs**	**Pinto et al. 2008**
TAF10*	TATA-Box Binding Protein Associated Factor 10	-	-	-
TBPL1*	TATA-Box Binding Protein Like 1	6	-	-
TBL2*	Transducin Beta Like 2	7	-	-

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
TCFL5*	Transcription Factor Like 5	20	-	-
TCP11*	T-Complex 11	6	-	-
TDRD5*	Tudor Domain Containing 5	-	-	-
TDRD6*	Tudor Domain Containing 6	-	-	Vasileva et al. 2009 [399]
TDRD7**	Tudor Domain Containing 7	-	-	-
TDRD9**	Tudor Domain Containing 9	-	-	-
TDRD10*	**Tudor Domain Containing 5**	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
TEKT1*	Tektin 1	-	-	-
TEKT2*	Tektin 2	-	-	-
TEKT3*	Tektin 3	-	-	-
TEKT4*	Tektin 4	-	-	-
TEKT5*	Tektin 5	-	CT database	Almeida et al. [11]
TEX9*	Testis Expressed 9	-	-	-
TEX14*	Testis Expressed 14	altered 5% in breast cancer [190]	CT database	Kulkarni et al. 2012 [211, 412]
TEX19*	Testis Expressed 19	-	-	Wang et al. 2001 [412]
TEX22*	Testis Expressed 22	-	-	-
TEX30*	Testis Expressed 30	13	-	-
TEX101*	Testis Expressed 101	-	-	-
TFDP2*	Transcription Factor Dp-2	-	-	-
TKTL1*	Transketolase-like-1	-	-	Wang et al. 2001 [412]
TMEM31*	Transmembrane Protein 31	-	cancer associated protein	Kamata et al. 2013 [186]
TMEM38B*	Transmembrane Protein 38B	-	-	-
TMEM53*	Transmembrane Protein 53	-	-	-
TMEM56*	Transmembrane Protein 56	-	-	-
TMEM97*	Transmembrane Protein 97	-	-	-

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
TMEM99*	Transmembrane Protein 99	-	-	-
TMEM108*	Transmembrane Protein 108	-	CT database	Almeida et al. [11]
TMEM116**	Transmembrane Protein 116	-	-	-
TMEM120a**	Transmembrane Protein 120a	-	-	-
TMEM146*	Transmembrane Protein 146	-	-	-
TMEM151B*	Transmembrane Protein 151B	-	-	-
TMEM190**	Transmembrane Protein 190	-	-	-
TMEM191a*	Transmembrane Protein 191a	-	-	-
TMEM191c*	Transmembrane Protein 191c	-	-	-
TMEM210*	Transmembrane Protein 210	-	-	-
TMPRSS12*	Transmembrane Serine Protease 12	CT database [11]	human mTECs	Gotter et al. 2004 [144]
TNP1*	Transition Protein 1	2	-	-
TP53RK*	TP53 Regulating Kinase	-	-	-
TP53RG5*	-		-	-
TPD52L3*	Tumor Protein D52 Like 3	-	-	-
TPP2*	Tripeptidyl Peptidase 2	-	-	-
TPPP2*	Tubulin Polymerization Promoting Protein Family Member 2	-	CT database	Almeida et al. [11]
TPRKB*	TP53RK Binding Protein	-	-	-
TPRN*	Taperin	-	-	-
TPTE*	Transmembrane Phosphatase With Tensin Homology	21, drugable target *1	-	Kulkarni et al. 2012 [211]
TPTE2P6*	Transmembrane Phosphoinositide 3-Phosphatase And Tensin Homolog 2 Pseudogene 6	-	-	-

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
TPX2**	TPX2, Microtubule Nucleation Factor	-	-	-
TRIM13*	Tripartite Motif Containing 13	13	-	-
TRIP12	Thyroid Hormone Receptor Interactor 12	2	-	-
TRIP13	Thyroid Hormone Receptor Interactor 13	5	testis genes	Peer et al. 2018
TSGA8*	testis specific gene A8	-	-	-
TSGA10*	testis specific gene A10	-	testis antigen	Afsharpad et al. 2019 [7]
TSGA10ip*	Testis Specific 10 Interacting Protein	-	-	-
TSSK1*	Testis Specific Serine Kinase 1	-	-	-
TSSK1B*	Testis Specific Serine Kinase 1B	-	-	-
TSSK2*	Testis Specific Serine Kinase 2	22	-	-
TSSK3*	Testis Specific Serine Kinase 3	-	-	-
TSSK4*	Testis Specific Serine Kinase 4	-	-	-
TSSK6*	Testis Specific Serine Kinase 6	-	CT database	Almeida et al. [11]
TSKS*	Testis-specific serine kinase	-	-	-
TSNAXIP1*	Translin Associated Factor X Interacting Protein 1	-	-	-
TSPAN16*	Tetraspanin 16	-	-	-
TSPYL5*	TSPY Like 5	altered 9% in breast cancer [376]	-	-
TSPYL6*	TSPY Like 6	-	-	-
TTC7A*	Tetratricopeptide repeat domain 7A	-		-
TTC12*	Tetratricopeptide repeat domain 12	-	-	-
TTC15*	Tetratricopeptide repeat domain 15	-	-	-

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
TTC16*	Tetratricopeptide repeat domain 16	-	-	-
TTC18*	Tetratricopeptide repeat domain 18	-	-	-
TTC21A*	Tetratricopeptide repeat domain 21 A	-	-	-
TTC23L*	Tetratricopeptide repeat domain 23 Like	-	-	-
TTC29*	Tetratricopeptide Repeat Domain 29	-	-	-
TTC39A**	Tetratricopeptide Repeat Domain 39A	-	-	-
TTK*	TTK Protein Kinase	-	CT database	Whitehurst et al. 2014 [425]
TTLL2*	Tubulin Tyrosine Ligase Like 2	-	-	-
TTLL4*	Tubulin Tyrosine Ligase Like 4	-	-	-
TTLL5*	Tubulin Tyrosine Ligase Like 5	-	-	-
TTLL6*	Tubulin Tyrosine Ligase Like 6	-	-	-
TTLL10*	Tubulin Tyrosine Ligase Like 10	-	-	-
TTLL13*	Tubulin Tyrosine Ligase Like 13	-	-	-
TUBA3A*	tubulin, alpha 3A	-	-	Wang et al. 2001 [412]
TUBA3C*	Tubulin Alpha 3C	13	-	-
TUBB2C**	Tubulin Beta 2C	-immune cells	-	-
TUBB4B**	Tubulin Beta 4B	-	-	-
TULP2*	tubby-like genes (TULPs)	-	CT database	Kulkarni et al. 2012 [211]
TUBG1**	Tubulin Gamma 1	-immune cells	-	-
UBQLN3*	Ubiquilin 3	11	-	-
UTRN*	Utrophin	6	-	-
XRCC6BP1*	ATP23 Metallopeptidase And ATP Synthase Assembly Factor Homolog	-	-	-
YBX2*	Y-Box Binding Protein 2	17	-	-

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference
YPEL1*	Yippee Like 1	22	-	-
ZPBP*	zona pellucida binding protein	7	-	-
ZMYND10*	Zinc Finger MYND-Type Containing 10	3	-	-
ZNF57*	Zinc Finger Protein 57	-	-	-
ZNF165*	Zinc Finger Protein 165	6	CT database	Kulkarni et al. 2012 [211]
ZNF200*	Zinc Finger Protein 200	-	-	-
ZNF217*	Zinc Finger Protein 217	altered 5% in breast cancer [27]	-	-
ZNF233*	Zinc Finger Protein 233	-	-	-
ZNF280B*	Zinc Finger Protein 280 B	-	-	-
ZNF280C*	Zinc Finger Protein 280 C	-	-	-
ZNF295AS1*	Zinc Finger Protein 295 AS1	-	-	-
ZNF367*	Zinc Finger Protein 367	-	-	-
ZNF467**	Zinc Finger Protein 467	-	**upregulated in mTECs versus cTECs**	**Gaertner et al. 2012**
ZNF541*	Zinc Finger Protein 541	-	-	-
ZNF546*	Zinc Finger Protein 546	-	-	-
ZNF610*	Zinc Finger Protein 610	-	-	-
ZNF677*	Zinc Finger Protein 677	-	-	-
ZNF683*	Zinc Finger Protein 683	-	-	-
ZNF689*	Zinc Finger Protein 689	-	-	-
ZNF829*	Zinc Finger Protein 829	-	-	-

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

gene name	gene description	chromosome	autoimmune	reference

Table 3.8: **testis-specific TRAs**, * tissue-restricted only to one tissue, ** tissue-restricted to two tissues, *** tissue-restricted to more than two tissues, *1 drugable target, not all of the genes here were annotated with literature. Pinto et al. 2008 (internal data), Gaertner et al. 2012 (internal data)

Table 3.8 – TRAs tissue-specific for the testis (CTAs)

3.6.10 Grave's disease

Grave's disease is an autoimmune disease, which is connected similar to hashimoto thyroiditis to autoimmune problems with the thyroid. Grave's disease is comming along with Grave's ophthalmopathy, goitre as well as hyperthyroidism, caused by a stimulation of the thryoid initiated through an interaction of the anti-TSHR [26]. Other autoantibodies known to be involved in Grave's diesease are DIO2 (Fig. 3.31) as well as for example BTG2 [300]. For more thyroid specific TRAs please refer to Table 3.7.

3.6.11 Systemic lupus erythematosus

Lupus erythematosus is an autoimmune disease with many different locations, which can not be clearly defined and connected to one specific tissue. In this case a systematic approach for finding new drug targets or potential autoantibodies is not so easy. In most patients lupus erythematosus is mainly diagnosed by the detection of antinuclear antibodies (ANA) assays [270]. Some autoantibodies however involved in lupus can be found in our TRA data, among them the example of NEK7 [241] as well as IFI44 [86] (Fig. 3.32).

3.7 Summary of finding TRAs

To sumarize the definition and finding of tissue-restricted antigens (TRAs) with the five times median gene expression method, we can find many already known TRAs involved in autoimmune diseases and we can also detect all genes which are expressed as non TRAs. Those are either housekeeping genes, such as genes, which are not higher expressed in any of the tissues than 3x the median gene expression and unspecific genes, which are higher expressed than 5x the median gene expression in more than five tissues. These genes are considered to be unspecific. Thus the method of finding tissue-restricted antigens in different datasets has been proven to be successful.

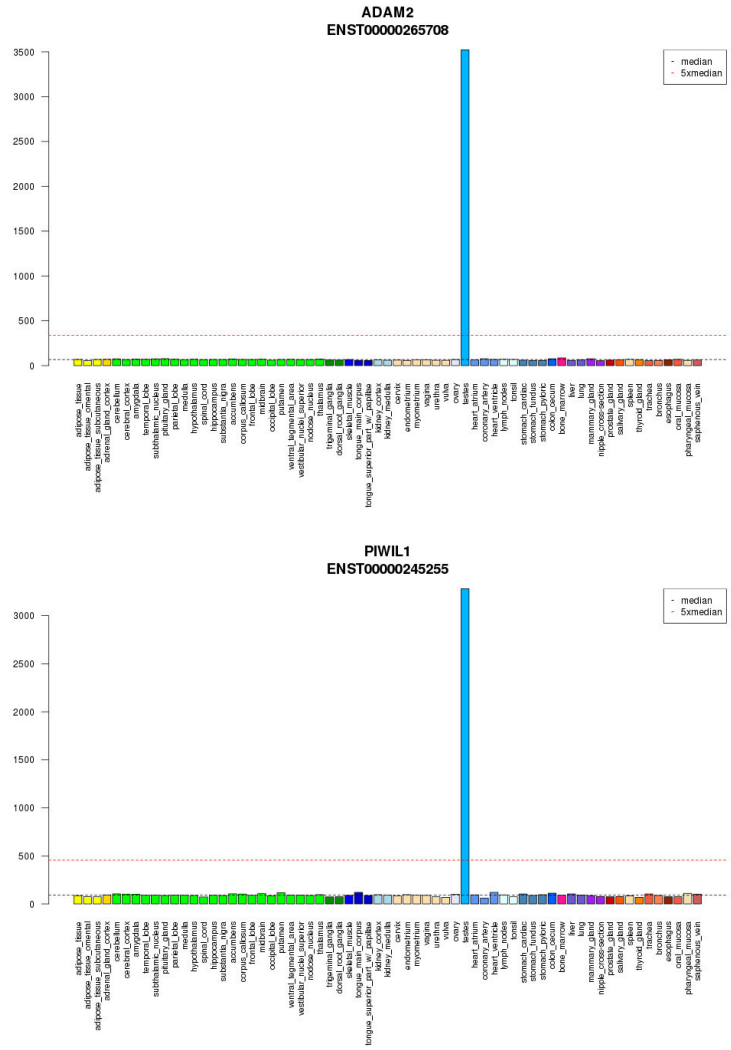


Figure 3.29: **ADAM2 and PIWIL1.** ADAM2 and PIWIL1 are known to be testis-specific antigens (CTAs) [211]. CTAs have been used in cancer immunotherapy because they are often up-regulated in cancer and can be treated very well, since there are no side effects besides potentially male infertility. Thus CTAs have been systematically tested as potential drug targets in immunotherapy.

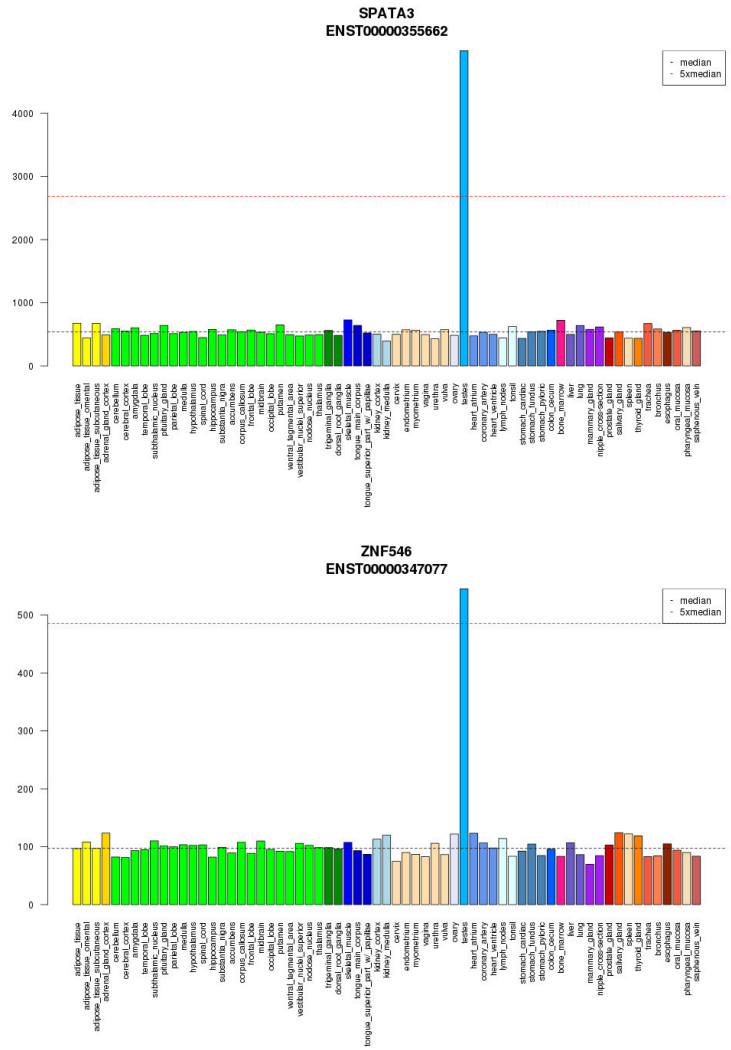


Figure 3.30: **SPATA3 and ZNF546.** SPATA3 and ZNF546 have not been known to be testis-specific antigens (CTAs). As many others we found them in our TRA data. These CTAs might be new and potential new targets for immunotherapy.

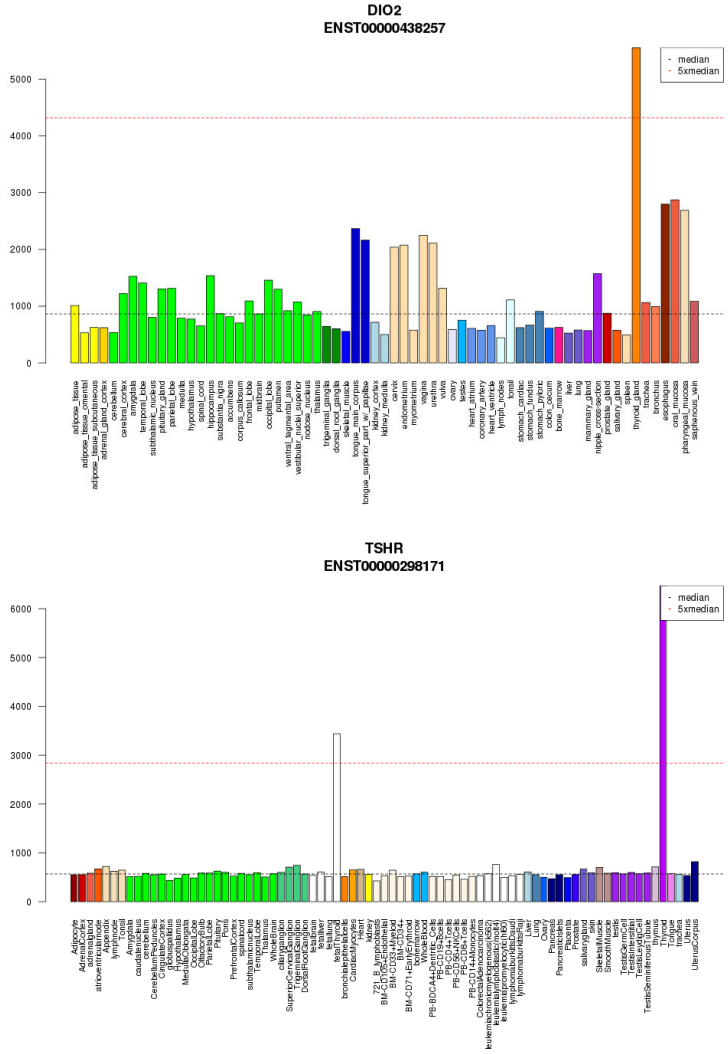


Figure 3.31: **DIO2 and TSHR are involved in autoimmune grave's disease.** Both DIO2 as well as TSHR are tissue-specific for the thyroid and known to be involved in grave's disease [300, 26]

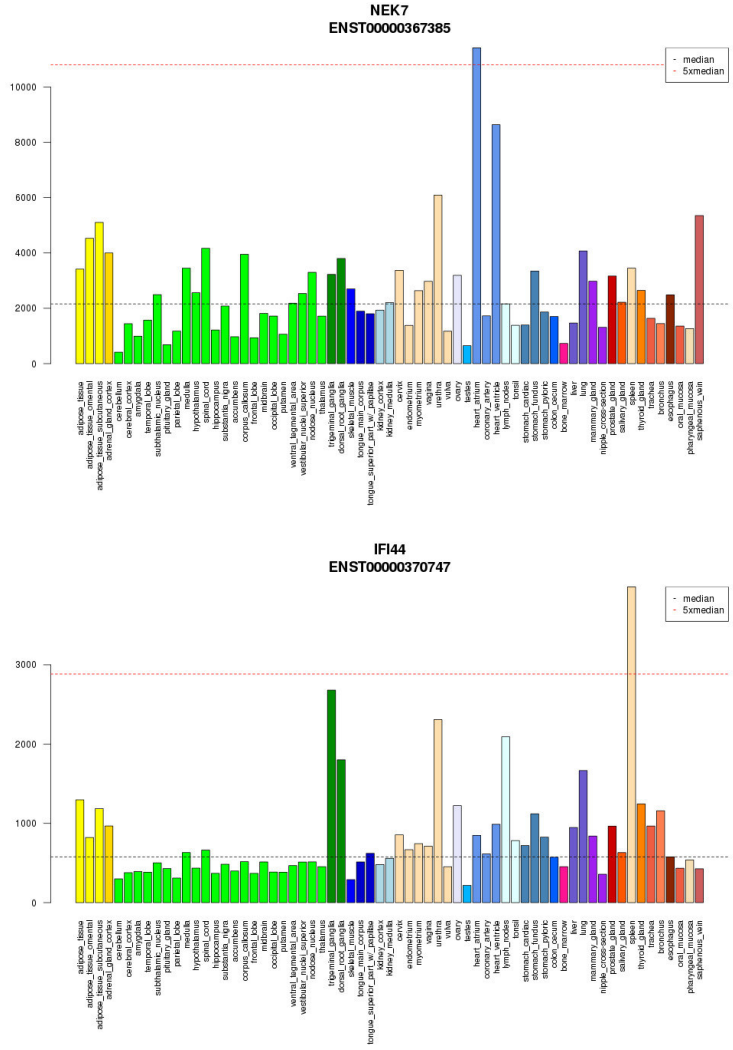


Figure 3.32: **NEK7 and IFI44 involved in lupus erythematosus.** NEK7 as well as IFI44 have been known to be involved in lupus erythematosus [241, 86]. While NEK7 is tissue-specific for the heart, IFI44 is tissue-specific for the spleen. Autoantibodies involved in lupus erythematosus are not directly connected to only one tissue type.

3.8 The roadmap of chemokines in different tissue types

The roadmap of chemokines gives a picture of where which chemokines are highly expressed in order to be tissue-specific. Chemokines are chemical attractants mainly for immune cells, in order to relocate immune cells and attract them to different tissue types. Also in cancer chemokines play an important role. While some chemokines are upregulated in certain cancer stages or types, others are down regulated. Therefore it is very important to know the role of chemokines in the healthy individual. For an overview of chemokines which are tissue-specifically expressed, please refer to Table 3.9. Each type of immune cell has certain receptors for chemokine signalling and is therefore attracted to certain tissues under certain circumstances, such as an inflammation, developmental stage as well as for example cancer. Most chemokines belong to certain chemokine families, declared by their names, there are CC-chemokines, CXC-chemokines, CX3C-chemokines, as well as CXX-chemokines. For a complete overview, please refer to Table 3.9. Chemokines with its corresponding tissue-specificity can be seen in Fig. 3.34 Similar to the TRA results in general also most tissue-specific chemokines can be found in testis.

symbol	species	tissue	citation
CCL1	human	• CNS, testis	• binds to CCR8 [268]
CCL2	human	• smooth muscle, cardiac myocytes	• produced by astrocytes in MS, attracts T-cells [96, 253]
	mouse	• macrophages, mast cells, microglia, osteoblasts, osteoclasts	• binds to CCR2, CCR5 [268]
	-		• recruits monocytes, NKT cells, monocytic MDSCs, protumor effect [268]
	-		• on stromal cells, immune cells, melanoma, activates TNF [440]
CCL3	mouse	• immune cells	• binds to CCR1, CCR4, CCR5 [268]
	-		• recruits monocytes, macrophages, promotes cancer extravasation [77]
	-		• on intratumoral myeloid derived suppressor cells, MDSCs, in melanoma CD8 T cell infiltrate, improved survival[440]
CCL4	human	• spleen	• binds to CCR1, CCR3, CCR5 [268]
	mouse	• immune cells	• on intratumoral myeloid derived suppressor cells, MDSCs, in melanoma CD8 T cell infiltrate, improved survival[440]

Table 3.9 – TRAs of the chemokine family

symbol	species	tissue	citation
CCL5	human	• whole blood, immune cells	• inflammatory chemokine, attracts T-cells [344]
		• bronchus, spleen, tonsil	• binds to CCR1, CCR3, CCR4, CCR5[268]
		• bronchus, spleen	• recruits monocytes, macrophages, promotes cancer invasion[268]
	human	• whole blood, immune cells	• recruits NK cells in melanoma[440]
		• bronchus, spleen, tonsil	-
		• bronchus, spleen	-
	mouse	• lymphnode, adiposetissue, immune cells, trachea	
		• immune cells, intestine small, lymph nodes, mammary gland, spleen	
CCL6	mouse	• large intestine	
CCL7	mouse	• immune cells, osteoblasts, osteoclasts	• binds to CCR1, CCR2, CCR3 [268]
CCL8	human	• adipose tissue, coronary artery, colon cecum	
CCL9	mouse	• adipose tissue, liver	
CCL11	human, mouse	• stomach, colon	• binds to CCR3 [268]
		• uterus, adipose tissue	
CCL12	mouse	• lymph nodes, macrophages, microglia, osteoclasts	
CCL14	human	• liver	
CCL15	human	• kidney, liver	
CCL16	human	• liver	
CCL17	human protein	• appendix, lymphnode, tonsil, urinary bladder	• binds to CCR4 [268]
CCL18		• lung, tonsil	• promotes invasion, metastasis [268]
CCL19	human	• lymphnode, thymus	• attract T-cells [344]
	human protein	• appendices, lymphnode, tonsil	• binds to CCR7[268]
	mouse	• lymphnode, trachea	
CCL20	human	• lung, stomach pyloric, tonsil	• endothelial cells, CNS in MS [344]
	-		• binds to CCR6 [268]
	-		• expressed by Th22 cells, recruited by CCL20, pro-tumor effects [268]
	-		• tumor macrophages, activates TNF, recruits DCs into melanoma [440]
CCL21	human	• lymphnode	• endothelial cells, CNS [344]

Table 3.9 – TRAs of the chemokine family

symbol	species	tissue	citation
	human	• appendices, fallopiantube, lymphnode, spleen, tonsil	• binds to CCR7[268]
CCL22	mouse	• DCs , lymph nodes	• binds to CCR4 [268]
CCL24	human	• colon transverse, small intestine terminal ileum, spleen	• binds to CCR3 [268]
CCL25	human	• thymus	• binds to CCR9 [268]
	mouse	• small intestine, thymus	• promotes chemoresistance, tumor invasion, metastasis[268]
	human	• CNS, small intestine, ileum, spleen, testis • small intestine, ileum, testis • small intestine, ileum	
CCL26	human	• adrenal gland, cervix ectocervix, fallopian tube, ovary, vagina • lymphnode, ovary, rectum	-
CCL27	human	• nipple cross section • testis	• binds to CCR10 [268]
	human protein	• skeletalmuscle, skin, testis	
CCL28	human	• bronchus, mammary gland, salivary gland, trachea • mammary gland, salivary gland, thyroid gland, trachea	• binds to CCR10[268]
	mouse	• intestine large, intestine small, salivary gland	
	human	• mammary gland, colon, salivary gland, pancreas, skin, thyroid	-
CXL1	no TRA	-	• binds to XCR1 [268]
CXCL1	human	• smooth muscle	• binds to CXCR1, CXCR2 [268]
		• bronchus, lung, spleen, trachea	• increases granulopesis[77]
		• appendices, spleen, urinarybladder	• expressed by Megakaryocytes, endothelial cells, cancer cells, attract neutrophils from bone marrow to the tumor[77]
CXCL2	human	• liver	• binds to CXCR2 [268]
		• pancreatic islets, smooth muscle	• increases granulopesis[77]
		• adipose tissue, mammary gland, fallopian tube, heart, liver, lung, salivary gland, pancreas	• expressed by megakaryocytes, endothelial cells, cancer cells, attract neutrophils from bone marrow to the tumor[77]

Table 3.9 – TRAs of the chemokine family

symbol	species	tissue	citation
		• adipose tissue, heart, liver, salivary gland	
	mouse	• immune cells, cornea, lens, microglia	
CXCL3	human	• smooth muscle	• binds to CXCR2 [268]
		• lung, stomach cardiac, stomach fundus, stomach pyloric	
		• lung, stomach pyloric	
		• adipose tissue, cervix, heart, lung, salivary gland, prostate, stomach	
		• adipose tissue, lung, salivary gland, stomach	
	mouse	• macrophages, microglia	
CXCL4	no TRA	-	• binds to CXCR3 [268]
CXCL5	human	• smooth muscle	• binds to CXCR2 [268]
		• immune cells, osteoblasts, uterus	• increases granulopesis [77]
		• CNS, lung, salivary gland, spleen, whole blood	• expressed by megakaryocytes, endothelial cells, cancer cells, attract neutrophils from bone marrow to the tumor[77]
	mouse	• vomeral nasal organ, medial olfactory epithelium	
CXCL6	human	• smooth muscle	
		• spleen, urethra	
CXCL7	no TRA	-	• binds to CXCR1, CXCR2 [268]
CXCL8	human	• adipose tissue, CNS, heart, atrial appendage, lung, salivary gland, prostate, whole blood	• recruits neutrophils, granulocytic MDSCs, promotes invasion, migration, apoptosis, resistance to hypoxia, angiogenesis[268]
	-		• recruits neutrophils, granulocytic MDSCs, increases immunogenicity of the tumor, antitumor effect [268]
	-		• increases granulopesis [77]
	-		• expressed by Megakaryocytes, endothelial cells, cancer cells, attract neutrophils from bone marrow to the tumor [77]
CXCL9	mouse	• adipose brown, immune cells , lymph nodes, mammary gland, spleen	• binds to CXCR3 [268]
	human protein	• appendices, lymphnode, tonsil	• recruits T cells, NK cells, angiogenesis inhibitor, antitumor[268]

Table 3.9 – TRAs of the chemokine family

symbol	species	tissue	citation
CXCL10	mouse	• immune cells , lymph nodes, osteoblasts, spleen	• binds to CXCR3 [268]
	human	• kidney cortex, lung, prostate, small intestine, terminal ileum, spleen, vagina	• recruits T cells, NK cells, angiogenesis inhibitor, anti-tumor[268]
	human protein	• appendices, lymphnode	
CXCL11	human	• CNS, lung, pancreas, prostate, spleen	• binds to CXCR3, CXCR7 [268]
	human protein	• appendices, lymphnode, stomach	
CXCL12	mouse	• bone, bonemarrow	• attract T-cells [344]
	human	• bone marrow, uterus, lymphnode, cardiac myocytes, thymus	• binds to CXCR4, CXCR7 [268]
		• adipose tissue, artery aorta, mammary gland, cervix, spleen, uterus, vagina	• recruits B cells, pDCs, Treg cells into the bone marrow, promotes proliferation, survival, invasion, metastasis, angiogenesis[268]
		• uterus, lymphnode, cardiac myocytes, thymus	• increases granulopesis [77]
		• adipose tissue, cervix, mammary gland, myometrium, lymph nodes, vagina	
		• bone marrow, adipose tissue subcutaneous, lymph nodes	
		• adipose tissue, spleen, uterus	
CXCL13	human	• spleen	• B-cell growth factor [422]
		• lymphnode, tonsil	• binds to CXCR5 [268]
		• bronchus, colon cecum, lymph nodes, spleen, stomach pyloric, tonsil	
	mouse	• lymphnode, adiposetissue, spleen, trachea	
		• lymph nodes, macrophages, mammary gland, prostate, spleen	
CXCL14	human	• skin	• recruits DCs, promotes invasion, motility [268]
	protein	• skin, kidney	• recruits DCs, inhibits proliferation, invasion, metastasis, increases apoptosis, anti-tumor[268]
	mouse	• snoutepidermis	
CXCL15	mouse	• lung	

Table 3.9 – TRAs of the chemokine family

symbol	species	tissue	citation
CXCL16	human	• lung • esophagus, lung, salivary gland, testis, whole blood • CNS, lung, spleen, whole blood • bladder, cervix ectocervix, testis	• binds to CXCR6 [268]
	mouse	• immune cells , lung, lymph nodes, microglia, osteoblasts, osteoclasts, uterus	
CXCL17	no TRA	-	• recruits granulocytic MDSCs, promotes angiogenesis [268]
CX3CL1	human	• mammary gland • artery aorta, artery coronary, CNS, mammary gland, lung • adipose tissue, artery aorta, CNS, breast, mammary gland, lung • artery aorta, artery coronary, mammary gland, lung, salivary gland • brain, lung, salivarygland	• binds to CX3CR1 [268]
	mouse	• CNS • CNS, intestine small, osteoblasts, pituitary, spinal cord	
CCR1	mouse	• immune cells, osteoclasts	• binds to CCL3, CCL4, CCL5, CCL7 [268]
	human	• adipose tissue, lung, spleen, whole blood	
	human protein	• appendices	
CCR2	human	• immune cells	• binds to CCL2, CCL7 [268]
	human protein	• appendices	• expressed by monocytes, recruited by CCL2, CCL5, tumor promotion[268]
	mouse	• bone, bone marrow, immune cells, lymph nodes	• expressed by monocytic myeloid-derived suppressor cells, recruited by CCL2, CXCL5, CXCL12, pro-tumor effects[268]
		• bone, bone marrow, immune cells, lymph nodes, uterus • bone marrow, granulocytes, macrophages	
	human	• appendices, lymphnode	

Table 3.9 – TRAs of the chemokine family

symbol	species	tissue	citation
		• colon, lung, salivary gland, small intestine, ileum, spleen, whole blood	
CCR3	mouse	• bone, bone marrow, macrophages, mast cells	• binds to CCL4, CCL5, CCL7, CCL11, CCL24[268]
	human	• skin, spleen, whole blood	
	human protein	• appendices, skeletal muscle, small intestine, stomach, urinary bladder	
CCR4	human protein	• appendices, gallbladder, lymphnode, tonsil, urinarybladder	• binds to CCL3, CCL5, CCL17, CCL22 [268]
	-		• expressed by Treg cells, recruited by CCL22, CCL28, protumor effects [268]
CCR5	mouse	• immune cells, lymph nodes, microglia, osteoclasts	• binds to CCL2, CCL3, CCL4, CCL5[268]
	human	• lung, intestine, ileum, spleen, whole blood • salivary gland	
	human protein	appendices, lymphnode	
CCR6	human	• spleen, tonsil	CD4+ T cells [96, 56]
		• small intestine, ileum, spleen, testis	• binds to CCL20 [268]
	human protein	• appendices, lymphnode, spleen, tonsil, urinarybladder	• Th17 cells, immature DCs, recruited by CCL20, tumor promotion[268]
	mouse	• immune cells, lymph nodes, spleen	• expressed by Th22 cells, recruited by CCL20, pro-tumor effects[268]
CCR7	human	• lymph nodes, spleen, tonsil	[54]
		• adipose tissue, small intestine, ileum, spleen, whole blood	• binds to CCL19, CCL21[268]
		• CNS, mammary gland, cervix	
	human protein	• appendices, lymphnode, spleen, tonsil, urinarybladder	
	mouse	• immune cells, lymph nodes, spleen	
CCR8	human	• salivary gland, spleen	• binds to CCL1 [268]
	human protein	• appendices, gallbladder, lymphnode, urinarybladder	
CCR9	human	• thymus	• lit. intestinal [54]
		• small intestine, ileum	• binds to CCL25 [268]
		• cervix, colon, fallopian tube, pituitary, vagina	

Table 3.9 – TRAs of the chemokine family

symbol	species	tissue	citation
	human protein	• appendices, duodenum, small intestine, spleen	
	mouse	• thymus • immune cells	
CCR10	human	• CNS, ovary, pituitary, uterus	• binds to CCL27, CCL28 [268]
		• cervix endocervix, gallopian tube, pituitary, testis, uterus	• expressed by Treg cells, recruited by CCL22, CCL28, protumor effects[268]
		• CNS, uterus	
	human protein	• colonrectum, fat	
CXR1	no TRA	-	• binds to XCL1 [268]
CXCR1	human	• whole blood	• binds to CXCL1, CXCL7 [268]
		• adipose tissue, fallopian tube, liver, lung, spleen, whole blood	• expressed by granulocytes myeloid-derived suppressor cells, recruited by CXCL8, pro tumor effects[268]
CXCR2	human	• whole blood	• binds to CXCL1, CXCL2, CXCL3, CXCL5, CXCL7 [268]
		• bone marrow, esophagus, oral mucosa, spleen	• expressed by granulocytes myeloid-derived suppressor cells, recruited by CXCL8, pro tumor effects[268]
	mouse	• bone, bone marrow, granulocytes, placenta	• expressed by monocytic myeloid-derived suppressor cells, recruited by CCL2, CXCL5, CXCL12, pro-tumor effects[268]
CXCR3	mouse	• immune cells, lymph nodes	• binds to CXCL9, CXCL10, CXCL11, CXCL4 [268]
	human	• colon , small intestine, ileum, spleen	• Th1 cells, CD8+ T cells, NK cells, recruited by CXCL9, CXCL10, CXCL11, tumor promotion[268]
CXCR4	human	• immune cells, bonemarrow, whole blood, lymphnode, thymus	• attract T-cells [344]
		• adrenal gland, cervix, lung, CNS, small intestine, oleum, spleen, whole blood	• binds to CXCL12 [268]
		• bone marrow, adrenal gland cortex, lung, lymph nodes, spleen, tonsil	• expressed by monocytic myeloid-derived suppressor cells, recruited by CCL2, CXCL5, CXCL12, pro-tumor effects[268]
	human protein	• appendices, bonemarrow, lymphnode, spleen, tonsil	• expressed by pDC, recruited by CXCL12, pro-tumor effects[268]

Table 3.9 – TRAs of the chemokine family

symbol	species	tissue	citation
	mouse	• bone, T cells, B cells, thymus	
CXCR5	mouse	• immune cells, lymph nodes, spleen	• binds to CXCL13 [268]
	human protein	• appendices, bonemarrow, lymphnode, spleen, tonsil	
CXCR6	human protein	• lymphnode	• binds to CXCL16 [268]
		• cervix ectocervix, minor salivary gland, ovary, vagina	
		• salivary gland, small intestine, terminal ileum, spleen	
	mouse	• immune cells, lymph nodes, spleen	
CXCR7	no TRA	-	• binds to CXCL12, CXCL11 [268]
CX3CR1	human	• immune cells	• binds to CX3CL1 [268]
		• CNS	
	human protein	• CNS	
	mouse	• mast cells, microglia	
	mouse	• immune cells, microglia, osteoblasts	

Table 3.9: **The roadmap of chemokines** Different groups of chemokines are tissue-specific for different tissue-types. Chemokines are chemoattractants for immune cells and guide them through their specific receptors to the tissues of their destination. This is true for the healthy individual as well as situations, such as inflammation, illnesses or cancer. Chemokines can be subgrouped according to their names in CC-chemokines, CXC-chemokines, CX3C-chemokines, as well as CXX-chemokines.

Table 3.9 – TRAs of the chemokine family

3.8.1 TRAs per tissue type

The distribution of TRAs per tissue-type varies per dataset. The biggest group of tissue-restricted antigens is in all five datasets covered by testis-specific antigens, in the mouse Novartis dataset we find more than 1,200 TRAs tissue-specific for the testis. In the human Novartis dataset, we have four different subgroups of testis genes, which also cover about the same amount of genes. In the mouse Lattin as well as in the human Roth dataset, we find about 4,000 testis-specific antigens and in the GTEX dataset, we even find more than 20,000 testis-specific antigens. The last dataset most probable also covers non coding RNA (Fig. 3.35 - 3.37).

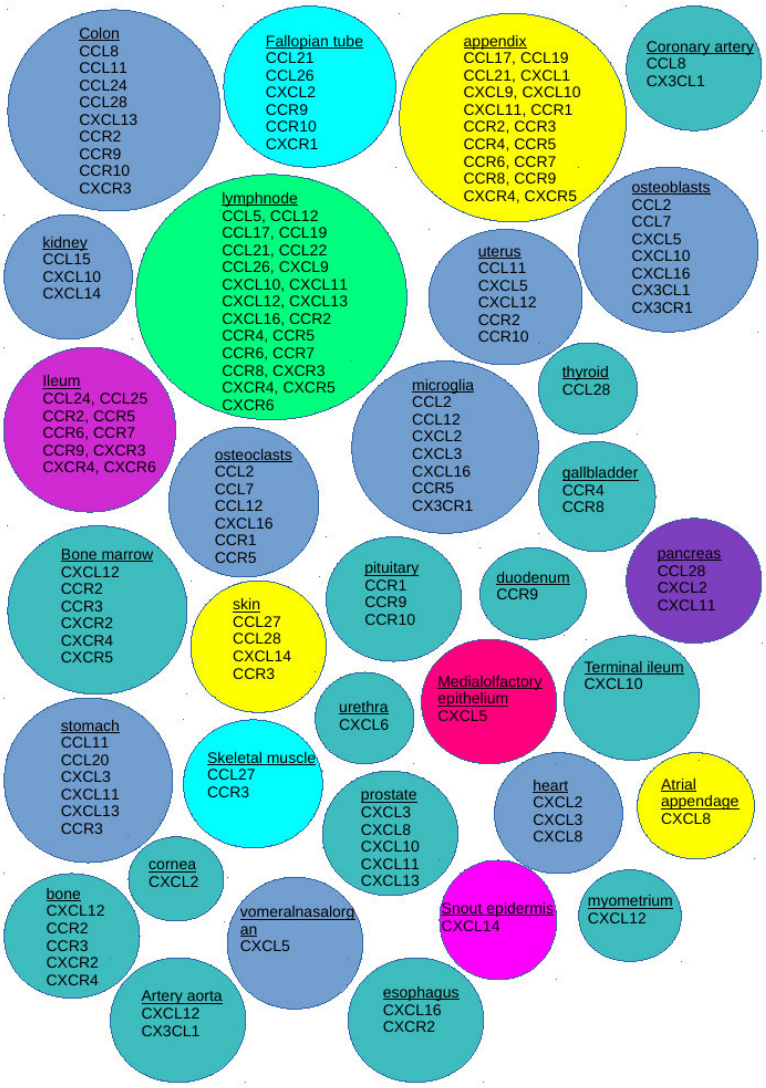


Figure 3.33: **Chemokines with its corresponding tissue-specificity.** Most chemokines are tissue-specific for the testis, some are tissue-specific for the immune cells (different cell types), smooth muscle, bronchus and trachea, the heart, central nervous system (CNS), kidney, ovary, mammary gland, spleen, lung, salivary gland, liver, thymus as well as whole blood. Chemokines which are tissue-specific for more than one tissue were not plotted here.

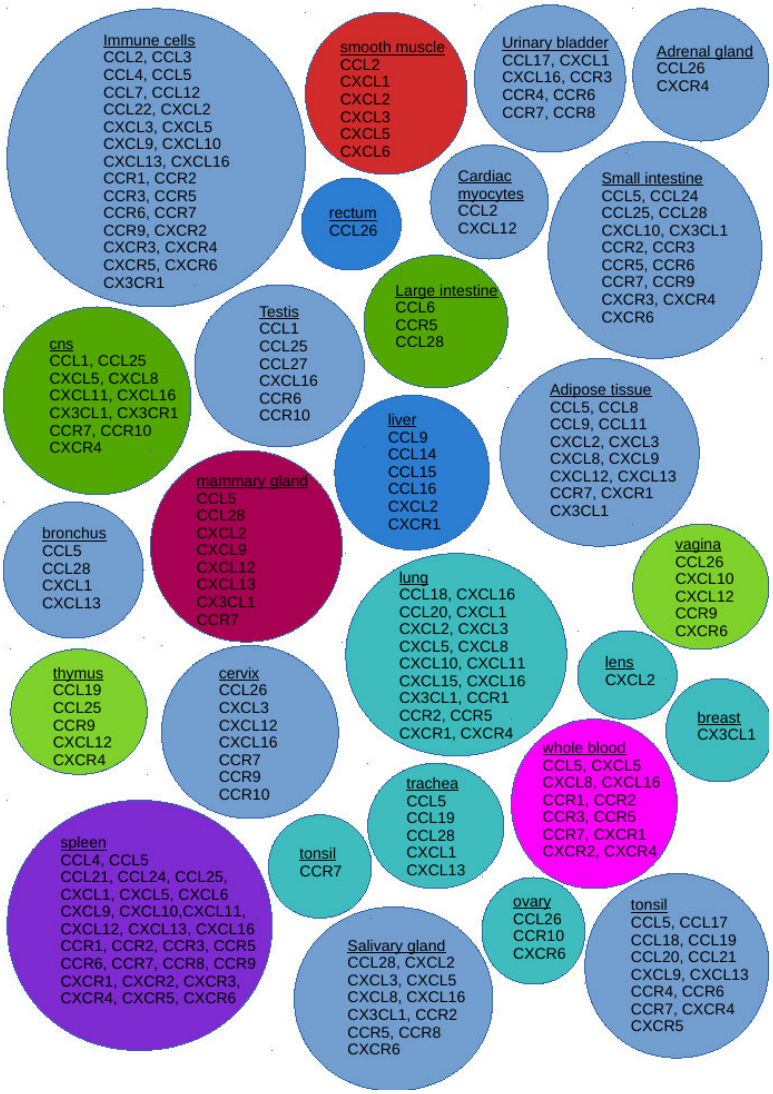


Figure 3.34: **Chemokines with its corresponding tissue-specificity.** Most chemokines are tissue-specific for the testis, some are tissue-specific for the immune cells (different cell types), smooth muscle, bronchus and trachea, the heart, central nervous system (CNS), kidney, ovary, mammary gland, spleen, lung, salivary gland, liver, thymus as well as whole blood. Chemokines which are tissue-specific for more than one tissue were not plotted here.

The second largest group of TRAs in the mouse Novartis dataset are tissue-restricted antigens tissue-specific for the oocyte. Other TRAs which often represent the same tissues are tissue-specific for the liver, kidney, skeletal muscle, placenta, spleen as well as bone marrow. For this calculation we have only considered tissues with maximum expression per TRA.

3.9 Chromosomal clustering of tissue-restricted antigens

In order to understand the common gene regulation of tissue-restricted antigens in medullary thymic epithelial cells (mTECs), we studied chromosomal clustering of TRAs. As it has been known, that chromosomal clustering might be an explanation for common gene regulation, due to common transcription factors, close proximity on a chromosomal level, chromatin folding, as well as open chromatin structures in gene neighborhoods this might be an explanation for TRA gene expression in the same cell type to otherwise genes of diverse tissue-type and function. For this study, we used two different methods the 10-gene window method developed by Roy et al. [328] as well as the sliding gene window method of fixed size developed by Gotter et al. [144]. While the 10-gene window method also takes gene loose and dense regions into account, the sliding gene window method of fixed size only counts genes within a direct neighborhood of a certain size, measured in kilo bases. The combination of both methods accounts for the complete picture of chromosomal clustering of TRAs.

3.9.1 The 10-gene window method

Chromosomal clustering of tissue-restricted antigens (TRAs) can be seen to be highly significant compared to 1,000 random gene lists of the same size as the TRA list in all datasets (Fig. 3.38 - 3.39) [102]. The number of TRAs within a sliding 10-gene window is summed up until the window encounters a region without any TRA. Once there is no TRA within the following 10-genes, the calculation drops of and starts with the next cluster calculation. The number of TRAs clustered is written down in an output file, including the TRAs within the cluster, which can be examined later on. The sizes and numbers of TRA clusters can be seen in Fig. 3.38 - 3.39. While we do not encounter any cluster bigger than 7, 8 or 10 genes in a direct neighborhood in the 1,000 randomly drawn gene lists, we can observe TRA clusters of up to 87, 22, 134 or 251 in our TRA datasets. This means, that we have significant TRA clustering with a p-value < 0.001. Within TRA clusters we often find members of whole gene families, which are shown in Fig. 4.3 and Fig. 4.4. Among them already partially known gene clusters, such as the KLK cluster, the S100 cluster as well as the CSN locus (Fig. 4.5 and Fig. 4.6) but also new clusters shown in Fig. 4.3 and Fig. 4.4.

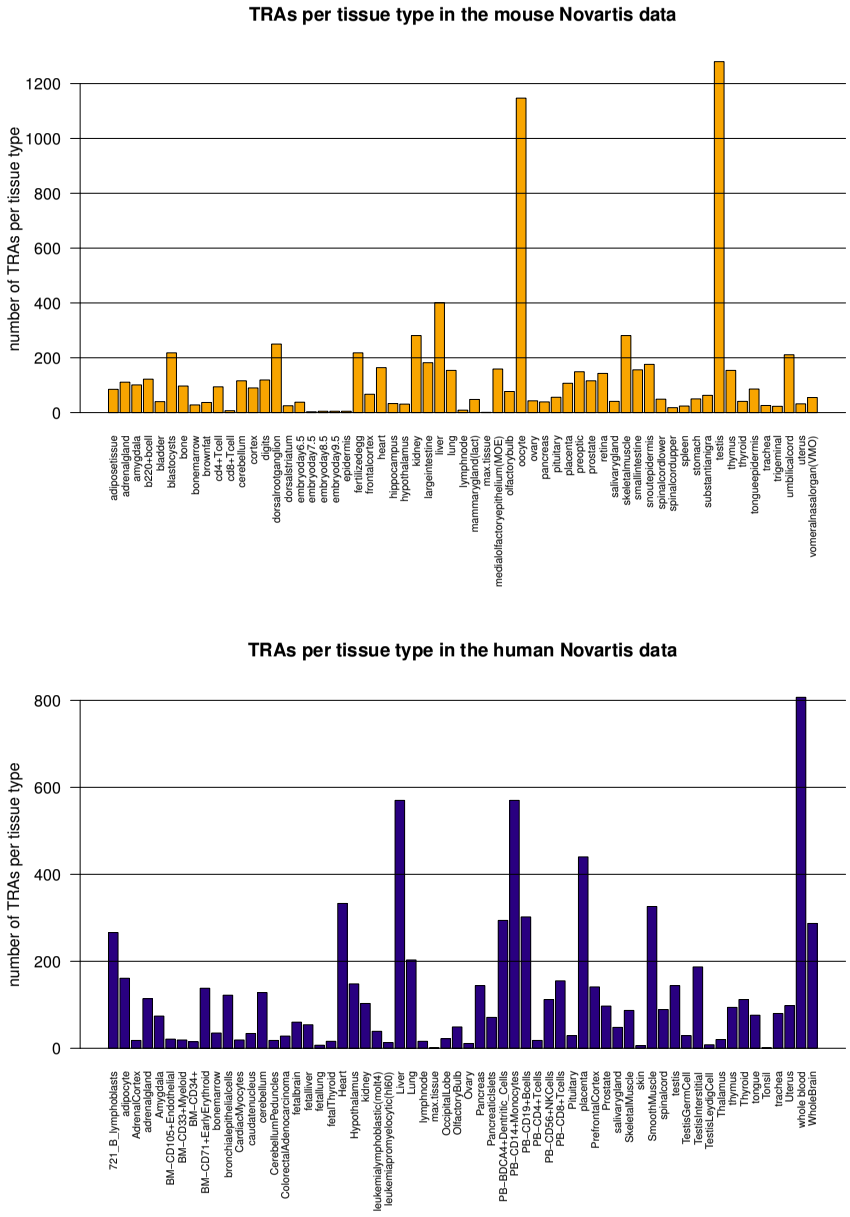


Figure 3.35: **TRAs per tissue-type in the mouse and human Novartis dataset.** The amount of TRAs found per tissue-type varies per dataset (orange bars). The biggest group of TRAs is tissue-specific for the testis as well as oocyte in the mouse Novartis dataset and tissue-specific for the testis, liver, placenta, smooth muscle, as well as heart in the human Novartis dataset (blue bars). In the human Novartis dataset there are different subtypes of testis-specific tissue-types, this is the explanation why this bar is not shown here as prominent, but the numbers of TRAs of all four testis-specific TRAs adds up to about the same amount as in the mouse Novartis dataset.

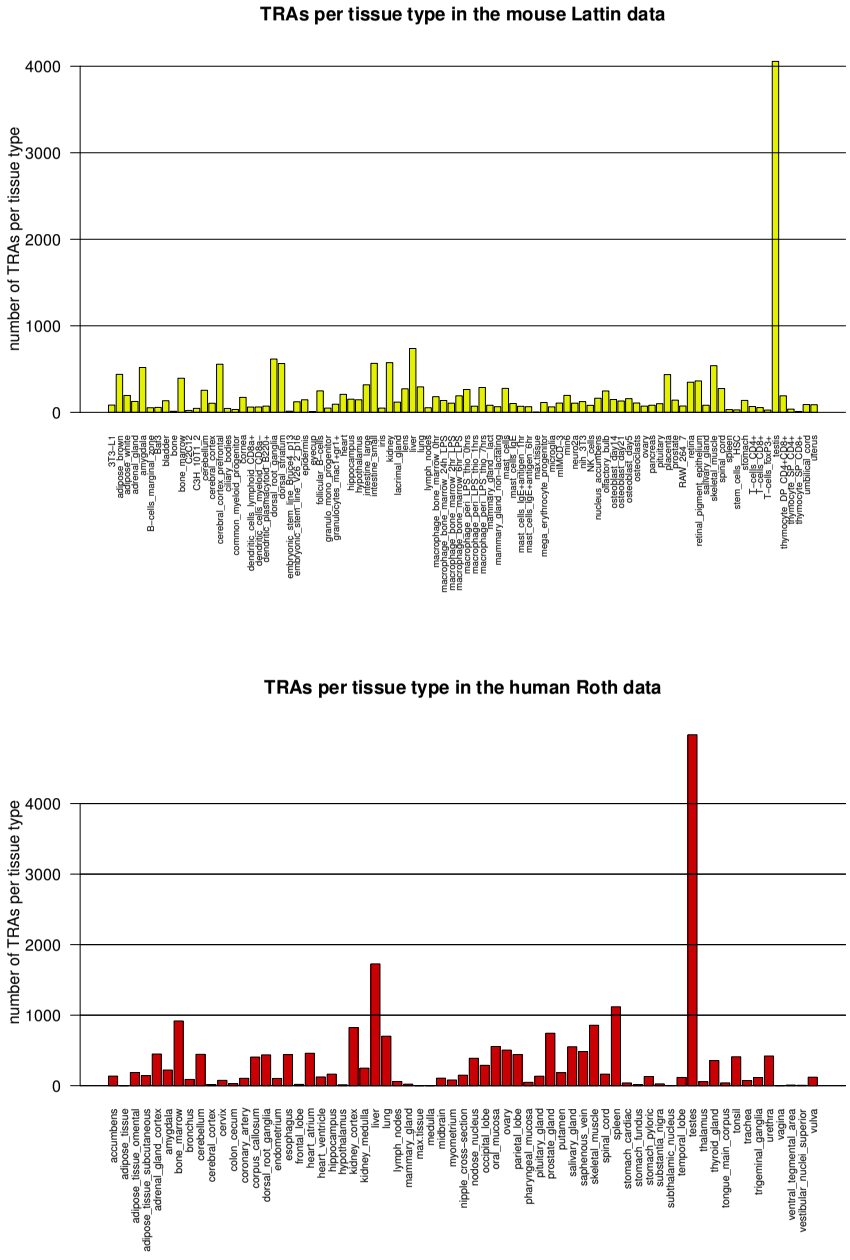


Figure 3.36: **TRAs per tissue-type in the mouse Lattin and human Roth dataset.** Most tissue-restricted antigens (TRAs) in the mouse Lattin dataset are tissue-specific for the testis (yellow bars). In the human Roth dataset the highest amount of TRAs is also tissue-specific for the testis (red bars), in this dataset also the group of liver, spleen and bone marrow specific TRAs is elevated compared to the other groups.

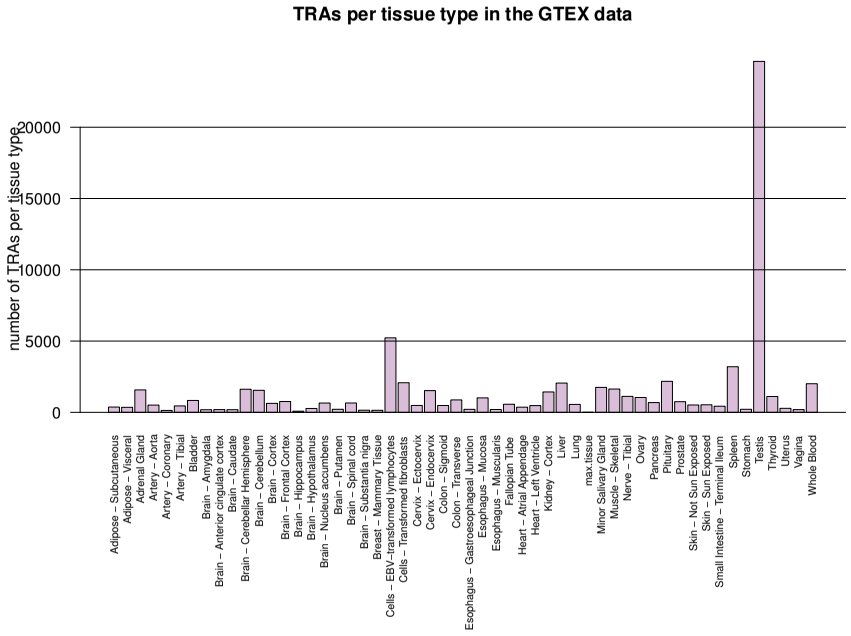


Figure 3.37: **TRAs per tissue-type in the human GTEX dataset.** The biggest group of TRAs in the human GTEX (RNAseq) dataset is tissue-specific for the testis (including non coding RNA). Further groups are adrenal gland, the CNS, spleen, pituitary as well as whole blood.

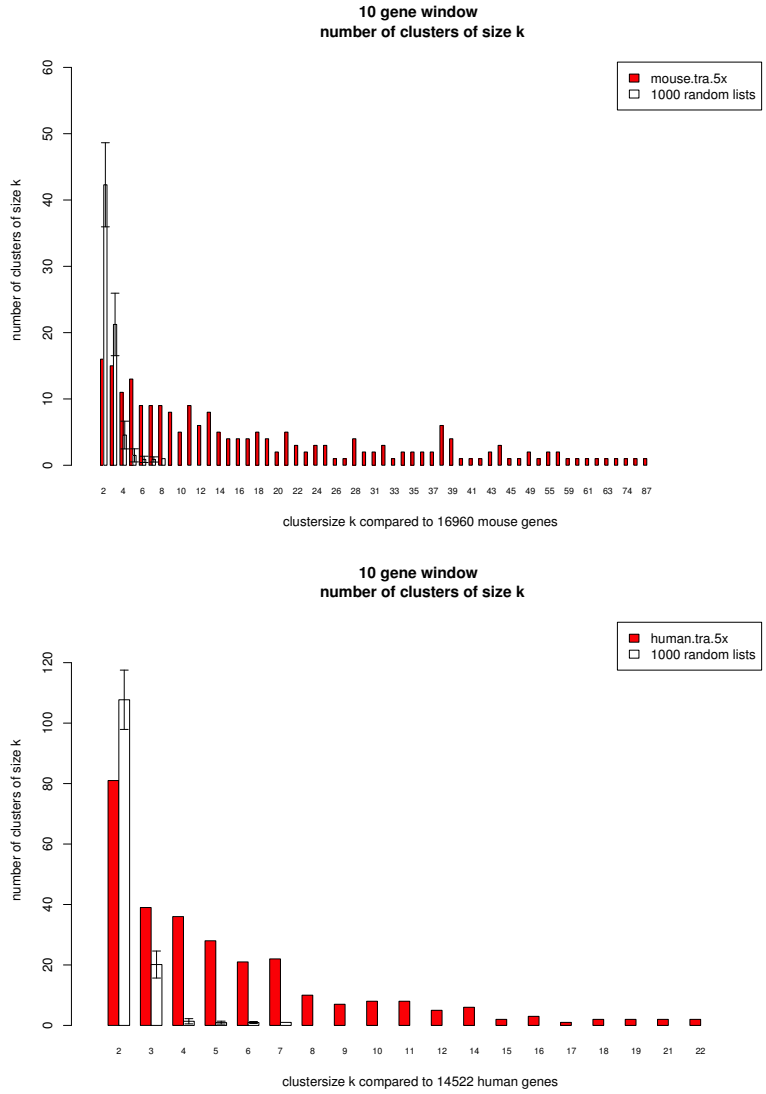


Figure 3.38: **Chromosomal clustering of TRAs in the mouse and human Novartis dataset.** In the mouse Novartis dataset, TRAs (red bars) are significantly clustered compared to 1,000 randomly drawn lists of the same size (white bars). In this dataset they account up to 87 genes within one TRA cluster compared to the maximum of 8 genes in the randomly drawn gene lists. Here the number of clusters of size k are drawn (y-axis) versus the size of the TRA clusters (x-axis). Chromosomal clustering of TRAs might explain the common regulatory mechanism of TRA gene expression in medullary thymic epithelial cells (mTECs). In the human Novartis dataset we find TRA clusters up to the size of 22 TRAs in a direct gene neighborhood compared to 7 genes in the 1,000 randomly drawn gene lists. This data has been recalculated according to Dinkelacker 2007 [102], the method taken from Gotter et al. 2004 [144].

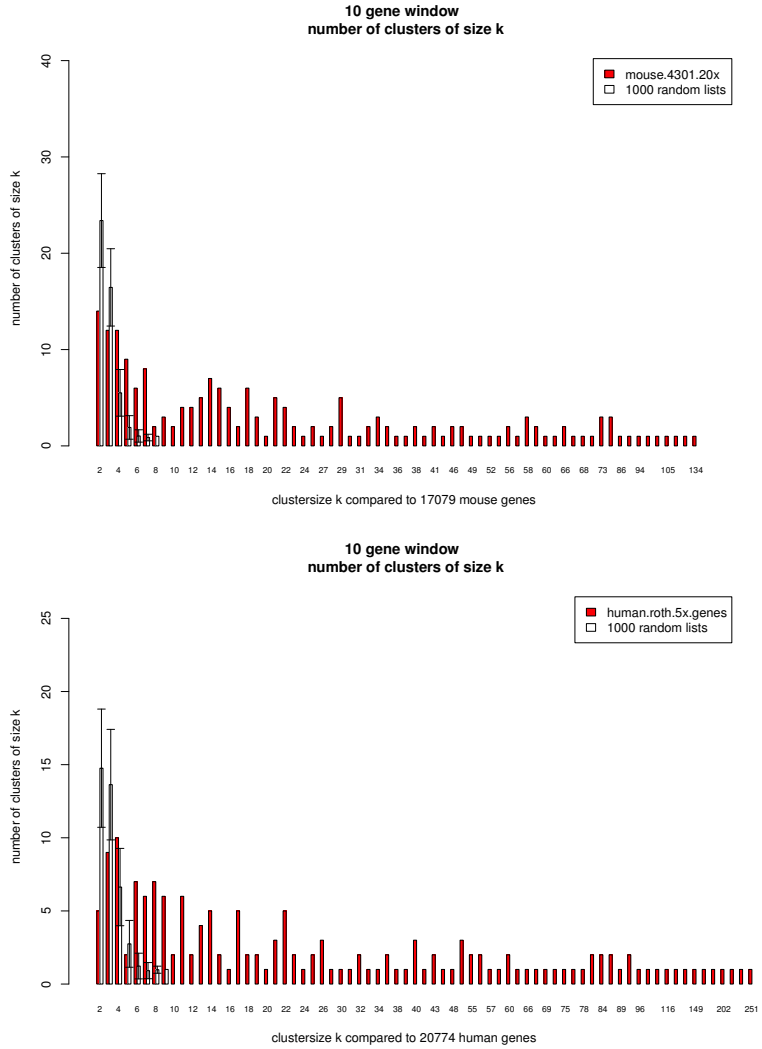


Figure 3.39: **Chromosomal clustering of TRAs in the mouse Lattin and human Roth dataset.** In the mouse Lattin dataset we find TRA clusters of sizes up to 134 TRAs in a direct neighborhood (red bars) compared to the maximum of 8 randomly drawn genes for 1000 randomly drawn gene lists of the same size (white bars). The number of clusters found is depicted in the y-axis, while the size of the clusters is shown on the x-axis. In the case of the human Roth dataset, we find TRA clusters of sizes of up to 251 TRAs in a direct neighborhood using the 10-gene window method, developed by Roy et al. [328] compared to the maximum of 9 genes in the 1,000 randomly drawn gene lists. Thus chromosomal clustering of TRAs is highly significant for all tested datasets, both in human and in mouse. This data has been recalculated according to Dinkelacker 2007 [102], the method taken from Gotter et al. 2004 [144].

3.9.2 The neighbourhood analysis

In the neighbourhood analysis, developed by Gotter et al. [144], we applied moving gene windows in the sizes of 50 kb, 100 kb, 200 kb, 500 kb, 800 kb, 2 mb up to 5 mb. For each window the number of gene pairs was calculated once for the TRA list as well as for 1,000 randomly drawn gene lists of the same size. In the case of the mouse Novartis dataset, we have highly significant clustering with a p-value < 0.001 for the sizes of 50 kb, 100 kb, 200 kb as well as 500 kb (Fig. 3.40). The red vertical line depicts the number of gene pairs of TRAs compared to the distribution of numbers of 1,000 randomly drawn gene lists (grey bars). In the case of the human Novartis dataset we observe the same chromosomal clustering of TRAs with a p-value < 0.001 for all window sizes (Fig. 3.41) (red vertical lines), compared to 1,000 randomly drawn lists (distribution of grey bars). The same tendency holds true for the other two datasets, sometimes depending on the cutoff (data not shown here) [102].

3.9.3 Chromosomal clustering of housekeeping genes and other functionally related genes

Since all our TRA clusters are highly significantly clustered on a chromosomal level, we also tested housekeeping genes as well as other functionally related gene groups for chromosomal clustering. Depending on the size of the gene group we could show chromosomal clustering also for housekeeping genes with the 10-gene window method compared to 1,000 randomly drawn gene lists, as well as for 211 actin cytoskeleton genes, 710 cell cycle genes, 854 apoptosis genes, as well as for 1,080 cytoskeleton genes. We could not show significant clustering for 17 TCA genes and only a slight enrichement of chromosomal clustering in 39 glycolysis genes, 41 caspase genes and 132 muscle specific genes (data not shown here). It seems as if chromosomal clustering of functionally related genes is rather a general tendency than an exception for tissue-restricted antigens.

3.9.4 Distribution of TRA clusters per chromosome

The size and the distribution of TRA clusters varies depending on the dataset and chromosomes. TRA clusters can generally be found on all chromosomes and are scattered over the varies regions (Fig. 3.42). As some chromosomes contain more genes than others, also some TRA clusters are bigger on these chromosomes than others.

3.9.5 TRA clusters are an intermingle of different tissue types

Different TRA clusters are very heterogeneous with respect to their characteristical tissue-type. While most TRA clusters consist of different tissue

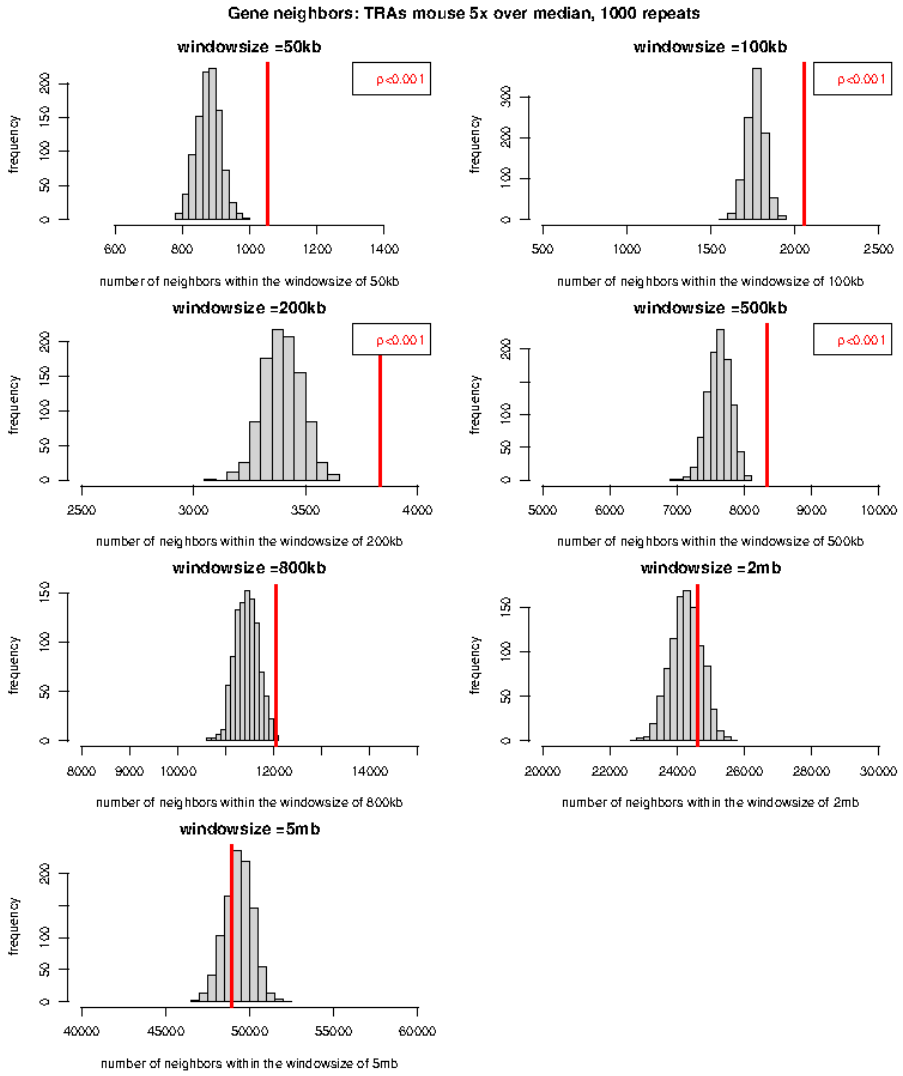


Figure 3.40: **Neighbourhood analysis of the mouse Novartis dataset.** In the neighborhood analysis of the mouse Novartis dataset, we can see TRA clustering with a p-value of < 0.001 for window sizes from 50 kb up to 500 kb (red vertical line) compared to 1,000 randomly drawn genes lists (grey bars). The bigger the window sizes get, the less significant is TRA clustering compared to the randomly drawn genes lists, which makes sence in the context of calculations. The frequency of numbers of neighbors are shown on the y-axis, compared to the number of neighbors found in direct neighborhood on the x-axis. For window sizes of 2 mb and 5 mb TRA clustering is not significant any more. This data has been recalculated according to Dinkelacker 2007 [102], the method taken from Gotter et al. 2004 [144].

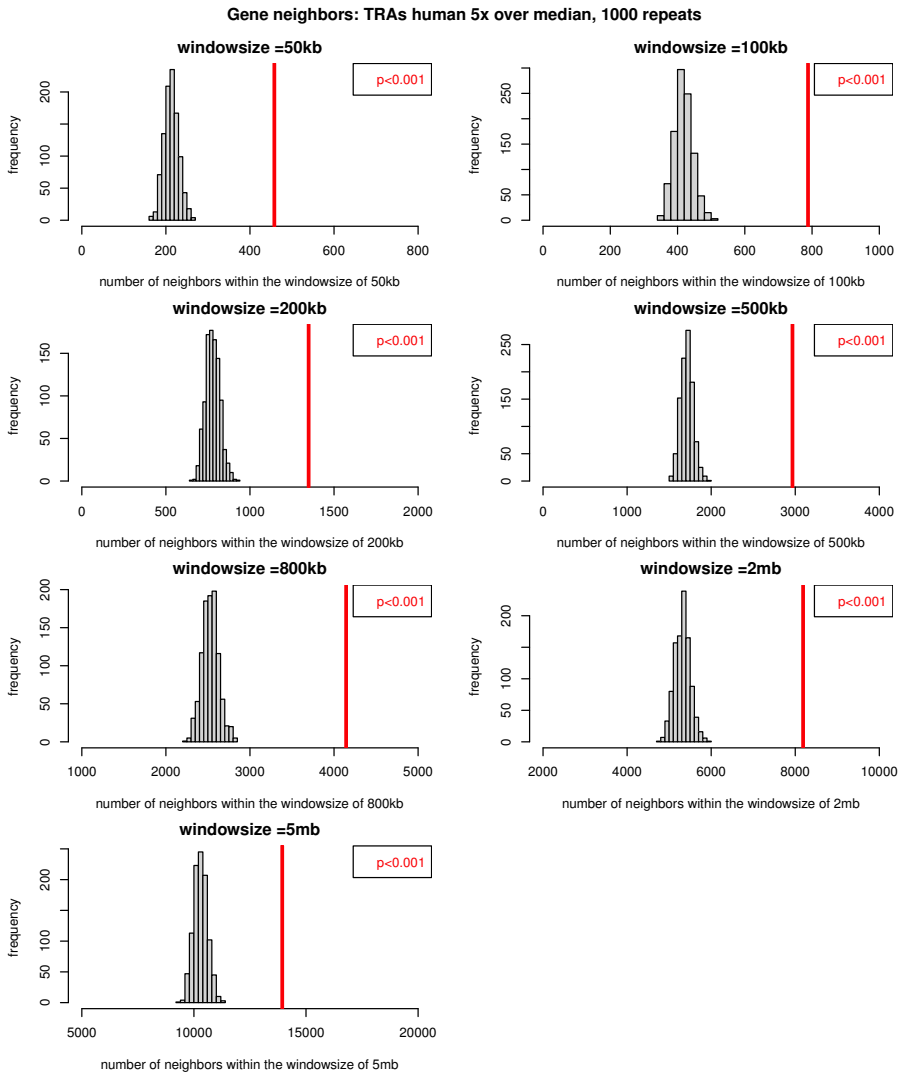
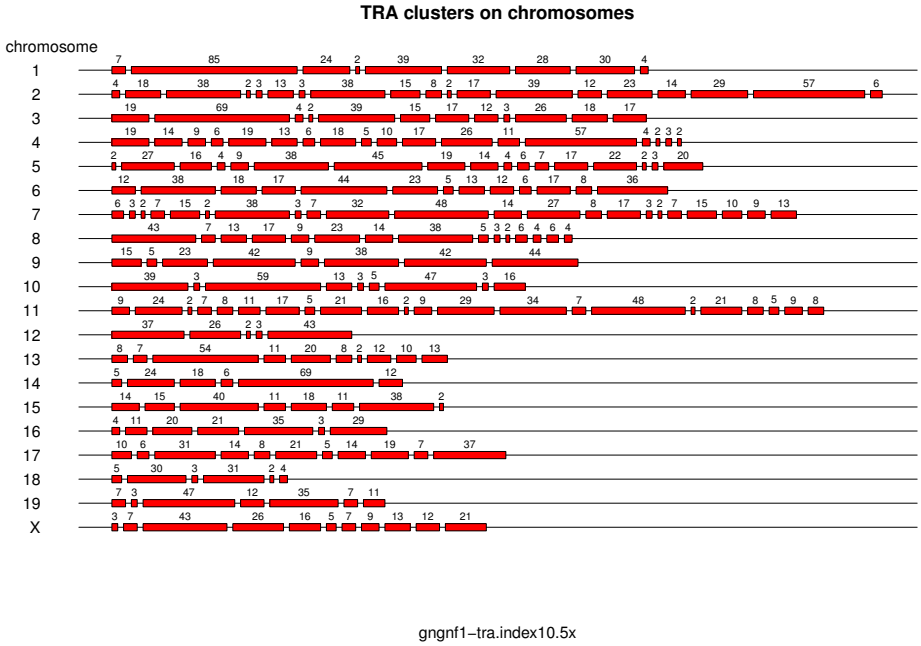


Figure 3.41: **Neighbourhood analysis of the human Novartis dataset.** In the human Novartis dataset, chromosomal clustering of TRAs is highly significant for all windowsizes from 50kb to 5mb (red vertical lines) compared to 1,000 randomly drawn gene lists of the same size (grey bars) with a p-value < 0.001. This data has been recalculated according to Dinkelacker 2007 [102], the method taken from Gotter et al. 2004 [144].



(a) mouse Novartis, 5x median

Figure 3.42: **Size and distribution of TRA clusters in the mouse Novartis dataset** TRA clusters can be found in different sizes spread out over all chromosomes. The cluster size varies from 3 adjacent genes to clusters of 85 TRAs in a direct neighborhood and are flanked of regions without TRAs. The distance between these TRA clusters are normalized here and do not represent the size of inter cluster regions, nor their exact distribution per chromosome.

specificities, other show regions of only one tissue-specificity, so called mono clusters (Fig. 3.43). In the case of the TRA clusters on chromosome 2 in the mouse, we see an intermingle of different tissue-types, while in the case of the TRA clusters on chromosome X in the mouse, we can find 43 genes in direct neighborhood tissue-specific for the testis. Although testis-specific TRAs are the largest group of all TRAs, this is still a surprisingly large group of functionally related genes. The existence of X-linked gene clusters has been known before, but the size of this gene cluster has been underestimated in the past [263, 196, 313, 388].

3.9.6 Conservation of chromosomal clustering of TRAs

Chromosomal clustering of tissue-restricted antigens (TRAs) is highly conserved between mouse human and rat. As can be seen in Fig. 3.44 as well as in Fig. 3.45, the gene order in general between human, mouse and rat is highly conserved. Not only TRAs (red lines) but also non TRAs (black lines) are either drawn in parallel or anti-parallel in all three species. Sometimes certain parts of the chromosomes are translocated to other regions, which can be seen in Fig. 3.44. This means, that also the gene order of TRAs being clustered is highly conserved between human, mouse and rat. Also in terms of tissue-type, tissue-specificity between different species seems to be conserved (Fig. 3.46).

Sometimes TRAs are co-expressed in certain tissue-type combinations, such as the liver and the kidney, but refer to the same tissue-specificity as can be seen in part a of Fig. 3.46. In case of gene duplication the tissue-specificity between species is highly conserved (Fig. 3.46). In the aspect of the evolutionary tree and conservation of the gene order of TRA clusters, we tried to follow the TRA genes (orthologs) to evolutionary more distant species, such as the kangaroo, platypus, frog, tetraodon, *Danio rerio*, sticklebacks, *Ciona intestinalis*, *Drosophila melanogaster* and *Caenorhabditis elegans* (Fig. 3.47, 3.48). The conservation of gene order loosens up the further we go down the evolutionary tree.

3.9.7 Tissue specificity in TRA clusters

Also in the single TRA clusters the diversity of different tissue-types in each cluster can be seen. Most of the time also here tissue-types are very diverse (Fig. 3.49). Fig. 3.50 and 3.51 show some examples of so called *"mono clusters"*. The examples in Fig. 3.50 show TRA clusters highly tissue-specific for the skin (Fig. 3.50 a), for the small intestine Fig. 3.50 b, or the placenta Fig. 3.51. But this is rather the exception than the normal case. Most of the time TRA clusters are very diverse in terms of tissue-specificity.

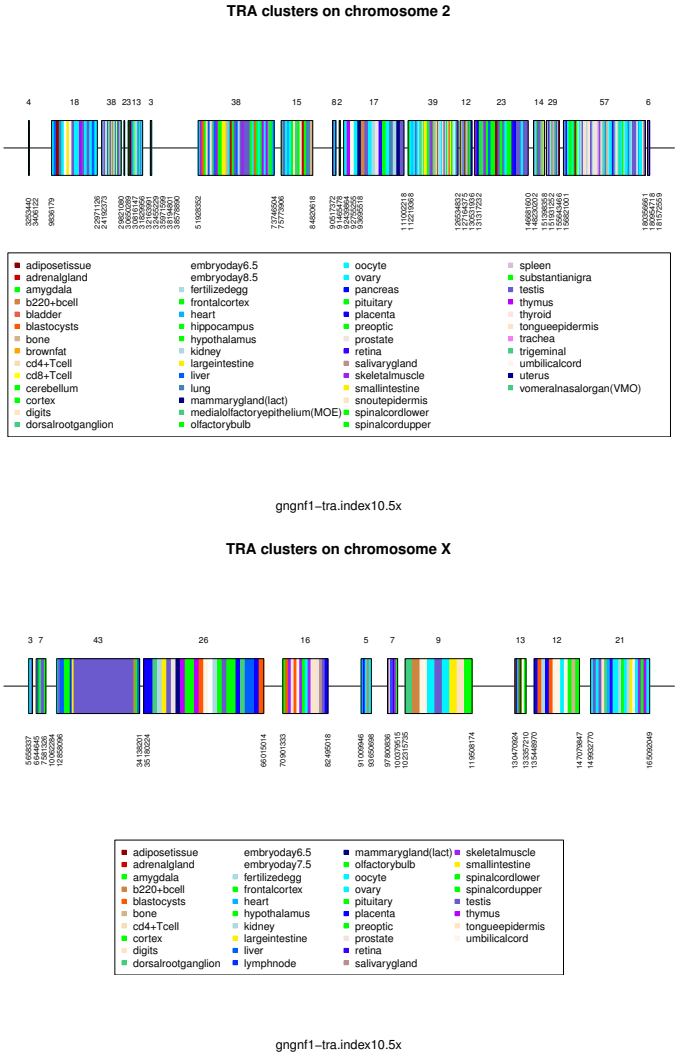


Figure 3.43: **TRA clusters in the mouse on chromosome 2 and the X chromosome.** TRA clusters are most of the time an intermingle of different tissue-types. Sometimes we can see regions of only one tissue-type, such as in this mono cluster here on chromsome X in the mouse. There are 43 different TRAs tissue-specific within one region for the testis. Since testis-specific genes account for the highest number of TRAs in general this might be an observation, which we expected, but also for other tissues this holds true. There are a few regions where we find mono clusters within TRA clusters.

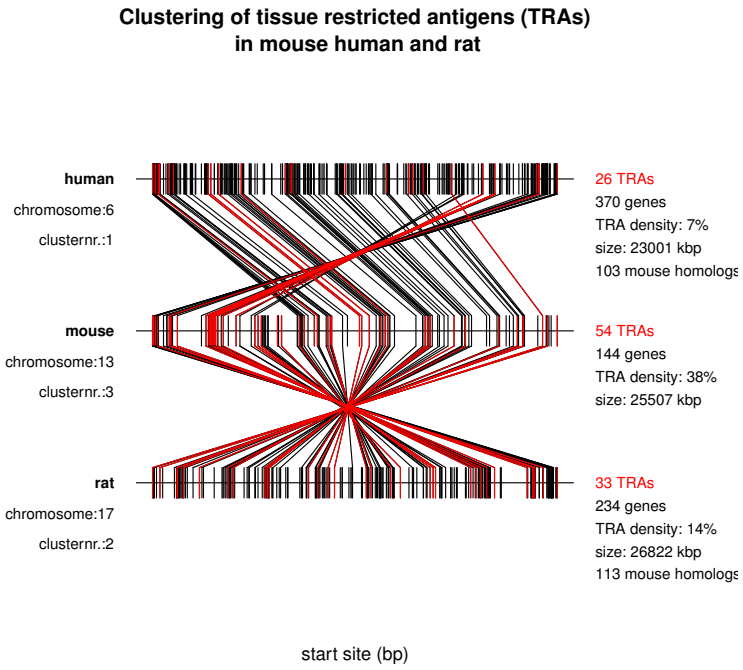


Figure 3.44: **Synteny plot between human, mouse and rat.** Looking at the synteny plot of TRA clusters between human, mouse and rat, it becomes evident that the gene order between these three species is highly conserved as well in terms of TRAs (red bars), as also in non-TRAs (black bars). most TRA clusters are either in parallel as here in the case of human and mouse or anti parallel as here in the case of mouse and rat. Some parts of the chromosomes have been shifted to other places, such as depicted in the case of the diagonal lines of the human and mouse cluster.

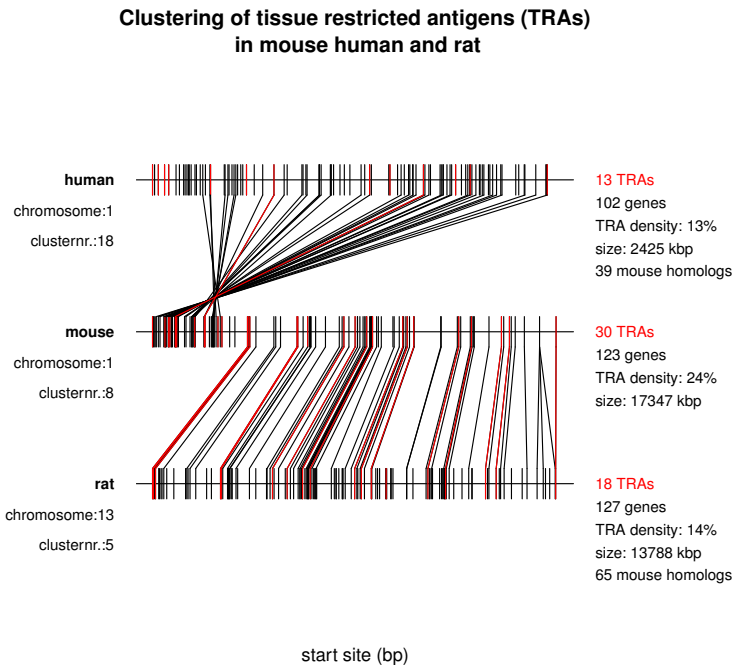


Figure 3.45: **Synteny plot between human, mouse and rat.** Another synteny plot between human, mouse and rat show that some parts of a TRA cluster matches to one TRA cluster in the other species and another to a TRA cluster in the other species. TRA clusters were matched upon the highest hitrate in terms of found homologous TRAs. All TRA clusters are always intermingled by non-TRAs (black lines). TRAs a drawn with red lines.

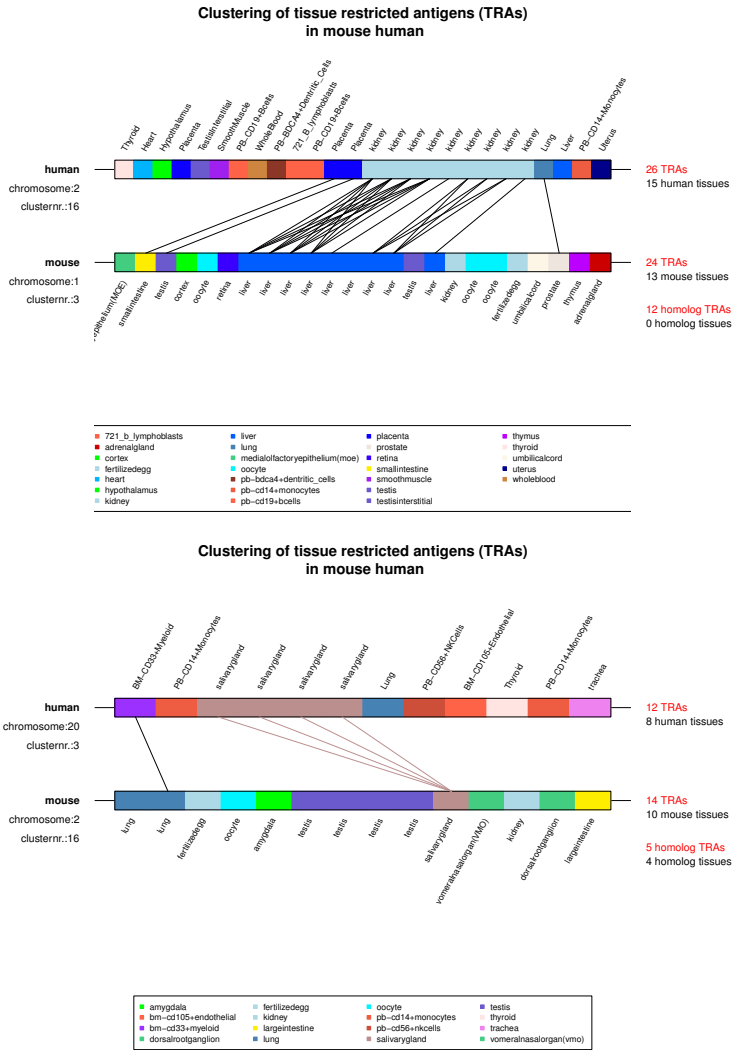


Figure 3.46: **Synteny plot mouse 1:3 human 2:16.** Also in terms of tissue specificity many TRAs are highly conserved within human and mouse in this respect, since only the max tissue was measured here, the tissue type is not always reflected by the colored lines, but also by the second highest tissue expressed in these TRAs. Many genes for example are tissue-specific for the liver but also for the kidney, as can be seen in the first example above. Sometimes also one gene in the mouse is homologue to several genes in the human, but tissue specific for the same tissue, as can be seen in the second example. Since this analysis is not so accurate however, we did not go into futher detail concerning this respect.

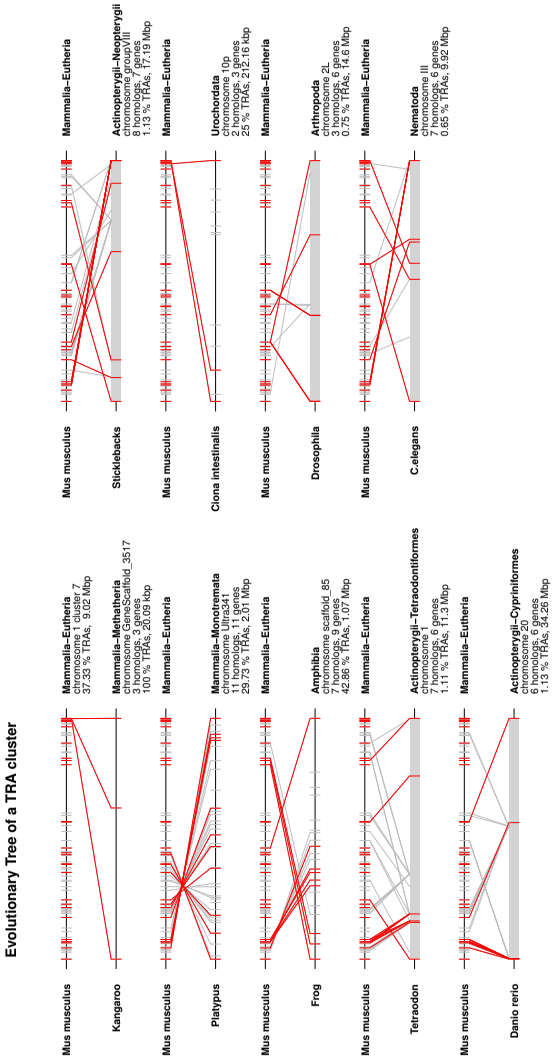


Figure 3.47: **TRAs in the evolutionary tree 1, chromosome 1, cluster 7.** TRA clutsers are evolutionary conserved within species, the conservation of gene order however gets lost the further down we reach in the evolutionary tree. TRAs are shown here in red lines, non TRAs in grey lines. We always chose the cluster with the highest score in terms of homologous genes.

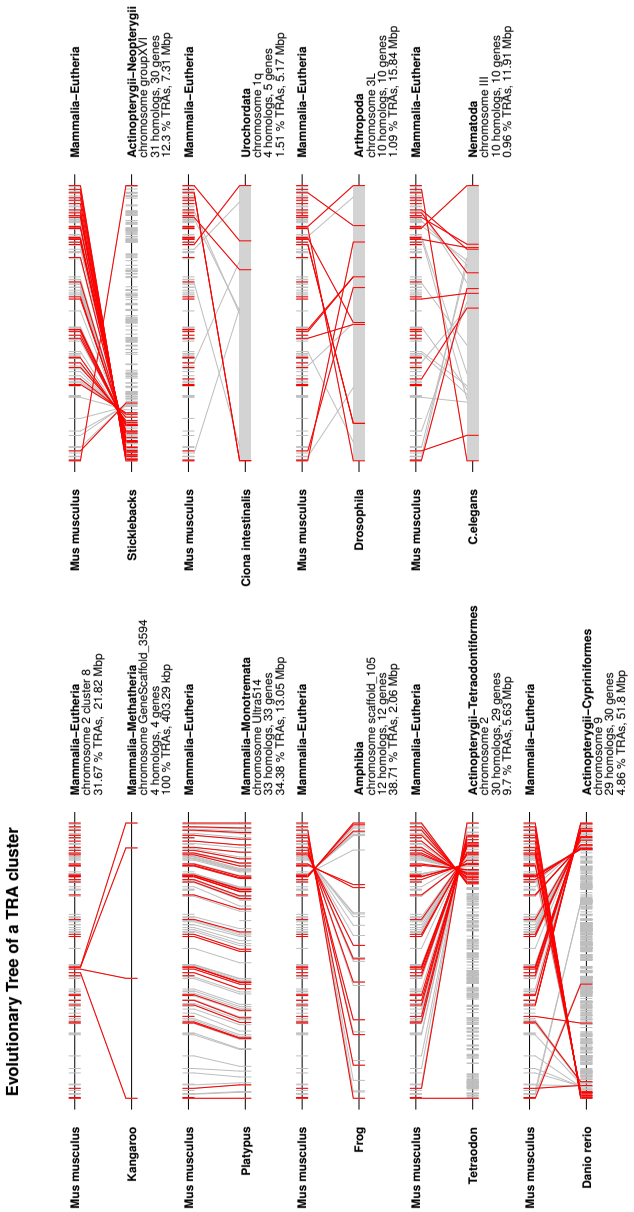
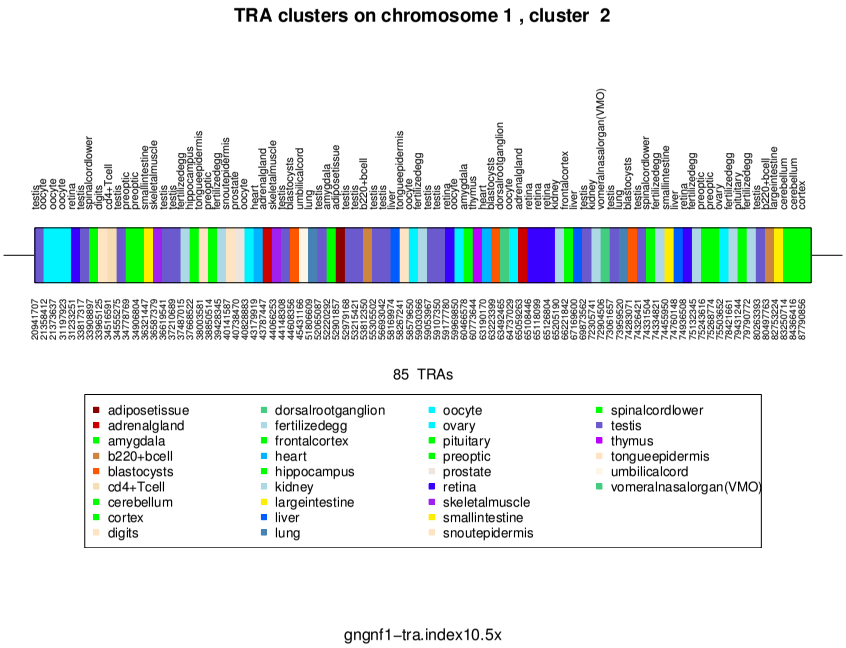


Figure 3.48: **TRAs in the evolutionary tree 2, chromosome 2, cluster 8.** Some times the gene order of TRA clusters is parallel between species, sometimes switched, parts of the genome is translocated to other regions as can be seen in the example of *Danio rerio*. In general gene order conservation gets lost the further away we come from the original species in this case the mouse.



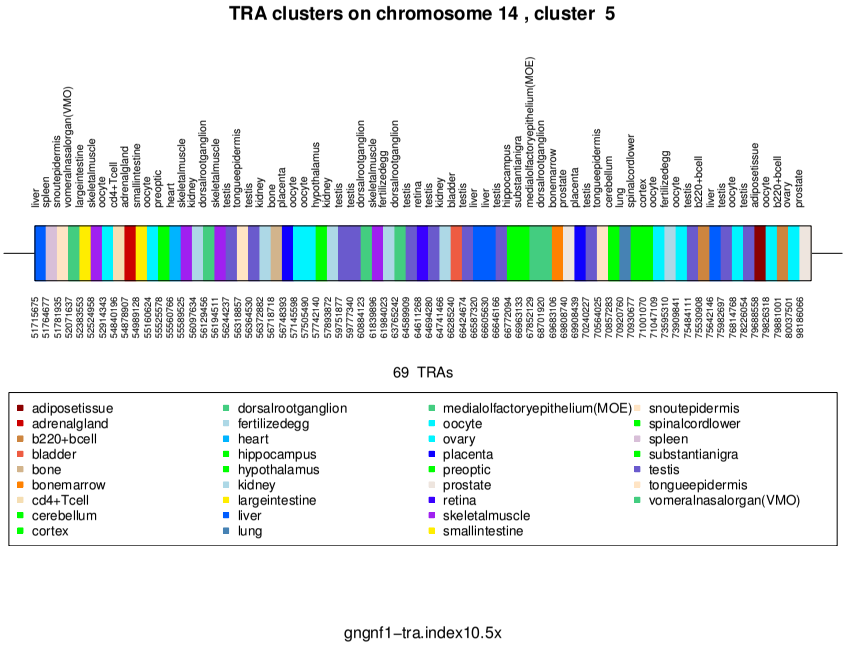


Figure 3.49: **Example of TRA clusters with great diversity in terms of tissue-specificity.** Most TRA clusters are a great intermingle of different tissue-types. The different colors represent TRAs tissue-specific for the different tissue-types.

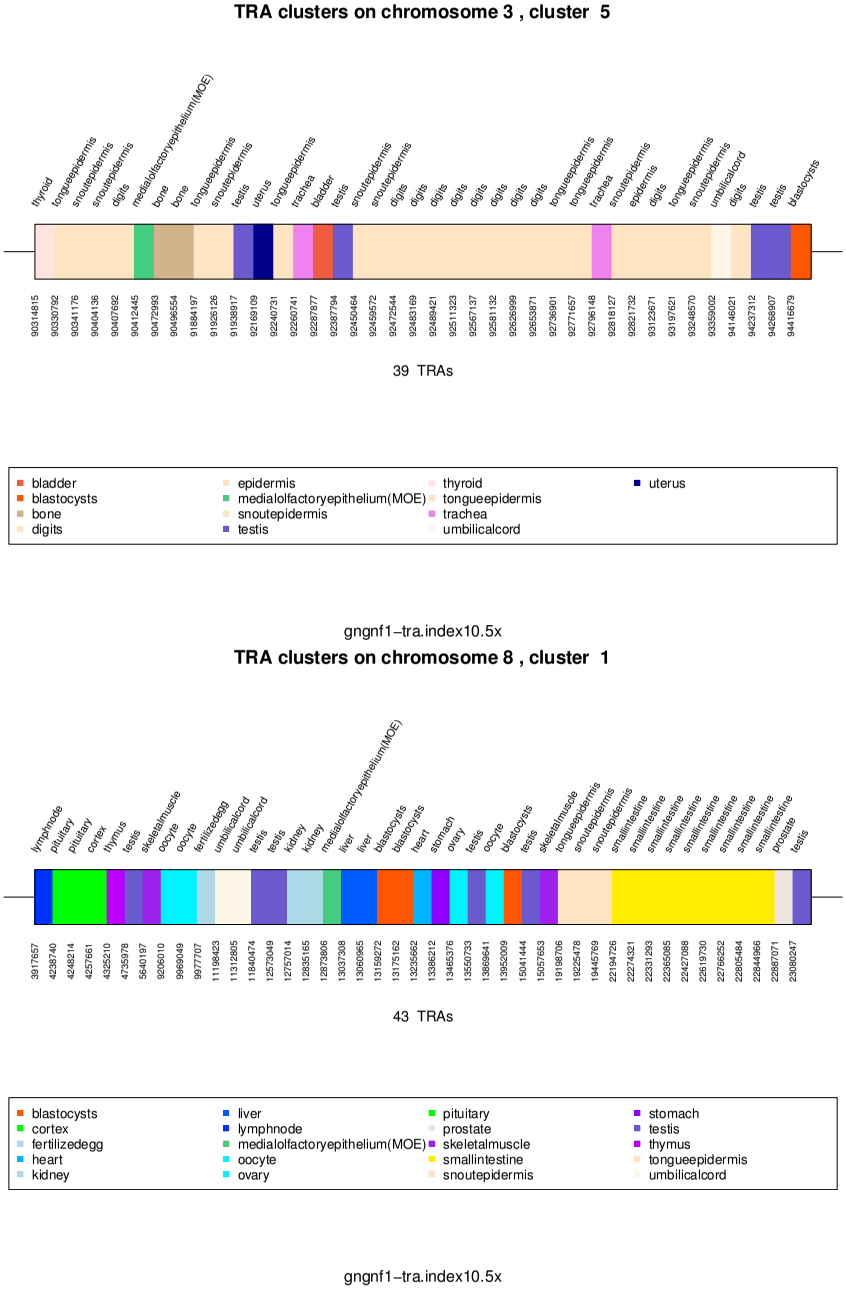


Figure 3.50: **Examples of TRA clusters with mono-clusters in terms of tissue-specificty.** These examples of TRA clusters show two cases of so called mono-clusters, where many TRAs within the cluster are tissue-specific for the same tissue-type. In the first case twenty-five out of thirty nine TRAs are tissue-specific for the skin, in the second example ten TRAs out of 43 TRAs in the cluster are tissue-specific for the small intestine. This picture of mono-clusters is rather the exception than the regular case.

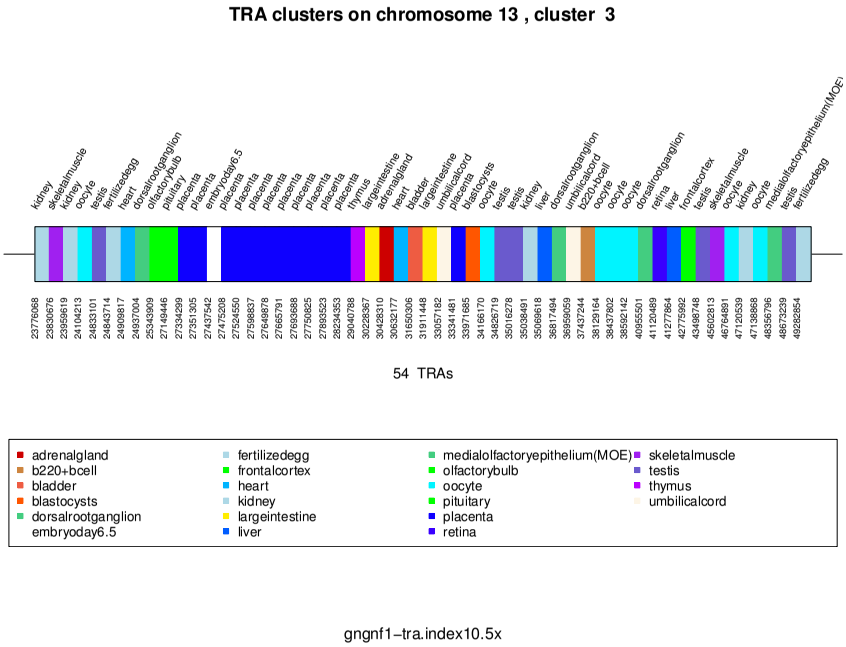


Figure 3.51: **Example of a TRA cluster tissue-specific for the placenta.** In the case of the TRA cluster number three on chromosome number thirteen in the mouse Novartis dataset, we see an example of a so called mono-cluster tissue-specific for the placenta, surrounded by TRAs tissue-specific for other tissues. In this case eleven out of fifty-four TRAs are tissue-specific for the placenta.

3.10 Intra-cluster homology analysis

In order to study, if the chromosomal clustering of TRAs is only due to gene duplication, we tested the sequence similarity of all TRAs within each cluster and calculated the percent identity of the genes per cluster. If there is no gene duplication the percent identity should vary around 25%. Looking at the homology plots (Fig. 3.52), we can see that the distribution of TRA clusters ordered by the size of the clusters varies around the 25%. While the smaller clusters start with a 40% range, the bigger ones are far below the 25% range. Some outlier dots reflect clear examples of gene duplications, but this is rather the exception than the norm. Both human and mouse Novartis datasets show similar results in this context. In order to test if this is dependent on the size of TRA clusters, we analyzed random clusters of different sizes with random genes, also these are grouped around the expected 25% range. This means that the number of gene duplications is not dependent on the size of the TRA cluster (Fig. 3.53). Some of the gene families found to be due to gene duplication are shown in Fig. 4.3, 4.4 and Fig. 4.5.

3.11 TRA-DB: database of tissue-restricted antigens

All TRAs found in this work were put into a database, TRA-DB. The database of tissue-restricted antigens (TRA-DB) can be found under: https://ibios.dkfz.de/tra/. Here genes can be queried upon their gene identifiers, chromosomes, tissue-types as well as according to different species. If genes belong to gene families all representatives of the gene families are found. The resulting query can be exported as a tab seperated file, with an annotated gene list, as well as a .pdf file including all TRA plots accross tissues and species. The background lists as well as TRA plots can be found on the CD in the technical appendix. Cutoffs for the TRA definition such as 3x, 5x, 10x as well as 20x the median can be chosen in the database. Also the number of genes exceeding the cutoff line can be chosen between one and five tissues over the 5x median line. Tissues of the same tissue-type, such as the central nervous system are regarded as one tissue-type and only counted as one (Fig. 3.1 - 3.4). The TRA-DB can be seen in Fig. 3.54 [102].

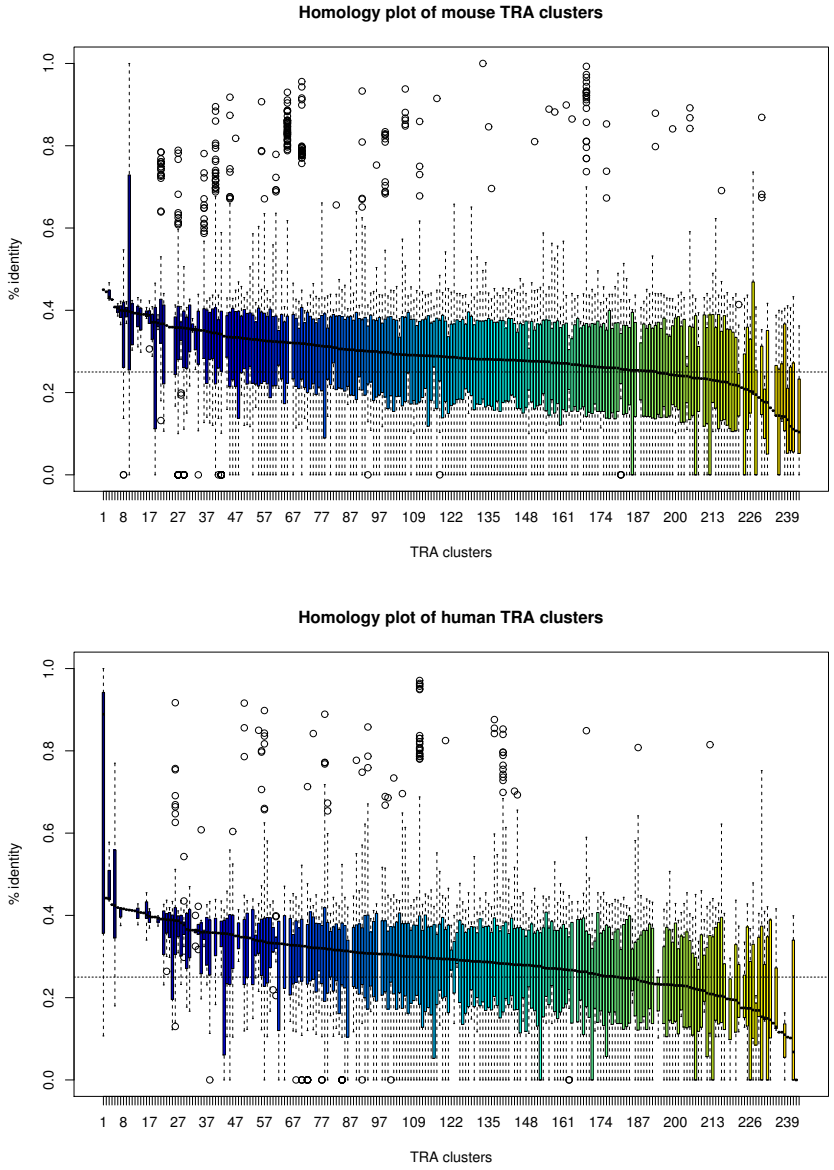


Figure 3.52: **Homology plot mouse and human.** Comparing all TRA clusters for the percentage of identity of TRAs within the clusters we can see that the main distribution of these clusters are scattered around the 25% line, which four base pairs are by nature. We ordered the TRA clusters according to their median line of % identity. The outliers (black dots on top) are the few TRA clusters which show a high % age of identity, these TRAs most likely belong to gene families. These plots have been recalculated and adapted from scripts from Prof. Dr. Benedikt Brors.

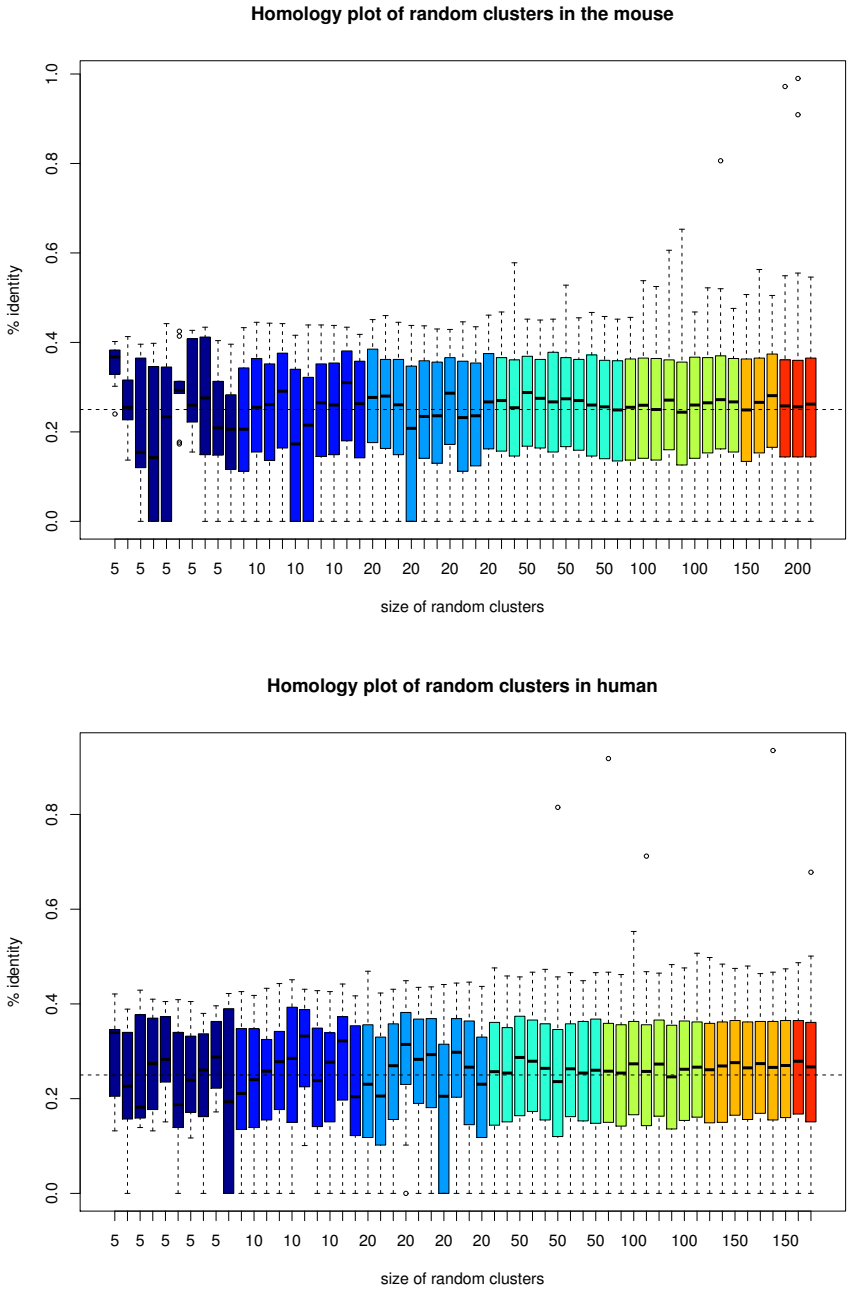


Figure 3.53: **Homology plot of random clusters in mouse and human.** The homology plot of TRA clusters was compared to randomly picked genes of different sizes, the same study of % identity was done to these clusters. Randomly picked clusters very well scatter along the % range of four base pairs, as would be expected. These plots have been recalculated from scripts from Prof. Dr. Benedikt Brors.

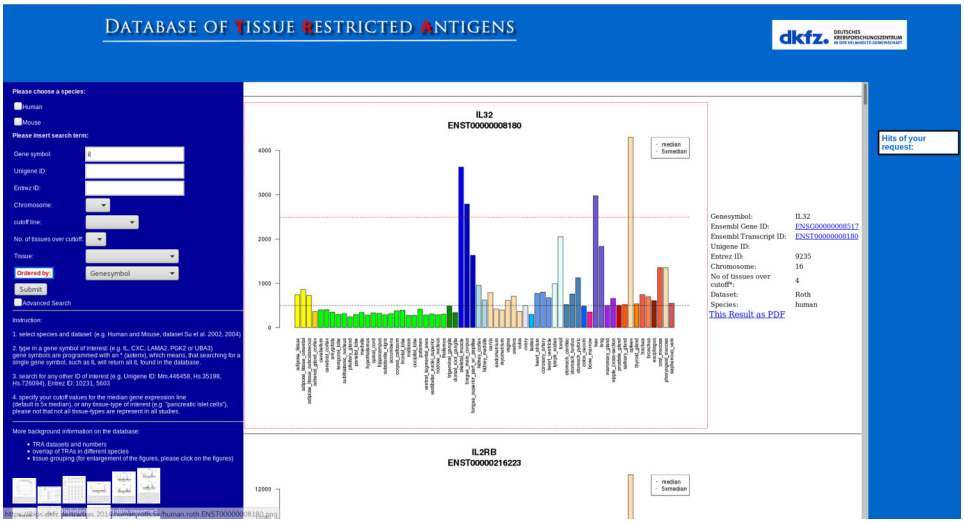


Figure 3.54: **The TRA database, TRA-DB.** The TRA database, TRA-DB is presenting all tissue-restricted antigens from all datasets. Here genes can be queried for datasets by gene names, tissue-types, gene identifiers, as well as different cutoff values. The background species can be selected upon the selection of different datasets and the results can be exported both as tables (.csv) files as well as plots (.pdf) files. Searches for gene names are programmed with an asterix, so that word completion for any word is automatically done. The database can be found under https://ibios.dkfz.de/tra/. This database has been re-calculated and re-established from a previous subversion from Dinkelacker 2007 [102].

4 Discussion

4.1 Chromosomal clustering of tissue-restricted antigens

In this study five different human and mouse datasets [371, 372, 218, 327, 3, 2] have been analyzed systematically for tissue-restricted antigens and chromosomal clustering of TRAs [102]. The outcoming TRAs have been studied in their known impact to autoimmune diseases. Furthermore the conservation in terms of gene order as well as chromosomal clustering of these TRA clusters has been studied in more detail. Within the TRA clusters new gene families could be identified, which have not be known before.

With an operational definition of tissue-restricted antigens as well as housekeeping genes, all TRAs as well as housekeeping genes have been identified for each dataset. Besides TRA clusters we also calculated the chromosomal clustering of housekeeping genes as a group, as well as subgroups of functionally related genes, such as cell cycle genes, glycolysis genes and cytoskeleton genes. All TRA plots for all different criteria, such as a cutoff value of 3x, 5x, 10x as well as 20x the median gene expression value, have been plotted and put into the TRA database, TRA-DB [102].

All TRA tables fully annotated can be found on CD in the technical appendix. TRAs tissue-specific for certain tissue-types have been extracted and analyzed for their known impact on tissue-specific autoimmune diseases. These lists have been analyzed on the most stringent cutoff criteria of only one tissue exceeding the 5x median gene expression line. Autoimmune relevant TRAs have been analyzed for the pancreatic islet cells, the thyroid, the central nervous system, the skeletal muscle, as well as the adrenal gland. Each in context of its equivalent autoimmune disease, such as diabetes type 1, hashimoto thyroiditis, multiple sclerosis, myasthenia gravis as well as addison's disease.

Futhermore testis-specific genes have been defined, which might play a role in the development and treatment of cancer as cancer testis antigens (CTAs). Here 440 different CTAs have been identified, of previously 225 CTAs known [425]. In the question of cancer treatment of immune vaccination against CTAs the gene expression of CTAs within the thymus has to be thoughtfully checked [75, 216, 41, 298]. Since a previously existing tolerance of T cells versus the potential antigens will decline the adaptation of the immune system to the cancer type and thus most probable lead to a failure of immune vaccination in cancer.

In our TRA list, we also analyzed tissue specific chemokines, as homing factors, which could be extracted via their identifiers. Chemokine ligands

as well as receptors follow a systematic nomenclature of CC-, CXC-, CX3C-chemokine ligands and receptors depending on the cystein residues in their N-terminus. Since our TRA database is programmed with an asterix, these gene lists could be extracted easily and be analyzed upon their tissue-type of main gene expression, which is in the context of homing factors for immune cells of chemokine ligands as well as their receptors very interesting. Also in the context of chemokine signalling in cancer and its migration of the tumor to the site of metastasis, depending on the type of chemokine ligand and receptor profile expressed in the tumor type.

For the definition of a TRA we used an operational definition, which can be adjusted depending on the scientific question by varying the criteria in the database. TRA calculation was mainly done on the basis of transcript level in all datasets, where this was possible. In some datasets, the transcript level was not available (e.g. in the protein data of the human protein atlas). We decided for this criteria in order to be able to detect TRAs with tissue-specificity including alternative splicing events [102].

The overlap of all TRAs between datasets of the same species was calculated, furthermore the impact on gene regulation of all TRAs defound here by the autoimmune regulator AIRE was calculated (data not shown here).

Chromosomal clustering of TRAs has been shown to be significantly in all datasets and for all criteria with two different methods. TRA clustering can explain a common regulation in gene expression of all tissue-restricted antigens in one cell type, such the medullary thymic epithelial cells. Co-regulation of those has been shown in the past for a few subgroups of TRA clusters, including certain gene families, more gene families could be found here.

It could furthermore be shown, that gene families are not the main, nor only reason for TRA clustering, and gene duplication can not be the reason for TRA clustering in general, since the homology plots calculated for all TRA clusters range around the expected 25% range of % identity for all genes, within each TRA cluster.

4.2 The selection of datasets and databases

As datasets we have chosen five different datasets, two in mouse and two in human, as well as one RNAseq human GTEX dataset. While the human and mouse Novartis dataset has only double measurements per tissue-type, and the human Roth dataset has a sample size of $n<4$, the human GTEX dataset includes measurements of up to 800 samples of the same tissue-type. Also the number of genes per chip varies due to the date it has been estab-

lished from 14,522 genes in 2002 up to 57,431 genes in 2015. The number of TRAs varies as well, but the calculation of overlaps in TRA definition brings the pool of potential tissue-restricted antigens back to a sufficient overlap of genes in all datasets. The core TRAs, which have been previously brought into context of autoimmune diseases, can be found in any dataset. In order to find tissue-restricted antigens in the dataset of the human protein atlas [390], the criteria of 5x the median cutoff did not proof to be well suited. This was drawn to the fact, that even well known autoantibodies, such as insulin could not be detected this way. We loosened the cutoff value up a little, to in this case 10x the median protein expression, in order to find the TRAs expected, such as insulin.

In the future this study might be extended to the FANTOM database [81] as well as to single cell data. In the case of single cell data, we will have to deal with the problem of drop outs, which will also be a problem considering the calculation of TRAs. Cell specific TRAs might be interesting though, and the same method of finding TRAs can also be applied here.

4.3 Number of TRAs and proportion

In our datasets we could identify 4,172 (24.6%) genes to be tissue-restricted in the mouse Novartis dataset [371, 372], 2,055 (14.15%) in the human Novartis dataset [371, 372], 8,924 (52.25%) in the mouse Lattin dataset [218], 6,515 (31.36%) in the human Roth dataset [327] and 27,339 (47.60%) in the human GTEX dataset [3, 2]. While the raising number of TRAs over time can be explained by the increasing number of genes being represented on the chips between 2002 and 2015, the percentage depends on the number and quality of probes on the chips. The number of TRAs is also dependent on the cutoff as well as the number of tissues represented on the chips, including the different tissue-groups existent per dataset. While the first mouse Novartis dataset covered about 16,960 genes in total, the human Novartis dataset represented the gene expresson of 14,522 genes in total, the mouse Lattin dataset 17,079 genes, the human Roth dataset 20,774 genes, and the human GTEX dataset 57,431 genes, including non-coding RNAs. As a calculation done in the human protein atlas data [390], we could find 7,241 (35.59%) of 20,343 proteins to be tissue-restricted. Comparing all TRAs, we found 1,091 TRAs to be commonly detected in all three human datasets, and 2,837 TRAs in the mouse datasets. Including the human protein atlas, we found 610 different TRAs in all human datasets (Fig. 4.2). The number of TRAs found is not only dependent on the tissues presented, the number of genes in general on the dataset, but also on the number of datasets calculated. The more datasets are involved, the smaller the number of common TRAs. Doing calculations on the overlap between orthologues human and mouse TRAs, we find a high concordance also between species.

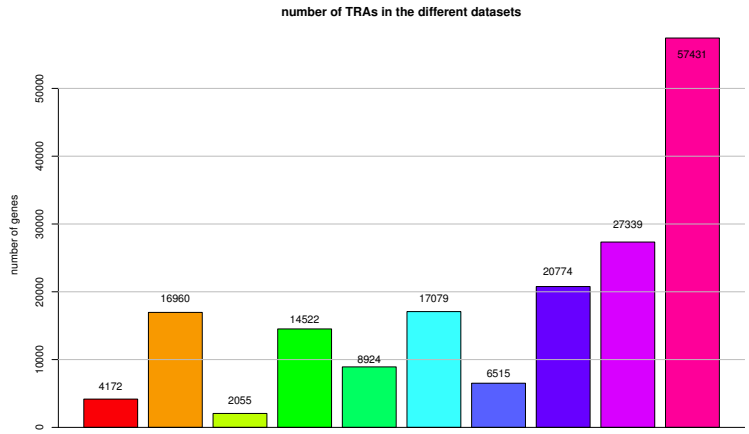


Figure 4.1: **TRA numbers and percentages in the different datasets.** In different datasets we find different numbers of tissue-restricted antigens by applying the same method for the detection of tissue-restricted antigens in a systematic way. Most numbers are dependent on the number of genes presented on the chips in the different datasets. In the mouse Novartis dataset (red, orange), we find 4,172 TRAs (red) out of 16,960 background genes on the chips (24.6%), in the human Novartis dataset (light green, green), we find 2,055 TRAs out of a total of 14,522 background genes (14,52%) on the chip, in the mouse Lattin dataset (light green, light blue), we find 8,924 TRAs out of 17,079 background genes (52.25%) on the chips, in the human Roth dataset (blue bars), we find 6,515 out of 20,77 genes (31.36%) on the chip, in the human GTEX RNAseq dataset (purple bars) we find 27,339 TRAs out of 57,431 genes (47.60%) on the chip.

Within our TRA lists, we can find almost all TRAs which have been previously identified. We could find for example the insulin gene (INS) [184], the thyroglobulin gene (TG) [202], the C-reactive protein (CRP) [202], SAP [202], RET S and many other tissue-restricted genes known to be important in autoimmune diseases [144]. Much of the data presented here has already been used in the context of our cooperation projects for example in the work of Derbinski et al. 2005, Pinto et al. 2013 as well as Brennecke et al. 2015 [99, 297, 46]. Besides this previous knowledge and pre-published data, we can detect many more and newly discovered TRAs in this work, which sorted to the tissues specifically expressed.

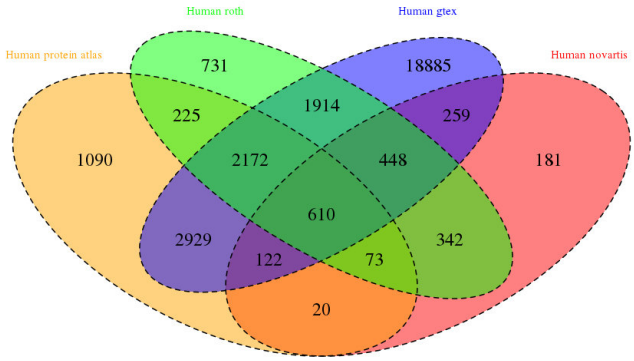


Figure 4.2: **Overlap of TRAs in four different human datasets.** In the study of four different human datasets, two microarray datasets [372, 327], one RNAseq dataset [3, 2] and the human protein atlas [390] we can calculated an overlap of 610 TRAs which can be detected in all four datasets. The number is restricted to the lowest background number of genes presented per chips. Due to the high number of genes in the human GTEX RNAseq dataset, the human GTEX dataset is also the one with the highest number of 18,885 genes which can only be found to be tissue-restricted in their dataset. The human Novartis dataset is the oldest dataset used in this study and therefore only shows 181 TRAs uniquely detected in this dataset. All in all the sum of all TRAs might lead more to the important results in this study than trying to be too restricted. Also because not all genes are presented in each tissue, nor is every tissue represented in each dataset.

4.4 Tissue-grouping, TRAs and housekeeping genes

In our data, we have excluded several cell types from our study, such as immune cells, embryonic tissues, as well as cancer cell lines. The scientific question of interest had been only related to tissue-restricted antigens in the context of the induction of central-self tolerance in the thymus. We incorporated the expression estimates in the excluded cell types as well as in cell lines still in the TRA plots, in order the visualize for example immune cell contribution to tissue-restricted expression in secondary lymphoid organs.

4.5 TRA clustering

For TRAs in all datasets, we calculated the chromosomal clustering with two different methods [144, 328]. While the first sliding 10-gene window method does not account for gene-dense and gene-poor regions, the second method with a sliding gene window of fixed size also takes this aspect into account. For both methods we could demonstrate significant chromosomal clustering of TRAs in all datasets. The cluster sizes were compared to 1,000 randomly drawn lists of the same size, and calculated over all chromosomes. Within the different datasets, we could find gene clusters including up to 251 TRAs in a direct neighborhood based on the human Roth dataset, 143 based on the mouse Lattin dataset, 22 based on the human Novartis dataset and 87 based on the mouse Novartis dataset [102].

Within these TRA clusters we could identify gene families to contribute to the gene clusters, such as the KLK cluster, the casein cluster, of the S100 cluster. But also new TRA clusters, which have not been identified before could be identified. A gene family was considered to be significantly contributing to a cluster if more than five members of the gene family were found within one TRA cluster (sliding 10-gene window method). Complete lists of all TRA clusters are in the technical appendix on CD. Examples of these clusters can be seen in Fig. 4.3 and Fig. 4.4. The fact that gene families arise from gene duplication does not account for TRA clustering in general as we could show by calculating the percent identity between all genes of a cluster (Fig. 3.52 and Fig. 3.53).

Among the gene families found in TRA clusters, we could identify the AKR1c genes, a group of genes, which catalyzes aldehyde and were tissue-specific for the kidney, the liver as well as the adrenal gland in the mouse, the CEACAM genes on chromosome 7, a family which is important for cell adhesion, mostly found in epithelial cells throughout the body, the PSG gene family on chromomosome 7 in the mouse, which is tissue-specific for the placenta, pregnancy specific glycoproteins, the CYP genes on chromomosome 19 in the mouse, belonging to the cytochrome p450 gene family, working

as enzymes in the intestine as well as the liver the MS4A gene family on chromosome 19 in the mouse, which is tissue-specific for the testis as well as many others, shown in Fig. 4.3, 4.4 and in the cluster tables on CD.

4.5.1 The kallikrein gene cluster

The KLK cluster, the S100 cluster as well as the casein cluster had already been studied before [99, 297, 46]. Therefore we analyzed these pre-existing clusters in our dataset. We could find the three KLK genes, KLK1, KLK2 as well as KLK3 in the human Novartis dataset to be tissue-restricted (Fig. 4.5). The up and down regulation of these genes in human mTECs has been measured by Pinto et al. 2008 and Gaertner et al. 2012 (unpublished data). Not all of these genes are always upregulated in mTECs.

In the mouse Novartis data, we find 10 KLK genes in a TRA cluster on chromosome 7. All of them are upregulated in mTECs versus cTECs in the thymus. The co-regulation of genes within one TRA clusters has convincinvly been demonstrated [46]. Some of these genes are represented by two transcripts for example the once presenting the genes *Klk1b1* to *Klk7*, *Klk8*, *Klk10*, *Klk11*. Interestingly these genes in our data seem to be tissue-specific for the thyroid (data not shown here).

In the human Roth dataset, we found the KLK genes within a TRA cluster of 93 TRAs on chromosome 19. These are tissue-specific for the salivary gland, the prostate gland, the oral mucosa, the central nervous system and the vulva, intermingled with TRAs that are tissue-specific for the testis. The KLK genes seem to be a very heterogeneous group in terms of tissue-types they represent.

In the human GTEX data, we find 15 KLK genes on chromosome 19 with 64 different transcripts. All of them are tissue-specific for different tissue-types. Some of these genes are upregulated in the thymus, include AIRE-dependent genes, and are highly expressed in mature CD 80 hi MHC II hi mTECs. Others are down-regulated between mTECs compared to cTECs and are expressed in the thymus. For the GTEX data, we have not calculated TRA clustering, so that the state of KLK genes within the TRA clusters cannot clearly be described.

symbol	**startside**	**tissue**	**mTECs vs cTECs**	**MHC II hi vs lo mTECs**
KLK1 (1 gene, 5 transcripts)	51322404	Pancreas	-0.25 (0.61) 1.55 (<0.01)	2.64 (<0.01)
KLK2 (1 gene, 18 transcripts)	51364824	Prostate	-0.003 (0.997)	
KLK3 (1 gene, 13 transcripts)	51358171		-0.11 (0.84)	

			0.09 (0.89) 0.12 (0.82)	

Table 4.1: **KLK cluster human Novartis dataset**, chromosome 19, 3 genes, 36 transcripts, mTECs versus cTECs (down/upregulated gene expression, adj. p-value), data from Martina Gaertner 2012 (AIRE, 3.897 (<0.01)/1.86 (<0.01)/0.11 (0.79)), MHC II hi versus MHC II lo AIRE dep. (down/up regulated gene expression, adj. p-value), data from Sheena Pinto 2008 (AIRE 1.39 (<0.01))

4.5.2 The S100 cluster

In the case of the S100 gene cluster, we find seven S100 genes on chromosome 1 in the human Novartis dataset (Table 4.2, Fig. 4.6). Seven of these genes code for 27 different transcripts. It has to be noted, that the number of transcripts is sometimes redundant to the same coding genes, so that transcript identifiers are not always representing true alternative reading frames. Most of the S100 genes are tissue-specific for the heart, bronchial epithelial cells or immune cells. Most of these genes are upregulated in the thymus. Some of them are AIRE-dependent.

We can also find nine different S100 genes on chromosome 3 in the mouse Novartis dataset, which are tissue-specific for the skin within a TRA cluster of sixty different genes. Within the same TRA cluster we can find LCE genes, which also form a gene family within the same TRA cluster (data not shown here).

In the human GTEX data, we found 15 S100 genes on chromosome 1, most of them differentially expressed in the thymus (data not shown here).

symbol	**startside**	**tissue**	**mTECs vs cTECs**	**MHC II hi vs lo mTECs**
S100A1 (1 gene, 3 transcripts)	153600402	heart	-0.21 (0.69)	-1.08 (<0.01)
S100A2 (1 gene, 6 transcripts)		bronchial epithelial cells	3.19 (0.01)	
S100A4 (1 gene, 6 transcripts)	153516089	immune cells	0.50 (0.27) 0.53 (0.36)	
S100A6 (1 gene, 5 transcripts)	153507075	bronchial epithelial cells	2.78 (<0.01)	
S100A9	153330330	immune cells	1.14(0.02)	4.22 (<0.01)

S100A10 (1 gene, 4 transcripts)	151955391	bronchial epithelial cells	1.20 (0.02)	2.23 (<0.01)
			1.14 (0.02)	
S100A12	153346184	immune cells	-1.19 (0.31)	

Table 4.2: **S100 cluster human Novartis dataset**, chromosone 1, 7 genes, 24 transcripts, mTECs versus cTECs (down/upregulated gene expression, adj. p-value), data from Martina Gaertner 2012 (AIRE, 3.897 (<0.01)/1.86 (<0.01)/0.11 (0.79)), MHC II hi versus MHC II lo AIRE dep. (down/upregulated gene expression, adj. p-value), data from Sheena Pinto 2008 (AIRE 1.39 (<0.01))

4.5.3 The casein cluster

A third gene cluster, that has been described is the casein cluster on chromomosome five in the mouse. We find five different representatives of *CSN1S1*, *CSN1S2a*, *CSN1S2b*, *CSN2* and *CSN3*, including seven different transcripts, most of these genes are upregulated in the thymus (Table 4.3). All casein genes are tissue-specific for the mammary gland (lact.).

symbol	**startside**	**tissue**	**mTECs vs cTECs**	**MHC II hi vs lo mTECs**
Csn1s1	87666224	mammary gland (lact)	1.88 (0.01)	1.94 (<0.01)
			1.58 (0.01)	
Csn1s2a	87774567			
Csn1s2b (1 gene, 2 transcripts)	87808082			
Csn2	87692624		1.34 (0.11)	-1.30 (<0.01)
Csn3 (1 gene, 2 transcripts)	87925579		0.43 (0.44)	1.28 (0.00)

Table 4.3: **Csn cluster mouse Novartis dataset**, chromosone 5, 5 genes, 7 transcripts, mTECs versus cTECs (down/upregulated gene expression, adj. p-value), data from Martina Gaertner 2012 (AIRE, 3.897 (<0.01)/1.86 (<0.01)/0.11 (0.79)), MHC II hi versus MHC II lo AIRE dep. (down/upregulated gene expression, adj. p-value), data from Sheena Pinto 2008 (AIRE 1.39 (<0.01))

While it has been an ongoing discussion if only tissue-specific genes or only housekeeping genes are clustered, we enlargened our analysis also on other functionally related gene groups and calculated the chromosomal clus-

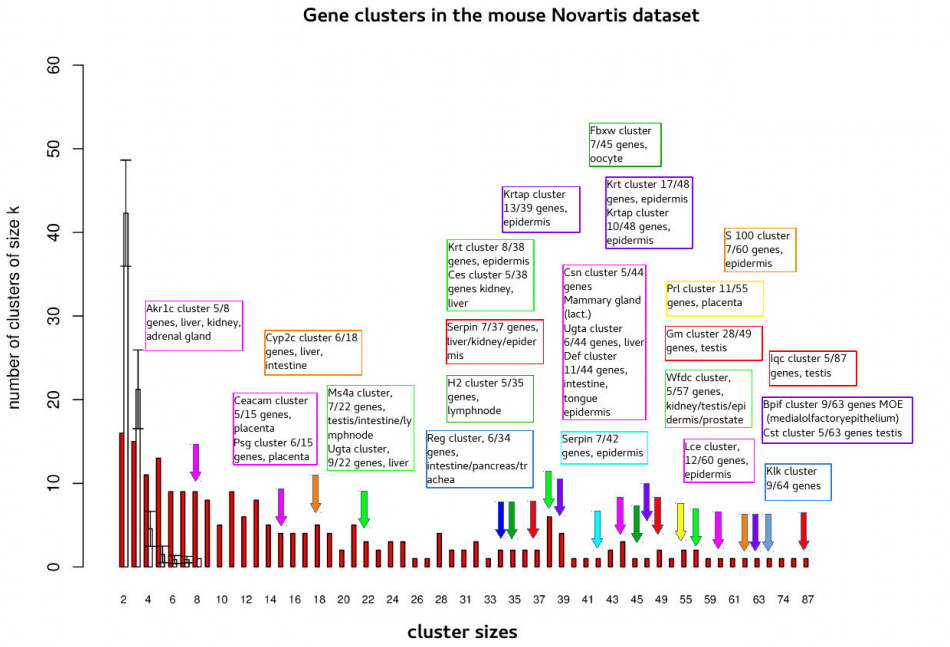


Figure 4.3: **Commonly regulated gene families in TRA clusters.** In the mouse Novartis dataset, we can find many gene families wich are most probably commonly regulated within the same TRA clusters. Some of these gene families have already been known, such as the klk cluster on chromosome 7 in the mouse, the S100 cluster in chromomosome 3 in the mouse and the casein locus on chromosome 5 in the mouse. The more detailed description of these genes and clusters can be found in the text as well as in Fig. 4.4 Different clusters are depicted with the arrows in the same color as the clusters described in the box. The clusters are sorted according to their size of the background TRA cluster. In some cases different gene families fall into the same TRA cluster. Not all TRA clusters are due to gene families most of them are an intermingle of different gene families as well as tissue-types. Gene families might have evolved due to gene duplication.

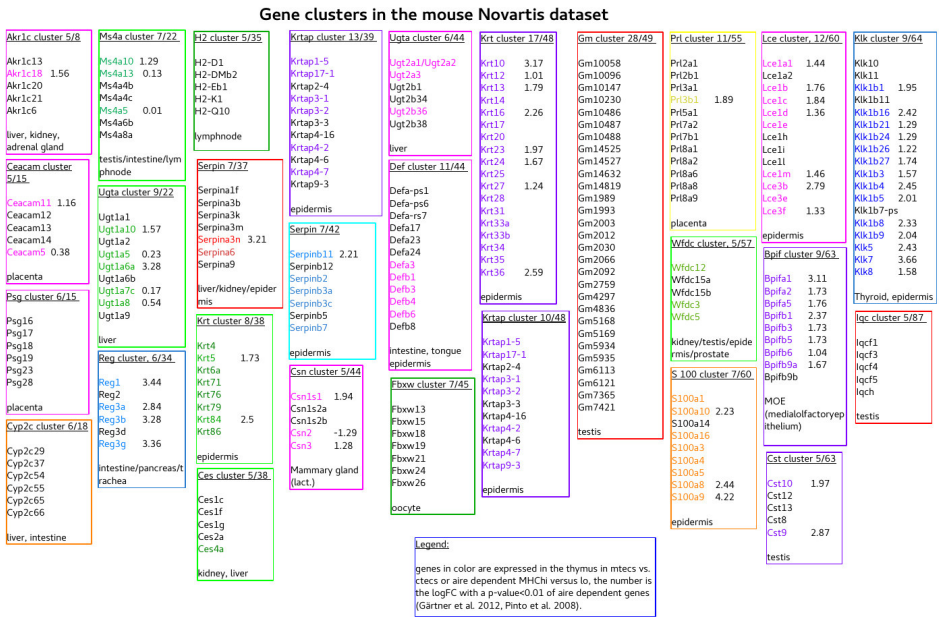


Figure 4.4: **Gene families within TRA clusters.** In the mouse Novartis dataset we could identify twenty-eigth different gene families of sizes higher than five within our TRA clusters. Some of these gene families have partly been known previously, such as the S100 gene family, the klk gene family as well as the casein locus. Genes which are upregulated in the thymus are shown here in color, this most probable gives an insight into a common regulatory mechanism on gene expression. Most members of a gene family are tissue-specific for the same tissue-type. The biggest gene family is the Gm gene family, tissue-specific for the testis. Log foldchanges as well as p-values have been measured by Martina Gaertner et al. 2012 (cTEC versus mTEC data) and Pinto et al. 2008 AIRE ko versus wt data in MHC II hi CD 80 hi mTECs. Further study on gene families especially in the context of autoimmune diseases might be very fruitful in the future.

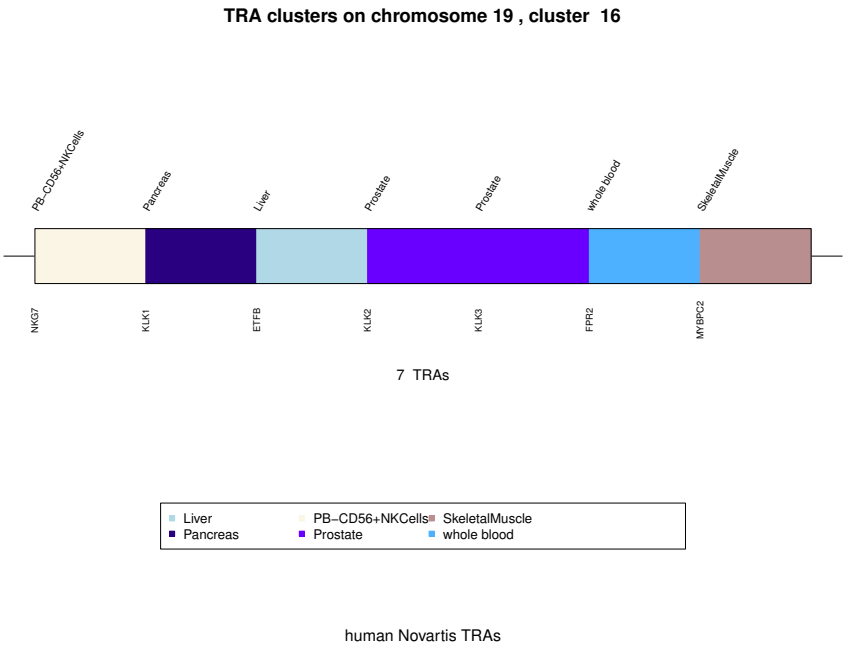


Figure 4.5: **Klk cluster in the human Novartis dataset.** In the human Novartis dataset, we can find the Klk gene cluster on chromomosome 19, where tree out of seven TRAs are kallikreins, tissue specific for the pancreas as well as the prostate (blue bars). The kallikrein cluster had been previously been detected by Derbinski et al. 2005 and further studied by Brennecke et al. 2015 [99, 46]. Both kallikreins as well as TRAs within a cluster seem to be co-regulated in the thymus. In the mouse Novartis dataset we find seventeen different transcripts coding for kallikreins. Most of them are tissue-specific for the thyroid klk1, others are tissue-specific for the snout-epidermis klk 5,7,8,10,11.

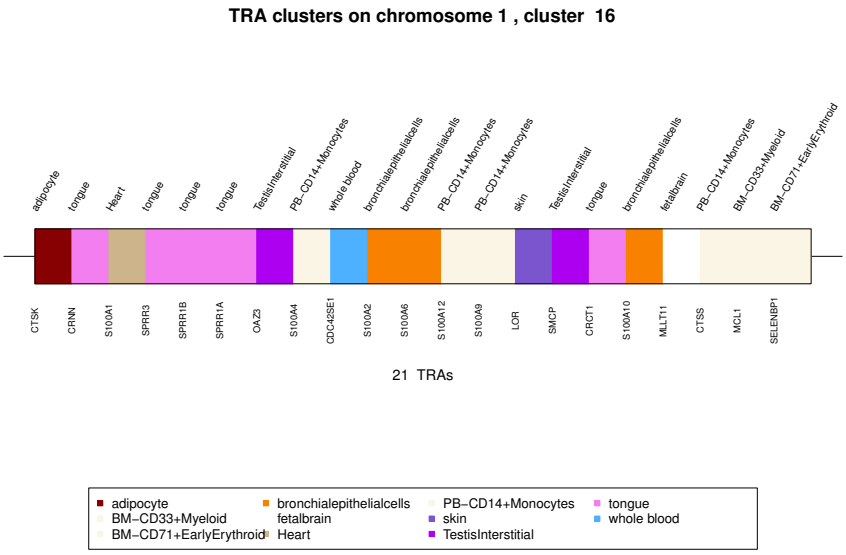


Figure 4.6: **S100 TRA cluster in the human Novatis dataset.** The S100 cluster can be found on chromosome 1 in the human Novartis dataset. Most S100 genes are tissue-specific for the tongue as well as other epithelial cells. They are located in a TRA cluster in this dataset including 21 different TRAs. The Immune cells have been plotted as a tissue-type here but were not considered for TRA calculation. Most S100 genes are also upregulated in the thymus.

tering of these. For this we took all housekeeping genes involved in our data, by defining a housekeeping gene as a gene, which is not higher expressed in any tissue than 3x the median gene expression over all tissues. Also here we could show chromosomal clustering of housekeeping genes in general but also some other gene groups, such as cytoskeleton genes, apoptosis genes and cell cycle genes. All of them are significantly clustered on the chromosome, as well as muscle-specific genes, actin cytoskeleton genes which are highly enriched in gene duplets in direct neighborhood. Some gene groups could not be shown to be clustered, such as seventeen genes involved in the tricarboxylic acid cycle (TCA), thirty nine glycolysis genes as well as fourty-one caspase genes. This might be due to the low number of genes involved in these groups.

4.6 Housekeeping genes numbers and percentages

In terms of the number of housekeeping genes as defined above, we determined the number of housekeeping genes in each dataset. We found 8,692 housekeeping genes in the mouse Novartis dataset, 9,938 in the human Novartis dataset, 2,863 in the mouse Lattin dataset and 9,148 in the human Roth dataset. Also the number of housekeeping genes have exceeded former attempts to define and find housekeeping genes. As has been discussed in the section about tissue-restricted antigens already, also the definition of a housekeeping gene to be a gene not to be expressed higher than 3x the median in any tissue is only an operational definition. Looking into the lists of housekeeping genes however this definition similar to the definition of tissue-restricted antigens seems to match the type of genes expected very well. We can find many genes involved in typical housekeeping functions.

4.7 Former attempts to find tissue-restricted antigens and define housekeeping genes

The previous work of determining TRAs has been mainly done by Derbinski et al. 2005 [99], who has been involved in the design of this study. In their work they defined TRAs to be genes, which are expressed in less than five tissues out of 45 in the mouse and this way found about 28% of all genes being upregulated in mTECs versus cTECs to be tissue-restricted [99, 102]. In their work TRAs have been hand selected. In their study they could identify 152 TRAs to be upregulated in mTECs [99, 102].

Werdelin et al. wrote in 1998 *"the number of tissue-specific proteins encoded by the genome may well be higher than the number of household proteins"* [424] and this is based upon the knowledge of gene numbers in general. This might explain why earlier estimates of the number of tissue-restricted antigens as well as housekeeping genes have been much smaller in general.

Eisenberg et al. have identified in 2003 a number of 575 housekeeping genes by defining a housekeeping genes as a gene, which is in their desription expressed *"constitutionally"* [111], whatever this means exactly in the context of housekeeping gene calculation.

Eisenberg et al. also used for this calculation the human as well as the mouse Novartis dataset from Su et al. 2002 and 2004 [371, 372], so the same dataset, but our number is much higher, which cannot only be explained by the way we annotated the probes. Zhang et al. defined a housekeeping gene in 2004 as a gene, *"which is always expressed in every tissue to maintain the cellular function"* [455] also this definition is not further clarified. Watson et al. had already thought about this definition in 1965 for the first time. Hurst et al. then later specified in 2002 their definition of a housekeeping gene as gene, which is expressed in 9 out of 14 different tissues, and Zhang et al. as well as Lercher et al. further specified this definition as a gene, which is expressed in at least 19 out of 60 tissues [455, 222]. But none of these studies considered tissue grouping as an aspect in the context of defining a gene to be tissue-specific, or a housekeeping gene. Since our definition of a housekeeping gene is considering a gene to be a housekeeping gene if the gene is not expressed higher than 3x the median gene expression over all tissues in any of the tissues, this might not fall into account in the context of housekeeping genes, but it matters a lot in the context of defining tissue-restricted antigens. Since most of the datasets have a strong bias towards many tissues of the central nervous system (CNS) this might influence the outcome a lot. Also all of these authors never really classified what the term "is expressed" or is "higher expressed as" really means, given the inherent variance in genome-wide measurements of gene expression.

4.8 A TRA database (TRA-DB)

We have established a database of tissue-restricted antigens, where each TRA can be searched by gene identifiers, tissue-restricted gene expression, species as well as for different expression criteria. The database can be found under https://ibios.dkfz.de/tra/ and provides both gene lists, as well as TRA plots, which can be exported from here via a comma seperated value file (.csv), or as pdf [102]. The database was established on the basis of transcript level, if possible, and serves as a ressource of all TRAs mainly in the context of autoimmune diseases. The database can be easily updated for newer data; the same method as previously used can be applied and also summed up with the gene expression of genes in the thymus. The correlation with AIRE regulated genes, versus AIRE independent genes, mTEC versus cTEC data as well as MHC II hi CD 80 hi versus undifferentiated data would be nice to be added in the future, in order to fill the knowledge gap between theoretically determined TRAs and true gene expression

of those in the thymus. The data is already existant we would only have to add a little bit of programming work here. The TRA database might help in order to find more autoimmune-related genes as well as for example more genes related to cancer in terms of tissue-specificity as well as for example cancer-testis antigens (CTAs).

Up to my knowledge there has never been a database on tissue-restricted antigens in this completion as well as a database, which is calculated on the basis of transcript level, rather than on gene level. There are databases on tissue-restricted genes, such as TIGER [233] as well as TissGDB [200] but none of them fullfills the same criteria and specificity in depth as our attempt, the scientific questions in this context also did not come from the question of central self tolerance, nor autoimmune diseases. Furthermore the database can also be used in the context of finding new drug targets, using wild card search, thus querying for a gene name A will also give the user all genes followed up by different numbers as well as characters, thus searching for gene families can be done very easily. This is also very helpful for finding new cancer testis antigens (CTAs), which often belong to gene families. Since not all datasets have been incooperated in the database yet, our gene lists were mostly calculated from our original TRA lists, which can be found in the technical appendix on CD. The annotation was done always on the newest standart of Ensembl Biomart at that time. Some of our TRA lists have been used previously in publications, such as Brennecke et al. 2015, Pinto et al. 2013 and Rattay et al. 2016 [46, 297, 311].

4.9 TRAs in autoimmune diseases

Since tissue-restricted antigens are known to be involved in autoimmune diseases, we searched for all prior known TRAs in our database. We could confirm most of the prior knowledge but also conclude some prior knowledge to be wrong. Especially the conclusion of certain tissue-types quite often seems to be confound in the previously published data. In our datasets, we could find, the insulin gene (INS) [144], glucagon (GLC), the pancreatic polypeptide (PPY), Somatostatin, Trypsin, the Elastase [184], we could find the myelin basic protein (MBP) [144], Albumin, thyroglobulin (TG), the thyroid peroxidase (TPO), the S100 family, CRP as well as SAP [202], PLP [203], IA-2 as well as MOG [98] but we could not find for example the genes GAD65 as well as GAD67 [366], nor the Acetylcholreceptor [333]. Instead we could find the acetylcholinesterase instead (TRA-DB). Furthermore we could find many TRAs to be tissue-specific for other tissues than previously expected as well as many cancer testis antigens being expressed in the bone marrow. Thus also for the negative control of previously found TRAs our database is suited well. The more exact findings are shown in the tissue-specific gene tables in the results part of this work. Examples of these are

for example the gene C9 previously connected by Nyalwidhe et al. 2017 to diabetes typ 1 [274], which is in our data tissue-specific for the liver, but not the pancreas as well as the gene ERBB3 found by Reddy et al. 2011 [312], which is in our data tissue-specific for the large and small intestine and the bladder but not for the pancreas.

For diabetes type 1, we could identify more than 70 different TRAs to be tissue-restricted for the pancreas and more specifically the pancreatic islet cells, many of them not previously known (Table 3.2), for myasthenia gravis (MG) we could identify more than one-hundred TRAs, which are tissue-specific for the sceletal muscle (Table 3.3), for multiple sclerosis we could identify 518 different TRAs tissue-specific for the central nervous system (CNS) among them also genes newly identified, such as AQP4 [177] (Table 3.4), for addisons disease we found more than one-hundred different genes tissue-specific for the adrenal gland, among them for example the gene CYP11A1, MRAP as well as STAR [118] (Table 3.5), for hashimoto thyroiditis we could find ninety-six TRAs tissue-specific for the thyroid, including all previously known genes, such as thyroglobulin (TG), the thyroid peroxidase (TPO) (Table 3.7), but also here we could prove many previously published genes to be wrong, such as PTPN22 and ITGAM which are tissue-specific for the bone marrow, but not the thyroid gland. Furthermore TRAs which are tissue-specific for the thyroid might also be involved in autoimmune grave's disease and not only in hashimoto thyroiditis [287, 257]. We could also find the S100 gene family [158] as shown earlier, if they are really involved as previously thought in autoimmune juvenile idiopathic arthritis (JIA) as has been thought earlier, remains to be clearified.

All in all we could identify 4,172 new tissue restricted antigens in the mouse Novartis dataset, 2,055 in the human Novartis dataset and many more in the other datasets analyed above. Upon previous annotation we can clearly state, which tissue these genes are mostly expressed in (maximum gene expression value and tissue type, max tiss), we can also say in which these genes are also expressed in, which might be important for example for the prediction of potential side effects on future drug targets. All these TRAs can now be sorted by tissue-type and then be analyzed in the context of tissue-specificity, AIRE regulation and autoimmune diseases. By looking at Fig. 4.7, we can clearly see that many if not all tissues of the body are represented by tissue-restricted antigens and most of them are upregulated by AIRE and expressed in medullary thymic epithelial cells (data not shown here). Most TRAs can be found to be tissue-specific for the testis, as well as the oocyte (Fig. 4.7), directly followed by TRAs tissue-specific for the liver, skeletal muscle, kidney, dorsalroot ganglion and fertilized egg.

Why these tissues have so many representatives compared to others still

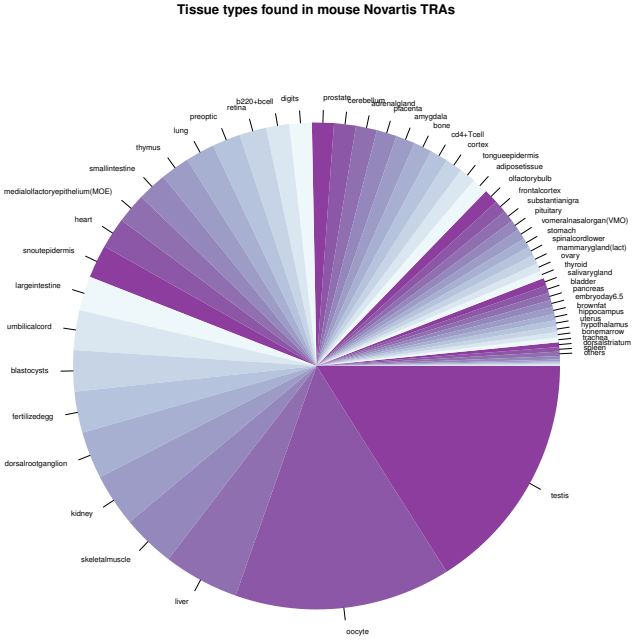


Figure 4.7: **Tissue types represented in the TRA data in mouse Novartis dataset** Most TRAs in the mouse Novartis dataset, are tissue-restricted for the testis, as well as the oocytes, directly followed by the liver, skeletal muscle, kidney as well as dorsal root ganglion. Basically all tissues of the body are represented by tissue-restricted antigens, most of them are presented in the thymus. For this calculation, we only used the maximum value of tissue-types.

remains unclear. One potential argumentation could be however the fact that in both cases it is a general immune deprived site, as in the case of spermatogenesis as well as the fact that gene expression seems to be regulated through similar effects, such as chromatin decondensation and broad gene expression through methylation of large areas, which might also explain why cancer testis antigens are often upregulated in cancer cells, which might be due to the broad histone methylation in cancer and thus the differential gene expression of cancer cells versus healthy cells.

4.10 TRAs and cancer testis antigens

Although autoimmune male infertility certainly is also a problem in the context of tissue-restricted antigens and central self-tolerance the finding of

cancer testis antigens in our data, is mainly interesting for immunotherapy of cancer. We have found in our data 440 cancer testis antigens (Table 3.8), many of them not known yet in the literature. As cancer testis antigens are often upregulated in cancer versus healthy tissue, side effects of immunotherapy against CTAs might be limited to male infertility. We have found many previously known cancer-testis antigens in our data, such as genes of the ADAM family, PIWIL1, SPATA3 and many others, but we have also proven previously published cancer testis antigens to be wrong, most of them are also expressed in the bone marrow and might be actually very harmful in case of cancer immunotherapy. Already Derbinski et al. 2001 as well as Kyewski et al. 2002 as well as Gotter et al. 2004 pointed out the fact, that immune vaccination against genes, which are potentially upregulated in the thymus might not be a successfull strategy for immunotherapy in general, because immune tolerance is already induced upon prior knowledge based on the negative selection of potentially autoreactive T cells in the thymus. As examples they gave the genes MUC1 as well as CEA and CEACAM6 [98, 215, 144]. Never the less Rosenberg et al. 1999 as well as Chomez et al. 2001 proposed this as a successfull possible strategy [325, 39, 73] and even went into clinical trials. In our data we tried to first identify all possible cancer testis antigens out of our TRA data and samplify its gene expression in the thymus upon previously measured data by Martine Gaertner et al. 2012 and Pinto et al. 2008. Both log FC as well as adjusted p-values are shown in (Table 3.8). More studies would have to be done before we can clearly state on which cancer testis antigens are exactly upregulated in which cancer type, as well as the outcome of potential upregulation of these, especially in the medullary part of the thymus. More analysis have to be done in this field before these strategies should be applied to the clinic and tested on patients. It does not become evident to the author that thoughts like this have not been previously applied.

4.11 TRAs and systemic lupus erythematosus

There are limitations however on the impact of tissue-restricted antigens, the negative selection of T cells in the thymus and the prediction of potential new drug targets or autoantibodies in the field of autoimmune diseases, one example is the difficult to solve problem of sysytemic lupus erythematosus. The problem of lupus in this context seems to be the fact that its reaction is not directed against a single tissue only in the body, but rather systemic against for example nuclear antigens, which we cannot find with our TRA data. It would be still interesting however to conduct systemic studies on tissue-specific autoimmune diseases and the context of tissue-restricted antigens, including the existence of autoantibodies in the blood of patients with systemic lupus erythematosus, but prior to this we cannot state any conclusion nor impact of TRAs for patients with systemic lupus

erythematosus. In general it would be very nice to systematically test now all known autoimmune patients both for their AIRE regulation as well as the immune status in terms of autoantibodies based on our lists of TRAs. There has been the doupt though that this might be more difficult than previously thought since also healthy patients seem to have autoantibodies in their blood without any autoimmune problem. This means systematically check all autoimmune patients for TRAs might over estimate the problem of potential autoimmune diseases. Both protein chips as well as later peptide chips might be an interesting tool to test autoimmune patients with our TRAs.

4.12 Evolutionary conservation of TRA clusters

It has been argued in the past, that TRA clusters are only clustered on a chromosomal level due to gene duplication. It can be shown however that gene families as have been shown previously can be found in TRA clusters and most of the time come from gene duplication, this cannot be the explanation of the general trade of chromosomal clustering of tissue-restricted antigens (TRAs). As the evolution of the adaptive immune system is an evolutionary speaken rather "new" event starting with the beginning of the vertebrates [215], we followed up upon the idea of tracking back the gene order of TRA clusters as well as non TRAs in general further down the evolutionary tree. In terms of the percent identity of genes found in TRA clusters, we could show that there are some gene families with high percent identity of sequence information within the TRA clusters, but that most of the genes range around the expected 25% range of four basepairs in the DNA information, also the testing upon dependence of this results in terms of the size of the clusters did not show any difference. We can there fore state, that gene duplication does not account for TRA clustering in general and that TRAs within a cluster do not only belong to gene families only.

It has been previously known that some functionally related genes are clustered, such as according to Zhang et al. 2004 74% of all housekeeping genes and 70% of all tissue-speicific genes [455], because they stem from in their view "multigene families" and evolve in their perspective slower if they are housekeeping genes, thatn tissue-specific genes [455], also spermatogonia genes have been known to be clustered on the X chromosome as was found by Wang et al. 2001 [412], Hurst et al. 2004 found clusters in different species in as well housekeeping as also tissue-specific genes [167], but non of them ever calculated the percent identity within all clusters found. Also it might still be the explanation of having a common gene pool for gene expression in medullary thymic epithelial cells, facilitating the gene expression by one or a few transcription factors such as Aire. Nevertheless the vast gene expression as well as chromosomal clustering cannot be explained by gene duplication,

nor by a long history view of evolutionary conservation, since our TRA clusters are highly conserved between species, such as human, mouse and rat, but get more and more dispersed the further we go down the evolutionary tree.

4.13 Alternative splicing

As has been mentioned above we have calculated tissue-restricted antigens (TRAs) mostly on the basis of transcripts, rather than on the gene level, since there has been known alternative splicing events, also important in the development of autoimmune dieseases [266, 187, 183]. According to Nagao et al. 2005 55% of all genes are alternatively spliced and therefore account for the vast complexity of the genome, as well as the proteome [266]. PLP1 is one of the many examples in the context of TRAs.

4.14 Conclusion

In conclusion it is possible to theoretically determine tissue-restricted antigens by systematically analyzing different datasets varying from microarray data as in the case of the human and mouse Novartis data [371, 372], RNAseq data [3, 2] as well as protein data, as in the case of the human protein atlas [390] by the means of a TRA is a gene which is higher expressed than 5x the median gene expression in at least one and not more than five tissues of all tissues in the whole dataset. Furthermore the calculation on the basis of transcript level seems to be more adequate in terms of the question of tissue-specificity and alternative splicing events, both in the context of autoimmune diseases as well as the gene expression of TRAs in the thymus. Most of our TRAs have been found in the context of previous knowledge of autoantibodies and autoimmune diseases, many TRAs are upregulated in the medullary part of the thymus. Many tissue-restricted antigens are upregulated as has been previously stated by the autoimmune regulator Aire, whose depletion in the case of knockout mice, or its mutation in the case of the autoimmune polyendocrine syndrome 1 (APS-1) or the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) leads to multiple autoimmune diseases and we could show that tissue-restricted antigens are chromosomally clustered which might be an explanation on their molecular pattern how they can be upregulated together in only one cell type at once, as in the case of medullary thymic epithelial cells (mTECs) in the process of the negative selection of potentially autoreactive T cells in the thymus. Many of the newly discovered tissue-restricted antigens will play a most critical role in the context of the different autoimmune diseases and more research on these can now be done in a more systematic way than before. I hope I could help with this work to elucidate more on the background of central self-tolerance, promiscous gene expression in the context

of negative selection of T cells in the thymus and thus hopefully help to cure autoimmune diseases in the future.

References

[1] Finnish-German APECED Consortium . An autoimmune disease, apeced, caused by mutations in a novel gene featuring two phd-type zinc-finger domains. *Nat Genet*, 17(4):399–403, Dec 1997.

[2] GTEx Consortium . Human genomics. the genotype-tissue expression (gtex) pilot analysis: multitissue gene regulation in humans. *Science*, 348(6235):648–660, May 2015.

[3] Philadelphia Pennsylvania USA. , GTEx ConsortiumNational Disease Research Interchange. The genotype-tissue expression (gtex) project. *Nat Genet*, 45(6):580–585, Jun 2013.

[4] J Aaltonen, N Horelli-Kuitunen, J B Fan, P Björses, J Perheentupa, R Myers, A Palotie, and L Peltonen. High-resolution physical and transcriptional mapping of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy locus on chromosome 21q22.3 by fish. *Genome research*, 7:820–829, August 1997.

[5] Nana Aduma, Hiroe Izumi, Shusei Mizushima, and Asato Kuroiwa. Knockdown of dead-box helicase 4 (ddx4) decreases the number of germ cells in male and female chicken embryonic gonads. *Reproduction, fertility, and development*, December 2018.

[6] Gijs B Afink, Geertruda Veenboer, Janine de Randamie, Remco Keijser, Christof Meischl, Hans Niessen, and Carrie Ris-Stalpers. Initial characterization of c16orf89, a novel thyroid-specific gene. *Thyroid : official journal of the American Thyroid Association*, 20:811–821, July 2010.

[7] Mandana Afsharpad, Mohammad Reza Nowroozi, Maryam Beigom Mobasheri, Mohsen Ayati, Leila Nekoohesh, Mojtaba Saffari, Kazem Zendehdel, and Mohammad Hossein Modarressi. Cancer-testis antigens as new candidate diagnostic biomarkers for transitional cell carcinoma of bladder. *Pathology oncology research : POR*, 25:191–199, January 2019.

[8] Takashi Akamizu. Monoclonal antibodies to thyroid specific autoantigens. *Autoimmunity*, 36:361–366, 2003.

[9] Taishin Akiyama, Shiori Maeda, Sayaka Yamane, Kaori Ogino, Michiyuki Kasai, Fumiko Kajiura, Mitsuru Matsumoto, and Jun-ichiro Inoue. Dependence of self-tolerance on traf6-directed development of thymic stroma. *Science (New York, N.Y.)*, 308:248–251, April 2005.

[10] Judith Alferink, Ivo Lieberam, Wolfgang Reindl, Andrea Behrens, Susanne Weiss, Norbert Hüser, Klaus Gerauer, Ralf Ross, Angelika B Reske-Kunz, Parviz Ahmad-Nejad, Hermann Wagner, and Irmgard Förster. Compartmentalized production of ccl17 in vivo: strong inducibility in peripheral dendritic cells contrasts selective absence from the spleen. *The Journal of experimental medicine*, 197:585–599, March 2003.

[11] Luiz Gonzaga Almeida, Noboru J Sakabe, Alice R deOliveira, Maria Cristina C Silva, Alex S Mundstein, Tzeela Cohen, Yao-Tseng Chen, Ramon Chua, Sita Gurung, Sacha Gnjatic, Achim A Jungbluth, Otávia L Caballero, Amos Bairoch, Eva Kiesler, Sarah L White, Andrew J G Simpson, Lloyd J Old, Anamaria A Camargo, and Ana Tereza R Vasconcelos. Ctdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic acids research*, 37:D816–D819, January 2009.

[12] Takanori Amano, Tomoko Sagai, Hideyuki Tanabe, Yoichi Mizushina, Hiromi Nakazawa, and Toshihiko Shiroishi. Chromosomal dynamics at the shh locus: limb bud-specific differential regulation of competence and active transcription. *Developmental cell*, 16:47–57, January 2009.

[13] Laís Helena Pereira Amaral, Natássia Elena Bufalo, Karina Colombera Peres, Icléia Siqueira Barreto, Antonio Hugo José Fróes Marques Campos, and Laura Sterian Ward. Id proteins may reduce aggressiveness of thyroid tumors. *Endocrine pathology*, November 2018.

[14] A C Anderson, L B Nicholson, K L Legge, V Turchin, H Zaghouani, and V K Kuchroo. High frequency of autoreactive myelin proteolipid protein-specific t cells in the periphery of naive mice: mechanisms of selection of the self-reactive repertoire. *The Journal of experimental medicine*, 191:761–770, March 2000.

[15] Kelly K Anderson, David A Ballok, Neena Prasad, Henry Szechtman, and Boris Sakic. Impaired response to amphetamine and neuronal degeneration in the nucleus accumbens of autoimmune mrl-lpr mice. *Behavioural brain research*, 166:32–38, January 2006.

[16] Mark S Anderson, Emily S Venanzi, Ludger Klein, Zhibin Chen, Stuart P Berzins, Shannon J Turley, Harald von Boehmer, Roderick Bronson, Andrée Dierich, Christophe Benoist, and Diane Mathis. Projection of an immunological self shadow within the thymus by the aire protein. *Science*, 298(5597):1395–1401, Nov 2002.

[17] Belén Ansoleaga, Mariona Jové, Agatha Schlüter, Paula Garcia-Esparcia, Jesús Moreno, Aurora Pujol, Reinald Pamplona, Manuel

Portero-Otín, and Isidre Ferrer. Deregulation of purine metabolism in alzheimer's disease. *Neurobiology of aging*, 36:68–80, January 2015.

[18] F Antequera and A Bird. Number of cpg islands and genes in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America*, 90:11995–11999, December 1993.

[19] Raffaella Avellino, Sabrina Carrella, Marinella Pirozzi, Maurizio Risolino, Francesco Giuseppe Salierno, Paola Franco, Patrizia Stoppelli, Pasquale Verde, Sandro Banfi, and Ivan Conte. mir-204 targeting of ankrd13a controls both mesenchymal neural crest and lens cell migration. *PloS one*, 8:e61099, 2013.

[20] Faezeh Azizi and Soudeh Ghafouri-Fard. Outer dense fiber proteins: Bridging between male infertility and cancer. *Archives of Iranian medicine*, 20:320–325, May 2017.

[21] T Bachrich, T Thalhammer, W Jäger, P Haslmayer, B Alihodzic, S Bakos, E Hitchman, A M Senderowicz, and E Penner. Characterization of autoantibodies against uridine-diphosphate glucuronosyltransferase in patients with inflammatory liver diseases. *Hepatology (Baltimore, Md.)*, 33:1053–1059, May 2001.

[22] Jin-Han Bae, Jongha Park, Kwang Mo Yang, Tae-Oh Kim, Joo Mi Yi, and IBD Study Group of Korean Association for Study of Intestinal Diseases (KASID). Detection of dna hypermethylation in sera of patients with crohn's disease. *Molecular medicine reports*, 9:725–729, February 2014.

[23] Y Baert, J-B Stukenborg, M Landreh, J De Kock, H Jörnvall, O Söder, and E Goossens. Derivation and characterization of a cytocompatible scaffold from human testis. *Human reproduction (Oxford, England)*, 30:256–267, February 2015.

[24] Shun Bai, Kaiqiang Fu, Huiqi Yin, Yiqiang Cui, Qiuling Yue, Wenbo Li, Le Cheng, Huanhuan Tan, Xiaofei Liu, Yueshuai Guo, Yingwen Zhang, Jie Xie, Wenxiu He, Yuanyuan Wang, Hua Feng, Changpeng Xin, Jinwen Zhang, Mingyan Lin, Bin Shen, Zheng Sun, Xuejiang Guo, Ke Zheng, and Lan Ye. Sox30 initiates transcription of haploid genes during late meiosis and spermiogenesis in mouse testes. *Development (Cambridge, England)*, 145, July 2018.

[25] Tanya Barrett, Stephen E Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F Kim, Maxim Tomashevsky, Kimberly A Marshall, Katherine H Phillippy, Patti M Sherman, Michelle Holko, Andrey Yefanov, Hyeseung Lee, Naigong Zhang, Cynthia L Robertson, Nadezhda Serova,

Sean Davis, and Alexandra Soboleva. Ncbi geo: archive for functional genomics data sets–update. *Nucleic Acids Res*, 41(Database issue):D991–D995, Jan 2013.

[26] Luigi Bartalena. Diagnosis and management of graves disease: a global overview. *Nature reviews. Endocrinology*, 9:724–734, December 2013.

[27] Aurélie Bellanger, Caterina F Donini, Julie A Vendrell, Jonathan Lavaud, Irma Machuca-Gayet, Maëva Ruel, Julien Vollaire, Evelyne Grisard, Balázs Győrffy, Ivan Bièche, Olivier Peyruchaud, Jean-Luc Coll, Isabelle Treilleux, Véronique Maguer-Satta, Véronique Josserand, and Pascale A Cohen. The critical role of the znf217 oncogene in promoting breast cancer metastasis to the bone. *The Journal of pathology*, 242:73–89, May 2017.

[28] Giovanna Berruti and Enzo Martegani. mubpy and msj-1, a deubiquitinating enzyme and a molecular chaperone specifically expressed in testis, associate with the acrosome and centrosome in mouse germ cells. *Annals of the New York Academy of Sciences*, 973:5–7, November 2002.

[29] Amol K Bhandage, Zhe Jin, Sergiy V Korol, Atieh S Tafreshiha, Priya Gohel, Charlotte Hellgren, Daniel Espes, Per-Ola Carlsson, Inger Sundström-Poromaa, and Bryndis Birnir. Expression of calcium release-activated and voltage-gated calcium channels genes in peripheral blood mononuclear cells is altered in pregnancy and in type 1 diabetes. *PloS one*, 13:e0208981, 2018.

[30] E Bhatia, K Balasubramanium, J Rajeswari, O Kordonouri, H Witt, O Landt, P Simon, and M M Lerch. Absence of association between spink1 trypsin inhibitor mutations and type 1 or 2 diabetes mellitus in india and germany. *Diabetologia*, 46:1710–1711, December 2003.

[31] Indrashis Bhattacharya, Sayon Basu, Bhola Shankar Pradhan, Hironmoy Sarkar, Perumal Nagarajan, and Subeer S Majumdar. Testosterone augments fsh signaling by upregulating the expression and activity of fsh-receptor in pubertal primate sertoli cells. *Molecular and cellular endocrinology*, 482:70–80, February 2019.

[32] Enrique Blanco, Miguel Pignatelli, Sergi Beltran, Adrià Punset, Silvia Pérez-Lluch, Florenci Serras, Roderic Guigó, and Montserrat Corominas. Conserved chromosomal clustering of genes governed by chromatin regulators in drosophila. *Genome biology*, 9:R134, 2008.

[33] K Blechschmidt, M Schweiger, K Wertz, R Poulson, H M Christensen, A Rosenthal, H Lehrach, and M L Yaspo. The mouse aire gene:

comparative genomic sequencing, gene organization, and expression. *Genome research*, 9:158–166, February 1999.

[34] Simona Blotta, Pierfrancesco Tassone, Rao H Prabhala, Piersandro Tagliaferri, David Cervi, Samir Amin, Jana Jakubikova, Yu-Tzu Tai, Klaus Podar, Constantine S Mitsiades, Alessandro Zullo, Brunella Franco, Kenneth C Anderson, and Nikhil C Munshi. Identification of novel antigens with induced immune response in monoclonal gammopathy of undetermined significance. *Blood*, 114:3276–3284, October 2009.

[35] Thomas Blumenthal and Kathy Seggerson Gleason. Caenorhabditis elegans operons: form and function. *Nature reviews. Genetics*, 4:112–120, February 2003.

[36] Bolstad BM. affyio: Tools for parsing affymetrix data files. r package version 1.46.0. 2017.

[37] Thomas Boehm, Stefanie Scheu, Klaus Pfeffer, and Conrad C Bleul. Thymic medullary epithelial cell differentiation, thymocyte emigration, and the control of autoimmunity require lympho-epithelial cross talk via ltbetar. *The Journal of experimental medicine*, 198:757–769, September 2003.

[38] Ewelina Bolcun-Filas, Emma Hall, Robert Speed, Mary Taggart, Corinne Grey, Bernard de Massy, Ricardo Benavente, and Howard J Cooke. Mutation of the mouse syce1 gene disrupts synapsis and suggests a link between synaptonemal complex structural components and dna repair. *PLoS genetics*, 5:e1000393, February 2009.

[39] T Boon, P G Coulie, and B Van den Eynde. Tumor antigens recognized by t cells. *Immunology today*, 18:267–268, June 1997.

[40] G Borsani, M T Bassi, M P Sperandeo, A De Grandi, A Buoninconti, M Riboni, M Manzoni, B Incerti, A Pepe, G Andria, A Ballabio, and G Sebastio. Slc7a7, encoding a putative permease-related protein, is mutated in patients with lysinuric protein intolerance. *Nature genetics*, 21:297–301, March 1999.

[41] Rinke Bos, Suzanne van Duikeren, Thorbald van Hall, Patricia Kaaijk, Richard Taubert, Bruno Kyewski, Ludger Klein, Cornelis J M Melief, and Rienk Offringa. Expression of a natural tumor antigen by thymic epithelial cells impairs the tumor-protective cd4+ t-cell repertoire. *Cancer research*, 65:6443–6449, July 2005.

[42] Fèlix Bosch and Laia Rosich. The contributions of paul ehrlich to pharmacology: a tribute on the occasion of the centenary of his nobel prize. *Pharmacology*, 82:171–179, 2008.

[43] Philippe Bousso, Nirav R Bhakta, Richard S Lewis, and Ellen Robey. Dynamics of thymocyte-stromal cell interactions visualized by two-photon microscopy. *Science (New York, N.Y.)*, 296:1876–1880, June 2002.

[44] Alexander M Boutanaev and Lev G Nemchinov. Are clustered genes in the genomes of arabidopsis and drosophila regulated differently? *Gene*, 491:284–288, January 2012.

[45] P Boël, C Wildmann, M L Sensi, R Brasseur, J C Renauld, P Coulie, T Boon, and P van der Bruggen. Bage: a new gene encoding an antigen recognized on human melanomas by cytolytic t lymphocytes. *Immunity*, 2:167–175, February 1995.

[46] Philip Brennecke, Alejandro Reyes, Sheena Pinto, Kristin Rattay, Michelle Nguyen, Rita Küchler, Wolfgang Huber, Bruno Kyewski, and Lars M Steinmetz. Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. *Nature immunology*, 16:933–941, September 2015.

[47] Lydie Brisson, Laurent Pouyet, Prudence N'guessan, Stéphane Garcia, Noëlla Lopes, Gilles Warcollier, Juan L Iovanna, and Alice Carrier. The thymus-specific serine protease tssp/prss16 is crucial for the antitumoral role of cd4(+) t cells. *Cell reports*, 10:39–46, January 2015.

[48] Sean G Brown, Melissa R Miller, Polina V Lishko, Douglas H Lester, Stephen J Publicover, Christopher L R Barratt, and Sarah Martins Da Silva. Homozygous in-frame deletion in catspere in a man producing spermatozoa with loss of catsper function and compromised fertilizing capacity. *Human reproduction (Oxford, England)*, 33:1812–1816, October 2018.

[49] Marcilei E Buim, Fernando A Soares, Alvaro S Sarkis, and Maria Aparecida Nagai. The transcripts of sfrp1,cep63 and eif4g2 genes are frequently downregulated in transitional cell carcinomas of the bladder. *Oncology*, 69:445–454, 2005.

[50] Debra T Burhans, Lakshmi Ramachandran, Jianxin Wang, Ping Liang, Hugh G Patterton, Michael Breitenbach, and William C Burhans. Non-random clustering of stress-related genes during evolution of the s. cerevisiae genome. *BMC evolutionary biology*, 6:58, July 2006.

[51] Leigh Ann Bush, John C Herr, Michael Wolkowicz, Nicholas E Sherman, Amy Shore, and Charles J Flickinger. A novel asparaginase-like protein is a sperm autoantigen in rats. *Molecular reproduction and development*, 62:233–247, June 2002.

[52] A J Butte, V J Dzau, and S B Glueck. Further defining housekeeping, or "maintenance," genes focus on "a compendium of gene expression in normal human tissues". *Physiological genomics*, 7:95–96, December 2001.

[53] Rachele Cagliani, Uberto Pozzoli, Diego Forni, Andrea Cassinotti, Matteo Fumagalli, Matteo Giani, Maria Fichera, Marta Lombardini, Sandro Ardizzone, Rosanna Asselta, Roberto de Franchis, Stefania Riva, Mara Biasin, Giacomo P Comi, Nereo Bresolin, Mario Clerici, and Manuela Sironi. Crohn's disease loci are common targets of protozoa-driven selection. *Molecular biology and evolution*, 30:1077–1087, May 2013.

[54] J J Campbell and E C Butcher. Chemokines in tissue-specific and microenvironment-specific lymphocyte homing. *Current opinion in immunology*, 12:336–341, June 2000.

[55] Alan B Cantor, Samuel G Katz, and Stuart H Orkin. Distinct domains of the gata-1 cofactor fog-1 differentially influence erythroid versus megakaryocytic maturation. *Molecular and cellular biology*, 22:4268–4279, June 2002.

[56] Yonghao Cao, Brittany A Goods, Khadir Raddassi, Gerald T Nepom, William W Kwok, J Christopher Love, and David A Hafler. Functional inflammatory profiles distinguish myelin-reactive t cells from patients with multiple sclerosis. *Science translational medicine*, 7:287ra74, May 2015.

[57] Ellen F Carney. Gatm mutations cause mitochondrial abnormalities and kidney failure. *Nature reviews. Nephrology*, 14:414, July 2018.

[58] S R Catford, M K O'Bryan, R I McLachlan, M B Delatycki, and L Rombauts. Germ cell arrest associated with asetx mutation in ataxia oculomotor apraxia type 2. *Reproductive biomedicine online*, January 2019.

[59] D P Cerretti, R F DuBose, R A Black, and N Nelson. Isolation of two novel metalloproteinase-disintegrin (adam) cdnas that show testis-specific gene expression. *Biochemical and biophysical research communications*, 263:810–815, October 1999.

[60] Chun-Chi Chang, Ya-Sian Chang, Hsi-Yuan Huang, Kun-Tu Yeh, Ta-Chih Liu, and Jan-Gowth Chang. Determination of the mutational landscape in taiwanese patients with papillary thyroid cancer by whole-exome sequencing. *Human pathology*, 78:151–158, August 2018.

[61] Ming-Chu Chang, I-Shiow Jan, Po-Chin Liang, Yung-Ming Jeng, Ching-Yao Yang, Yu-Wen Tien, Jau-Min Wong, and Yu-Ting Chang. Human cationic trypsinogen but not serine peptidase inhibitor, kazal type 1 variants increase the risk of type 1 autoimmune pancreatitis. *Journal of gastroenterology and hepatology*, 29:2038–2042, December 2014.

[62] N X Chao, L Z Li, G R Luo, W G Zhong, R S Huang, R Fan, and F L Zhao. Cancer-testis antigen gage-1 expression and serum immunoreactivity in hepatocellular carcinoma. *Nigerian journal of clinical practice*, 21:1361–1367, October 2018.

[63] Mayank Chauhan, Tanel Punga, and Anna Rostedt Punga. Muscle-specific regulation of the mtor signaling pathway in musk antibody seropositive (musk+) experimental autoimmune myasthenia gravis (eamg). *Neuroscience research*, 77:102–109, 2013.

[64] Hanbo Chen and Paul C Boutros. Venndiagram: a package for the generation of highly-customizable venn and euler diagrams in r. *BMC bioinformatics*, 12:35, January 2011.

[65] Hsin-Hsiung Chen, Wen-Pin Chen, Wan-Lun Yan, Yuan-Chun Huang, Szu-Wei Chang, Wen-Mei Fu, Ming-Jai Su, I-Shing Yu, Tzung-Chieh Tsai, Yu-Ting Yan, Yeou-Ping Tsao, and Show-Li Chen. Nrip is newly identified as a z-disc protein, activating calmodulin signaling for skeletal muscle contraction and regeneration. *Journal of cell science*, 128:4196–4209, November 2015.

[66] Liang Chen, Yang Xu, Wen-li Zuo, Hui-xia Yang, Qin-ping Liao, Zhong-cheng Xin, and Ying-lu Guo. [effect of ras signal regulating cox7a2 expression and its sub-cellular location in testicular interstitial cell]. *Beijing da xue xue bao. Yi xue ban = Journal of Peking University. Health sciences*, 44:507–510, August 2012.

[67] Lin Chen, Peter B Barker, Robert G Weiss, Peter C M van Zijl, and Jiadi Xu. Creatine and phosphocreatine mapping of mouse skeletal muscle by a polynomial and lorentzian line-shape fitting cest method. *Magnetic resonance in medicine*, 81:69–78, January 2019.

[68] Ying Chen, Qing Wang, Mi Wang, and Ming Li. Overexpressed lrig3 gene ameliorates prostate cancer through suppression of cell invasion and migration. *International journal of biological macromolecules*, 124:1–9, November 2018.

[69] An Ning Cheng, Yu-Kang Lo, Yi-Sheng Lin, Tswen-Kei Tang, Chun-Hua Hsu, John T-A Hsu, and Alan Yueh-Luen Lee. Identification of novel cdc7 kinase inhibitors as anti-cancer agents that target the

interaction with dbf4 by the fragment complementation and drug repositioning approach. *EBioMedicine*, 36:241–251, October 2018.

[70] C Cho, P Primakoff, J M White, and D G Myles. Chromosomal assignment of four testis-expressed mouse genes from a new family of transmembrane proteins (adams) involved in cell-cell adhesion and fusion. *Genomics*, 34:413–417, June 1996.

[71] Hyosuk Cho, Gong-Qing Shen, Xiaofeng Wang, Fan Wang, Stephen Archacki, Yabo Li, Gang Yu, Susmita Chakrabarti, Qiuyun Chen, and Qing Kenneth Wang. Long noncoding rna anril regulates endothelial cell activities associated with coronary artery disease by up-regulating clip1, ezr, and lyve1 genes. *The Journal of biological chemistry*, January 2019.

[72] Inchul Choi, Jong-Min Woo, Sunghee Hong, Yong-Keun Jung, Do Han Kim, and Chunghee Cho. Identification and characterization of adam32 with testis-predominant gene expression. *Gene*, 304:151–162, January 2003.

[73] P Chomez, O De Backer, M Bertrand, E De Plaen, T Boon, and S Lucas. An overview of the mage gene family with the identification of all human members of the family. *Cancer research*, 61:5544–5551, July 2001.

[74] Uttio Roy Chowdhury, Rajeev S Samant, Oystein Fodstad, and Lalita A Shevde. Emerging role of nuclear protein 1 (nupr1) in cancer biology. *Cancer metastasis reviews*, 28:225–232, June 2009.

[75] Silvie Cloosen, Janna Arnold, Marco Thio, Gerard M J Bos, Bruno Kyewski, and Wilfred T V Germeraad. Expression of tumor-associated differentiation antigens, muc1 glycoforms and cea, in human thymic epithelial cells: implications for self-tolerance and tumor therapy. *Cancer research*, 67:3919–3926, April 2007.

[76] Pierre Close, Magali Gillard, Aurélie Ladang, Zheshen Jiang, Jessica Papuga, Nicola Hawkes, Laurent Nguyen, Jean-Paul Chapelle, Fabrice Bouillenne, Jesper Svejstrup, Marianne Fillet, and Alain Chariot. Derp6 (elp5) and c3orf75 (elp6) regulate tumorigenicity and migration of melanoma cells as subunits of elongator. *The Journal of biological chemistry*, 287:32535–32545, September 2012.

[77] Seth B Coffelt, Max D Wellenstein, and Karin E de Visser. Neutrophils in cancer: neutral no more. *Nature reviews. Cancer*, 16:431–446, July 2016.

[78] B A Cohen, R D Mitra, J D Hughes, and G M Church. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature genetics*, 26:183–186, October 2000.

[79] Melvin Cohn, N Av Mitchison, William E Paul, Arthur M Silverstein, David W Talmage, and Martin Weigert. Reflections on the clonal-selection theory. *Nature reviews. Immunology*, 7:823–830, October 2007.

[80] Dan Cong, Mengzi He, Silin Chen, Xiaoli Liu, Xiaodong Liu, and Hui Sun. Expression profiles of pivotal micrornas and targets in thyroid papillary carcinoma: an analysis of the cancer genome atlas. *OncoTargets and therapy*, 8:2271–2277, 2015.

[81] FANTOM Consortium, the RIKEN PMI, CLST (DGT), Alistair R R Forrest, Hideya Kawaji, Michael Rehli, J Kenneth Baillie, Michiel J L de Hoon, Vanja Haberle, Timo Lassmann, Ivan V Kulakovskiy, Marina Lizio, Masayoshi Itoh, Robin Andersson, Christopher J Mungall, Terrence F Meehan, Sebastian Schmeier, Nicolas Bertin, Mette Jørgensen, Emmanuel Dimont, Erik Arner, Christian Schmidl, Ulf Schaefer, Yulia A Medvedeva, Charles Plessy, Morana Vitezic, Jessica Severin, Colin A Semple, Yuri Ishizu, Robert S Young, Margherita Francescatto, Intikhab Alam, Davide Albanese, Gabriel M Altschuler, Takahiro Arakawa, John A C Archer, Peter Arner, Magda Babina, Sarah Rennie, Piotr J Balwierz, Anthony G Beckhouse, Swati Pradhan-Bhatt, Judith A Blake, Antje Blumenthal, Beatrice Bodega, Alessandro Bonetti, James Briggs, Frank Brombacher, A Maxwell Burroughs, Andrea Califano, Carlo V Cannistraci, Daniel Carbajo, Yun Chen, Marco Chierici, Yari Ciani, Hans C Clevers, Emiliano Dalla, Carrie A Davis, Michael Detmar, Alexander D Diehl, Taeko Dohi, Finn Drabløs, Albert S B Edge, Matthias Edinger, Karl Ekwall, Mitsuhiro Endoh, Hideki Enomoto, Michela Fagiolini, Lynsey Fairbairn, Hai Fang, Mary C Farach-Carson, Geoffrey J Faulkner, Alexander V Favorov, Malcolm E Fisher, Martin C Frith, Rie Fujita, Shiro Fukuda, Cesare Furlanello, Masaaki Furino, Jun-ichi Furusawa, Teunis B Geijtenbeek, Andrew P Gibson, Thomas Gingeras, Daniel Goldowitz, Julian Gough, Sven Guhl, Reto Guler, Stefano Gustincich, Thomas J Ha, Masahide Hamaguchi, Mitsuko Hara, Matthias Harbers, Jayson Harshbarger, Akira Hasegawa, Yuki Hasegawa, Takehiro Hashimoto, Meenhard Herlyn, Kelly J Hitchens, Shannan J Ho Sui, Oliver M Hofmann, Ilka Hoof, Furni Hori, Lukasz Huminiecki, Kei Iida, Tomokatsu Ikawa, Boris R Jankovic, Hui Jia, Anagha Joshi, Giuseppe Jurman, Bogumil Kaczkowski, Chieko Kai, Kaoru Kaida, Ai Kaiho, Kazuhiro Kajiyama, Mutsumi

Kanamori-Katayama, Artem S Kasianov, Takeya Kasukawa, Shintaro Katayama, Sachi Kato, Shuji Kawaguchi, Hiroshi Kawamoto, Yuki I Kawamura, Tsugumi Kawashima, Judith S Kempfle, Tony J Kenna, Juha Kere, Levon M Khachigian, Toshio Kitamura, S Peter Klinken, Alan J Knox, Miki Kojima, Soichi Kojima, Naoto Kondo, Haruhiko Koseki, Shigeo Koyasu, Sarah Krampitz, Atsutaka Kubosaki, Andrew T Kwon, Jeroen F J Laros, Weonju Lee, Andreas Lennartsson, Kang Li, Berit Lilje, Leonard Lipovich, Alan Mackay-Sim, Ri-ichiroh Manabe, Jessica C Mar, Benoit Marchand, Anthony Mathelier, Niklas Mejhert, Alison Meynert, Yosuke Mizuno, David A de Lima Morais, Hiromasa Morikawa, Mitsuru Morimoto, Kazuyo Moro, Efthymios Motakis, Hozumi Motohashi, Christine L Mummery, Mitsuyoshi Murata, Sayaka Nagao-Sato, Yutaka Nakachi, Fumio Nakahara, Toshiyuki Nakamura, Yukio Nakamura, Kenichi Nakazato, Erik van Nimwegen, Noriko Ninomiya, Hiromi Nishiyori, Shohei Noma, Shohei Noma, Tadasuke Noazaki, Soichi Ogishima, Naganari Ohkura, Hiroko Ohimiya, Hiroshi Ohno, Mitsuhiro Ohshima, Mariko Okada-Hatakeyama, Yasushi Okazaki, Valerio Orlando, Dmitry A Ovchinnikov, Arnab Pain, Robert Passier, Margaret Patrikakis, Helena Persson, Silvano Piazza, James G D Prendergast, Owen J L Rackham, Jordan A Ramilowski, Mamoon Rashid, Timothy Ravasi, Patrizia Rizzu, Marco Roncador, Sugata Roy, Morten B Rye, Eri Saijyo, Antti Sajantila, Akiko Saka, Shimon Sakaguchi, Mizuho Sakai, Hiroki Sato, Suzana Savvi, Alka Saxena, Claudio Schneider, Erik A Schultes, Gundula G Schulze-Tanzil, Anita Schwegmann, Thierry Sengstag, Guojun Sheng, Hisashi Shimoji, Yishai Shimoni, Jay W Shin, Christophe Simon, Daisuke Sugiyama, Takaai Sugiyama, Masanori Suzuki, Naoko Suzuki, Rolf K Swoboda, Peter A C 't Hoen, Michihira Tagami, Naoko Takahashi, Jun Takai, Hiroshi Tanaka, Hideki Tatsukawa, Zuotian Tatum, Mark Thompson, Hiroo Toyodo, Tetsuro Toyoda, Elvind Valen, Marc van de Wetering, Linda M van den Berg, Roberto Verado, Dipti Vijayan, Ilya E Vorontsov, Wyeth W Wasserman, Shoko Watanabe, Christine A Wells, Louise N Winteringham, Ernst Wolvetang, Emily J Wood, Yoko Yamaguchi, Masayuki Yamamoto, Misako Yoneda, Yohei Yonekura, Shigehiro Yoshida, Susan E Zabierowski, Peter G Zhang, Xiaobei Zhao, Silvia Zucchelli, Kim M Summers, Harukazu Suzuki, Carsten O Daub, Jun Kawai, Peter Heutink, Winston Hide, Tom C Freeman, Boris Lenhard, Vladimir B Bajic, Martin S Taylor, Vsevolod J Makeev, Albin Sandelin, David A Hume, Piero Carninci, and Yoshihide Hayashizaki. A promoter-level mammalian expression atlas. *Nature*, 507:462–470, March 2014.

[82] Jaques A Courtade, Agnieszka M Klimek-Abercrombie, Yi-Chun Chen, Nirja Patel, Phoebe Y T Lu, Cate Speake, Paul C Orban,

Behzad Najafian, Graydon Meneilly, Carla J Greenbaum, Garth L Warnock, Constadina Panagiotopoulos, and C Bruce Verchere. Measurement of pro-islet amyloid polypeptide (1-48) in diabetes and islet transplants. *The Journal of clinical endocrinology and metabolism*, 102:2595–2603, July 2017.

[83] Sara Carmela Credendino, Nicole Lewin, Miriane de Oliveira, Swaraj Basu, Barbara D'Andrea, Elena Amendola, Luigi Di Guida, Antonio Nardone, Remo Sanges, Mario De Felice, and Gabriella De Vita. Tissue- and cell type-specific expression of the long noncoding rna klhl14-as in mouse. *International journal of genomics*, 2017:9769171, 2017.

[84] T Cremer and C Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews. Genetics*, 2:292–301, April 2001.

[85] Jennifer A Crookshank, Daniel Serrano, Gen-Sheng Wang, Christopher Patrick, Baylie S Morgan, Marie-France Pare, and Fraser W Scott. Changes in insulin, glucagon and er stress precede immune activation in type 1 diabetes. *The Journal of endocrinology*, August 2018.

[86] Mary K Crow. Interferon-alpha: a therapeutic target in systemic lupus erythematosus. *Rheumatic diseases clinics of North America*, 36:173–86, x, February 2010.

[87] Z Cui, A Agarwal, B F da Silva, R Sharma, and E Sabanegh. Evaluation of seminal plasma proteomics and relevance of fsh in identification of nonobstructive azoospermia: A preliminary study. *Andrologia*, 50:e12999, June 2018.

[88] Manhong Dai, Pinglang Wang, Andrew D Boyd, Georgi Kostov, Brian Athey, Edward G Jones, William E Bunney, Richard M Myers, Terry P Speed, Huda Akil, Stanley J Watson, and Fan Meng. Evolving gene/-transcript definitions significantly alter the interpretation of genechip data. *Nucleic Acids Res*, 33(20):e175, 2005.

[89] Cajsa Davegårdh, Christa Broholm, Alexander Perfilyev, Tora Henriksen, Sonia García-Calzón, Lone Peijs, Ninna Schiøler Hansen, Petr Volkov, Rasmus Kjøbsted, Jørgen F P Wojtaszewski, Maria Pedersen, Bente Klarlund Pedersen, Dov B Ballak, Charles A Dinarello, Bas Heinhuis, Leo A B Joosten, Emma Nilsson, Allan Vaag, Camilla Scheele, and Charlotte Ling. Abnormal epigenetic changes during differentiation of human skeletal muscle stem cells from obese subjects. *BMC medicine*, 15:39, February 2017.

[90] O De Backer, K C Arden, M Boretti, V Vantomme, C De Smet, S Czekay, C S Viars, E De Plaen, F Brasseur, P Chomez, B Van den Eynde, T Boon, and P van der Bruggen. Characterization of the gage genes that are expressed in various human cancers and in normal testis. *Cancer research*, 59:3157–3165, July 1999.

[91] E De Plaen, K Arden, C Traversari, J J Gaforio, J P Szikora, C De Smet, F Brasseur, P van der Bruggen, B Lethé, and C Lurquin. Structure, chromosomal localization, and expression of 12 genes of the mage family. *Immunogenetics*, 40:360–369, 1994.

[92] Elzo de Wit and Wouter de Laat. A decade of 3c technologies: insights into nuclear organization. *Genes & development*, 26:11–24, January 2012.

[93] Ignacio Del Valle, Federica Buonocore, Andrew J Duncan, Lin Lin, Martino Barenco, Rahul Parnaik, Sonia Shah, Mike Hubank, Dianne Gerrelli, and John C Achermann. A genomic atlas of human adrenal and gonad development. *Wellcome open research*, 2:25, April 2017.

[94] Francesca Demichelis, Sunita R Setlur, Samprit Banerjee, Dimple Chakravarty, Jin Yun Helen Chen, Chen X Chen, Julie Huang, Himisha Beltran, Derek A Oldridge, Naoki Kitabayashi, Birgit Stenzel, Georg Schaefer, Wolfgang Horninger, Jasmin Bektic, Arul M Chinnaiyan, Sagit Goldenberg, Javed Siddiqui, Meredith M Regan, Michale Kearney, T David Soong, David S Rickman, Olivier Elemento, John T Wei, Douglas S Scherr, Martin A Sanda, Georg Bartsch, Charles Lee, Helmut Klocker, and Mark A Rubin. Identification of functionally active, low frequency copy number variants at 15q21.3 and 12q21.31 associated with prostate cancer risk. *Proceedings of the National Academy of Sciences of the United States of America*, 109:6686–6691, April 2012.

[95] Korcan Demir, Ayça Altıncık, and Ece Böber. Severe short stature due to 3-m syndrome with a novel obsl1 gene mutation. *Journal of pediatric endocrinology & metabolism : JPEM*, 26:147–150, 2013.

[96] Calliope A Dendrou, Lars Fugger, and Manuel A Friese. Immunopathology of multiple sclerosis. *Nature reviews. Immunology*, 15:545–558, September 2015.

[97] Heather C Denroche and C Bruce Verchere. Iapp and type 1 diabetes: implications for immunity, metabolism and islet transplants. *Journal of molecular endocrinology*, 60:R57–R75, February 2018.

[98] J. Derbinski, A. Schulte, B. Kyewski, and L. Klein. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nat Immunol*, 2(11):1032–1039, Nov 2001.

[99] Jens Derbinski and Bruno Kyewski. Linking signalling pathways, thymic stroma integrity and autoimmunity. *Trends Immunol*, 26(10):503–506, Oct 2005.

[100] Edward Dere, Daniel J Spade, Susan J Hall, Aimee Altemus, James D Smith, Jonathan A Phillips, Jeffrey S Moffit, Kerry T Blanchard, and Kim Boekelheide. Identification of sperm mrna biomarkers associated with testis injury during preclinical testing of pharmaceutical compounds. *Toxicology and applied pharmacology*, 320:1–7, April 2017.

[101] J M Devaney, S Wang, S Funda, J Long, D J Taghipour, R Tbaishat, P Furbert-Harris, M Ittmann, and B Kwabi-Addo. Identification of novel dna-methylated genes that correlate with human prostate cancer and high-grade prostatic intraepithelial neoplasia. *Prostate cancer and prostatic diseases*, 16:292–300, December 2013.

[102] Maria Agnes Dinkelacker. A database of genes that are expressed in a tissue-restricted manner to analyse promiscous gene expression in medullary thymic epithelial cells. Diplomarbeit, Albert-Ludwigs-Universität, Freiburg, Germany, 2007.

[103] Dijana Djureinovic, Björn M Hallström, Masafumi Horie, Johanna Sofia Margareta Mattsson, Linnea La Fleur, Linn Fagerberg, Hans Brunnström, Cecilia Lindskog, Katrin Madjar, Jörg Rahnenführer, Simon Ekman, Elisabeth Ståhle, Hirsh Koyi, Eva Brandén, Karolina Edlund, Jan G Hengstler, Mats Lambe, Akira Saito, Johan Botling, Fredrik Pontén, Mathias Uhlén, and Patrick Micke. Profiling cancer testis antigens in non-small-cell lung cancer. *JCI insight*, 1:e86837, July 2016.

[104] Gabrielle C Douglas, Moira K O'Bryan, Mark P Hedger, David K L Lee, Michael A Yarski, A Ian Smith, and Rebecca A Lew. The novel angiotensin-converting enzyme (ace) homolog, ace2, is selectively expressed by adult leydig cells of the testis. *Endocrinology*, 145:4703–4711, October 2004.

[105] Pan Du, Warren A Kibbe, and Simon M Lin. nuid: a universal naming scheme of oligonucleotides for illumina, affymetrix, and other microarrays. *Biol Direct*, 2:16, 2007.

[106] Erica Dugnani, Valeria Sordi, Silvia Pellegrini, Raniero Chimienti, Ilaria Marzinotto, Valentina Pasquale, Daniela Liberati, Gianpaolo

Balzano, Claudio Doglioni, Michele Reni, Alessandra Gandolfi, Massimo Falconi, Vito Lampasona, and Lorenzo Piemonti. Gene expression analysis of embryonic pancreas development master regulators and terminal cell fate markers in resected pancreatic cancer: A correlation with clinical outcome. *Pancreatology : official journal of the International Association of Pancreatology (IAP) ... [et al.]*, September 2018.

[107] Chihiro Ebihara, Ken Ebihara, Megumi Aizawa-Abe, Tomoji Mashimo, Tsutomu Tomita, Mingming Zhao, Valentino Gumbilai, Toru Kusakabe, Yuji Yamamoto, Daisuke Aotani, Sachiko Yamamoto-Kataoka, Takeru Sakai, Kiminori Hosoda, Tadao Serikawa, and Kazuwa Nakao. Seipin is necessary for normal brain development and spermatogenesis in addition to adipogenesis. *Human molecular genetics*, 24:4238–4249, August 2015.

[108] Pablo C Echeverría, Pierre-André Briand, and Didier Picard. A remodeled hsp90 molecular chaperone ensemble with the novel cochaperone aarsd1 is required for muscle differentiation. *Molecular and cellular biology*, 36:1310–1321, April 2016.

[109] Ron Edgar, Michael Domrachev, and Alex E Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res*, 30(1):207–210, Jan 2002.

[110] M Egerton, R Scollay, and K Shortman. Kinetics of mature t-cell development in the thymus. *Proceedings of the National Academy of Sciences of the United States of America*, 87:2579–2582, April 1990.

[111] Eli Eisenberg and Erez Y Levanon. Human housekeeping genes are compact. *Trends in genetics : TIG*, 19:362–365, July 2003.

[112] Weiyi Fang, Qingping Jiang, and Zhen Liu. [screening of proteins regulated by ccdc19 gene in nasopharyngeal carcinoma using proteomics technology]. *Nan fang yi ke da xue xue bao = Journal of Southern Medical University*, 32:1127–1130, August 2012.

[113] Sofia Farkona, Antoninus Soosaipillai, Panagiota Filippou, Christos Liaskos, Dimitrios P Bogdanos, Eleftherios P Diamandis, and Ivan M Blasutig. Novel immunoassays for detection of cuzd1 autoantibodies in serum of patients with inflammatory bowel diseases. *Clinical chemistry and laboratory medicine*, 55:1574–1581, August 2017.

[114] Shu Feng, Ryan E Dodge, and Courtney E Francis. Myasthenia gravis associated with orbital marginal zone lymphoma. *Journal of neuro-ophthalmology : the official journal of the North American Neuro-Ophthalmology Society*, January 2019.

[115] Lara P Fernández, Arístides López-Márquez, Angel M Martínez, Gonzalo Gómez-López, and Pilar Santisteban. New insights into foxe1 functions: identification of direct foxe1 targets in thyroid cells. *PloS one*, 8:e62849, 2013.

[116] Bruno Ferraz-de Souza and John C Achermann. Disorders of adrenal development. *Endocrine development*, 13:19–32, 2008.

[117] R Fischel and D Eilat. Structure and binding properties of a monoclonal anti-idiotypic autoantibody to anti-dna with epibody activity. *Journal of immunology (Baltimore, Md. : 1950)*, 149:3089–3096, November 1992.

[118] Christa E Flück. Mechanisms in endocrinology: Update on pathogenesis of primary adrenal insufficiency: beyond steroid enzyme deficiency and autoimmune adrenal destruction. *European journal of endocrinology*, 177:R99–R111, September 2017.

[119] Jason D Fontenot and Alexander Y Rudensky. A well adapted regulatory contrivance: regulatory t cell development and the forkhead family transcription factor foxp3. *Nature immunology*, 6:331–337, April 2005.

[120] Lily Francis, Zong Sheng Guo, Zuqiang Liu, Roshni Ravindranathan, Julie A Urban, Magesh Sathaiah, Deepa Magge, Pawel Kalinski, and David L Bartlett. Modulation of chemokines in the tumor microenvironment enhances oncolytic virotherapy for colorectal cancer. *Oncotarget*, 7:22174–22185, April 2016.

[121] J Frayne, J A Jury, H L Barker, and L Hall. Rat mdc family of proteins: sequence analysis, tissue distribution, and expression in prepubertal and adult rat testis. *Molecular reproduction and development*, 48:159–167, October 1997.

[122] Yoshitaka Fujihara, Asami Oji, Kanako Kojima-Kita, Tamara Larasati, and Masahito Ikawa. Co-expression of sperm membrane proteins cmtm2a and cmtm2b is essential for adam3 localization and male fertility in mice. *Journal of cell science*, 131, October 2018.

[123] Norihiko Furuya, Shin-Ichi Ikeda, Shigeto Sato, Sanae Soma, Junji Ezaki, Juan Alejandro Oliva Trejo, Mitsue Takeda-Ezaki, Tsutomu Fujimura, Eri Arikawa-Hirasawa, Norihiro Tada, Masaaki Komatsu, Keiji Tanaka, Eiki Kominami, Nobutaka Hattori, and Takashi Ueno. Park2/parkin-mediated mitochondrial clearance contributes to proteasome activation during slow-twitch muscle atrophy via nfe2l1 nuclear translocation. *Autophagy*, 10:631–641, April 2014.

[124] Jenna L Galloway and Leonard I Zon. Ontogeny of hematopoiesis: examining the emergence of hematopoietic cells in the vertebrate embryo. *Current topics in developmental biology*, 53:139–158, 2003.

[125] Shujie Gan, Shenlong Qiu, Yiwen Feng, Yanping Zhang, Qin Qian, Zhong Wan, and Jingdong Tang. Identification of genes associated with the effect of inflammation on the neurotransmission of vascular smooth muscle cell. *Experimental and therapeutic medicine*, 13:1303–1312, April 2017.

[126] Fleur C Garton, Peter J Houweling, Damjan Vukcevic, Lyra R Meehan, Fiona X Z Lee, Monkol Lek, Kelly N Roeszler, Marshall W Hogarth, Chrystal F Tiong, Diana Zannino, Nan Yang, Stephen Leslie, Paul Gregorevic, Stewart I Head, Jane T Seto, and Kathryn N North. The effect of actn3 gene doping on skeletal muscle performance. *American journal of human genetics*, 102:845–857, May 2018.

[127] Laurent Gautier, Leslie Cope, Benjamin M Bolstad, and Rafael A Irizarry. affy–analysis of affymetrix genechip data at the probe level. *Bioinformatics*, 20(3):307–315, Feb 2004.

[128] Krista A Geister, Michelle L Brinkmeier, Leonard Y Cheung, Jennifer Wendt, Melissa J Oatley, Daniel L Burgess, Kenneth M Kozloff, James D Cavalcoli, Jon M Oatley, and Sally A Camper. Line-1 mediated insertion into poc1a (protein of centriole 1 a) causes growth insufficiency and male infertility in mice. *PLoS genetics*, 11:e1005569, October 2015.

[129] Yingcai Geng, Yu Song, Zhenming Zhang, Hanlu Zhang, Yi Huang, and Yun Wang. [the role of tnip1 in the pathogenesis of myasthenia gravis among patients with thymoma]. *Zhonghua yi xue yi chuan xue za zhi = Zhonghua yixue yichuanxue zazhi = Chinese journal of medical genetics*, 33:615–618, October 2016.

[130] R. C. Gentleman, V. J. Carey, W. Huber, and Sandrine Dudoit Rafael A. Irizarry. *Bioinformatic and Computational Biology Solutions Using R and Bioconductor*. Springer, 2005.

[131] T J Gibson, C Ramu, C Gemünd, and R Aasland. The apeced polyglandular autoimmune syndrome protein, aire-1, contains the sand domain and is probably a transcription factor. *Trends in biochemical sciences*, 23:242–244, July 1998.

[132] Matthew J Giefer, Mark E Lowe, Steven L Werlin, Bridget Zimmerman, Michael Wilschanski, David Troendle, Sarah Jane Schwarzenberg, John F Pohl, Joseph Palermo, Chee Y Ooi, Veronique D Morinville, Tom K Lin, Sohail Z Husain, Ryan Himes, Melvin B Heyman,

Tanja Gonska, Cheryl E Gariepy, Steven D Freedman, Douglas S Fishman, Melena D Bellin, Bradley Barth, Maisam Abu-El-Haija, and Aliye Uc. Early-onset acute recurrent and chronic pancreatitis is associated with prss1 or ctrc gene mutations. *The Journal of pediatrics*, 186:95–100, July 2017.

[133] Alexander Giese, Rony Jude, Heidi Kuiper, Terje Raudsepp, Francois Piumi, Alexandra Schambony, Gérard Guérin, Bhanu P Chowdhary, Ottmar Distl, Edda Töpfer-Petersen, and Tosso Leeb. Molecular characterization of the equine testis-specific protein 1 (tpx1) and acidic epididymal glycoprotein 2 (aeg2) genes encoding members of the cysteine-rich secretory protein (crisp) family. *Gene*, 299:101–109, October 2002.

[134] K Giesecke, H Hamann, K F Stock, A Woehlke, H Sieme, and O Distl. Evaluation of spata1-associated markers for stallion fertility. *Animal genetics*, 40:359–365, August 2009.

[135] Cesidio Giuliani, Ines Bucci, and Giorgio Napolitano. The role of the transcription factor nuclear factor-kappa b in thyroid autoimmunity and cancer. *Frontiers in endocrinology*, 9:471, 2018.

[136] M F Gjerstorff and H J Ditzel. An overview of the gage cancer/testis antigen family with the inclusion of newly identified members. *Tissue antigens*, 71:187–192, March 2008.

[137] M A Goldman. The chromatin domain as a unit of gene regulation. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 9:50–55, 1988.

[138] A W Goldrath and M J Bevan. Selecting and maintaining a diverse t-cell repertoire. *Nature*, 402:255–262, November 1999.

[139] Hugo Gomez-Rueda, Rebeca Palacios-Corona, Hugo Gutiérrez-Hermosillo, and Victor Trevino. A robust biomarker of differential correlations improves the diagnosis of cytologically indeterminate thyroid cancers. *International journal of molecular medicine*, 37:1355–1362, May 2016.

[140] Qingli Gong, Xue Li, Qixing Gong, Wenyuan Zhu, Guoxin Song, and Yan Lu. Hashimoto's thyroiditis could be secondary to vitiligo: the possibility of antigen crossover and oxidative stress between the two diseases. *Archives of dermatological research*, 308:277–281, May 2016.

[141] David U Gorkin, Danny Leung, and Bing Ren. The 3d genome in transcriptional regulation and pluripotency. *Cell stem cell*, 14:762–775, June 2014.

[142] Manjunath Goroshi, Anurag R Lila, Swati S Jadhav, Sushil Sonawane, Priya Hira, Shetteppa Goroshi, Mahadev N Garle, Abhay Dalvi, Pragati Sathe, Tushar R Bandgar, and Nalini S Shah. Percentage arterial enhancement: An objective index for accurate identification of parathyroid adenoma/hyperplasia in primary hyperparathyroidism. *Clinical endocrinology*, 87:791–798, December 2017.

[143] R P Gorter, E Nutma, M-C Jahrei, J C de Jonge, R A Quinlan, P van der Valk, J M van Noort, W Baron, and S Amor. Heat shock proteins are differentially expressed in brain and spinal cord: implications for multiple sclerosis. *Clinical and experimental immunology*, 194:137–152, November 2018.

[144] Jörn Gotter, Benedikt Brors, Manfred Hergenhahn, and Bruno Kyewski. Medullary epithelial cells of the human thymus express a highly diverse selection of tissue-specific genes colocalized in chromosomal clusters. *The Journal of experimental medicine*, 199:155–166, January 2004.

[145] Ksenia J Groh, René Schönenberger, Rik I L Eggen, Helmut Segner, and Marc J-F Suter. Analysis of protein expression in zebrafish during gonad differentiation by targeted proteomics. *General and comparative endocrinology*, 193:210–220, November 2013.

[146] Jirka Grosse, Markus Wehland, Jessica Pietsch, Herbert Schulz, Katrin Saar, Norbert Hübner, Christoph Eilles, Johann Bauer, Khalil Abou-El-Ardat, Sarah Baatout, Xiao Ma, Manfred Infanger, Ruth Hemmersbach, and Daniela Grimm. Gravity-sensitive signaling drives 3-dimensional formation of multicellular thyroid cancer spheroids. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 26:5124–5140, December 2012.

[147] Huang-Tao Guan, Jin-Ming Lü, Chun-Lin Wu, Hong-Gang Li, Chang-Hong Zhu, and Cheng-Liang Xiong. [detecting testis- and epididymis-specific methylated promoters in human cell-free seminal dna by medip-qpcr]. *Zhonghua nan ke xue = National journal of andrology*, 19:977–983, November 2013.

[148] Giulia Guerriero, Samantha Trocchia, Fagr K Abdel-Gawad, and Gaetano Ciarcia. Roles of reactive oxygen species in the spermatogenesis regulation. *Frontiers in endocrinology*, 5:56, 2014.

[149] Ekaterina S Gushchanskaya, Artem V Artemov, Sergey V Ulyanov, Maria D Logacheva, Aleksey A Penin, Elena S Kotova, Sergey B Akopov, Lev G Nikolaev, Olga V Iarovaia, Eugene D Sverdlov, Alexey A Gavrilov, and Sergey V Razin. The clustering of cpg islands

may constitute an important determinant of the 3d organization of interphase chromosomes. *Epigenetics*, 9:951–963, July 2014.

[150] Dohyun Han, Sungyoon Moon, Hyunsoo Kim, Sung-E Choi, Soo-Jin Lee, Kyong Soo Park, Heesook Jun, Yup Kang, and Youngsoo Kim. Detection of differential proteomes associated with the development of type 2 diabetes in the zucker rat model using the itraq technique. *Journal of proteome research*, 10:564–577, February 2011.

[151] S Y Han, L Zhou, A Upadhyaya, S H Lee, K L Parker, and J DeJong. Tfiiaalpha/beta-like factor is encoded by a germ cell-specific gene whose expression is up-regulated with other general transcription factors during spermatogenesis in the mouse. *Biology of reproduction*, 64:507–517, February 2001.

[152] M-S Hashemi, H Mozdarani, K Ghaedi, and M H Nasr-Esfahani. Among seven testis-specific molecular markers, spem1 appears to have a significant clinical value for prediction of sperm retrieval in azoospermic men. *Andrology*, 6:890–895, November 2018.

[153] Junyi He, Na Shen, and Xinsheng Huang. Thyroid carcinoma cells produce plgf to enhance metastasis. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, 36:8601–8607, November 2015.

[154] Yu He, Dan Zhang, Yanfang Yang, Xixi Wang, Xinyu Zhao, Peng Zhang, Hongxia Zhu, Ningzhi Xu, and Shufang Liang. A double-edged function of ddx3, as an oncogene or tumor suppressor, in cancer progression (review). *Oncology reports*, 39:883–892, March 2018.

[155] M Heino, P Peterson, J Kudoh, K Nagamine, A Lagerstedt, V Ovod, A Ranki, I Rantala, M Nieminen, J Tuukkanen, H S Scott, S E Antonarakis, N Shimizu, and K Krohn. Autoimmune regulator is expressed in the cells regulating immune tolerance in thymus medulla. *Biochemical and biophysical research communications*, 257:821–825, April 1999.

[156] M Heino, P Peterson, J Kudoh, N Shimizu, S E Antonarakis, H S Scott, and K Krohn. Apeced mutations in the autoimmune regulator (aire) gene. *Human mutation*, 18:205–211, September 2001.

[157] Mizue Hisano, Hiroshi Ohta, Yoshitake Nishimune, and Masami Nozaki. Methylation of cpg dinucleotides in the open reading frame of a testicular germ cell-specific intronless gene, tact1/actl7b, represses its expression in somatic cells. *Nucleic acids research*, 31:4797–4804, August 2003.

[158] Dirk Holzinger, Dirk Foell, and Christoph Kessel. The role of s100 proteins in the pathogenesis and monitoring of autoinflammatory diseases. *Molecular and cellular pediatrics*, 5:7, September 2018.

[159] Yu Hong, Hai-Feng Li, Geir Olve Skeie, Fredrik Romi, Hong-Jun Hao, Xu Zhang, Xiang Gao, Jone Furlund Owe, and Nils Erik Gilhus. Autoantibody profile and clinical characteristics in a cohort of chinese adult myasthenia gravis patients. *Journal of neuroimmunology*, 298:51–57, September 2016.

[160] L L Hsiao, F Dangond, T Yoshida, R Hong, R V Jensen, J Misra, W Dillon, K F Lee, K E Clark, P Haverty, Z Weng, G L Mutter, M P Frosch, M E MacDonald, E L Milford, C P Crum, R Bueno, R E Pratt, M Mahadevappa, J A Warrington, G Stephanopoulos, G Stephanopoulos, and S R Gullans. A compendium of gene expression in normal human tissues. *Physiological genomics*, 7:97–104, December 2001.

[161] Jiankang Huang, Ting Zhang, Hongmei Wang, and Yuwu Zhao. Treatment of experimental autoimmune myasthenia gravis rats with fty720 and its effect on th1/th2 cells. *Molecular medicine reports*, 17:7409–7414, May 2018.

[162] Min Huang, Hui Wang, Jianmin Li, Zuomin Zhou, Yong Du, Min Lin, and Jiahao Sha. Involvement of alf in human spermatogenesis and male infertility. *International journal of molecular medicine*, 17:599–604, April 2006.

[163] Zaohua Huang, Polina V Danshina, Kathleen Mohr, Weidong Qu, Summer G Goodson, Thomas M O'Connell, and Deborah A O'Brien. Sperm function, protein phosphorylation, and metabolism differ in mice lacking successive sperm-specific glycolytic enzymes. *Biology of reproduction*, 97:586–597, October 2017.

[164] Wolfgang Huber, Vincent J Carey, Robert Gentleman, Simon Anders, Marc Carlson, Benilton S Carvalho, Hector Corrada Bravo, Sean Davis, Laurent Gatto, Thomas Girke, Raphael Gottardo, Florian Hahne, Kasper D Hansen, Rafael A Irizarry, Michael Lawrence, Michael I Love, James MacDonald, Valerie Obenchain, Andrzej K Oleś, Hervé Pagès, Alejandro Reyes, Paul Shannon, Gordon K Smyth, Dan Tenenbaum, Levi Waldron, and Martin Morgan. Orchestrating high-throughput genomic analysis with bioconductor. *Nature methods*, 12:115–121, February 2015.

[165] Wolfgang Huber, Anja von Heydebreck, Holger Sültmann, Annemarie Poustka, and Martin Vingron. Variance stabilization applied to mi-

croarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18 Suppl 1:S96–104, 2002.

[166] M E Huot, R Mazroui, P Leclerc, and E W Khandjian. Developmental expression of the fragile x-related 1 proteins in mouse testis: association with microtubule elements. *Human molecular genetics*, 10:2803–2811, November 2001.

[167] Laurence D Hurst, Csaba Pál, and Martin J Lercher. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet*, 5(4):299–310, Apr 2004.

[168] Tihana Ibrahimpasic, Bin Xu, Iñigo Landa, Snjezana Dogan, Sumit Middha, Venkatraman Seshan, Shyam Deraje, Diane L Carlson, Jocelyn Migliacci, Jeffrey A Knauf, Brian Untch, Michael F Berger, Luc Morris, R Michael Tuttle, Timothy Chan, James A Fagin, Ronald Ghossein, and Ian Ganly. Genomic alterations in fatal forms of non-anaplastic thyroid cancer: Identification of , javax.xml.bind.jaxbelement@197c37f3, and , javax.xml.bind.jaxbelement@718f1c90, as novel thyroid cancer genes associated with tumor virulence. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 23:5970–5980, October 2017.

[169] Masahito Ikawa, Keizo Tokuhiro, Ryo Yamaguchi, Adam M Benham, Taku Tamura, Ikuo Wada, Yuhkoh Satouh, Naokazu Inoue, and Masaru Okabe. Calsperin is a testis-specific chaperone required for sperm fertility. *The Journal of biological chemistry*, 286:5639–5646, February 2011.

[170] Jose M Irimia, Vincent S Tagliabracci, Catalina M Meyer, Dyann M Segvich, Anna A DePaoli-Roach, and Peter J Roach. Muscle glycogen remodeling and glycogen phosphate metabolism following exhaustive exercise of wild type and laforin knockout mice. *The Journal of biological chemistry*, 290:22686–22698, September 2015.

[171] Rafael A Irizarry, Bridget Hobbs, Francois Collin, Yasmin D Beazer-Barclay, Kristen J Antonellis, Uwe Scherf, and Terence P Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England)*, 4:249–264, April 2003.

[172] Kosuke Ishikawa, Sakura Azuma, Shuntaro Ikawa, Yasuyuki Morishita, Jin Gohda, Taishin Akiyama, Kentaro Semba, and Jun ichiro Inoue. Cloning and characterization of xenopus laevis drg2, a member of the developmentally regulated gtp-binding protein subfamily. *Gene*, 322:105–112, December 2003.

[173] Keita Iyama, Michiko Matsuse, Norisato Mitsutake, Tatiana Rogounovitch, Vladimir Saenko, Keiji Suzuki, Mai Ashizawa, Chiyo Ookouchi, Satoshi Suzuki, Hiroshi Mizunuma, Toshihiko Fukushima, Shinichi Suzuki, and Shunichi Yamashita. Identification of three novel fusion oncogenes, sqstm1/ntrk3, afap1l2/ret, and ppfibp2/ret, in thyroid cancers of young patients in fukushima. *Thyroid : official journal of the American Thyroid Association*, 27:811–818, June 2017.

[174] Nirmala Jagadish, Deepak Parashar, Namita Gupta, Sumit Agarwal, Aditi Sharma, Rukhsar Fatima, Vaishali Suri, Rajive Kumar, Anju Gupta, Nirmal Kumar Lohiya, and Anil Suri. A novel cancer testis antigen target a-kinase anchor protein (akap4) for the early diagnosis and immunotherapy of colon cancer. *Oncoimmunology*, 5:e1078965, February 2016.

[175] Madhumita Jagannathan-Bogdan and Leonard I Zon. Hematopoiesis. *Development (Cambridge, England)*, 140:2463–2467, June 2013.

[176] Charles A Janeway, Paul Travers, Mark Walport, and Mark Shlomchik. *Immunobiology: The Immune System in Health and Disease - fifth edition*. Garland Publishing Group, 2001.

[177] S Jarius, F Paul, O Aktas, N Asgari, R C Dale, J de Seze, D Franciotta, K Fujihara, A Jacob, H J Kim, I Kleiter, T Kümpfel, M Levy, J Palace, K Ruprecht, A Saiz, C Trebst, B G Weinshenker, and B Wildemann. Mog encephalomyelitis: international recommendations on diagnosis and antibody testing. *Journal of neuroinflammation*, 15:134, May 2018.

[178] Gui-Yang Jiang, Xiu-Peng Zhang, Liang Wang, Xu-Yong Lin, Juan-Han Yu, En-Hua Wang, and Yong Zhang. Fbxo25 promotes cell proliferation, invasion, and migration of nsclc. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, 37:14311–14319, October 2016.

[179] Samantha Jilek, Myriam Schluep, Giuseppe Pantaleo, and Renaud A Du Pasquier. Mobp-specific cellular immune responses are weaker than mog-specific cellular immune responses in patients with multiple sclerosis and healthy subjects. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 34:539–543, April 2013.

[180] Hong-Jian Jin, Segun Jung, Auditi R DebRoy, and Ramana V Davuluri. Identification and validation of regulatory snps that modulate transcription factor chromatin binding and gene expression in prostate cancer. *Oncotarget*, 7:54616–54626, August 2016.

[181] Bente B Johansson, Karianne Fjeld, Khadija El Jellas, Anny Gravdal, Monica Dalva, Erling Tjora, Helge Ræder, Rohit N Kulkarni, Stefan Johansson, Pål R Njølstad, and Anders Molven. The role of the carboxyl ester lipase (cel) gene in pancreatic disease. *Pancreatology : official journal of the International Association of Pancreatology (IAP) ... [et al.]*, 18:12–19, January 2018.

[182] Jonathan B Johnnidis, Emily S Venanzi, Debra J Taxman, Jenny P-Y Ting, Christophe O Benoist, and Diane J Mathis. Chromosomal clustering of genes controlled by the aire transcription factor. *Proceedings of the National Academy of Sciences of the United States of America*, 102:7233–7238, May 2005.

[183] Jason M Johnson, John Castle, Philip Garrett-Engele, Zhengyan Kan, Patrick M Loerch, Christopher D Armour, Ralph Santos, Eric E Schadt, Roland Stoughton, and Daniel D Shoemaker. Genome-wide survey of human alternative pre-mrna splicing with exon junction microarrays. *Science (New York, N.Y.)*, 302:2141–2144, December 2003.

[184] C Jolicoeur, D Hanahan, and K M Smith. T-cell tolerance toward a transgenic beta-cell antigen and transcription of endogenous pancreatic genes in thymus. *Proceedings of the National Academy of Sciences of the United States of America*, 91:6707–6711, July 1994.

[185] Allison Jordan and Miriam Freimer. Recent advances in understanding and managing myasthenia gravis. *F1000Research*, 7, 2018.

[186] Yuko Kamata, Akiko Kuhara, Takeo Iwamoto, Kazumi Hayashi, Shigeo Koido, Takahiro Kimura, Shin Egawa, and Sadamu Homma. Identification of hla class i-binding peptides derived from unique cancer-associated proteins by mass spectrometric analysis. *Anticancer research*, 33:1853–1859, May 2013.

[187] Z Kan, E C Rouchka, W R Gish, and D J States. Gene structure prediction and alternative splicing analysis using genomically aligned ests. *Genome research*, 11:889–900, May 2001.

[188] Madiha Kanwal, Xiao-Jie Ding, Zhans-Han Ma, Lian-Wei Li, Ping Wang, Ying Chen, Yun-Chao Huang, and Yi Cao. Characterization of germline mutations in familial lung cancer from the chinese population. *Gene*, 641:94–104, January 2018.

[189] Justin C Kao, Margherita Milone, Duygu Selcen, Xin-Ming Shen, Andrew G Engel, and Teerin Liewluck. Congenital myasthenic syndromes in adult neurology clinic: A long road to diagnosis and therapy. *Neurology*, 91:e1770–e1777, November 2018.

[190] Kristen L Karlin, Gourish Mondal, Jessica K Hartman, Siddhartha Tyagi, Sarah J Kurley, Chris S Bland, Tiffany Y T Hsu, Alexander Renwick, Justin E Fang, Ilenia Migliaccio, Celetta Callaway, Amritha Nair, Rocio Dominguez-Vidana, Don X Nguyen, C Kent Osborne, Rachel Schiff, Li-Yuan Yu-Lee, Sung Y Jung, Dean P Edwards, Susan G Hilsenbeck, Jeffrey M Rosen, Xiang H-F Zhang, Chad A Shaw, Fergus J Couch, and Thomas F Westbrook. The oncogenic stp axis promotes triple-negative breast cancer via degradation of the rest tumor suppressor. *Cell reports*, 9:1318–1332, November 2014.

[191] Anastasia Katsarou, Soffia Gudbjörnsdottir, Araz Rawshani, Dana Dabelea, Ezio Bonifacio, Barbara J Anderson, Laura M Jacobsen, Desmond A Schatz, and Åke Lernmark. Type 1 diabetes mellitus. *Nature reviews. Disease primers*, 3:17016, March 2017.

[192] Stefan H E Kaufmann. Remembering emil von behring: from tetanus treatment to antibody cooperation with phagocytes. *mBio*, 8, February 2017.

[193] Raghuveer Kavarthapu and Maria L Dufau. Germ cell nuclear factor (gcnf/rtr) regulates transcription of gonadotropin-regulated testicular rna helicase (grth/ddx25) in testicular germ cells–the androgen connection. *Molecular endocrinology (Baltimore, Md.)*, 29:1792–1804, December 2015.

[194] Eiji Kawasaki, Yasunori Sera, Norio Abiru, Mikako Yamauchi, Masako Ozaki, Kenichi Yamakawa, Takahiro Abe, Shigeo Uotani, Hironori Yamasaki, Yoshihiko Yamaguchi, and Katsumi Eguchi. Distinct ia-2 autoantibody epitope recognition between childhood-onset and adult-onset type 1 diabetes. *Annals of the New York Academy of Sciences*, 958:235–240, April 2002.

[195] Christoph Kessel, Dirk Holzinger, and Dirk Foell. Phagocyte-derived s100 proteins in autoinflammation: putative role in pathogenesis and usefulness as biomarkers. *Clinical immunology (Orlando, Fla.)*, 147:229–241, June 2013.

[196] Pavel P Khil, Natalya A Smirnova, Peter J Romanienko, and R Daniel Camerini-Otero. The mouse x chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nature genetics*, 36:642–646, June 2004.

[197] Ekyune Kim, Ki-Eun Park, Ji-Su Kim, Dong Chul Baek, Jae-Woong Lee, Sang-Rae Lee, Myeong-Su Kim, Sang-Hyun Kim, Chan-Shick Kim, Deog-Bon Koo, Han-Seok Kang, Zae-Young Ryoo, and Kyu-Tae Chang. Importance of the porcine adam3 disintegrin domain in

sperm-egg interaction. *The Journal of reproduction and development*, 55:156–162, April 2009.

[198] Ji Won Kim, Oh Yun Kwon, and Myoung Hee Kim. Differentially expressed genes and morphological changes during lengthened immobilization in rat soleus muscle. *Differentiation; research in biological diversity*, 75:147–157, February 2007.

[199] John J Kim, Krithika Rajagopalan, Basil Hussain, Brenten H Williams, Prakash Kulkarni, and Steven M Mooney. Cetn1 is a cancer testis antigen with expression in prostate and pancreatic cancers. *Biomarker research*, 1:22, June 2013.

[200] Pora Kim, Aekyung Park, Guangchun Han, Hua Sun, Peilin Jia, and Zhongming Zhao. Tissgdb: tissue-specific gene database in cancer. *Nucleic acids research*, 46:D1031–D1038, January 2018.

[201] Katherine Y King and Margaret A Goodell. Inflammatory modulation of hscs: viewing the hsc as a foundation for the immune response. *Nat Rev Immunol*, 11(10):685–692, Oct 2011.

[202] L Klein, T Klein, U Rüther, and B Kyewski. Cd4 t cell tolerance to human c-reactive protein, an inducible serum protein, is mediated by medullary thymic epithelium. *The Journal of experimental medicine*, 188:5–16, July 1998.

[203] L Klein and B Kyewski. Self-antigen presentation by thymic stromal cells: a subtle division of labor. *Current opinion in immunology*, 12:179–186, April 2000.

[204] Ludger Klein, Maria Hinterberger, Gerald Wirnsberger, and Bruno Kyewski. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nat Rev Immunol*, 9(12):833–844, Dec 2009.

[205] Kazuhiko Kogawa, Seiho Nagafuchi, Hitoshi Katsuta, Jun Kudoh, Sadafumi Tamiya, Yumiko Sakai, Nobuyoshi Shimizu, and Mine Harada. Expression of aire gene in peripheral monocyte/dendritic cell lineage. *Immunology letters*, 80:195–198, March 2002.

[206] Hyun Kyung Kong, Sae Jeong Park, Ye Sol Kim, Kyoung Min Kim, Hyun-Woo Lee, Hyeok-Gu Kang, Yu Mi Woo, Eun Young Park, Je Yeong Ko, Hiromu Suzuki, Kyung-Hee Chun, Erwei Song, Kyu Yun Jang, and Jong Hoon Park. Epigenetic activation of ly6k predicts the presence of metastasis and poor prognosis in breast carcinoma. *Oncotarget*, 7:55677–55689, August 2016.

[207] Satu O A Koskinen, Heikki Kyröläinen, Riina Flink, Harri P Selänne, Sheila S Gagnon, Juha P Ahtiainen, Bradley C Nindl, and Maarit Lehti. Human skeletal muscle type 1 fibre distribution and response of stress-sensing proteins along the titin molecule after submaximal exhaustive exercise. *Histochemistry and cell biology*, 148:545–555, November 2017.

[208] Mina Kozai, Yuki Kubo, Tomoya Katakai, Hiroyuki Kondo, Hiroshi Kiyonari, Karin Schaeuble, Sanjiv A Luther, Naozumi Ishimaru, Izumi Ohigashi, and Yousuke Takahama. Essential role of ccl21 in establishment of central self-tolerance in t cells. *The Journal of experimental medicine*, 214:1925–1935, July 2017.

[209] H Kubota, G M Hynes, S M Kerr, and K R Willison. Tissue-specific subunit of the mouse cytosolic chaperonin-containing tcp-1. *FEBS letters*, 402:53–56, January 1997.

[210] F Kuchenbauer, C Schoch, E Holler, T Haferlach, W Hiddemann, and S Schnittger. A rare case of acute myeloid leukemia with a chic2-etv6 fusiongen and multiple other molecular aberrations. *Leukemia*, 19:2366–2368, December 2005.

[211] Prakash Kulkarni, Takumi Shiraishi, Krithika Rajagopalan, Robert Kim, Steven M Mooney, and Robert H Getzenberg. Cancer/testis antigens and urological malignancies. *Nature reviews. Urology*, 9:386–396, June 2012.

[212] Atul Kumar, Lars Paeger, Kosmas Kosmas, Peter Kloppenburg, Angelika A Noegel, and Vivek S Peche. Neuronal actin dynamics, spine density and neuronal dendritic complexity are regulated by cap2. *Frontiers in cellular neuroscience*, 10:180, 2016.

[213] P G Kumar, M Laloraya, C Y Wang, Q G Ruan, A Davoodi-Semiromi, K J Kao, and J X She. The autoimmune regulator (aire) is a dna-binding protein. *The Journal of biological chemistry*, 276:41357–41364, November 2001.

[214] Bruno Kyewski and Jens Derbinski. Self-representation in the thymus: an extended view. *Nat Rev Immunol*, 4(9):688–698, Sep 2004.

[215] Bruno Kyewski, Jens Derbinski, Jörn Gotter, and Ludger Klein. Promiscuous gene expression and central t-cell tolerance: more than meets the eye. *Trends Immunol*, 23(7):364–371, Jul 2002.

[216] Bruno Kyewski and Ludger Klein. A central role for central tolerance. *Annual review of immunology*, 24:571–606, 2006.

[217] Malin Lando, Marit Holden, Linn C Bergersen, Debbie H Svendsrud, Trond Stokke, Kolbein Sundfør, Ingrid K Glad, Gunnar B Kristensen, and Heidi Lyng. Gene dosage, expression, and ontology analysis identifies driver genes in the carcinogenesis and chemoradioresistance of cervical cancer. *PLoS genetics*, 5:e1000719, November 2009.

[218] Jane E Lattin, Kate Schroder, Andrew I Su, John R Walker, Jie Zhang, Tim Wiltshire, Kaoru Saijo, Christopher K Glass, David A Hume, Stuart Kellie, and Matthew J Sweet. Expression analysis of g protein-coupled receptors in mouse macrophages. *Immunome Res*, 4:5, 2008.

[219] Ho-Sun Lee, Yunkyeong Kang, Kyung Tae, Gyu-Un Bae, Jong Y Park, Yoon Hee Cho, and Mihi Yang. Proteomic biomarkers for bisphenol a-early exposure and women's thyroid cancer. *Cancer research and treatment : official journal of Korean Cancer Association*, 50:111–117, January 2018.

[220] Katie S Lennard, Ryan W Goosen, and Jonathan M Blackburn. Bacterially-associated transcriptional remodelling in a distinct genomic subtype of colorectal cancer provides a plausible molecular basis for disease development. *PloS one*, 11:e0166282, 2016.

[221] Martin J Lercher, Thomas Blumenthal, and Laurence D Hurst. Co-expression of neighboring genes in caenorhabditis elegans is mostly due to operons and duplicate genes. *Genome research*, 13:238–243, February 2003.

[222] Martin J Lercher, Araxi O Urrutia, and Laurence D Hurst. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature genetics*, 31:180–183, June 2002.

[223] Ake Lernmark and Helena Elding Larsson. Immune therapy in type 1 diabetes mellitus. *Nature reviews. Endocrinology*, 9:92–103, February 2013.

[224] Hui Li, Xiaoyan Li, Jian Guo, Guifu Wu, Chunping Dong, Yaling Pang, Shan Gao, and Yangwei Wang. Identification of biomarkers and mechanisms of diabetic cardiomyopathy using microarray data. *Cardiology journal*, September 2018.

[225] Yifei Li, Ran Xia, Bo Zhang, and Chunsheng Li. Chronic atrophic gastritis: A review. *Journal of environmental pathology, toxicology and oncology : official organ of the International Society for Environmental Toxicology and Cancer*, 37:241–259, 2018.

[226] Leilei Liang, Jingchao Xu, Meng Wang, Gaoran Xu, Ning Zhang, Guangzhi Wang, and Yongfu Zhao. Lncrna hcp5 promotes follicu-

lar thyroid carcinoma progression via mirnas sponge. *Cell death & disease*, 9:372, March 2018.

[227] Teerin Liewluck, Xin-Ming Shen, Margherita Milone, and Andrew G Engel. Endplate structure and parameters of neuromuscular transmission in sporadic centronuclear myopathy associated with myasthenia. *Neuromuscular disorders : NMD*, 21:387–395, June 2011.

[228] Amanda P Liggins, Seah H Lim, Elizabeth J Soilleux, Karen Pulford, and Alison H Banham. A panel of cancer-testis genes exhibiting broad-spectrum expression in haematological malignancies. *Cancer immunity*, 10:8, August 2010.

[229] Shou-Ren Lin, Yu-Chi Li, Man-Ling Luo, Huan Guo, Tian-Tian Wang, Jian-Bo Chen, Qian Ma, Yan-Li Gu, Zhi-Mao Jiang, and Yao-Ting Gui. Identification and characteristics of the testes-specific gene, ccdc38, in mice. *Molecular medicine reports*, 14:1290–1296, August 2016.

[230] Inka Lindner, Ulf Helwig, Diana Rubin, Yun Li, Eva Fisher, Heiner Boeing, Matthias Möhlig, Joachim Spranger, Andreas Pfeiffer, Jochen Hampe, Stefan Schreiber, Frank Döring, and Jürgen Schrezenmeir. Putative association between a new polymorphism in exon 3 (arg109cys) of the pancreatic colipase gene and type 2 diabetes mellitus in two independent caucasian study populations. *Molecular nutrition & food research*, 49:972–976, October 2005.

[231] Guoping Liu, Kainan Wu, and Yuan Sheng. Elucidation of the molecular mechanisms of anaplastic thyroid carcinoma by integrated mirna and mrna analysis. *Oncology reports*, 36:3005–3013, November 2016.

[232] Lian-Bin Liu, Jing Huang, Jin-Ping Zhong, Gui-Lin Ye, Ling Xue, Mao-Hua Zhou, Gang Huang, and Shao-Jin Li. High expression of ccdc34 is associated with poor survival in cervical cancer patients. *Medical science monitor : international medical journal of experimental and clinical research*, 24:8383–8390, November 2018.

[233] Xiong Liu, Xueping Yu, Donald J Zack, Heng Zhu, and Jiang Qian. Tiger: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics*, 9:271, 2008.

[234] Wei Long, Guanting Lu, Wenbai Zhou, Yuqi Yang, Bin Zhang, Hong Zhou, Lihua Jiang, and Bin Yu. Targeted next-generation sequencing of thirteen causative genes in chinese patients with congenital hypothyroidism. *Endocrine journal*, 65:1019–1028, October 2018.

[235] A Lounici Boudiaf, D Bouziane, M Smara, Y Meddour, E M Haffaf, B Oudjit, S Chaib Mamouzi, and S Aouichat Bouguerra. Could znt8 antibodies replace ica, gad, ia2 and insulin antibodies in the diagnosis of type 1 diabetes? *Current research in translational medicine*, 66:1–7, March 2018.

[236] Weng-Onn Lui, Theodoros Foukakis, Johan Lidén, Srinivasan R Thoppe, Trisha Dwight, Anders Höög, Jan Zedenius, Göran Wallin, Mark Reimers, and Catharina Larsson. Expression profiling reveals a distinct transcription signature in follicular thyroid carcinomas with a pax8-ppar(gamma) fusion oncogene. *Oncogene*, 24:1467–1476, February 2005.

[237] A P López-Cardona, I Ibarra-Lecue, R Laguna-Barraza, S Pérez-Cerezales, L Urigüen, N Agirregoitia, A Gutiérrez-Adán, and E Agirregoitia. Effect of chronic thc administration in the reproductive organs of male mice, spermatozoa and in vitro fertilization. *Biochemical pharmacology*, 157:294–303, November 2018.

[238] J-Matthias Löhr, Ralf Faissner, Dirk Koczan, Peter Bewerunge, Claudio Bassi, Benedikt Brors, Roland Eils, Luca Frulloni, Anette Funk, Walter Halangk, Ralf Jesenofsky, Ralf Jesnowski, Lars Kaderali, Jörg Kleeff, Burkhard Krüger, Markus M Lerch, Ralf Lösel, Mauro Magnani, Michael Neumaier, Stephanie Nittka, Miklós Sahin-Tóth, Julian Sänger, Sonja Serafini, Martina Schnölzer, Hermann-Josef Thierse, Silke Wandschneider, Giuseppe Zamboni, and Günter Klöppel. Autoantibodies against the exocrine pancreas in autoimmune pancreatitis: gene and protein expression profiling and immunoassays identify pancreatic enzymes as a major target of the inflammatory process. *The American journal of gastroenterology*, 105:2060–2071, September 2010.

[239] Carlson M. Go.db: A set of annotation maps describing the entire gene ontology. r package version 3.4.1. 2017.

[240] Yuhang Ma, Jing Gao, Jiajing Yin, Liping Gu, Xing Liu, Su Chen, Qianfang Huang, Huifang Lu, Yuemin Yang, Hu Zhou, Yufan Wang, and Yongde Peng. Identification of a novel function of adipocyte plasma membrane-associated protein (apmap) in gestational diabetes mellitus by proteomic analysis of omental adipose tissue. *Journal of proteome research*, 15:628–637, February 2016.

[241] Zhen-Zhen Ma, Hong-Sheng Sun, Ji-Cai Lv, Lei Guo, and Qing-Rui Yang. Expression and clinical significance of the nek7-nlrp3 inflammasome signaling pathway in patients with systemic lupus erythematosus. *Journal of inflammation (London, England)*, 15:16, 2018.

[242] G Maccaferri, J D Roberts, P Szucs, C A Cottingham, and P Somogyi. Cell surface domain specific postsynaptic currents evoked by identified gabaergic neurones in rat hippocampus in vitro. *The Journal of physiology*, 524 Pt 1:91–116, April 2000.

[243] Emeaga Maheswaran, Christina B Pedersen, Henrik J Ditzel, and Morten F Gjerstorff. Lack of adam2, calr3 and sage1 cancer/testis antigen expression in lung and breast cancer. *PloS one*, 10:e0134967, 2015.

[244] Zabed Mahmud, Somaya Zahran, Philip B Liu, Bela Reiz, Brandon Y H Chan, Andrej Roczkowsky, Christian-Scott E McCartney, Peter L Davies, Liang Li, Richard Schulz, and Peter M Hwang. Structure and proteolytic susceptibility of the inhibitory c-terminal tail of cardiac troponin i. *Biochimica et biophysica acta. General subjects*, January 2019.

[245] Virginie Maire, Faisal Mahmood, Guillem Rigaill, Mengliang Ye, Amélie Brisson, Fariba Némati, David Gentien, Gordon C Tucker, Sergio Roman-Roman, and Thierry Dubois. Lrp8 is overexpressed in estrogen-negative breast cancers and a potential target for these tumors. *Cancer medicine*, December 2018.

[246] Oksana Maksimenko and Pavel Georgiev. Mechanisms and proteins involved in long-distance interactions. *Frontiers in genetics*, 5:28, 2014.

[247] L Mannucci, F Emma, M Markert, C Bachmann, O Boulat, R Carrozzo, G Rizzoni, and C Dionisi-Vici. Increased no production in lysinuric protein intolerance. *Journal of inherited metabolic disease*, 28:123–129, 2005.

[248] Atsushi Masamune. Genetics of pancreatitis: the 2014 update. *The Tohoku journal of experimental medicine*, 232:69–77, February 2014.

[249] Mehrdad Matloubian, Charles G Lo, Guy Cinamon, Matthew J Lesneski, Ying Xu, Volker Brinkmann, Maria L Allende, Richard L Proia, and Jason G Cyster. Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on s1p receptor 1. *Nature*, 427:355–360, January 2004.

[250] Chie Matsuda, Kimihiko Kameyama, Atsushi Suzuki, Wataru Mishima, Satoshi Yamaji, Harumasa Okamoto, Ichizo Nishino, and Yukiko K Hayashi. Affixin activates rac1 via betapix in c2c12 myoblast. *FEBS letters*, 582:1189–1196, April 2008.

[251] Yoshiaki Matsumoto, Junji Itou, Fumiaki Sato, and Masakazu Toi. Sall4 - khdrbs3 network enhances stemness by modulating cd44 splic-

ing in basal-like breast cancer. *Cancer medicine*, 7:454–462, February 2018.

[252] N Mattsson, M Axelsson, S Haghighi, C Malmeström, G Wu, R Anckarsäter, S Sankaranarayanan, U Andreasson, S Fredrikson, A Gundersen, L Johnsen, T Fladby, A Tarkowski, E Trysberg, A Wallin, H Anckarsäter, J Lycke, O Andersen, A J Simon, K Blennow, and H Zetterberg. Reduced cerebrospinal fluid bace1 activity in multiple sclerosis. *Multiple sclerosis (Houndmills, Basingstoke, England)*, 15:448–454, April 2009.

[253] Lior Mayo, Sunia A Trauger, Manon Blain, Meghan Nadeau, Bonny Patel, Jorge I Alvarez, Ivan D Mascanfroni, Ada Yeste, Pia Kivisäkk, Keith Kallas, Benjamin Ellezam, Rohit Bakshi, Alexandre Prat, Jack P Antel, Howard L Weiner, and Francisco J Quintana. Regulation of astrocyte activation by glycolipids drives chronic cns inflammation. *Nature medicine*, 20:1147–1156, October 2014.

[254] Andrew McQuillin, Mie Rizig, and Hugh M D Gurling. A microarray gene expression study of the molecular pharmacology of lithium carbonate on mouse brain mrna to understand the neurobiology of mood stabilization and treatment of bipolar affective disorder. *Pharmacogenetics and genomics*, 17:605–617, August 2007.

[255] Karen J Meaburn and Tom Misteli. Cell biology: chromosome territories. *Nature*, 445:379–781, January 2007.

[256] R Meccariello, G Berruti, R Chianese, R De Santis, F Di Cunto, D Scarpa, G Cobellis, I Zucchetti, R Pierantoni, F Altruda, and S Fasano. Structure of msj-1 gene in mice and humans: a possible role in the regulation of male reproduction. *General and comparative endocrinology*, 156:91–103, March 2008.

[257] Shuai Meng, Wen Zhang, Li-Juan Guan, Fatuma-Said Muhali, Jiao-Zhen Zhou, Rong-Hua Song, Jian Xu, and Jin-An Zhang. Proteomic analysis reveals aberrant expression of calr and hspa5 in thyroid tissues of graves' disease. *Clinical biochemistry*, 50:40–45, January 2017.

[258] Michael A Miller, Asher D Cutter, Ikuko Yamamoto, Samuel Ward, and David Greenstein. Clustered organization of reproductive genes in the c. elegans genome. *Current biology : CB*, 14:1284–1290, July 2004.

[259] Anna L Mitchell and Simon H S Pearce. Autoimmune addison disease: pathophysiology and genetic complexity. *Nature reviews. Endocrinology*, 8:306–316, January 2012.

[260] L Mittaz, C Rossier, M Heino, P Peterson, K J Krohn, A Gos, M A Morris, J Kudoh, N Shimizu, S E Antonarakis, and H S Scott. Isolation and characterization of the mouse aire gene. *Biochemical and biophysical research communications*, 255:483–490, February 1999.

[261] S E Monstad, J K Nøstbakken, and C A Vedeler. Crmp5 antibodies found in a patient with limbic encephalitis and myasthenia gravis. *Journal of neurology, neurosurgery, and psychiatry*, 80:241–242, February 2009.

[262] Lili Mu, Yao Zhang, Bo Sun, Jinghua Wang, Xiaoli Xie, Na Li, Jia Zhang, Qingfei Kong, Yumei Liu, Zhijuan Han, Guangyou Wang, Zheng Fu, Bo Yu, Guozhong Li, and Hulun Li. Activation of the receptor for advanced glycation end products (rage) exacerbates experimental autoimmune myasthenia gravis symptoms. *Clinical immunology (Orlando, Fla.)*, 141:36–48, October 2011.

[263] Jacob L Mueller, Shantha K Mahadevaiah, Peter J Park, Peter E Warburton, David C Page, and James M A Turner. The mouse x chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nature genetics*, 40:794–799, June 2008.

[264] A Márquez, M C Cénit, C Núñez, J L Mendoza, C Taxonera, M Díaz-Rubio, M Bartolomé, R Arroyo, M Fernández-Arquero, E G de la Concha, and E Urcelay. Effect of bsn-mst1 locus on inflammatory bowel disease and multiple sclerosis susceptibility. *Genes and immunity*, 10:631–635, October 2009.

[265] K Nagamine, P Peterson, HS Scott, J Kudoh, S Minoshima, M Heino, KJ Krohn, MD Lalioti, PE Mullis, SE Antonarakis, K Kawasaki, S Asakawa, F Ito, and N Shimizu. Positional cloning of the apeced gene. *Nat Genet*, 17(4):393–398, Dec 1997.

[266] Kazuaki Nagao, Naoyuki Togawa, Katsunori Fujii, Hideki Uchikawa, Yoichi Kohno, Masao Yamada, and Toshiyuki Miyashita. Detecting tissue-specific alternative splicing and disease-associated aberrant splicing of the ptch gene with exon junction microarrays. *Human molecular genetics*, 14:3379–3388, November 2005.

[267] Prashantha Nagaraja, Kavya Parashivamurthy, Nandini Sidnal, Siddappa Mali, Dakshyani Nagaraja, and Sivarami Reddy. Analysis of gene expression on ngn3 gene signaling pathway in endocrine pancreatic cancer. *Bioinformation*, 9:739–747, 2013.

[268] Nisha Nagarsheth, Max S Wicha, and Weiping Zou. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nature reviews. Immunology*, 17:559–572, September 2017.

[269] Michael T Nguyen, Daniel P Delaney, and Thomas F Kolon. Gene expression alterations in cryptorchid males using spermatozoal microarray analysis. *Fertility and sterility*, 92:182–187, July 2009.

[270] Alison L Nightingale, Julie E Davidson, Charles T Molta, Hong J Kan, and Neil J McHugh. Presentation of sle in uk primary care using the clinical practice research datalink. *Lupus science & medicine*, 4:e000172, 2017.

[271] Masuhiro Nishimura and Shinsaku Naito. Tissue-specific mrna expression profiles of human solute carrier transporter superfamilies. *Drug metabolism and pharmacokinetics*, 23:22–44, 2008.

[272] Pengxia Niu, Yinghui Wei, Qian Gao, Xue Zhang, Yanqing Hu, Yiqing Qiu, Yulian Mu, and Kui Li. Male fertility potential molecular mechanisms revealed by itraq-based quantitative proteomic analysis of the epididymis from wip1, javax.xml.bind.jaxbelement@51507e0f, mice. *Omics : a journal of integrative biology*, 23:54–66, January 2019.

[273] Seyedmehdi Nourashrafeddin and Batool Hosseini Rashidi. Gonadotropin regulation of retinoic acid activity in the testis. *Acta medica Iranica*, 56:34–42, January 2018.

[274] Julius O Nyalwidhe, Wojciech J Grzesik, Tanya C Burch, Michele L Semeraro, Tayab Waseem, Ivan C Gerling, Raghavendra G Mirmira, Margaret A Morris, and Jerry L Nadler. Comparative quantitative proteomic analysis of disease stratified laser captured microdissected human islets identifies proteins and pathways potentially related to type 1 diabetes. *PloS one*, 12:e0183908, 2017.

[275] Junko Odajima, Siddharth Saini, Piotr Jung, Yasmine Ndassa-Colday, Scott Ficaro, Yan Geng, Eugenio Marco, Wojciech Michowski, Yaoyu E Wang, James A DeCaprio, Larisa Litovchick, Jarrod Marto, and Piotr Sicinski. Proteomic landscape of tissue-specific cyclin e functions in vivo. *PLoS genetics*, 12:e1006429, November 2016.

[276] Satoshi Okano, Lingyun Zhou, Toshimasa Kusaka, Kazuhide Shibata, Kazuhiro Shimizu, Xu Gao, Yuko Kikuchi, Yoshiyuki Togashi, Tomonori Hosoya, Satoru Takahashi, Osamu Nakajima, and Masayuki Yamamoto. Indispensable function for embryogenesis, expression and regulation of the nonspecific form of the 5-aminolevulinate synthase gene in mouse. *Genes to cells : devoted to molecular & cellular mechanisms*, 15:77–89, January 2010.

[277] Keisuke Okita, Tomoko Ichisaka, and Shinya Yamanaka. Generation of germline-competent induced pluripotent stem cells. *Nature*, 448:313–317, July 2007.

[278] B Orelle, V Keim, L Masciotra, J C Dagorn, and J L Iovanna. Human pancreatitis-associated protein. messenger rna cloning and expression in pancreatic diseases. *The Journal of clinical investigation*, 90:2284–2291, December 1992.

[279] Stuart H Orkin and Leonard I Zon. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*, 132:631–644, February 2008.

[280] Li Ou, Michael J Przybilla, and Chester B Whitley. Proteomic analysis of mucopolysaccharidosis i mouse brain with two-dimensional polyacrylamide gel electrophoresis. *Molecular genetics and metabolism*, 120:101–110, 2017.

[281] K Ozaki, T Kuroki, S Hayashi, and Y Nakamura. Isolation of three testis-specific genes (tsa303, tsa806, tsa903) by a differential mrna display method. *Genomics*, 36:316–319, September 1996.

[282] Falcon S Pagès H, Carlson M and Li N. Annotationdbi: Annotation database interface. r package version 1.38.2. 2017.

[283] Zhi-qiang Pan, Zhao-qin Fang, and Wen-li Lu. [characteristics of gene expression of adrenal cortical steroid synthetase and its regulatory factor in mice with h22 liver cancer of different patterns]. *Zhongguo Zhong xi yi jie he za zhi Zhongguo Zhongxiyi jiehe zazhi = Chinese journal of integrated traditional and Western medicine*, 31:85–89, January 2011.

[284] Suchismita Panda, Meenakshi Setia, Navjot Kaur, Varsha Shepal, Vivek Arora, Divya Kumari Singh, Abir Mondal, Abhishek Teli, Madhura Tathode, Rajendra Gajula, L C Padhy, and Anjali Shiras. Noncoding rna ginir functions as an oncogene by associating with centrosomal proteins. *PLoS biology*, 16:e2004204, October 2018.

[285] Sunil K Panigrahi, Ana Vasileva, and Debra J Wolgemuth. Sp1 transcription factor and gata1 cis-acting elements modulate testis-specific expression of mouse cyclin a1. *PloS one*, 7:e47862, 2012.

[286] Fresia Pareja, Alissa H Brandes, Thais Basili, Pier Selenica, Felipe C Geyer, Dan Fan, Arnaud Da Cruz Paula, Rahul Kumar, David N Brown, Rodrigo Gularte-Mérida, Barbara Alemar, Rui Bi, Raymond S Lim, Ino de Bruijn, Sho Fujisawa, Rui Gardner, Elvin Feng, Anqi Li, Edaise M da Silva, John R Lozada, Pedro Blecua, Leona Cohen-Gould, Achim A Jungbluth, Emad A Rakha, Ian O Ellis, Maria I A Edelweiss, Juan Palazzo, Larry Norton, Travis Hollmann, Marcia Edelweiss, Brian P Rubin, Britta Weigelt, and Jorge S Reis-Filho. Loss-of-function mutations in atp6ap1 and atp6ap2 in granular cell tumors. *Nature communications*, 9:3533, August 2018.

[287] Eunkuk Park, Jaehoon Jung, Osamu Araki, Katsuhiko Tsunekawa, So Young Park, Jeonghyun Kim, Masami Murakami, Seon-Yong Jeong, and Sihoon Lee. Concurrent tshr mutations and dio2 t92a polymorphism result in abnormal thyroid hormone metabolism. *Scientific reports*, 8:10090, July 2018.

[288] H Parker, M J J Rose-Zerilli, A Parker, T Chaplin, R Wade, A Gardiner, M Griffiths, A Collins, B D Young, D G Oscier, and J C Strefford. 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. *Leukemia*, 25:489–497, March 2011.

[289] Raymond C Pasek, Erik Malarkey, Nicolas F Berbari, Neeraj Sharma, Robert A Kesterson, Laura L Tres, Abraham L Kierszenbaum, and Bradley K Yoder. Coiled-coil domain containing 42 (ccdc42) is necessary for proper sperm development and male fertility in the mouse. *Developmental biology*, 412:208–218, April 2016.

[290] Kashyap A Patel, Bridget Knight, Aftab Aziz, Tarig Babiker, Avades Tamar, Joanna Findlay, Sue Cox, Ioannis Dimitropoulos, Carolyn Tysoe, Vijay Panicker, and Bijay Vaidya. Utility of systematic tshr gene testing in adults with hyperthyroidism lacking overt autoimmunity and diffuse uptake on thyroid scintigraphy. *Clinical endocrinology*, October 2018.

[291] Catriona Paul, Serena Teng, and Philippa T K Saunders. A single, mild, transient scrotal heat stress causes hypoxia and oxidative stress in mouse testes, which induces germ cell death. *Biology of reproduction*, 80:913–919, May 2009.

[292] Florencia Pauli, Yueyi Liu, Yoona A Kim, Pei-Jiun Chen, and Stuart K Kim. Chromosomal clustering and gata transcriptional regulation of intestine-expressed genes in c. elegans. *Development (Cambridge, England)*, 133:287–295, January 2006.

[293] Zhengtong Pei, Zhenzhen Jia, and Paul A Watkins. The second member of the human and murine bubblegum family is a testis- and brainstem-specific acyl-coa synthetase. *The Journal of biological chemistry*, 281:6632–6641, March 2006.

[294] Pärt Peterson. Sirt-ainly aire. *Nat Immunol*, 16(7):680–681, Jul 2015.

[295] Konstantin A Petrov, Evgeny E Nikolsky, and Patrick Masson. Autoregulation of acetylcholine release and micro-pharmacodynamic mechanisms at neuromuscular junction: Selective acetylcholinesterase inhibitors for therapy of myasthenic syndromes. *Frontiers in pharmacology*, 9:766, 2018.

[296] Iva Petrovchich and James M Ford. Genetic predisposition to gastric cancer. *Seminars in oncology*, 43:554–559, October 2016.

[297] Sheena Pinto, Chloé Michel, Hannah Schmidt-Glenewinkel, Nathalie Harder, Karl Rohr, Stefan Wild, Benedikt Brors, and Bruno Kyewski. Overlapping gene coexpression patterns in human medullary thymic epithelial cells generate self-antigen diversity. *Proc Natl Acad Sci U S A*, 110(37):E3497–E3505, Sep 2013.

[298] Sheena Pinto, Daniel Sommermeyer, Chloé Michel, Susanne Wilde, Dolores Schendel, Wolfgang Uckert, Thomas Blankenstein, and Bruno Kyewski. Mis-initiation of intrathymic mart-1 transcription and biased tcr usage explain the high frequency of mart-1-specific t cells. *Eur J Immunol*, May 2014.

[299] J Pitkänen, A Rebane, J Rowell, A Murumägi, P Ströbel, K Möll, M Saare, J Heikkilä, V Doucas, A Marx, and P Peterson. Cooperative activation of transcription by autoimmune regulator aire and cbp. *Biochemical and biophysical research communications*, 333:944–953, August 2005.

[300] Tereza Planck, Bushra Shahida, Marketa Sjögren, Leif Groop, Bengt Hallengren, and Mikael Lantz. Association of btg2, cyr61, zfp36, and scd gene polymorphisms with graves' disease and ophthalmopathy. *Thyroid : official journal of the American Thyroid Association*, 24:1156–1161, July 2014.

[301] Mirele Daiana Poleti, Cristina Tschorny Moncau, Bárbara Silva-Vignato, Alessandra Fernandes Rosa, Adalfredo Rocha Lobo, Thaís Regiani Cataldi, João Alberto Negrão, Saulo L Silva, Joanir Pereira Eler, and Júlio César de Carvalho Balieiro. Label-free quantitative proteomic analysis reveals muscle contraction and metabolism proteins linked to ultimate ph in bovine skeletal muscle. *Meat science*, 145:209–219, November 2018.

[302] Anello Marcello Poma, Riccardo Giannini, Paolo Piaggi, Clara Ugolini, Gabriele Materazzi, Paolo Miccoli, Paolo Vitti, and Fulvio Basolo. A six-gene panel to label follicular adenoma, low- and high-risk follicular thyroid carcinoma. *Endocrine connections*, 7:124–132, January 2018.

[303] Lőrinc Pongor, Máté Kormos, Christos Hatzis, Lajos Pusztai, András Szabó, and Balázs Győrffy. A genome-wide approach to link genotype to clinical outcome by utilizing next generation sequencing and gene chip data of 6,697 breast cancer patients. *Genome medicine*, 7:104, October 2015.

[304] Eleonora Porcu, Marco Medici, Giorgio Pistis, Claudia B Volpato, Scott G Wilson, Anne R Cappola, Steffan D Bos, Joris Deelen, Martin den Heijer, Rachel M Freathy, Jari Lahti, Chunyu Liu, Lorna M Lopez, Ilja M Nolte, Jeffrey R O'Connell, Toshiko Tanaka, Stella Trompet, Alice Arnold, Stefania Bandinelli, Marian Beekman, Stefan Böhringer, Suzanne J Brown, Brendan M Buckley, Clara Camaschella, Anton J M de Craen, Gail Davies, Marieke C H de Visser, Ian Ford, Tom Forsen, Timothy M Frayling, Laura Fugazzola, Martin Gögele, Andrew T Hattersley, Ad R Hermus, Albert Hofman, Jeanine J Houwing-Duistermaat, Richard A Jensen, Eero Kajantie, Margreet Kloppenburg, Ee M Lim, Corrado Masciullo, Stefano Mariotti, Cosetta Minelli, Braxton D Mitchell, Ramaiah Nagaraja, Romana T Netea-Maier, Aarno Palotie, Luca Persani, Maria G Piras, Bruce M Psaty, Katri Räikkönen, J Brent Richards, Fernando Rivadeneira, Cinzia Sala, Mona M Sabra, Naveed Sattar, Beverley M Shields, Nicole Soranzo, John M Starr, David J Stott, Fred C G J Sweep, Gianluca Usala, Melanie M van der Klauw, Diana van Heemst, Alies van Mullem, Sita H Vermeulen, W Edward Visser, John P Walsh, Rudi G J Westendorp, Elisabeth Widen, Guangju Zhai, Francesco Cucca, Ian J Deary, Johan G Eriksson, Luigi Ferrucci, Caroline S Fox, J Wouter Jukema, Lambertus A Kiemeney, Peter P Pramstaller, David Schlessinger, Alan R Shuldiner, Eline P Slagboom, André G Uitterlinden, Bijay Vaidya, Theo J Visser, Bruce H R Wolffenbuttel, Ingrid Meulenbelt, Jerome I Rotter, Tim D Spector, Andrew A Hicks, Daniela Toniolo, Serena Sanna, Robin P Peeters, and Silvia Naitza. A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. *PLoS genetics*, 9:e1003266, 2013.

[305] Mark Preciados, Changwon Yoo, and Deodutta Roy. Estrogenic endocrine disrupting chemicals influencing nrf1 regulated gene networks in the development of complex human brain diseases. *International journal of molecular sciences*, 17, December 2016.

[306] E Pure, C J Durie, C K Summerill, and J C Unkeless. Identification of soluble fc receptors in mouse serum and the conditioned medium of stimulated b cells. *The Journal of experimental medicine*, 160:1836–1849, December 1984.

[307] Andrea Párniczky, Eszter Hegyi, Anna Zsófia Tóth, Ákos Szücs, Andrea Szentesi, Áron Vincze, Ferenc Izbéki, Balázs Csaba Németh, Péter Hegyi, and Miklós Sahin-Tóth. Genetic analysis of human chymotrypsin-like elastases 3a and 3b (cela3a and cela3b) to assess the role of complex formation between proelastases and procarboxypepti-

dases in chronic pancreatitis. *International journal of molecular sciences*, 17:2148, December 2016.

[308] Gentleman R and Biocore. geneplotter: Graphics related functions for bioconductor. r package version 1.54.0. 2017.

[309] Ranjith Ramasamy, Alex Ridgeway, Larry I Lipshultz, and Dolores J Lamb. Integrative dna methylation and gene expression analysis identifies discoidin domain receptor 1 association with idiopathic nonobstructive azoospermia. *Fertility and sterility*, 102:968–973.e3, October 2014.

[310] Sajid Rashid, Rolf Breckle, Mike Hupe, Stephanie Geisler, Nadine Doerwald, and Juergen Neesen. The murine dnali1 gene encodes a flagellar protein that interacts with the cytoplasmic dynein heavy chain 1. *Molecular reproduction and development*, 73:784–794, June 2006.

[311] Kristin Rattay, Hannah Verena Meyer, Carl Herrmann, Benedikt Brors, and Bruno Kyewski. Evolutionary conserved gene co-expression drives generation of self-antigen diversity in medullary thymic epithelial cells. *Journal of autoimmunity*, 67:65–75, February 2016.

[312] M V Prasad Linga Reddy, H Wang, S Liu, B Bode, J C Reed, R D Steed, S W Anderson, L Steed, D Hopkins, and J-X She. Association between type 1 diabetes and gwas snps in the southeast us caucasian population. *Genes and immunity*, 12:208–212, April 2011.

[313] Valerie Reinke. Sex and the genome. *Nature genetics*, 36:548–549, June 2004.

[314] Sarah J Richardson, Noel G Morgan, and Alan K Foulis. Pancreatic pathology in type 1 diabetes mellitus. *Endocrine pathology*, 25:80–92, March 2014.

[315] Michael A Rieger and Timm Schroeder. Hematopoiesis. *Cold Spring Harbor perspectives in biology*, 4, December 2012.

[316] John Ian Risinger, Gadisetti V R Chandramouli, G Larry Maxwell, Mary Custer, Svetlana Pack, Dmitri Loukinov, Olga Aprelikova, Tracy Litzi, David S Schrump, Susan K Murphy, Andrew Berchuck, Victor Lobanenkov, and J Carl Barrett. Global expression analysis of cancer/testis genes in uterine cancers reveals a high incidence of boris expression. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 13:1713–1719, March 2007.

[317] Matthew E Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. limma powers differential ex-

pression analyses for rna-sequencing and microarray studies. *Nucleic acids research*, 43:e47, April 2015.

[318] Lauren L Ritterhouse, Lori J Wirth, Gregory W Randolph, Peter M Sadow, Douglas S Ross, Whitney Liddy, and Jochen K Lennerz. Ros1 rearrangement in thyroid cancer. *Thyroid : official journal of the American Thyroid Association*, 26:794–797, June 2016.

[319] Michael D Roberts, Matthew A Romero, Christopher B Mobley, Petey W Mumford, Paul A Roberson, Cody T Haun, Christopher G Vann, Shelby C Osburn, Hudson H Holmes, Rory A Greer, Christopher M Lockwood, Hailey A Parry, and Andreas N Kavazis. Skeletal muscle mitochondrial volume and myozenin-1 protein differences exist between high versus low anabolic responders to resistance training. *PeerJ*, 6:e5338, 2018.

[320] João Paulo Vilela Rodrigues, Fabiana Angelo Marques, Ana Maria Rosa Freato Gonçalves, Marília Silveira de Almeida Campos, Tiago Marques Dos Reis, Manuela Roque Siani Morelo, Andrea Fontoura, Beatriz Maria Pereira Girolineto, Helen Palmira Miranda de Camargo Souza, Maurílio de Souza Cazarim, Lauro César da Silva Maduro, and Leonardo Régis Leira Pereira. Analysis of clinical pharmacist interventions in the neurology unit of a brazilian tertiary teaching hospital. *PloS one*, 14:e0210779, 2019.

[321] Kryssia Isabel Rodriguez-Castro, Marilisa Franceschi, Chiara Miraglia, Michele Russo, Antonio Nouvenne, Gioacchino Leandro, Tiziana Meschi, Gian Luigi De' Angelis, and Francesco Di Mario. Autoimmune diseases in autoimmune atrophic gastritis. *Acta bio-medica : Atenei Parmensis*, 89:100–103, December 2018.

[322] Christian Rogon, Anna Ulbricht, Michael Hesse, Simon Alberti, Preethi Vijayaraj, Diana Best, Ian R Adams, Thomas M Magin, Bernd K Fleischmann, and Jörg Höhfeld. Hsp70-binding protein hspbp1 regulates chaperone expression at a posttranslational level and is essential for spermatogenesis. *Molecular biology of the cell*, 25:2260–2271, August 2014.

[323] Mary F Rooney, Richard K Porter, Lisa M Katz, and Emmeline W Hill. Skeletal muscle mitochondrial bioenergetics and associations with myostatin genotypes in the thoroughbred horse. *PloS one*, 12:e0186247, 2017.

[324] Judith A Rosales-Aviña, Jorge Torres-Flores, Adriana Aguilar-Lemarroy, Carmen Gurrola-Díaz, Georgina Hernández-Flores, Pablo C Ortiz-Lazareno, José M Lerma-Díaz, Ruth de Celis, Óscar González-Ramella, Esperanza Barrera-Chaires, Alejandro Bravo-Cuellar, and

Luis F Jave-Suárez. Meis1, prep1, and pbx4 are differentially expressed in acute lymphoblastic leukemia: association of meis1 expression with higher proliferation and chemotherapy resistance. *Journal of experimental & clinical cancer research : CR*, 30:112, December 2011.

[325] S A Rosenberg. A new era for cancer immunotherapy based on the genes that encode cancer antigens. *Immunity*, 10:281–287, March 1999.

[326] Fabio M V Rossi, Stephane Y Corbel, Jasmeen S Merzaban, Douglas A Carlow, Klaus Gossens, Jeffrey Duenas, Leslie So, Lin Yi, and Hermann J Ziltener. Recruitment of adult thymic progenitors is regulated by p-selectin and its ligand psgl-1. *Nature immunology*, 6:626–634, June 2005.

[327] Richard B Roth, Peter Hevezi, Jerry Lee, Dorian Willhite, Sandra M Lechner, Alan C Foster, and Albert Zlotnik. Gene expression analyses reveal molecular relationships among 20 regions of the human cns. *Neurogenetics*, 7(2):67–80, May 2006.

[328] Peter J Roy, Joshua M Stuart, Jim Lund, and Stuart K Kim. Chromosomal clustering of muscle-expressed genes in caenorhabditis elegans. *Nature*, 418(6901):975–979, Aug 2002.

[329] Chong-Mei Ruan, Jing Wang, Yong-Xin Yang, Jun-Jie Hu, You-Ji Ma, Yong Zhang, and Xing-Xu Zhao. Proteomic analysis of tianzhu white yak (bos grunniens) testis at different sexual developmental stages. *Animal science journal = Nihon chikusan Gakkaiho*, January 2019.

[330] Giuditta Rurale, Luca Persani, and Federica Marelli. javax.xml.bind.jaxbelement@3620387d, and thyroid: A pleiotropic candidate gene for congenital hypothyroidism. *Frontiers in endocrinology*, 9:730, 2018.

[331] U Sahin, O Türeci, H Schmitt, B Cochlovius, T Johannes, R Schmits, F Stenner, G Luo, I Schobert, and M Pfreundschuh. Human neoplasms elicit multiple specific immune responses in the autologous host. *Proceedings of the National Academy of Sciences of the United States of America*, 92:11810–11813, December 1995.

[332] A Sakula. Robert koch (1843-1910): Founder of the science of bacteriology and discoverer of the tubercle bacillus : A study of his life and work. *The Canadian veterinary journal = La revue veterinaire canadienne*, 24:124–127, April 1983.

[333] A M Salmon, C Bruand, A Cardona, J P Changeux, and S Berrih-Aknin. An acetylcholine receptor alpha subunit promoter confers intrathymic expression in transgenic mice. implications for tolerance of

a transgenic self-antigen and for autoreactivity in myasthenia gravis. *The Journal of clinical investigation*, 101:2340–2350, June 1998.

[334] Emmanouil Saloustros, Paraskevi Salpea, Chen-Feng Qi, Lina A Gugliotti, Kitman Tsang, Sisi Liu, Matthew F Starost, Herbert C Morse, and Constantine A Stratakis. Hematopoietic neoplasms in prkar2a-deficient mice. *Journal of experimental & clinical cancer research : CR*, 34:143, November 2015.

[335] Julia Salzman, Robert J Marinelli, Peter L Wang, Ann E Green, Julie S Nielsen, Brad H Nelson, Charles W Drescher, and Patrick O Brown. Esrra-c11orf20 is a recurrent gene fusion in serous ovarian carcinoma. *PLoS biology*, 9:e1001156, September 2011.

[336] Luna Samanta, Ashok Agarwal, Nirlipta Swain, Rakesh Sharma, Banu Gopalan, Sandro C Esteves, Damayanthi Durairajanayagam, and Edmund Sabanegh. Proteomic signatures of sperm mitochondria in varicocele: Clinical use as biomarkers of varicocele associated infertility. *The Journal of urology*, 200:414–422, August 2018.

[337] Rickard Sandberg and Ola Larsson. Improved precision and accuracy for microarrays using updated probe set definitions. *BMC bioinformatics*, 8:48, February 2007.

[338] Nobuya Sasaki, Yayoi Hosoda, Aogu Nagata, Ming Ding, Ji-Ming Cheng, Tomomi Miyamoto, Shinya Okano, Atsushi Asano, Ichiro Miyoshi, and Takashi Agui. A mutation in tpst2 encoding tyrosylprotein sulfotransferase causes dwarfism associated with hypothyroidism. *Molecular endocrinology (Baltimore, Md.)*, 21:1713–1721, July 2007.

[339] Matthew J Scanlan, Claudia M Gordon, Barbara Williamson, Sang-Yull Lee, Yao-Tseng Chen, Elisabeth Stockert, Achim Jungbluth, Gerd Ritter, Dirk Jäger, Elke Jäger, Alexander Knuth, and Lloyd J Old. Identification of cancer/testis genes by database mining and mrna expression analysis. *International journal of cancer*, 98:485–492, April 2002.

[340] Susan M Schlenner and Hans-Reimer Rodewald. Early t cell development and the pitfalls of potential. *Trends in immunology*, 31:303–310, August 2010.

[341] Benedikt Schoser, Saiju Jacob, David Hilton-Jones, Wolfgang Müller-Felber, Christian Kubisch, Detlef Claus, Hans H Goebel, Giuseppe Vita, Angela Vincent, Antonio Toscano, and Peter Van den Bergh. Immune-mediated rippling muscle disease with myasthenia gravis: a report of seven patients with long-term follow-up in two. *Neuromuscular disorders : NMD*, 19:223–228, March 2009.

[342] Hans-Juergen Schulten, Zuhoor Al-Mansouri, Ibtisam Baghallab, Nadia Bagatian, Ohoud Subhi, Sajjad Karim, Hosam Al-Aradati, Abdulmonem Al-Mutawa, Adel Johary, Abdulrahman A Meccawy, Khalid Al-Ghamdi, Osman Al-Hamour, Mohammad Hussain Al-Qahtani, and Jaudah Al-Maghrabi. Comparison of microarray expression profiles between follicular variant of papillary thyroid carcinomas and follicular adenomas of the thyroid. *BMC genomics*, 16 Suppl 1:S7, 2015.

[343] Sandra R Schulze and Lori L Wallrath. Gene regulation by chromatin structure: paradigms established in drosophila melanogaster. *Annual review of entomology*, 52:171–192, 2007.

[344] Nils Schweingruber, Henrike J Fischer, Lisa Fischer, Jens van den Brandt, Anna Karabinskaya, Verena Labi, Andreas Villunger, Benedikt Kretzschmar, Peter Huppke, Mikael Simons, Jan P Tuckermann, Alexander Flügel, Fred Lühder, and Holger M Reichardt. Chemokine-mediated redirection of t cells constitutes a critical mechanism of glucocorticoid therapy in autoimmune cns responses. *Acta neuropathologica*, 127:713–729, May 2014.

[345] E W Scott, M C Simon, J Anastasi, and H Singh. Requirement of transcription factor pu.1 in the development of multiple hematopoietic lineages. *Science (New York, N.Y.)*, 265:1573–1577, September 1994.

[346] D Selcen, V C Juel, L D Hobson-Webb, E C Smith, D E Stickler, A V Bite, K Ohno, and A G Engel. Myasthenic syndrome caused by plectinopathy. *Neurology*, 76:327–336, January 2011.

[347] Bushra Shahida, Tereza Planck, Peter Åsman, and Mikael Lantz. Study of deiodinase type 2 polymorphisms in graves' disease and ophthalmopathy in a swedish population. *European thyroid journal*, 7:289–293, November 2018.

[348] Xinwei She, Carol A Rohl, John C Castle, Amit V Kulkarni, Jason M Johnson, and Ronghua Chen. Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. *BMC genomics*, 10:269, June 2009.

[349] Celeste A Shelton and David C Whitcomb. Genetics and treatment options for recurrent acute and chronic pancreatitis. *Current treatment options in gastroenterology*, 12:359–371, September 2014.

[350] Shulin Shen, Jinzi Wang, Jihong Liang, and Chunhui Zhu. Low-expressed testis-specific calcium-binding protein cbp86-iv (cabyr) is observed in idiopathic asthenozoospermia. *World journal of urology*, 33:633–638, May 2015.

[351] Yong-Hyun Shin, Youngsok Choi, Serpil Uckac Erdin, Svetlana A Yatsenko, Malgorzata Kloc, Fang Yang, P Jeremy Wang, Marvin L Meistrich, and Aleksandar Rajkovic. Hormad1 mutation disrupts synaptonemal complex formation, recombination, and chromosome segregation in mammalian meiosis. *PLoS genetics*, 6:e1001190, November 2010.

[352] Eiichi Shoguchi, Manabu Fujie, and Mayuko Hamada. No chromosomal clustering of housekeeping genes in the marine chordate ciona intestinalis. *Marine genomics*, 4:151–157, September 2011.

[353] Rong-Yaun Shyu, Chang-Chieh Wu, Chun-Hua Wang, Tzung-Chieh Tsai, Lu-Kai Wang, Mao-Liang Chen, Shun-Yuan Jiang, and Fu-Ming Tsai. H-rev107 regulates prostaglandin d2 synthase-mediated suppression of cellular invasion in testicular cancer cells. *Journal of biomedical science*, 20:30, May 2013.

[354] Matt Silver, Eva Janousova, Xue Hua, Paul M Thompson, Giovanni Montana, and Alzheimer's Disease Neuroimaging Initiative. Identification of gene pathways implicated in alzheimer's disease using longitudinal imaging phenotypes with sparse regression. *NeuroImage*, 63:1681–1694, November 2012.

[355] Matthew J Simmonds. Gwas in autoimmune thyroid disease: redefining our understanding of pathogenesis. *Nature reviews. Endocrinology*, 9:277–287, May 2013.

[356] Andrew J G Simpson, Otavia L Caballero, Achim Jungbluth, Yao-Tseng Chen, and Lloyd J Old. Cancer/testis antigens, gametogenesis and cancer. *Nature reviews. Cancer*, 5:615–625, August 2005.

[357] Rebecca Sinnott, Leah Winters, Brittany Larson, Daniela Mytsa, Patrick Taus, Kathryn M Cappell, and Angelique W Whitehurst. Mechanisms promoting escape from mitotic stress-induced tumor cell death. *Cancer research*, 74:3857–3869, July 2014.

[358] Sherona R Sirkisoon, Richard L Carpenter, Tadas Rimkus, Ashley Anderson, Alexandria Harrison, Allison M Lange, Guangxu Jin, Kounosuke Watabe, and Hui-Wen Lo. Interaction between stat3 and gli1/tgli1 oncogenic transcription factors promotes the aggressiveness of triple-negative breast cancers and her2-enriched breast cancer. *Oncogene*, 37:2502–2514, May 2018.

[359] E G Skurikhin, A V Pakhomova, O V Pershina, V A Krupin, N N Ermakova, E S Pan, A I Kudryashova, L A Ermolaeva, E S Khmelevskaya, V E Goldberg, V V Zhdanov, and A M Dygai. Role of sertoli and leydig cells in the regulation of spermatogonial stem

cell and development of reproductive disorders in male c57bl/6 mice with type 1 diabetes mellitus. *Bulletin of experimental biology and medicine*, 164:127–131, December 2017.

[360] Natalya A Smirnova, Peter J Romanienko, Pavel P Khil, and R Daniel Camerini-Otero. Gene expression profiles of spo11-/- mouse testes with spermatocytes arrested in meiotic prophase i. *Reproduction (Cambridge, England)*, 132:67–77, July 2006.

[361] Tatiana A Soboleva, Brian J Parker, Maxim Nekrasov, Gene Hart-Smith, Ying Jin Tay, Wei-Quan Tng, Marc Wilkins, Daniel Ryan, and David J Tremethick. A new link between transcriptional initiation and pre-mrna splicing: The rna binding histone variant h2a.b. *PLoS genetics*, 13:e1006633, February 2017.

[362] Neal Sondheimer, Stacy Hewson, Jessie M Cameron, Gino R Somers, Jane Dunning Broadbent, Marcello Ziosi, Catarina Maria Quinzii, and Ali B Naini. Novel recessive mutations in coq4 cause severe infantile cardiomyopathy and encephalopathy associated with coq, javax.xml.bind.jaxbelement@58cf6dd3, deficiency. *Molecular genetics and metabolism reports*, 12:23–27, September 2017.

[363] Rong-Hua Song, Qian Li, Xi Jia, Qiu-Ming Yao, Bin Wang, and Jin-An Zhang. Polymorphisms of fam167a-blk region confer risk of autoimmune thyroid disease. *DNA and cell biology*, 37:932–940, November 2018.

[364] Sarah F C Soper, Godfried W van der Heijden, Tara C Hardiman, Mary Goodheart, Sandra L Martin, Peter de Boer, and Alex Bortvin. Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Developmental cell*, 15:285–297, August 2008.

[365] Elin Pettersen Sorgjerd. Type 1 diabetes-related autoantibodies in different forms of diabetes. *Current diabetes reviews*, July 2018.

[366] M Sospedra, X Ferrer-Francesch, O Domínguez, M Juan, M Foz-Sala, and R Pujol-Borrell. Transcription of a broad range of self-antigens in human thymus suggests a role for central mechanisms in tolerance toward peripheral antigens. *Journal of immunology (Baltimore, Md. : 1950)*, 161:5918–5929, December 1998.

[367] M I Stefanou, L Komorowski, S Kade, A Bornemann, U Ziemann, and M Synofzik. A case of late-onset, thymoma-associated myasthenia gravis with ryanodine receptor and titin antibodies and concomitant granulomatous myositis. *BMC neurology*, 16:172, September 2016.

[368] L Stein, P Sternberg, R Durbin, J Thierry-Mieg, and J Spieth. Wormbase: network access to the genome and biology of caenorhabditis elegans. *Nucleic acids research*, 29:82–86, January 2001.

[369] Leighton Stein, Jenniffer Rothschild, Jesse Luce, John K Cowell, Geraldine Thomas, Tatania I Bogdanova, Mycola D Tronko, and Lesleyann Hawthorn. Copy number and gene expression alterations in radiation-induced papillary thyroid carcinoma from chernobyl pediatric patients. *Thyroid : official journal of the American Thyroid Association*, 20:475–487, May 2010.

[370] Janusz Strzelczyk, Magdalena Szumska, Aleksandra Damasiewicz-Bodzek, Anna Krywult, Michał Długaszek, Justyna Czubilińska, Kaja Gawlik, Konrad Synowiec, Krystyna Tyrpień-Golder, Karolina Poczkaj, and Beata Kos-Kudła. Determination of advanced glycation end-products and antibodies against them (anti-cml and anti-cel) in the serum of graves' orbitopathy patients before and after methylprednisolone treatment. *Endokrynologia Polska*, 67:383–389, 2016.

[371] Andrew I Su, Michael P Cooke, Keith A Ching, Yaron Hakak, John R Walker, Tim Wiltshire, Anthony P Orth, Raquel G Vega, Lisa M Sapinoso, Aziz Moqrich, Ardem Patapoutian, Garret M Hampton, Peter G Schultz, and John B Hogenesch. Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci U S A*, 99(7):4465–4470, Apr 2002.

[372] Andrew I Su, Tim Wiltshire, Serge Batalov, Hilmar Lapp, Keith A Ching, David Block, Jie Zhang, Richard Soden, Mimi Hayakawa, Gabriel Kreiman, Michael P Cooke, John R Walker, and John B Hogenesch. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A*, 101(16):6062–6067, Apr 2004.

[373] Ayana Suzuki, Mitsuyoshi Hirokawa, Nami Takada, Miyoko Higuchi, Aki Tanaka, Toshitetsu Hayashi, Seiji Kuma, and Akira Miyauchi. Utility of monoclonal pax8 antibody for distinguishing intrathyroid thymic carcinoma from follicular cell-derived thyroid carcinoma. *Endocrine journal*, September 2018.

[374] Damian Szklarczyk, John H Morris, Helen Cook, Michael Kuhn, Stefan Wyder, Milan Simonovic, Alberto Santos, Nadezhda T Doncheva, Alexander Roth, Peer Bork, Lars J Jensen, and Christian von Mering. The string database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research*, 45:D362–D368, January 2017.

[375] Hiroaki Tachiwana, Akihisa Osakabe, Hiroshi Kimura, and Hitoshi Kurumizaka. Nucleosome formation with the testis-specific histone h3

variant, h3t, by human nucleosome assembly proteins in vitro. *Nucleic acids research*, 36:2208–2218, April 2008.

[376] Afsoon Taghavi, Mohammad Esmaeil Akbari, Mohammad Hashemi-Bahremani, Nahid Nafissi, Ahad Khalilnezhad, Seyed Mohammad Poorhosseini, Feyzollah Hashemi-Gorji, and Vahid Reza Yassaee. Gene expression profiling of the 8q22-24 position in human breast cancer: , javax.xml.bind.jaxbelement@38de3d7f, , , javax.xml.bind.jaxbelement@7d63fde5, , , javax.xml.bind.jaxbelement@496301b, and , javax.xml.bind.jaxbelement@7860c627, genes are implicated in oncogenesis, while , javax.xml.bind.jaxbelement@319c2991, and , javax.xml.bind.jaxbelement@53727692, genes may predict a risk of metastasis. *Oncology letters*, 12:3845–3855, November 2016.

[377] Yousuke Takahama. Journey through the thymus: stromal guides for t-cell development and selection. *Nature reviews. Immunology*, 6:127–135, February 2006.

[378] Maciej Tarnowski, Michał Czerewaty, Anna Deskur, Krzysztof Safranow, Wojciech Marlicz, Elżbieta Urasińska, Mariusz Z Ratajczak, and Teresa Starzyńska. Expression of cancer testis antigens in colorectal cancer: New prognostic and therapeutic implications. *Disease markers*, 2016:1987505, 2016.

[379] Manuela Tavian, Katia Biasch, Lidia Sinka, Judith Vallet, and Bruno Péault. Embryonic origin of human hematopoiesis. *The International journal of developmental biology*, 54:1061–1065, 2010.

[380] Sofia Tedelind, Kseniia Poliakova, Amanda Valeta, Ruth Hunegnaw, Eyoel Lemma Yemanaberhan, Nils-Erik Heldin, Junichi Kurebayashi, Ekkehard Weber, Nataša Kopitar-Jerala, Boris Turk, Matthew Bogyo, and Klaudia Brix. Nuclear cysteine cathepsin variants in thyroid carcinoma cells. *Biological chemistry*, 391:923–935, August 2010.

[381] Sofia Thomaidou, Arnaud Zaldumbide, and Bart O Roep. Islet stress, degradation and autoimmunity. *Diabetes, obesity & metabolism*, 20 Suppl 2:88–94, September 2018.

[382] Raffi Topakian, Fritz Zimprich, Stephan Iglseder, Norbert Embacher, Michael Guger, Karl Stieglbauer, Dieter Langenscheidt, Jakob Rath, Stefan Quasthoff, Philipp Simschitz, Julia Wanschitz, David Windisch, Petra Müller, Dierk Oel, Günther Schustereder, Stefan Einsiedler, Christian Eggers, and Wolfgang Löscher. High efficacy of rituximab for myasthenia gravis: a comprehensive nationwide study in austria. *Journal of neurology*, January 2019.

[383] D Torrents, J Mykkänen, M Pineda, L Feliubadaló, R Estévez, R de Cid, P Sanjurjo, A Zorzano, V Nunes, K Huoponen, A Reinikainen, O Simell, M L Savontaus, P Aula, and M Palacín. Identification of slc7a7, encoding y+lat-1, as the lysinuric protein intolerance gene. *Nature genetics*, 21:293–296, March 1999.

[384] Cecilie Totland, Torbjørn Kråkenes, Kibret Mazengia, Mette Haugen, and Christian Vedeler. Expression of the onconeural protein cdr1 in cerebellum and ovarian cancer. *Oncotarget*, 9:23975–23986, May 2018.

[385] Eric S Tucker, Samantha Segall, Deepak Gopalakrishna, Yongqin Wu, Mike Vernon, Franck Polleux, and Anthony-Samuel Lamantia. Molecular specification and patterning of progenitor cells in the lateral and medial ganglionic eminences. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28:9504–9518, September 2008.

[386] G Tufo, A W E Jones, Z Wang, J Hamelin, N Tajeddine, D D Esposti, C Martel, C Boursier, C Gallerne, C Migdal, C Lemaire, G Szabadkai, A Lemoine, G Kroemer, and C Brenner. The protein disulfide isomerases pdia4 and pdia6 mediate resistance to cisplatin-induced cell death in lung adenocarcinoma. *Cell death and differentiation*, 21:685–695, May 2014.

[387] Kenneth S K Tung, Jessica Harakal, Hui Qiao, Claudia Rival, Jonathan C H Li, Alberta G A Paul, Karen Wheeler, Patcharin Pramoonjago, Constance M Grafer, Wei Sun, Robert D Sampson, Elissa W P Wong, Prabhakara P Reddi, Umesh S Deshmukh, Daniel M Hardy, Huanghui Tang, C Yan Cheng, and Erwin Goldberg. Egress of sperm autoantigen from seminiferous tubules maintains systemic tolerance. *The Journal of clinical investigation*, 127:1046–1060, March 2017.

[388] James M A Turner. Meiotic sex chromosome inactivation. *Development (Cambridge, England)*, 134:1823–1831, May 2007.

[389] O Türeci, U Sahin, C Zwick, M Koslowski, G Seitz, and M Pfreundschuh. Identification of a meiosis-specific protein as a member of the class of cancer/testis antigens. *Proceedings of the National Academy of Sciences of the United States of America*, 95:5211–5216, April 1998.

[390] Mathias Uhlén, Linn Fagerberg, Björn M Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, IngMarie Olsson, Karolina Edlund, Emma Lundberg, Sanjay Navani, Cristina Al-Khalili Szigyarto, Jacob Odeberg, Dijana Djureinovic, Jenny Ottosson Takanen, Sophia

Hober, Tove Alm, Per-Henrik Edqvist, Holger Berling, Hanna Tegel, Jan Mulder, Johan Rockberg, Peter Nilsson, Jochen M Schwenk, Marica Hamsten, Kalle von Feilitzen, Mattias Forsberg, Lukas Persson, Fredric Johansson, Martin Zwahlen, Gunnar von Heijne, Jens Nielsen, and Fredrik Pontén. Proteomics. tissue-based map of the human proteome. *Science (New York, N.Y.)*, 347:1260419, January 2015.

[391] Sergey V Ulianov, Alexey A Gavrilov, and Sergey V Razin. Nuclear compartments, genome folding, and enhancer-promoter communication. *International review of cell and molecular biology*, 315:183–244, 2015.

[392] Cindy Ung, Angie V Sanchez, Lishuang Shen, Samaneh Davoudi, Tina Ahmadi, Daniel Navarro-Gomez, Ching J Chen, Heather Hancock, Alan Penman, Suzanne Hoadley, Mark Consugar, Carlos Restrepo, Vinay A Shah, Joseph F Arboleda-Velasquez, Lucia Sobrin, Xiaowu Gai, and Leo A Kim. Whole exome sequencing identification of novel candidate genes in patients with proliferative diabetic retinopathy. *Vision research*, 139:168–176, October 2017.

[393] Brian R Untch, Vanessa Dos Anjos, Maria E R Garcia-Rendueles, Jeffrey A Knauf, Gnana P Krishnamoorthy, Mahesh Saqcena, Umeshkumar K Bhanot, Nicholas D Socci, Alan L Ho, Ronald Ghossein, and James A Fagin. Tipifarnib inhibits hras-driven dedifferentiated thyroid cancers. *Cancer research*, 78:4642–4657, August 2018.

[394] Gerold Untergasser, Roland Gander, Claudia Lilg, Günter Lepperdinger, Eugen Plas, and Peter Berger. Profiling molecular targets of tgf-beta1 in prostate fibroblast-to-myofibroblast transdifferentiation. *Mechanisms of ageing and development*, 126:59–69, January 2005.

[395] Urbanek. R graphics device using cairo graphics library. 2015.

[396] P van der Bruggen, C Traversari, P Chomez, C Lurquin, E De Plaen, B Van den Eynde, A Knuth, and T Boon. A gene encoding an antigen recognized by cytolytic t lymphocytes on a human melanoma. *Science (New York, N.Y.)*, 254:1643–1647, December 1991.

[397] Joris Vanderlocht, Mart van der Cruys, Frans Stals, Liesbeth Bakker-Jonges, and Jan Damoiseaux. Multiplex autoantibody detection for autoimmune liver diseases and autoimmune gastritis. *Journal of immunological methods*, 448:21–25, September 2017.

[398] Guadalupe Vargas-Ortega, Lourdes Balcázar-Hernández, Baldomero González-Virla, Claudia Ramírez-Rentería, Oriana Nieto-Guzmán, Ana Pamela Garrido-Mendoza, Marco Antonio Flores-Maya, Moisés

Mercado, and Mendoza-Zubieta Victoria. Symptomatic primary hyperparathyroidism as a risk factor for differentiated thyroid cancer. *Journal of thyroid research*, 2018:9461079, 2018.

[399] Ana Vasileva, Daniela Tiedau, Adriana Firooznia, Thomas Müller-Reichert, and Rolf Jessberger. Tdrd6 is required for spermiogenesis, chromatoid body architecture, and regulation of mirna expression. *Current biology : CB*, 19:630–639, April 2009.

[400] George Vasmatzis, Xue Wang, James B Smadbeck, Stephen J Murphy, Katherine B Geiersbach, Sarah H Johnson, Athanasios G Gaitatzes, Yan W Asmann, Farhad Kosari, Mitesh J Borad, Daniel J Serie, Sarah A McLaughlin, Jennifer M Kachergus, Brian M Necela, and E Aubrey Thompson. Chromoanasynthesis is a common mechanism that leads to erbb2 amplifications in a cohort of early stage her2, javax.xml.bind.jaxbelement@6c5dbb9b, breast cancer samples. *BMC cancer*, 18:738, July 2018.

[401] V E Velculescu, S L Madden, L Zhang, A E Lash, J Yu, C Rago, A Lal, C J Wang, G A Beaudry, K M Ciriello, B P Cook, M R Dufault, A T Ferguson, Y Gao, T C He, H Hermeking, S K Hiraldo, P M Hwang, M A Lopez, H F Luderer, B Mathews, J M Petroziello, K Polyak, L Zawel, and K W Kinzler. Analysis of human transcriptomes. *Nature genetics*, 23:387–388, December 1999.

[402] Alexander E Vinogradov. Compactness of human housekeeping genes: selection for economy or genomic design? *Trends in genetics : TIG*, 20:248–253, May 2004.

[403] H Vira, V Pradhan, V Umare, A Chaudhary, A Rajadhyksha, M Nadkar, K Ghosh, and A Nadkarni. Role of mmp-7 in the pathogenesis of systemic lupus erythematosus (sle). *Lupus*, 26:937–943, August 2017.

[404] Lucian Visan, Ioana A Visan, Andreas Weishaupt, Harald H Hofstetter, Klaus V Toyka, Thomas Hünig, and Ralf Gold. Tolerance induction by intrathymic expression of p0. *Journal of immunology (Baltimore, Md. : 1950)*, 172:1364–1370, February 2004.

[405] D I Vodolazhsky, D S Kutilin, Kh A Mogushkova, and O I Kit. Specific features of transcription activity of cancer-testis antigens in patients with metastatic and non-metastatic breast cancer. *Bulletin of experimental biology and medicine*, July 2018.

[406] Anna-Lena Volckmar, Jonas Leichsenring, Christa Flechtenmacher, Nicole Pfarr, Udo Siebolts, Martina Kirchner, Jan Budczies, Michael Bockmayr, Kathrin Ridinger, Katja Lorenz, Esther Herpel, Aurelia Noske, Wilko Weichert, Frederick Klauschen, Peter Schirmacher,

Roland Penzel, Volker Endris, and Albrecht Stenzinger. Tubular, lactating, and ductal adenomas are devoid of med12 exon2 mutations, and ductal adenomas show recurrent mutations in gnas and the pi3k-akt pathway. *Genes, chromosomes & cancer*, 56:11–17, January 2017.

[407] Burçak Vural, Elçin Sehitoğlu, Filiz Cavuş, Nazlı Yalçınkaya, Hazal Haytural, Melike Küçükerden, Canan Ulusoy, Elif Uğurel, Selin Turan, Leyla Bulut, Recai Türkoğlu, Erkingül Shugaiv, Murat Kürtüncü, Sükrü Atakan, Ali O Güre, Ahmet Gül, Mefkure Eraksoy, Gülşen Akman-Demir, and Erdem Tüzün. Mitochondrial carrier homolog 1 (mtch1) antibodies in neuro-behçet's disease. *Journal of neuroimmunology*, 263:139–144, October 2013.

[408] Lucy S K Walker and Abul K Abbas. The enemy within: keeping self-reactive t cells at bay in the periphery. *Nature reviews. Immunology*, 2:11–19, January 2002.

[409] C Y Wang, J D Shi, A Davoodi-Semiromi, and J X She. Cloning of aire, the mouse homologue of the autoimmune regulator (aire) gene responsible for autoimmune polyglandular syndrome type 1 (asp1). *Genomics*, 55:322–326, February 1999.

[410] Fang Wang, Chang Liu, Xiuhua Jia, Xiangju Liu, Yinglei Xu, Shengli Yan, Xuewen Jia, Zuzhou Huang, Shiguo Liu, and Maosheng Gu. Next-generation sequencing of nkx2.1, foxe1, pax8, nkx2.5, and tshr in 100 chinese patients with congenital hypothyroidism and athyreosis. *Clinica chimica acta; international journal of clinical chemistry*, 470:36–41, July 2017.

[411] Hao Wang, Peng Zhang, and Chun-Ting Wang. [analysis of expression pattern of a novel testis-highly expressed gene biot2-l and the primary study on its role in testis development]. *Sichuan da xue xue bao. Yi xue ban = Journal of Sichuan University. Medical science edition*, 40:853–856, September 2009.

[412] P J Wang, J R McCarrey, F Yang, and D C Page. An abundance of x-linked genes expressed in spermatogonia. *Nature genetics*, 27:422–426, April 2001.

[413] Pinglang Wang, Fei Ding, Hsienyuan Chiang, Robert C Thompson, Stanley J Watson, and Fan Meng. Probematchdb–a web database for finding equivalent probes across microarray platforms and species. *Bioinformatics*, 18(3):488–489, Mar 2002.

[414] Renjie Wang, Qingjie Ma, Linlin Ji, Yue Yao, Mengshi Ma, and Qiang Wen. mir-622 suppresses tumor formation by directly targeting vegfa

in papillary thyroid carcinoma. *OncoTargets and therapy*, 11:1501–1509, 2018.

[415] J A Warrington, A Nair, M Mahadevappa, and M Tsyganskaya. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiological genomics*, 2:143–147, April 2000.

[416] Clive Wasserfall, Harry S Nick, Martha Campbell-Thompson, Dawn Beachy, Leena Haataja, Irina Kusmartseva, Amanda Posgai, Maria Beery, Christopher Rhodes, Ezio Bonifacio, Peter Arvan, and Mark Atkinson. Persistence of pancreatic insulin mrna expression and proinsulin protein in type 1 diabetes pancreata. *Cell metabolism*, 26:568–575.e3, September 2017.

[417] Emma A Webb, Meena Balasubramanian, Nadja Fratzl-Zelman, Wayne A Cabral, Hannah Titheradge, Atif Alsaedi, Vrinda Saraff, Julie Vogt, Trevor Cole, Susan Stewart, Nicola J Crabtree, Brandi M Sargent, Sonja Gamsjaeger, Eleftherios P Paschalis, Paul Roschger, Klaus Klaushofer, Nick J Shaw, Joan C Marini, and Wolfgang Högler. Phenotypic spectrum in osteogenesis imperfecta due to mutations in tmem38b: Unraveling a complex cellular defect. *The Journal of clinical endocrinology and metabolism*, 102:2019–2028, June 2017.

[418] Claudia C Weber and Laurence D Hurst. Support for multiple classes of local expression clusters in drosophila melanogaster, but no evidence for gene order conservation. *Genome biology*, 12:R23, 2011.

[419] Ulrich H Weidle, Vesna Evtimova, Saverio Alberti, Emanuela Guerra, Nikos Fersis, and Sepp Kaul. Cell growth stimulation by crash, an asparaginase-like protein overexpressed in human tumors and metastatic breast cancers. *Anticancer research*, 29:951–963, April 2009.

[420] Paul Weinberger, Sithara Raju Ponny, Hongyan Xu, Shan Bai, Robert Smallridge, John Copland, and Ashok Sharma. Cell cycle m-phase genes are highly upregulated in anaplastic thyroid carcinoma. *Thyroid : official journal of the American Thyroid Association*, 27:236–252, February 2017.

[421] Frank U Weiss, Mariya E Skube, and Markus M Lerch. Chronic pancreatitis: an update on genetic risk factors. *Current opinion in gastroenterology*, 34:322–329, September 2018.

[422] Hartmut Wekerle. B cells in multiple sclerosis. *Autoimmunity*, 50:57–60, February 2017.

[423] Xianjie Wen and Yisheng Yang. Emerging roles of glis3 in neonatal diabetes, type 1 and type 2 diabetes. *Journal of molecular endocrinology*, 58:R73–R85, February 2017.

[424] O Werdelin, U Cordes, and T Jensen. Aberrant expression of tissue-specific proteins in the thymus: a hypothesis for the development of central tolerance. *Scandinavian journal of immunology*, 47:95–100, February 1998.

[425] Angelique W Whitehurst. Cause and consequence of cancer/testis antigen activation in cancer. *Annual review of pharmacology and toxicology*, 54:251–272, 2014.

[426] Janica Lea Wiederstein, Hendrik Nolte, Stefan Günther, Tanja Piller, Martina Baraldo, Sawa Kostin, Wilhelm Bloch, Natalie Schindler, Marco Sandri, Bert Blaauw, Thomas Braun, Soraya Hölper, and Marcus Krüger. Skeletal muscle-specific methyltransferase mettl21c trimethylates p97 and regulates autophagy-associated protein breakdown. *Cell reports*, 23:1342–1356, May 2018.

[427] Ruth R E Williams. Transcription and the territory: the ins and outs of gene positioning. *Trends in genetics : TIG*, 19:298–302, June 2003.

[428] H Witt, W Luck, H C Hennies, M Classen, A Kage, U Lass, O Landt, and M Becker. Mutations in the gene encoding the serine protease inhibitor, kazal type 1 are associated with chronic pancreatitis. *Nature genetics*, 25:213–216, June 2000.

[429] Bartosz Wojtas, Aleksandra Pfeifer, Malgorzata Oczko-Wojciechowska, Jolanta Krajewska, Agnieszka Czarniecka, Aleksandra Kukulska, Markus Eszlinger, Thomas Musholt, Tomasz Stokowy, Michal Swierniak, Ewa Stobiecka, Ewa Chmielik, Dagmara Rusinek, Tomasz Tyszkiewicz, Monika Halczok, Steffen Hauptmann, Dariusz Lange, Michal Jarzab, Ralf Paschke, and Barbara Jarzab. Gene expression (mrna) markers for differentiating between malignant and benign follicular thyroid tumours. *International journal of molecular sciences*, 18, June 2017.

[430] Beata Wojtczak, Bartosz Pula, Agnieszka Gomulkiewicz, Mateusz Olbromski, Marzena Podhorska-Okolow, Paweł Domoslawski, Marek Bolanowski, Jacek Daroszewski, and Piotr Dziegiel. Metallothionein isoform expression in benign and malignant thyroid lesions. *Anticancer research*, 37:5179–5185, September 2017.

[431] Brian M Wolpin, Cosmeri Rizzato, Peter Kraft, Charles Kooperberg, Gloria M Petersen, Zhaoming Wang, Alan A Arslan, Laura Beane-Freeman, Paige M Bracci, Julie Buring, Federico Canzian, Eric J

Duell, Steven Gallinger, Graham G Giles, Gary E Goodman, Phyllis J Goodman, Eric J Jacobs, Aruna Kamineni, Alison P Klein, Laurence N Kolonel, Matthew H Kulke, Donghui Li, Núria Malats, Sara H Olson, Harvey A Risch, Howard D Sesso, Kala Visvanathan, Emily White, Wei Zheng, Christian C Abnet, Demetrius Albanes, Gabriella Andreotti, Melissa A Austin, Richard Barfield, Daniela Basso, Sonja I Berndt, Marie-Christine Boutron-Ruault, Michelle Brotzman, Markus W Büchler, H Bas Bueno-de Mesquita, Peter Bugert, Laurie Burdette, Daniele Campa, Neil E Caporaso, Gabriele Capurso, Charles Chung, Michelle Cotterchio, Eithne Costello, Joanne Elena, Niccola Funel, J Michael Gaziano, Nathalia A Giese, Edward L Giovannucci, Michael Goggins, Megan J Gorman, Myron Gross, Christopher A Haiman, Manal Hassan, Kathy J Helzlsouer, Brian E Henderson, Elizabeth A Holly, Nan Hu, David J Hunter, Federico Innocenti, Mazda Jenab, Rudolf Kaaks, Timothy J Key, Kay-Tee Khaw, Eric A Klein, Manolis Kogevinas, Vittorio Krogh, Juozas Kupcinskas, Robert C Kurtz, Andrea LaCroix, Maria T Landi, Stefano Landi, Loic Le Marchand, Andrea Mambrini, Satu Mannisto, Roger L Milne, Yusuke Nakamura, Ann L Oberg, Kouros Owzar, Alpa V Patel, Petra H M Peeters, Ulrike Peters, Raffaele Pezzilli, Ada Piepoli, Miquel Porta, Francisco X Real, Elio Riboli, Nathaniel Rothman, Aldo Scarpa, Xiao-Ou Shu, Debra T Silverman, Pavel Soucek, Malin Sund, Renata Talar-Wojnarowska, Philip R Taylor, George E Theodoropoulos, Mark Thornquist, Anne Tjønneland, Geoffrey S Tobias, Dimitrios Trichopoulos, Pavel Vodicka, Jean Wactawski-Wende, Nicolas Wentzensen, Chen Wu, Herbert Yu, Kai Yu, Anne Zeleniuch-Jacquotte, Robert Hoover, Patricia Hartge, Charles Fuchs, Stephen J Chanock, Rachael S Stolzenberg-Solomon, and Laufey T Amundadottir. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nature genetics*, 46:994–1000, September 2014.

[432] Chunlei Wu, Xuefeng Jin, Ginger Tsueng, Cyrus Afrasiabi, and Andrew I Su. Biogps: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res*, 44(D1):D313–D316, Jan 2016.

[433] Chunlei Wu, Ian Macleod, and Andrew I Su. Biogps and mygene.info: organizing online, gene-centric information. *Nucleic Acids Res*, 41(Database issue):D561–D565, Jan 2013.

[434] Chunlei Wu, Camilo Orozco, Jason Boyer, Marc Leglise, James Goodale, Serge Batalov, Christopher L Hodge, James Haase, Jeff Janes, Jon W Huss, and Andrew I Su. Biogps: an extensible and

customizable portal for querying and organizing gene annotation resources. *Genome biology*, 10:R130, 2009.

[435] J J Wyrick, J G Aparicio, T Chen, J D Barnett, E G Jennings, R A Young, S P Bell, and O M Aparicio. Genome-wide distribution of orc and mcm proteins in s. cerevisiae: high-resolution mapping of replication origins. *Science (New York, N.Y.)*, 294:2357–2360, December 2001.

[436] Endi Xia, Sotaro Kanematsu, Yusuke Suenaga, Asmaa Elzawahry, Hitomi Kondo, Noriko Otsuka, Yasumitsu Moriya, Toshihiko Iizasa, Mamoru Kato, Ichiro Yoshino, and Sana Yokoi. Microrna induction by copy number gain is associated with poor outcome in squamous cell carcinoma of the lung. *Scientific reports*, 8:15363, October 2018.

[437] Kaipeng Xie, Chenyang Fu, Suli Wang, Hanzi Xu, Siyu Liu, Yang Shao, Zhen Gong, Xiaoli Wu, Bo Xu, Jing Han, Juan Xu, Pengfei Xu, Xuemei Jia, and Jiangping Wu. Cancer-testis antigens in ovarian cancer: implication for biomarkers and therapeutic targets. *Journal of ovarian research*, 12:1, January 2019.

[438] X Xie, R Schmits, C Renner, D Preuss, B Kubuschok, and M Pfreundschuh. Systematic search and molecular characterization of the antigenic targets of myeloma immunoglobulins: a monoclonal iga from a female patient targeting sperm-specific cylicin ii. *Cancer immunity*, 1:11, December 2001.

[439] Tongbai Xu, Dongsheng Li, Yuan He, Fuliang Zhang, Man Qiao, and Yanhua Chen. The expression level of csdap1 in lung cancer and its clinical significance. *Oncology letters*, 16:4361–4366, October 2018.

[440] Andrew O Yam and Tatyana Chtanova. The ins and outs of chemokine-mediated immune cell trafficking in skin cancer. *Frontiers in immunology*, 10:386, 2019.

[441] Hiroshi Y Yamada and Chinthalapally V Rao. Brd8 is a potential chemosensitizing target for spindle poisons in colorectal cancer therapy. *International journal of oncology*, 35:1101–1109, November 2009.

[442] Fumiichiro Yamamoto and Miyako Yamamoto. Scanning copy number and gene expression on the 18q21-qter chromosomal region by the systematic multiplex pcr and reverse transcription-pcr methods. *Electrophoresis*, 28:1882–1895, June 2007.

[443] Ping Yang, Weilong Zhang, Jing Wang, Yuanyuan Liu, Ran An, and Hongmei Jing. Genomic landscape and prognostic analysis of mantle cell lymphoma. *Cancer gene therapy*, 25:129–140, June 2018.

[444] Damian B Yap, David C Walker, Leah M Prentice, Steven McKinney, Gulisa Turashvili, Katrin Mooslehner-Allen, Teresa Ruiz de Algara, John Fee, Xavier d'Anglemont de Tassigny, William H Colledge, and Samuel Aparicio. Mll5 is required for normal spermatogenesis. *PloS one*, 6:e27127, 2011.

[445] Mengsi Ye, Yun He, Hao Lin, Shouxing Yang, Yuhui Zhou, Lingling Zhou, Jinwei Zhong, Guangrong Lu, Jihang Zheng, Zhan-Xiong Xue, and Zhen-Zhai Cai. High expression of atonal homolog 8 predicts a poor clinical outcome in patients with colorectal cancer and contributes to tumor progression. *Oncology reports*, 37:2955–2963, May 2017.

[446] Ying Ye, Ya-Nan Song, Sai-Fei He, Ju-Hua Zhuang, Guo-Yu Wang, and Wei Xia. Gins2 promotes cell proliferation and inhibits cell apoptosis in thyroid cancer by regulating cited2 and loxl2. *Cancer gene therapy*, September 2018.

[447] Guo Nan Yin, Ho Won Lee, Je-Yoel Cho, and Kyoungho Suk. Neuronal pentraxin receptor in cerebrospinal fluid as a potential biomarker for neurodegenerative diseases. *Brain research*, 1265:158–170, April 2009.

[448] Dallah Yoo, Han-Joon Kim, Jin Sook Lee, Sangmoon Lee, Soo Yeon Kim, Murim Choi, Jong-Hee Chae, and Beomseok Jeon. Early-onset generalized dystonia starting in the lower extremities in a patient with a novel ano3 variant. *Parkinsonism & related disorders*, 50:124–125, May 2018.

[449] K Yoshimoto, T Fujimoto, A Itaya-Hironaka, T Miyaoka, S Sakuramoto-Tsuchida, A Yamauchi, M Takeda, T Kasai, K Nakagawara, A Nonomura, and S Takasawa. Involvement of autoimmunity to reg, a regeneration factor, in patients with primary sjögren's syndrome. *Clinical and experimental immunology*, 174:1–9, October 2013.

[450] Liqin Yuan, Yuzhong Xiao, Qiuzhi Zhou, Dongmei Yuan, Baiping Wu, Gannong Chen, and Jianlin Zhou. Proteomic analysis reveals that mael, a component of nuage, interacts with stress granule proteins in cancer cells. *Oncology reports*, 31:342–350, January 2014.

[451] Jin-Sung Yuk, Eun-Ju Park, Yong-Soo Seo, Hee Jin Kim, Seon-Young Kwon, and Won I Park. Graves disease is associated with endometriosis: A 3-year population-based cross-sectional study. *Medicine*, 95:e2975, March 2016.

[452] Fernando Tadeu Zamunér, Bruno Takao Real Karia, Cleyton Zanardo de Oliveira, Carlos Roberto Dos Santos, André Lopes Carvalho, and

André Luiz Vettore. A comprehensive expression analysis of cancer testis antigens in head and neck squamous cell carcinoma revels magea3/6 as a marker for recurrence. *Molecular cancer therapeutics*, 14:828–834, March 2015.

[453] Geng Zhang, Peijun He, Hanson Tan, Anuradha Budhu, Jochen Gaedcke, B Michael Ghadimi, Thomas Ried, Harris G Yfantis, Dong H Lee, Anirban Maitra, Nader Hanna, H Richard Alexander, and S Perwez Hussain. Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 19:4983–4993, September 2013.

[454] Huiyuan Zhang, Qinyi Wang, Junzhong Gu, Le Yin, Shenghui Liang, Lida Wu, Hao Xu, Chao Zhao, and Yuchun Gu. Elevated mitochondrial slc25a29 in cancer modulates metabolic status by increasing mitochondria-derived nitric oxide. *Oncogene*, 37:2545–2558, May 2018.

[455] Liqing Zhang and Wen-Hsiung Li. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Molecular biology and evolution*, 21:236–239, February 2004.

[456] Wenzhong Zheng, Zihao Zou, Shouren Lin, Xiang Chen, Feixiang Wang, Xianxin Li, and Jican Dai. Identification and functional analysis of spermatogenesis-associated gene modules in azoospermia by weighted gene coexpression network analysis. *Journal of cellular biochemistry*, September 2018.

[457] Cheng-Ji Zhou, Nobuya Inagaki, Samuel J Pleasure, Li-Xia Zhao, Sakae Kikuyama, and Seiji Shioda. Atp-binding cassette transporter abca2 (abc2) expression in the developing spinal cord and pns during myelination. *The Journal of comparative neurology*, 451:334–345, September 2002.

[458] Guangjin Zhou, Jian Wang, Yue Zhang, Chaomin Zhong, Jun Ni, Li Wang, Jinhu Guo, Kexiong Zhang, Long Yu, and Shouyuan Zhao. Cloning, expression and subcellular localization of hn1 and hn1l genes, as well as characterization of their orthologs, defining an evolutionarily conserved gene family. *Gene*, 331:115–123, April 2004.

[459] Jian Zhou, Jieyan Pan, Sigrid Eckardt, N Adrian Leu, K John McLaughlin, and P Jeremy Wang. Nxf3 is expressed in sertoli cells, but is dispensable for spermatogenesis. *Molecular reproduction and development*, 78:241–249, April 2011.

[460] Shao Hu Zhou, Yu Fei Deng, Zhi Wei Weng, Hao Wei Weng, and Zhi Dan Liu. Traditional chinese medicine as a remedy for male infertility: A review. *The world journal of men's health*, January 2019.

[461] Shasha Zhou, Yihang Shen, Min Zheng, Linlin Wang, Raymond Che, Wanning Hu, and Pin Li. Dna methylation of mettl7a gene body regulates its transcriptional level in thyroid cancer. *Oncotarget*, 8:34652–34660, May 2017.

[462] G Z Zhu, Y Lin, D G Myles, and P Primakoff. Identification of four novel adams with potential roles in spermatogenesis and fertilization. *Gene*, 234:227–237, July 1999.

[463] Hongling Zhu, Qian Xi, Lianyong Liu, Jingnan Wang, and Mingjun Gu. Quantitative assessment of common genetic variants on foxe1 and differentiated thyroid cancer risk. *PloS one*, 9:e87332, 2014.

[464] Tongyu Zhu, Jing Yuan, Yuzhi Wang, Cuiping Gong, Yudou Xie, and Hong Li. Mir-661 contributed to cell proliferation of human ovarian cancer cells by repressing inpp5j expression. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, 75:123–128, October 2015.

[465] Xiaoxiao Zuo, Ling Chen, Lifeng Liu, Zhe Zhang, Xiaojin Zhang, Qing Yu, Lu Feng, Xinhan Zhao, and Tianjie Qin. Identification of a panel of complex autoantigens (lgals3, phb2, muc1, and gk2) in combination with ca15-3 for the diagnosis of early-stage breast cancer. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, 37:1309–1317, January 2016.

[466] O Zwermann, A C Braun, E Lalli, P Sassone-Corsi, F Beuschlein, and M Reincke. Regulation of human mc2-r gene expression by creb, crem, and icer in the adrenocortical cell line y1. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme*, 39:560–566, August 2007.

5 Technical appendix

5.1 Part A: programming code

In the thechnical appendix, there is all programming code, used in this thesis. The ready to use R scripts are on CD, as well as in a printed form in the technical appendix, part A: programming code as a .pdf file.

All programming code is owned by the author and can only be used by demanding and citing the authors rights, as well as the title of this thesis.

All R scripts were programmed in the open source language R, ready to use as downloadable packages from the CRAN network, as well as bioconductor.

All packages, perl scripts, shell scripts, PHP language of the database TRADB as well as annotation packages refere the standart of the most updated version at the time of programming.

If you want to use any of the here presented R code, please ask the author for permission. And cite the title above.

5.1.1 Data retrieval Microarray data

For the retrieval of data we used the GEO database (gene expression omnibus) for all microarray datasets, with the GEO accession numbers, refered to in the methods part of this work.

The RNA sequencing data was drawn from the GTEX consortium and downloaded ad ready to use RPKM values.

The annotation of the microarray data was done with the most actual package of the brainarray database and later gene annotation was done by the biomart database.

The pre processing steps, as well as quality control of the microarray data is described in all detail in the methods part of this work, as well as commented in the original R scripts.

All data processing as well as plotting, as well as most of the programming has been done in the open source programming language R, some scripts as well as programs were run in Perl or started via shell scripts directly from the comment line.

All different datasets have been imported into R, as well as been processed in the exact same way, following the same workflows, sometimes with differ-

ent version of the most updated annotation packages, since we always used the newest annotation versions of all packages.

The programming code has been commented well, so that re-programming should be farely easy for everyone firm with the open source programming language R. Please ask the author for permission.

For the R script of how to import and process Microarray data, do quality control, do normalization, calculation of the mean vsnrma and averaging over double measurements, please refere to the script.

Skript: analysis_mouse_gngnf1_chips.R

5.1.2 Data retrieval GTEX data

Data retieval and download of the human GTEX Sequencing data was done from the GTEX Consortium. The data was downloaded as ready to use RPKM values, for annotation their annotation was used.

Skript: analysis_human_gtex_data.R

5.1.3 Calculating tissue-restricted antigens (TRAs)

For the calculation of tissue-restricetd antigens (TRAs) for all datasets a stringent criteria was defound. A gene or transcript is considered to be tissue-restricted if it exceeds in at least one, but not more than five tissues of all tissues of the dataset the cutoff of 5x the median gene expression over all tissues. TRAs were plotted as .png and TRA tables were established for each dataset, in order to import it into the TRA-DB.

Similar tissues were grouped together to groups and only considered as one tissue in the TRA calculation. The exact definition can be found in the methods part of this thesis.

All programming code can be found in the technical appendix on CD in the following scripts.

Skript: calc_tras_mouse_gngnf1.R

Skript: calc_tras_human_GTEX_data.R

5.1.4 Chromosomal clustering of TRAs

For the calculation of chromosomal clustering of tissue-restricted antigens (TRAs), as well as housekeeping genes or other functionally related genes, please refer to the following R script in the technical appendix on CD.

For the cluster analysis of chromosomal clustering, we used two different methods, the first method is the 10-gene window method, which takes a hash table with all genes of interest, as for example all TRAs as well as a table of all background genes annotated for gene name, chromosome and startsite as an input and is calculated with a perl script, which can be started by the command line. How to use it, is described in the R scripts above. The perl script

calculates the number duplets, triplets, quadruplets, etc. in a sliding 10-gene window method, as well in the TRA list or gene list of interest, as also in 1000 randomly picked genes with the same length of interest. The calculation of the TRA lists is started with the shell script

And the 1000 randomly drawn gene lists are started with the shell script

For the second clustering method we used a sliding-gene window method of fixed window sizes of different sizes of kb windows, it calculates the number of neighbors within a sliding gene window of fixed size. The same is done for 1000 randomly drawn gene lists of the same length. For this we used the two following scripts

Skript: calc_clustering_gngnf1.R

Skript: observed_ntuples_mouse.pl

Skript: start_perl_batch_mouse.sh

Skript: validate_ntuples_mouse.pl

Skript: permute_chrloc.R

Skript: start_R_batch.sh

Script: dist.genloc.R

5.1.5 TRA plots and plotting chromosomal clustering of TRAs

TRA clusters were plotted in different versions and views and further documented in TRA cluster tables.

Skript: cluster_plots_gngnf1.R

Skript: cluster_table_gngnf1.R

Script: barplots.R

Skript: errorbars.R

Script: chromosomenmap.R

Script: ebars.R

Script: extract.chromosome.R

Script: plotten.eps.R

Script: plotten.png.R

Script: plotten.R

Script: plot_ybmat.R

5.1.6 Gene annotation and TRA tables

This script is a documentation of gene and transcript numbers of the different versions of annotations.

Script: annotations.R

Script: merge.table.pl

Script: findRedundant.R

Script: pasteList.R

5.1.7 Plot TRA clusters

Script: plot.cluster.R

Script: human.print.cluster.R

Script: print.cluster1.R

Script: print.cluster.R

Script: print.two.clusters.R

Script: stats.R

5.1.8 Aire genes in TRAs

Aire regulated genes were calculated from an Illumina Microarray dataset by Sheena Pinto et al. and the overlap with TRAs was calculated.

Script: aire.genes.in.tras.R

5.1.9 Homology plots

Script: homology_in_clusters.R

Script: homology.R

5.1.10 Synteny maps

This script calculates the synteny maps between human, mouse and rat.

Script: synteny_maps.R

Script: print.synteny.maps.R

Script: print.synteny.maps_tissues.R

5.1.11 Calculate paired t-tests

Script: paired_ttest.R

Script: venn6dim.R

5.2 Part B: Additional figures

Additional figures are included in the technical appendix, Part B on CD.

5.2.1 Quality control

Figure 1 - Quality control - single chip control in the rat Novartis dataset

Figure 2 - Quality control - single chip control in the human Roth dataset 1

Figure 3 - Quality control - single chip control in the human Roth dataset 2

Figure 4 - Quality control - single chip control in the mouse Lattin dataset

5.2.2 Quality control - RNA degradation plots

Figure 5 - Quality control - RNA degradation plot in the human Novartis dataset

Figure 6 - Quality control - RNA degradation plot in the mouse Lattin dataset

Figure 7 - Quality control - RNA degradation plot in the human Roth dataset

Figure 8 - Quality control - RNA degradation plot in the rat Novartis dataset

5.2.3 Boxplots of all datasets

Figure 9 - Quality control - Boxplot in the human Novartis dataset before normalization

Figure 10 - Quality control - Boxplot in the human Novartis dataset after vsnrma normalization

Figure 11 - Quality control - Boxplot in the human Novartis dataset mean vsnrma

Figure 12 - Quality control - Boxplot in the mouse Lattin dataset before normalization

Figure 13 - Quality control - Boxplot in the mouse Lattin dataset after vsnrma normalization

Figure 14 - Quality control - Boxplot in the mouse Lattin dataset mean vsnrma

Figure 15 - Quality control - Boxplot in the human Roth dataset before normalization

Figure 16 - Quality control - Boxplot in the human Roth dataset after vsnrma normalization

Figure 17 - Quality control - Boxplot in the human Roth dataset mean vsnrma

5.2.4 Density plot before and after vsnrma normalization

Figure 18 - Quality control - Density plot before and after vsnrma normalization in the human Novartis dataset

Figure 19 - Quality control - Density plot before and after vsnrma normalization in the mouse Lattin dataset

Figure 20 - Quality control - Density plot before and after vsnrma normalization in the human Roth dataset

5.2.5 Tissue types in the different datasets

Figure 21 - Tissue types in the mouse Lattin dataset

Figure 22 - Tissue types in the human Roth dataset

5.2.6 Saturationplots in the different datasets

Figure 23 - Saturationplot in the human Novartis dataset

Figure 24 - Saturationplot in the mouse Lattin dataset

Figure 25 - Saturationplot in the human Roth dataset

Figure 26 - Saturationplot in the human GTEX dataset 1

Figure 27 - Saturationplot in the human GTEX dataset 2

5.2.7 Some example TRAs in all different datasets

Figure 28a - Complement component C2, TRA example in the human Novartis dataset

Figure 28b - CD53 cell surface protein for signal transduction, TRA example in the human Novartis dataset

Figure 28c - CD3D part of the T cell receptor complex, TRA example in the human Novartis dataset

Figure 29a - HLA-DPB1, TRA example in the human Novartis dataset

Figure 29b - APOD apolipoprotein D, TRA example in the human Novartis dataset

Figure 30a - CHFR, checkpoint protein from the cell cycle, non-TRA example in the human Novartis dataset

Figure 30b - HBB, hemoglobin subunit, non-TRA example in the human Novartis dataset

Figure 31a - ANKRD, ankyrin repeat domain, TRA example in the mouse Lattin dataset

Figure 31b - Ehd1, epidermal growth factor receptor, TRA example in the mouse Lattin dataset

Figure 31c - Gm20459, paralemmin A kinase anchor protein, TRA example in the mouse Lattin dataset

Figure 32a - Fam3b, TRA example in the mouse Lattin dataset

Figure 32b - Fyco1, TRA example in the mouse Lattin dataset

Figure 33a - Bcl2l13, non-TRA example in the mouse Lattin dataset

Figure 33b - Fbp2, fructose bisphosphatase 2 gene, non-TRA example in the mouse Lattin dataset

Figure 34a - C2, complement component, TRA example in the human Roth dataset

Figure 34b - APOA1, Apolipoprotein A1, TRA example in the human Roth dataset

Figure 34c - C11orf54, TRA example in the human Roth dataset

Figure 35a - ABLIM1, actin binding LIM Protein 1, TRA example in the

human Roth dataset

Figure 35b - CST6 protein, TRA example in the human Roth dataset

Figure 36a - ACBD4 gene, non-TRA example in the human Roth dataset

Figure 36b - FAM3B, a pancreatic derived factor, non-TRA example in the human Roth dataset

Figure 37a - B1 like protein, TRA example in the human GTEX dataset

Figure 37b - BTG2 Protein, TRA example in the human GTEX dataset

Figure 37c - GLIS1, family zinc finger 1 protein, TRA example in the human GTEX dataset

Figure 38a - ERICH3 gene, TRA example in the human GTEX dataset, Transcript 1

Figure 38b - MTMR9LP, myotubularin related protein 9 like pseudogene, TRA example in the human GTEX dataset

Figure 39a - ERICH3 gene, non-TRA example in the human GTEX dataset, Transcript 2

Figure 39b - NBL1 gene, non-TRA example in the human GTEX dataset

5.2.8 TRA clusters per chromosome in the human and mouse Novartis dataset

Figure 40 - The distribution of tissue specificity on chromosome number 1 and 2 in the mouse Novartis dataset.

Figure 41 - The distribution of tissue specificity on chromosome number 3 and 4 in the mouse Novartis dataset.

Figure 42 - The distribution of tissue specificity on chromosome number 5 and 6 in the mouse Novartis dataset.

Figure 43 - The distribution of tissue specificity on chromosome number 7 and 8 in the mouse Novartis dataset.

Figure 44 - The distribution of tissue specificity on chromosome number 9 and 10 in the mouse Novartis dataset.

Figure 45 - The distribution of tissue specificity on chromosome number 11 and 12 in the mouse Novartis dataset.

Figure 46 - The distribution of tissue specificity on chromosome number 13 and 14 in the mouse Novartis dataset.

Figure 47 - The distribution of tissue specificity on chromosome number 15 and 16 in the mouse Novartis dataset.

Figure 48 - The distribution of tissue specificity on chromosome number 17 and 18 in the mouse Novartis dataset.

Figure 49 - The distribution of tissue specificity on chromosome number 19 and X in the mouse Novartis dataset.

Figure 50 - The distribution of tissue specificity on chromosome number 1 and 2 in the human Novartis dataset.

Figure 51 - The distribution of tissue specificity on chromosome number 3 and 4 in the human Novartis dataset.

Figure 52 - The distribution of tissue specificity on chromosome number 5 and 6 in the human Novartis dataset.

Figure 53 - The distribution of tissue specificity on chromosome number 7 and 8 in the human Novartis dataset.

Figure 54 - The distribution of tissue specificity on chromosome number 9 and 10 in the human Novartis dataset.

Figure 55 - The distribution of tissue specificity on chromosome number 11 and 12 in the human Novartis dataset.

Figure 56 - The distribution of tissue specificity on chromosome number 13 and 14 in the human Novartis dataset.

Figure 57 - The distribution of tissue specificity on chromosome number 15 and 16 in the human Novartis dataset.

Figure 58 - The distribution of tissue specificity on chromosome number 17 and 18 in the human Novartis dataset.

Figure 59 - The distribution of tissue specificity on chromosome number

19 and 20 in the human Novartis dataset.

Figure 60 - The distribution of tissue specificity on chromosome number 21 and 22 in the human Novartis dataset.

Figure 61 - The distribution of tissue specificity on chromosome number X and Y in the human Novartis dataset.

5.2.9 Synteny plots of TRA clusters between human, mouse and rat

Figure 62 - Synteny plot of TRA cluster nr. 2 on chromosome 1 in and cluster nr. 3 on chromosome 13 in the mouse Novartis data.

Figure 63 - Synteny plot of TRA cluster nr. 3 on chromosome 1 and cluster nr. 6 on chromosome 1 in the mouse Novartis data.

Figure 64 - Synteny plot of TRA cluster nr. 7 on chromosome 1 and cluster nr. 8 on chromosome 1 in the mouse Novartis data.

Figure 65 - Synteny plot of TRA cluster nr. 3 on chromosome 19 and cluster nr. 13 on chromosome 2 in the mouse Novartis data.

Figure 66 - Synteny plot of the TRA cluster nr. 8 on chromosome 2 in the mouse Novartis data.

5.2.10 Synteny maps on tissue-specificity in TRA clusters between human and mouse

Figure 67 - Synteny plot with tissue specificity in TRA clusters, TRA cluster nr. 3 on chromosome nr. 1 and cluster nr. 8 on chromosome nr. 2 in the mouse Novartis data.

Figure 68 - Synteny plot with tissue specificity in TRA clusters, TRA cluster nr. 16 on chromosome nr. 2 and cluster nr. 11 on chromosome nr. 4 in the mouse Novartis data.

Figure 69 - Synteny plot with tissue specificity in TRA clusters, TRA cluster nr. 13 on chromosome nr. 6 and cluster nr. 10 on chromosome nr. 7 in the mouse Novartis data.

Figure 70 - Synteny plot with tissue specificity in TRA clusters, TRA cluster nr. 20 on chromosome nr. 7 and cluster nr. 4 on chromosome nr. 9 in the mouse Novartis data.

Figure 71 - Synteny plot with tissue specificity in TRA clusters, TRA cluster nr. 8 on chromosome nr. 9 and cluster nr. 2 on chromosome nr. 14 in the mouse Novartis data.

Figure 72 - Synteny plot with tissue specificity in TRA clusters, TRA cluster nr. 7 on chromosome nr. 15 and cluster nr. 5 on chromosome nr. 19 in the mouse Novartis data.

5.2.11 conservation of TRAs in the evolutionary tree

Figure 73 - TRAs in the evolutionary Tree, cluster nr. 2 on chromosome nr. 5 in the mouse Novartis data.

Figure 74 - TRAs in the evolutionary Tree, cluster nr. 6 on chromosome nr. 5 in the mouse Novartis data.

Figure 75 - TRAs in the evolutionary Tree, cluster nr. 7 on chromosome nr. 5 in the mouse Novartis data.

Figure 76 - TRAs in the evolutionary Tree, cluster nr. 2 on chromosome nr. 6 in the mouse Novartis data.

Figure 77 - TRAs in the evolutionary Tree, cluster nr. 3 on chromosome nr. 6 in the mouse Novartis data.

Figure 78 - TRAs in the evolutionary Tree, cluster nr. 9 on chromosome nr. 7 in the mouse Novartis data.

Figure 79 - TRAs in the evolutionary Tree, cluster nr. 1 on chromosome nr. 8 in the mouse Novartis data.

5.2.12 TRA clusters in the mouse

Figure 80 - TRA clusters on chromosome nr. 2, 3 and 4 - overview in the mouse Novartis dataset.

Figure 81 - TRA clusters on chromosome nr. 5, 6 and 7 - overview in the mouse Novartis dataset.

Figure 82 - TRA clusters on chromosome nr. 8, 10 and 11 - overview in the mouse Novartis dataset.

Figure 83 - TRA clusters on chromosome nr. 12, 13 and 14 - overview

in the mouse Novartis dataset.

Figure 84 - TRA clusters on chromosome nr. 15, 16 and 17 - overview in the mouse Novartis dataset.

Figure 85 - TRA clusters on chromosome nr. 18, 19 and X - overview in the mouse Novartis dataset.

5.2.13 TRA clustered genes in the mouse

Figure 86 - TRA clusters gene wise on chromosome nr. 2, 3 and 4 in the mouse Novartis dataset.

Figure 87 - TRA clusters gene wise on chromosome nr. 5, 6 and 7 in the mouse Novartis dataset.

Figure 88 - TRA clusters gene wise on chromosome nr. 8, 10 and 11 in the mouse Novartis dataset.

Figure 89 - TRA clusters gene wise on chromosome nr. 12, 13 and 14 in the mouse Novartis dataset.

Figure 90 - TRA clusters gene wise on chromosome nr. 15, 16 and 17 in the mouse Novartis dataset.

Figure 91 - TRA clusters gene wise on chromosome nr. 18, 19 and X in the mouse Novartis dataset.

5.2.14 TRA clustered tissues in the mouse

Figure 92 - Tissue types in TRA cluster nr. 5 and nr. 6 on chromosome nr. 1 in the mouse Novartis data.

Figure 93 - Tissue types in TRA cluster nr. 8 on chromosome nr. 1 and cluster nr. 3 on chromosome nr. 2 in the mouse Novartis data.

Figure 94 - Tissue types in TRA cluster nr. 8 and nr. 13 on chromosome nr. 2 in the mouse Novartis data.

Figure 95 - Tissue types in TRA cluster nr. 18 on chromosome nr. 2 and cluster nr. 2 on chromosome nr. 3 in the mouse Novartis data.

Figure 96 - Tissue types in TRA cluster nr. 5 on chromosome nr. 3 and cluster nr. 15 on chromosome nr. 4 in the mouse Novartis data.

Figure 97 - Tissue types in TRA cluster nr. 5 and nr. 7 on chromosome nr. 5 in the mouse Novartis data.

Figure 98 - Tissue types in TRA cluster nr. 2 and nr. 5 on chromosome nr. 6 in the mouse Novartis data.

Figure 99 - Tissue types in TRA cluster nr. 13 on chromosome nr. 6 and cluster nr. 7 on chromosome nr. 7 in the mouse Novartis data.

Figure 100 - Tissue types in TRA cluster nr. 10 and nr. 11 on chromosome nr. 7 in the mouse Novartis data.

Figure 101 - Tissue types in TRA cluster nr. 1 and nr. 9 on chromosome nr. 8 in the mouse Novartis data.

Figure 102 - Tissue types in TRA cluster nr. 4 and nr. 6 on chromosome nr. 9 in the mouse Novartis data.

Figure 103 - Tissue types in TRA cluster nr. 7 and nr. 8 on chromosome nr. 9 in the mouse Novartis data.

Figure 104 - Tissue types in TRA cluster nr. 1 and nr. 3 on chromosome nr. 10 in the mouse Novartis data.

Figure 105 - Tissue types in TRA cluster nr. 7 on chromosome nr. 10 and cluster nr. 16 on chromosome nr. 11 in the mouse Novartis data.

Figure 106 - Tissue types in TRA cluster nr. 18 on chromosome nr. 11 and cluster nr. 1 on chromosome nr. 12 in the mouse Novartis data.

Figure 107 - Tissue types in TRA cluster nr. 5 on chromosome nr. 12 and cluster nr. 3 on chromosome nr. 13 in the mouse Novartis data.

Figure 108 - Tissue types in TRA cluster nr. 3 and nr. 7 on chromosome nr. 15 in the mouse Novartis data.

Figure 109 - Tissue types in TRA cluster nr. 5 on chromosome nr. 16 and cluster nr. 3 on chromosome nr. 17 in the mouse Novartis data.

Figure 110 - Tissue types in TRA cluster nr. 11 on chromosome nr. 17 and cluster nr. 2 on chromosome nr. 18 in the mouse Novartis data.

Figure 111 - Tissue types in TRA cluster nr. 4 on chromosome nr. 18

and cluster nr. 3 on chromosome nr. 19 in the mouse Novartis data.

Figure 112 - Tissue types in TRA cluster nr. 5 on chromosome nr. 19 and cluster nr. 3 on chromosome nr. X in the mouse Novartis data.