THESIS ON NATURAL AND EXACT SCIENCES B180

Regulator of G Protein Signalling 16 (RGS16): Functions in Immunity and Genomic Location in an Ancient MHC-Related Evolutionarily Conserved Synteny Group

JAANUS SUURVÄLI

TUT
PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Science
Department of Gene Technology

The dissertation was accepted for the defence of the degree of Doctor of Philosophy in Gene Technology on October 28, 2014.

Supervisor: **Assistant Professor Sirje Rüütel Boudinot, PhD**
Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia

Opponents: **Professor Pärt Peterson, PhD**
Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia
Professor Lars T. Hellman, PhD
Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

Defence of the thesis: December 9, 2014.

Declaration:
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology, has not been submitted for any other academic degree.

/Jaanus Suurväli/

ISSN 1406-4723
ISBN 978-9949-23-701-2 (publication)
ISBN 978-9949-23-702-9 (PDF)

LOODUS- JA TÄPPISTEADUSED B180

G valkude signaaliülekande regulaator 16 (RGS16): osalus immuunvastuses ja genoomne asukoht ürgses MHC-seoselises evolutsiooniliselt konserveerunud sünteenses grupis

JAANUS SUURVÄLI

TTÜ
KIRJASTUS

CONTENTS

INTRODUCTION

G protein-coupled receptors (GPCRs) are central to the eukaryotic signalling systems. The pathways initiated by ligands binding to them vary greatly and can have numerous physiological effects. In human the GPCRs are essential for the nervous system, for the immune system, and for the endocrine system. Hence, signalling via GPCRs needs to be tightly controlled by regulatory molecules such as Regulators of G protein Signalling (RGS).

Most members of the R4 subclass of RGS proteins are associated with functions in the immune system, although they are also found in the brain and other organs. One of them, RGS16, is even directly targeted by a virus that causes immune deficiency in pigs. RGS16 is also induced during antiviral responses and controls the expression of a number of genes associated with antiviral functions. However, it is also upregulated by other types of infection and impacts on MAP kinase activity.

RGS16 and most of the other R4 RGS are located close together in the genomes of both humans and other vertebrates, thus defining a conserved genomic region (the 'R4 RGS region'). The best studied region in vertebrate genomes that is involved in immune responses is the MHC which encodes for genes involved in antigen presentation (among others) and possesses four large paralogous segments in most vertebrate species. Intriguingly, early works have suggested the R4 RGS region to be associated with one of these duplicated copies based on its genomic location.

Susumu Ohno postulated already in 1970 that gene and genome duplication is one of the main driving forces behind evolution of new species. He further suggested that vertebrates may have undergone genome duplications during their early evolution, followed by a period of extensive chromosomal rearrangements, gene re-specialization and gene loss in order to restore a stable diploid state. This theory has gained much support during the last two decades and is now generally accepted as it provides the most plausible explanation for the four copies of Hox gene clusters, four copies of regions paralogous to the MHC, and a number of other similar cases in many vertebrate genomes.

With full sequences of many vertebrates available and the reconstruction of ancestral chordate linkage groups it appeared that many large genomic regions are conserved across vertebrates, even chordates. In fact, some conserved syntenies have even been found between vertebrates and some invertebrate species, suggesting they have been conserved throughout metazoan evolution. There are at least three theories explaining such conservation, involving conserved regulatory sequences of key developmental genes, beneficial functional co-regulation, or simply the contingent absence of recombination in the conserved regions. The proto-MHC or somewhat related regions have been identified in the amphioxus, tunicates, sea urchin, and even fruit fly. Such studies suggest that other regions derived from the same ancestral chromosome as the MHC encode for genes with immune functions in vertebrate genomes.

The aims of the present study were double. First, the impact of RGS16 on monocyte activation was assessed to get more insight on the roles of R4 RGS proteins in the immune system. Second, the hypothesis of an evolutionary connection between the R4 RGS region and the ancestral proto-MHC was tested by determining this gene set's three paralogous copies and tracking them throughout metazoan evolution. The results presented here suggest that the link between R4 RGS and proto-MHC has been well conserved across metazoans and that this region comprises of a new set of useful markers for tracking the MHC in invertebrate and ancestral species.

ORIGINAL PUBLICATIONS

The thesis is based on the following publications, which will be referred to in text by their Roman numerals:

I **Suurväli, J.,** Pahtma, M., Saar, R., Paalme, V., Nutt, A., Tiivel, T., Saaremäe, M., Fitting, C., Cavaillon, J.-M., Rüütel Boudinot, S. (2014). RGS16 restricts the pro-inflammatory response of monocytes. *Scandinavian Journal of Immunology*, in press. DOI: 10.1111/sji.12250

II **Suurväli, J.,** Robert, J., Boudinot, P., Rüütel Boudinot, S. (2013). R4 regulators of G protein signaling (RGS) identify an ancient MHC-linked synteny group. *Immunogenetics* 65(2):145 - 56. DOI: 10.1007/s00251-012-0661-x

III **Suurväli, J.,** Jouneau, L., Thépot, D., Grusea, S., Pontarotti, P., Du Pasquier, L., Rüütel Boudinot, S., Boudinot, P. (2014). The Proto-MHC of Placozoans, a Region Specialized in Cellular Stress and Ubiquitination/Proteasome Pathways. The *Journal of Immunology* 193(6):2891 - 901. DOI: 10.4049/jimmunol.1401177

ABBREVIATIONS

2R WGD	two whole genome duplications at the early vertebrate evolution
AIRE	autoimmune regulator
BDNF	brain-derived neurotrophic factor
CTX	cortical thymocyte marker of *Xenopus*
GAP	guanosine triphosphatase activating protein
GC	germinal centre
GDI	guanine nucleotide dissociation inhibitor
GDP	guanosine diphosphate
GEF	guanine nucleotide exchange factor
GPCR	G protein-coupled receptor
GRK	G protein-coupled receptor kinase
GTP	guanosine triphosphate
GTPase	guanosine triphosphatase
IFN	interferon
IgSF	immunoglobulin superfamily
IL	interleukin
IVNS1ABP	influenza virus non-structural protein 1A binding protein
JAM	junctional adhesion molecule
JN	JAM-NECTIN paralogy groups
KO	knockout (mouse)
LPS	lipopolysaccharide
MAPK	mitogen-activated protein kinase
MHC	major histocompatibility complex
MS	multiple sclerosis
NAHR	non-allelic homologous recombination
NF-κB	nuclear factor kappa-light-chain-enhancer of B cells
NGF	nerve growth factor
NHEJ	non-homologous end-joining
NK	natural killer (cell)
NT	neurotrophin paralogy groups
Pam3	Pam3CysSK4, a synthetic triacylated lipopeptide
PBMC	peripheral blood mononuclear cell
RAG	recombination-activating gene
RGS	regulator of G protein signalling
RH	regulator of G protein signalling homology domain
siRNA	small interfering ribonucleic acid
TAP	transporter associated with antigen processing
TLR	toll-like receptor
TNF	tumour necrosis factor
TRIM	tripartite motif
WGD	whole genome duplication
wt	wild type

GLOSSARY

2R WGD hypothesis	"Early evolution of vertebrates involved two consecutive rounds of whole genome duplication"
Agranular mono-nuclear myeloid cells	Promonocytes, monocytes, and monocyte derived cells, e.g. macrophages and dendritic cells
Allotetraploidy	Hybridization of two diploid genomes
Ancestral linkage group	A set of genes linked in a number of genomes, suggestive of its presence in the common ancestor
Autotetraploidy	Duplication of a diploid genome
Class I, II, III	Three main regions of the vertebrate MHC
Dosage balance hypothesis	"Changes in copy number of certain genes are selected against since it is balanced with that of other genes"
Fragile site	Genomic region that is prone to DNA damage
Immunoproteasome	Proteasome that is encoded for by genes linked to the MHC and involved in antigen processing
Linkage disequilibrium	Non-random association of alleles at two or more genetic loci
Neofunctionalization	One of the copies resulting from gene duplications gains a new function and is thus retained
Neo-subfunctionalization	Duplicated gene copies retain distinct functions of the original gene, but also obtain new functions
Ohnologs	Paralogs that originate from whole genome duplications
Paralogons	Paralogy regions derived from genome duplications
Proto-MHC, NT, JN	Invertebrate / ancestral regions that correspond to vertebrate MHC, NT, JN paralogons, respectively
Polyploidization	The process of a genome becoming polyploid
Re-diploidization	The restoration of diploid state
Scaffold	Genome fragment not assigned to a chromosome yet
Subfunctionalization	Copies of a gene keep distinct functions of the original gene and are thus retained
Synteny	Conservation of a genomic region between two or more species
Tetrapod	Four-limbed vertebrate
Ur-metazoan	The hypothetical common ancestor of metazoans

Different phyla and species discussed in this thesis are defined on Figure 1.

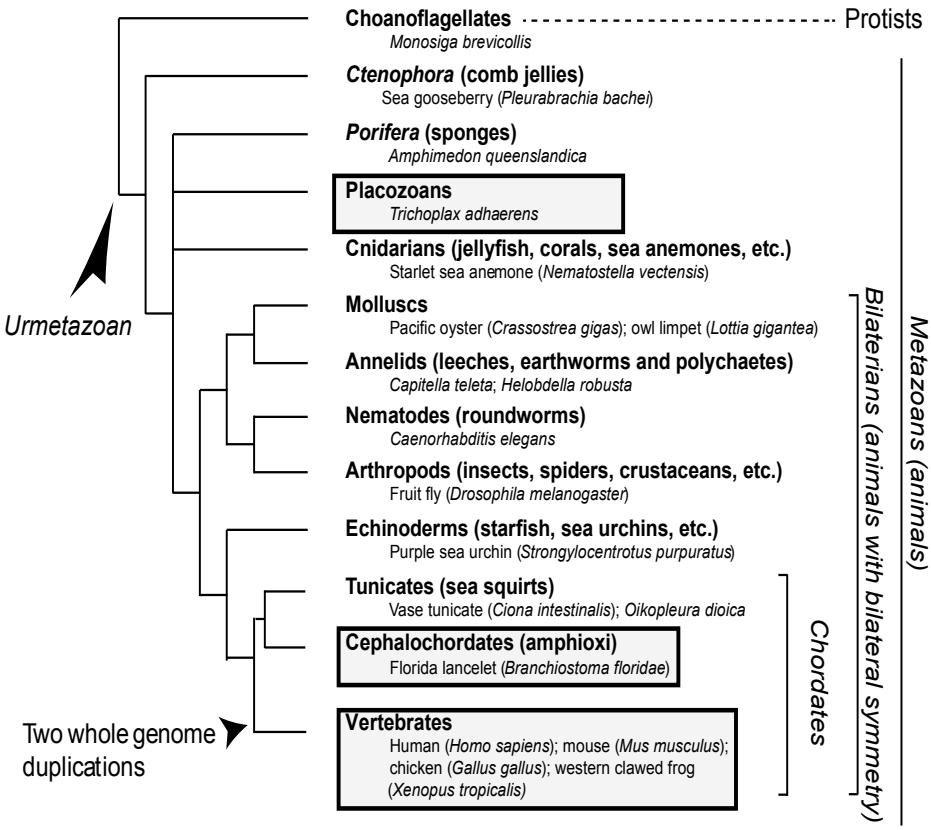


Figure 1. Phylogenetic relationships of animals discussed in the dissertation. Vertebrates, cephalochordates, and placozoans are marked with grey boxes. The hypothetical metazoan ancestor – the Ur-metazoan – is marked at the stem of metazoans. Modified from Publication III, Figure 1.

1. REVIEW OF THE LITERATURE

1.1 R4 Regulators of G protein Signalling: structures, functions and genomic context

1.1.1 G proteins and G protein-coupled receptors

The superfamily of G protein-coupled receptors (GPCRs) is found across eukaryotes and is classified into 5-6 families based on sequence homology and functional similarities (Fredriksson *et al.*, 2003; King *et al.*, 2003; Eichinger *et al.*, 2005; Wettschureck and Offermanns, 2005; Römpler *et al.*, 2007; Xue *et al.*, 2008; Krishnan *et al.*, 2012; Jékely, 2013; Krishnan *et al.*, 2013; Urano *et al.*, 2013; de Mendoza *et al.*, 2014; Isberg *et al.*, 2014). In general, vertebrate genomes encode for much more GPCRs than invertebrate genomes (Römpler *et al.*, 2007). In prokaryotes structurally similar sensory receptors have been identified (such as bacteriorhodopsins) and have been proposed to represent molecules that eventually evolved into eukaryote GPCRs (Pertseva and Shpakov, 2009; Ernst *et al.*, 2014; Zhang *et al.*, 2014). GPCRs are also encoded by the genomes of some large DNA viruses (such as herpesviruses and poxviruses), which is likely the result of horizontal gene transfer (Vischer *et al.*, 2006). In any case it appears that the presence of true GPCRs is specific to eukaryotes. It has been suggested that the genome of the common ancestor of all eukaryotes already encoded for all the components of a functional GPCR signalling pathway (de Mendoza *et al.*, 2014), and that even the ligands for GPCRs have co-evolved with the receptors themselves (Mirabeau and Joly, 2013).

While G protein-coupled receptors can have diverse ligand binding sites, a typical GPCR contains an extracellular N-terminus, seven membrane spanning helices, and an intracellular C-terminus (Venkatakrishnan *et al.*, 2014). An agonist binding to the extracellular ligand binding site induces conformational change in the receptor, enabling the cytosolic part of the protein to mediate the dissociation of small heterotrimeric G proteins into Gα and Gβγ subunits (Cabrera-Vera *et al.*, 2003). The dissociated G protein subunits can then activate intracellular pathways depending on the subunit subtype (Figure 2); different phylogenetic clades can express different subtypes due to evolutionary processes such as lineage-specific expansions and gene loss (Anantharaman *et al.*, 2011). The signal is eventually terminated by re-association of the G protein components (Cabrera-Vera *et al.*, 2003).

A heterotrimeric G protein's structural integrity is mainly determined by the state of its Gα subunit. There are four classes of Gα subunits in humans ($G\alpha_i$, $G\alpha_q$, $G\alpha_s$, and $G\alpha_{12/13}$) (Lattin *et al.*, 2007; Lagman *et al.*, 2012), but more in teleost fish and invertebrates (Lagman *et al.*, 2012). In a fully assembled Gαβγ heterotrimer the Gα subunit is bound to GDP; the G protein dissociates into Gα and Gβγ subunits when the GDP is replaced by GTP (Cabrera-Vera *et al.*, 2003).

The Gβγ dimer can be bound and stabilized by molecular chaperons related to phosducin (Lukov *et al.*, 2005). The Gα subunit itself contains an intrinsic GTPase activity, eventually resulting in GTP being hydrolysed back to GDP and subsequent Gαβγ re-association (Cabrera-Vera *et al.*, 2003; Wettschureck and Offermanns, 2005). In line with this it has been determined that the cytoplasmic tail of an activated GPCR contains a guanine nucleotide exchange factor (GEF) activity and thus catalyses the GDP→GTP exchange on the bound Gα subunit (Cabrera-Vera *et al.*, 2003). Each GPCR can couple to (and thus activate) only select types of Gαβγ heterotrimers, allowing for signal specificity (Cabrera-Vera *et al.*, 2003; Wong, 2003). Furthermore, GPCRs can form homo- and heterodimers with coupling and ligand specificities different from the individual receptors forming the dimer (Prinster *et al.*, 2005; Franco *et al.*, 2007; Kasai and Kusumi, 2014). Rapid switching between monomeric and homodimeric states has been observed (Kasai and Kusumi, 2014).

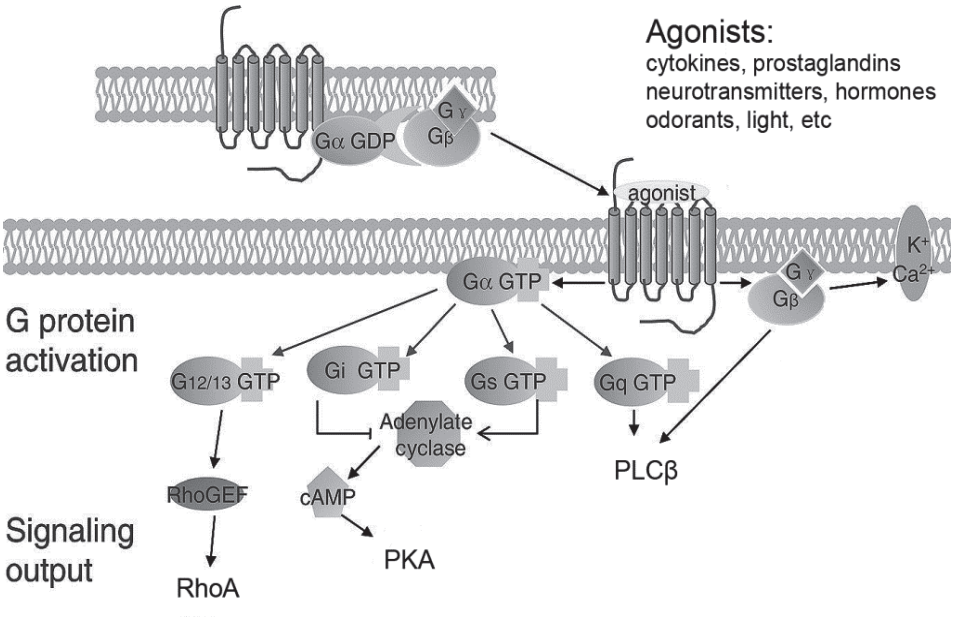


Figure 2. Simplified overview of the different pathways initiated by G protein signalling. The Gβγ subunit and different types of $G\alpha_i$ subunits are all involved in intracellular signalling via different mechanisms. Adapted by permission of Macmillan Publishers Ltd from Sun and Ye, 2012.

The Gβγ subunits are mainly associated with ion channel regulation (Figure 2), but also have some other functions (Sun and Ye, 2012). The alpha subunits have different activities depending on the subunit class: $G\alpha_i$ subunits inhibit adenylate cyclase activity, $G\alpha_s$ subunits enhance adenylate cyclase activity, $G\alpha_q$ subunits activate phospholipase C and $G\alpha_{12/13}$ subunits modulate the activity of Rho GTPases (Sun and Ye, 2012) (Figure 2).

The human genome (Lander *et al.*, 2001) encodes for nearly a thousand different G protein-coupled receptors (Vassilatis *et al.*, 2003; Wettschureck and Offermanns, 2005). There are sixteen genes encoding for Gα subunits (divided into four classes), five for Gβ and thirteen for Gγ present in the human genome (Lagman *et al.*, 2012). Our immune, nervous, and endocrine systems all rely greatly on GPCRs – for example, blood pressure is controlled by GPCR signalling, taste and smell are sensed via GPCRs, neurotransmitters function by binding to GPCRs, and immune cells are recruited to the sites of infection using GPCRs (Cabrera-Vera *et al.*, 2003; Wettschureck and Offermanns, 2005). Among others, the GPCRs include (Cabrera-Vera *et al.*, 2003; Wettschureck and Offermanns, 2005):

- Neurotransmitter receptors (e. g. for serotonin, dopamine, glutamate)
- Chemokine receptors (e. g. CXCR4, CCR3)
- Hormone receptors
- Opioid and cannabinoid receptors
- Adrenergic receptors
- Olfactory receptors
- Visual receptors (e. g. rhodopsin)
- Taste receptors
- Pheromone receptors
- Other receptors (e. g. Frizzled, Smoothened)

Malfunction of GPCR signalling can result in a wide spectrum of pathologies and thus ~40 % of all clinically approved drugs have been developed to target G protein-coupled receptors; some reports suggest the number to be even higher (Schöneberg *et al.*, 2004; Heng *et al.*, 2013; Stevens *et al.*, 2013; Wang and Lewis, 2013).

1.1.2 Modulation of G protein signalling

As described above, G protein signalling involves a fine-tuned balance between two states of the Gα subunit of the heterotrimeric G proteins. On the one hand, the GPCR exhibits a GEF activity leading to dissociation of the G protein. On the other hand, the Gα subunit itself contains a GTPase activity to reverse the dissociation. Various scaffolding proteins may tether signalling components near the GPCR (Ritter and Hall, 2009) and mechanisms are in place to ensure tight regulation and restriction of the GPCR-induced pathways.

- **Steric hindrance**. For example, phosducin and related proteins stabilize the betagamma dimer, but this interaction can also block its signalling activities (Lukov *et al.*, 2005; Partridge *et al.*, 2006). Guanine nucleotide dissociation inhibitors (GDI) can bind to the Gα subunit and block its activation (Siderovski and Willard, 2005).
- **G protein-coupled receptor kinases (GRK).** GRKs are kinases that are activated by binding to an activated GPCR, after which they can phosphorylate a variety of substrates and may even have functions other than kinase activity (Ribas *et al.*, 2007; Tesmer, 2009; Gurevich *et al.*,

2012; Watari *et al.*, 2014). Phosphorylation of the GPCR by GRKs results in β-arrestin binding to the receptor, which reduces G protein coupling and can even lead to internalization of the GPCR (Ribas *et al.*, 2007; Liggett, 2011; Watari *et al.*, 2014). Ubiquitination of β-arrestin is required for these processes as deubiquitinated β-arrestin can bind to GPCRs only transiently (Shenoy *et al.*, 2007). In plants, sugar has been also shown to induce GPCR phosphorylation and internalization (Phan *et al.*, 2013). However, even internalized GPCRs in endosomes retain some signalling activity (Irannejad *et al.*, 2013).

- **Regulators of G protein Signalling (RGS).** The main function of RGS proteins is generally associated with an RGS homology (RH) domain (also known as the RGS domain or 'RGS-box') that has a GTPase-activating protein (GAP) activity directed at the Gα subunits (Tesmer, 2009). Therefore the role of RGS is to accelerate the reassembly of the dissociated heterotrimeric G proteins, thus leading to inhibition of the G protein subunit-mediated signalling pathways (Ross and Wilkie, 2000). The RGS proteins have different expression patterns, different Gα subtype specificities and therefore often also different functions (Soundararajan *et al.*, 2008).
- **'Activators of G protein Signalling'.** These are a small group of proteins that can interact with both α and βγ subunits and can have GDI or GEF activity on the Gα subunit. Examples are the small G proteins Dexras1 and Rhes, as well as the GPSM1 and GPSM2 proteins (De Vries *et al.*, 2000; Cismowski *et al.*, 2001; Kerov *et al.*, 2005; Thapliyal *et al.*, 2008; Harrison and He, 2011).

The abovementioned regulatory mechanisms are further fine-tuned on both post-transcriptional and post-translational levels (Riddle *et al.*, 2005; Bansal *et al.*, 2007; Ribas *et al.*, 2007; Kach *et al.*, 2012). For example, RGS activity can be modulated by binding to the scaffolding protein spinophilin (Wang *et al.*, 2005; Bansal *et al.*, 2007; Fujii *et al.*, 2008; Ma *et al.*, 2012).

The role of RH domains in proteins other than RGS is less straightforward (Tesmer, 2009). For example, GRKs specialize in phosphorylation, although they have an RH domain, and only GRK2 exhibits a weak GAP activity (Carman *et al.*, 1999; Watari *et al.*, 2014). Other known RH domain proteins include axins as well as some Rho GEFs (e. g. ARHGEF1), sorting nexins, and others (Siderovski *et al.*, 1999; Ross and Wilkie, 2000; Siderovski and Willard, 2005; Bansal *et al.*, 2007; Tesmer, 2009). Most of the RH domain proteins are associated with protein-protein interactions related to G protein signalling pathways (Tesmer, 2009).

There are 21 known RGS proteins encoded in the human genome for which the main function is based on GAP activity (Siderovski and Willard, 2005). While classically the RH domain proteins have been divided into at least six subfamilies (up to nine), these RGS proteins themselves are divided between four of them depending on their domain structure and phylogenetic relationships

(Siderovski *et al.*, 1999; Ross and Wilkie, 2000; Siderovski and Willard, 2005; Bansal *et al.*, 2007; Tesmer, 2009):

- A/RZ - RGS 17, 19, 20
- B/R4 - RGS 1, 2, 3, 4, 5, 8, 13, 16, 18, 21
- C/R7 - RGS 6, 7, 9, 11
- D/R12 - RGS 10, 12, 14

The testis-specific protein RGS22 contains 3 RH domains and is not associated with any of the subfamilies (Hu *et al.*, 2008; Tesmer, 2009).

RGS proteins belonging to the A/RZ and B/R4 subfamilies have very simple domain structures, consisting of a single RGS domain accompanied by C- and N-terminal regions (Siderovski *et al.*, 1999; Ross and Wilkie, 2000; Siderovski and Willard, 2005; Bansal *et al.*, 2007; Tesmer, 2009). The N termini of RZ RGS are characterized by a poly-cysteine tail while the R4 RGS generally have an amphipathic helix in their N termini that enhances membrane association (Tesmer, 2009).

Both R7 and R12 RGS contain additional structural features besides the RH domain. One of the domains present in R7 RGS is highly similar to the G protein subunit Gγ and forms dimers with $G\beta_5$ (Posner *et al.*, 1999; Anderson *et al.*, 2009; Tesmer, 2009). RGS12 and RGS14 (R12 subfamily members) contain a C-terminal GoLoco domain with GDI activity (Kimple *et al.*, 2001), providing an additional opportunity to inhibit GPCR signalling. RGS10 is another RGS with a very simple structure, but its N-terminal end does not contain features characteristic to R4 and RZ RGS and it is in fact considered as an R12 RGS (Tesmer, 2009).

The presence of RGS homology domains is not specific to vertebrate or even metazoan genomes. A phylogenetic analysis of the GPCR pathway components in 75 genomes indicated that proteins with RH domains are present in nearly all species that encode for heterotrimeric G proteins (de Mendoza *et al.*, 2014), i. e. all metazoans, fungi (except for *Microsporidia*, a group of parasitic unicellular fungi), and a number of protists (de Mendoza *et al.*, 2014). However, in most lineages other than metazoans it is more common for the RH domain to be associated with the GPCRs themselves rather than as part of a separate protein (Phan *et al.*, 2013; de Mendoza *et al.*, 2014).

1.1.3 R4 RGS proteins and their functional implications

The largest subfamily of RGS proteins found in vertebrate species is the R4 RGS (Bansal *et al.*, 2007). In the human genome the genes for R4 RGS can be found on two chromosomes: *RGS3* is located on chromosome 9q32 while all the other R4 RGS (RGS1,2,4,5,8,13,16,18 and 21) are on chromosome 1q, suggesting that they originate from tandem duplications of a single gene (Snow *et al.*, 1998; Sierra *et al.*, 2002). It is therefore likely that the vertebrate ancestors had originally only one or two R4 RGS genes.

Although R4 RGS themselves have very simple structures, they can undergo post-translational modifications such as phosphorylation, palmitoylation,

arginylation and ubiquitination (Bansal *et al.*, 2007; Kach *et al.*, 2012). RGS4, 5 and 16 in particular are among the very few known targets for the N-end degradation pathway (Lee *et al.*, 2005). In this pathway proteolytic excision of a methionine leads to nitric oxide-dependent oxidation of the N-terminal cysteine, making it a suitable target for arginylation (Hu *et al.*, 2005). This chain of events creates a viable substrate for ubiquitination by specific ubiquitin ligases, e. g. ATE1, and subsequent degradation by the proteasome (Hu *et al.*, 2005; Lee *et al.*, 2005; Brower and Varshavsky, 2009). The same cysteine is also a target for palmitoylation so it seems the fate of these three R4 RGS proteins largely depends on post-translational modifications of the N-terminal residues (Druey *et al.*, 1999; Bastin *et al.*, 2012).

Many of the R4 RGS proteins have prominent roles in the mammalian (and likely other vertebrate) immune systems (Bansal *et al.*, 2007). However, they can also regulate the signalling of GPCRs not directly related to the immune system (Bansal *et al.*, 2007). Furthermore, functions of R4 RGS are not limited to GAP activity as other types of interactions have also been shown (Bansal *et al.*, 2007). For instance, RGS3 interacts with the $G\beta_1\gamma_2$ G protein subunit and RGS16 can inhibit PI3 kinase activity by binding to it (Shi *et al.*, 2001; Liang *et al.*, 2009). The next sections discuss in detail the different functions of R4 RGS in human and mouse as well as their expression patterns by focusing on the seven R4 RGS that in human are located at chromosome 1q25-q31.

RGS1 is one of the key RGS in the immune system and is expressed by all leukocyte subsets (Kveberg *et al.*, 2005; Bansal *et al.*, 2007). It restricts the migration of both B and T lymphocytes toward chemokine gradients (Moratz *et al.*, 2004a; Moratz *et al.*, 2004b; Agenès *et al.*, 2005; Bansal *et al.*, 2007), is highly expressed in germinal centres (Hong *et al.*, 1993), and is differentially regulated in a number of malignancies (Sethakorn and Dulin, 2013). It has been proposed that in resting B cells the *RGS1* locus is associated with centromeric heterochromatin (Jefferson *et al.*, 2010). In human and mouse agranular mononuclear myeloid cells activation with both viral and non-viral Toll-Like Receptor (TLR) agonists, or even with the pro-inflammatory cytokine tumour necrosis factor (TNF), potently induces RGS1 expression (Shi *et al.*, 2004; Barker *et al.*, 2005; Wen *et al.*, 2005; Gu *et al.*, 2009; Riekenberg *et al.*, 2009; Ottoboni *et al.*, 2013). In septic heart tissue RGS1 is induced in a MAPK (mitogen-activated protein kinase) - dependent manner (Panetta *et al.*, 1999). The expression of RGS1 is also upregulated in autoinflammatory diseases, e. g. spondylarthritis, psoriasis, lichen planus, and atopic dermatitis (Gu *et al.*, 2009; Li *et al.*, 2013; Rivas Bejarano and Valdecantos, 2013; Xie *et al.*, 2014). In chickens the upregulation of RGS1 is associated with resistance to infectious bursal disease virus (Koren *et al.*, 2008). Specific *RGS1* polymorphisms are considered a risk factor for autoimmune diseases such as celiac disease, type I diabetes and multiple sclerosis (MS) (Hunt *et al.*, 2008; Smyth *et al.*, 2008; Romanos *et al.*, 2009; International Multiple Sclerosis Genetics

Consortium, 2010; Lindén *et al.*, 2013). In MS RGS1 polymorphisms have been also associated with earlier onset of the disease and reduced attack severity, as well as increased amounts of the chemokine CXCL13 in cerebrospinal fluid (Johnson *et al.*, 2010; Lindén *et al.*, 2013; Mowry *et al.*, 2013).

RGS2 is expressed almost ubiquitously (Kehrl and Sinnarajah, 2002; Kveberg *et al.*, 2005; Bansal *et al.*, 2007) and is a key player in cardiovascular regulation, mostly by restriction of angiotensin II - induced signalling (Calò *et al.*, 2004; Hercule *et al.*, 2007). Like many other R4 RGS, it is dysregulated in various cancers (Sethakorn and Dulin, 2013). Lack of RGS2 impairs nitric oxide mediated vasodilatation in response to acetylcholine while vascular smooth muscle cells of *Rgs2* knockout (KO) mice proliferate more readily than their wild type (wt) counterparts (Osei-Owusu *et al.*, 2012; Momen *et al.*, 2014). In the retina, lack of RGS2 leads to decreased intraocular pressure due to increased actin filament assembly in the ciliary muscle cells and increased drainage of the aqueous humour (Inoue-Mochita *et al.*, 2009). Interferon gamma (IFNγ) inhibits cardiac hypertrophy at least partly by inducing RGS2; the overexpression of RGS2 in general can have results such as urinary incontinence due to smooth muscle inhibition (Jin *et al.*, 2005; Lin *et al.*, 2009; Nunn *et al.*, 2010; Chrobak *et al.*, 2013). RGS2 deficiency correlates with hypertension in both humans and rodents (Heximer *et al.*, 2003; Hercule *et al.*, 2007; da Costa Goncalves *et al.*, 2008; Yono *et al.*, 2010). Polymorphisms of the *RGS2* gene associated with hypertension (Zhang *et al.*, 2013), but also with metabolic syndrome (Freson *et al.*, 2007). RGS2 is normally highly expressed in adipocyte development (Cheng *et al.*, 2008) and *Rgs2* KO mice are much leaner than wt mice (Nunn *et al.*, 2011), while fat chickens have higher RGS2 expression than lean chickens (Sibut *et al.*, 2011). The impact on blood pressure of mice is at least partially elicited in kidneys – wt kidneys transplanted to KO mice normalize the blood pressure while KO kidneys in wt mice induce hypertension (Gurley *et al.*, 2010). Moreover, RGS2 is induced in kidney fibrosis to restrict angiotensin II-mediated disease progression (Jang *et al.*, 2014). RGS2 is also overexpressed in ischemic astrocytes and in cirrhotic cardiomyopathy (Endale *et al.*, 2010; Nunn *et al.*, 2010). In the nervous system, upregulation of RGS2 has been associated with epilepsy and is reversible by treatment with carbamazepine, an anticonvulsant and mood stabilizer (Almgren *et al.*, 2008). *Rgs2* KO mice are more anxious and have decreased male aggression (Oliveira-Dos-Santos *et al.*, 2000). In humans RGS2 and one of its polymorphisms are associated with anxiety disorders, suicidal behaviour, schizophrenia and parkinsonian side effects of antipsychotics used in treatment of schizophrenia (Leygraf *et al.*, 2006; Greenbaum *et al.*, 2007; Campbell *et al.*, 2008; Cui *et al.*, 2008; Amstadter *et al.*, 2009; Mouri *et al.*, 2010; Otowa *et al.*, 2011; Lifschytz *et al.*, 2012). A recent study linked RGS2 expression with Parkinson's Disease itself (Dusonchet *et al.*, 2014). Accordingly, RGS2 modulates the activity of LRRK2 kinase associated with Parkinson's Disease and can protect against neurotoxicity of its mutations (Dusonchet *et al.*, 2014).

In the immune system, engaging the T cell receptor generally upregulates RGS2 expression (Heximer *et al.*, 1997; Oliveira-Dos-Santos *et al.*, 2000). T cells from *Rgs2* KO mice have significantly impaired proliferation and IL-2 production (Oliveira-Dos-Santos *et al.*, 2000). The promonocytic cell line THP-1 (Abrink *et al.*, 1994) upregulates RGS2 in response to the yeast *Candida albicans* (Barker *et al.*, 2005). Stimulating murine Natural Killer cell (NK) receptors induces RGS2 regardless of whether the engaged receptor is inhibitory or activating (Kveberg *et al.*, 2005), suggesting a possible link with decreased migration toward the chemokine CXCL12 which is also observed in activation of both types of NK receptors (Inngjerdingen *et al.*, 2003; Kveberg *et al.*, 2005). *Rgs2* KO mice have impaired antiviral response to infection with lymphocytic choriomeningitis virus (Oliveira-Dos-Santos *et al.*, 2000). In the macrophages RGS2 is overexpressed by treatment with viral TLR agonists and IFNβ, but downregulated by treatment with non-viral TLR agonists (Giorelli *et al.*, 2002; Riekenberg *et al.*, 2009; Lee *et al.*, 2010). *Rgs2* KO macrophages treated with TLR2 agonists respond with stronger upregulation of the anti-inflammatory nitric oxide than wt macrophages (Byeon *et al.*, 2012). Therefore, even though RGS2 has been mainly studied in the context of cardiovascular regulation, it is apparent that it is also involved in immune responses, which seems to be common for the R4 RGS proteins.

RGS8 is mainly expressed in the central nervous system with the highest expression in cerebellar Purkinje cells (Gold *et al.*, 1997; Saitoh *et al.*, 2003; Saitoh and Odagiri, 2003). Although *Rgs8* KO mice do not substantially differ from wt mice (Kuwata *et al.*, 2007), Purkinje cells upregulate RGS8 expression during early development and downregulate it in the neurodegenerative disease spinocerebellar ataxia (Saitoh and Odagiri, 2003; Rodriguez-Lebron *et al.*, 2013). RGS8 in the pancreas is associated with β-cell expansion in diabetes models and early development (Villasenor *et al.*, 2010). In the immune system RGS8 is only expressed by the NK cells, in which it can be induced by treatment with IL-2 (Kveberg *et al.*, 2005). Engaging the murine activating NK receptor Ly49D reduces RGS8 expression up to two-fold (Kveberg *et al.*, 2005).

RGS13 is mainly associated with lymphatic tissues and lungs (Johnson and Druey, 2002; Shi *et al.*, 2004; Chng *et al.*, 2009). It is upregulated in adult T cell leukaemia/lymphoma (Pise-Masison *et al.*, 2009), but downregulated in B cell malignancies (Islam *et al.*, 2003; Cahir-McFarland *et al.*, 2004; Iwaki *et al.*, 2011). In germinal centres (GCs) RGS13 is induced to restrict B cell migration toward chemokine CXCL12 and CXCL13 gradients and overabundant expression of RGS13 in the GCs can result in autoreactive B cell development (Shi *et al.*, 2002; Hsu *et al.*, 2008; Wang *et al.*, 2013). In mast cells RGS13 inhibits both CXCL12-induced migration and Immunoglobulin E-mediated allergic responses (Bansal *et al.*, 2008a; Bansal *et al.*, 2008b). In rat brains RGS13 is mostly detected in the hippocampus and in Purkinje cells (Grafstein-Dunn *et al.*, 2001).

RGS16 is another R4 RGS that is often dysregulated in tumours (Sethakorn and Dulin, 2013). It was first cloned from the retina where it modulates phototransduction (Chen *et al.*, 1996; Snow *et al.*, 1998). It is also highly expressed in other tissues, most notably brain, liver, heart and lungs, but is also present in most leukocyte subsets (Chen *et al.*, 1997; Lippert *et al.*, 2003; Larminie *et al.*, 2004; Kveberg *et al.*, 2005; Bansal *et al.*, 2007). The best described role of RGS16 in the brain is in the suprachiasmatic nucleus in which it contributes to regulation of the circadian clock (Doi *et al.*, 2011; Hayasaka *et al.*, 2011). In the liver RGS16 is induced during fasting and downregulated during feeding (Huang *et al.*, 2006). Its physiological role in the liver is restriction of fatty acid metabolism – RGS16 KO mice have fatty livers while RGS16 overexpressing mice have many fatty acid metabolites in the blood (Pashkov *et al.*, 2011). Furthermore, RGS16 knockdown mice show significantly less anticipatory activity compared to wt when their feeding time approaches (Hayasaka *et al.*, 2011). In the pancreas RGS16 is associated with β-cell expansion in diabetes models and early development (Villasenor *et al.*, 2010). RGS16 expression in the heart is upregulated in response to cell activation in an interleukin (IL) 1β and MAPK-dependent manner, resulting in decreased signalling by endothelin, a molecule associated with regulating pain transmission and blood pressure (Panetta *et al.*, 1999; Patten *et al.*, 2002; Patten *et al.*, 2003; Kohan *et al.*, 2011; Smith *et al.*, 2014). Vascular smooth muscle cells upregulate RGS16 in response to stimulation with LPS or sphingosine-1-phosphate (Stuebe *et al.*, 2008; Hendriks-Balk *et al.*, 2009a; Hendriks-Balk *et al.*, 2009b). In various leukocyte subsets RGS16 is inducible by cell activation by both TLR agonists and different cytokines, although it is barely responsive in macrophages (Kveberg *et al.*, 2005; Riekenberg *et al.*, 2009; Xie *et al.*, 2010). In lymphocytes, RGS16 upregulation results in decreased cell migration (Beadling *et al.*, 1999; Lippert *et al.*, 2003; Estes *et al.*, 2004; Agenès *et al.*, 2005; Hsu *et al.*, 2008; Shankar *et al.*, 2012), which in germinal centres can lead to autoreactivity (Hsu *et al.*, 2008; Xie *et al.*, 2010; Ding *et al.*, 2013). In megakaryocytes RGS16 not only inhibits migration, but also controls cell maturation and inhibits platelet-activating factor-induced activation of MAP kinases (Zhang *et al.*, 1999; Berthebaud *et al.*, 2005). Despite the very weak presence of RGS16 in macrophages and unactivated peripheral blood mononuclear cells (PBMC) (Giorelli *et al.*, 2002; Kveberg *et al.*, 2005) its expression can be induced by cell activation (e. g. TLR agonists, heat shock, or pro-inflammatory mediators) in promonocytes (THP-1 cell line (Abrink *et al.*, 1994)), in PBMCs, and in dendritic cells (Fong *et al.*, 2000; Frevel *et al.*, 2003; Perrier *et al.*, 2004; Shi *et al.*, 2004; Barker *et al.*, 2005; Wong *et al.*, 2008; Riekenberg *et al.*, 2009; Timmusk *et al.*, 2009). In porcine PBMCs infected by circovirus RGS16 has been shown to interact with a virally encoded apoptosis-inducing protein (Timmusk *et al.*, 2009). Previously published RGS16 knockdown experiments using small interfering RNA (siRNA) have resulted in a pro-inflammatory profile, e. g. decreased expression of antiviral genes (a result

obtained from murine bone marrow-derived dendritic cells treated with the TLR2 agonist Pam3CysSK4 (Gat-Viks *et al.*, 2013) and enhanced activity of nuclear factor kappa-light-chain-enhancer of B cells (NF-κB) (result from a breast cancer cell line) (Vasilatos *et al.*, 2013).

RGS18 is most abundantly found in the bone marrow and expressed by hematopoietic stem cells, by megakaryocytes / platelets and by agranular mononuclear myeloid cells (Nagata *et al.*, 2001; Park *et al.*, 2001; Shi *et al.*, 2004; Iwai *et al.*, 2007), although weak RGS18 expression is also detectable in B cells, mast cells and NK cells, in which it is upregulated by activation of the inhibitory NK receptor Ly49A similar to RGS16 (Kveberg *et al.*, 2005). RGS18 is a negative regulator of osteoclast formation and is required for megakaryopoiesis (Iwai *et al.*, 2007; Louwette *et al.*, 2012). In response to cell activation dendritic cells downregulate the constitutively expressed RGS18 while megakaryocytes upregulate it (Shi *et al.*, 2004; Brass and Ma, 2012). Interestingly, RGS18 also controls the development of cilia by inner ear cells as demonstrated in zebrafish *Danio rerio* and the african clawed frog *Xenopus laevis* (Louwette *et al.*, 2012).

RGS21 is the most specialized and least studied among the R4 RGS proteins. RGS21 is specifically expressed in taste buds and lung airway epithelial cells where it restricts signalling via bitter taste receptors (Deshpande *et al.*, 2010; Cohen *et al.*, 2012; Kimple *et al.*, 2014).

In conclusion, previous works studying the R4 RGS that are located as a cluster on human chromosome 1q25-q31 seem to mainly associate them with immune functions, although they are also active elsewhere in the organism (reviewed by Bansal *et al.,* 2007). RGS2 and RGS21 are notable exceptions, as RGS2 is mostly associated with regulating blood pressure (Tsang *et al.*, 2010) and RGS21 is only found in a very specific subset of cells (Cohen *et al.*, 2012). As for the rest of the R4 RGS in the region – RGS1, RGS8, RGS13, RGS16, and RGS18 – many functions appear to be shared between them and thus due to the functional redundancy between them the loss of function in merely one of these regulatory proteins does not result in detrimental effects, although there can be changes in the phenotype in the context of different diseases.

1.1.4 The genomic R4 RGS region

Most of the human R4 RGS are located on chromosome 1q: *RGS4* and *5* on chromosome 1q23 are separated from *RGS8* and *16* by a distance of 20 Mb; *RGS8* and *16* themselves are separated from *RGS1, 2, 13, 18 a*nd *21* by a distance of 10 Mb. The genomic region containing RGS1, 2, 8, 13, 16, 18, 21 is defined here as the "R4 RGS region". The next sections focus on describing some of the features of the R4 RGS region that in human is in most part located at 1q25.3 – 1q31.3.

The R4 RGS region is located near two fragile sites in the genome where the chromosome easily breaks (at 1q25.3 and 1q32.1). In an evolutionary perspective, the R4 RGS region likely is – at least partly – "protected" from rearrangements as its sequence contains a highly conserved non-coding element together with its target developmental regulatory gene (that could be *RNF2* or *HMCN1*), forming a genomic regulatory block (Kikuta *et al.*, 2007). This is supported by numerous reports and case studies showing that large changes of this region are relatively uncommon, often deleterious, can result in different developmental disorders, and are commonly associated with tumours, e. g. leukaemia, breast cancer, and prostate cancer (Steffensen *et al.*, 1977; Chen *et al.*, 1989; Petković *et al.*, 1992; Baumgarten *et al.*, 1993; Jarvis *et al.*, 1999; Smedley *et al.*, 2000; Aboura *et al.*, 2002; Kawamata *et al.*, 2002; Hidalgo *et al.*, 2003; Schaid, 2004; Mark *et al.*, 2005; Nishimura *et al.*, 2005; Della Monica *et al.*, 2007; Wiechec *et al.*, 2008; Prontera *et al.*, 2011; Wiechec *et al.*, 2011; Cambosu *et al.*, 2013; Sethakorn and Dulin, 2013; Wiechec *et al.*, 2013).

One of the functions of the R4 RGS region is connected to G protein signalling regulation, which is not only due to the seven R4 RGS genes defining the region. Near *RGS8* and *RGS16* is another RH domain gene, RGS-like (*RGSL*) with mostly unknown function (Wiechec *et al.*, 2008). The mutations of *RGSL* are associated with breast cancer and the region in general contains many cancer susceptibility loci (Wiechec *et al.*, 2011). Another gene in the region, *PDC*, is also involved in GPCR signalling: it encodes for phosducin, which stabilizes the betagamma dimers of heterotrimeric G proteins. In cardiovascular system phosducin activity regulates blood pressure as its KO mice are clearly hypertensive (Beetz *et al.*, 2009). Another gene, *RGL1*, encodes for a Ral guanine nucleotide dissociation stimulator-like protein.

Although, as discussed above, the R4 RGS are deeply involved in immune responses, a systemic survey shows that there are other genes besides them in the R4 RGS region that are associated with immune function. Not far from the five R4 RGS genes at 1q31 is another cluster of duplicated genes, encoding for complement factor H and its homologs. These genes modulate the complement system and their mutations result in complement damaging self cells and causing systemic inflammation (Dragon-Durey *et al.*, 2009; Skerka *et al.*, 2013; Chaudhary *et al.*, 2014; Józsi and Meri, 2014). *CFHR4*, one of the genes in this cluster, is also upregulated in response to infection with

human herpesvirus 1 (Miyazaki *et al.*, 2011). The R4 RGS region includes susceptibility loci for both different types of tumours and a number of autoinflammatory and autoimmune disorders, e. g. psoriasis, celiac disease, type I diabetes, multiple sclerosis, lupus erythematosus and Sjögren's Syndrome (Johanneson *et al.*, 2002; Hunt *et al.*, 2008; Smyth *et al.*, 2008; Gu *et al.*, 2009; Romanos *et al.*, 2009; Schulte-Pelkum *et al.*, 2009; International Multiple Sclerosis Genetics Consortium, 2010; Villasenor *et al.*, 2010; Li *et al.*, 2013; Lindén *et al.*, 2013; Skerka *et al.*, 2013; Józsi and Meri, 2014; Kim-Howard *et al.*, 2014; Xie *et al.*, 2014). Changes in RGS13 and RGS16 expression are associated with the development of autoreactive germinal centres in BXD2 autoimmune mice (Hsu *et al.*, 2008; Xie *et al.*, 2010; Hwang *et al.*, 2013). The R4 RGS region also includes one gene encoding for a protein involved in antigen presentation (*MR1*) that is specifically used to present both microbially derived ligands and vitamin B metabolites to mucosal-associated invariant T cells in the gut, as only these cells express the invariant T cell receptor capable of MR1 recognition (Kjer-Nielsen *et al.*, 2012; Le Bourhis *et al.*, 2013; Meierovics *et al.*, 2013). Examples of other immune-related genes present in the region are *NCF2*, which encodes for the p67 subunit of the NADPH oxidase capable of producing superoxide (Kim-Howard *et al.*, 2014), and genes for proteins involved in prostaglandin biosynthesis, e. g. cytosolic phospholipase a2 (gene *PLA2G4A*) and cyclooxygenase 2 (gene *PTGS2*) (Kosaka *et al.*, 1994; Miyashita *et al.*, 1995; Järving *et al.*, 2004). Some of the genes in the region are bound by viral proteins during infection, such as *IVNS1ABP* (Influenza Virus Non-Structural protein 1A Binding protein) (Wolff *et al.*, 1998). Antiviral genes are also present, most notably, a gene for the well-characterized antiviral effector ribonuclease L (RNase L) (Sadler and Williams, 2008). RNA helicase A (gene *DHX9*) can function as a sensor of viral nucleic acids (Kim *et al.*, 2010; Zhang *et al.*, 2011) and R4 RGS themselves are also involved in antiviral responses (see Chapter 1.1.3). Most notably, RGS16 is induced by viruses, binds directly to a viral protein in porcine PBMCs and controls the induction of a number of immune response genes in response to Sendai virus infection (Timmusk *et al.*, 2009; Gat-Viks *et al.*, 2013), even though type I interferon can restrict the induction of RGS16 expression (Giorelli *et al.*, 2002).

Previous works have shown that many genes of the human R4 RGS region besides RGS have paralogs near the location of *RGS3* on chromosome 9q32-34, suggestive of ancestral duplication of the entire region (Fredriksson *et al.*, 2003; Hokamp *et al.*, 2003). This specific location on human chromosome 9 corresponds to one of the regions that have been identified as the "MHC paralogons" (Flajnik and Kasahara, 2010).

1.2 The R4 RGS region and other MHC-related genomic regions: a common history during vertebrate evolution

The immune system of vertebrates is based on a large number of genes with specialized functions (receptors, signalling factors, transcription factors, cytokines etc.). This high diversity was produced in the context of the evolution of genomes from the first metazoans to modern vertebrates, and this history is a frame that can be used to understand the selection pressures exerted by environment (including pathogens) on the genes of immunity.

1.2.1 The vertebrate whole genome duplication model

In 1970, years before the first genomes were fully sequenced, Susumu Ohno postulated in his classical book "Evolution by gene duplication" that allelic mutations of already existing gene loci cannot possibly account for major changes in evolution (Ohno, 1970). Although there are reports of completely new genes rising *de novo* (Knowles and McLysaght, 2009; Kaessmann, 2010; Zhao *et al.*, 2014), most of the major evolutionary events are largely driven by duplication of pre-existing genetic material and re-specialization of the duplicate genes (Ohno, 1970, 1999). Based on his observations that vertebrates have more DNA in general than invertebrates he took this idea one step further by proposing that vertebrate genomes may be the result of whole genome duplication (WGD) events, resulting in polyploidy and subsequent re-diploidization (Ohno, 1970).

Subsequent studies have found evidence of large duplicated segments in human (Schughart *et al.*, 1989; Lundin, 1993; Garcia-Fernández and Holland, 1994; Holland *et al.*, 1994; Hallböök *et al.*, 1998; Kasahara, 1998; Pébusque *et al.*, 1998; Abi-Rached *et al.*, 2002; Larhammar *et al.*, 2002; McLysaght *et al.*, 2002; Fredriksson *et al.*, 2003; Hokamp *et al.*, 2003; Lundin *et al.*, 2003; Danchin and Pontarotti, 2004a, 2004b; Olinski *et al.*, 2005, 2006; Nakatani *et al.*, 2007; Putnam *et al.*, 2008; Zucchetti *et al.*, 2009; Lagman *et al.*, 2012; Lagman *et al.*, 2013), most notably the four clusters of Hox genes on chromosomes 7, 17, 12, and 2, suggestive of two rounds of duplications (Schughart *et al.*, 1989; Garcia-Fernández and Holland, 1994; Holland *et al.*, 1994; Larhammar *et al.*, 2002). The Hox genes are essential for controlling the identity of segments forming early in development (Kmita and Duboule, 2003). The four clusters of Hox genes, as well as many other four-copy paralogous regions ('paralogons') found in the genomes of human, mouse, and most other vertebrate species, are strongly suggestive of two consecutive WGD events. Moreover, invertebrates such as *Drosophila melanogaster* have only a single Hox cluster (and a single copy of other duplicated regions) in their genome, providing further support to the WGD theory (Garcia-Fernández and Holland, 1994). The hypothesis of two rounds of consecutive WGD events during early vertebrate evolution is usually referred to as the "2R WGD hypothesis" or "Ohno's hypothesis", and in honor of Susumu Ohno

the paralogous genes derived from such events are referred to as 'ohnologs' in the literature.

When exactly during the early evolution of vertebrates did the genome duplications take place? The genomes of teleost fish suggest that duplications happened before the common ancestor of tetrapods and teleost fish; in fact teleost fish genomes have even been subjected to an additional round of WGD (Taylor *et al.*, 2001; Brunet *et al.*, 2006). It was long thought that the first duplication must have occurred in the common ancestor of jawed and jawless vertebrates and the second one right after their divergence (Sidow, 1996; Hallböök *et al.*, 1998; Ohno, 1999; Abi-Rached *et al.*, 2002; Escriva *et al.*, 2002; Furlong and Holland, 2002; Putnam *et al.*, 2008; Flajnik and Kasahara, 2010). However, recent sequencing of the lamprey genome has confirmed the alternative theory of both events happening in a common ancestor of all vertebrates (Kuraku *et al.*, 2009; Smith *et al.*, 2013).

Critics of Ohno's theory have proposed an alternative mechanism to explain the duplicated regions involving a series of tandem duplications followed by translocations (Hughes and Friedman, 2003; Ajmal *et al.*, 2014). Moreover, vertebrates do not actually have four copies of every single gene and usually paralogons have only two or three copies of most ohnologs (Holland, 2003). Phylogenetic trees made by the critics for many genes do not support the 2R WGD theory, providing tree topology and diverging times inconsistent with the hypothesis (Hughes, 1998, 1999; Ajmal *et al.*, 2014; Ambreen *et al.*, 2014). However, tandem duplications and WGD events are not mutually exclusive (Kasahara, 1997). It is clear that a lot of the duplicated genes are lost after duplications while others may be kept as is or reprogrammed with new functions (Figure 3) (Wolfe, 2001; Sémon and Wolfe, 2007; Gout *et al.*, 2009). Zebrafish for example has many genes that originate from the duplications but have been lost in mammals (Postlethwait, 2007). The discrepancies surrounding phylogenetic trees could be explained by the quick succession of WGD events (Gibson and Spring, 2000), and by the fact that molecular clock speed is not uniform; even different paralogs in the same species could evolve at different speeds (Flajnik and Kasahara, 2001; Lundin *et al.*, 2003; Flajnik and Kasahara, 2010). By constructing ancestral linkage groups and conducting thorough statistical analyses it has been demonstrated that the distribution of ohnologs in vertebrate genomes is not random and can be best explained by polyploidization events (Abi-Rached *et al.*, 2002; McLysaght *et al.*, 2002; Hokamp *et al.*, 2003; Danchin and Pontarotti, 2004b, 2004a; Nakatani *et al.*, 2007; Putnam *et al.*, 2008). Large scale chromosomal rearrangements following and/or preceding the duplications have been incorporated into the 2R WGD hypothesis (Lundin, 1993; Kasahara, 1999a; Kasahara *et al.*, 2007; Nakatani *et al.*, 2007; Sémon and Wolfe, 2007; Hufton *et al.*, 2008; Putnam *et al.*, 2008) and 2R WGD is not considered mutually exclusive with tandem duplications (Flajnik and Kasahara, 2010).

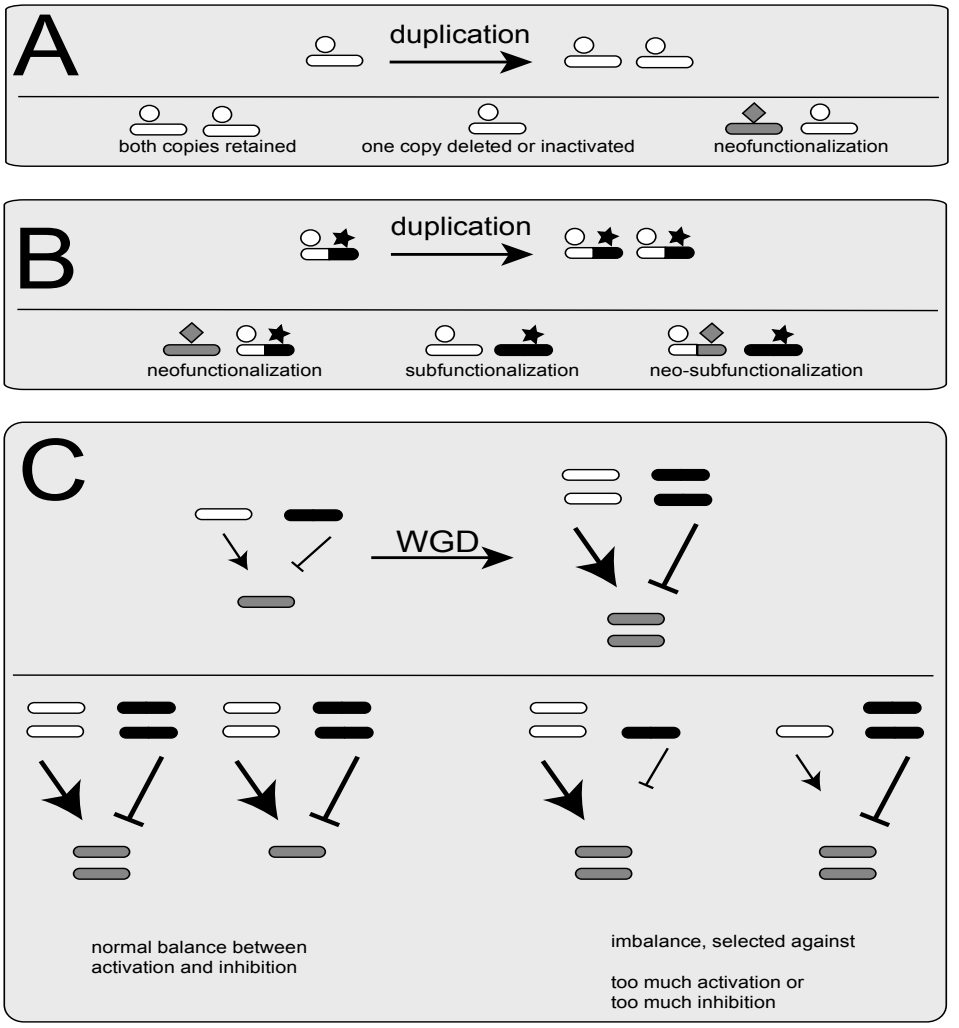


Figure 3. Fates of duplicated genes. (A) When a gene is duplicated, the duplicates can either be lost, retained, or neofunctionalized (a new function is then acquired by the gene). Examples of gene retention include dosage balanced genes and, alternatively, genes that are essential for the organism; in this last case, having two copies would provide a "back-up" when the essential gene is mutated. **(B)** For genes with more than one function, another possibility is subfunctionalisation in which the duplicate genes take over different functions from the parent gene(s). In addition, different mechanisms can be combined, such as neo-subfunctionalization in which the subfunctionalized duplicates may have additional new functions as well. **(C)** For dosage balanced genes the copy numbers need to be balanced with that of specific other genes. The most reliable way to retain the correct ratio for such genes is WGD, after which the duplicates are preferentially retained since gene loss could lead to imbalance. For example, an inhibitor and activator could be dosage balanced with each other if the fine-tuned regulation of their target is so important that its dysregulation can have deleterious effects. Figure adapted from Sémon and Wolfe, 2007.

Broadly speaking, the two main types of polyploidy are alloploidy and autoploidy. Alloploids form by hybridization of two different individuals, while autoploids result from duplications in a single genome (Comai, 2005). Somatic polyploidy is caused by cell fusion or cell division abnormalities (Gentric and Desdouets, 2014) and is commonly observed in cancers (Storchova and Pellman, 2004). Even certain types of healthy somatic cells can be polyploid, such as megakaryocytes in the immune system and hepatocytes in the liver (Winkelmann *et al.*, 1987; Biesterfeld *et al.*, 1994; Gentric and Desdouets, 2014). For an entire organism to become fully autopolyploid the polyploidization event must happen in germline cells and is mainly associated with either cell cycle / cell division abnormalities or with two haploid sperms fertilizing one haploid oocyte (Egozcue *et al.*, 2002; Rosenbusch, 2008). Most of the time polyploidy is lethal for the embryo, and triploidy accounts for approximately 15% of all spontaneous abortions caused by chromosomal abnormalities in human (Iliopoulos *et al.*, 2005; Wang *et al.*, 2014). The rare children born triploid or tetraploid are generally retarded, malformed, and die soon after birth (Fryns *et al.*, 1977; Shiono *et al.*, 1988; Niemann-Seyde *et al.*, 1993; Guc-Scekic *et al.*, 2002; Iliopoulos *et al.*, 2005; Stefanova *et al.*, 2010). However, there are reports of successfully thriving polyploid tetrapod species (a good example is *Xenopus laevis*) (Bisbee *et al.*, 1977; Wertheim *et al.*, 2013; Beçak, 2014), and polyploidy is even more commonly observed in plants (Adams and Wendel, 2005; Wertheim *et al.*, 2013).

Polyploidization can have a number of effects including problems in cell division and genomic instability (Comai, 2005; Hau *et al.*, 2006; Zheng *et al.*, 2012), and thus leads to extensive chromosomal rearrangements, e. g. translocations, as described for the 2R WGD events (Nakatani *et al.*, 2007; Putnam *et al.*, 2008), until the genome stabilizes, often by re-diploidization (Ohno, 1970; Comai, 2005). However, it should be noted that stress-induced DNA damage associated with chromosomal rearrangements is itself considered one of the driving forces leading to polyploidization (Chow and Poon, 2010; Pandit *et al.*, 2013), which supports the idea that the 2R WGD could have happened in rapid succession (Gibson and Spring, 2000). In any case, polyploidization initially leads to multiple fold expression of all genes in the genome. It has been suggested that during the period of post WGD rearrangements the constraints on gene retention may be relaxed, which explains the massive gene loss and pseudogenization of this period (Gout *et al.*, 2009).

Loss of gene duplicates can often be attributed simply to one gene being sufficient for the biological function and the duplicates becoming redundant. However, there are different possibilities that could lead to retention of duplicated genes (Figure 3).

- In some cases more copies of a gene can result in increased metabolic performance (Gout *et al.*, 2009) or a back-up solution in case the other copy has a deleterious mutation (Sémon and Wolfe, 2007).

- In other cases copies may be re-specialized (Wolfe, 2001; Sémon and Wolfe, 2007). In subspecialization two copies of a single ancestral gene inherit different sets of functions from the parent gene. In neospecialization both genes gain novel functionalities, and in sub-neospecialization these two are combined (Wolfe, 2001; Sémon and Wolfe, 2007) (Figure 3 A, B). Subspecialization also includes different tissue-specific expression patterns for duplicate genes, as can be seen for some of the R4 RGS (*e.g.*, RGS21 and RGS8) (von Buchholtz *et al.*, 2004; Kveberg *et al.*, 2005).
- Finally, the dosage balance hypothesis states that dosage balance constraints can keep different genes at a fixed copy number ratio (Sémon and Wolfe, 2007; Makino and McLysaght, 2010) (Figure 3C). Accordingly, loss of one of these genes would result in functional imbalance and deleterious effects (Figure 3C); to avoid this, such balanced genes would have to all either lose or gain a copy simultaneously (as happens during polyploidization). The eventually retained ohnologs are significantly enriched in dosage balanced genes, their copy number variations result in different diseases and are thus selected against (Makino and McLysaght, 2010; McLysaght *et al.*, 2014). Moreover, antecedents of such genes seem to be dosage balanced even in invertebrates, as in the tunicate *Ciona* it has been shown that genes that have not went through lineage specific expansions and duplications in this species are more likely to have retained ohnologs in vertebrates (Makino and McLysaght, 2010). Finally, the retained ohnologs in vertebrates are associated with copy number variation deserts in which nearby 'bystander genes' are also unlikely to have copy number variations (Makino *et al.*, 2013).

For the aforementioned chromosomal rearrangements double-stranded breaks in the chromosome are required. Such breaks often occur in the same spots repeatedly, and even in different lineages (Murphy *et al.*, 2005; Hinsch and Hannenhalli, 2006; Zhao and Bourque, 2009). The locations of such "fragile sites" or "breakpoints" tend to correlate with evolutionary "hotspots" between syntenic blocks (Ruiz-Herrera *et al.*, 2006; Gu *et al.*, 2008) that can be used to reconstruct ancestral genomes (Abi-Rached *et al.*, 2002; Danchin *et al.*, 2003; Vienne *et al.*, 2003a; Danchin *et al.*, 2004; Danchin and Pontarotti, 2004a, 2004b; Ruiz-Herrera *et al.*, 2006; Nakatani *et al.*, 2007; Putnam *et al.*, 2008; Alekseyev and Pevzner, 2009). These are called fragile sites because they are prone to breakage by stress-induced DNA damage, which has been associated with an ancient mechanism of increased mutagenesis and fast adaptation to handle sudden changes in environment (Galhardo *et al.*, 2007). Many of the hotspots are located subtelomerically, a distribution that has been demonstrated to be non-random using computational approaches (Mackiewicz *et al.*, 2013). Furthermore, the hotspots are associated with methylation deserts in the genome (Li *et al.*, 2012) and can be specified by *cis*-acting sequences that

interact with chromatin-modifying proteins such as PRDM9 in humans and mice (Baudat *et al.*, 2010). The evolutionarily conserved syntenic blocks located between the fragile sites are characterized by the presence of highly conserved noncoding *cis*-regulatory elements accompanied by both their target developmental genes and unrelated "bystander" genes (Kikuta *et al.*, 2007; Irimia *et al.*, 2012). The mechanism for double stranded break formation in fragile sites is generally associated with the replication fork stalling at DNA secondary structure, leading to late replication and replication errors (Glover *et al.*, 2005; Lee *et al.*, 2007; Gu *et al.*, 2008; Dillon *et al.*, 2010).

Chromosomal breaks can be also induced by transposon activity, which is mostly mediated by retroelements in vertebrate species (Gasior *et al.*, 2006; Pan and Zhang, 2007; Villarreal, 2009). In humans retrotransposon activity can cause at least fifty different diseases (Kaer and Speek, 2013).

Two main mechanisms have been proposed for the repair of breaks and are thus also responsible for the combination of such fragments from two different chromosomes. These mechanisms are non-homologous end-joining (NHEJ) and homologous recombination. Homologous recombination can also be non-allelic (NAHR, Non-Allelic Homologous Recombination) (Gu *et al.*, 2008).

- **NHEJ** is used to repair breaks in chromosome when the broken ends don't match: it bridges two "broken" chromosome fragments and ligates them together, leaving a "scar" (such as microdeletions) since it doesn't check for similarities with the original sequence (Lieber *et al.*, 2003; Gu *et al.*, 2008). Usually breakage tends to occur at genomic fragile sites characterized by large amounts of tandem repeats reducing DNA stability (Ruiz-Herrera *et al.*, 2006).
- **NAHR** can happen between two highly similar genomic regions (>92% identity) (Chen *et al.*, 2007) and may result in duplications, deletions, and insertions (Myers *et al.*, 2005; Gu *et al.*, 2008). The concerted evolution of paralogous genes is generally explained by gene conversion driven by NAHR (Liao, 1999). In human meiosis the minimal length of uninterrupted homology in two interacting sequences required for NAHR to happen is between 300 and 500 bp (Reiter *et al.*, 1998; Gu *et al.*, 2008), in mouse the length is >200 bp (Liskay *et al.*, 1987; Waldman and Liskay, 1988). With NAHJ a double-stranded break in DNA can be repaired according to homology with a similar sequence even if it is not the same gene from the sister chromatid (Helleday *et al.*, 2007). Recombination mechanisms for repair are utilized when NHEJ fails to repair the break (Shibata *et al.*, 2011), and the fragile sites are also considered recombination hotspots with many gene conversion events due to the high number of segmental duplications associated with them (Kauppi *et al.*, 2004; Cole *et al.*, 2014).

1.2.2 The Major Histocompatibility Complex

Functionally related genes often cluster together and form co-regulated units. While in prokaryotes such units can be transcribed as a single mRNA (operons) the eukaryotic equivalents are often kept together simply by evolutionary pressures (Makino and McLysaght, 2008; Osbourn and Field, 2009). However, it is not clear whether the conserved synteny is kept across wide branches of the tree of life by positive selection of clustering of genes with linked functions, or by a contingent absence of recombination. One of the best examples of such perplexing situation is the Major Histocompatibility Complex (MHC), a conserved genomic region packed with genes involved in immune responses (The MHC sequencing consortium, 1999). The human MHC encodes for over a hundred genes in a single region sized approximately four megabases (The MHC sequencing consortium, 1999) and includes many of the hallmark genes associated with adaptive immunity – such as those for the Class I and Class II molecules directly involved in antigen presentation and the *TAP* genes involved in antigen processing (The MHC sequencing consortium, 1999).

The MHC is divided into three main regions – Class I, Class II, and Class III, although more divisions have been proposed (Gruen and Weissman, 1997). Class I encodes for molecules that present epitopes derived from intracellular proteins to the circulating T cells and function as inhibitory ligands for the NK cells (Moretta *et al.*, 1996). Class II encodes for molecules that present epitopes from antigens digested by the cell. Both regions also contain additional genes, e. g. B30.2 domain proteins (Henry *et al.*, 1997) and genes involved in antigen processing such, e. g. for transporter associated with antigen processing (TAP) and the immunoproteasome (Tanaka and Kasahara, 1998; Krüger and Kloetzel, 2012). Class III is perhaps the most exciting as it has been suggested to be the most gene-dense region in the human genome with 72% of the region transcribed and 14% of the sequence coding (Xie *et al.*, 2003). The genes of Class III are mainly involved in innate immune responses, encoding for complement components and genes involved in inflammatory responses (Gruen and Weissman, 1997; The MHC sequencing consortium, 1999). Additionally, some authors have proposed to extend the MHC to include the adjacent large clusters of histones and olfactory G protein-coupled receptors (GPCRs), which are physically linked to the MHC in many species (Horton *et al.*, 2004). The GPCR cluster in particular exhibits strong linkage disequilibrium and polymorphic genes (Ehlers *et al.*, 2000). Genes in the MHC region itself generally also exhibit a lot of polymorphisms and strong linkage disequilibrium, leading to numerous possible alleles in the population. Many of the MHC alleles are associated with a variety of pathologies, including autoimmune disorders and susceptibility to infections (Blackwell *et al.*, 2009; Trowsdale and Knight, 2013).

While there are variations in the structure of the MHC in different vertebrates, the key elements of the complex (including typical class I and II,

as well as molecules involved in antigen processing) have been found in all jawed vertebrates, i. e. in all animals expressing immunoglobulins (Ig) and T cell receptor (Flajnik and Kasahara, 2001, 2010). In humans and many other mammalian species the Class III region is located between Class I and Class II (Kumanovics *et al.*, 2003; Xie *et al.*, 2003); one antigen-presenting non-classical Class I molecule (*MR1*) is located separately from the rest of Class I on a different chromosome (Chromosome 1 in humans and mice) (Tsukamoto *et al.*, 2013). However, the regional architecture of the MHC is different in other vertebrates, and it has been suggested that in mammals there has been a transposition of either Class I or Class II across the Class III region (Nonaka *et al.*, 1997; Kasahara, 1999a). Additionally, in mammals the genes for immunoproteasome and TAP are linked to Class II genes (Flajnik and Kasahara, 2010). In fact, Class I and Class II regions originally result from a tandem duplication of one single ancestral region (Kasahara, 1999a; Ohta *et al.*, 2000) as seen in shark (Ohta *et al.*, 2000; Ohta *et al.*, 2002). A good example of the similarity of Class I and Class II in mammals is the marsupial animal wallaby (*Macropus eugenii*), in which the classical Class I has been replaced by a second copy of the Class II region (including two copies of *TAP*). Non-classical Class I genes are still present; the two Class II regions are separated by Class III genes (Siddle *et al.*, 2011). Moreover, antigen-presenting Class II genes in this animal are separated from the rest of Class II. It should be noted, however, that this situation is not characteristic to all marsupials. In the opossum *Monodelphis domestica* the genes encoding Class I and Class II antigen presenting molecules are interspersed while Class I "framework genes" are located on the other side of Class III, as in human and mouse (Belov *et al.*, 2006). Phylogenetic analyses have demonstrated that the eutherian (most mammals), marsupial, and monotreme (platypus, echidna) Class I genes represent three evolutionary lines derived from the same ancestral loci and evolving in parallel (Miska *et al.*, 2002). On the other hand, the Class II genes of all three groups are more closely related, with monotremes positioning at the base of the mammalian clade (Belov *et al.*, 2003).

Chicken (*Gallus gallus*) possesses what is called a "minimal MHC" (all 'non-essential' genes have been lost, including immunoproteasome) (Kaufman *et al.*, 1999; Flajnik and Kasahara, 2001). It contains merely 19 genes total and is thus approximately twenty times more compact than the human MHC. As in other non-mammalian vertebrates, the Class III genes are located outside of Class I and II; Class I has a single dominantly expressed gene for antigen presentation (Walker *et al.*, 2011). Chicken genome in general is very compact and devoid of repeats; it has been even proposed that it may be similar to the ancestral configuration (Burt, 2002). In contrast to the general compactness of the chicken genome, the olfactory GPCR cluster linked to MHC is much larger in chicken that it is in human, containing twice the amount of receptors (Miller *et al.*, 2014). An array of scavenger receptors has been also reported in the near proximity (Miller *et al.*, 2014). Such 'minimal MHC' is also

characteristic to a number of other avian species, including quail, pheasants and the black grouse (Shiina *et al.*, 2004; Ye *et al.*, 2012; Strand *et al.*, 2013). The number of some genes in 'minimal MHC', such as NK receptors, has been expanded in quail (Shiina *et al.*, 2004; Rogers *et al.*, 2005). However, it should be noted that the contracted MHC is not characteristic to all birds. Passerine birds like warblers and zebra finch have a much larger MHC region than chicken (Westerdahl *et al.*, 2000; Balakrishnan *et al.*, 2010). On the other hand, in birds of prey (such as owls and hawks) the region is more reminiscent of the chicken MHC, although not as "minimal" (Alcaide *et al.*, 2007).

In the frog *Xenopus* Class I genes are linked with a single antigen presenting Class II gene as well as antigen processing genes - TAP and proteasome (Nonaka *et al.*, 1997; Ohta *et al.*, 2003). Amphibian non-classical Class I are expanded separately from the rest of the MHC (Flajnik and Kasahara, 2001; Zhao *et al.*, 2013). In the Class III region an immunoproteasome gene not present in the MHC itself in human has been found (*PSMB10*) (Ohta *et al.*, 2006). In contrast, salamanders (e. g. axolotl) have many Class I genes, but only a single Class II and thus rely mostly on Class I responses (Sammut *et al.*, 1999).

In teleost fish the Class I and Class II genes are generally not linked (Bingulac-Popovic *et al.*, 1997; Hansen *et al.*, 1999). In zebrafish (*Danio rerio*) the MHC is mostly fragmented although a core set of genes has been kept together (Bingulac-Popovic *et al.*, 1997; Sambrook *et al.*, 2005). In the rice fish medaka (*Oryzias latipes*) the Class II region is fragmented while Class I is mostly intact (Bannai and Nonaka, 2013). Remarkably, pipefish *(Syngnathus typhle) a*nd the cod (*Gadus morhua*) have lost Class II genes completely and instead have many genes for Class I and receptors involved in innate immunity (Star *et al.*, 2011; Haase *et al.*, 2013), a situation reminiscent of the amphibian axolotl.

While teleost fish have an MHC configuration very different from that in other vertebrates (separate Class I and Class II), it does not represent the MHC of ancestral species. This can be concluded from studies of nurse shark, a cartilagenous fish. In nurse shark the Class I, Class II and Class III genes are all linked, suggesting that the situation in fish is derived and not ancestral (Ohta *et al.*, 2000; Ohta *et al.*, 2002).

Jawless vertebrates (lamprey and hagfish) have evolved an alternative adaptive immunity based on leucine-rich repeats instead of immunoglobulins. These species also lack the Class I and Class II genes, although a common ancestor of both jawed and jawless vertebrates may have possessed the necessary components for both types of immunity (Boehm *et al.*, 2012).

1.2.3 Origin of the MHC and its paralogons

The MHC is one of the best studied regions in the vertebrate genomes and its conserved paralogons are among the most extensively studied paralogon sets (Kasahara, 1997; Flajnik and Kasahara, 2001, 2010). Although the entire region is named after and defined by the Class I and Class II genes involved in adaptive

immunity, vertebrate-like adaptive immunity does not exist in invertebrates (Abi-Rached *et al.*, 1999; Azumi *et al.*, 2003; Du Pasquier, 2004a; Kasahara *et al.*, 2004; Darbo *et al.*, 2008). The Class I and Class II genes are in fact found only in jawed vertebrates and are located in most cases on a single chromosome instead of four. However, the MHC genomic region also contains many genes involved in innate immunity, most notably complement components, genes involved in inflammatory responses (such as TNF), the antiviral tripartite motif (TRIM) / B30.2 domain proteins (van der aa *et al.*, 2009; Boudinot *et al.*, 2011) and even proteasome components (Horton *et al.*, 2004; Darbo *et al.*, 2008). Tracking the genes not involved in antigen presentation has led to the identification of three regions paralogous to the MHC (Kasahara *et al.*, 1996; Katsanis *et al.*, 1996; Kasahara, 1997; Hughes, 1998; Flajnik and Kasahara, 2001, 2010) that in human are at the following genomic locations (Kasahara, 1997; Flajnik and Kasahara, 2001, 2010) (Figure 4):

- 6p21-22 / 15q13-26
- 1p11-32 / 1q21-25
- 5q11-23 / 9p13-24 / 9q32-34
- 19p13

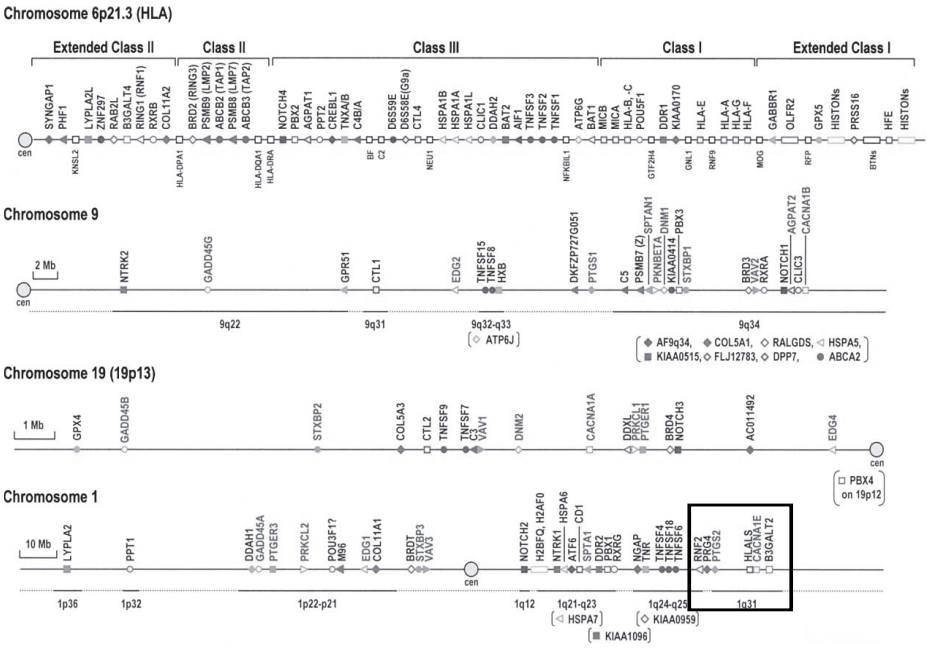


Figure 4. The MHC paralogons as described in 2001 by Flajnik and Kasahara. Black-bordered box marks the R4 RGS region. Reprinted from Flajnik and Kasahara, 2001, with permission from Elsevier.

Earlier works positioned the paralogons slighty differently and did not include the later discovered fragments on chromosomes 5 and 15 that result from translocations following the WGD (Flajnik and Kasahara, 2001) (Figure 4).

Tracking the MHC in ancestral species has been of particular interest with the following questions – is it possible to determine the ancestral genes that eventually gave rise to the Class I / Class II molecules and to the adaptive immunity as we know it? Was MHC formed from the reshuffling of existing proteins and their domains accompanied by other events during the period of extensive genomic rearrangements after the WGD, as suggested previously (Abi-Rached *et al.*, 1999)? Were other genes, especially those involved in antigen processing, important to structure the modern MHC? And in this case, what did the region from which the modern MHC eventually formed (i. e. the "proto-MHC") look like? The genomes of invertebrates may provide a valuable asset in studying the formation of the proto-MHC (and other paralogon sets) as they share a common ancestry with the vertebrates yet can have very differently evolved genomes (most notably, were not duplicated by the vertebrate-specific WGD) (Du Pasquier, 2001; Abi-Rached *et al.*, 2002; Azumi *et al.*, 2003; Danchin *et al.*, 2003; Danchin *et al.*, 2004; Danchin and Pontarotti, 2004a; Flajnik and Du Pasquier, 2004; Holland *et al.*, 2008). However, it should be also taken into account that at least some of the genes involved in adaptive (and maybe even innate) immunity may have a viral (or bacterial) origin and result from horizontal gene transfer events (Lander *et al.*, 2001; Villarreal, 2009, 2011). Accordingly, viruses that 'mimic' the proteins of our defence systems may actually represent the origin of the same defence systems (Villarreal, 2009, 2011), which is further supported by reports implicating horizontal transfer events as important players in eukaryote evolution (Lander *et al.*, 2001; Liu *et al.*, 2010; de la Casa-Esperón, 2012; Wallau *et al.*, 2012). For example, the *RAG* (Recombination-Activating Gene) genes important for the huge variability found in T Cell Receptor and the immoglobulins (Hsu, 2009) are of viral (or bacterial) origin, belong to the TnsB transposase family (Dreyfus *et al.*, 1999; Kapitonov and Jurka, 2005; Villarreal, 2009), and have independently entered the genomes of jawed vertebrates, amphioxus and echinoderms (Kapitonov and Jurka, 2005; Fugmann *et al.*, 2006; Villarreal, 2009; Flajnik and Kasahara, 2010; Litman *et al.*, 2010). Taking into account both observations such as this and the need to combat the many fast-evolving pathogens, it is thus apparent that the adaptive immune system evolved by selective pressures on ancestral invertebrate genes in concert with viral activity. In the process, viruses themselves have also evolved and even obtained immune-related genes from their host organisms, such as the chemokine receptor (GPCR) homologs found in large DNA viruses (Vischer *et al.*, 2006).

Several propositions have been made concerning the invertebrate origins of the adaptive immunity and antigen recognition, including various proteins with immunoglobulin domains (the immunoglobulin superfamily (IgSF)), as well as

molecular chaperones (such as heat shock proteins) and even proteasome along with tripartif motif (TRIM) proteins targeting viral proteins for degradation (Niedermann *et al.*, 1997; Marino *et al.*, 1998; Basu and Srivastava, 2000; Flajnik and Kasahara, 2001; Léonard *et al.*, 2001; Bartl *et al.*, 2003; Du Pasquier *et al.*, 2004; Litman *et al.*, 2010; Boudinot *et al.*, 2011). Huge variability in some immune-related genes (a characteristic of the antigen presenting molecules) has been found from invertebrates as well (i. e. The sea urchin 185/333 genes (Buckley and Smith, 2007), arthropod Dscam (Shi and Lee, 2012; Jin *et al.*, 2013) and mollusc FREPs (Léonard *et al.*, 2001; Moné *et al.*, 2010)), yet the genes do not appear to be directly related to the vertebrate antigen presenting ones.

Aside the concerns surrounding adaptive immunity there is another reason to study proto-MHC in the invertebrate species. As the vertebrate MHC paralogons contain a great amount of genes involved in innate immunity (The MHC sequencing consortium, 1999; Horton *et al.*, 2004; Shiina *et al.*, 2007), it is thought that tracking the proto-MHC will provide insights into both the innate immunity and the ancestral configuration of genomic region encoding for many genes involved in it (Flajnik and Kasahara, 2001; Levasseur and Pontarotti, 2010). Tracking the proto-MHC in animals that diverged from the ancestors of vertebrates in the distant past can thus potentially give insights on the formation of the immune system as such in the first place. Locations encoding for homologs of multiple MHC paralogon-linked genes have been already identified in a several invertebrate species using different markers to track the proto-MHC, i. e. genes encoding for proteasome subunits (Rast *et al.*, 2000; Abi-Rached *et al.*, 2002; Azumi *et al.*, 2003; Danchin *et al.*, 2003; Vienne *et al.*, 2003b; Danchin and Pontarotti, 2004a).

The closest species related to vertebrates are tunicates and lancelets (amphioxus) – these three groups form the single Chordates clade morphologically characterized by the presence of a notochord (Putnam *et al.*, 2008). The amphioxus proto-MHC and its proteins were initially characterized on a set of genome fragments which were later combined into a single chromosome (Holland *et al.*, 2001; Abi-Rached *et al.*, 2002; Vienne *et al.*, 2003b; Castro *et al.*, 2004). This was then used to predict an ancestral configuration of the region that gave rise to amphioxus proto-MHC and vertebrate MHC paralogons (Figure 5) (Danchin *et al.*, 2004; Danchin and Pontarotti, 2004a).

In tunicates the situation is more complicated (Azumi *et al.*, 2003; Kasahara *et al.*, 2004; Denoeud *et al.*, 2010). While tunicates are phylogenetically the closest relatives of vertebrates (Delsuc *et al.*, 2006; Vienne and Pontarotti, 2006; Putnam *et al.*, 2008), their genome has gone through extensive rearrangements and one species, *Oikopleura dioica*, has been even reported to have almost no apparent conserved synteny with vertebrates (Denoeud *et al.*, 2010). However, in the tunicate genus *Ciona* fragments of a putative proto-MHC have been identified (Kasahara *et al.*, 2004).

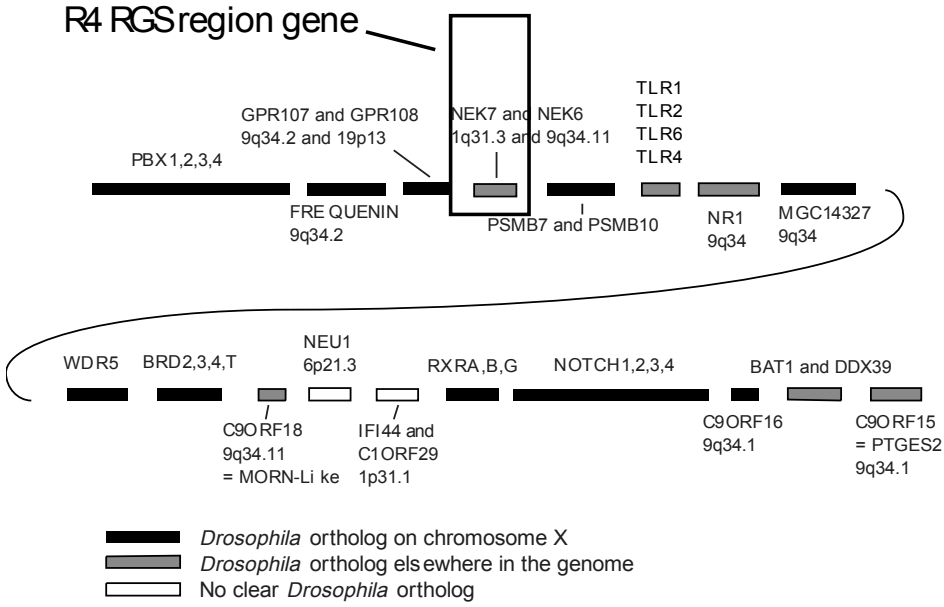


Figure 5. The "minimal Ur-euchordate proto-MHC" predicted from vertebrate and amphioxus MHC regions. Different colours mark correspondence with the fruit fly proto-MHC. Adapted from Danchin and Pontarotti, 2004. Figure reprinted with permission from Elsevier.

Other invertebrates where the proto-MHC has been described include the purple sea urchin *Strongylocentrotus purpuratus* and the fruit fly *Drosophila melanogaster* (Rast *et al.*, 2000; Trachtulec and Forejt, 2001; Danchin *et al.*, 2003; Danchin and Pontarotti, 2004a) (Figure 5). Very limited synteny has been also proposed for the nematode worm *Caenorhabditis elegans* (Trachtulec and Forejt, 2001). In the sea urchin a linkage has been found between homologs for two different complement genes (C2 and C3/4/5) and also between the orthologs of two classical MHC paralogon markers *PBX* and *NOTCH*, both of which exist in four copies in the human genome – one ohnolog on each of the MHC paralogons (Rast *et al.*, 2000). In *Drosophila* a linkage was found between the homologs for *RXR* (another MHC paralogon marker) and *NOTCH* (Trachtulec and Forejt, 2001), this was then used to deduce a 'minimal ancestral *RXR-NOTCH* region' (Danchin *et al.*, 2003). Later the analysis was further expanded to include the rest of the MHC region: the *Drosophila* proto-MHC was mapped to chromosome X and was found to also include the orthologs for proteasome and *PBX* genes (Danchin and Pontarotti, 2004a).

Since the animals described above already had a proto-MHC, the origins of the region could be sought from even more distant species and maybe even from phyla close to the hypothetical Ur-metazoan, common ancestor of all animals (Müller *et al.*, 2002; Schierwater and Desalle, 2007; Richter and King, 2013).

Toward the base of Metazoans, sequenced genomes are available for reprentatives of the non-bilaterian phyla *Cnidaria* (including jellyfish, corals, sea anemones and myxozoans) (Putnam *et al.*, 2007; Feng *et al.*, 2014), *Ctenophora* (sea combs) (Moroz *et al.*, 2014), *Porifera* (sponges) (Gauthier *et al.*, 2010; Srivastava *et al.*, 2010b), and *Placozoa* (Srivastava *et al.*, 2008). The true most basal group of metazoan is still a matter of debate although the general consensus seems to be that it is not *Cnidaria* (Dohrmann and Wörheide, 2013). Moreover, *Ctenophora* has been recently described as an animal group that has evolved alternative signalling systems in parallel to the rest of the metazoans and thus likely differs greatly from the hypothetical Ur-metazoan (Moroz *et al.*, 2014). It is therefore more likely that the most basal-like metazoan could be represented by a species belonging to either *Porifera* or *Placozoa*. In particular, the placozoan *Trichoplax adhaerens* (or at least the sequenced individual, they are genetically widely different and even their mitochondrial genomes can vary greatly in both structure and size (Osigus *et al.*, 2013)) has been reported to have very well conserved synteny with vertebrates (Srivastava *et al.*, 2008; Simakov *et al.*, 2013).

1.2.4 Other genomic complexes related to the MHC

There are many other sets of ohnologs conserved in vertebrates in addition to those located on the MHC and Hox paralogon sets, likely representing remnants of the 2R WGD (McLysaght *et al.*, 2002; Nakatani *et al.*, 2007; Putnam *et al.*, 2008). Among those, two sets have been linked to the MHC, and appear to share a common ancestral region (Flajnik *et al.*, 2012).

(1) The genomic locations 14q and 12p, as well as 19q all belong to the same set of paralogons (Hallböök *et al.*, 1998; Olinski *et al.*, 2005, 2006). The so-called "neurotrophin paralogons" (NT paralogons) were initially identified by four ohnologs from the same neurotrophin family: the Nerve Growth Factor (NGF) on the chromosome 1 NT paralogon in human; the Brain-Derived Neurotrophic Factor (BDNF) on the chromosome 11 NT paralogon in human, and two more neurotrophins mapping to the other two NT paralogons (NTF3 and NTF4) (Hallböök *et al.*, 1998). Analysis of insulin genes in the human genome was able to further confirm the paralogon set (Olinski *et al.*, 2005, 2006). The NT paralogons are even more fragmented than the MHC ones and have been identified at the following genomic locations (Olinski *et al.*, 2005, 2006; Flajnik and Kasahara, 2010; Flajnik *et al.*, 2012):

- 1p13 / 1q32-44 / 2p12-23 / 20p11-12
- 11p12-15 / 11q12-13 / 11q23-24
- 12p11-13 / 12q22-24 / 14q11-32
- 19q13

(2) Louis Du Pasquier and colleagues were looking for MHC-related regions in tunicates from the genus *Ciona* (Azumi *et al.*, 2003; Du Pasquier, 2004b, 2004a; Du Pasquier *et al.*, 2004; Kasahara *et al.*, 2004; Zucchetti *et al.*, 2009). Proteins with IgSF domains were considered likely candidates to provide new

insights about the origin of adaptive immunity (Du Pasquier *et al.*, 2004; Zucchetti *et al.*, 2009). *Ciona* was found to encode for a number of adhesion molecules that have a characteristic IgSF domain, and for which the vertebrate homologs are virus receptors (Du Pasquier, 2004b, 2004a; Du Pasquier *et al.*, 2004; Kasahara *et al.*, 2004; Zucchetti *et al.*, 2009). Moreover, the IgSF domains in these proteins belong to a subset typically associated with adaptive immunity (V and C1 domains) (Du Pasquier, 2004a; Du Pasquier *et al.*, 2004; Kasahara *et al.*, 2004). The vertebrate homologs of these *Ciona* genes belong to two main families (the junctional adhesion molecule (JAM) / cortical thymocyte marker of *Xenopus* (CTX) family and the nectin family) and were found to cluster to four specific regions encoding (among others) important immune-related genes (Du Pasquier *et al.*, 2004; Kasahara *et al.*, 2004; Zucchetti *et al.*, 2009). These observations led to the identification of another conserved set of MHC-related paralogons. The JAM-NECTIN (JN) paralogons are located in the human genome as follows (Zucchetti *et al.*, 2009; Flajnik and Kasahara, 2010; Flajnik *et al.*, 2012):

- 1q23-25
- 3q13-22
- 11q23-24
- 21q21-22 / 19q13

A translocated fragment from the chromosome 3 JN paralogon to chromosome X has been also identified (Zucchetti *et al.*, 2009).

Careful examination reveals many connections between these three sets of human MHC-related paralogons (MHC, NT, and JN). The MHC paralogon on chromosome 1 overlaps with the NT paralogon at 1p13, and with one of the JN paralogons at 1q23-25. On 11q23-24 and 19q13 the NT and JN paralogon fragments correspond fairly well to each other. Furthermore, all of these three paralogon sets contain genes for critical components of adaptive immunity. The most prominent of such genes encoded in the MHC paralogons are likely the Class I and Class II genes (both classical and non-classical). Class I and Class II molecules present antigens to the T cell receptor (Smith-Garvin *et al.*, 2009), which is partly (alpha chain) encoded at the 'ambiguous' NT/JN region of 14q11. Class I is also a ligand for the NK cell receptors (Sawicki *et al.*, 2001) that in human are encoded on the NT paralogon at chromosome 12p. The transcription factor AIRE (**A**uto**I**mmune **RE**gulator), which is largely responsible for the negative selection of new T cells in the thymus to avoid self-reactivity (Laan and Peterson, 2013; Kisand *et al.*, 2014), is encoded at the JN paralogon on human chromosome 21q (Nagamine *et al.*, 1997). CD79A (MB-1) of the B cell receptor complex (Herren and Burrows, 2002) is encoded at 19q13 (NT/JN) and the heavy chain locus of antibodies is located at chromosome 14 (NT) (Tomlinson *et al.*, 1995). Finally, members of the B7 family, proteins containing V and C1 IgSF domains and able to modulate T cell activation, are found across all 3 paralogon sets: MHC, NT, and JN (Flajnik *et al.*, 2012).

In addition to genes involved in adaptive immunity, the three sets of paralogons also have a puzzling number of genes pivotal for the innate responses. A good example is the interferon machinery, one of the primary innate mechanisms for combating viruses (Lin and Young, 2014). In the human genome, the cluster of type I IFN genes is found on chromosome 9p21 (MHC). The gene for IFNγ (type III IFN) is found on chromosome 12q15 (24 Mb from the NT paralogon) and the type 3 IFN genes are located on 19q13 (a region corresponding to both NT and JN). Chromosome 21q22 contains the type I IFN receptor genes *IFNAR1* and *IFNAR2*, as well as *IFNGR2* (one of type II IFN receptor genes) and *IL10RB* (one of type III IFN receptor genes). The other receptor genes for type II and type III IFN are located on chromosome 6 and chromosome 1, respectively (although not on the MHC paralogons). Many of the best characterized IFN-responsive genes are similarly located in these regions - for example, in human the 2'-5' oligoadenylate synthetase genes that can be found across metazoans (Päri *et al.*, 2014) are located on chromosome 12q24 (NT), protein kinase R (Pfaller *et al.*, 2011) is encoded on chromosome 2p22 (NT), and the gene for MxA (Horisberger, 1995) is located on chromosome 21q22 (JN).

Taken together, it appears that many important immune functions are scattered across three specific sets of paralogons, a notion that could be further strengthened by statistical tests addressing the possibility of such a skewed distribution being random.

How to explain that one genomic region apparently corresponds to two paralogons, with some genes belonging to, i.e, MHC-related sets of ohnologs and others to NT-related sets of ohnologs? Several lines of evidence have led researchers to hypothesize that all three sets of paralogons may in fact be derived from a single ancestral genomic region, which likely went through "local" duplication and rearrangement events prior to the WGD itself (Kasahara, 1997, 1999b; Flajnik and Kasahara, 2001; Olinski *et al.*, 2005, 2006) (Figure 6). In fact, there are some families such as the B7 family with members spread across two or more of the three sets (Hansen *et al.*, 2009; Flajnik *et al.*, 2012), which could be explained by ancestral duplications of the gene into what later became the origin sites for different paralogons.

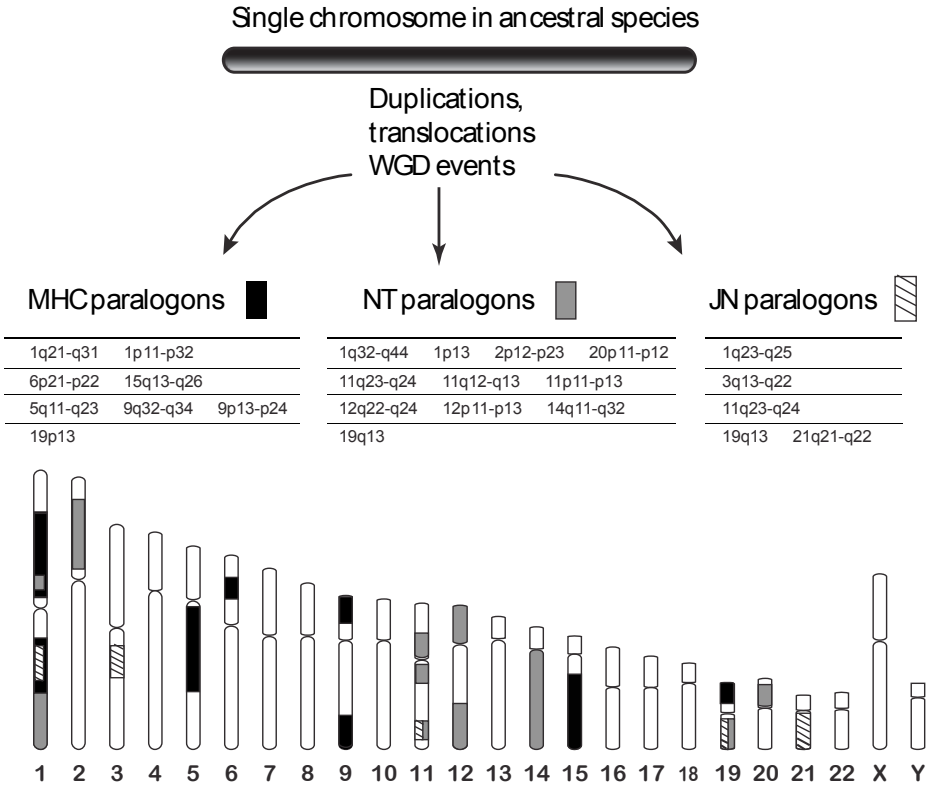


Figure 6. The three sets of MHC-related paralogons. Two rounds of whole genome duplications along with chromosomal rearrangements eventually led to the distribution of paralogon fragments seen in human genome. Modified from Publication III, Figure 1.

Kasahara suggested already in the year 1999 that the MHC and NT might have a common ancestor (Kasahara, 1999b). More recent reviews also tackle this idea and discuss the uncanny amount of similarities between the nervous system and the immune system (Okada and asai, 2008; Zhu *et al.*, 2008; Kioussis and Pachnis, 2009). A set of indirect evidences and observations are consistent with this hypothesis: for example, chickens actually have NK receptors in their minimal MHC close to the Class I and Class II genes, while the similar mammalian NK receptors are found on an NT paralogon (Kaufman *et al.*, 1999); the marsupial MHC includes a homolog of the gene *OSCAR* encoded on 19q13 where JN and NT paralogons overlap in human (Belov *et al.*, 2006; Siddle *et al.*, 2011); overall, genes with specific functions of both adaptive and innate immunity can be found across all three sets as described above, which would be consistent with this hypothesis. The final conclusive piece of evidence, however, is likely to come from the study of ancestral and invertebrate genomes near the base of metazoans. Finding genes from MHC, NT, and/or JN paralogons clustered together in an ancestral animal would greatly strengthen the hypothetical connections between the three regions.

1.2.5 Interest of R4 RGS region genes in relation to three sets of paralogons involved in immunity (MHC, NT, and JN)

As mentioned above, Hokamp and colleagues suggested that the region containing R4 RGS genes on human chromosome 1q belongs to one of the largest paralogous blocks in the human genome with its paralogy region on chromosome 9q32-34 (Hokamp *et al.*, 2003). In fact this region on human chromosome 9 corresponds to one of the MHC paralogons and contains the only R4 RGS gene not located on chromosome 1 (*RGS3*) (Figure 7). The current names of genes indicated in the work of Hokamp and colleagues include the R4 RGS genes themselves, *PTGS2, LAMC2, PDC, aRPC5, FAM129A, GLT25D2 a*nd *NPHS2* (Hokamp *et al.*, 2003).

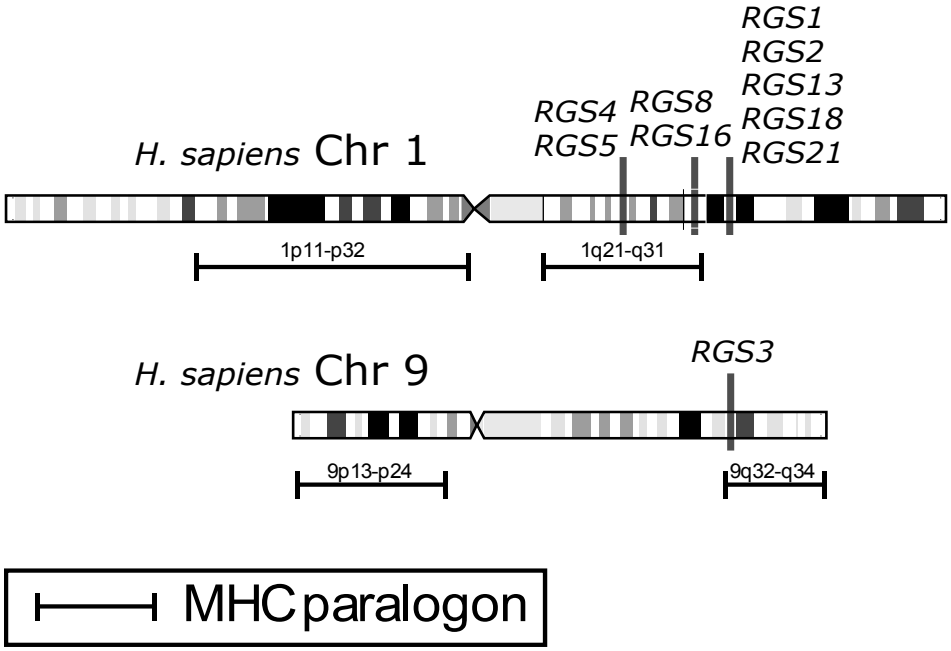


Figure 7. R4 RGS region in the human genome map to MHC paralogons on chromosomes 1 and 9.

Other studies have also identified at least some of the R4 RGS region genes as ohnologs with copies on the MHC paralogons (Katsanis *et al.*, 1996; Kasahara, 1999b; Flajnik and Kasahara, 2001; Abi-Rached *et al.*, 2002; Vienne *et al.*, 2003b; Danchin *et al.*, 2004) (Figure 4). The calcium channel *CACNA1E* in particular has been suggested as a marker for tracking the MHC paralogons in both vertebrates and amphioxus (Katsanis *et al.*, 1996; Kasahara, 1999b; Flajnik and Kasahara, 2001; Abi-Rached *et al.*, 2002; Vienne *et al.*, 2003b; Danchin *et al.*, 2004). *PTGS2* (the gene for cyclooxygenase 2) is another gene mentioned

in previous works (Katsanis *et al.*, 1996; Kasahara, 1999b; Flajnik and Kasahara, 2001) and has been found present in invertebrate genomes (Järving *et al.*, 2004; Hansen *et al.*, 2014). *RGL1* and its homologs (Kasahara, 1999a; Abi-Rached *et al.*, 2002) as well as *B3GALT2* have been described on MHC paralogons (Kasahara, 1999a; Flajnik and Kasahara, 2001). Flajnik and Kasahara also identified *RNF2*, *PRG4* and *MR1* as MHC paralogon markers (Flajnik and Kasahara, 2001). *NEK7* is another gene found in both amphioxus proto-MHC and on vertebrate MHC paralogons (Abi-Rached *et al.*, 2002; Vienne *et al.*, 2003b; Danchin *et al.*, 2004; Danchin and Pontarotti, 2004a). Intriguingly, the genomic region on chromosome 1q31 including most of the R4 RGS genes themselves, as well as previously described markers *PRG4, PDC, PTGS2, B3GALT2 a*nd *NEK7* has not been considered as part of the MHC paralogon on human chromosome 1 in more recent reviews (Flajnik and Kasahara, 2010; Flajnik *et al.*, 2012) and most of the studies mentioned above were conducted more than ten years ago.

The R4 RGS region in the human genome, as defined in this thesis, is located on chromosome 1q25-q31. This location seems to be a "hotspot" connecting all three MHC-related paralogon sets discussed above – one of the MHC paralogons is found at chromosome 1q21-q25, a JN paralogon is at chromosome 1q23-q25, and NT paralogon at chromosome 1q32-44. Kasahara pointed out already in 1999 that *CACNA1E* has a homolog on the chromosome 12 NT paralogon (Kasahara, 1999b). When mapping GPCR paralogy sets in 2003, Lundin and colleagues suggested the human chromosome 1 segment containing the R4 RGS region to have paralogy groups on chromosomes 5p-q21, 6p21-25, 9, 15q11-26, 19p, 2p22-25, 11p, 11q13-23, 12, 14q, 19q, 20p (Fredriksson *et al.*, 2003). This selection includes most of the regions associated with the MHC and NT paralogon sets described above. Even the R4 RGS themselves have functions in both immunity and the nervous system, exemplified by their polymorphisms being associated with both CNS and immune disorders as described in Chapter 1.1. It is therefore likely that the R4 RGS region and its genomic location between different MHC-related paralogon sets could be a remnant of the proposed ancestral linkage between different MHC-related paralogons that was described above.

2. AIM OF THE STUDY

This study aims to describe the *RGS16* gene and its genomic context from both functional aspects related to immunity and the evolutionary perspective. The following tasks were undertaken:

1) To investigate the impact of RGS16 on monocyte pro-inflammatory responses (**Publication I**)

2) To track the genomic neighbourhood of *RGS16* throughout evolution and search for its origin in ancestral species (**Publication II and III**).

3) To test the hypothesis of an evolutionarily conserved connection between the genomic neighbourhood of *RGS16* and other known immune-related regions (**Publication II and III**)

3. MATERIALS AND METHODS

Detailed descriptions of materials and methods are provided in the corresponding publications. An overview of the methods is as follows:

- Animal experiments (**Publications I and II**)
- Isolation of organs and cell populations (**Publications I and II**)
- Cell culture and transfections (**Publication I**)
- RNAi (**Publication I**)
- Enzyme-linked immunosorbent assays (ELISAs) (**Publication I**)
- RNA extraction (**Publications I and II**)
- cDNA synthesis (**Publications I and II**)
- Quantitative PCR (qPCR) (**Publications I and II**)
- *in silico* analysis of publicly available genomes (**Publications II and III**)
- Gene ontology enrichment analysis (**Publication III**)
- Phylogeny construction (**Publications II and III**)
- Statistical analyses (**Publications I, II and III**)

4. RESULTS AND DISCUSSION

4.1 RGS16 regulates immune responses (Publications I, II)

4.1.1 RGS16 in inflammatory responses (Publication I)

The importance of R4 RGS proteins in the immune responses of mammalian myeloid cells is a complex issue. RGS16 is not detectable in rat macrophages (Kveberg *et al.*, 2005) and is barely responsive to LPS in bone-marrow derived macrophages (Riekenberg *et al.*, 2009), while RGS1, RGS2, RGS13 and RGS18 are all downregulated during differentiation of monocytes into macrophages (Saeed *et al.*, 2014). RGS8 and RGS21 are not even expressed by agranular mononuclear myeloid cells (von Buchholtz *et al.*, 2004; Kveberg *et al.*, 2005). In contrast, different R4 RGS (e. g. RGS16) are expressed in dendritic cells which derive from the same monocyte progenitors as macrophages (described in Chapter 1.1.3). The expression of RGS16 in particular can inhibit MAPK activity, NF-κB, and PI3 kinase, and is induced during immune responses (Panetta *et al.*, 1999; Zhang *et al.*, 1999; Fong *et al.*, 2000; Patten *et al.*, 2002; Frevel *et al.*, 2003; Patten *et al.*, 2003; Perrier *et al.*, 2004; Shi *et al.*, 2004; Barker *et al.*, 2005; Berthebaud *et al.*, 2005; Hendriks-Balk *et al.*, 2009b; Liang *et al.*, 2009; Riekenberg *et al.*, 2009; Timmusk *et al.*, 2009; Xie *et al.*, 2010; Gat-Viks *et al.*, 2013; Vasilatos *et al.*, 2013).

We hypothesized that RGS16 induction may restrict the production of pro-inflammatory mediators. To better elucidate the roles that RGS16 may play in activation and response of monocytic cells, we activated THP-1 monocytic cells with the synthetic TLR2 agonist $Pam3CysSK_4$ (Pam3) in these cells after transfection with either RGS16 expression vector or RGS16-specific small interfering RNA (siRNA). The expression of pro-inflammatory cytokines (IL-1β, IL-6, IL-8 and TNF) and RGS16 itself was compared to the one in control transfected cells. Our results clearly demonstrated at the protein level that the overexpression of RGS16 leads to a significant decrease in activation-induced pro-inflammatory cytokine production, and that its downregulation increases the pro-inflammatory response (Publication I, Figure 1 and 2). However, mRNA level expression of these cytokines was apparently not affected (Publication I, Supplementary Figure S1), suggesting that RGS16 elicits its inhibitory function either post-transcriptionally or post-translationally. RGS16 siRNA-mediated enhancement of the activation-induced late expression of anti-inflammatory IL-10 was also observed (Publication I, Figure 2). In general, this cytokine is induced in inflammatory immune responses as part of a negative feedback loop (Lampropoulou *et al.*, 2010; Saraiva and O'Garra, 2010). Therefore RGS16 inhibits not only pro-inflammatory cytokine expression but also its restriction mechanisms, suggesting a complex loop of interactions involving this protein (Publication I, Figure 4).

To gain additional insight for RGS16 function in agranular mononuclear myeloid cells we compared wt mice with *Rgs16* KO mice obtained from Dr. Kirk Druey, Bethesda, Maryland. Different macrophage populations were isolated and tested for cytokine expression upon activation with LPS or Pam3CysSK4 (Publication I, Figure 3, Supplementary Data). No significant differences were found in alveolar or in peritoneal macrophages (Publication I, Supplementary Data); this is further supported by previous works showing that macrophages do not express significant amounts of RGS16 and respond very weakly to activation with LPS (Kveberg *et al.*, 2005; Riekenberg *et al.*, 2009). However, bone marrow-derived cells from *Rgs16* KO mice were found to express higher levels of the cytokine CXCL1 than wt (Publication I, Figure 3). These cells were derived from the bone marrow progenitors *in vitro*, which might explain the observed difference from primary macrophages. Moreover, RGS16 is responsive to cell activation in other agranular mononuclear myeloid cells such as monocytes and dendritic cells (Giorelli *et al.*, 2002; Shi *et al.*, 2004; Barker *et al.*, 2005; Wong *et al.*, 2008; Timmusk *et al.*, 2009; Gat-Viks *et al.*, 2013). In agreement with the current result, an enhancement of CXCL1 expression has also been observed for LPS-treated spleen cells from *Rgs16* KO mice as compared to similarly treated wt spleen cells (Pahtma, 2011). The observations in primary cells suggest that the results of the current study concerning cytokine production in activated THP-1 cells might have a general relevance. Additional experiments will be needed to gain statistical support for the observations made in *Rgs16* KO mice.

What could be the mechanism through which RGS16 affects the inflammatory response in such a way? It is known that GPCR signalling can activate MAP kinases via $G\alpha_i$ and $G\alpha_q$ activation (Blaukat *et al.*, 2000; Fukuhara *et al.*, 2000; Arai *et al.*, 2003). RGS16 can target and inhibit signalling of both of these types of Gα subunits (Druey *et al.*, 1999; Bansal *et al.*, 2007), and has been shown to inhibit p38 MAPK activity (Zhang *et al.*, 1999; Liu *et al.*, 2005; Vasilatos *et al.*, 2013). The MAP kinases have been long recognized as key players in pro-inflammatory responses (including in monocytes) (Obata *et al.*, 2000; Guha and Mackman, 2001), identifying the RGS16-mediated restriction of p38 MAPK activity as a likely candidate for the mechanism responsible for our observations. While the receptors for LPS and Pam3 are not GPCRs themselves (Takeuchi *et al.*, 1999; Beutler, 2005; Lattin *et al.*, 2007; Kayagaki *et al.*, 2013), signalling via them leads to secretion of inflammatory mediators out of which some are GPCR ligands.

RGS16 also directly binds and inhibits the PI3 Kinase (Liang *et al.*, 2009), which can differentially modulate the pro-inflammatory responses and has some functional redundancy with MAPK (Günzl *et al.*, 2010; Goc *et al.*, 2011; Hochdörfer *et al.*, 2011). Therefore, the involvement of PI3K in the effects observed in this study cannot be excluded either. Further studies will be needed to identify the exact mechanisms involved in RGS16-mediated reduction of cytokine production.

4.1.2 RGS16 in antiviral responses (Publication II)

In addition to bacterial compounds, RGS16 is also induced by viral nucleic acids (Timmusk *et al.*, 2009). Furthermore, in circovirus-infected porcine PBMCs it interacts directly with a viral protein (Timmusk *et al.*, 2009), suggesting that either the virus targets RGS16 to inhibit the antiviral response or RGS16 binds viral proteins to restrict their activity.

In Publication II it is shown that the expression of RGS16 in frog spleens is inhibited *in vivo* by infection with Frog Virus 3, a large double-stranded DNA virus that can be lethal to amphibian tadpoles (Publication II, Figure 4) (Gantress *et al.*, 2003; De Jesús Andino *et al.*, 2012; Morrison *et al.*, 2014). In comparison, the expression of RGS16 in humans is downregulated in response to hepatitis C virus and influenza, but upregulated in response to herpesvirus or measles virus (Publication II, Supplementary Data 4). This suggests that either RGS16 expression is induced differentially in response to different viral infections, or viruses often interfere with the normal expression patterns of RGS16. We favour the hypothesis of viruses targeting RGS16 and its expression, as this would explain the direct interaction with circovirus protein (Timmusk *et al.*, 2009). A reason for viruses to target RGS16 was provided by a recent study published in Nature Biotechnology (Gat-Viks *et al.*, 2013). In this work a computational approach was used together with siRNA experiments to search for responsiveness quantitative trait loci from the mouse genome and the gene *Rgs16* was identified as the causal variant in the chromosome 1 viral stimulus-specific responsiveness quantitative trait locus in bone marrow-derived dendritic cells (Gat-Viks *et al.*, 2013). *Rgs16* siRNA specifically inhibited the expression of genes related to antiviral responses when cells were infected with Sendai virus or treated with an agent mimicking double-stranded RNA (Gat-Viks *et al.*, 2013). *Rgs16* knockdown together with Pam3 treatment decreased the expression of genes of the same antiviral module while *Rgs16* knockdown together with LPS treatment had no apparent effect on the expression of the same set of genes (Gat-Viks *et al.*, 2013).

Due to the antiviral effects of RGS16, the fact that a protein of circovirus specifically binds to RGS16 suggests that this interaction may be partly responsible for the circovirus's success in inhibiting the host's immune responses (Timmusk *et al.*, 2009). In pilot experiments of a mouse model for circovirus infection *Rgs16* KO mice also had higher virus titres than wt mice (Pahtma, 2011), suggesting that restriction of *Rgs16* expression is indeed required for successful viral infection.

In conclusion, it appears that RGS16 is induced in virus infections to enhance the antiviral response via an unidentified pathway. Accordingly, viruses could inhibit the immune response partially by targeting RGS16 itself (e. g. circovirus (Timmusk *et al.*, 2009)) or other targets that regulate its expression (e. g. influenza virus). However, the situation is even more complex than that as

type I interferon generally associated with antiviral responses has been shown to inhibit RGS16 induction (Giorelli *et al.*, 2002).

Taken together, these data show that RGS16 affects immune responses at several levels:

1) RGS16 expression inhibits G protein signalling pathways and their effects, including migration.
2) In bacterial infections (mimicked *in vitro* by Pam3 and LPS treatment) RGS16 induction likely leads to inhibition of p38 MAP kinase activity, providing a restriction mechanism for pro-inflammatory responses. It can also lead to reduced expression of antiviral genes (Gat-Viks *et al.*, 2013).
3) In viral infections RGS16 induction leads to increased expression of antiviral genes (Gat-Viks *et al.*, 2013). The function of RGS16 in this is not simple since interferon can inhibit its expression (Giorelli *et al.*, 2002), but viruses seem to also target either RGS16 or its expression (Timmusk *et al.*, 2009).

These observations suggest a complex interplay involving RGS16. A model of known interactions in proinflammatory and antiviral responses is presented on Figure 8. Future works will aim to identify the unidentified mechanisms and variables in this network.

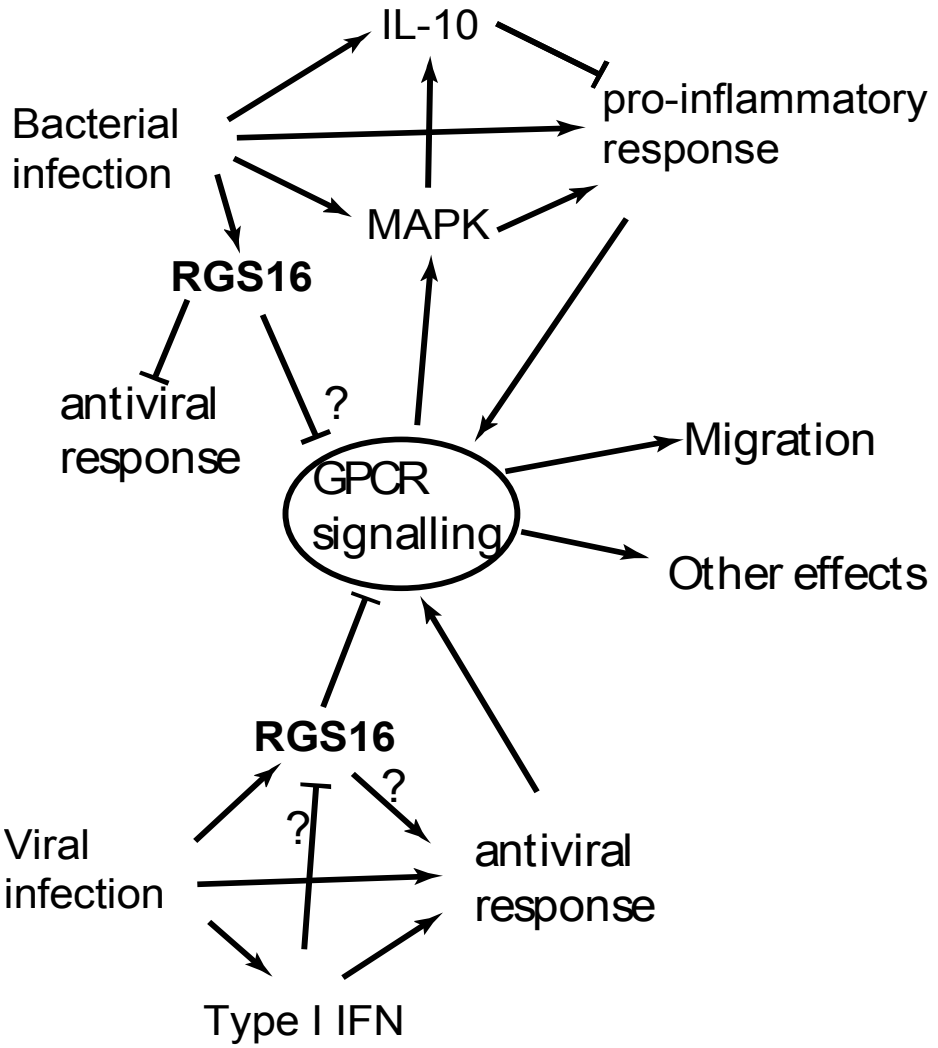


Figure 8. Simplified model of RGS16 in antiviral and pro-inflammatory responses. The expression of RGS16 is induced by both viral and bacterial TLR agonists (Timmusk *et al.*, 2009) and inhibits GPCR signalling pathways. Bacterial infection induces a pro-inflammatory response, and in this context RGS16 inhibits the expression of antiviral genes by an unidentified mechanism (Gat-Viks *et al.*, 2013). RGS16 also restricts cytokine production, likely by inhibiting MAPK activation via modulation of GPCR signalling. In viral infections RGS16 is induced to enhance the antiviral response via an unidentified pathway (Gat-Viks *et al.*, 2013), but its induction can be inhibited by type I interferon (Giorelli *et al.*, 2002).

4.2 The genomic neighbourhood of RGS16 (Publications II, III)

4.2.1 R4 RGS region genes are involved in tetrapod antiviral responses (Publication II)

An initial survey of the genetic neighbourhood of RGS16 revealed a number of genes that either interact with viruses directly or are involved in antiviral responses. The most notable of these is the gene encoding for RNase L (Sadler and Williams, 2008), which is located next to the genes for RGS8 and RGS16 in tetrapods (Wiechec *et al.*, 2008; Timmusk *et al.*, 2009). Other examples include *IVNS1ABP* that was even named by its interaction with the non-structural protein 1A of Influenza Virus (Wolff *et al.*, 1998) and RGS16 that was discussed above in Chapter 4.1.2 (Timmusk *et al.*, 2009) We selected a list of genes with immune function from this region and analysed their expression across microarrays from the ArrayExpress database (Publication II, Online Resource 4). *MR1, IER5, NCF2* and *PTGS2* were overexpressed during various viral infections, others were modulated differentially depending on viruses (Publication II, Online Resource 4). Additionally, as described in chapters 1.1.3 and 4.1.2, R4 RGS themselves have also many interactions with viruses.

We thus hypothesized that antiviral responses may exert an evolutionary pressure keeping the genes of R4 RGS region together across evolution. To test this hypothesis, we investigated whether the orthologs of these genes in a distant vertebrate species were also modulated by viral infection. For this a ranavirus infection model in *Xenopus,* developed by professor Jacques Robert and colleagues (Gantress *et al.*, 2003; De Jesús Andino *et al.*, 2012), was used. The mRNA expression of five R4 RGS region genes in addition to *RGS16* itself (described above) was analysed in the spleen and kidney of virus-infected frogs – *DHX9* (RNA helicase A), *IER5, STX6, NCF2* (NADPH oxidase subunit) and *PTGS2* (cyclooxygenase 2). Our results indicated differential regulation of these genes; the strongest effect was seen for *NCF2* and *PTGS2* in the kidney (~40-fold induction) (Publication II, Figure 4), although classically these are associated with inflammatory responses. Other tested genes were less affected.

Overall, we demonstrated that some genes of the R4 RGS region are differentially modulated by viral infections in humans and frogs, while they do not all belong to the same gene families. Within MHC and related regions, this is not the only example of gene groups that are conservatively kept together across evolution and share related functions. Although a very large region, the MHC itself may be seen as such a region (The MHC sequencing consortium, 1999), with B30.2 proteins, heat shock proteins, immunoproteasome, complement, genes for antigen processing and presentation, etc. (Flajnik and Kasahara, 2001). In fact, the MHC-related paralogon sets have been similarly shown to contain a number of genes that are related by their function in immunity (including antiviral response), but not necessarily by structure (Du Pasquier, 2004b, 2004a;

Flajnik and Du Pasquier, 2004). However, the pressures keeping such synteny groups together remain unclear.

4.2.2 R4 RGS region has an ancient origin (Publications II, III)

The human 'R4 RGS region' as we defined it (Publication II) contains 70 genes from *FAM163A* to *NEK7* (Chromosome 1: 180,154,834 - 198,322,420), including the R4 RGS genes RGS1, 2, 3, 8, 13, 16, 18, and 21. This region is located between two genomic fragile sites and corresponds to a single vertebrate ancestral genomic block (Nakatani *et al.*, 2007; Putnam *et al.*, 2008). Such blocks often contain highly conserved arrays of *cis*-regulatory noncoding elements and their target genes together with bystander genes that are not regulated by these elements (Kikuta *et al.*, 2007; Irimia *et al.*, 2012). We were able to confirm the conserved linkage of this region between mammals (humans and mice), reptilians/birds (chicken and lizard) and amphibians (*Xenopus*) (Publication II, Figure 1, Supplementary Material 1). Generally even the order of genes had been conserved, although in chicken the region had been broken into three translocated regions (Publication II, Figure 1). In the teleost fish the region was found in two copies, both scattered across a single chromosome. However, even in the scattered context of teleost fish the R4 RGS region genes on each fragment are linked together in tetrapod species (Publication II, Figure 1), suggesting that the R4 RGS region was present in the common ancestor of teleost fish and tetrapods.

We further tested the coherence of this region in the genome of the amphioxus *Branchiostoma floridae*, a cephalochordate, which has been sequenced (Holland *et al.*, 2008; Putnam *et al.*, 2008), but not fully assembled into chromosomes. We found seven genomic scaffolds containing two or more R4 RGS region genes (Publication II, Figure 3, Supplementary Material 3), and four of these had been previously mapped to a single chromosome (Vienne *et al.*, 2003b; Castro *et al.*, 2004), suggesting that the orthologs of human R4 RGS region genes were already kept together in the common ancestor of all vertebrates and cephalochordates. Of note is that a single R4 RGS-like domain itself was also found from the amphioxus (as part of a protein with domain structure characteristic of an RH domain Rho GEF) while in chickens and teleost fish the genes for RGS4 and RGS5 were also located next to the other R4 RGS (in human and mouse they are further away from the other R4 RGS on the same chromosome) (Publication II, Figure 1). Taken together, these observations support previous works suggesting that the R4 RGS on human chromosome 1 could stem from duplication events involving a single gene (Snow *et al.*, 1998; Sierra *et al.*, 2002).

As mentioned before, the R4 RGS region is located between two "fragile sites" that are sensitive to DNA damage and prone to formation of double-stranded breaks (Nakatani *et al.*, 2007; Putnam *et al.*, 2008). Mutations that have originated from such events have been detected from a number of genes of the R4 RGS region in various disorders, including RGS16,

one of the R4 RGS themselves (Kawamata *et al.*, 2002; Pichon *et al.*, 2004; Wiechec *et al.*, 2008; Francis *et al.*, 2012; Medjeral-Thomas *et al.*, 2014). As discussed in Chapter 1.2, such genomic locations are often involved in genome rearrangement events, which may have promoted RGS gene duplication. In fact the R4 RGS region also contains a second cluster of highly homologous genes that are often mutated – complement factor H and its homologs (Francis *et al.*, 2012; Medjeral-Thomas *et al.*, 2014). However, there is a remarkable conservation of a core set of genes in the R4 RGS region that can be attributed to the presence of an evolutionarily conserved regulatory block near *RNF2*. Such blocks are generally identified by a regulatory non-coding sequence together with its developmental target gene (Kikuta *et al.*, 2007), which in this case could be *RNF2* itself as it is required for craniofacial development (van der Velden *et al.*, 2013), or *HMCN1* that has been associated with limb / fin development (Carney *et al.*, 2010). Kikuta's findings have been further confirmed by a study showing that the list of 'essential genes' the deletion of which leads to lethality or sterility is significantly enriched in developmental genes, and that such developmental genes remain essential even when duplicated into multiple copies (Makino *et al.*, 2009).

As we found a counterpart of the R4 RGS region in the amphioxus, we investigated whether its origin might be even more ancient and looked for it in various invertebrates. In fact, our initial survey of different invertebrate genomes gave the most promising results from some of the animals that are most distant from vertebrates – the cnidarian *Nematostella vectensis* and the placozoan *Trichoplax adhaerens* (but not the demosponge a*mphimedon queenslandica*), compared to more derived species such as the arthropod *Drosophila melanogaster*. In particular, a single scaffold (scaffold 2) of *T. adhaerens* genome was found to encode for putative orthologs (identified as reciprocal best blast hits) for 14 of the 70 R4 RGS genes across a segment of four megabases, including the LIM homeobox gene *LHX4* that has been previously described in *T. adhaerens* and *N. vectensis* (Srivastava *et al.*, 2010a). Ten genes out of these 14 had been also found linked in amphioxus (Table 1). An RH domain protein was also identified on the same scaffold, but its structure is more similar to the R7 RGS proteins as it contains an additional G protein gamma domain. On the human chromosome 1 one R7 RGS gene is also present, but it is located further toward the telomeric region of the chromosome and not in the defined R4 RGS region. In any case, these results suggest that the R4 RGS region was already constituted in the placozoans and is thus very ancient. The genome of the demosponge *A. queenslandica* may represent a case in which this linkage has been lost secondarily or, maybe, an even more ancestral confirmation as the phylogenetic relationships of sponges and placozoans are not entirely clear yet (Dohrmann and Wörheide, 2013).

Table 1. List of human R4 RGS region genes with orthologs on *T.adhaerens* scaffold 2. The *T. adhaerens* R7 RGS is marked with an asterisk since only its RH domain is shared with proteins in the human R4 RGS region.

human gene	*T. adhaerens* gene	gene start on scaffold 2	gene end on scaffold 2	Amphioxus EntrezID	Amphioxus scaffold
*R7 RGS**	*TriadG52866*	197,869	202,848		
LAMC	*TriadG21436*	2,288,747	2,293,368	*7241772*	34
DHX9	*TriadG20896*	2,326,402	2,331,505	*7242207*	34
SWT1	*TriadG63628*	2,374,744	2,378,052	*7255400*	34
CDC73	*TriadG53180*	2,748,848	2,752,804	*7221437*	200
LHX4	*TriadG20649*	3,694,880	3,718,801	*7247591*	209
NEK7	*TriadG21160*	3,991,137	3,995,789	*7206693*	166
GLUL	*TriadG49883*	4,233,724	4,236,607		
RGL1	*TriadG53399*	4,365,993	4,374,104	*7218548*	136
C1orf27	*TriadG53416*	4,504,423	4,508,451		
ASPM	*TriadG53434*	4,613,305	4,625,611	*7246315*	209
EDEM3	*TriadG20429*	4,957,269	4,962,797		
ARPC5	*TriadG53486*	5,060,428	5,061,610	*7231377*	166
AXDND1	*TriadG53503*	5,189,125	5,196,454		
QSOX1	*TriadG53524*	5,339,787	5,345,352	*7246644*	209

4.3 From the R4 RGS region to other conserved immune-related regions (Publications II, III)

4.3.1 The R4 RGS region is located on one of the MHC paralogons and can be used to track the proto-MHC in invertebrate species (Publications II, III)

The MHC is a conserved genomic region that is related to many immune functions in vertebrates and has three paralogons which in human are located at **(1)** 6p21-22 / 15q13-26, **(2)** 1p11-32 / 1q21-25, **(3)** 5q11-23 / 9p13-24 / 9q32-34 and **(4)** 19p13 (Kasahara, 1997; Flajnik and Kasahara, 2001; Danchin and Pontarotti, 2004a; Flajnik and Kasahara, 2010). We observed that the genomic location of the R4 RGS region (1q25.3 – 1q31.3) is in fact in close proximity to the MHC paralogon on human chromosome 1 and hypothesized that both regions may be linked.

In order to test this hypothesis we searched the human genome for paralogs of the R4 RGS region genes and confirmed their phylogenetic relationships. We found that many genes in the R4 RGS region (35 out of 70) had closely related paralogs (ohnologs) on one or more of the MHC paralogons (Publication II, Figure 2, Online Resource 2). Out of these 35 genes, 23 had at least one of the corresponding human ohnologs on the chromosome 9 paralogon,

which supports the results from earlier studies (Kasahara, 1999a; Flajnik and Kasahara, 2001; Fredriksson *et al.*, 2003; Hokamp *et al.*, 2003). Additionally, the only R4 RGS gene which in human is not located on chromosome 1, *RGS3*, is found on the chromosome 9 MHC paralogon (Publication II, Figure 2, Online Resource 2).

We were also able to confirm that the R4 RGS region had been linked to an MHC paralogon throughout vertebrate evolution, and that the amphioxus orthologs of these genes mainly map to the scaffolds corresponding to the previously identified proto-MHC region of this animal (Publication II, Figure 2 and 3) (Abi-Rached *et al.*, 2002; Vienne *et al.*, 2003b; Castro *et al.*, 2004; Danchin and Pontarotti, 2004a). In fact, one gene of the region encoding for a kinase (*NEK7*) had been previously mapped to a minimal ancestral proto-MHC from which vertebrate and amphioxus MHC regions both derived (Danchin and Pontarotti, 2004a) (Figure 5). This led us to hypothesize that the R4 RGS region may consist of a set of useful markers that can be used to track the proto-MHC in invertebrate species. However, *NEK7* in *Drosophila melanogaster* is not located in the putative proto-MHC at chromosome X (Danchin and Pontarotti, 2004a) (Figure 5).

Checking a number of invertebrate species, we identified an R4 RGS-like region on *T. adhaerens* genomic scaffold 2 (Table 1). A systematic survey of this scaffold revealed genes for which the vertebrate orthologs map to MHC paralogons. Moreover, the same scaffold has been previous described as a highly conserved synteny group corresponding to the MHC paralogon on Chromosome 9; the ohnologs corresponding to 23 R4 RGS region genes were located in this region (Srivastava *et al.*, 2008) (Publication II, Figure 2; Publication III, Table I). One of these 23 genes was the aforementioned *NEK7* (Table 1). Altogether, these observations suggested the presence of a proto-MHC on *Trichoplax adhaerens* scaffold 2.

In order to perform a comprehensive analysis of this hypothesis we obtained a list of all human genes encoded on the four MHC paralogons and used them for reciprocal delta-blast search against the *T. adhaerens* genome. Highly significant clustering of the orthologs of these genes was detected on genomic locations across six of the scaffolds (on scaffolds 2, 3, 7, 9, 10 and 15) (Publication III, Figure 2 and 3, Supplementary Table I and II). The largest and most conserved of such clusters was located on scaffold 2 as expected and also contained most of the orthologs for the R4 RGS region genes (Table 1; Publication III, Figure 2). An R4 RGS gene itself was identified on scaffold 7, at one of the other (smaller) regions considered as part of the proto-MHC.

The validity of our approach was confirmed by a similar analysis against chicken genome. The chicken genome was selected for our control survey beacuse it was divergent enough from the human genome to represent a real test for our approach, since its MHC and related regions had been studied and mapped (Kaufman *et al.*, 1999; Walker *et al.*, 2011; Miller *et al.*, 2014). Our approach successfully identified the MHC paralogons, with the expected

exception of the MHC itself (Publication III, Supplemental Figure 1). Indeed, the chicken MHC is very minimal and contains much less genes than in human (Kaufman *et al.*, 1999), hence the local concentration of best blast hits of human MHC related genes was not high enough to be detected by our approach. While these results provided a good validation of the method, they also pointed out a limit regarding the detection of fragments with a limited number of relevant genes.

In the choanoflagellate *Monosiga brevicollis* (a unicellular non-metazoan) the proto-MHC could not be found (Publication III, Figure 5), suggesting that the proto-MHC is a metazoan innovation (Publication III, Figure 5). However, even in the choanoflagellate nine genes involved in the synteny between human chromosome 9 and *T. adhaerens* scaffold 2 were located on a single scaffold (Publication III, Figure 5). While none of the human homologs of these nine genes is from the R4 RGS region itself, four of them are involved in the regulation of G protein activity: *VAV2* is a GEF while *TBC1D13*, *GARNL3* and *RABGAP1* are GAPs. Furthermore, one of these nine genes shares a function with *EDEM3* from the R4 RGS region (*MAN1B1*, an alpha mannosidase). These observations might reflect that G protein regulation is another function that could be associated with this set of genes throughout evolution.

In conclusion it can be said that the R4 RGS region genes can be used as a novel set of MHC paralogon markers to identify a proto-MHC in evolutionarily distant metazoan species.

4.3.2 The ancestral proto-MHC region contained genes relating to protein degradation, stress response, and antiviral responses (Publication III)

How to explain that the proto-MHC was kept as a relatively conserved genomic unit across evolution from placozoans to arthropods, cephalochordates, and vertebrates? The simplest explanation would involve the absence of fragile sites or evolutionary hotspots. However, the MHC has been rearranged in different vertebrates. Alternative hypotheses to explain the conservation of an MHC region include the presence of highly conserved critical regulatory element stogether with their developmental target genes (Kikuta *et al.*, 2007; Makino *et al.*, 2009; Irimia *et al.*, 2012), and also co-regulation of genes in the region with related functions; functional relationships between genes are preferentially retained in *cis* even after WGD (Makino and McLysaght, 2012).

The presence of long-range regulatory non-coding *cis*-acting elements is considered as one of the main evolutionary forces keeping syntenic relationships intact across evolution as breakage of the relationship between such elements and their developmental target genes often results in detrimental effects (Kikuta *et al.*, 2007). As *T. adhaerens* genome encodes for many non-coding sequences (Hertel *et al.*, 2009), we then tested *T. adhaerens* proto-MHC for such non-coding regulatory elements that would be conserved in at least one more

representative metazoan genome, and may control a key transcription factor. However, no such sequence with obvious regulatory potential was found. In fact, *T. adhaerens* non-coding DNA has little in common with human genome and even the transposon content of *T adhaerens* is very low (Hertel *et al.*, 2009; Wang *et al.*, 2010). Even though there are many tandem repeats in *T. adhaerens*, the most common repeat motif is unique to this animal and not found in other sequenced animals: the pentanucleotide ACAGT (Wang *et al.*, 2010). Therefore the ancestral proto-MHC was not associated with obvious conserved non-coding elements that could be involved in regulating key developmental switches.

We next studied the functional content of *T. adhaerens* proto-MHC. In particular, we expected to find genes encoding for proteasome subunits as the proteasome genes might have constituted the primordial frame on which antigen processing and adaptive immunity were built (Flajnik and Kasahara, 2001). Indeed, the primary proteasome is mainly associated with degradation of aberrant, misfolded, or unneeded proteins, but the MHC-associated proteasome – 'immunoproteasome' – is responsible for the antigen processing for peptide presentation (Saveanu *et al.*, 2002). In fact, in non-mammalian vertebrate species the proteasome genes are associated with the Class III region that is mostly involved in innate immunity (Ohta *et al.*, 2006; Flajnik and Kasahara, 2010). Genes encoding for proteasome subunits have been found from the proto-MHC regions of both amphioxus and fruit fly (Danchin and Pontarotti, 2004a) and their presence in this region thus seems to predate the appearance of vertebrates (and antigen presentation on MHC Class I and Class II molecules) by hundreds of millions of years. From the *T. adhaerens* proto-MHC we found reciprocal best blast hit orthologs of *PSMB7* (on scaffold 2, in a region corresponding to the gene found in human MHC and invertebrate proto-MHC regions), *PSMD5* (also on scaffold 2) and *PSMA4* (on the scaffold 15 proto-MHC fragment), suggesting that the proteasome was linked to this region already near the dawn of metazoan evolution.

The list of genes orthologous to the human MHC paralogon genes and located at the *T. adhaerens* proto-MHC was next analysed for functional content. The genes involved in conserved synteny relationships were found not to be significantly enriched in human transcription factor homologs. Gene ontology analysis found the terms relating to proteasome and ubiquitination as significantly overrepresented in the list of reciprocal best blast hits of *Trichoplax adhaerens* proto-MHC as compared to all genes on human MHC paralogons, further supporting the idea that the proteasome functionality has been kept in the proto-MHC region throughout metazoan evolution.

In addition to the aforementioned proteasome, the corresponding set of genes had likely implications in immunity through stress response and DNA repair. Other significantly overrepresented terms were related to RNA metabolism and gene expression. In contrast, functions traditionally associated with vertebrate immunity (especially with adaptive immunity) were underrepresented, as expected since genes of the adaptive immunity are indeed absent from

the genome of *T. adhaerens*. Thus the ancestral genomic region that eventually evolved into vertebrate MHC paralogons already contained genes related to proteasome and stress response, but obtained most of its other genes involved in vertebrate immunity later during evolution.

What could be the rationale behind the clustering of immune-related genes near proteasome and stress response? In fact, previous studies have linked stress response, DNA repair, ubiquitination, and proteasome all to immune responses, particularly against viruses. Stress may be caused by environmental factors, but also by infections. In the pacific oyster there are thousands of genes involved in responses to different stressors and stimuli (Zhang *et al.*, 2012). Proteins that are misfolded or otherwise aberrant due to cellular stress are tagged by ubiquitination, which then targets them to proteasome for degradation (Hilt and Wolf, 1996; Fang *et al.*, 2011). Cellular stress often results in DNA damage and subsequent triggering of DNA repair mechanisms, which can be greatly decreased by proteasome inhibition (Cron *et al.*, 2013). An explanation for this is provided by the importance of ubiquitination-proteasome pathway in making the DNA at the double-stranded breaks accessible to repair enzymes by removing histones and other bound proteins (Bergink and Jentsch, 2009; Jackson and Durocher, 2013). Furthermore, in *Caenorhabditis elegans* it has been demonstrated that DNA damage in germline cells leads to MAPK activation and subsequent secretion of peptides that induce innate immune responses in somatic tissues by activating the ubiquitination-proteasome system (Ermolaeva *et al.*, 2013). Therefore it is clear that proteasome, ubiquitination, stress response and DNA repair are all parts of an ancient mechanism involved in defence regulation. As for DNA repair, the enzymes involved in DNA repair can function as antiviral sensors / effectors in both humans and bacteria, although by different mechanisms (Babu *et al.*, 2011; Lilley *et al.*, 2011; Ferguson *et al.*, 2012).

In fact, genes from the R4 RGS region may also be lined to these defence functions. One of the stress-related genes in *T. adhaerens* proto-MHC is a homolog of the human DHX9, located close to RGS16 and RNase L in human. DHX9 encodes for RNA helicase A, for which RNA helicase activity is only one of the known functions (Koh *et al.*, 2014). This enzyme unwinds triple helical DNA structures, which could otherwise greatly contribute to genomic fragility and thus also DNA damage (Jain *et al.*, 2013). Moreover, RNA helicase A can also function as a sensor of viral nucleic acids (Kim *et al.*, 2010; Zhang *et al.*, 2011). *EDEM3* that was also found in *T. adhaerens* accelerates the degradation of misfolded glycoproteins and its knockdown results in enhanced hepatitis C virus production in the infected cells (Hirao *et al.*, 2006; Saeed *et al.*, 2011). Next to the cluster of R4 RGS in tetrapod species there is a gene for a deubiquitinase - *UCHL5*, and a gene for an E3 ubiquitin ligase is also found in the region (*RNF2*). The human *QSOX1* – another gene with an ortholog found in *T. adhaerens* - is inducible by cellular stress and inhibits autophagy (Morel *et*

al., 2007; Poillet *et al.*, 2014). RGS16, one of the R4 RGS themselves, is inducible by heat stress (Wong *et al.*, 2008).

In conclusion, we propose that the MHC originates from the functional clustering of genes related to innate (antiviral) defence mechanisms, such as ubiquitination/proteasome and stress response genes, as well as some genes from the R4 RGS region.

4.3.3 Identification of a second MHC-related region in *Trichoplax adhaerens,* the proto-NT (Publication III)

As described in detail in Chapter 1.4, two other sets of paralogons had been previously associated with the MHC paralogons. The NT paralogons were intitially discovered by the presence of paralogs for the neurotrophins NGF and BDNF (Hallböök *et al.*, 1998), but later found to include many immune-related genes including the leukocyte receptor complex and NK receptors (Olinski *et al.*, 2005, 2006; Flajnik and Kasahara, 2010; Flajnik *et al.*, 2012). The JN paralogons were discovered by studying IgSF proteins in tunicates and their homologs in vertebrates (Du Pasquier *et al.*, 2004; Zucchetti *et al.*, 2009; Flajnik *et al.*, 2012). Later studies have suggested that these two sets together with the MHC paralogons may represent rearranged pieces or duplications of a single ancestral region (Kasahara, 1997, 1999b; Flajnik and Kasahara, 2001; Olinski *et al.*, 2005, 2006), which was further supported by finding members of some protein families such as the B7 proteins across all three sets of paralogons (Flajnik *et al.*, 2012).

Our search in chicken and *Trichoplax adhaerens* in fact also included genes located on NT and JN paralogons. Similar to MHC paralogons, we were able to successfully detect the corresponding regions in chicken (Publication III, Supplemental Figure 1). The proto-NT region was mostly detected on scaffolds 1, 6, 9, 12 and a number of smaller scaffolds (Publication III, Figure 2). This is in contrast to proto-MHC which was found on scaffolds 2, 3, 7, 9, and 10, suggesting that the two regions were mostly separated already in *T. adhaerens*. Although clearly JN-related regions were not found (other than the very small scaffold 35) (Publication III, Figure 2), we found genes that have orthologs associated with JN paralogons from both proto-MHC and proto-NT regions, supporting a strong connection between the three sets.

While our approach in *T. adhaerens* detected many regions for which most of the detected reciprocal best blast hits belonged to either MHC-related or NT-related paralogon sets, other regions had no clear dominance of one gene set over the other. These were found on scaffolds 1, 3, 9, and 10 and contained a mixture of genes from two or more of these paralogy sets (Publication III, Figure 2). Notably, one third of scaffold 9 was associated mainly with MHC, another third with NT, and the last third located between MHC and NT had mixed content from both (Publication III, Figure 2).

Finally, we were able to show that these two MHC-related gene sets might also be involved in both proteasome and antiviral responses. While the reciprocal best blast hits for three human proteasome genes mapped to the proto-MHC as described above, there was also a number of other proteasome genes that where not reciprocal best blast hits, but were the best blast hit in *T. adhaerens* for human proteasome genes. Remarkably, most of them appeared associated with the identified MHC-related, NT-related, and mixed content regions (Figure 9), suggesting that the proteasome function is also common to other MHC-related gene sets. Furthermore, when searching for the genomic locations of B30.2 genes in *T. adhaerens*, we found that most of them were also associated with the identified regions (Figure 9; Publication III, Figure 4). The B30.2 proteins are remarkable in that many of them are ubiquitin ligases and they are associated with both the MHC and with innate/antiviral responses (Abi-Rached *et al.*, 1999; Rhodes *et al.*, 2005; Darbo *et al.*, 2008; Boudinot *et al.*, 2011; D'Cruz *et al.*, 2013).

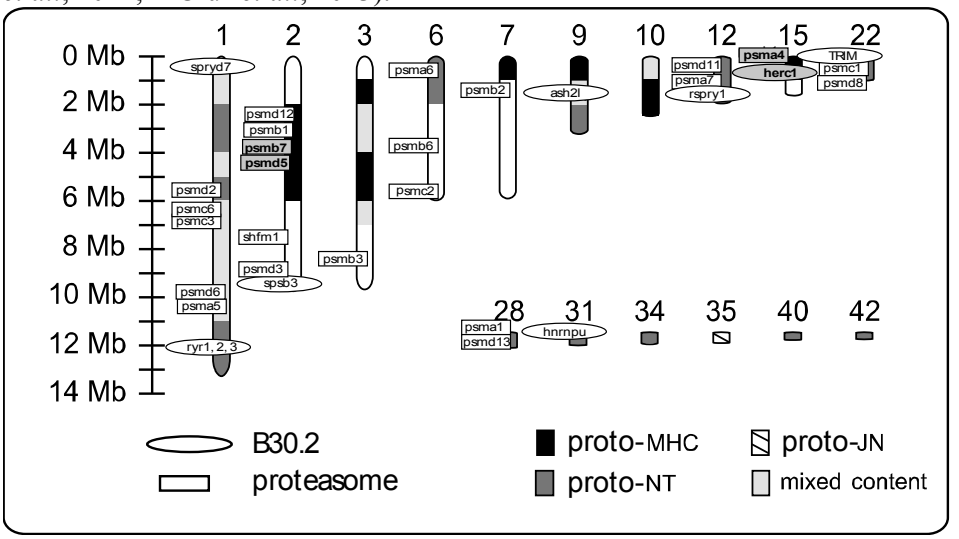


Figure 9. B30.2 and proteasome genes on the *T. adhaerens* MHC-related scaffolds. The identified Proto-MHC, Proto-NT, Proto-JN and mixed content regions are marked with different colours. B30.2 and proteasome genes that were reciprocal best blast hits between human and *T. adhaerens* are labelled with dark-coloured markers. Modified from Publication III, Figure 2.

As an additional control, we searched the genome of the choanoflagellate *Monosiga brevicollis* (a non-metazoan, protist species) for the presence of proteasome genes and MHC-related regions. Our approach gave no significant results from this organism, suggesting that the functional clustering of MHC-related genes was a metazoan innovation. While one of its scaffolds seemed to contain some genes with human orthologs on chromosome 9 and *T. adhaerens* orthologs on scaffold 2, the proteasome genes were found scattered across the entire *M.* brevicollis genome (Publication III, Figure 5). This provides

even further support for the hypothesis of ancient linkage between the MHC-related paralogon sets (Kasahara, 1997, 1999b; Flajnik and Kasahara, 2001; Olinski *et al.*, 2005, 2006) being metazoan-specific. Although *T. adhaerens* is a very basal-like animal, it is in fact also derived from the hypothetical Ur-metazoan, in which the gene sets may have been even more strongly interspersed.

4.4 Future perspectives

4.4.1 RGS16 and its implications in immune regulation

In the present study we showed that RGS16 is involved in both antiviral and inflammatory immune responses. However, pathways in which RGS16 is involved mostly remain to be identified and the main role of RGS16 in primary monocytes remains elusive. RGS16 inhibits GPCR signalling, but the mechanism linking it with its downstream effects of induction of viral response genes or reduced pro-inflammatory cytokine production is not obvious. Even the idea that it restricts pro-inflammatory responses by inhibiting MAPK activity is mostly a speculation and will require experimental confrontation. Of particular interest is the direct interaction of RGS16 with a circovirus protein described in (Timmusk *et al.*, 2009), which might bring further light on its involvement in antiviral responses. The mouse model of PCV2 infection described in (Pahtma, 2011) will be a very useful tool for these studies.

In addition to responses to pathogens, RGS16 has been also associated with various autoinflammatory disorders (Li *et al.*, 2013), confirming its involment in immune regulation. An intriguing thought is that RGS16 could be also involved in autoimmune diseases, similar to RGS1 that is also associated with immune regulation (Hunt *et al.*, 2008; Smyth *et al.*, 2008; Romanos *et al.*, 2009; International Multiple Sclerosis Genetics Consortium, 2010; Johnson *et al.*, 2010; Lindén *et al.*, 2013; Mowry *et al.*, 2013). Examples of possible interaction partners of RGS16 in such diseases include chemokine receptors and cannabinoid receptors, both of which have been previously associated with Multiple Sclerosis (Rossi *et al.*, 2013; Cheng and Chen, 2014). Pilot experiments using a mouse model of MS on *Rgs16* KO mice in collaboration with Dr. Simon Fillatreau, DRFZ, Berlin have shown that the disruption of the *Rgs16* gene is protective against disease development, at least in females (Sirje Rüütel Boudinot, personal communications). Furthermore, preliminary data of our laboratory suggest that PBMCs from untreated MS patients, as compared to patients receiving treatment or healthy individuals, express less RGS16 (Kärdi, 2014). This could lead to increased migration toward the inflammatory lesions in central nervous system (Cheng and Chen, 2014) and, according to the results presented here, might also contribute to increased pro-inflammatory activity. It would be of great interest to further study RGS16 expression as a novel marker of MS in humans. The mechanisms for RGS16 involvement in MS will

be further studied in *Rgs16* KO mice in the context of both the disease and its various treatment options e. g. drugs based on Type I IFN or cannabinoids.

4.4.2 Evolution of the genomic regions involved in immunity

While the current study focused to the analysis of the proto-MHC region in the placozoan *Trichoplax adhaerens*, the proto-NT and "mixed content" regions received less attention. In the future we are going to analyse these genomic regions also in detail using a similar approach. In addition, it would be of interest to confront the hypothesis that the proto-MHC, proto-NT and proto-JN originated from rearrangements and duplications of one single region (Kasahara, 1997, 1999b; Flajnik and Kasahara, 2001; Olinski *et al.*, 2005, 2006).

In *T. adhaerens* the region studied is incredibly similar to the vertebrate corresponding genomic regions. It would be of interest to track its evolution across different species from basal-like animals to human. In particular, a recent study showed that molluscs and annelids (e. g. The owl limpet *Lottia gigantea*) have very well conserved synteny with most of the ancestral linkage groups identified from studies of the amphioxus (Simakov *et al.*, 2013). Hence, the owl limpet is one of the species that will be extensively studied in our future analyses. Other interesting species to consider include the recently sequenced Ctenophore, who has developed alternative signalling molecules in parallel to other metazoans (Moroz *et al.*, 2014), and maybe arthropods (e. g. fruit fly or horseshoe crab) although the synteny groups are certainly less conserved in these organisms (Simakov *et al.*, 2013; Nossa *et al.*, 2014). For such study in multiple species, reciprocal blast searches may not be the most efficient approach. In collaboration with Dr Pierre Pontarroti (CNRS, Marseille, France) we consider to use the program OrthoMCL instead to produce a matrix of orthology between a dozen invertebrates (including own limpet and ctenophore) for the relevant genes (Chen *et al.*, 2006; Fischer *et al.*, 2011). The final selection of genomes will be done based on four main criteria:

(1) Quality of the genomes that will be determined by analysis of twenty randomly chosen long genes.
(2) Previous studies showing a good synteny between the selected genome and vertebrate genomes, especially for the proto-MHC or related regions.
(3) Species in which there have been a lot of genomic rearrangements (based on literature) will have to be avoided as much as possible, but some will be kept considering (2), such as *Ciona* and *Drosophila*.
(4) A wide diversity across different phyla of invertebrates

Recent studies by Lagman and colleagues reported the identification of additional sets of paralogons that seems to be mostly associated with the same chromosomes as the MHC, NT, and JN (Lagman *et al.*, 2012; Lagman *et al.*, 2013). One of these sets is mainly identified by components of G protein

signalling, including GPCRs such as opsins and vasopressin receptors, as well as Gα subunits (Lagman *et al.*, 2013). Other involved genes include ohnologs corresponding to the CACNA1 ion channels, the neuronal adhesion molecule L1CAM, plexins, filamin, and ubiquitin pathway components (Lagman *et al.*, 2013). In human this "new" paralogy set has been identified at chromosomes 1, 3, 7/12, and X and on chromosome 1 it may roughly correspond to the NT paralogon with genes located at 1p13 as well as 1q32 (Lagman *et al.*, 2013), in proximity to the R4 RGS region. Gβ subunits are associated with another paralogon set that in human is located mainly at chromosome 1, 3, 12 and 17 (Lagman *et al.*, 2012), paralogy groups between chromosome 3 and 17 has been also discussed in earlier works (Larhammar *et al.*, 2002; McLysaght *et al.*, 2002). It would be interesting to test whether these paralogon sets might also be connected to the ancestral MHC-related region. It is possible that the RGS genes were originally associated with the region that became the GPCR paralogons, but then during ancestral genomic rearrangements ended up where they are now.

Besides the *in silico* studies it would be equally important to test whether the "immune-related" genes found in *T. adhaerens* are truly immune related, or the antiviral responses were acquired later during evolution. Experiments on *T. adhaerens* and other basal-like invertebrate species would be able to confirm our results experimentally. It would be especially interesting to experimentally compare the RGS proteins of vertebrates – such as RGS16 - and those found in other organisms such as amoeba (Eichinger *et al.*, 2005).

CONCLUSIONS

1. RGS16 is involved in the tetrapod immune responses. In humans and mice RGS16 restricts the proinflammatory response of monocytic cells.

2. In addition to RGS16 itself, many of its neighbours on the genome are also involved in antiviral pathways. The origin of this entire region – the R4 RGS region – can be tracked throughout metazoan evolution to placozoans.

3. The R4 RGS region genes are linked to the MHC-related gene sets in the genomes of both vertebrates and their invertebrate antecedants. This can be used to detect ancestral MHC-related regions in evolutionarily distant species.

4. Using these markers a proto-MHC was found in the placozoan *Trichoplax adhaerens*, as well as a region that mainly corresponds to a proto-neurotrophin genomic segment. The gene set located in the proto-MHC seems to be significantly involved in the proteasome and stress response pathways.

REFERENCES

Abi-Rached, L., McDermott, M. F., Pontarotti, P. (1999). The MHC big bang. Immunol Rev 167, 33-44.

Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P., Inoko, H. (2002). Evidence of en bloc duplication in vertebrate genomes. Nat Genet 31, 100-105.

Aboura, A., Coulomb-L'Herminé, A., Audibert, F., Capron, F., Frydman, R., Tachdjian, G. (2002). De novo interstitial direct duplication 1(q23.1q31.1) in a fetus with Pierre Robin sequence and camptodactyly. Am J Med Genet 108, 153-159.

Abrink, M., Gobl, A. E., Huang, R., Nilsson, K., Hellman, L. (1994). Human cell lines U-937, THP-1 and Mono Mac 6 represent relatively immature cells of the monocyte-macrophage cell lineage. Leukemia 8, 1579-1584.

Adams, K. L., Wendel, J. F. (2005). Polyploidy and genome evolution in plants. Curr Opin Plant Biol 8, 135-141.

Agenès, F., Bosco, N., Mascarell, L., Fritah, S., Ceredig, R. (2005). Differential expression of regulator of G-protein signalling transcripts and in vivo migration of CD4+ naive and regulatory T cells. Immunology 115, 179-188.

Ajmal, W., Khan, H., Abbasi, A. A. (2014). Phylogenetic investigation of human FGFR-bearing paralogons favors piecemeal duplication theory of vertebrate genome evolution. Mol Phylogenet Evol.

Alcaide, M., Edwards, S. V., Negro, J. J. (2007). Characterization, polymorphism, and evolution of MHC class II B genes in birds of prey. J Mol Evol 65, 541-554.

Alekseyev, M. A., Pevzner, P. A. (2009). Breakpoint graphs and ancestral genome reconstructions. Genome Res 19, 943-957.

Almgren, M., Nyengaard, J. R., Persson, B., Lavebratt, C. (2008). Carbamazepine protects against neuronal hyperplasia and abnormal gene expression in the megencephaly mouse. Neurobiol Dis 32, 364-376.

Ambreen, S., Khalil, F., Abbasi, A. A. (2014). Integrating large-scale phylogenetic datasets to dissect the ancient evolutionary history of vertebrate genome. Mol Phylogenet Evol 78C, 1-13.

Amstadter, A. B., Koenen, K. C., Ruggiero, K. J., Acierno, R., Galea, S., Kilpatrick, D. G., Gelernter, J. (2009). Variation in RGS2 is associated with suicidal ideation in an epidemiological study of adults exposed to the 2004 Florida hurricanes. Arch Suicide Res 13, 349-357.

Anantharaman, V., Abhiman, S., de Souza, R. F., Aravind, L. (2011). Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system. Gene 475, 63-78.

Anderson, G. R., Posokhova, E., Martemyanov, K. A. (2009). The R7 RGS protein family: multi-subunit regulators of neuronal G protein signaling. Cell Biochem Biophys 54, 33-46.

Arai, K., Maruyama, Y., Nishida, M., Tanabe, S., Takagahara, S., Kozasa, T., Mori, Y., Nagao, T., Kurose, H. (2003). Differential requirement of G alpha12, G alpha13, G alphaq, and G beta gamma for endothelin-1-induced c-Jun NH2 terminal kinase and extracellular signal-regulated kinase activation. Mol Pharmacol 63, 478-488.

Azumi, K., De Santis, R., De Tomaso, A., Rigoutsos, I., Yoshizaki, F., Pinto, M. R., Marino, R., Shida, K., Ikeda, M., et al. (2003). Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". Immunogenetics 55, 570-581.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., et al. (2011). A dual function of the CRISPR Cas system in bacterial antivirus immunity and DNA repair. Mol Microbiol 79, 484-502.

Balakrishnan, C. N., Ekblom, R., Völker, M., Westerdahl, H., Godinez, R., Kotkiewicz, H., Burt, D. W., Graves, T., Griffin, D. K., et al. (2010). Gene duplication and fragmentation in the zebra finch major histocompatibility complex. BMC Biol 8, 29.

Bannai, H. P., Nonaka, M. (2013). Comprehensive analysis of medaka major histocompatibility complex (MHC) class II genes: implications for evolution in teleosts. Immunogenetics 65, 883-895.

Bansal, G., Druey, K. M., Xie, Z. (2007). R4 RGS proteins: regulation of G protein signaling and beyond. Pharmacology and Therapeutics 116, 473-495.

Bansal, G., DiVietro, J. A., Kuehn, H. S., Rao, S., Nocka, K. H., Gilfillan, A. M., Druey, K. M. (2008a). RGS13 controls G protein-coupled receptor-evoked responses of human mast cells. J Immunol 181, 7882-7890.

Bansal, G., Xie, Z., Rao, S., Nocka, K. H., Druey, K. M. (2008b). Suppression of immunoglobulin E-mediated allergic responses by regulator of G protein signaling 13. Nat Immunol 9, 73-80.

Barker, K. S., Liu, T., Rogers, P. D. (2005). Coculture of THP-1 human mononuclear cells with Candida albicans results in pronounced changes in host gene expression. J Infect Dis 192, 901-912.

Bartl, S., Baish, M., Weissman, I. L., Diaz, M. (2003). Did the molecules of adaptive immunity evolve from the innate immune system? Integr Comp Biol 43, 338-346.

Bastin, G., Singh, K., Dissanayake, K., Mighiu, A. S., Nurmohamed, A., Heximer, S. P. (2012). Amino-terminal cysteine residues differentially influence RGS4 protein plasma membrane targeting, intracellular trafficking, and function. J Biol Chem 287, 28966-28974.

Basu, S., Srivastava, P. K. (2000). Heat shock proteins: the fountainhead of innate and adaptive immune responses. Cell Stress Chaperones 5, 443-451.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., de Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 327, 836-840.

Baumgarten, E., Wegner, R. D., Fengler, R., Koch, H., Henze, G. (1993). Partial trisomy 1q, an uncommon chromosomal aberration in erythroleukemia. Leuk Lymphoma 10, 237-240.

Beadling, C., Druey, K. M., Richter, G., Kehrl, J. H., Smith, K. A. (1999). Regulators of G protein signaling exhibit distinct patterns of gene expression and target G protein specificity in human lymphocytes. J Immunol 162, 2677-2682.

Beçak, M. L. (2014). Polyploidy and epigenetic events in the evolution of Anura. Genet Mol Res 13, 5995-6014.

Beetz, N., Harrison, M. D., Brede, M., Zong, X., Urbanski, M. J., Sietmann, A., Kaufling, J., Lorkowski, S., Barrot, M., et al. (2009). Phosducin influences sympathetic activity and prevents stress-induced hypertension in humans and mice. J Clin Invest 119, 3597-3612.

Belov, K., Lam, M. K., Hellman, L., Colgan, D. J. (2003). Evolution of the major histocompatibility complex: Isolation of class II beta cDNAs from two monotremes, the platypus and the short-beaked echidna. Immunogenetics 55, 402-411.

Belov, K., Deakin, J. E., Papenfuss, A. T., Baker, M. L., Melman, S. D., Siddle, H. V., Gouin, N., Goode, D. L., Sargeant, T. J., et al. (2006). Reconstructing an ancestral

mammalian immune supercomplex from a marsupial major histocompatibility complex. PLoS Biol 4, e46.

Bergink, S., Jentsch, S. (2009). Principles of ubiquitin and SUMO modifications in DNA repair. Nature 458, 461-467.

Berthebaud, M., Rivière, C., Jarrier, P., Foudi, A., Zhang, Y., Compagno, D., Galy, A., Vainchenker, W., Louache, F. (2005). RGS16 is a negative regulator of SDF 1-CXCR4 signaling in megakaryocytes. Blood 106, 2962-2968.

Beutler, B. (2005). The Toll-like receptors: analysis by forward genetic methods. Immunogenetics 57, 385-392.

Biesterfeld, S., Gerres, K., Fischer-Wein, G., Böcking, A. (1994). Polyploidy in non neoplastic tissues. J Clin Pathol 47, 38-42.

Bingulac-Popovic, J., Figueroa, F., Sato, A., Talbot, W. S., Johnson, S. L., Gates, M., Postlethwait, J. H., Klein, J. (1997). Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, Danio rerio. Immunogenetics 46, 129-134.

Bisbee, C. A., Baker, M. A., Wilson, A. C., Haji-Azimi, I., Fischberg, M. (1977). Albumin phylogeny for clawed frogs (Xenopus). Science 195, 785-787.

Blackwell, J. M., Jamieson, S. E., Burgner, D. (2009). HLA and infectious diseases. Clin Microbiol Rev 22, 370-385, Table of Contents.

Blaukat, A., Barac, A., Cross, M. J., Offermanns, S., Dikic, I. (2000). G protein coupled receptor-mediated mitogen-activated protein kinase activation through cooperation of Galpha(q) and Galpha(i) signals. Mol Cell Biol 20, 6837 6848.

Boehm, T., McCurley, N., Sutoh, Y., Schorpp, M., Kasahara, M., Cooper, M. D. (2012). VLR-based adaptive immunity. Annu Rev Immunol 30, 203-220.

Boudinot, P., van der aa, L. M., Jouneau, L., Du Pasquier, L., Pontarotti, P., Briolat, V., Benmansour, A., Levraud, J. P. (2011). Origin and evolution of TRIM proteins: new insights from the complete TRIM repertoire of zebrafish and pufferfish. PLoS One 6, e22022.

Brass, L. F., Ma, P. (2012). Applying the brakes to platelet activation. Blood 119, 3651-3652.

Brower, C. S., Varshavsky, A. (2009). Ablation of arginylation in the mouse N-end rule pathway: loss of fat, higher metabolic rate, damaged spermatogenesis, and neurological perturbations. PLoS ONE 4, e7757.

Brunet, F. G., Roest Crollius, H., Paris, M., Aury, J. M., Gibert, P., Jaillon, O., Laudet, V., Robinson-Rechavi, M. (2006). Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Mol Biol Evol 23, 1808 1816.

Buckley, K. M., Smith, L. C. (2007). Extraordinary diversity among members of the large gene family, 185/333, from the purple sea urchin, Strongylocentrotus purpuratus. BMC Mol Biol 8, 68.

Burt, D. W. (2002). Origin and evolution of avian microchromosomes. Cytogenet Genome Res 96, 97-112.

Byeon, S. E., Lee, J., Kim, J. H., Yang, W. S., Kwak, Y. S., Kim, S. Y., Choung, E. S., Rhee, M. H., Cho, J. Y. (2012). Molecular mechanism of macrophage activation by red ginseng acidic polysaccharide from Korean red ginseng. Mediators Inflamm 2012, 732860.

Cabrera-Vera, T. M., Vanhauwe, J., Thomas, T. O., Medkova, M., Preininger, A., Mazzoni, M. R., Hamm, H. E. (2003). Insights into G protein structure, function, and regulation. Endocr Rev 24, 765-781.

Cahir-McFarland, E. D., Carter, K., Rosenwald, A., Giltnane, J. M., Henrickson, S. E., Staudt, L. M., Kieff, E. (2004). Role of NF-kappa B in cell survival and transcription

of latent membrane protein 1-expressing or Epstein Barr virus latency III-infected cells. J Virol 78, 4108-4119.

Calò, L. A., Pagnin, E., Davis, P. A., Sartori, M., Ceolotto, G., Pessina, A. C., Semplicini, A. (2004). Increased expression of regulator of G protein signaling-2 (RGS 2) in Bartter's/Gitelman's syndrome. A role in the control of vascular tone and implication for hypertension. J Clin Endocrinol Metab 89, 4153-4157.

Cambosu, F., Capobianco, G., Fogu, G., Bandiera, P., Pirino, A., Moro, M. A., Sanna, R., Soro, G., Dessole, M., Montella, A. (2013). Partial trisomy of the long arm of chromosome 1: prenatal diagnosis, clinical evaluation and cytogenetic findings. Case report and review of the literature. J Obstet Gynaecol Res 39, 592-597.

Campbell, D. B., Lange, L. A., Skelly, T., Lieberman, J., Levitt, P., Sullivan, P. F. (2008). Association of RGS2 and RGS5 variants with schizophrenia symptom severity. Schizophrenia Research 101, 67-75.

Carman, C. V., Parent, J. L., Day, P. W., Pronin, A. N., Sternweis, P. M., Wedegaertner, P. B., Gilman, A. G., Benovic, J. L., Kozasa, T. (1999). Selective regulation of Galpha(q/11) by an RGS domain in the G protein-coupled receptor kinase, GRK2. J Biol Chem 274, 34483-34492.

Carney, T. J., Feitosa, N. M., Sonntag, C., Slanchev, K., Kluger, J., Kiyozumi, D., Gebauer, J. M., Coffin Talbot, J., Kimmel, C. B., et al. (2010). Genetic analysis of fin development in zebrafish identifies furin and hemicentin1 as potential novel fraser syndrome disease genes. PLoS Genet 6, e1000907.

Castro, L. F., Furlong, R. F., Holland, P. W. (2004). An antecedent of the MHC linked genomic region in amphioxus. Immunogenetics 55, 782-784.

Chaudhary, P., Hepgur, M., Sarkissian, S., Smith, R. J., Weitz, I. C. (2014). Atypical haemolytic-uraemic syndrome due to heterozygous mutations of CFH/CFHR1-3 and complement factor H 479. Blood Transfus 12, 111-113.

Chen, C., Zheng, B., Han, J., Lin, S. C. (1997). Characterization of a novel mammalian RGS protein that binds to Galpha proteins and inhibits pheromone signaling in yeast. J Biol Chem 272, 8679-8685.

Chen, C. K., Wieland, T., Simon, M. I. (1996). RGS-r, a retinal specific RGS protein, binds an intermediate conformation of transducin and enhances recycling. Proc Natl Acad Sci U S A 93, 12885-12889.

Chen, F., Mackey, A. J., Stoeckert, C. J., Jr., Roos, D. S. (2006). OrthoMCL DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res 34, D363-368.

Chen, J. M., Cooper, D. N., Chuzhanova, N., Férec, C., Patrinos, G. P. (2007). Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8, 762-775.

Chen, L. C., Dollbaum, C., Smith, H. S. (1989). Loss of heterozygosity on chromosome 1q in human breast cancer. Proc Natl Acad Sci U S A 86, 7204 7207.

Cheng, W., Chen, G. (2014). Chemokines and chemokine receptors in multiple sclerosis. Mediators Inflamm 2014, 659206.

Cheng, Y. S., Lee, T. S., Hsu, H. C., Kou, Y. R., Wu, Y. L. (2008). Characterization of the transcriptional regulation of the regulator of G protein signaling 2 (RGS2) gene during 3T3-L1 preadipocyte differentiation. J Cell Biochem 105, 922-930.

Chng, W. J., Remstein, E. D., Fonseca, R., Bergsagel, P. L., Vrana, J. A., Kurtin, P. J., Dogan, A. (2009). Gene expression profiling of pulmonary mucosa associated lymphoid tissue lymphoma identifies new biologic insights with potential diagnostic and therapeutic applications. Blood 113, 635-645.

Chow, J., Poon, R. Y. (2010). DNA damage and polyploidization. Adv Exp Med Biol 676, 57-71.

Chrobak, I., Lenna, S., Stawski, L., Trojanowska, M. (2013). Interferon gamma promotes vascular remodeling in human microvascular endothelial cells by upregulating endothelin (ET)-1 and transforming growth factor (TGF) beta2. J Cell Physiol 228, 1774-1783.

Cismowski, M. J., Takesono, A., Bernard, M. L., Duzic, E., Lanier, S. M. (2001). Receptor-independent activators of heterotrimeric G-proteins. Life Sci 68, 2301-2308.

Cohen, S. P., Buckley, B. K., Kosloff, M., Garland, A. L., Bosch, D. E., Cheng, G., Jr., Radhakrishna, H., Brown, M. D., Willard, F. S., et al. (2012). Regulator of G-protein signaling-21 (RGS21) is an inhibitor of bitter gustatory signaling found in lingual and airway epithelia. J Biol Chem 287, 41706-41719.

Cole, F., Baudat, F., Grey, C., Keeney, S., de Massy, B., Jasin, M. (2014). Mouse tetrad analysis provides insights into recombination mechanisms and hotspot evolutionary dynamics. Nat Genet 46, 1072-1080.

Comai, L. (2005). The advantages and disadvantages of being polyploid. Nat Rev Genet 6, 836-846.

Cron, K. R., Zhu, K., Kushwaha, D. S., Hsieh, G., Merzon, D., Rameseder, J., Chen, C. C., D'Andrea, A. D., Kozono, D. (2013). Proteasome inhibitors block DNA repair and radiosensitize non-small cell lung cancer. PLoS One 8, e73710.

Cui, H., Nishiguchi, N., Ivleva, E., Yanagi, M., Fukutake, M., Nushida, H., Ueno, Y., Kitamura, N., Maeda, K., Shirakawa, O. (2008). Association of RGS2 gene polymorphisms with suicide and increased RGS2 immunoreactivity in the postmortem brain of suicide victims. Neuropsychopharmacology 33, 1537 1544.

D'Cruz, A. A., Babon, J. J., Norton, R. S., Nicola, N. A., Nicholson, S. E. (2013). Structure and function of the SPRY/B30.2 domain proteins involved in innate immunity. Protein Sci 22, 1-10.

da Costa Goncalves, A., Luft, F. C., Gross, V. (2008). Fine tuning of blood pressure by the regulator of G protein signaling (RGS) 2. J am Soc Hypertens 2, 403-409.

Danchin, E., Vitiello, V., Vienne, A., Richard, O., Gouret, P., McDermott, M. F., Pontarotti, P. (2004). The major histocompatibility complex origin. Immunol Rev 198, 216-232.

Danchin, E. G., Abi-Rached, L., Gilles, A., Pontarotti, P. (2003). Conservation of the MHC-like region throughout evolution. Immunogenetics 55, 141-148.

Danchin, E. G., Pontarotti, P. (2004a). Towards the reconstruction of the bilaterian ancestral pre-MHC region. Trends Genet 20, 587-591.

Danchin, E. G., Pontarotti, P. (2004b). Statistical evidence for a more than 800 million-year-old evolutionarily conserved genomic region in our genome. J Mol Evol 59, 587-597.

Darbo, E., Danchin, E. G., Mc Dermott, M. F., Pontarotti, P. (2008). Evolution of major histocompatibility complex by "en bloc" duplication before mammalian radiation. Immunogenetics 60, 423-438.

De Jesús Andino, F., Chen, G., Li, Z., Grayfer, L., Robert, J. (2012). Susceptibility of *Xenopus laevis* tadpoles to infection by the ranavirus Frog Virus 3 correlates with a reduced and delayed innate immune response in comparison with adult frogs. Virology 432, 435-443.

de la Casa-Esperón, E. (2012). Horizontal transfer and the evolution of host pathogen interactions. Int J Evol Biol 2012, 679045.

de Mendoza, A., Sebé-Pedró, A., Ruiz-Trillo, I. (2014). The evolution of the GPCR signaling system in eukaryotes: modularity, conservation, and the transition to metazoan multicellularity. Genome Biol Evol 6, 606-619.

De Vries, L., Fischer, T., Tronchère, H., Brothers, G. M., Strockbine, B., Siderovski, D. P., Farquhar, M. G. (2000). Activator of G protein signaling 3 is a guanine dissociation inhibitor for Galpha i subunits. Proc Natl Acad Sci U S A 97, 14364-14369.

Della Monica, M., Lonardo, F., Faravelli, F., Pierluigi, M., Luquetti, D. V., De Gregori, M., Zuffardi, O., Scarano, G. (2007). A case of autism with an interstitial 1q deletion (1q23.3-24.2) and a de novo translocation of chromosomes 1q and 5q. Am J Med Genet a 143A, 2733-2737.

Delsuc, F., Brinkmann, H., Chourrout, D., Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439, 965-968.

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J. M., Da Silva, C., Brinkmann, H., Mikhaleva, J., Olsen, L. C., Jubin, C., et al. (2010). Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. Science 330, 1381-1385.

Deshpande, D. A., Wang, W. C., McIlmoyle, E. L., Robinett, K. S., Schillinger, R. M., An, S. S., Sham, J. S., Liggett, S. B. (2010). Bitter taste receptors on airway smooth muscle bronchodilate by localized calcium signaling and reverse obstruction. Nat Med 16, 1299-1304.

Dillon, L. W., Burrow, A. A., Wang, Y. H. (2010). DNA instability at chromosomal fragile sites in cancer. Curr Genomics 11, 326-337.

Ding, Y., Li, J., Wu, Q., Yang, P., Luo, B., Xie, S., Druey, K. M., Zajac, A. J., Hsu, H. C., Mountz, J. D. (2013). IL-17RA is essential for optimal localization of follicular Th cells in the germinal center light zone to promote autoantibody-producing B cells. J Immunol 191, 1614-1624.

Dohrmann, M., Wörheide, G. (2013). Novel scenarios of early animal evolution--is it time to rewrite textbooks? Integr Comp Biol 53, 503-511.

Doi, M., Ishida, A., Miyake, A., Sato, M., Komatsu, R., Yamazaki, F., Kimura, I., Tsuchiya, S., Kori, H., et al. (2011). Circadian regulation of intracellular G-protein signalling mediates intercellular synchrony and rhythmicity in the suprachiasmatic nucleus. Nat Commun 2, 327.

Dragon-Durey, M. A., Blanc, C., Marliot, F., Loirat, C., Blouin, J., Sautes-Fridman, C., Fridman, W. H., Frémeaux-Bacchi, V. (2009). The high frequency of complement factor H related CFHR1 gene deletion is restricted to specific subgroups of patients with atypical haemolytic uraemic syndrome. J Med Genet 46, 447-450.

Dreyfus, D. H., Jones, J. F., Gelfand, E. W. (1999). Asymmetric DDE (D35E)-like sequences in the RAG proteins: implications for V(D)J recombination and retroviral pathogenesis. Med Hypotheses 52, 545-549.

Druey, K. M., Ugur, O., Caron, J. M., Chen, C. K., Backlund, P. S., Jones, T. L. (1999). Amino-terminal cysteine residues of RGS16 are required for palmitoylation and modulation of Gi- and Gq-mediated signaling. J Biol Chem 274, 18836-18842.

Du Pasquier, L. (2001). The immune system of invertebrates and vertebrates. Comp Biochem Physiol B Biochem Mol Biol 129, 1-15.

Du Pasquier, L. (2004a). Speculations on the origin of the vertebrate immune system. Immunol Lett 92, 3-9.

Du Pasquier, L. (2004b). Innate immunity in early chordates and the appearance of adaptive immunity. C R Biol 327, 591-601.

Du Pasquier, L., Zucchetti, I., De Santis, R. (2004). Immunoglobulin superfamily receptors in protochordates: before RAG time. Immunol Rev 198, 233-248.

Dusonchet, J., Li, H., Guillily, M., Liu, M., Stafa, K., Derada Troletti, C., Boon, J. Y., Saha, S., Glauser, L., et al. (2014). A Parkinson's disease gene regulatory network identifies the signaling protein RGS2 as a modulator of LRRK2 activity and neuronal toxicity. Hum Mol Genet.

Egozcue, S., Blanco, J., Vidal, F., Egozcue, J. (2002). Diploid sperm and the origin of triploidy. Hum Reprod 17, 5-7.

Ehlers, A., Beck, S., Forbes, S. A., Trowsdale, J., Volz, A., Younger, R., Ziegler, A. (2000). MHC-linked olfactory receptor loci exhibit polymorphism and contribute to extended HLA/OR-haplotypes. Genome Res 10, 1968-1978.

Eichinger, L., Pachebat, J. A., Glöckner, G., Rajandream, M. A., Sucgang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., et al. (2005). The genome of the social amoeba Dictyostelium discoideum. Nature 435, 43-57.

Endale, M., Kim, S. D., Lee, W. M., Kim, S., Suk, K., Cho, J. Y., Park, H. J., Wagley, Y., Kim, S., et al. (2010). Ischemia induces regulator of G protein signaling 2 (RGS2) protein upregulation and enhances apoptosis in astrocytes. Am J Physiol Cell Physiol 298, C611-623.

Ermolaeva, M. A., Segref, A., Dakhovnik, A., Ou, H. L., Schneider, J. I., Utermöhlen, O., Hoppe, T., Schumacher, B. (2013). DNA damage in germ cells induces an innate immune response that triggers systemic stress resistance. Nature 501, 416-420.

Ernst, O. P., Lodowski, D. T., Elstner, M., Hegemann, P., Brown, L. S., Kandori, H. (2014). Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. Chem Rev 114, 126-163.

Escriva, H., Manzon, L., Youson, J., Laudet, V. (2002). Analysis of lamprey and hagfish genes reveals a complex history of gene duplications during early vertebrate evolution. Mol Biol Evol 19, 1440-1450.

Estes, J. D., Thacker, T. C., Hampton, D. L., Kell, S. A., Keele, B. F., Palenske, E. A., Druey, K. M., Burton, G. F. (2004). Follicular dendritic cell regulation of CXCR4-mediated germinal center CD4 T cell migration. J Immunol 173, 6169-6178.

Fang, N. N., Ng, A. H., Measday, V., Mayor, T. (2011). Hul5 HECT ubiquitin ligase plays a major role in the ubiquitylation and turnover of cytosolic misfolded proteins. Nat Cell Biol 13, 1344-1352.

Feng, J. M., Xiong, J., Zhang, J. Y., Yang, Y. L., Yao, B., Zhou, Z. G., Miao, W. (2014). New phylogenomic and comparative analyses provide corroborating evidence that Myxozoa is Cnidaria. Mol Phylogenet Evol.

Ferguson, B. J., Mansur, D. S., Peters, N. E., Ren, H., Smith, G. L. (2012). DNA-PK is a DNA sensor for IRF-3-dependent innate immunity. Elife 1, e00047.

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Shanmugam, D., Roos, D. S., Stoeckert, C. J., Jr. (2011). Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr Protoc Bioinformatics Chapter 6, Unit 6 12 11-19.

Flajnik, M. F., Kasahara, M. (2001). Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. Immunity 15, 351 362.

Flajnik, M. F., Du Pasquier, L. (2004). Evolution of innate and adaptive immunity: can we draw a line? Trends Immunol 25, 640-644.

Flajnik, M. F., Kasahara, M. (2010). Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet 11, 47 59.

Flajnik, M. F., Tlapakova, T., Criscitiello, M. F., Krylov, V., Ohta, Y. (2012). Evolution of the B7 family: co-evolution of B7H6 and NKp30, identification of a new

B7 family member, B7H7, and of B7's historical relationship with the MHC. Immunogenetics 64, 571-590.

Fong, C. W., Zhang, Y., Neo, S. Y., Lin, S. C. (2000). Specific induction of RGS16 (regulator of G-protein signalling 16) mRNA by protein kinase C in CEM leukaemia cells is mediated via tumour necrosis factor alpha in a calcium-sensitive manner. Biochem J 352 Pt 3, 747-753.

Francis, N. J., McNicholas, B., Awan, A., Waldron, M., Reddan, D., Sadlier, D., Kavanagh, D., Strain, L., Marchbank, K. J., et al. (2012). A novel hybrid CFH/CFHR3 gene generated by a microhomology-mediated deletion in familial atypical hemolytic uremic syndrome. Blood 119, 591-601.

Franco, R., Casadó, V., Cortés, A., Ferrada, C., Mallol, J., Woods, A., Lluis, C., Canela, E. I., Ferré, S. (2007). Basic concepts in G-protein-coupled receptor homo- and heterodimerization. ScientificWorldJournal 7, 48-57.

Fredriksson, R., Lagerström, M. C., Lundin, L. G., Schiöth, H. B. (2003). The G protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol 63, 1256 1272.

Freson, K., Stolarz, K., Aerts, R., Brand, E., Brand-Herrmann, S. M., Kawecka-Jaszcz, K., Kuznetsova, T., Tikhonoff, V., Thijs, L., et al. (2007). 391 C to G substitution in the regulator of G-protein signalling-2 promoter increases susceptibility to the metabolic syndrome in white European men: consistency between molecular and epidemiological studies. J Hypertens 25, 117-125.

Frevel, M. A., Bakheet, T., Silva, A. M., Hissong, J. G., Khabar, K. S., Williams, B. R. (2003). p38 Mitogen-activated protein kinase-dependent and independent signaling of mRNA stability of aU-rich element-containing transcripts. Mol Cell Biol 23, 425-436.

Fryns, J. P., van de Kerckhove, A., Goddeeris, P., van den Berghe, H. (1977). Unusually long survival in a case of full triploidy of maternal origin. Hum Genet 38, 147-155.

Fugmann, S. D., Messier, C., Novack, L. A., Cameron, R. A., Rast, J. P. (2006). An ancient evolutionary origin of the Rag1/2 gene locus. Proc Natl Acad Sci U S A 103, 3728-3733.

Fujii, S., Yamazoe, G., Itoh, M., Kubo, Y., Saitoh, O. (2008). Spinophilin inhibits the binding of RGS8 to M1-mAChR but enhances the regulatory function of RGS8. Biochem Biophys Res Commun 377, 200-204.

Fukuhara, S., Marinissen, M. J., Chiariello, M., Gutkind, J. S. (2000). Signaling from G protein-coupled receptors to ERK5/Big MAPK 1 involves Galpha q and Galpha 12/13 families of heterotrimeric G proteins. Evidence for the existence of a novel Ras and Rho independent pathway. J Biol Chem 275, 21730-21736.

Furlong, R. F., Holland, P. W. (2002). Were vertebrates octoploid? Philos Trans R Soc Lond B Biol Sci 357, 531-544.

Galhardo, R. S., Hastings, P. J., Rosenberg, S. M. (2007). Mutation as a stress response and the regulation of evolvability. Crit Rev Biochem Mol Biol 42, 399-435.

Gantress, J., Maniero, G. D., Cohen, N., Robert, J. (2003). Development and characterization of a model system to study amphibian immune responses to iridoviruses. Virology 311, 254-262.

Garcia-Fernández, J., Holland, P. W. (1994). Archetypal organization of the amphioxus Hox gene cluster. Nature 370, 563-566.

Gasior, S. L., Wakeman, T. P., Xu, B., Deininger, P. L. (2006). The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol 357, 1383-1393.

Gat-Viks, I., Chevrier, N., Wilentzik, R., Eisenhaure, T., Raychowdhury, R., Steuerman, Y., Shalek, A. K., Hacohen, N., Amit, I., Regev, A. (2013). Deciphering molecular circuits from genetic variation underlying transcriptional responsiveness to stimuli. Nat Biotechnol 31, 342-349.

Gauthier, M. E., Du Pasquier, L., Degnan, B. M. (2010). The genome of the sponge amphimedon queenslandica provides new perspectives into the origin of Toll-like and interleukin 1 receptor pathways. Evol Dev 12, 519 533.

Gentric, G., Desdouets, C. (2014). Polyploidization in liver tissue. Am J Pathol 184, 322-331.

Gibson, T. J., Spring, J. (2000). Evidence in favour of ancient octaploidy in the vertebrate genome. Biochem Soc Trans 28, 259-264.

Giorelli, M., Livrea, P., Defazio, G., Iacovelli, L., Capobianco, L., Picascia, A., Sallese, M., Martino, D., Aniello, M. S., et al. (2002). Interferon beta-1a counteracts effects of activation on the expression of G protein-coupled receptor kinases 2 and 3, beta-arrestin-1, and regulators of G protein signalling 2 and 16 in human mononuclear leukocytes. Cell Signal 14, 673-678.

Glover, T. W., Arlt, M. F., Casper, A. M., Durkin, S. G. (2005). Mechanisms of common fragile site instability. Hum Mol Genet 14 Spec No. 2, R197-205.

Goc, A., Al-Husein, B., Kochuparambil, S. T., Liu, J., Heston, W. W., Somanath, P. R. (2011). PI3 kinase integrates akt and MAP kinase signaling pathways in the regulation of prostate cancer. Int J Oncol 38, 267-277.

Gold, S. J., Ni, Y. G., Dohlman, H. G., Nestler, E. J. (1997). Regulators of G protein signaling (RGS) proteins: region-specific expression of nine subtypes in rat brain. J Neurosci 17, 8024-8037.

Gout, J. F., Duret, L., Kahn, D. (2009). Differential retention of metabolic genes following whole-genome duplication. Mol Biol Evol 26, 1067-1072.

Grafstein-Dunn, E., Young, K. H., Cockett, M. I., Khawaja, X. Z. (2001). Regional distribution of regulators of G-protein signaling (RGS) 1, 2, 13, 14, 16, and GAIP messenger ribonucleic acids by in situ hybridization in rat brain. Brain Res Mol Brain Res 88, 113-123.

Greenbaum, L., Strous, R. D., Kanyas, K., Merbl, Y., Horowitz, A., Karni, O., Katz, E., Kotler, M., Olender, T., et al. (2007). Association of the RGS2 gene with extrapyramidal symptoms induced by treatment with antipsychotic medication. Pharmacogenet Genomics 17, 519-528.

Gruen, J. R., Weissman, S. M. (1997). Evolving views of the major histocompatibility complex. Blood 90, 4252-4265.

Gu, J., Wei, Y. L., Wei, J. C., Huang, F., Jan, M. S., Centola, M., Frank, M. B., Yu, D. (2009). Identification of RGS1 as a candidate biomarker for undifferentiated spondylarthritis by genome-wide expression profiling and real-time polymerase chain reaction. Arthritis Rheum 60, 3269-3279.

Gu, W., Zhang, F., Lupski, J. R. (2008). Mechanisms for human genomic rearrangements. Pathogenetics 1, 4.

Guc-Scekic, M., Milasin, J., Stevanovic, M., Stojanov, L. J., Djordjevic, M. (2002). Tetraploidy in a 26-month-old girl (cytogenetic and molecular studies). Clin Genet 61, 62-65.

Guha, M., Mackman, N. (2001). LPS induction of gene expression in human monocytes. Cell Signal 13, 85-94.

Günzl, P., Bauer, K., Hainzl, E., Matt, U., Dillinger, B., Mahr, B., Knapp, S., Binder, B. R., Schabbauer, G. (2010). Anti-inflammatory properties of the PI3K pathway are mediated by IL-10/DUSP regulation. J Leukoc Biol 88, 1259-1269.

Gurevich, E. V., Tesmer, J. J., Mushegian, A., Gurevich, V. V. (2012). G protein coupled receptor kinases: more than just kinases and not only for GPCRs. Pharmacol Ther 133, 40-69.

Gurley, S. B., Griffiths, R. C., Mendelsohn, M. E., Karas, R. H., Coffman, T. M. (2010). Renal actions of RGS2 control blood pressure. J am Soc Nephrol 21, 1847-1851.

Haase, D., Roth, O., Kalbe, M., Schmiedeskamp, G., Scharsack, J. P., Rosenstiel, P., Reusch, T. B. (2013). Absence of major histocompatibility complex class II mediated immunity in pipefish, Syngnathus typhle: evidence from deep transcriptome sequencing. Biol Lett 9, 20130044.

Hallböök, F., Lundin, L. G., Kullander, K. (1998). Lampetra fluviatilis neurotrophin homolog, descendant of a neurotrophin ancestor, discloses the early molecular evolution of neurotrophins in the vertebrate subphylum. J Neurosci 18, 8700-8711.

Hansen, J. D., Strassburger, P., Thorgaard, G. H., Young, W. P., Du Pasquier, L. (1999). Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, Oncorhynchus mykiss. J Immunol 163, 774-786.

Hansen, J. D., Du Pasquier, L., Lefranc, M. P., Lopez, V., Benmansour, A., Boudinot, P. (2009). The B7 family of immunoregulatory receptors: a comparative and evolutionary perspective. Mol Immunol 46, 457-472.

Hansen, K., Varvas, K., Järving, I., Samel, N. (2014). Novel membrane associated prostaglandin E synthase-2 from crustacean arthropods. Comp Biochem Physiol B Biochem Mol Biol 174, 45-52.

Harrison, L. M., He, Y. (2011). Rhes and aGS1/Dexras1 affect signaling by dopamine D1 receptors through adenylyl cyclase. J Neurosci Res 89, 874-882.

Hau, P. M., Siu, W. Y., Wong, N., Lai, P. B., Poon, R. Y. (2006). Polyploidization increases the sensitivity to DNA-damaging agents in mammalian cells. FEBS Lett 580, 4727-4736.

Hayasaka, N., Aoki, K., Kinoshita, S., Yamaguchi, S., Wakefield, J. K., Tsuji Kawahara, S., Horikawa, K., Ikegami, H., Wakana, S., et al. (2011). Attenuated food anticipatory activity and abnormal circadian locomotor rhythms in Rgs16 knockdown mice. PLoS ONE 6, e17655.

Helleday, T., Lo, J., van Gent, D. C., Engelward, B. P. (2007). DNA double strand break repair: from mechanistic understanding to cancer treatment. DNA Repair (Amst) 6, 923-935.

Hendriks-Balk, M. C., Hajji, N., van Loenen, P. B., Michel, M. C., Peters, S. L., Alewijnse, A. E. (2009a). Sphingosine-1-phosphate regulates RGS2 and RGS16 mRNA expression in vascular smooth muscle cells. Eur J Pharmacol 606, 25-31.

Hendriks-Balk, M. C., Tjon-Atsoi, M., Hajji, N., Alewijnse, A. E., Peters, S. L. M. (2009b). LPS differentially affects vasoconstrictor responses: a potential role for RGS16? Journal of Physiology and Biochemistry 65, 71-83.

Heng, B. C., Aubel, D., Fussenegger, M. (2013). An overview of the diverse roles of G-protein coupled receptors (GPCRs) in the pathophysiology of various human diseases. Biotechnol adv 31, 1676-1694.

Henry, J., Ribouchon, M. T., Offer, C., Pontarotti, P. (1997). B30.2-like domain proteins: a growing family. Biochem Biophys Res Commun 235, 162 165.

Hercule, H. C., Tank, J., Plehm, R., Wellner, M., da Costa Goncalves, A. C., Gollasch, M., Diedrich, A., Jordan, J., Luft, F. C., Gross, V. (2007). Regulator of G

protein signalling 2 ameliorates angiotensin II-induced hypertension in mice. Exp Physiol 92, 1014-1022.

Herren, B., Burrows, P. D. (2002). B cell-restricted human mb-1 gene. Immunol Res 26, 35-43.

Hertel, J., de Jong, D., Marz, M., Rose, D., Tafer, H., Tanzer, A., Schierwater, B., Stadler, P. F. (2009). Non-coding RNA annotation of the genome of Trichoplax adhaerens. Nucleic Acids Res 37, 1602-1615.

Heximer, S. P., Cristillo, A. D., Forsdyke, D. R. (1997). Comparison of mRNA expression of two regulators of G-protein signaling, RGS1/BL34/1R20 and RGS2/G0S8, in cultured human blood mononuclear cells. DNA Cell Biol 16, 589-598.

Heximer, S. P., Knutsen, R. H., Sun, X., Kaltenbronn, K. M., Rhee, M. H., Peng, N., Oliveira-dos-Santos, A., Penninger, J. M., Muslin, A. J., et al. (2003). Hypertension and prolonged vasoconstrictor signaling in RGS2-deficient mice. J Clin Invest 111, 1259.

Hidalgo, A., Monroy, A., Arana, R. M., Taja, L., Vázquez, G., Salcedo, M. (2003). Chromosomal imbalances in four new uterine cervix carcinoma derived cell lines. BMC Cancer 3, 8.

Hilt, W., Wolf, D. H. (1996). Proteasomes: destruction as a programme. Trends Biochem Sci 21, 96-102.

Hinsch, H., Hannenhalli, S. (2006). Recurring genomic breaks in independent lineages support genomic fragility. BMC Evol Biol 6, 90.

Hirao, K., Natsuka, Y., Tamura, T., Wada, I., Morito, D., Natsuka, S., Romero, P., Sleno, B., Tremblay, L. O., et al. (2006). EDEM3, a soluble EDEM homolog, enhances glycoprotein endoplasmic reticulum-associated degradation and mannose trimming. J Biol Chem 281, 9650-9658.

Hochdörfer, T., Kuhny, M., Zorn, C. N., Hendriks, R. W., Vanhaesebroeck, B., Bohnacker, T., Krystal, G., Huber, M. (2011). Activation of the PI3K pathway increases TLR-induced TNF-alpha and IL-6 but reduces IL-1beta production in mast cells. Cell Signal 23, 866-875.

Hokamp, K., McLysaght, A., Wolfe, K. H. (2003). The 2R hypothesis and the human genome sequence. J Struct Funct Genomics 3, 95-110.

Holland, L. Z., Rached, L. A., Tamme, R., Holland, N. D., Kortschak, D., Inoko, H., Shiina, T., Burgtorf, C., Lardelli, M. (2001). Characterization and developmental expression of the amphioxus homolog of Notch (AmphiNotch): evolutionary conservation of multiple expression domains in amphioxus and vertebrates. Dev Biol 232, 493-507.

Holland, L. Z., Albalat, R., Azumi, K., Benito-Gutiérrez, E., Blow, M. J., Bronner Fraser, M., Brunet, F., Butts, T., Candiani, S., et al. (2008). The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res 18, 1100-1111.

Holland, P. W., Garcia-Fernández, J., Williams, N. A., Sidow, A. (1994). Gene duplications and the origins of vertebrate development. Dev Suppl, 125 133.

Holland, P. W. (2003). More genes in vertebrates? J Struct Funct Genomics 3, 75-84.

Hong, J. X., Wilson, G. L., Fox, C. H., Kehrl, J. H. (1993). Isolation and characterization of a novel B cell activation gene. J Immunol 150, 3895-3904.

Horisberger, M. A. (1995). Interferons, Mx genes, and resistance to influenza virus. Am J Respir Crit Care Med 152, S67-71.

Horton, R., Wilming, L., Rand, V., Lovering, R. C., Bruford, E. A., Khodiyar, V. K., Lush, M. J., Povey, S., Talbot, C. C., Jr., et al. (2004). Gene map of the extended human MHC. Nat Rev Genet 5, 889-899.

Hsu, E. (2009). V(D)J recombination: of mice and sharks. Adv Exp Med Biol 650, 166-179.

Hsu, H. C., Yang, P., Wang, J., Wu, Q., Myers, R., Chen, J., Yi, J., Guentert, T., Tousson, A., et al. (2008). Interleukin 17-producing T helper cells and interleukin 17 orchestrate autoreactive germinal center development in autoimmune BXD2 mice. Nat Immunol 9, 166-175.

Hu, R. G., Sheng, J., Qi, X., Xu, Z., Takahashi, T. T., Varshavsky, A. (2005). The N end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. Nature 437, 981-986.

Hu, Y., Xing, J., Chen, L., Guo, X., Du, Y., Zhao, C., Zhu, Y., Lin, M., Zhou, Z., Sha, J. (2008). RGS22, a novel testis-specific regulator of G-protein signaling involved in human and mouse spermiogenesis along with GNA12/13 subunits. Biol Reprod 79, 1021-1029.

Huang, J., Pashkov, V., Kurrasch, D. M., Yu, K., Gold, S. J., Wilkie, T. M. (2006). Feeding and fasting controls liver expression of a regulator of G protein signaling (Rgs16) in periportal hepatocytes. Comp Hepatol 5, 8.

Hufton, A. L., Groth, D., Vingron, M., Lehrach, H., Poustka, A. J., Panopoulou, G. (2008). Early vertebrate whole genome duplications were predated by a period of intense genome rearrangement. Genome Res 18, 1582 1591.

Hughes, A. L. (1998). Phylogenetic tests of the hypothesis of block duplication of homologous genes on human chromosomes 6, 9, and 1. Mol Biol Evol 15, 854-870.

Hughes, A. L. (1999). Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. J Mol Evol 48, 565-576.

Hughes, A. L., Friedman, R. (2003). 2R or not 2R: testing hypotheses of genome duplication in early vertebrates. J Struct Funct Genomics 3, 85-93.

Hunt, K. A., Zhernakova, A., Turner, G., Heap, G. A., Franke, L., Bruinenberg, M., Romanos, J., Dinesen, L. C., Ryan, A. W., et al. (2008). Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet 40, 395-402.

Hwang, I. Y., Hwang, K. S., Park, C., Harrison, K. A., Kehrl, J. H. (2013). Rgs13 constrains early B cell responses and limits germinal center sizes. PLoS One 8, e60139.

Iliopoulos, D., Vassiliou, G., Sekerli, E., Sidiropoulou, V., Tsiga, A., Dimopoulou, D., Voyiatzis, N. (2005). Long survival in a 69,XXX triploid infant in Greece. Genet Mol Res 4, 755-759.

Inngjerdingen, M., Rolstad, B., Ryan, J. C. (2003). Activating and inhibitory Ly49 receptors modulate NK cell chemotaxis to CXC chemokine ligand (CXCL) 10 and CXCL12. J Immunol 171, 2889-2895.

Inoue-Mochita, M., Inoue, T., Epstein, D. L., Blumer, K. J., Rao, P. V. (2009). RGS2-deficient mice exhibit decreased intraocular pressure and increased retinal ganglion cell survival. Mol Vis 15, 495-504.

International Multiple Sclerosis Genetics Consortium. (2010). IL12A, MPHOSPH9/CDK2AP1 and RGS1 are novel multiple sclerosis susceptibility loci. Genes Immun 11, 397-405.

Irannejad, R., Tomshine, J. C., Tomshine, J. R., Chevalier, M., Mahoney, J. P., Steyaert, J., Rasmussen, S. G., Sunahara, R. K., El-Samad, H., et al. (2013). Conformational biosensors reveal GPCR signalling from endosomes. Nature 495, 534 538.

Irimia, M., Tena, J. J., Alexis, M. S., Fernandez-Minan, A., Maeso, I., Bogdanovic, O., de la Calle-Mustienes, E., Roy, S. W., Gomez-Skarmeta, J. L., Fraser, H. B. (2012).

Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res 22, 2356-2367.

Isberg, V., Vroling, B., van der Kant, R., Li, K., Vriend, G., Gloriam, D. (2014). GPCRDB: an information system for G protein-coupled receptors. Nucleic Acids Res 42, D422-425.

Islam, T. C., Asplund, A. C., Lindvall, J. M., Nygren, L., Liden, J., Kimby, E., Christensson, B., Smith, C. I., Sander, B. (2003). High level of cannabinoid receptor 1, absence of regulator of G protein signalling 13 and differential expression of Cyclin D1 in mantle cell lymphoma. Leukemia 17, 1880-1890.

Iwai, K., Koike, M., Ohshima, S., Miyatake, K., Uchiyama, Y., Saeki, Y., Ishii, M. (2007). RGS18 acts as a negative regulator of osteoclastogenesis by modulating the acid sensing OGR1/NFAT signaling pathway. J Bone Miner Res 22, 1612-1620.

Iwaki, S., Lu, Y., Xie, Z., Druey, K. M. (2011). p53 negatively regulates RGS13 protein expression in immune cells. J Biol Chem 286, 22219-22226.

Jackson, S. P., Durocher, D. (2013). Regulation of DNA damage responses by ubiquitin and SUMO. Mol Cell 49, 795-807.

Jain, A., Bacolla, A., Del Mundo, I. M., Zhao, J., Wang, G., Vasquez, K. M. (2013). DHX9 helicase is involved in preventing genomic instability induced by alternatively structured DNA in human cells. Nucleic Acids Res 41, 10345 10357.

Jang, H. S., Kim, J. I., Noh, M., Rhee, M. H., Park, K. M. (2014). Regulator of G protein signaling 2 (RGS2) deficiency accelerates the progression of kidney fibrosis. Biochim Biophys Acta 1842, 1733-1741.

Järving, R., Järving, I., Kurg, R., Brash, A. R., Samel, N. (2004). On the evolutionary origin of cyclooxygenase (COX) isozymes: characterization of marine invertebrate COX genes points to independent duplication events in vertebrate and invertebrate lineages. J Biol Chem 279, 13624-13633.

Jarvis, A., Sharma, P., Watson, N., Smith, A. (1999). Two children with acute lymphoblastic leukemia and "jumping" translocations: both involve 1q23 as the donor breakpoint. Cancer Genet Cytogenet 114, 112-116.

Jefferson, A., Colella, S., Moralli, D., Wilson, N., Yusuf, M., Gimelli, G., Ragoussis, J., Volpi, E. V. (2010). Altered intra-nuclear organisation of heterochromatin and genes in ICF syndrome. PLoS One 5, e11364.

Jékely, G. (2013). Global view of the evolution and diversity of metazoan neuropeptide signaling. Proc Natl Acad Sci U S A 110, 8702-8707.

Jin, H., Li, W., Yang, R., Ogasawara, A., Lu, H., Paoni, N. F. (2005). Inhibitory effects of interferon-gamma on myocardial hypertrophy. Cytokine 31, 405-414.

Jin, X. K., Li, W. W., Wu, M. H., Guo, X. N., Li, S., Yu, A. Q., Zhu, Y. T., He, L., Wang, Q. (2013). Immunoglobulin superfamily protein Dscam exhibited molecular diversity by alternative splicing in hemocytes of crustacean, Eriocheir sinensis. Fish Shellfish Immunol 35, 900-909.

Johanneson, B., Lima, G., von Salomé, J., Alarcón-Segovia, D., Alarcón Riquelme, M. E., Collaborative Group on the Genetics of SLE, the Biomed II Collaboration on the Genetics of SLE and Sjögrens syndrome. (2002). A major susceptibility locus for systemic lupus erythemathosus maps to chromosome 1q31. Am J Hum Genet 71, 1060 1071.

Johnson, B. A., Wang, J., Taylor, E. M., Caillier, S. J., Herbert, J., Khan, O. A., Cross, A. H., De Jager, P. L., Gourraud, P. A., et al. (2010). Multiple sclerosis susceptibility alleles in african americans. Genes Immun 11, 343-350.

Johnson, E. N., Druey, K. M. (2002). Functional characterization of the G protein regulator RGS13. J Biol Chem 277, 16768-16774.

Józsi, M., Meri, S. (2014). Factor H-related proteins. Methods Mol Biol 1100, 225 236.

Kach, J., Sethakorn, N., Dulin, N. O. (2012). A finer tuning of G protein signaling through regulated control of RGS proteins. Am J Physiol Heart Circ Physiol 303, H19 35.

Kaer, K., Speek, M. (2013). Retroelements in human disease. Gene 518, 231 241.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. Genome Res 20, 1313-1326.

Kapitonov, V. V., Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol 3, e181.

Kärdi, L. (2014). Testing regulator of G protein signaling 16 as a potential new biomarker in patients with Multiple Sclerosis. Master's Thesis. Tallinn University of Technology.

Kasahara, M., Hayashi, M., Tanaka, K., Inoko, H., Sugaya, K., Ikemura, T., Ishibashi, T. (1996). Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. Proc Natl Acad Sci U S A 93, 9096-9101.

Kasahara, M. (1997). New insights into the genomic organization and origin of the major histocompatibility complex: role of chromosomal (genome) duplication in the emergence of the adaptive immune system. Hereditas 127, 59 65.

Kasahara, M. (1998). What do the paralogous regions in the genome tell us about the origin of the adaptive immune system? Immunol Rev 166, 159-175.

Kasahara, M. (1999a). Genome dynamics of the major histocompatibility complex: insights from genome paralogy. Immunogenetics 50, 134-145.

Kasahara, M. (1999b). The chromosomal duplication model of the major histocompatibility complex. Immunol Rev 167, 17-32.

Kasahara, M., Suzuki, T., Pasquier, L. D. (2004). On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. Trends Immunol 25, 105-111.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., et al. (2007). The medaka draft genome and insights into vertebrate genome evolution. Nature 447, 714-719.

Kasai, R. S., Kusumi, A. (2014). Single-molecule imaging revealed dynamic GPCR dimerization. Curr Opin Cell Biol 27, 78-86.

Katsanis, N., Fitzgibbon, J., Fisher, E. M. (1996). Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. Genomics 35, 101-108.

Kaufman, J., Milne, S., Göbel, T. W., Walker, B. A., Jacob, J. P., Auffray, C., Zoorob, R., Beck, S. (1999). The chicken B locus is a minimal essential major histocompatibility complex. Nature 401, 923-925.

Kauppi, L., Jeffreys, A. J., Keeney, S. (2004). Where the crossovers are: recombination distributions in mammals. Nat Rev Genet 5, 413-424.

Kawamata, N., Sakajiri, S., Sugimoto, K. J., Isobe, Y., Kobayashi, H., Oshimi, K. (2002). A novel chromosomal translocation t(1;14)(q25;q32) in pre B acute lymphoblastic leukemia involves the LIM homeodomain protein gene, Lhx4. Oncogene 21, 4983-4991.

Kayagaki, N., Wong, M. T., Stowe, I. B., Ramani, S. R., Gonzalez, L. C., Akashi Takamura, S., Miyake, K., Zhang, J., Lee, W. P., et al. (2013). Noncanonical inflammasome activation by intracellular LPS independent of TLR4. Science 341, 1246 1249.

Kehrl, J. H., Sinnarajah, S. (2002). RGS2: a multifunctional regulator of G protein signaling. Int J Biochem Cell Biol 34, 432-438.

Kerov, V. S., Natochin, M., Artemyev, N. O. (2005). Interaction of transducin-alpha with LGN, a G-protein modulator expressed in photoreceptor cells. Mol Cell Neurosci 28, 485-495.

Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A. Z., Engstrom, P. G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., et al. (2007). Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res 17, 545-555.

Kim-Howard, X., Sun, C., Molineros, J. E., Maiti, A. K., Chandru, H., Adler, A., Wiley, G. B., Kaufman, K. M., Kottyan, L., et al. (2014). Allelic heterogeneity in NCF2 associated with systemic lupus erythematosus (SLE) susceptibility across four ethnic populations. Hum Mol Genet 23, 1656-1668.

Kim, T., Pazhoor, S., Bao, M., Zhang, Z., Hanabuchi, S., Facchinetti, V., Bover, L., Plumas, J., Chaperot, L., et al. (2010). Aspartate-glutamate-alanine histidine box motif (DEAH)/RNA helicase A helicases sense microbial DNA in human plasmacytoid dendritic cells. Proc Natl Acad Sci U S A 107, 15181-15186.

Kimple, A. J., Garland, A. L., Cohen, S. P., Setola, V., Willard, F. S., Zielinski, T., Lowery, R. G., Tarran, R., Siderovski, D. P. (2014). RGS21, a regulator of taste and mucociliary clearance? Laryngoscope 124, E56-63.

Kimple, R. J., De Vries, L., Tronchère, H., Behe, C. I., Morris, R. A., Gist Farquhar, M., Siderovski, D. P. (2001). RGS12 and RGS14 GoLoco motifs are G alpha(i) interaction sites with guanine nucleotide dissociation inhibitor activity. J Biol Chem 276, 29275-29281.

King, N., Hittinger, C. T., Carroll, S. B. (2003). Evolution of key cell signaling and adhesion protein families predates animal origins. Science 301, 361-363.

Kioussis, D., Pachnis, V. (2009). Immune and nervous systems: more than just a superficial similarity? Immunity 31, 705-710.

Kisand, K., Peterson, P., Laan, M. (2014). Lymphopenia-induced proliferation in aire-deficient mice helps to explain their autoimmunity and differences from human patients. Front Immunol 5, 51.

Kjer-Nielsen, L., Patel, O., Corbett, A. J., Le Nours, J., Meehan, B., Liu, L., Bhati, M., Chen, Z., Kostenko, L., et al. (2012). MR1 presents microbial vitamin B metabolites to MAIT cells. Nature 491, 717-723.

Kmita, M., Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. Science 301, 331-333.

Knowles, D. G., McLysaght, A. (2009). Recent de novo origin of human protein coding genes. Genome Res 19, 1752-1759.

Koh, H. R., Xing, L., Kleiman, L., Myong, S. (2014). Repetitive RNA unwinding by RNA helicase A facilitates RNA annealing. Nucleic Acids Res 42, 8556-8564.

Kohan, D. E., Rossi, N. F., Inscho, E. W., Pollock, D. M. (2011). Regulation of blood pressure and salt homeostasis by endothelin. Physiol Rev 91, 1-77.

Koren, E., Zhou, H., Cahaner, A., Heller, E. D., Pitcovski, J., Lamont, S. J. (2008). Unique co-expression of immune cell-related genes in IBDV resistant chickens indicates

the activation of specific cellular host-response mechanisms. Dev Biol (Basel) 132, 153 159.

Kosaka, T., Miyata, A., Ihara, H., Hara, S., Sugimoto, T., Takeda, O., Takahashi, E., Tanabe, T. (1994). Characterization of the human gene (PTGS2) encoding prostaglandin-endoperoxide synthase 2. Eur J Biochem 221, 889-897.

Krishnan, A., Almen, M. S., Fredriksson, R., Schiöth, H. B. (2012). The origin of GPCRs: identification of mammalian like Rhodopsin, adhesion, Glutamate and Frizzled GPCRs in fungi. PLoS One 7, e29817.

Krishnan, A., Almén, M. S., Fredriksson, R., Schiöth, H. B. (2013). Remarkable similarities between the hemichordate (Saccoglossus kowalevskii) and vertebrate GPCR repertoire. Gene 526, 122-133.

Krüger, E., Kloetzel, P. M. (2012). Immunoproteasomes at the interface of innate and adaptive immune responses: two faces of one enzyme. Curr Opin Immunol 24, 77-83.

Kumanovics, A., Takada, T., Lindahl, K. F. (2003). Genomic organization of the mammalian MHC. Annu Rev Immunol 21, 629-657.

Kuraku, S., Meyer, A., Kuratani, S. (2009). Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? Mol Biol Evol 26, 47-59.

Kuwata, H., Nakao, K., Harada, T., Matsuda, I., Aiba, A. (2007). Generation of RGS8 null mutant mice by Cre/loxP system. Kobe J Med Sci 53, 275-281.

Kveberg, L., Ryan, J. C., Rolstad, B., Inngjerdingen, M. (2005). Expression of regulator of G protein signalling proteins in natural killer cells, and their modulation by Ly49A and Ly49D. Immunology 115, 358-365.

Laan, M., Peterson, P. (2013). The many faces of aire in central tolerance. Front Immunol 4, 326.

Lagman, D., Sundström, G., Ocampo Daza, D., Abalo, X. M., Larhammar, D. (2012). Expansion of transducin subunit gene families in early vertebrate tetraploidizations. Genomics 100, 203-211.

Lagman, D., Ocampo Daza, D., Widmark, J., Abalo, X. M., Sundström, G., Larhammar, D. (2013). The vertebrate ancestral repertoire of visual opsins, transducin alpha subunits and oxytocin/vasopressin receptors was established by duplication of their shared genomic region in the two rounds of early vertebrate genome duplications. BMC Evol Biol 13, 238.

Lampropoulou, V., Calderon-Gomez, E., Roch, T., Neves, P., Shen, P., Stervbo, U., Boudinot, P., Anderton, S. M., Fillatreau, S. (2010). Suppressive functions of activated B cells in autoimmune diseases reveal the dual roles of Toll-like receptors in immunity. Immunol Rev 233, 146-161.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Larhammar, D., Lundin, L. G., Hallböök, F. (2002). The human Hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. Genome Res 12, 1910-1920.

Larminie, C., Murdock, P., Walhin, J. P., Duckworth, M., Blumer, K. J., Scheideler, M. A., Garnier, M. (2004). Selective expression of regulators of G protein signaling (RGS) in the human central nervous system. Brain Research Molecular Brain Research 122, 24-34.

Lattin, J., Zidar, D. A., Schroder, K., Kellie, S., Hume, D. A., Sweet, M. J. (2007). G-protein-coupled receptor expression, function, and signaling in macrophages. J Leukoc Biol 82, 16-32.

Le Bourhis, L., Dusseaux, M., Bohineust, A., Bessoles, S., Martin, E., Premel, V., Coré, M., Sleurs, D., Serriari, N. E., et al. (2013). MAIT cells detect and efficiently lyse bacterially-infected epithelial cells. PLoS Pathog 9, e1003681.

Lee, H. K., Yeo, S., Kim, J. S., Lee, J. G., Bae, Y. S., Lee, C., Baek, S. H. (2010). Protein kinase C-eta and phospholipase D2 pathway regulates foam cell formation via regulator of G protein signaling 2. Mol Pharmacol 78, 478-485.

Lee, J. A., Carvalho, C. M., Lupski, J. R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131, 1235-1247.

Lee, M. J., Tasaki, T., Moroi, K., An, J. Y., Kimura, S., Davydov, I. V., Kwon, Y. T. (2005). RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. Proc Natl Acad Sci U S A 102, 15030-15035.

Léonard, P. M., Adema, C. M., Zhang, S. M., Loker, E. S. (2001). Structure of two FREP genes that combine IgSF and fibrinogen domains, with comments on diversity of the FREP gene family in the snail Biomphalaria glabrata. Gene 269, 155-165.

Levasseur, A., Pontarotti, P. (2010). Was the ancestral MHC involved in innate immunity? Eur J Immunol 40, 2682-2685.

Leygraf, A., Hohoff, C., Freitag, C., Willis-Owen, S. A., Krakowitzky, P., Fritze, J., Franke, P., Bandelow, B., Fimmers, R., et al. (2006). Rgs 2 gene polymorphisms as modulators of anxiety in humans? J Neural Transm 113, 1921-1925.

Li, J., Harris, R. A., Cheung, S. W., Coarfa, C., Jeong, M., Goodell, M. A., White, L. D., Patel, A., Kang, S. H., et al. (2012). Genomic hypomethylation in the human germline associates with selective structural mutability in the human genome. PLoS Genet 8, e1002692.

Li, X., Li, J., Yang, Y., Hou, R., Liu, R., Zhao, X., Yan, X., Yin, G., An, P., et al. (2013). Differential gene expression in peripheral blood T cells from patients with psoriasis, lichen planus, and atopic dermatitis. J Am Acad Dermatol 69, e235-243.

Liang, G., Bansal, G., Xie, Z., Druey, K. M. (2009). RGS16 inhibits breast cancer cell growth by mitigating phosphatidylinositol 3-kinase signaling. J Biol Chem 284, 21719-21727.

Liao, D. (1999). Concerted evolution: molecular mechanism and biological implications. Am J Hum Genet 64, 24-30.

Lieber, M. R., Ma, Y., Pannicke, U., Schwarz, K. (2003). Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Biol 4, 712-720.

Lifschytz, T., Broner, E. C., Zozulinsky, P., Slonimsky, A., Eitan, R., Greenbaum, L., Lerer, B. (2012). Relationship between Rgs2 gene expression level and anxiety and depression-like behaviour in a mutant mouse model: serotonergic involvement. Int J Neuropsychopharmacol 15, 1307-1318.

Liggett, S. B. (2011). Phosphorylation barcoding as a mechanism of directing GPCR signaling. Sci Signal 4, pe36.

Lilley, C. E., Chaurushiya, M. S., Boutell, C., Everett, R. D., Weitzman, M. D. (2011). The intrinsic antiviral defense to incoming HSV-1 genomes includes specific DNA repair proteins and is counteracted by the viral protein ICP0. PLoS Pathog 7, e1002084.

Lin, F. C., Young, H. A. (2014). Interferons: Success in anti-viral immunotherapy. Cytokine Growth Factor Rev.

Lin, G., Shindel, A. W., Banie, L., Deng, D., Wang, G., Hayashi, N., Lin, C. S., Lue, T. F. (2009). Molecular mechanisms related to parturition-induced stress urinary incontinence. Eur Urol 55, 1213-1222.

Lindén, M., Khademi, M., Lima Bomfim, I., Piehl, F., Jagodic, M., Kockum, I., Olsson, T. (2013). Multiple sclerosis risk genotypes correlate with an elevated cerebrospinal fluid level of the suggested prognostic marker CXCL13. Mult Scler 19, 863-870.

Lippert, E., Yowe, D. L., Gonzalo, J. A., Justice, J. P., Webster, J. M., Fedyk, E. R., Hodge, M., Miller, C., Gutierrez-Ramos, J. C., et al. (2003). Role of regulator of G protein signaling 16 in inflammation-induced T lymphocyte migration and activation. J Immunol 171, 1542-1555.

Liskay, R. M., Letsou, A., Stachelek, J. L. (1987). Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics 115, 161-167.

Litman, G. W., Rast, J. P., Fugmann, S. D. (2010). The origins of vertebrate adaptive immunity. Nat Rev Immunol 10, 543-553.

Liu, H., Fu, Y., Jiang, D., Li, G., Xie, J., Cheng, J., Peng, Y., Ghabrial, S. A., Yi, X. (2010). Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. J Virol 84, 11876-11887.

Liu, T., Bohlken, A., Kuljaca, S., Lee, M., Nguyen, T., Smith, S., Cheung, B., Norris, M. D., Haber, M., et al. (2005). The retinoid anticancer signal: mechanisms of target gene regulation. Br J Cancer 93, 310-318.

Louwette, S., Labarque, V., Wittevrongel, C., Thys, C., Metz, J., Gijsbers, R., Debyser, Z., Arnout, J., Van Geet, C., Freson, K. (2012). Regulator of G-protein signaling 18 controls megakaryopoiesis and the cilia-mediated vertebrate mechanosensory system. FASEB J 26, 2125-2136.

Lukov, G. L., Hu, T., McLaughlin, J. N., Hamm, H. E., Willardson, B. M. (2005). Phosducin-like protein acts as a molecular chaperone for G protein betagamma dimer assembly. EMBO J 24, 1965-1975.

Lundin, L. G. (1993). Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics 16, 1 19.

Lundin, L. G., Larhammar, D., Hallböök, F. (2003). Numerous groups of chromosomal regional paralogies strongly indicate two genome doublings at the root of the vertebrates. J Struct Funct Genomics 3, 53-63.

Ma, P., Cierniewska, A., Signarvic, R., Cieslak, M., Kong, H., Sinnamon, A. J., Neubig, R. R., Newman, D. K., Stalker, T. J., Brass, L. F. (2012). A newly identified complex of spinophilin and the tyrosine phosphatase, SHP-1, modulates platelet activation by regulating G protein-dependent signaling. Blood 119, 1935-1945.

Mackiewicz, D., de Oliveira, P. M., Moss de Oliveira, S., Cebrat, S. (2013). Distribution of recombination hotspots in the human genome--a comparison of computer simulations with real data. PLoS One 8, e65272.

Makino, T., McLysaght, A. (2008). Interacting gene clusters and the evolution of the vertebrate immune system. Mol Biol Evol 25, 1855-1862.

Makino, T., Hokamp, K., McLysaght, A. (2009). The complex relationship of gene duplication and essentiality. Trends Genet 25, 152-155.

Makino, T., McLysaght, A. (2010). Ohnologs in the human genome are dosage balanced and frequently associated with disease. Proc Natl Acad Sci U S A 107, 9270 9274.

Makino, T., McLysaght, A. (2012). Positionally biased gene loss after whole genome duplication: evidence from human, yeast, and plant. Genome Res 22, 2427-2435.

Makino, T., McLysaght, A., Kawata, M. (2013). Genome-wide deserts for copy number variation in vertebrates. Nat Commun 4, 2283.

Marino, R., Pinto, M. R., Cotelli, F., Lamia, C. L., De Santis, R. (1998). The hsp70 protein is involved in the acquisition of gamete self-sterility in the ascidian Ciona intestinalis. Development 125, 899-907.

Mark, H. F., Wyandt, H., Pan, A., Milunsky, J. M. (2005). Constitutional partial 1q trisomy mosaicism and Wilms tumor. Cancer Genet Cytogenet 162, 166-171.

McLysaght, A., Hokamp, K., Wolfe, K. H. (2002). Extensive genomic duplication during early chordate evolution. Nat Genet 31, 200-204.

McLysaght, A., Makino, T., Grayton, H. M., Tropeano, M., Mitchell, K. J., Vassos, E., Collier, D. A. (2014). Ohnologs are overrepresented in pathogenic copy number mutations. Proc Natl Acad Sci U S A 111, 361-366.

Medjeral-Thomas, N., Malik, T. H., Patel, M. P., Toth, T., Cook, H. T., Tomson, C., Pickering, M. C. (2014). A novel CFHR5 fusion protein causes C3 glomerulopathy in a family without Cypriot ancestry. Kidney Int 85, 933-937.

Meierovics, A., Yankelevich, W. J., Cowley, S. C. (2013). MAIT cells are critical for optimal mucosal immune responses during in vivo pulmonary bacterial infection. Proc Natl Acad Sci U S A 110, E3119-3128.

Miller, M. M., Robinson, C. M., Abernathy, J., Goto, R. M., Hamilton, M. K., Zhou, H., Delany, M. E. (2014). Mapping genes to chicken microchromosome 16 and discovery of olfactory and scavenger receptor genes near the major histocompatibility complex. J Hered 105, 203-215.

Mirabeau, O., Joly, J. S. (2013). Molecular evolution of peptidergic signaling systems in bilaterians. Proc Natl Acad Sci U S A 110, E2028-2037.

Miska, K. B., Harrison, G. A., Hellman, L., Miller, R. D. (2002). The major histocompatibility complex in monotremes: an analysis of the evolution of Mhc class I genes across all three mammalian subclasses. Immunogenetics 54, 381-393.

Miyashita, A., Crystal, R. G., Hay, J. G. (1995). Identification of a 27 bp 5' flanking region element responsible for the low level constitutive expression of the human cytosolic phospholipase a2 gene. Nucleic Acids Res 23, 293-301.

Miyazaki, D., Haruki, T., Takeda, S., Sasaki, S., Yakura, K., Terasaka, Y., Komatsu, N., Yamagami, S., Touge, H., et al. (2011). Herpes simplex virus type 1 induced transcriptional networks of corneal endothelial cells indicate antigen presentation function. Invest Ophthalmol Vis Sci 52, 4282-4293.

Momen, A., Afroze, T., Sadi, A. M., Khoshbin, A., Zhang, H., Choi, J., Gu, S., Zaidi, S. H., Heximer, S. P., Husain, M. (2014). Enhanced proliferation and altered calcium handling in RGS2-deficient vascular smooth muscle cells. J Recept Signal Transduct Res, 1-8.

Moné, Y., Gourbal, B., Duval, D., Du Pasquier, L., Kieffer-Jaquinod, S., Mitta, G. (2010). A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. PLoS Negl Trop Dis 4.

Moratz, C., Harrison, K., Kehrl, J. H. (2004a). Regulation of chemokine induced lymphocyte migration by RGS proteins. Methods Enzymol 389, 15-32.

Moratz, C., Hayman, J. R., Gu, H., Kehrl, J. H. (2004b). Abnormal B-cell responses to chemokines, disturbed plasma cell localization, and distorted immune tissue architecture in Rgs1-/- mice. Mol Cell Biol 24, 5767-5775.

Morel, C., Adami, P., Musard, J. F., Duval, D., Radom, J., Jouvenot, M. (2007). Involvement of sulfhydryl oxidase QSOX1 in the protection of cells against oxidative stress-induced apoptosis. Exp Cell Res 313, 3971-3982.

Moretta, L., Mingari, M. C., Pende, D., Bottino, C., Biassoni, R., Moretta, A. (1996). The molecular basis of natural killer (NK) cell recognition and function. J Clin Immunol 16, 243-253.

Moroz, L. L., Kocot, K. M., Citarella, M. R., Dosung, S., Norekian, T. P., Povolotskaya, I. S., Grigorenko, A. P., Dailey, C., Berezikov, E., et al. (2014). The ctenophore genome and the evolutionary origins of neural systems. Nature 510, 109-114.

Morrison, E. A., Garner, S., Echaubard, P., Lesbarrères, D., Kyle, C. J., Brunetti, C. R. (2014). Complete genome analysis of a frog virus 3 (FV3) isolate and sequence comparison with isolates of differing levels of virulence. Virol J 11, 46.

Mouri, K., Hishimoto, A., Fukutake, M., Nishiguchi, N., Shirakawa, O., Maeda, K. (2010). Association study of RGS2 gene polymorphisms with panic disorder in Japanese. Kobe J Med Sci 55, E116-121.

Mowry, E. M., Carey, R. F., Blasco, M. R., Pelletier, J., Duquette, P., Villoslada, P., Malikova, I., Roger, E., Kinkel, R. P., et al. (2013). Multiple sclerosis susceptibility genes: associations with relapse severity and recovery. PLoS One 8, e75416.

Müller, W. E., Böhm, M., Grebenjuk, V. A., Skorokhod, A., Müller, I. M., Gamulin, V. (2002). Conservation of the positions of metazoan introns from sponges to humans. Gene 295, 299-309.

Murphy, W. J., Larkin, D. M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auvil, L., Beever, J. E., Chowdhary, B. P., Galibert, F., et al. (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309, 613-617.

Myers, S., Bottolo, L., Freeman, C., McVean, G., Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. Science 310, 321 324.

Nagamine, K., Peterson, P., Scott, H. S., Kudoh, J., Minoshima, S., Heino, M., Krohn, K. J., Lalioti, M. D., Mullis, P. E., et al. (1997). Positional cloning of the APECED gene. Nat Genet 17, 393-398.

Nagata, Y., Oda, M., Nakata, H., Shozaki, Y., Kozasa, T., Todokoro, K. (2001). A novel regulator of G-protein signaling bearing GAP activity for Galphai and Galphaq in megakaryocytes. Blood 97, 3051-3060.

Nakatani, Y., Takeda, H., Kohara, Y., Morishita, S. (2007). Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res 17, 1254-1265.

Niedermann, G., Grimm, R., Geier, E., Maurer, M., Realini, C., Gartmann, C., Soll, J., Omura, S., Rechsteiner, M. C., et al. (1997). Potential immunocompetence of proteolytic fragments produced by proteasomes before evolution of the vertebrate immune system. J Exp Med 186, 209-220.

Niemann-Seyde, S. C., Rehder, H., Zoll, B. (1993). A case of full triploidy (69,XXX) of paternal origin with unusually long survival time. Clin Genet 43, 79-82.

Nishimura, T., Nishida, N., Itoh, T., Komeda, T., Fukuda, Y., Ikai, I., Yamaoka, Y., Nakao, K. (2005). Discrete breakpoint mapping and shortest region of overlap of chromosome arm 1q gain and 1p loss in human hepatocellular carcinoma detected by semiquantitative microsatellite analysis. Genes Chromosomes Cancer 42, 34-43.

Nonaka, M., Namikawa, C., Kato, Y., Sasaki, M., Salter-Cid, L., Flajnik, M. F. (1997). Major histocompatibility complex gene mapping in the amphibian Xenopus implies a primordial organization. Proc Natl Acad Sci U S A 94, 5789 5791.

Nossa, C. W., Havlak, P., Yue, J. X., Lv, J., Vincent, K. Y., Brockmann, H. J., Putnam, N. H. (2014). Joint assembly and genetic mapping of the atlantic horseshoe crab genome reveals ancient whole genome duplication. Gigascience 3, 9.

Nunn, C., Zou, M. X., Sobiesiak, A. J., Roy, A. A., Kirshenbaum, L. A., Chidiac, P. (2010). RGS2 inhibits beta-adrenergic receptor-induced cardiomyocyte hypertrophy. Cell Signal 22, 1231-1239.

Nunn, C., Zhao, P., Zou, M. X., Summers, K., Guglielmo, C. G., Chidiac, P. (2011). Resistance to age-related, normal body weight gain in RGS2 deficient mice. Cell Signal 23, 1375-1386.

Obata, T., Brown, G. E., Yaffe, M. B. (2000). MAP kinase pathways activated by stress: the p38 MAPK pathway. Crit Care Med 28, N67-77.

Ohno, S., (1970). Evolution by gene duplication. Allen & Unwin, London.

Ohno, S. (1999). Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. Semin Cell Dev Biol 10, 517-522.

Ohta, Y., Okamura, K., McKinney, E. C., Bartl, S., Hashimoto, K., Flajnik, M. F. (2000). Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. Proc Natl Acad Sci U S A 97, 4712-4717.

Ohta, Y., McKinney, E. C., Criscitiello, M. F., Flajnik, M. F. (2002). Proteasome, transporter associated with antigen processing, and class I genes in the nurse shark Ginglymostoma cirratum: evidence for a stable class I region and MHC haplotype lineages. J Immunol 168, 771-781.

Ohta, Y., Powis, S. J., Lohr, R. L., Nonaka, M., Pasquier, L. D., Flajnik, M. F. (2003). Two highly divergent ancient allelic lineages of the transporter associated with antigen processing (TAP) gene in Xenopus: further evidence for co-evolution among MHC class I region genes. Eur J Immunol 33, 3017-3027.

Ohta, Y., Goetz, W., Hossain, M. Z., Nonaka, M., Flajnik, M. F. (2006). Ancestral organization of the MHC revealed in the amphibian Xenopus. J Immunol 176, 3674-3685.

Okada, K., Asai, K. (2008). Expansion of signaling genes for adaptive immune system evolution in early vertebrates. BMC Genomics 9, 218.

Olinski, R. P., Lundin, L. G., Hallböök, F. (2005). Genome duplication-driven evolution of gene families: insights from the formation of the insulin family. Ann N Y Acad Sci 1040, 426-428.

Olinski, R. P., Lundin, L. G., Hallböök, F. (2006). Conserved synteny between the Ciona genome and human paralogons identifies large duplication events in the molecular evolution of the insulin-relaxin gene family. Mol Biol Evol 23, 10-22.

Oliveira-Dos-Santos, A. J., Matsumoto, G., Snow, B. E., Bai, D., Houston, F. P., Whishaw, I. Q., Mariathasan, S., Sasaki, T., Wakeham, A., et al. (2000). Regulation of T cell activation, anxiety, and male aggression by RGS2. Proc Natl Acad Sci U S A 97, 12272-12277.

Osbourn, A. E., Field, B. (2009). Operons. Cell Mol Life Sci 66, 3755-3775.

Osei-Owusu, P., Sabharwal, R., Kaltenbronn, K. M., Rhee, M. H., Chapleau, M. W., Dietrich, H. H., Blumer, K. J. (2012). Regulator of G protein signaling 2 deficiency causes endothelial dysfunction and impaired endothelium-derived hyperpolarizing factor-mediated relaxation by dysregulating Gi/o signaling. J Biol Chem 287, 12541 12549.

Osigus, H. J., Eitel, M., Bernt, M., Donath, A., Schierwater, B. (2013). Mitogenomics at the base of Metazoa. Mol Phylogenet Evol 69, 339-351.

Otowa, T., Shimada, T., Kawamura, Y., Sugaya, N., Yoshida, E., Inoue, K., Yasuda, S., Liu, X., Minato, T., et al. (2011). Association of RGS2 variants with panic disorder in a Japanese population. Am J Med Genet B Neuropsychiatr Genet 156B, 430 434.

Ottoboni, L., Frohlich, I. Y., Lee, M., Healy, B. C., Keenan, B. T., Xia, Z., Chitnis, T., Guttmann, C. R., Khoury, S. J., et al. (2013). Clinical relevance and functional consequences of the TNFRSF1A multiple sclerosis locus. Neurology 81, 1891-1899.

Pahtma, M. (2011). Regulator of G protein signaling (RGS) 16: impact on tumor development, viral infection and cytokine expression in mice. Master's Thesis. Tallinn University of Technology.

Pan, D., Zhang, L. (2007). Quantifying the major mechanisms of recent gene duplications in the human and mouse genomes: a novel strategy to estimate gene duplication rates. Genome Biol 8, R158.

Pandit, S. K., Westendorp, B., de Bruin, A. (2013). Physiological significance of polyploidization in mammalian cells. Trends Cell Biol 23, 556 566.

Panetta, R., Guo, Y., Magder, S., Greenwood, M. T. (1999). Regulators of G protein signaling (RGS) 1 and 16 are induced in response to bacterial lipopolysaccharide and stimulate c-fos promoter expression. Biochem Biophys Res Commun 259, 550-556.

Päri, M., Kuusksalu, A., Lopp, A., Kjaer, K. H., Justesen, J., Kelve, M. (2014). Enzymatically active 2',5'-oligoadenylate synthetases are widely distributed among Metazoa, including protostome lineage. Biochimie 97, 200 209.

Park, I. K., Klug, C. A., Li, K., Jerabek, L., Li, L., Nanamori, M., Neubig, R. R., Hood, L., Weissman, I. L., Clarke, M. F. (2001). Molecular cloning and characterization of a novel regulator of G-protein signaling from mouse hematopoietic stem cells. J Biol Chem 276, 915-923.

Partridge, J. G., Puhl, H. L., 3rd, Ikeda, S. R. (2006). Phosducin and phosducin-like protein attenuate G-protein-coupled receptor-mediated inhibition of voltage-gated calcium channels in rat sympathetic neurons. Mol Pharmacol 70, 90-100.

Pashkov, V., Huang, J., Parameswara, V. K., Kedzierski, W., Kurrasch, D. M., Tall, G. G., Esser, V., Gerard, R. D., Uyeda, K., et al. (2011). Regulator of G protein signaling (RGS16) inhibits hepatic fatty acid oxidation in a carbohydrate response element-binding protein (ChREBP)-dependent manner. J Biol Chem 286, 15116-15125.

Patten, M., Bünemann, J., Thoma, B., Krämer, E., Thoenes, M., Stübe, S., Mittmann, C., Wieland, T. (2002). Endotoxin induces desensitization of cardiac endothelin-1 receptor signaling by increased expression of RGS4 and RGS16. Cardiovascular Research 53, 156-164.

Patten, M., Stübe, S., Thoma, B., Wieland, T. (2003). Interleukin-1beta mediates endotoxin- and tumor necrosis factor alpha-induced RGS16 protein expression in cultured cardiac myocytes. Naunyn Schmiedeberg's Arch Pharmacol 368, 360-365.

Pébusque, M. J., Coulier, F., Birnbaum, D., Pontarotti, P. (1998). Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. Mol Biol Evol 15, 1145-1159.

Perrier, P., Martinez, F. O., Locati, M., Bianchi, G., Nebuloni, M., Vago, G., Bazzoni, F., Sozzani, S., Allavena, P., Mantovani, A. (2004). Distinct transcriptional programs activated by interleukin-10 with or without lipopolysaccharide in dendritic cells: induction of the B cell-activating chemokine, CXC chemokine ligand 13. J Immunol 172, 7031-7042.

Pertseva, M. N., Shpakov, A. O. (2009). The prokaryotic origin and evolution of eukaryotic chemosignaling systems. Neurosci Behav Physiol 39, 793-804.

Petković, I., Konja, J., Nakić, M. (1992). Cytogenetic analysis in children with acute nonlymphocytic leukemia. Cancer Genet Cytogenet 58, 155-159.

Pfaller, C. K., Li, Z., George, C. X., Samuel, C. E. (2011). Protein kinase PKR and RNA adenosine deaminase aDAR1: new roles for old players as modulators of the interferon response. Curr Opin Immunol 23, 573-582.

Phan, N., Urano, D., Srba, M., Fischer, L., Jones, A. M. (2013). Sugar induced endocytosis of plant 7TM-RGS proteins. Plant Signal Behav 8, e22814.

Pichon, B., Vankerckhove, S., Bourrouillou, G., Duprez, L., Abramowicz, M. J. (2004). A translocation breakpoint disrupts the ASPM gene in a patient with primary microcephaly. Eur J Hum Genet 12, 419-421.

Pise-Masison, C. A., Radonovich, M., Dohoney, K., Morris, J. C., O'Mahony, D., Lee, M. J., Trepel, J., Waldmann, T. A., Janik, J. E., Brady, J. N. (2009). Gene expression profiling of aTL patients: compilation of disease-related genes and evidence for TCF4 involvement in BIRC5 gene expression and cell viability. Blood 113, 4016-4026.

Poillet, L., Pernodet, N., Boyer-Guittaut, M., Adami, P., Borg, C., Jouvenot, M., Delage-Mourroux, R., Despouy, G. (2014). QSOX1 inhibits autophagic flux in breast cancer cells. PLoS One 9, e86641.

Posner, B. A., Gilman, A. G., Harris, B. A. (1999). Regulators of G protein signaling 6 and 7. Purification of complexes with gbeta5 and assessment of their effects on G protein-mediated signaling pathways. J Biol Chem 274, 31087 31093.

Postlethwait, J. H. (2007). The zebrafish genome in context: ohnologs gone missing. J Exp Zool B Mol Dev Evol 308, 563-577.

Prinster, S. C., Hague, C., Hall, R. A. (2005). Heterodimerization of G protein-coupled receptors: specificity and functional significance. Pharmacol Rev 57, 289-298.

Prontera, P., Clerici, G., Bernardini, L., Schippa, M., Capalbo, A., Manes, I., Giuffrida, M. G., Barbieri, M. G., Ardisia, C., Donti, E. (2011). Prenatal diagnosis and molecular characterization of an interstitial 1q24.3-31.3 deletion: case report and review. Genet Couns 22, 41-48.

Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., et al. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317, 86-94.

Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064-1071.

Rast, J. P., Oliveri, P., Davidson, E. H. (2000). Conserved linkage among sea urchin homologs of genes encoded in the vertebrate MHC region. In: Kasahara, M. (Ed.), Major Histocompatibility Complex. Springer Japan, pp. 66-74.

Reiter, L. T., Hastings, P. J., Nelis, E., De Jonghe, P., Van Broeckhoven, C., Lupski, J. R. (1998). Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. Am J Hum Genet 62, 1023-1033.

Rhodes, D. A., de Bono, B., Trowsdale, J. (2005). Relationship between SPRY and B30.2 protein domains. Evolution of a component of immune defence? Immunology 116, 411-417.

Ribas, C., Penela, P., Murga, C., Salcedo, A., García-Hoz, C., Jurado Pueyo, M., Aymerich, I., Mayor, F., Jr. (2007). The G protein-coupled receptor kinase (GRK)

interactome: role of GRKs in GPCR regulation and signaling. Biochim Biophys acta 1768, 913-922.

Richter, D. J., King, N. (2013). The genomic and cellular foundations of animal origins. Annu Rev Genet 47, 509-537.

Riddle, E. L., Schwartzman, R. A., Bond, M., Insel, P. A. (2005). Multi tasking RGS proteins in the heart: the next therapeutic target? Circ Res 96, 401-411.

Riekenberg, S., Farhat, K., Debarry, J., Heine, H., Jung, G., Wiesmüller, K. H., Ulmer, A. J. (2009). Regulators of G-protein signalling are modulated by bacterial lipopeptides and lipopolysaccharide. FEBS J 276, 649 659.

Ritter, S. L., Hall, R. A. (2009). Fine-tuning of GPCR activity by receptor interacting proteins. Nat Rev Mol Cell Biol 10, 819-830.

Rivas Bejarano, J. J., Valdecantos, W. C. (2013). Psoriasis as autoinflammatory disease. Dermatol Clin 31, 445-460.

Rodriguez-Lebron, E., Liu, G., Keiser, M., Behlke, M. A., Davidson, B. L. (2013). Altered Purkinje cell miRNA expression and SCA1 pathogenesis. Neurobiol Dis 54, 456-463.

Rogers, S. L., Göbel, T. W., Viertlboeck, B. C., Milne, S., Beck, S., Kaufman, J. (2005). Characterization of the chicken C-type lectin-like receptors B-NK and B-lec suggests that the NK complex and the MHC share a common ancestral region. J Immunol 174, 3475-3483.

Romanos, J., Barisani, D., Trynka, G., Zhernakova, A., Bardella, M. T., Wijmenga, C. (2009). Six new coeliac disease loci replicated in an Italian population confirm association with coeliac disease. J Med Genet 46, 60-63.

Römpler, H., Stäubert, C., Thor, D., Schulz, A., Hofreiter, M., Schöneberg, T. (2007). G protein-coupled time travel: evolutionary aspects of GPCR research. Mol Interv 7, 17-25.

Rosenbusch, B. E. (2008). Mechanisms giving rise to triploid zygotes during assisted reproduction. Fertil Steril 90, 49-55.

Ross, E. M., Wilkie, T. M. (2000). GTPase-activating proteins for heterotrimeric G proteins: regulators of G protein signaling (RGS) and RGS-like proteins. Annu Rev Biochem 69, 795-827.

Rossi, S., Bozzali, M., Bari, M., Mori, F., Studer, V., Motta, C., Buttari, F., Cercignani, M., Gravina, P., et al. (2013). Association between a genetic variant of type 1 cannabinoid receptor and inflammatory neurodegeneration in multiple sclerosis. PLoS One 8, e82848.

Ruiz-Herrera, A., Castresana, J., Robinson, T. J. (2006). Is mammalian chromosomal evolution driven by regions of genome fragility? Genome Biol 7, R115.

Sadler, A. J., Williams, B. R. (2008). Interferon-inducible antiviral effectors. Nat Rev Immunol 8, 559-568.

Saeed, M., Suzuki, R., Watanabe, N., Masaki, T., Tomonaga, M., Muhammad, A., Kato, T., Matsuura, Y., Watanabe, H., et al. (2011). Role of the endoplasmic reticulum-associated degradation (ERAD) pathway in degradation of hepatitis C virus envelope proteins and production of virus particles. J Biol Chem 286, 37264-37273.

Saeed, S., Quintin, J., Kerstens, H. H., Rao, N. A., Aghajanirefah, A., Matarese, F., Cheng, S. C., Ratter, J., Berentsen, K., et al. (2014). Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. Science 345, 1251086.

Saitoh, O., Masuho, I., Itoh, M., Abe, H., Komori, K., Odagiri, M. (2003). Distribution of regulator of G protein signaling 8 (RGS8) protein in the cerebellum. Cerebellum 2, 154-160.

Saitoh, O., Odagiri, M. (2003). RGS8 expression in developing cerebellar Purkinje cells. Biochem Biophys Res Commun 309, 836-842.

Sambrook, J. G., Figueroa, F., Beck, S. (2005). A genome-wide survey of Major Histocompatibility Complex (MHC) genes and their paralogues in zebrafish. BMC Genomics 6, 152.

Sammut, B., Du Pasquier, L., Ducoroy, P., Laurens, V., Marcuz, A., Tournefier, A. (1999). Axolotl MHC architecture and polymorphism. Eur J Immunol 29, 2897-2907.

Saraiva, M., O'Garra, A. (2010). The regulation of IL-10 production by immune cells. Nat Rev Immunol 10, 170-181.

Saveanu, L., Fruci, D., van Endert, P. (2002). Beyond the proteasome: trimming, degradation and generation of MHC class I ligands by auxiliary proteases. Mol Immunol 39, 203-215.

Sawicki, M. W., Dimasi, N., Natarajan, K., Wang, J., Margulies, D. H., Mariuzza, R. A. (2001). Structural basis of MHC class I recognition by natural killer cell receptors. Immunol Rev 181, 52-65.

Schaid, D. J. (2004). The complex genetic epidemiology of prostate cancer. Hum Mol Genet 13 Spec No 1, R103-121.

Schierwater, B., Desalle, R. (2007). Can we ever identify the Urmetazoan? Integr Comp Biol 47, 670-676.

Schöneberg, T., Schulz, A., Biebermann, H., Hermsdorf, T., Römpler, H., Sangkuhl, K. (2004). Mutant G-protein-coupled receptors as a cause of human diseases. Pharmacol Therap 104, 173-206.

Schughart, K., Kappen, C., Ruddle, F. H. (1989). Duplication of large genomic regions during the evolution of vertebrate homeobox genes. Proc Natl Acad Sci U S A 86, 7067-7071.

Schulte-Pelkum, J., Fritzler, M., Mahler, M. (2009). Latest update on the Ro/SS-A autoantibody system. Autoimmun Rev 8, 632-637.

Sémon, M., Wolfe, K. H. (2007). Consequences of genome duplication. Curr Opin Genet Dev 17, 505-512.

Sethakorn, N., Dulin, N. O. (2013). RGS expression in cancer: oncomining the cancer microarray data. J Recept Signal Transduct Res 33, 166-171.

Shankar, S. P., Wilson, M. S., DiVietro, J. A., Mentink-Kane, M. M., Xie, Z., Wynn, T. A., Druey, K. M. (2012). RGS16 attenuates pulmonary Th2/Th17 inflammatory responses. J Immunol 188, 6347-6356.

Shenoy, S. K., Barak, L. S., Xiao, K., Ahn, S., Berthouze, M., Shukla, A. K., Luttrell, L. M., Lefkowitz, R. J. (2007). Ubiquitination of beta-arrestin links seven-transmembrane receptor endocytosis and ERK activation. J Biol Chem 282, 29549 29562.

Shi, C. S., Lee, S. B., Sinnarajah, S., Dessauer, C. W., Rhee, S. G., Kehrl, J. H. (2001). Regulator of G-protein signaling 3 (RGS3) inhibits Gbeta1gamma 2-induced inositol phosphate production, mitogen-activated protein kinase activation, and akt activation. J Biol Chem 276, 24293-24300.

Shi, G. X., Harrison, K., Wilson, G. L., Moratz, C., Kehrl, J. H. (2002). RGS13 regulates germinal center B lymphocytes responsiveness to CXC chemokine ligand (CXCL)12 and CXCL13. J Immunol 169, 2507-2515.

Shi, G. X., Harrison, K., Han, S. B., Moratz, C., Kehrl, J. H. (2004). Toll like receptor signaling alters the expression of regulator of G protein signaling proteins in dendritic cells: implications for G protein-coupled receptor signaling. J Immunol 172, 5175-5184.

Shi, L., Lee, T. (2012). Molecular diversity of Dscam and self recognition. Adv Exp Med Biol 739, 262-275.

Shibata, A., Conrad, S., Birraux, J., Geuting, V., Barton, O., Ismail, A., Kakarougkas, A., Meek, K., Taucher-Scholz, G., et al. (2011). Factors determining DNA double-strand break repair pathway choice in G2 phase. EMBO J 30, 1079-1092.

Shiina, T., Shimizu, S., Hosomichi, K., Kohara, S., Watanabe, S., Hanzawa, K., Beck, S., Kulski, J. K., Inoko, H. (2004). Comparative genomic analysis of two avian (quail and chicken) MHC regions. J Immunol 172, 6751 6763.

Shiina, T., Briles, W. E., Goto, R. M., Hosomichi, K., Yanagiya, K., Shimizu, S., Inoko, H., Miller, M. M. (2007). Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. J Immunol 178, 7162-7172.

Shiono, H., Azumi, J., Fujiwara, M., Yamazaki, H., Kikuchi, K. (1988). Tetraploidy in a 15-month-old girl. Am J Med Genet 29, 543-547.

Sibut, V., Hennequet-Antier, C., Le Bihan-Duval, E., Marthey, S., Duclos, M. J., Berri, C. (2011). Identification of differentially expressed genes in chickens differing in muscle glycogen content and meat quality. BMC Genomics 12, 112.

Siddle, H. V., Deakin, J. E., Coggill, P., Whilming, L. G., Harrow, J., Kaufman, J., Beck, S., Belov, K. (2011). The tammar wallaby major histocompatibility complex shows evidence of past genomic instability. BMC Genomics 12, 421.

Siderovski, D. P., Strockbine, B., Behe, C. I. (1999). Whither goest the RGS proteins? Crit Rev Biochem Mol Biol 34, 215-251.

Siderovski, D. P., Willard, F. S. (2005). The GAPs, GEFs, and GDIs of heterotrimeric G-protein alpha subunits. Int J Biol Sci 1, 51-66.

Sidow, A. (1996). Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev 6, 715-722.

Sierra, D. A., Gilbert, D. J., Householder, D., Grishin, N. V., Yu, K., Ukidwe, P., Barker, S. A., He, W., Wensel, T. G., et al. (2002). Evolution of the regulators of G protein signaling multigene family in mouse and human. Genomics 79, 177-185.

Simakov, O., Marletaz, F., Cho, S. J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D. H., Larsson, T., Lv, J., et al. (2013). Insights into bilaterian evolution from three spiralian genomes. Nature 493, 526-531.

Skerka, C., Chen, Q., Fremeaux-Bacchi, V., Roumenina, L. T. (2013). Complement factor H related proteins (CFHRs). Mol Immunol 56, 170-180.

Smedley, D., Sidhar, S., Birdsall, S., Bennett, D., Herlyn, M., Cooper, C., Shipley, J. (2000). Characterization of chromosome 1 abnormalities in malignant melanomas. Genes Chromosomes Cancer 28, 121-125.

Smith-Garvin, J. E., Koretzky, G. A., Jordan, M. S. (2009). T cell activation. Annu Rev Immunol 27, 591-619.

Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., Yandell, M. D., Manousaki, T., Meyer, A., et al. (2013). Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat Genet 45, 415-421, 421e411-412.

Smith, T. P., Haymond, T., Smith, S. N., Sweitzer, S. M. (2014). Evidence for the endothelin system as an emerging therapeutic target for the treatment of chronic pain. J Pain Res 7, 531-545.

Smyth, D. J., Plagnol, V., Walker, N. M., Cooper, J. D., Downes, K., Yang, J. H., Howson, J. M., Stevens, H., McManus, R., et al. (2008). Shared and distinct genetic variants in type 1 diabetes and celiac disease. New Engl J Med 359, 2767-2777.

Snow, B. E., Antonio, L., Suggs, S., Siderovski, D. P. (1998). Cloning of a retinally abundant regulator of G-protein signaling (RGS-r/RGS16): genomic structure and chromosomal localization of the human gene. Gene 206, 247-253.

Soundararajan, M., Willard, F. S., Kimple, A. J., Turnbull, A. P., Ball, L. J., Schoch, G. A., Gileadi, C., Fedorov, O. Y., Dowler, E. F., et al. (2008). Structural diversity in the RGS domain and its interaction with heterotrimeric G protein alpha-subunits. Proc Natl Acad Sci U S A 105, 6457-6462.

Srivastava, M., Begovic, E., Chapman, J., Putnam, N. H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., et al. (2008). The Trichoplax genome and the nature of placozoans. Nature 454, 955-960.

Srivastava, M., Larroux, C., Lu, D. R., Mohanty, K., Chapman, J., Degnan, B. M., Rokhsar, D. S. (2010a). Early evolution of the LIM homeobox gene family. BMC Biol 8, 4.

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E., Mitros, T., Richards, G. S., Conaco, C., Dacre, M., et al. (2010b). The Amphimedon queenslandica genome and the evolution of animal complexity. Nature 466, 720-726.

Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrom, M., Gregers, T. F., Rounge, T. B., Paulsen, J., Solbakken, M. H., et al. (2011). The genome sequence of atlantic cod reveals a unique immune system. Nature 477, 207-210.

Stefanova, I., Jenderny, J., Kaminsky, E., Mannhardt, A., Meinecke, P., Grozdanova, L., Gillessen-Kaesbach, G. (2010). Mosaic and complete tetraploidy in live-born infants: two new patients and review of the literature. Clin Dysmorphol 19, 123-127.

Steffensen, D. M., Chu, E. H., Speert, D. P., Wall, P. M., Meilinger, K., Kelch, R. P. (1977). Partial trisomy of the long arm of human chromosome 1 as demostrated by in situ hybridization with 5S ribosomal RNA. Hum Genet 36, 25-33.

Stevens, R. C., Cherezov, V., Katritch, V., Abagyan, R., Kuhn, P., Rosen, H., Wüthrich, K. (2013). The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. Nat Rev Drug Discov 12, 25-34.

Storchova, Z., Pellman, D. (2004). From polyploidy to aneuploidy, genome instability and cancer. Nat Rev Mol Cell Biol 5, 45-54.

Strand, T., Wang, B., Meyer-Lucht, Y., Höglund, J. (2013). Evolutionary history of black grouse major histocompatibility complex class IIB genes revealed through single locus sequence-based genotyping. BMC Genet 14, 29.

Stuebe, S., Wieland, T., Kraemer, E., Stritzky, A., Schroeder, D., Seekamp, S., Vogt, A., Chen, C. K., Patten, M. (2008). Sphingosine-1-phosphate and endothelin-1 induce the expression of rgs16 protein in cardiac myocytes by transcriptional activation of the rgs16 gene. Naunyn-Schmiedeberg's Arch Pharmacol 376, 363-373.

Sun, L., Ye, R. D. (2012). Role of G protein-coupled receptors in inflammation. Acta Pharmacol Sin 33, 342-350.

Takeuchi, O., Hoshino, K., Kawai, T., Sanjo, H., Takada, H., Ogawa, T., Takeda, K., Akira, S. (1999). Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. Immunity 11, 443-451.

Tanaka, K., Kasahara, M. (1998). The MHC class I ligand-generating system: roles of immunoproteasomes and the interferon-gamma-inducible proteasome activator PA28. Immunol Rev 163, 161-176.

Taylor, J. S., Van de Peer, Y., Braasch, I., Meyer, A. (2001). Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond B Biol Sci 356, 1661-1679.

Tesmer, J. J. (2009). Structure and function of regulator of G protein signaling homology domains. Prog Mol Biol Transl Sci 86, 75-113.

Thapliyal, A., Bannister, R. A., Hanks, C., Adams, B. A. (2008). The monomeric G proteins aGS1 and Rhes selectively influence Galphai-dependent signaling to modulate N-type (CaV2.2) calcium channels. Am J Physiol Cell Physiol 295, C1417 1426.

The MHC sequencing consortium. (1999). Complete sequence and gene map of a human major histocompatibility complex. Nature 401, 921-923.

Timmusk, S., Merlot, E., Lövgren, T., Järvekülg, L., Berg, M., Fossum, C. (2009). Regulator of G protein signalling 16 is a target for a porcine circovirus type 2 protein. J Gen Virol 90, 2425-2436.

Tomlinson, I. M., Cook, G. P., Walter, G., Carter, N. P., Riethman, H., Buluwela, L., Rabbitts, T. H., Winter, G. (1995). A complete map of the human immunoglobulin VH locus. Ann N Y Acad Sci 764, 43-46.

Trachtulec, Z., Forejt, J. (2001). Synteny of orthologous genes conserved in mammals, snake, fly, nematode, and fission yeast. Mamm Genome 12, 227-231.

Trowsdale, J., Knight, J. C. (2013). Major histocompatibility complex genomics and human disease. Annu Rev Genomics Hum Genet 14, 301-323.

Tsang, S., Woo, A. Y., Zhu, W., Xiao, R. P. (2010). Deregulation of RGS2 in cardiovascular diseases. Front Biosci (Schol Ed) 2, 547-557.

Tsukamoto, K., Deakin, J. E., Graves, J. A., Hashimoto, K. (2013). Exceptionally high conservation of the MHC class I-related gene, MR1, among mammals. Immunogenetics 65, 115-124.

Urano, D., Chen, J. G., Botella, J. R., Jones, A. M. (2013). Heterotrimeric G protein signalling in the plant kingdom. Open Biol 3, 120186.

van der aa, L. M., Levraud, J. P., Yahmi, M., Lauret, E., Briolat, V., Herbomel, P., Benmansour, A., Boudinot, P. (2009). A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish. BMC Biol 7, 7.

van der Velden, Y. U., Wang, L., Querol Cano, L., Haramis, A. P. (2013). The polycomb group protein ring1b/rnf2 is specifically required for craniofacial development. PLoS One 8, e73997.

Vasilatos, S. N., Katz, T. A., Oesterreich, S., Wan, Y., Davidson, N. E., Huang, Y. (2013). Crosstalk between lysine-specific demethylase 1 (LSD1) and histone deacetylases mediates antineoplastic efficacy of HDAC inhibitors in human breast cancer cells. Carcinogenesis 34, 1196-1207.

Vassilatis, D. K., Hohmann, J. G., Zeng, H., Li, F., Ranchalis, J. E., Mortrud, M. T., Brown, A., Rodriguez, S. S., Weller, J. R., et al. (2003). The G protein coupled receptor repertoires of human and mouse. Proc Natl Acad Sci U S A 100, 4903-4908.

Venkatakrishnan, A., Flock, T., Prado, D. E., Oates, M. E., Gough, J., Madan Babu, M. (2014). Structured and disordered facets of the GPCR fold. Curr Opin Struct Biol 27C, 129-137.

Vienne, A., Rasmussen, J., Abi-Rached, L., Pontarotti, P., Gilles, A. (2003a). Systematic phylogenomic evidence of en bloc duplication of the ancestral 8p11.21 8p21.3-like region. Mol Biol Evol 20, 1290-1298.

Vienne, A., Shiina, T., Abi-Rached, L., Danchin, E., Vitiello, V., Cartault, F., Inoko, H., Pontarotti, P. (2003b). Evolution of the proto-MHC ancestral region: more evidence for the plesiomorphic organisation of human chromosome 9q34 region. Immunogenetics 55, 429-436.

Vienne, A., Pontarotti, P. (2006). Metaphylogeny of 82 gene families sheds a new light on chordate evolution. Int J Biol Sci 2, 32-37.

Villarreal, L. P. (2009). The source of self: genetic parasites and the origin of adaptive immunity. Ann N Y Acad Sci 1178, 194-232.

Villarreal, L. P. (2011). Viral ancestors of antiviral systems. Viruses 3, 1933 1958.

Villasenor, A., Wang, Z. V., Rivera, L. B., Ocal, O., Asterholm, I. W., Scherer, P. E., Brekken, R. A., Cleaver, O., Wilkie, T. M. (2010). Rgs16 and Rgs8 in embryonic endocrine pancreas and mouse models of diabetes. Disease Models and Mechanisms 3, 567-580.

Vischer, H. F., Vink, C., Smit, M. J. (2006). A viral conspiracy: hijacking the chemokine system through virally encoded pirated chemokine receptors. Curr Top Microbiol Immunol 303, 121-154.

von Buchholtz, L., Elischer, A., Tareilus, E., Gouka, R., Kaiser, C., Breer, H., Conzelmann, S. (2004). RGS21 is a novel regulator of G protein signalling selectively expressed in subpopulations of taste bud cells. Eur J Neurosci 19, 1535-1544.

Waldman, A. S., Liskay, R. M. (1988). Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. Mol Cell Biol 8, 5350 5357.

Walker, B. A., Hunt, L. G., Sowa, A. K., Skjødt, K., Göbel, T. W., Lehner, P. J., Kaufman, J. (2011). The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. Proc Natl Acad Sci U S A 108, 8396-8401.

Wallau, G. L., Ortiz, M. F., Loreto, E. L. (2012). Horizontal transposon transfer in eukarya: detection, bias, and perspectives. Genome Biol Evol 4, 689 699.

Wang, B. T., Chong, T. P., Boyar, F. Z., Kopita, K. A., Ross, L. P., El Naggar, M. M., Sahoo, T., Wang, J. C., Hemmat, M., et al. (2014). Abnormalities in spontaneous abortions detected by G-banding and chromosomal microarray analysis (CMA) at a national reference laboratory. Mol Cytogenet 7, 33.

Wang, C. I., Lewis, R. J. (2013). Emerging opportunities for allosteric modulation of G-protein coupled receptors. Biochem Pharmacol 85, 153-162.

Wang, J. H., New, J. S., Xie, S., Yang, P., Wu, Q., Li, J., Luo, B., Ding, Y., Druey, K. M., et al. (2013). Extension of the germinal center stage of B cell development promotes autoantibodies in BXD2 mice. Arthritis Rheum 65, 2703 2712.

Wang, S., Zhang, L., Meyer, E., Bao, Z. (2010). Genome-wide analysis of transposable elements and tandem repeats in the compact placozoan genome. Biol Direct 5, 18.

Wang, X., Zeng, W., Soyombo, A. A., Tang, W., Ross, E. M., Barnes, A. P., Milgram, S. L., Penninger, J. M., Allen, P. B., et al. (2005). Spinophilin regulates Ca2+ signalling by binding the N-terminal domain of RGS2 and the third intracellular loop of G-protein-coupled receptors. Nat Cell Biol 7, 405-411.

Watari, K., Nakaya, M., Kurose, H. (2014). Multiple functions of G protein coupled receptor kinases. J Mol Signal 9, 1.

Wen, W., Chen, S., Cao, Y., Zhu, Y., Yamamoto, Y. (2005). HIV-1 infection initiates changes in the expression of a wide array of genes in U937 promonocytes and HUT78 T cells. Virus Res 113, 26-35.

Wertheim, B., Beukeboom, L. W., van de Zande, L. (2013). Polyploidy in animals: effects of gene expression on sex determination, evolution and ecology. Cytogenet Genome Res 140, 256-269.

Westerdahl, H., Wittzell, H., von Schantz, T. (2000). Mhc diversity in two passerine birds: no evidence for a minimal essential Mhc. Immunogenetics 52, 92-100.

Wettschureck, N., Offermanns, S. (2005). Mammalian G proteins and their cell type specific functions. Physiol Rev 85, 1159-1204.

Wiechec, E., Overgaard, J., Hansen, L. L. (2008). A fragile site within the HPC1 region at 1q25.3 affecting RGS16, RGSL1, and RGSL2 in human breast carcinomas. Genes Chromosomes Cancer 47, 766-780.

Wiechec, E., Wiuf, C., Overgaard, J., Hansen, L. L. (2011). High-resolution melting analysis for mutation screening of RGSL1, RGS16, and RGS8 in breast cancer. Cancer Epidemiology, Biomarkers and Prevention 20, 397-407.

Wiechec, E., Overgaard, J., Kjeldsen, E., Hansen, L. L. (2013). Chromosome 1q25.3 copy number alterations in primary breast cancers detected by multiplex ligation dependent probe amplification and allelic imbalance assays and its comparison with fluorescent in situ hybridization assays. Cell Oncol (Dordr) 36, 113-120.

Winkelmann, M., Pfitzer, P., Schneider, W. (1987). Significance of polyploidy in megakaryocytes and other cells in health and tumor disease. Klin Wochenschr 65, 1115 1131.

Wolfe, K. H. (2001). Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet 2, 333-341.

Wolff, T., O'Neill, R. E., Palese, P. (1998). NS1-Binding protein (NS1-BP): a novel human protein that interacts with the influenza a virus nonstructural NS1 protein is relocalized in the nuclei of infected cells. J Virol 72, 7170-7180.

Wong, H. R., Odoms, K., Sakthivel, B. (2008). Divergence of canonical danger signals: the genome-level expression patterns of human mononuclear cells subjected to heat shock or lipopolysaccharide. BMC Immunol 9, 24.

Wong, S. K. (2003). G protein selectivity is regulated by multiple intracellular regions of GPCRs. Neurosignals 12, 1-12.

Xie, S., Li, J., Wang, J. H., Wu, Q., Yang, P., Hsu, H. C., Smythies, L. E., Mountz, J. D. (2010). IL-17 activates the canonical NF-kappaB signaling pathway in autoimmune B cells of BXD2 mice to upregulate the expression of regulators of G protein signaling 16. J Immunol 184, 2289-2296.

Xie, S., Chen, Z., Wang, Q., Song, X., Zhang, L. (2014). Comparisons of gene expression in normal, lesional, and non-lesional psoriatic skin using DNA microarray techniques. Int J Dermatol.

Xie, T., Rowen, L., Aguado, B., Ahearn, M. E., Madan, A., Qin, S., Campbell, R. D., Hood, L. (2003). Analysis of the gene-dense major histocompatibility complex class III region and its comparison to mouse. Genome Res 13, 2621-2636.

Xue, C., Hsueh, Y. P., Heitman, J. (2008). Magnificent seven: roles of G protein-coupled receptors in extracellular sensing in fungi. FEMS Microbiol Rev 32, 1010-1032.

Ye, Q., He, K., Wu, S. Y., Wan, Q. H. (2012). Isolation of a 97-kb minimal essential MHC B locus from a new reverse-4D BAC library of the golden pheasant. PLoS One 7, e32154.

Yono, M., Yoshida, M., Yamamoto, Y., Imanishi, A., Fukagawa, A., Latifpour, J., Eto, M. (2010). Identification of potential therapeutic targets in hypertension-associated bladder dysfunction. BJU Int 105, 877-883.

Zhang, C., Wang, L., Liao, Q., Zhang, L., Xu, L., Chen, C., Ye, H., Xu, X., Ye, M., Duan, S. (2013). Genetic associations with hypertension: meta-analyses of six candidate genetic variants. Genet Test Mol Biomarkers 17, 736-742.

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., et al. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490, 49-54.

Zhang, Y., Neo, S. Y., Han, J., Yaw, L. P., Lin, S. C. (1999). RGS16 attenuates galphaq-dependent p38 mitogen-activated protein kinase activation by platelet-activating factor. J Biol Chem 274, 2851-2857.

Zhang, Z., Yuan, B., Lu, N., Facchinetti, V., Liu, Y. J. (2011). DHX9 pairs with IPS 1 to sense double-stranded RNA in myeloid dendritic cells. J Immunol 187, 4501 4508.

Zhang, Z., Jin, Z., Zhao, Y., Zhang, Z., Li, R., Xiao, J., Wu, J. (2014). Systematic study on G-protein couple receptor prototypes: did they really evolve from prokaryotic genes? IET Syst Biol 8, 154-161.

Zhao, H., Bourque, G. (2009). Recovering genome rearrangements in the mammalian phylogeny. Genome Res 19, 934-942.

Zhao, L., Saelao, P., Jones, C. D., Begun, D. J. (2014). Origin and spread of de novo genes in Drosophila melanogaster populations. Science 343, 769-772.

Zhao, M., Wang, Y., Shen, H., Li, C., Chen, C., Luo, Z., Wu, H. (2013). Evolution by selection, recombination, and gene duplication in MHC class I genes of two Rhacophoridae species. BMC Evol Biol 13, 113.

Zheng, L., Dai, H., Zhou, M., Li, X., Liu, C., Guo, Z., Wu, X., Wu, J., Wang, C., et al. (2012). Polyploid cells rewire DNA damage response networks to overcome replication stress-induced barriers for tumour progression. Nat Commun 3, 815.

Zhu, B., Pennack, J. A., McQuilton, P., Forero, M. G., Mizuguchi, K., Sutcliffe, B., Gu, C. J., Fenton, J. C., Hidalgo, A. (2008). Drosophila neurotrophins reveal a common mechanism for nervous system formation. PLoS Biol 6, e284.

Zucchetti, I., De Santis, R., Grusea, S., Pontarotti, P., Du Pasquier, L. (2009). Origin and evolution of the vertebrate leukocyte receptors: the lesson from tunicates. Immunogenetics 61, 463-481.

ACKNOWLEDGEMENTS

This work was mainly carried out in the Tallinn University of Technology, Department of Gene Technology. I wish to express my gratitude to my supervisor, assistant Professor Sirje Rüütel Boudinot, for all the support and opportunities she has provided me with during my doctoral studies. I am thankful to her for helping me in opening the door to the international scientific community, for the chances to both learn and teach, and for the high degree of liberty in research.

I have immensely enjoyed my time in the molecular immunology lab in no small part thanks to my co-workers who have always provided a friendly, fun and helpful environment both at work and after it. Anu Nutt, Liina Kärdi, Lilian Järvekülg, Merle Saaremäe, Merlis Pahtma, Reet Rumvolt, Regina Saar, Toomas Tiivel, Viiu Paalme, and too many others to list them all – a big thank you to all of you!

I also thank other colleagues for being a part of this road – Liivi Tiismus, Kerli Vallmann, Cecilia Sarmiento and Anne Menert are just a few of them.

I am deeply grateful to Dr. Pierre Boudinot (INRA) whose invaluable advice has helped me find new research interests, and has resulted in two publications and fruitful collaborations. I am also thankful to Dr. Boudinot for agreeing to review my thesis prior to submission, and to Dr. Richard Tamme, who was the official pre-reviewer.

My appreciation goes to Professors Jean-Marc Cavaillon (Institut Pasteur) and Louis Du Pasquier (University of Basel), to Dr. Pierre Pontarotti (CNRS), and to my other co-authors for excellent help and advice with research. I would also like to express my gratitude to Professor Cavaillon for briefly accepting me into his lab in Institut Pasteur, Paris in 2012, during which a part of the work presented in this thesis was conducted.

I thank Dr. Kirk Druey (NIAID, NIH) for the *Rgs*16 knockout mice and assistant Professor Jacques Robert (University of Rochester) for the frogs.

I am thankful to all my teachers and students during my time in the university. Special mentions go to Dr. Allan Olspert and Professor Erkki Truve (my first supervisors), and to Ksenia Kuznetsova, Liina Kärdi, Liis Loorits, Merlis Pahtma and Tanel Vinkel, whose theses I had the chance to supervise, and who were all brilliant students.

I would like to thank the "Carrot Club" (Deniss-Eduard Juganson, Kadi Tommis, Kerli Vallmann, Kristiina Remmel, Liivi Tiismus, Maris Sepp, Priit ahte and Sven Kõllamets) for being friends and sticking together for the past ten years. Our scarce meetings have always been a great source of inspiration.

Finally, my greatest gratitude goes to my family, especially my parents and my grandmother Eha Pranti, for all the unconditional love and support over the years.

SUMMARY

RGS16 and other members of the R4 RGS family of regulatory proteins are key regulators of G protein-coupled pathways in the mammalian immune system. Most of the R4 RGS genes in human are located on a relatively short genomic stretch at chromosome 1q25.3–q31.3 that contains many immune-related genes in addition to the R4 RGS, e. g. *RNaseL a*nd the gene for cyclooxygenase 2. The aforementioned RGS16 is induced by cell activation, dysregulated in cancers, corresponds to a quantitative trait locus regulating antiviral responses, and is also implicated in inflammatory responses. It is not inducible in macrophages, but interacts directly with a circovirus protein in monocytic cells.

It is generally accepted that the vertebrate genomes are in fact the products of two rounds of tetraploidization during early evolution, followed by subsequent re-diploidization. Therefore, while there are not four copies of each individual gene in our genomes, there are four copies (paralogons) to be found for most genomic regions. The R4 RGS region studied in the present work is located close to an MHC paralogon on chromosome 1 and flanked on the other side by a paralogon belonging to one of the other MHC-related paralogon sets. The MHC and related regions are central to the vertebrate adaptive immune systems as they contain gene for the antigen presenting proteins, the leukocyte receptor complex and Natural Killer cell receptors, among others. However, these regions are also involved in many innate immune responses and encode for genes involved in responses against both viral and non-viral infections.

In the present work it is shown that RGS16 restricts proinflammatory responses, likely by its previously described function in inhibiting MAP kinase activity. RGS16 overexpression in THP-1 promonocytic cells led to inhibition of the production of Pam3 activation-induced proinflammatory cytokines while siRNA-mediated knockdown of RGS16 enhanced the Pam3-induced cytokine production. The bone marrow-derived cells of *Rgs16* knockout mice responded to activation with LPS or Pam3 by producing larger amounts of CXCL1 than their wild type counterparts.

It is also shown here that viruses modulate RGS16 expression even in amphibians like the western clawed frog (*Xenopus tropicalis*) as seen from the analysis of Frog Virus 3-induced gene expression in this animal. Infection with Frog Virus 3 led to RGS16 downregulation and differential modulation of other R4 RGS region genes, suggesting that they are involved in antiviral responses in an evolutionarily distant vertebrate species.

The R4 RGS region in general was found to be conserved throughout vertebrate evolution, and linked to the proto-MHC in amphioxus. Blast searches and phylogenetic analyses revealed that 35 out 70 genes in the region have ohnologs on the MHC paralogons, and 23 genes out of these have one ohnolog on chromosome 9 in human. In the amphioxus the orthologs of R4 RGS region genes mostly mapped to the same scaffold as the proto-MHC as revealed by blast searches and phylogenetic analyses.

Finally, genes of R4 RGS region were used to track the proto-MHC in invertebrate species and successfully identified its presence in the placozoan *Trichoplax adhaerens*. This finding was further confirmed by large scale genomic comparisons, leading to the discovery of other MHC-related regions and genes from this basal-like animal as well. The methods employed here included reciprocal blast searches, synteny analyses, gene ontology analyses and analysis of non-coding sequences. The results obtained are in agreement with other works that also show *Trichoplax adhaerens* to have well conserved synteny with human.

Follow up studies for the present work will likely continue in two directions. RGS16 could be further studied experimentally both in the context of antiviral responses to determine the exact mechanisms involved in its activity, and in the context of inflammation, for example in inflammatory autoimmune disorders such as Multiple Sclerosis. In parallel, it would be interesting to continue tracking the proto-MHC and related regions, e. g. The R4 RGS region, throughout invertebrate evolution using mainly computational approaches.

KOKKUVÕTE

RGS16 ja teised R4 RGS valkude perekonna esindajad on imetajate immuunsüsteemis olulised võtmemolekulid kontrollimaks G valk-seoseliste retseptorite signaaliradasid. Enamus inimese R4 RGS valkudest paikneb lühikesel genoomsel alal kromosoomil 1q25.3-q31.3, kus lisaks R4 RGS geenide endi leidub veel teisigi immuunfunktsiooniga geene, näiteks *RNaseL* ja tsüklooksügenaas 2 geen. RGS16 tootmine indutseeritakse rakkude aktivatsioonil, see reguleerib viirusvastuses osalevate geenide avaldumist ja osaleb põletikulises immuunvastuses. RGS16 ei ole küll indutseeritav makrofaagides, kuid monotsüüdid toodavad seda vastusena infektsioonidele ning sea monotsüütides on näidatud selle seondumist ühe sigade tsirkoviiruse valguga. Lisaks on RGS16 normaalne talitlus sageli häiritud kasvajate korral.

Üldlevinud arvamuse kohaselt on selgroogsete genoomid välja kujunenud kahe kogu genoomi haarava duplikatsiooni tulemusena, millele järgnes genoomi taandumine uuesti diploidseks. Seega, ehkki meil ei ole kõiki geene neli koopiat, on meil paljud genoomi piirkonnad esindatud neljas korduses (paralogonitega). Käesolevas töös uuritud R4 RGS piirkond paikneb genoomis ühe MHC paralogoni ning ühe MHC-seoselise regiooni paralogonite vahelisel alal. MHC ning MHC-seoseliste regioonide paralogonid on selgroogsete immuunsüsteemis kesksel kohal, kuna neil asuvad nii antigeene esitlevate valkude geenid, leukotsüütide retseptorikompleks (inglise keeles Leukocyte Receptor Complex, LRC), kui ka loomulike tapjarakkude (inglise keeles 'Natural Killer') retseptorite geenid. Lisaks on samad genoomsed piirkonnad seotud kaasasündinud immuunvastusega ning kodeerivad valke, mis on vajalikud nii viiruste kui muude patogeenide vastaseks kaitseks.

Käesolevas töös on kirjeldatud RGS16 rolli põletikulises immuunvastuses, mis on tõenäoliselt vahendatud juba eelnevalt RGS16-le omistatud MAP kinaaside aktiivsuse tõkestamise funktsiooni poolt. RGS16 ületootmine THP-1 promonotsüütide rakuliinis vähendas rakkudes sünteetilise triatsüleeritud lipopeptiidi Pam3 poolt indutseeritud põletikuliste tsütokiinide tootmist. RGS16 tootmise allasurumine siRNA abil aga suurendas rakkude võimet Pam3 mõjul põletikulisi tsütokiine toota. Puuduva *Rgs16* geeniga hiirte luuüdist kasvatatud rakud tootsid bakterite endotoksiini või Pam3 mõjul rohkem tsütokiini CXCL1 kui tavalistest hiirtest eraldatud vastavad rakud.

Lisaks ilmneb käesolevast tööst väikese kannuskonna (*Xenopus tropicalis*) näitel, et viirused mõjutavad RGS16 tootmist isegi kahepaiksetes. Konnaviirus 3 poolt selles loomas põhjustatud geenide avaldumise analüüsi tulemusena selgus, et nakatunud konnades oli RGS16 tootmine vähenenud ning lisaks olid mõjutatud ka ka teised R4 RGS piirkonna geenid, millest võib järeldada nende seotust viirusvastusega evolutsiooniliselt kaugetes selgroogsetes.

Selgus, et R4 RGS piirkond on konserveerunud läbi selgroogsete evolutsiooni ja on ka selgrootus keelikloomas - süstikkalas - seotud proto-MHC piirkonnaga genoomis. Blast'i otsingud ning fülogeneetiline analüüs näitasid, et

35-l geenil 70-st antud piirkonnas asuvast geenist on genoomiduplikatsioonide tulemusena tekkinud paraloogid MHC paralogonitel, inimese genoomis on 23-l neist geenidest üks paraloogidest kromosoomil 9. Lisaks selgus Blast'i otsingutest ning fülogeneetilisest analüüsist, et süstikkalas asub enamus R4 RGS piirkonna geenidele vastavaist geenidest proto-MHC-ga ühes piirkonnas.

Järgnevalt kasutati R4 RGS piirkonna geene otsimaks selgrootute genoomidest proto-MHC piirkonda. Naastlooma *Trichoplax adhaerens* genoomist see leitigi. Tulemust kinnitati genoomide bioinformaatilise võrdluse abil, mis viis tollest väga lihtsa ehitusega loomast ka mitmete teiste MHC-seoseliste piirkondade ning nende geenide avastamiseni. Selleks kasutati järgmiseid meetodeid: vastastikused Blast otsingud, genoomipiirkondade sarnasuse võrdlus, geenide funktsioonide analüüs ning mittekodeerivate järjestuste analüüs. Saadud tulemused toetavad teisi töid, kus on samuti näidatud, et naastlooma genoomil on selgroogsete genoomidega palju sarnasusi.

Käesolevas töös saadud tulemuste põhjal on uurimistööd võimalik jätkata peamiselt kahes suunas. Esiteks tuleks RGS16 omadusi uurida katseliselt edasi nii viirusvastuses täpsete mehhanismide leidmiseks kui ka põletiku kontekstis põletikuliste autoimmuunhaiguste nagu hulgiskleroos näitel. Paralleelselt oleks huvitav jätkata bioinformaatiliselt proto-MHC ja sellega seotud regioonide – nagu näiteks R4 RGS piirkond - kaardistamisega selgrootute loomade genoomides.

CURRICULUM VITAE

Name: Jaanus Suurväli
Date and place of birth: 18.12.1984, Tallinn, Estonia
Citizenship: Estonian

Contact information

Address: Tallinn University of Technology (TUT), Department of Gene Technology, Akadeemia tee 15, 12618 Tallinn, Estonia
Tel: +372 620 4447
e-mail: jaanus.suurvali@ttu.ee

Education

2008 – 2014	PhD student, TUT, Faculty of Science, Department of Gene Technology. Thesis topic: *"Regulator of G protein Signalling 16 (RGS16): Genomic Location, Evolutionary Implications and Functions in Immunity"*. Supervisor Sirje Rüütel Boudinot.
2012 – 2014	ING.PAED.IGIP, qualification and diploma of an International Engineering Educator
2005 – 2008	MSc, TUT, Faculty of Science, Department of Gene Technology. Thesis topic: *"Mutational analysis of the Cocksfoot mottle virus ribosomal frameshifting signal"*. Supervisors Allan Olspert and Erkki Truve.
2002 – 2005	BSc, TUT, Faculty of Science, Department of Gene Technology. Thesis topic: *"Analysis of cocksfoot mottle virus cDNA clone containing GFP fused to the coat protein"*. Supervisors Allan Olspert and Erkki Truve.
1999 – 2002	Tallinn Secondary Science School
1992 – 1999	Tallinna Lasnamäe Üldgümnaasium

Language skills

Fluent	Estonian, English
Basic	Russian

Professional employment

2012 – …	TUT, Department of Gene Technology, Early Stage Researcher
2008 – 2012	TUT, Department of Gene Technology, Engineer
2008 – 2009	Competence Centre of Food and Fermentation Technologies, Researcher

Supervised theses

Ksenia Kuznetsova, BSc (2014). "Testing G protein signalling pathway components VAV1, CXCR4, RGS1 as new potential biomarkers in patients with Multiple Sclerosis" (co-supervisor Sirje Rüütel Boudinot).

Liina Kärdi, MSc (2014). "Testing RGS16 as a potential new biomarker in patients with Multiple Sclerosis" (co-supervisor Sirje Rüütel Boudinot).

Merlis Pahtma, MSc (2011). "Regulator of G protein signaling (RGS) 16: impact on tumor development, viral infection and cytokine expression in mice" (co-supervisor Sirje Rüütel Boudinot).

Tanel Vinkel, BSc (2011). "Identification of microorganisms responsible for anaerobic degradation in residual sludge of wastewater treatment plant" (co-supervisor Anne Menert).

Liis Loorits, BSc (2011). "Identification of microorganisms connected with anammox process in an anaerobic treatment system" (co-supervisor Anne Menert).

Merlis Pahtma, BSc (2009). "Investigation of human monocyte activation and RGS16 role in it" (co-supervisor Sirje Rüütel Boudinot).

Publications

Suurväli, J., Pahtma, M., Saar, R., Paalme, V., Nutt, A., Tiivel, T., Saaremäe, M., Fitting, C., Cavaillon, J.-M., Rüütel Boudinot, S. (2014). RGS16 restricts the pro-inflammatory response of monocytes. *Scandinavian Journal of Immunology*, in press. DOI: 10.1111/sji.12250

Suurväli, J., Jouneau, L., Thépot, D., Grusea, S., Pontarotti, P., Du Pasquier, L., Rüütel Boudinot, S., Boudinot, P. (2014). The Proto-MHC of Placozoans, a Region Specialized in Cellular Stress and Ubiquitination/Proteasome Pathways. The *Journal of Immunology* 193(6):2891 - 901.

Suurväli, J., Robert, J., Boudinot, P., Rüütel Boudinot, S. (2013). R4 regulators of G protein signaling (RGS) identify an ancient MHC-linked synteny group. *Immunogenetics* 65(2):145 - 56.

Saar, R., Põdersoo, D., Järvelaid, M., Tuubel, L., **Suurväli, J.,** Nutt, A., Saaremäe, M., Saar, T., Rüütel Boudinot, S. (2013). The Estonian H1N1 influenza 2009 outbreak was highly underestimated. *Proceedings of the Estonian Academy of Sciences* 61(4):320 - 329.

Tamm, T., **Suurväli, J.,** Lucchesi, J., Olspert, A., Truve, E. (2009). Stem-loop structure of Cocksfoot mottle virus RNA is indispensable for programmed −1 ribosomal frameshifting. *Virus Research* 146(1-2): 73 - 80.

ELULOOKIRJELDUS

Ees- ja perekonnanimi: Jaanus Suurväli
Sünniaeg ja -koht: 18.12.1984, Tallinn, Eesti
Kodakondsus: Eesti

Kontaktandmed

Aadress: Tallinna Tehnikaülikool (TTÜ), Geenitehnoloogia Instituut, Akadeemia tee 15, 12618 Tallinn, Eesti
Tel: +372 620 4447
e-mail: jaanus.suurvali@ttu.ee

Hariduskäik

2008 – 2014	doktorant, TTÜ, Matemaatika-loodusteaduskond, Geenitehnoloogia Instituut. Lõputöö teema: "*G valkude signaaliülekande regulaator 16 (RGS16): asukoht genoomis, evolutsiooniline tähtsus ja osalus immuunvastuses*". Juhendaja Sirje Rüütel Boudinot.
2012 – 2014	ING.PAED.IGIP, Rahvusvahelise insenerpedagoogi diplom
2005 – 2008	MSc, TTÜ, Matemaatika-loodusteaduskond, Geenitehnoloogia Instituut. Lõputöö teema: "*Keraheina laiguviiruse ribosomaalse raaminihke signaaljärjestuse mutageneesanalüüs*". Juhendajad Allan Olspert ja Erkki Truve.
2002 – 2005	BSc, TTÜ, Matemaatika-loodusteaduskond, Geenitehnoloogia Instituut. Lõputöö teema: *"GFP-kattevalgu liitvalku sisaldava keraheina laiguviiruse cDNA klooni infektsioonilisuse analüüs"*. Juhendajad Allan Olspert ja Erkki Truve.
1999 – 2002	Tallinna Reaalkool
1992 – 1999	Tallinna Lasnamäe Üldgümnaasium

Keelteoskus

Kõrgtase	Eesti keel, inglise keel
Algtase	Vene keel

Teenistuskäik

2012 – …	TTÜ, Geenitehnoloogia Instituut, nooremteadur
2008 – 2012	TTÜ, Geenitehnoloogia Instituut, insener
2008 – 2009	Toidu- ja Fermentatsioonitehnoloogia arenduskeskus, teadur

Juhendamised

Ksenia Kuznetsova, BSc (2014). "G valkude signaaliraja komponentide VAV1, CXCR4, RGS1 testimine potentsiaalsete uute biomarkeritena Sclerosis Multiplex`i patsientidel" (kaasjuhendaja Sirje Rüütel Boudinot).

Liina Kärdi, MSc (2014). "RGS16 kui võimaliku uue biomarkeri testimine Sclerosis Multiplex'i haigetel" (kaasjuhendaja Sirje Rüütel Boudinot).

Merlis Pahtma, MSc (2011). "G valkude signaliseerimise regulaator 16: mõju kasvaja arengule, viirusinfektsioonile ja tsütokiinide ekspressioonile hiirtes" (kaasjuhendaja Sirje Rüütel Boudinot).

Tanel Vinkel, BSc (2011). "Anaeroobse lagundamisega seotud mikroorganismide identifitseerimine reoveepuhasti settes" (kaasjuhendaja Anne Menert).

Liis Loorits, BSc (2011). "Anammox protsessiga seotud mikroorganismide identifitseerimine anaeroobses puhastis" (kaasjuhendaja Anne Menert).

Merlis Pahtma, BSc (2009). "Inimese monotsüütide aktivatsiooni uurimine ja selle seos RGS16 valguga" (kaasjuhendaja Sirje Rüütel Boudinot).

Publikatsioonid

Suurväli, J., Pahtma, M., Saar, R., Paalme, V., Nutt, A., Tiivel, T., Saaremäe, M., Fitting, C., Cavaillon, J.-M., Rüütel Boudinot, S. (2014). RGS16 restricts the pro-inflammatory response of monocytes. *Scandinavian Journal of Immunology*, ilmumas. DOI: 10.1111/sji.12250

Suurväli, J., Jouneau, L., Thépot, D., Grusea, S., Pontarotti, P., Du Pasquier, L., Rüütel Boudinot, S., Boudinot, P. (2014). The Proto-MHC of Placozoans, a Region Specialized in Cellular Stress and Ubiquitination/Proteasome Pathways. The *Journal of Immunology* 193(6):2891 - 901.

Suurväli, J., Robert, J., Boudinot, P., Rüütel Boudinot, S. (2013). R4 regulators of G protein signaling (RGS) identify an ancient MHC-linked synteny group. *Immunogenetics* 65(2):145 - 56.

Saar, R., Põdersoo, D., Järvelaid, M., Tuubel, L., **Suurväli, J.,** Nutt, A., Saaremäe, M., Saar, T., Rüütel Boudinot, S. (2013). The Estonian H1N1 influenza 2009 outbreak was highly underestimated. *Proceedings of the Estonian Academy of Sciences* 61(4):320 - 329.

Tamm, T., **Suurväli, J.,** Lucchesi, J., Olspert, A., Truve, E. (2009). Stem-loop structure of Cocksfoot mottle virus RNA is indispensable for programmed −1 ribosomal frameshifting. *Virus Research* 146(1-2): 73 - 80.

PUBLICATION I

RGS16 restricts the pro-inflammatory response of monocytes

Suurväli, J., Pahtma, M., Saar, R., Paalme, V., Nutt, A., Tiivel, T., Saaremäe, M., Fitting, C., Cavaillon, J.-M., Rüütel Boudinot, S.

Scandinavian Journal of Immunology (2014), in press.
DOI: 10.1111/sji.12250

Article Type: Regular Article

RGS16 restricts the pro-inflammatory response of monocytes

Jaanus Suurväli[1], Merlis Pahtma[1], Regina Saar[1], Viiu Paalme[1], Anu Nutt[1], Toomas Tiivel[1], Merle Saaremäe[1], Catherine Fitting[2], Jean-Marc Cavaillon[2], Sirje Rüütel Boudinot[1*]

Running title: RGS16 inhibits monocyte activation

Scientific heading: Basic Immunology

[1]*Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia*

[2]*Unit Cytokines & Inflammation, Institut Pasteur, Paris, France.*

[*]Corresponding author

Sirje Rüütel Boudinot,

Tallinn University of Technology

Department of Gene Technology

Akadeemia tee 15

Tallinn 12618

ESTONIA

sirje.boudinot@gmail.com

Office: +372 6204447

ABSTRACT

Immune cells express powerful and harmful effectors that require tight regulation. Heterotrimeric G proteins are critical mediators in translating extracellular signals into cell responses, which need a fine-tuned regulation for the control of cell activation. Regulator of G protein signaling 16 (RGS16) has been identified as a key factor of G protein-mediated activation in lymphocytes, modulating inflammatory and survival responses of various cell

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/sji.12250

types. However, data about the expression of this regulatory protein in monocytes are scarce, and it has remained unclear whether activation and migration of these cells is regulated by RGS16. In this study the impact of RGS16 on the production of inflammatory cytokines by activated human monocytes was investigated *in vitro* using the human promonocytic cell line THP-1 as a model. Gain and loss of function experiments showed that RGS16 overexpression reduces the expression of pro-inflammatory cytokines IL-1β, IL-6, IL-8 and TNFα while RGS16 knockdown by RNAi upregulates IL-1β, IL-6, and TNFα but not IL-8. RGS16 knockdown was also shown to enhance Pam3-mediated induction of the anti-inflammatory cytokine IL-10. Our results indicate that RGS16 restricts the activation-induced pro-inflammatory profile in myeloid cells.

INTRODUCTION

G protein-coupled receptors (GPCR) are essential for cell-to-cell communications and are targeted by 60 % of all clinically approved drugs [1]. An agonist binding to the GPCR induces conformational change in the receptor and results in the exchange of GDP to GTP on the Gα subunits of small heterotrimeric G proteins [2, 3]. The resulting separate Gα and Gβγ subunits both induce downstream signaling events; the signal is eventually terminated by hydrolysis of GTP to GDP by the intrinsic GTPase activity of Gα subunit, and by subsequent reassembly of the G protein [2-5]. GTP hydrolysis on Gα subunit can be greatly accelerated by Regulators of G Protein Signaling (RGS), a family of proteins characterised by a common conserved RGS domain, that contains a GTPase Activating Protein (GAP) activity [4, 6, 7]. RGS proteins are further divided into six subfamilies based on the overall domain structure [6, 7].

Regulator of G-protein Signaling 16 (RGS16) belongs to the B/R4 subfamily of the RGS proteins which co-evolved with the genes constituting the major histocompatibility complex and its paralogous genomic regions [8, 9]. RGS16 was initially cloned from the retina [10, 11] and its expression is highest in the brain, heart, liver and lungs [12-15]. RGS16 is also expressed by most immune cell subsets, although it is reportedly not detectable from resting macrophages [16]. We have previously demonstrated RGS16 expression in porcine and human peripheral blood mononuclear cells (PBMCs) where it is induced in response to cell activation by Toll-Like Receptor (TLR) agonists or Concanavalin A [17, 18]. Transcription

factors that control RGS16 expression include ChREBP, ERβ, NANOG, p53, RUNX2, STAT3, and YY1 [12, 19-23], as well as the circadian regulators CLOCK, BMAL, and DBP [24]; post-translational modifications are required to fully mature RGS16 into the functional stable form [25-28].

Results from *Rgs16* knockout and knockdown mice indicate that RGS16 in the brain, liver, and the immune system has a significant impact on the circadian rhythm, liver metabolism and immune cell migration, respectively [23, 24, 29-31]. RGS16 restricts T cell migration by inhibiting signaling via the G protein-coupled chemokine receptors CXCR4 and CCR4 [30, 32]. Naïve T cells are more responsive to the CXCR4 ligand CXCL12 than activated and regulatory T cells due to lower RGS16 expression [33]. RGS16 modulation of the CXCR4/CXCL12 axis has been described in many different cell types, including B and T lymphocytes and megakaryocytes [31-37]. Moreover, RGS16 is involved in GC formation [31, 34, 35] and in psoriasis [38].

Activation with TLR agonists or pro-inflammatory mediators induces RGS16 expression in a number of different cell types including cardiomyocytes and immune cells [9, 12, 16-18, 33, 35, 37, 39-46]. RGS16 siRNA reverses histone deacetylase inhibitor-induced downregulation of NF-κB protein expression [46] and has differential effects on mouse dendritic cell activation depending on which TLR has been engaged [45]. Overexpression of RGS16 has been shown to inhibit Mitogen-Activated Protein Kinase (MAPK) activity [12] and in breast cancers RGS16 is often downregulated by mutations and methylation, resulting in increased pro-survival MAPK and PI3K activity [46-49]. Moreover, concurrent downregulation of RGS16 is needed for retinoid-induced neuroblastoma growth inhibition to be efficiently countered by the siRNA of DUSP6, a negative MAPK regulator [50].

The induction of MAP kinases is only one aspect of cell activation when using TLR agonists such as the synthetic lipopeptide $Pam3CysSK_4$ (Pam3) [51]. In monocytes and other cells the production and secretion of pro-inflammatory cytokines is induced, although viral TLR agonists (e.g. double-stranded RNA) induce interferon production instead [51]. Interestingly,

treatment of monocytes with Pam3 has been also shown to eventually induce production of the anti-inflammatory/regulatory cytokine IL-10 in addition to the pro-inflammatory cytokines IL-1β, IL-6, TNFα, and IL-8 [52]. IL-10 is induced during immune responses after the early and fast pro-inflammatory cytokine burst and dampens inflammation, as discussed in [53].

We have previously shown that RGS16 in Porcine Circovirus type 2 (PCV2) - infected porcine PBMCs interacts with the viral apoptin ORF3 [17]. The genomic region surrounding *RGS16* contains many genes involved in antiviral response [9, 45]. However, RGS16 expression in both PBMCs and the promonocytic THP-1 cell line is also induced by agonists of TLR2 and TLR4, which are related to responses against bacterial infections [17, 18, 54]. RGS16 response to cell activation was similar in THP-1 cells and PBMCs confirming activated THP-1 cells as a viable model for studying RGS16 in activated PBMCs [18]; the presence of fetal bovine serum (FBS) in the cell culture media was shown to be required for optimal RGS16 induction [18]. Other studies have shown RGS16 expression in THP-1 cells to be also induced by heat shock [55] and, in a TLR2/TLR4 dependent manner, by coincubation with *Candida albicans* (alongside TNFα and IL-8) [43].

We previously proposed that RGS16 can modulate the activation-induced pro-inflammatory profile in monocytes and the human THP-1 cell line [18]. In the current study we tested and validated this hypothesis in THP-1 cells using gain and loss of function experiments. The expression of pro-inflammatory cytokines by TLR2 agonist Pam3 was investigated at mRNA and protein level in THP-1 cells in which RGS16 was overexpressed or suppressed.

MATERIALS AND METHODS

Ethics statement

Animal handling and maintenance was performed according to the interdisciplinary principles and guidelines for the use of animals in research, testing, and education (FELASA) prepared by the Ad Hoc Committee on Animal Research (The New York Academy of

Sciences, New York, N.Y). The animal experiments described in this study were authorized by the Ethical and Animal Welfare Committee of Estonia (Tartu University, ERC nr 181T-1).

Bone marrow-derived cells (BMDCs) from RGS16-/- and wild type mice

Rgs16 knockout (Rgs16-/-) mice (C57BL/6 genetic background) were a generous gift from Dr. Kirk M. Druey, M.D., NIAID, Bethesda, MD, USA. Wild type C57BL/6 mice were used as controls. Prior to cell isolation for activation experiments, mice were anesthetized with carbon dioxide and sacrificed by cervical dislocation. Bone marrow was extracted from femurs isolated from 8 weeks old mice. Cells were washed once with RPMI medium (GE Healthcare, PAA) and resuspended in RPMI containing 10 % FBS (Life Technologies). To remove differentiated cells, bone marrow was plated on 10 cm diameter dishes and non adherent cells harvested after overnight incubation. Cells were then centrifuged at 500 G for 5 minutes and resuspended in 10 ml of new medium adding 1:10 M-CSF containing supernatant (a kind gift from Dr. Simon Fillatreau, Immune regulation, DRFZ, Berlin). Cells were plated on Petri dishes, and differentiated BMDCs were yielded with 0.02 M EDTA in PBS after 7 days. BMDCs were plated to 24-well plates so that each well contained 500,000 cells in 1 ml RPMI medium (GE Healthcare, PAA) supplemented with 10 % FBS (Life Technologies). Cell activation was performed with by adding either 1 µg/ml of LPS (TLR grade lipopolysaccharide from Escherichia coli 0111:B4, Enzo Life Sciences) or 10 ng/ml of Pam3 (Pam3CysSK4, EMC Microcollections). Control cells were left untreated. Cell supernatants were collected for cytokine measurements at the timepoints 2h, 4h, 6h, 8h and 24h.

THP-1 human promonocytes

THP-1 cells were obtained from Institut Pasteur, Paris, France. Cells were cultured in RPMI 1640 GlutaMAX™ medium (Life Technologies) supplemented with 5 % FBS (Biowest) and subcultured twice per week.

RGS16 overexpression

5×10^5 THP-1 cells per well were seeded to 6-well plates in 2 mL of RPMI 1640 GlutaMAX™ medium (Life Technologies) supplemented with 5 % FBS (Biowest). Cells were transfected

with pcDNA3-RGS16 (RGS16 cloned into the pcDNA3 vector) or with an empty pcDNA3 vector (Life Technologies) using Lipofectamine® LTX Reagent (Life Technologies) according to the manufacturer's recommendations. Following an initial incubation for 4 h at 37 °C in a CO_2 incubator, cell activation was induced by adding Pam3 (EMC Microcollections) to final concentration of 100 ng/ml. Timepoints used for cell and supernatant collection were designated according to the time incubated with Pam3 as 0 h (prior to Pam3 addition), 2 h, 24 h and 48 h post activation.

siRNA transfections

All siRNA reagents were obtained from Life Technologies. THP-1 cells were seeded to 24-well plates so that each well contained 10^5 THP-1 cells in 500 µl RPMI 1640 GlutaMAX™ (Life Technologies). Cells were transfected with siPORT™ NeoFX Transfection Agent using 5 nM of either Silencer® Select siRNA specific to RGS16 (Catalog number #4392420) or 5 nM of Silencer® Select Negative Control No. 1 siRNA according to the manufacturer's recommendations. 17 h post siRNA transfection FBS was added to final concentration of 5 % and the cells were activated with Pam3 as described above. Control cells had no transfection carried out with them, but were treated similarly in regards to serum (17 h in serumfree medium, then addition of serum to final concentration of 5 %). Cells were harvested for mRNA analysis at timepoints 2 h and 24 h post activation. Cell culture medium was collected for ELISA at 24 h and 48 h post activation.

Detection of cytokine expression in THP-1 and BMDC cell supernatants

DuoSet ELISA Kits (R&D Systems) were used to analyze the cell culture supernatants for the expression of IL-1β, IL-6, IL-8, and TNFα according to the manufacturer's recommendations. When necessary, the supernatants were diluted 5 - 50 times prior to the analysis to match the detection range of the assay. For siRNA experiments expression of the anti-inflammatory cytokine IL-10 was also analyzed with a DuoSet ELISA Kit (R&D Systems). For murine BMDCs Duoset ELISA kits for mouse pro-inflammatory cytokines KC (aka CXCL1), TNFα, IL-1β and IL-6 were used. The manufacturer's protocol was followed in all cases with the slight

modification of MilliQ water replacing PBS in the wash buffer. Optical densities were measured on a Multiskan FC microplate reader (Thermo Fisher Scientific).

Construction of cDNA libraries

Total RNA was isolated from cell pellets obtained after removal of the supernatant by combining the recommended protocols for TRIzol reagent (Life Technologies) and RNeasy Mini kit (Qiagen) as described in [17]. Complementary DNA was synthesized with RevertAid™ Reverse Transcriptase (Thermo Fisher Scientific) according to the manufacturer's protocol using 200 ng of RNA per reaction as a template.

Quantitative real-time PCR

cDNA samples were analyzed by qPCR using the human primers for RGS16, IL-1β, IL-6, IL-8, IL-10, TNFα, and the housekeeping gene GAPDH (Table I). Samples were run in triplicate reactions on a 7900HT Fast Real-Time PCR System (Applied Biosystems) using 5x HOT FIREPol® EvaGreen® qPCR Mix Plus (ROX) (Solis Biodyne) at the following thermocycler conditions: 95 °C for 15 min, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Relative gene expression was determined using the $2^{-\Delta\Delta Ct}$ method by normalizing to GAPDH expression.

Statistical analysis

Statistical analyses were performed using Microsoft Excel 2010 and the software package R 3.0.2. [28]. Population distribution normality was evaluated using Shapiro-Wilk and one-sample Kolmogorov-Smirnov tests. Significances were evaluated using Welch two sample t-test for data following normal distribution (qPCR) and Mann-Whitney U test for other data (ELISA).

RESULTS

RGS16 overexpression in THP-1 cells restricts the induction of pro-inflammatory response by TLR2 agonist Pam3

To test the impact of RGS16 on human monocyte activation, THP-1 cells were transfected by an expression plasmid of human RGS16 and then incubated with Pam3, a TLR2 agonist inducing potent activation of human monocytes, including increased production of the cytokines IL-1β, IL-6, IL-8, TNFα, and IL-10 [52, 56]. Parallel control transfections with a plasmid expressing GFP showed the transfection efficiency to be 15 - 20 %. Transfection resulted in a time-dependent increase in RGS16 mRNA expression, differences between RGS16-overexpressing and control vector-transfected THP-1 cells were statistically significant at 2h and 24 h post activation with Pam3 (6 h and 28 h post transfection, respectively) (Figure 1A). RGS16 mRNA expression level increased in a time-dependent manner after Pam3 treatment in transfected as well as in control cells, indicating that the endogenous RGS16 gene was also induced by the activation.

The mRNA level expression profiles of IL-1β, IL-6, IL-8, TNFα and IL-10 showed that IL-1β, IL-6 and IL-8, but not IL-10 and TNFα, were induced at the mRNA level by Pam3 treatment (Figure S1). TNFα mRNA expression at the 48 h time point was significantly inhibited by RGS16 overexpression. However, we did not observe significant differences between RGS16-overexpressing and control vector-transfected THP-1 cells for other cytokines (Figure S1). At the protein level, testing cell supernatants by ELISA for the pro-inflammatory cytokines IL-1β, IL-6, IL-8 and TNFα revealed that Pam3 was also inducing the cells to produce more cytokines on protein level (Table II, Figure 1B-E). This could be seen from a time-dependent increase in the amounts of IL-1β, IL-6, IL-8 and TNFα in the cell culture media, which was significantly hampered in cells overexpressing RGS16. THP-1 cells transfected with the pcDNA3-RGS16 expression vector 4 hours prior to Pam3 activation produced lower amounts of pro-inflammatory cytokines as compared to THP-1 cells transfected with an empty pcDNA3 control vector (Table II, Figure 1B-E). The differences were statistically significant at 48 h post activation for IL-1β, at the timepoints 2 h and 48 h post activation for IL-8 and TNFα, and in all observed timepoints (2 h, 24 h and 48 h post activation) for IL-6 (Table II, Figure 1B-E).

Taken together, these results suggest that RGS16 restricts pro-inflammatory cytokine production by THP-1 cells on the protein level upon Pam3 activation.

RGS16 loss of function experiments confirm the results from overexpression experiments

To confirm these observations by loss of function experiments, we performed RGS16 knockdown experiments using RNA interference. THP-1 cells were pretreated with either RGS16-specific Silencer® Select siRNA or Silencer® Select Negative Control siRNA for 17 h, and then activated by Pam3. RGS16 silencing was effective and led to a statistically significant reduction (about two-fold) in Pam3 activation-induced RGS16 mRNA expression (Figure 2A). We therefore quantified pro-inflammatory cytokines by ELISA in cell supernatants harvested 24h and 48h post activation to assess the impact of RGS16 knockdown (Figure 2B-E). Expression of the anti-inflammatory cytokine IL-10 was also investigated (Figure 2F).

We confirmed in these experiments that pro-inflammatory cytokines IL-1β, IL-6, IL-8 and TNFα were induced by Pam3 treatment (Figure 2B-E). Interestingly, RGS16 knockdown led to a higher induction of IL-1β, IL-6 and TNFα in at least at one time point, as compared to THP-1 cells transfected with nonspecific control siRNA (Figure 2B, C, E). For IL-1β and IL-6 the differences were statistically significant 24h and 48h post-activation, while only at the second point for TNFα (Figure 2B, C, E). IL-8 was not significantly affected by RGS16 down-regulation (Figure 2D). IL-10 was produced in very low amounts – close to the detection limit of the ELISA kit at both timepoints; however, the amount was increased 48h post activation, which was significantly enhanced by RGS16 knockdown in THP-1 cells (Figure 2F).

To further assess the impact of RGS16 on cytokine expression in a primary culture system of myeloid cells, we used BMDCs from *Rgs16* knockout mice (Rgs16-/-) or from wild type mice. Interestingly, unstimulated BMDCs derived from Rgs16-/- mice expressed higher level of KC until 8 hours after starting culture, compared to wt-derived BMDCs. Furthermore, after stimulation with LPS or Pam3, BMDCs derived from Rgs16-/- mice express clearly higher

level of KC as shown in the kinetics experiment. In contrast, IL-1β, IL-6 and TNFα did not show significant differences of expression between Rgs16-/- and wt (Figure 3). We also analyzed the cytokine production by alveolar and peritoneal macrophages prepared from wt versus Rgs16-/- mice. Mouse alveolar and peritoneal macrophages from wt and Rgs16-/- mice did not show clear difference of production of KC, TNFα, IL-1β and IL-6 after stimulation by LPS (Figure S2).

DISCUSSION

RGS16 is an important regulator of activation which is reportedly expressed by most immune cell subsets, with the important exception of macrophages in rats [16]. RGS16 is up-regulated by cell activation in human monocytes, e.g. in the THP-1 cell line [17, 18, 43, 54, 55]. RGS16 expression in monocytes is likely involved in the control of cell migration through inhibition of G protein-coupled receptor signaling [4, 6, 7]. However, RGS16 may also modulate cytokine production during inflammatory responses [18], as it inhibits the MAPK and PI3K/Akt pathways [12, 46-50]. In this work we assessed the impact of RGS16 overexpression and knockdown on the production of pro-inflammatory cytokines by activated monocytes. To this purpose, a potent synthetic TLR2 agonist (Pam3) was used to induce monocyte activation, which leads to MAP kinase phosphorylation as well as to the production of cytokines such as IL-1β, IL-6, IL-8, TNFα, and IL-10 [52, 56, 57] [18].

Pam3-induced secretion of pro-inflammatory cytokines was significantly downregulated by RGS16 overexpression, indicating that RGS16 restricts the inflammatory response of THP-1 monocytes. This conclusion was also supported by the increased production of inflammatory cytokines by cells in which RGS16 expression has been knocked down by RNAi. Importantly, we also showed that BMDCs from RGS16-/- mice produce higher amounts of the chemokine KC, the mouse counterpart of CXCL1 which has neutrophil chemoattractant activity. Thus, RGS16 appears as an important factor to control activation of monocytes and their implication in the inflammatory response, as previously demonstrated for lung Th2/Th17 inflammatory responses [30]. Although RGS16 expression is generally not detected in murine macrophages, it could be inducible during activation and then play a role in cytokine

regulation. However, we did not detect any clear differences in the pro-inflammatory cytokine profiles of murine peritoneal and alveolar macrophages from wild type and Rgs16-/- mice, even upon activation by LPS.

It is interesting to note that RGS16 knockdown also enhanced Pam3-induced IL-10 production. As IL-10 is a regulatory cytokine which generally controls inflammation, this result suggests the involvement of RGS16 in a complex feedback loop to attenuate inflammation. Such a model for a role of RGS16 in Pam3 treatment-induced monocyte activation is presented in Figure 4, implicating a regulatory motif called "feed-forward loop" (FFL) [58]. FFLs consist of three signals: a regulator X that regulates Y, and a target Z, which is regulated by both X and Y. In our model, RGS16 (X) inhibits pro-inflammatory cytokines (Z) and also inhibits IL-10 (Y), which itself is an inhibitor of pro-inflammatory cytokines (Z). This motif – “type 2 incoherent FFL” [58] (Figure 4) - would tune the control of inflammation by RGS16 as other FFLs known to regulate cytokine production [53]. The notion that RGS16 can be involved in such regulatory patterns should be considered in the context of different cell types in which RGS16 is induced by activation [9, 12, 16-18, 33, 35, 37, 39-46]. Although such a model is a strong oversimplification of the intricate biological reality, it helps to realize that the regulation mediated by RGS16 is likely complex and regulated by multiple control loops.

Intriguingly, these results from ELISA experiments measuring cytokine concentration in Pam3-activated THP-1 supernatants were generally not reflected by the mRNA level measurements. This observation may be explained by a transient regulation of expression, not reflected by the qPCR measurements. An alternative explanation could involve post-transcriptional processes, e.g. mRNA stabilization. Notably, RGS16 inhibits the p38 MAP kinase [12, 39], which is involved in the stabilization of pro-inflammatory cytokine mRNAs [54, 59, 60]. Post-translational modifications, such as IL-1β maturation by inflammasome activation [61, 62] might also explain discrepancies between mRNA levels and cytokine secreted amounts.

Overall, our results show that RGS16 attenuates the inflammatory response of monocytes, and is likely involved in complex regulatory loops. Further work using primary monocytes

and macrophages will be necessary to get a complete view of the importance of RGS16 on cytokine production by myeloid cells.

ACKNOWLEDGEMENTS

This work was supported by Estonian Science Foundation grant ETF8914, by the Competence Center of Cancer Research, and by Estonian Research Council Targeted Financing Project nr SF0140066s09. The authors would like to express their gratitude to the EGIDE/PARROT exchange program between France and Estonia. The authors are grateful to Dr. Kirk Druey for the generous gift of *Rgs16* knockout mice, and to Dr. Simon Fillatreau for providing critical materials used for the study.

REFERENCES

1 Gonzalez-Maeso J. *Anxious interactions.* Nat Neurosci. 2010 May;13:524-6.

2 Vilardaga JP. *Theme and variations on kinetics of GPCR activation/deactivation.* J Recept Signal Transduct Res. 2010 Oct;30:304-12.

3 Latek D, Modzelewska A, Trzaskowski B, Palczewski K, Filipek S. *G protein-coupled receptors--recent advances.* Acta Biochim Pol. 2012;59:515-29.

4 Dohlman HG, Thorner J. *RGS proteins and signaling by heterotrimeric G proteins.* J Biol Chem. 1997 Feb 14;272:3871-4.

5 Jones MB, Siderovski DP, Hooks SB. *The G betagamma dimer as a novel source of selectivity in G-protein signaling: GGL-ing at convention.* Mol Interv. 2004 Aug;4:200-14.

6 Ross EM, Wilkie TM. *GTPase-activating proteins for heterotrimeric G proteins: regulators of G protein signaling (RGS) and RGS-like proteins.* Annu Rev Biochem. 2000;69:795-827.

7 Fisher RA. *Molecular Biology of RGS Proteins.* Progress in Molecular Biology and Translational Science. 2009;86:1-366.

8 Bansal G, Druey KM, Xie Z. *R4 RGS proteins: regulation of G-protein signaling and beyond.* Pharmacol Ther. 2007 Dec;116:473-95.

9 Suurväli J, Robert J, Boudinot P, Rüütel Boudinot S. *R4 regulators of G protein signaling (RGS) identify an ancient MHC-linked synteny group.* Immunogenetics. 2013 Feb;65:145-56.

10 Chen CK, Wieland T, Simon MI. *RGS-r, a retinal specific RGS protein, binds an intermediate conformation of transducin and enhances recycling.* Proc Natl Acad Sci U S A. 1996 Nov 12;93:12885-9.

11 Snow BE, Antonio L, Suggs S, Siderovski DP. *Cloning of a retinally abundant regulator of G-protein signaling (RGS-r/RGS16): genomic structure and chromosomal localization of the human gene.* Gene. 1998 Jan 12;206:247-53.

12 Buckbinder L, Velasco-Miguel S, Chen Y *et al. The p53 tumor suppressor targets a novel regulator of G protein signaling.* Proc Natl Acad Sci U S A. 1997 Jul 22;94:7868-72.

13 Chen C, Zheng B, Han J, Lin SC. *Characterization of a novel mammalian RGS protein that binds to Galpha proteins and inhibits pheromone signaling in yeast.* J Biol Chem. 1997 Mar 28;272:8679-85.

14 Larminie C, Murdock P, Walhin JP *et al. Selective expression of regulators of G-protein signaling (RGS) in the human central nervous system.* Brain Res Mol Brain Res. 2004 Mar 17;122:24-34.

15 Hu RG, Sheng J, Qi X, Xu Z, Takahashi TT, Varshavsky A. *The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators.* Nature. 2005 Oct 13;437:981-6.

16 Kveberg L, Ryan JC, Rolstad B, Inngjerdingen M. *Expression of regulator of G protein signalling proteins in natural killer cells, and their modulation by Ly49A and Ly49D.* Immunology. 2005 Jul;115:358-65.
17 Timmusk S, Merlot E, Lövgren T, Järvekülg L, Berg M, Fossum C. *Regulator of G protein signalling 16 is a target for a porcine circovirus type 2 protein.* J Gen Virol. 2009 Oct;90:2425-36.
18 Suurväli J, Pahtma M, Saar R *et al.* Functional importance of RGS16 signaling in the activation of human monocytes. In: Schmidt R, ed. *Proceedings of the 2nd European Congress of Immunology (ECI).* Berlin, Germany: Medimond S.r.l., Bologna, Italy, 2009:41-5.
19 Stuebe S, Wieland T, Kraemer E *et al. Sphingosine-1-phosphate and endothelin-1 induce the expression of rgs16 protein in cardiac myocytes by transcriptional activation of the rgs16 gene*. Naunyn Schmiedebergs Arch Pharmacol. 2008 Jan;376:363-73.
20 Teplyuk NM, Galindo M, Teplyuk VI *et al. Runx2 regulates G protein-coupled signaling pathways to control growth of osteoblast progenitors.* J Biol Chem. 2008 Oct 10;283:27585-97.
21 Bourillot PY, Aksoy I, Schreiber V *et al. Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog.* Stem Cells. 2009 Aug;27:1760-71.
22 Zhao C, Putnik M, Gustafsson JA, Dahlman-Wright K. *Microarray analysis of altered gene expression in ERbeta-overexpressing HEK293 cells.* Endocrine. 2009 Oct;36:224-32.
23 Pashkov V, Huang J, Parameswara VK *et al. Regulator of G protein signaling (RGS16) inhibits hepatic fatty acid oxidation in a carbohydrate response element-binding protein (ChREBP)-dependent manner.* J Biol Chem. 2011 Apr 29;286:15116-25.
24 Doi M, Ishida A, Miyake A *et al. Circadian regulation of intracellular G-protein signalling mediates intercellular synchrony and rhythmicity in the suprachiasmatic nucleus.* Nat Commun. 2011;2:327.
25 Derrien A, Zheng B, Osterhout JL *et al. Src-mediated RGS16 tyrosine phosphorylation promotes RGS16 stability.* J Biol Chem. 2003 May 2;278:16107-16.
26 Druey KM, Ugur O, Caron JM, Chen CK, Backlund PS, Jones TL. *Amino-terminal cysteine residues of RGS16 are required for palmitoylation and modulation of Gi- and Gq-mediated signaling.* J Biol Chem. 1999 Jun 25;274:18836-42.
27 Hiol A, Davey PC, Osterhout JL *et al. Palmitoylation regulates regulators of G-protein signaling (RGS) 16 function. I. Mutation of amino-terminal cysteine residues on RGS16 prevents its targeting to lipid rafts and palmitoylation of an internal cysteine residue.* J Biol Chem. 2003 May 23;278:19301-8.
28 Osterhout JL, Waheed AA, Hiol A *et al. Palmitoylation regulates regulator of G-protein signaling (RGS) 16 function. II. Palmitoylation of a cysteine residue in the RGS box is critical for RGS16 GTPase accelerating activity and regulation of Gi-coupled signalling.* J Biol Chem. 2003 May 23;278:19309-16.
29 Hayasaka N, Aoki K, Kinoshita S *et al. Attenuated food anticipatory activity and abnormal circadian locomotor rhythms in Rgs16 knockdown mice.* PLoS ONE. 2011;6:e17655.
30 Shankar SP, Wilson MS, DiVietro JA *et al. RGS16 attenuates pulmonary Th2/Th17 inflammatory responses.* J Immunol. 2012 Jun 15;188:6347-56.
31 Ding Y, Li J, Wu Q *et al. IL-17RA is essential for optimal localization of follicular Th cells in the germinal center light zone to promote autoantibody-producing B cells.* J Immunol. 2013 Aug 15;191:1614-24.

32 Lippert E, Yowe DL, Gonzalo JA *et al. Role of regulator of G protein signaling 16 in inflammation-induced T lymphocyte migration and activation.* J Immunol. 2003 Aug 1;171:1542-55.
33 Agenes F, Bosco N, Mascarell L, Fritah S, Ceredig R. *Differential expression of regulator of G-protein signalling transcripts and in vivo migration of CD4+ naive and regulatory T cells.* Immunology. 2005 Jun;115:179-88.
34 Hsu HC, Yang P, Wang J *et al. Interleukin 17-producing T helper cells and interleukin 17 orchestrate autoreactive germinal center development in autoimmune BXD2 mice.* Nat Immunol. 2008 Feb;9:166-75.
35 Xie S, Li J, Wang JH *et al. IL-17 activates the canonical NF-kappaB signaling pathway in autoimmune B cells of BXD2 mice to upregulate the expression of regulators of G-protein signaling 16.* J Immunol. 2010 Mar 1;184:2289-96.
36 Estes JD, Thacker TC, Hampton DL *et al. Follicular dendritic cell regulation of CXCR4-mediated germinal center CD4 T cell migration.* J Immunol. 2004 Nov 15;173:6169-78.
37 Berthebaud M, Riviere C, Jarrier P *et al. RGS16 is a negative regulator of SDF-1-CXCR4 signaling in megakaryocytes.* Blood. 2005 Nov 1;106:2962-8.
38 Li X, Li J, Yang Y *et al. Differential gene expression in peripheral blood T cells from patients with psoriasis, lichen planus, and atopic dermatitis.* J Am Acad Dermatol. 2013 Nov;69:e235-43.
39 Zhang Y, Neo SY, Han J, Yaw LP, Lin SC. *RGS16 attenuates galphaq-dependent p38 mitogen-activated protein kinase activation by platelet-activating factor.* J Biol Chem. 1999 Jan 29;274:2851-7.
40 Giorelli M, Livrea P, Defazio G *et al. Interferon beta-1a counteracts effects of activation on the expression of G-protein-coupled receptor kinases 2 and 3, beta-arrestin-1, and regulators of G-protein signalling 2 and 16 in human mononuclear leukocytes.* Cell Signal. 2002 Aug;14:673-8.
41 Patten M, Stube S, Thoma B, Wieland T. *Interleukin-1beta mediates endotoxin- and tumor necrosis factor alpha-induced RGS16 protein expression in cultured cardiac myocytes.* Naunyn Schmiedebergs Arch Pharmacol. 2003 Nov;368:360-5.
42 Shi GX, Harrison K, Han SB, Moratz C, Kehrl JH. *Toll-like receptor signaling alters the expression of regulator of G protein signaling proteins in dendritic cells: implications for G protein-coupled receptor signaling.* J Immunol. 2004 May 1;172:5175-84.
43 Barker KS, Liu T, Rogers PD. *Coculture of THP-1 human mononuclear cells with Candida albicans results in pronounced changes in host gene expression.* J Infect Dis. 2005 Sep 1;192:901-12.
44 Orth JH, Lang S, Taniguchi M, Aktories K. *Pasteurella multocida toxin-induced activation of RhoA is mediated via two families of G{alpha} proteins, G{alpha}q and G{alpha}12/13.* J Biol Chem. 2005 Nov 4;280:36701-7.
45 Gat-Viks I, Chevrier N, Wilentzik R *et al. Deciphering molecular circuits from genetic variation underlying transcriptional responsiveness to stimuli.* Nat Biotechnol. 2013 Apr;31:342-9.
46 Vasilatos SN, Katz TA, Oesterreich S, Wan Y, Davidson NE, Huang Y. *Crosstalk between lysine-specific demethylase 1 (LSD1) and histone deacetylases mediates antineoplastic efficacy of HDAC inhibitors in human breast cancer cells.* Carcinogenesis. 2013 Jun;34:1196-207.
47 Wiechec E, Overgaard J, Hansen LL. *A fragile site within the HPC1 region at 1q25.3 affecting RGS16, RGSL1, and RGSL2 in human breast carcinomas.* Genes Chromosomes Cancer. 2008 Sep;47:766-80.

48 Liang G, Bansal G, Xie Z, Druey KM. *RGS16 inhibits breast cancer cell growth by mitigating phosphatidylinositol 3-kinase signaling*. J Biol Chem. 2009 Aug 7;284:21719-27.
49 Wiechec E, Wiuf C, Overgaard J, Hansen LL. *High-resolution melting analysis for mutation screening of RGSL1, RGS16, and RGS8 in breast cancer*. Cancer Epidemiol Biomarkers Prev. 2011 Feb;20:397-407.
50 Liu T, Bohlken A, Kuljaca S *et al. The retinoid anticancer signal: mechanisms of target gene regulation*. Br J Cancer. 2005 Aug 8;93:310-8.
51 Kumar H, Kawai T, Akira S. *Toll-like receptors and innate immunity*. Biochem Biophys Res Commun. 2009 Oct 30;388:621-5.
52 Funderburg NT, Jadlowsky JK, Lederman MM, Feng Z, Weinberg A, Sieg SF. *The Toll-like receptor 1/2 agonists Pam(3) CSK(4) and human beta-defensin-3 differentially induce interleukin-10 and nuclear factor-kappaB signalling patterns in human monocytes*. Immunology. 2011 Oct;134:151-60.
53 Lampropoulou V, Calderon-Gomez E, Roch T *et al. Suppressive functions of activated B cells in autoimmune diseases reveal the dual roles of Toll-like receptors in immunity*. Immunol Rev. 2010 Jan;233:146-61.
54 Frevel MA, Bakheet T, Silva AM, Hissong JG, Khabar KS, Williams BR. *p38 Mitogen-activated protein kinase-dependent and -independent signaling of mRNA stability of AU-rich element-containing transcripts*. Mol Cell Biol. 2003 Jan;23:425-36.
55 Wong HR, Odoms K, Sakthivel B. *Divergence of canonical danger signals: the genome-level expression patterns of human mononuclear cells subjected to heat shock or lipopolysaccharide*. BMC Immunol. 2008;9:24.
56 Carta S, Tassi S, Pettinati I, Delfino L, Dinarello CA, Rubartelli A. *The rate of interleukin-1beta secretion in different myeloid cells varies with the extent of redox response to Toll-like receptor triggering*. J Biol Chem. 2011 Aug 5;286:27069-80.
57 Vasselon T, Hanlon WA, Wright SD, Detmers PA. *Toll-like receptor 2 (TLR2) mediates activation of stress-activated MAP kinase p38*. J Leukoc Biol. 2002 Mar;71:503-10.
58 Alon U. *Network motifs: theory and experimental approaches*. Nat Rev Genet. 2007 Jun;8:450-61.
59 Campbell J, Ciesielski CJ, Hunt AE *et al. A novel mechanism for TNF-alpha regulation by p38 MAPK: involvement of NF-kappa B with implications for therapy in rheumatoid arthritis*. J Immunol. 2004 Dec 1;173:6928-37.
60 Langer M, Malykhin A, Maeda K *et al. Bacillus anthracis peptidoglycan stimulates an inflammatory response in monocytes through the p38 mitogen-activated protein kinase pathway*. PLoS ONE. 2008;3:e3706.
61 Gross O, Thomas CJ, Guarda G, Tschopp J. *The inflammasome: an integrated view*. Immunol Rev. 2011 Sep;243:136-51.
62 Ghiringhelli F, Apetoh L, Tesniere A *et al. Activation of the NLRP3 inflammasome in dendritic cells induces IL-1beta-dependent adaptive immunity against tumors*. Nat Med. 2009 Oct;15:1170-8.

TABLES

Table I. Human primers used for qPCR.

	Forward primer	Reverse primer
GAPDH	GAGTCAACGGATTTGGTCGT	TTGATTTTGGAGGGATCTCG
IL-1β	TCCTGCGTGTTGAAAGATGATAA	CAAATCGCTTTTCCATCTTCTTC
IL-6	TGAAAGCAGCAAAGAGGCACTGGCAGA	GACCAGTGATGATTTTCACCAGGCAAGTCTCC
IL-8	TCTTGGCAGCCTTCCTGATT	TTTCTGTGTTGGCGCAGTGT
IL-10	CCAAGACCCAGACATCAAGGCGCATGT	AGGCATTCTTCACCTGCTCCACGGC
RGS16	CCTGGAGAGAGCCAAAGAGTT	GCAGGTCGAACGACTCTCTC
TNFa	CACGCTCTTCTGCCTGCTG	GATGATCTGACTGCCTGGGC

Table II. Pam3-induced cytokine production by RGS16-overexpressing and control vector-transfected THP-1 cells. Timepoints indicate hours post activation.

	IL-1β, ng / ml		IL-6, ng / ml		IL-8, ng / ml		TNFα, ng / ml	
sample	mean	SD	mean	SD	mean	SD	mean	SD
2h pcDNA3	0.7	0.3	41.3	3.2	980.7	47.9	488.8	32.4
2h RGS16	1.06	0.7	23.1	2.4	642.6	60.8	243.3	18.9
24h pcDNA3	314.4	70.7	2890.9	209.6	55116.1	1 370.9	1300.7	86.6
24h RGS16	314.6	23.9	2068.2	105.2	52206.9	1 344.8	1223.8	99.4
48h pcDNA3	631.8	50.2	7960.3	523.6	67636.0	2 030.4	2011.9	63.7
48h RGS16	469.1	27.7	4762.5	349.8	61560.2	1 890.3	1379.4	36.6

FIGURE LEGENDS

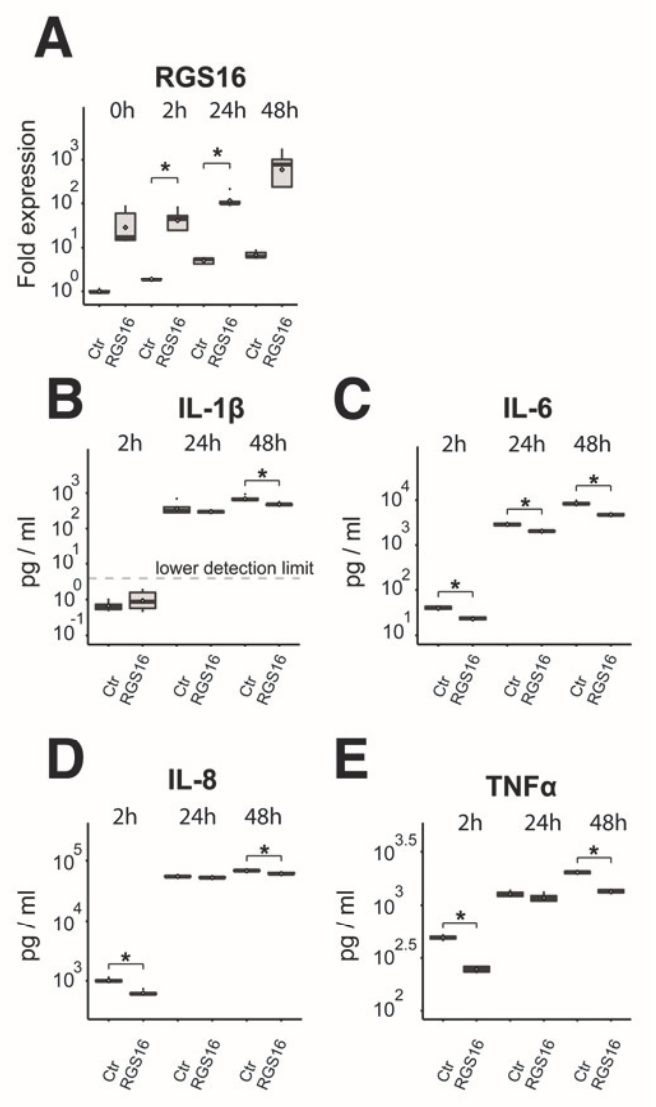
Figure 1. Effects of RGS16 overexpression on monocyte activation. THP-1 cells transfected with either pcDNA3-RGS16 („RGS16" on the figure) or with an empty pcDNA3 control vector („Ctr" on the figure) were activated using Pam3 in the final concentration of 100 ng/ml 4 hours post transfection. Hours on the figure indicate activation timepoints. **(A)** RGS16 mRNA expression relative to the expression in cells transfected with control vector and harvested at

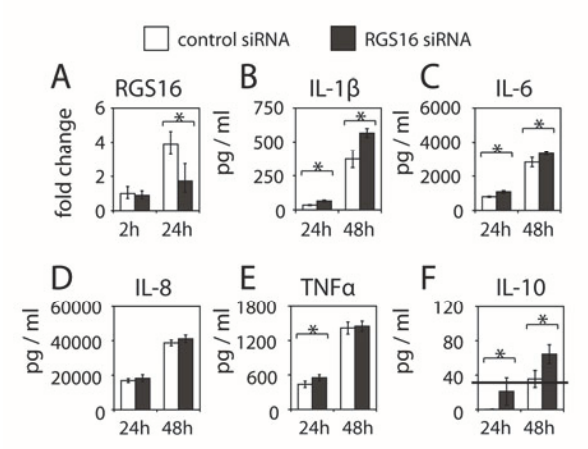
the timepoint 0 h (just before adding Pam3). (**B – E)** Cytokine production quantified by ELISA. Outliers are presented using empty circles and mean of the group is shown by empty diamond. Asterisks mark statistically significant results ($p < 0.05$) evaluated using Welch two sample t-test for data following normal distribution (A) and Mann-Whitney U test for other data (B-E); $n \geq 5$.

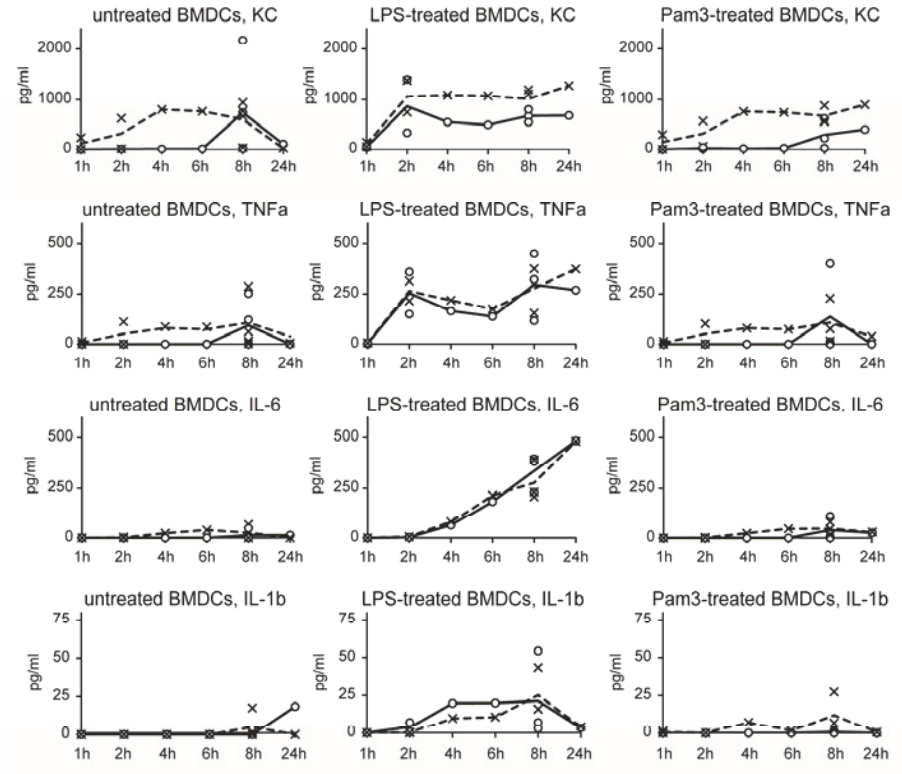
Figure 2. Effects of RGS16 silencing on monocyte activation. Cells pretreated for 17 h with 5 nM of either RGS16-specific siRNA or Silencer® Select Negative Control siRNA were activated using Pam3 in the final concentration of 100 ng/ml. Hours on the figure indicate activation timepoints. (**A)** RGS16 mRNA expression in activated cells relative to cells treated with control siRNA at the same timepoint, normalized to the housekeeping gene GAPDH. (**B – F)** Cytokine expression measured either 24 h or 48 h after cell activation using Pam3 in the final concentration of 100 ng/ml. (**F)** Horizontal line marks the lower detection limit of the IL-10 DuoSet ELISA kit (31.2 pg/ml). Asterisks mark statistically significant results ($p < 0.05$) evaluated using Welch two sample t-test for data following normal distribution (A) and Mann-Whitney U test for other data (B - F); $n \geq 4$.

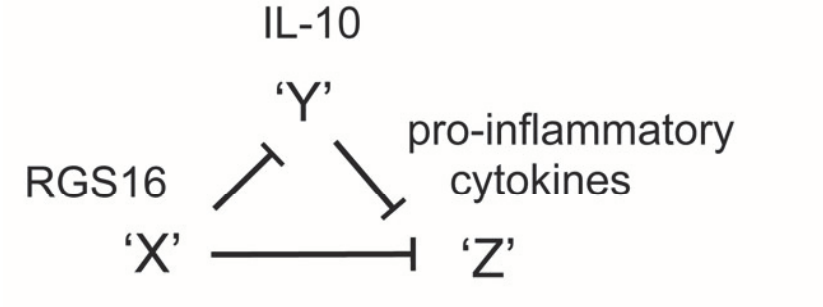
Figure 3. **Effect of Rgs16 gene disruption on cytokine production by LPS or Pam3-treated bone marrow-derived cells (BMDCs), quantified by ELISA.** BMDCs (500 000 cells/ml), derived from wild-type or Rgs16 knock-out mice were activated using LPS (1 μg/ml) or Pam3 (10 ng/ml). Open circles represent the kinetics of cytokine production by wild-type BMDC-s and crosses represent the kinetics of cytokine production by BMDC-s derived from Rgs16 knock-out mice. Solid and dashed lines connect mean values for each time-point for wild-type and Rgs16 knock-out BMDC-s, respectively.

Figure 4. Model for RGS16 regulation of cytokine production during monocyte activation. The functional relationships between RGS16, IL-10 and pro-inflammatory cytokines are modeled as a “type 2 incoherent feed-forward loop” [58]: the effector X (RGS16) inhibits both the target Z (pro-inflammatory cytokines) and its inhibitor Y (IL-10).









Incoherent Feed Forward Regulatory Loop

PUBLICATION II

R4 regulators of G protein signaling (RGS) identify an ancient MHC-linked synteny group

Suurväli, J., Robert, J., Boudinot, P., Rüütel Boudinot, S.

Immunogenetics (2013), 65(2):145-56.
DOI: 10.1007/s00251-012-0661-x

Immunogenetics (2013) 65:145–156
DOI 10.1007/s00251-012-0661-x

ORIGINAL PAPER

R4 regulators of G protein signaling (RGS) identify an ancient MHC-linked synteny group

Jaanus Suurväli · Jacques Robert · Pierre Boudinot · Sirje Rüütel Boudinot

Received: 21 June 2012 / Accepted: 22 October 2012 / Published online: 6 November 2012

Abstract Regulators of G protein signaling (RGS) are key regulators of G protein signaling. RGS proteins of the R4 RGS group are composed of a mere RGS domain and are mainly involved in immune response modulation. In both human and mouse, most genes encoding the R4 RGS proteins are located in the same region of chromosome 1. We show here that the *RGS1*/*RGS16* neighborhood constitutes a synteny group well conserved across tetrapods and closely linked to the MHC paralogon of chromosome 1. Genes located in the *RGS1*/*RGS16* region have paralogs close to the MHC on chromosome 6 or close to the other MHC paralogons. In amphioxus, a cephalochordate, these genes possess orthologs that are located in the same scaffolds as a number of markers defining the proto-MHC in this species (Abi-Rached et al., Nat Genet 31:100–115, 2002). We therefore propose that the *RGS1*/*RGS16* region provides useful markers to investigate the origins and the evolution of the MHC. In addition, we show that some genes of the region appear to have immune functions not only in human, but also in *Xenopus*.

Keyword Regulators of G protein signaling · Tetrapod evolution · *Branchiostoma floridae* · Proto-MHC · *Xenopus tropicalis*

Electronic supplementary material The online version of this article (doi:10.1007/s00251-012-0661-x) contains supplementary material, which is available to authorized users.

J. Suurväli (✉) · S. Rüütel Boudinot
Department of Gene Technology,
Tallinn University of Technology,
Akadeemia tee 15,
Tallinn 12618, Estonia
e-mail: jaanus.suurvali@ttu.ee

J. Robert
Department of Microbiology and Immunology,
University of Rochester Medical Center,
Rochester, NY 14642, USA

P. Boudinot
INRA, Molecular Virology and Immunology,
Domaine de Vilvert,
78350 Jouy-en-Josas, France

Introduction

Regulators of G protein signaling (RGS) are key factors in the regulation of a multitude of processes that are initiated by ligand binding to a G protein-coupled receptor (GPCR). They regulate the GDP/GTP exchange on the Gα subunit of the G proteins bound to GPCRs (reviewed by Bansal et al. 2007), thus controlling a vast number of processes requiring cell-to-cell communication, including development, neurotransmission, chemotaxis, endocrine regulation, visual and olfactory sensing, and, in fungi, even mating-type exclusion.

It has been suggested that the components of GPCR signaling pathways have a prokaryotic origin, the signaling systems of unicellular eukaryotes representing a transitional stage between those found in prokaryotes and higher eukaryotes (Pertseva and Shpakov 2009). Moreover, the presence of the RGS domain has been reported in most, if not all, major eukaryotic groups (Anantharaman et al. 2011), and RGS domains therefore constitute very old structural units. The evolution of both RGS and Gα proteins has been dominated by multiple independent lineage-specific expansions (Anantharaman et al. 2011), leading to an extensive diversification of their domain architectures. It is generally accepted that an expansion of RGS domains occurred in metazoans with a core list of 12 proteins present in the metazoan common ancestor (Sierra et al. 2002). The RGS protein family, while defined by the presence of a functional

RGS domain, is divided into several subfamilies based on sequence homology and the presence or absence of additional protein domains (Bansal et al. 2007; Siderovski and Willard 2005; Sierra et al. 2002). The smallest RGS proteins do not contain any other domains besides the core RGS domain itself and form the R4 RGS subfamily (Bansal et al. 2007).

Many members of the R4 *RGS* subfamily are involved in chemotaxis and immunity to infections (reviewed by Bansal et al. 2007). Bacterial endotoxin has been shown to induce the expression of *RGS1*, *RGS4*, and *RGS16* and to downregulate *RGS2* and *RGS18* (Lee et al. 2010; Patten et al. 2002; Riekenberg et al. 2009; Shi et al. 2004; Timmusk et al. 2009). *RGS2* is upregulated during the intracellular infection of phagocytes with the gram-negative bacteria *Brucella abortus* and is in fact required for the efficient replication and survival of the pathogen (Kim et al. 2012). Several R4 *RGS* can be either upregulated or downregulated by type I interferons (Giorelli et al. 2002; Tran et al. 2010) and by viral infection (Cahir-McFarland et al. 2004). For example, we have previously shown that *RGS16* is upregulated by the interferon inducers poly I:C and CpG as well as by LPS and PHA in porcine PBMCs (Timmusk et al. 2009). Additionally, we demonstrated a direct interaction between RGS16 and a viral protein from the porcine circovirus type 2, which causes immunodeficiency in pigs (Timmusk et al. 2009).

In human, genes encoding the ten R4 RGS proteins are located in groups of two or more genes on chromosome 1, except for *RGS3* which is located on chromosome 9: *RGS4* and *RGS5* are located on 1q23.3, *RGS8* and *RGS16* on 1q25.3, and *RGS1*, *RGS2*, *RGS13*, *RGS18*, and *RGS21* on 1q31.2. This configuration likely results from multiple tandem duplications followed by fragmentation of the region (Sierra et al. 2002). The location of most *RGS* genes is close to an MHC paralogon on human chromosome 1, which suggests that they may be located in a paralogous region of the extended MHC. The human genome contains three paralogous regions (or "paralogons") of the MHC region; while the MHC itself is located on chromosome 6, its paralogons are for the most part found on chromosomes 1, 9, and 19 (Flajnik and Kasahara 2001; Kasahara et al. 1996; Kasahara 1999; Katsanis et al. 1996). It is generally accepted that these four regions were produced by the two rounds of whole genome duplication that occurred between the emergence of urochordates and the rise of jawed vertebrates. The MHC is a very large genomic region where many genes are involved in immunity or inflammation (Beck et al. 1999; Trowsdale 2001). It is remarkable that this region and its three paralogons were kept during the evolution of vertebrates as synteny blocks, suggesting that a conserved linkage of genes involved in immunity could be favorable. Alternatively, this configuration could be a mere historical legacy of the vertebrate whole genome duplications. Tracking the history and the limits of these regions is of pivotal importance to understand the origins of the immune system through the evolution of vertebrate genomes.

In this study, we showed that R4 RGS genes identify a region which is conserved across vertebrate genomes and represents an ancient synteny group. This region is linked to one of the paralogons of the MHC and might, therefore, provide useful markers to trace the evolution of this genetic complex. In the cephalochordate amphioxus (*Branchiostoma floridae*), the counterpart of the R4 RGS region was indeed found closely linked to the proto-MHC.

Materials and methods

Phylogenetic analysis for the identification of paralogs/ohnologs

Protein sequences for the 76 genes of the RGS1/RGS16 region were obtained from the human assembly GRCh37.9 in the GenBank RefSeq database. If more than one protein sequence was available for a given gene, the longest isoform was used. To find paralogs, sequences were then blasted using the first iteration of Domain Enhanced Lookup Time Accelerated Basic Local Alignment Search Tool (DELTA-BLAST; Boratyn et al. 2012) against the reference sequences of the human genome to build a list of unique sequences that would display both sequence similarity and similar domain organization to each sequence of the initial set. In most cases, sequences were selected from DELTA-BLAST hits using a cutoff determined by a sharp drop in either the *E* values or protein identity. When such a cutoff did not appear, 20 hits with the lowest *E* values were chosen for phylogenetic analysis. In rare cases like the one of zinc finger proteins (*ZNF648* in the human *RGS1*/*RGS16* region), there were too many results all very similar to each other and a relevant phylogenetic analysis could not be performed.

The MEGA5 software package was used to compute phylogenetic trees using the retrieved sequences (Tamura et al. 2011; Online Resource 2). The evolutionary history was inferred using the maximum likelihood method based on the JTT matrix-based model (Jones et al. 1992). The bootstrap consensus tree was inferred from 1,000 replicates (Felsenstein 1985). The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with <95 % site coverage were eliminated. That is, fewer than 5 % alignment gaps, missing data, and ambiguous bases were allowed at any position.

Ensembl BioMart was used to determine the genomic location of each gene used to build the phylogenies. Ohnologs were determined by visual analysis of the trees

and confrontation of the trees and genomic locations (Online Resource 2). Genes (1) that appeared as close paralogs of the *RGS1*/*RGS16* genes in the phylogeny and (2) that were also located in MHC paralogons on chromosomes 1, 6, 9, or 19 as well as 5 and 15 (reviewed by Flajnik and Kasahara 2010) were considered to be potential ohnologs.

Determining the synteny groups in vertebrates

In order to check the conservation of the region in vertebrates other than human, the synteny groups between human (*Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), anole lizard (*Anolis carolinensis*), Western clawed frog (*Xenopus tropicalis*), stickleback (*Gasterosteus aculeatu*), and Japanese medaka (*Oryzias latipes*) were determined from the Ensembl genome assemblies version 68.01 of the Genomicus browser described in detail in Muffato et al. (2010) and updated with some additional genes present in GenBank but not Ensembl itself. To confirm the identity of *RGS1*/*RGS16* region genes shown by the Genomicus browser, we referred to the online database PhylomeDB (Huerta-Cepas et al. 2011; Online Material 1). To confirm the identity of NOTCH2 genes and RGS genes themselves, the maximum likelihood phylogenies were constructed with a bootstrap value of 1,000 using MEGA5 (Online Material 1).

Determining the synteny groups in amphioxus

Cosmids were previously used to identify the amphioxus proto-MHC (Abi-Rached et al. 2002). To determine the equivalent regions in the current genome assembly, these cosmid sequences were blasted against the genomic RefSeq sequences at National Center for Biotechnology Information (NCBI). Amphioxus scaffolds with the best coverage and lowest *E* values were considered. The most similar sequences to the MHC marker genes from these scaffolds were retrieved using protein BLAST as described above.

Protein sequences corresponding to the 76 genes of the human (GRCh37.9) RGS1/RGS16 region were retrieved from GenBank as described above and blasted against *B. floridae* assembly version 2 using protein BLAST (blastp) with the default settings. The best hits were then blasted back against the human assembly with the same settings in order to confirm whether the result had been true or false positive.

Sequences for the previously cloned amphioxus *NOTCH*-like protein (Holland et al. 2001) and other markers of the proto-MHC (Abi-Rached et al. 2002) located on scaffolds with sequences homologous to *RGS1*/*RGS16* genes were treated similarly. These scaffolds were identified by blasting the mRNA and cosmid sequences of the proto-MHC markers from previous works (Abi-Rached et al. 2002; Holland et al. 2001) against the amphioxus reference genomic sequences using NCBI megaBLAST.

We refer to the online database PhylomeDB (Huerta-Cepas et al. 2011, Supplementary Material 3) for many of our genes of interest. Additional phylogenies were initially created using the online tool FIGENIX (Gouret et al. 2005; Paganini and Gouret 2012) to get a first idea of the homologs in other organisms. To build the final version of these trees, sequences for relevant genes were retrieved from GenBank and Ensembl and a maximum likelihood tree was built using the MEGA5 software package as described above for ohnolog identification (Online Resource 3). Amphioxus orthologs of the human genes were individually confirmed by visual analysis of phylogenies and examination of the domain structure of encoded proteins using NCBI or by EMBL/EBI InterProScan (Quevillon et al. 2005).

Animal handling

Frog virus 3 (FV3, *Iridoviridae*) was grown in and purified from baby hamster kidney (BHK-21) cell lines incubated at 30 °C during the infection, and virus titer was determined on fathead minnow cells as previously described (Chen et al. 2011). Three-year-old *X. tropicalis* outbred animals were obtained from the *Xenopus* Research Resource for Immunology at the University of Rochester (http://www.urmc.rochester.edu/smd/mbi/xenopus/index.htm) and injected intraperitoneally with one million plaque-forming units of FV3 in a volume of 0.1 ml of amphibian PBS. Tissues were collected 6 and 9 days postinfection. All animals were handled under strict laboratory and UCAR regulations (approval number 100577/2003-151), minimizing discomfort at all times.

Sample preparation for qPCR

RNA was extracted from the spleens and kidneys collected from FV3-infected and control frogs using Trizol reagent following the manufacturer's protocol (Invitrogen). Total RNA (0.5 to 1.0 μg) in 20 μl was used to synthesize cDNA with the iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA).

Quantitative PCR

Quantitative polymerase chain reactions (qPCR) were carried out in a total volume of 10 μl containing 5 μl Maxima™ SYBR Green/ROX qPCR Master Mix (2×) (Fermentas), 0.5 μM (0.25 μl) forward primer, 0.5 μM (0.25 μl) reverse primer (Table 1), 1 μl 10× diluted cDNA, and 3.5 μl nuclease-free water. Thermocycler conditions were as follows: incubation for 2 min at 50 °C, incubation for 10 min at 95 °C, 40 cycles of denaturation

Table 1 *X. tropicalis* primers used for qPCR

Gene	Forward primer	Reverse primer
gapdh	xtGAPDH_ex6F: TGTTGGGGTGAACCATGACAA	xtGAPDH_ex7-ex8R: AAGGCATGGACGGTAGTCATCA
dhx9	xtDHX9_ex20F: GACCTGGTTTCTGCTTCCATTTATGT	xtDHX9_ex21R: TCGTGCAGTGGAGTACGGAATATC
ier5	xtIER5_F: CTCTTGCAGCCCAGACGACA	xtIER5_R: TTTATGCTGGGACGCGGAGT
ncf2	xtNCF2_ex10F: CCGGATGCTACTCCTCCTCAACTA	xtNCF2_ex11-ex12R: CAGCCACTGCTTCTGTTTCCTGTA
ptgs2	xtPTGS2_ex1F: TGATCGTACTACCCGCCGCT	xtPTGS2_ex2R: TGGCAGGGATTTGAACAGCA
rgs16	xtRGS16_ex4: GATGGGTAAAAAGGCTCCCCA	xtRGS16_ex5: TCTCTGGCCCGATGGTCAATA
stx6	xtSTX6_ex3F: GCTGAACTGAGACAAAGGAAAGCC	xtSTX6_ex4R: TTGAACAGAAGGGCTTGTCATCC

for 15 s at 95 °C, and annealing/extension for 1 min at 60 °C, followed by dissociation curve analysis. *gapdh* was used as the housekeeping gene for determining relative gene expression with the $2^{-\Delta\Delta Ct}$ method.

Statistical analysis

All experiments were repeated three times. qPCR was run in triplicates, results from which were averaged (after outlier removal) to obtain the Ct value for each sample. After correcting these values in reference to the housekeeping gene *gapdh*, the resulting ΔCt values were used for *p* value calculation using the Student's unpaired two-tailed *t* test.

Results and discussion

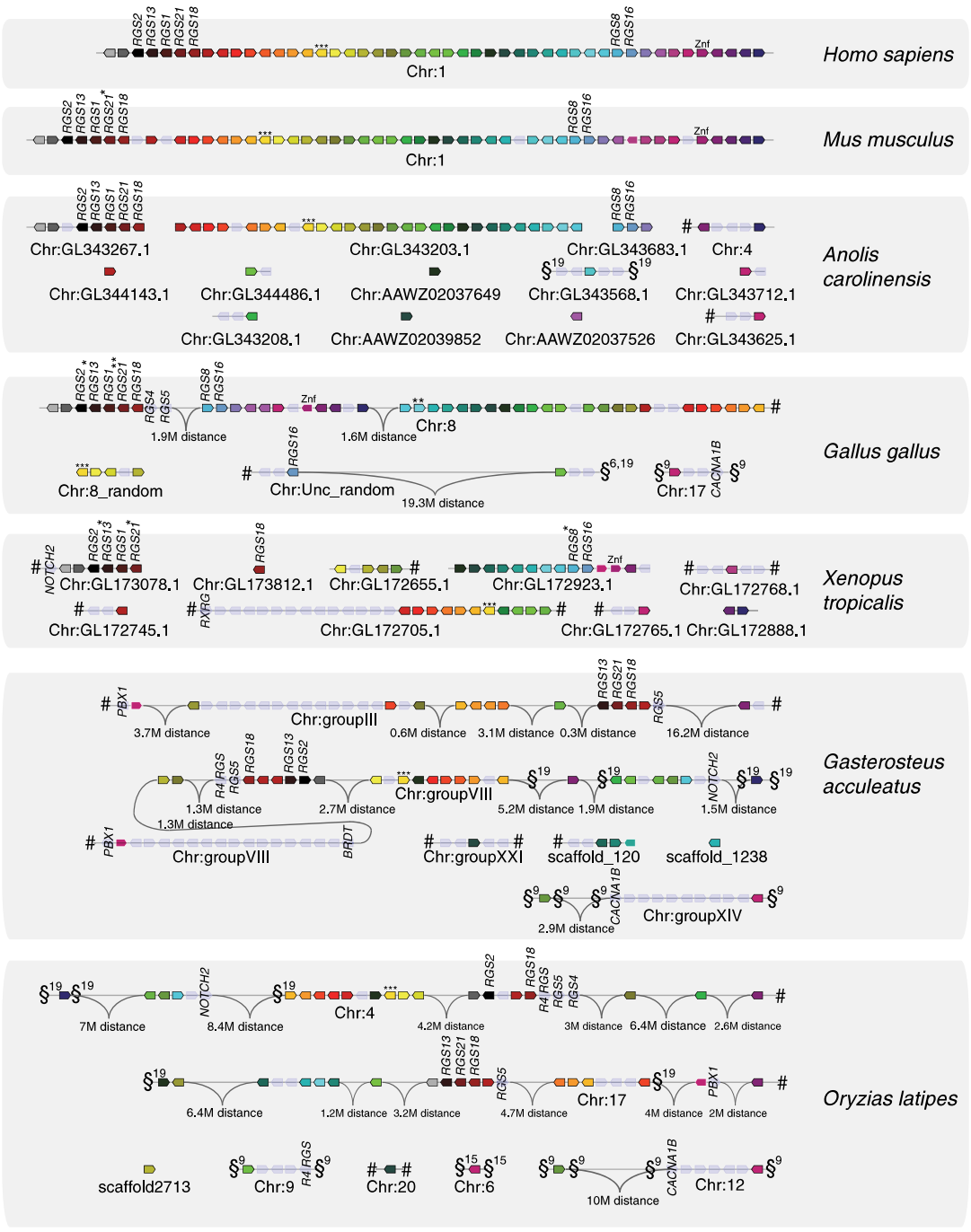
The *RGS1*/*RGS16* region is conserved across vertebrate species

Most of the human R4 *RGS* genes (*RGS1*, *RGS2*, *RGS8*, *RGS13*, *RGS16*, *RGS18*, and *RGS21*) can be found in the proximity of type I interferon-inducible genes *RGS1* and *RGS16*. For the purpose of the current study, we defined the *RGS1*/*RGS16* region in human as the region located between the genes *FAM163A* and *NEK7*, i.e., at 1q25.2–q31.3. This region comprises all genes located between *RGS16* and *RGS2*—including other R4 *RGS* genes—and 20 additional genes on both sides (Online Resource 1). This region belongs to a single ancestral genomic block (Nakatani et al. 2007; Putnam et al. 2008). Following Putnam et al. (2008), this block corresponds to the human genome segment 1.14, which covers the region 1q23.3–32.1.

To determine whether this region was conserved in other vertebrates, the orthologs of all its genes were searched in mouse, chicken, the lizard *A. carolinensis*, and *X. tropicalis*, as well as in two fish species, the stickleback and the medaka. Gene locations were determined in the respective genomes using the Genomicus browser, which supports viewing of the synteny groups and ancestral blocks in different orthologous and paralogous regions (Muffato et al. 2010). We found that the gene set encoded in the human *RGS1*/*RGS16* region constitutes a synteny group remarkably well conserved across tetrapod species (Fig. 1; Online Resource 1). In the chicken, this region has been broken into three subregions later rearranged through intrachromosomal recombinations and inversions. However, most R4 *RGS* genes are grouped in a unique cluster on chromosome 8, including *RGS4* and *RGS5* that are not located in the region in human and mice. Regarding *A. carolinensis* and *X. tropicalis*, although their chromosomes have not been fully assembled, many of the genes of the *RGS1*/*RGS16* region are located on three to four large scaffolds, suggesting that the synteny group is at least partly conserved in these species (Fig. 1).

We also retrieved the linkage group in two different teleost fish species. Duplicated copies of the *RGS1*/*RGS16* region could be retrieved in several linkage groups (Fig. 1). For example, R4 *RGS* were found in paralogous regions containing MHC markers and located in medaka chromosomes 4 and 17, which originated from the same ancestral chromosome through the fish-specific whole genome duplication (Kasahara et al. 2007). Importantly, these observations

Fig. 1 The *RGS1*/*RGS16* region defined on human chromosome 1 is conserved as a synteny group in vertebrate genomes. Synteny groups were determined from Ensembl genome assemblies using the Genomicus database and browser version 68.01 (Muffato et al. 2010). Orthologous genes are indicated by the *same color* in different species. Empty nodes on the figure represent genes not present in the human version of the *RGS1*/*RGS16* region. Gene names on the figure refer to the closest human ortholog. *Broken lines* for a given species indicate different scaffolds. *Number sign (#)* indicates that the next genes on the scaffold are not located in the neighborhood in human. *§9*/*§19* genomic region containing orthologs of genes on the human MHC paralogons on chromosomes 9 or19, respectively. *Single asterisk (*)* gene not present in Genomicus/Ensembl, but present in the NCBI assembly, *double asterisks (**)* gene present in NCBI but not in Genomicus/Ensembl, *triple asterisks (***)* in the current version of Ensembl and Genomicus, human *HMCN1* is not recognized as an ortholog of the gene in other species. Due to high similarity between these proteins, it is shown as such on the figure. Zinc finger protein (*Znf*) representation is based on a Znf gene being located at the correct genetic context, not on phylogenetic analysis. MR1 orthologs in chicken, lizard, and *Xenopus* are shown in an additional figure (Online Resource 1). The orthology of the *NOTCH* and *RGS* genes across different species was confirmed by phylogenetic analysis (Online Resource 1)



indicate that the *RGS1*/*RGS16* region was present in the common ancestor of fish and tetrapods, which fits with the evolution of genome blocks conserved through vertebrate evolution described in Nakatani et al. (2007).

The *RGS1*/*RGS16* region belongs to the extended MHC paralogon on human chromosome 1 and has paralogs on chromosomes 6, 9, and 19

MHC had been first defined functionally as "a group of genes coding for molecules that provide the context of recognition of foreign antigen by T lymphocytes" (Klein 1986), but this definition did not take the genomic dimension into account. The MHC may be also described as a large genetic region of more than 100 genes with a large proportion involved in immunity (Danchin et al. 2004), integrating the presence of many genes in addition to the MHC class I and class II genes. Such a genomic definition leads to the notion of "MHC paralogons" describing the four regions on human chromosomes 1, 6, 9, and 19, which were produced from the original proto-MHC through two rounds of global genome duplication that occurred in early vertebrate evolution (Ohno 1970; Kasahara et al. 1996). The main paralogons were initially found on four chromosomes, but additional smaller fragments were later identified. In fact, the MHC paralogons in human are, therefore, located in a total of at least eight genomic locations on six different chromosomes (1p11–p32 and 1q21–25; 6p21–6p22 and 15q13–15q26; 9p13–p24, 9q32–q34, and 5q11–q23; and 19p13.1–13.3) (reviewed in Flajnik and Kasahara 2010; Flajnik et al. 2012). The main MHC paralogons are commonly tracked by the presence of sets of markers that have paralog copies on some or all MHC paralogons. These markers include genes such as *brd*, *cacna1*, *notch*, *pbx*, and *rxr*.

In the human genome, the *RGS1*/*RGS16* region is located on the edge of an MHC paralogon at 1q21–25 and includes *CACNA1E*, a well-known marker gene that has paralogs on two of the other main MHC paralogons. In addition, the *RGS1*/*RGS16* region, which contains multiple *RGS* genes encoding proteins expressed in the central nervous system, is flanked on the other side by one of the main neurotrophin paralogons at 1q32–q44 (Hallböök et al. 2006). In addition to *CACNA1E*, other markers for the chromosome 1 MHC paralogon are located in the *RGS1*/*RGS16* neighborhood (Fig. 2; Online Resource 2) including (1) the gene *MR1* that is a paralog of the *HLA* genes found in the MHC on chromosome 6 and (2) the gene *IER5* which has paralogs on chromosomes 9 and 19. Our phylogenetic analyses (Supplementary Fig. S1) show that many genes in the *RGS1*/*RGS16* region have in fact paralogs on chromosomes 5, 6, 9, 15, and/or 19 (Fig. 2). When these paralogs were located in the extended MHC area or its counterparts, we considered them as ohnologs. Figure 2 shows that ohnologs of the *RGS1*/*RGS16* genes are indeed distributed within the MHC paralogons. Notably, a gene encoding an R4 RGS protein (*RGS3*) is found in the MHC paralogon on human chromosome 9, which is also the chromosome that was found to contain the highest amount of ohnologs for the *RGS1*/*RGS16* region genes (Fig. 2; Online Resource 2). Interestingly, most of these paralogs mapped to chromosome 9, which is consistent with the idea that, among MHC paralogons, 9q34 retained much more markers than the other duplicated regions (Vienne et al. 2003).

The relationship between the *RGS1*/*RGS16* region and MHC paralogons is further supported by observations from other vertebrate genomes: sequences similar to genes from the *RGS1*/*RGS16* region are indeed often found at close proximity of classical MHC marker homologs (Figs. 1 and 3).

In chicken, a copy of glutamine synthase *GLUL*, a gene of the *RGS1*/*RGS16* region, is located near the MHC marker *CACNA1B* on the small 11 Mb chicken chromosome 17, supporting the notion that avian microchromosomes provide a good representation of the ancestral configuration (Burt 2002; Nakatani et al. 2007). Additionally, a copy of *GLUL* and *CACNA1B* are also found close to each other in both stickleback and medaka, and an additional MHC marker, a *NOTCH2*-like sequence, maps together with the *RGS1*/*RGS16* region genes in both *Xenopus* and fish. The MHC paralogon found on human chromosome 1 appears to have been split into two parts, resulting in several of the genes including the human *NOTCH2* located in the region 1p11–1p32, while other markers are located closer to the *RGS1*/*RGS16* region at the location 1q21–25. However, this is likely a result of recent intrachromosomal rearrangements (Kasahara 1999) and thus does not represent the ancestral configuration. Notably, a *NOTCH2*-like sequence is not only located next to the R4 *RGS* genes in *X. tropicalis* but is also located next to some other genes of the *RGS1*/*RGS16* region in both medaka and stickleback (Fig. 1). Phylogenetic analysis indicates that these *NOTCH2*-like sequences are indeed orthologs of the human *NOTCH2* (Online Resource 1).

A copy of the MHC marker gene *PBX* is next to a *GLUL* homolog in both medaka and stickleback, and in stickleback, a homolog of the marker *BRDT* is located at close proximity to a second pair of *GLUL* and *PBX* (Fig. 1). More generally, the fish *RGS1*/*RGS16* region contains a number of homologs to genes found in the human MHC paralogon on chromosome 9 (Fig. 1).

Several genes of the *RGS1*/*RGS16* region also have paralogs on "minor" MHC paralogous regions, as defined in Flajnik et al. (2012). These regions include, for example, human 19q and 12p (Fig. 2; Online Resource 2), which contain the leukocyte receptor complex and NK cell receptor genes, respectively. Our findings are well consistent with previous studies, which have provided clear evidence that such regions were indeed linked to the MHC before the divergence of tetrapods and teleosts (Flajnik and Du Pasquier 2008; Ohashi et al. 2010).

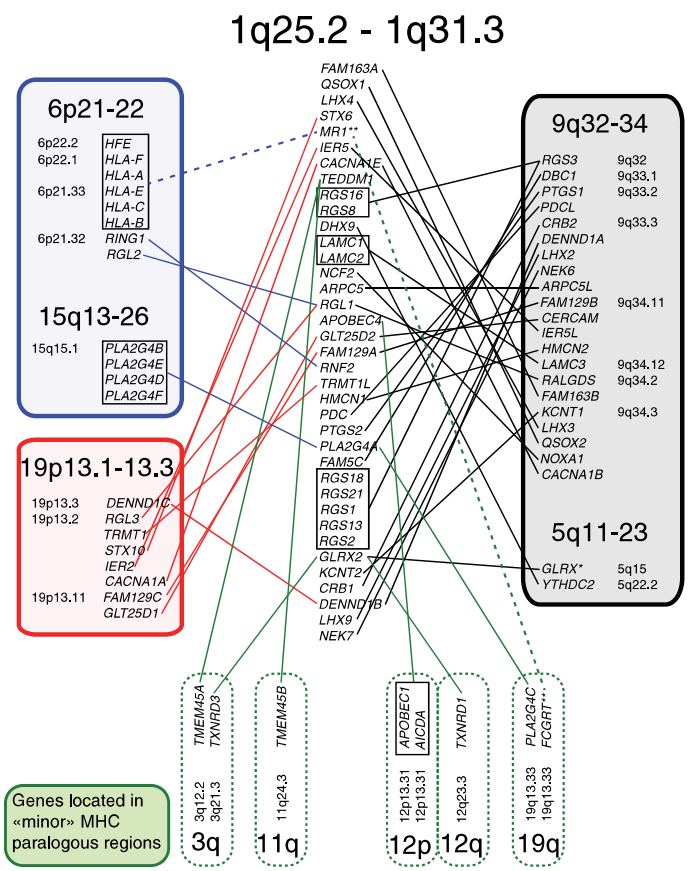


Fig. 2 Ohnologs of genes located in the *RGS1*/*RGS16* region. According to phylogenetic analysis, 35 out of 76 genes located around *RGS1* and *RGS16* in human have copies on the different MHC paralogons, confirming that this region constitutes an extension of the paralogon on the human chromosome 1. The highest number of ohnologs could be retrieved from the MHC paralogon split between genomic locations 9q32–34 and 5q11–23. These results represent a phylogenetic analysis with the maximum likelihood method and bootstrap value of 1,000. *Single asterisk (*)* the gene *GLRX* was not among the closest paralogs of *GLRX2* according to our phylogenetic analysis, yet it was shown as the only paralog found in the database. *Double asterisk (**)* since phylogenetic analysis indicates that MR1 is rather a recent duplicate of the classical MHC molecules than an ancient paralog, it is linked to those genes by a *dashed line* on the figure

The *RGS1*/*RGS16* region origin can be tracked to cephalochordates

The origin of MHC as an antigen-presenting unit is generally seen as the result of the co-option of preexisting genes combined with the recruitment of new genes. The conserved clustering of many genes inside MHC paralogons across vertebrate evolution strongly suggests that (some of) the preexisting genes were already located in the same region before the emergence of the MHC. Selection pressures keeping them in a conserved synteny group are still a matter of debate; it may be either related to immunity-related functions, which could be co-regulated, or to the presence of genomic elements in the vicinity that would determine critical developmental patterns and then freeze the regional architecture (Engström et al. 2007).

Since the *RGS1*/*RGS16* region belongs to the same conserved block as the ancestral MHC, its genes could represent useful markers to track the origin of the proto-MHC before the emergence of vertebrates, as shown for example with a gene orthologous to *LHX4* (and *LHX3*) which mapped near the MHC marker *PBX* in *Ciona* (Kasahara et al. 2004). To further test this idea, we searched for homologs of these genes in genomic scaffolds of the amphioxus, in which a proto-MHC has been previously identified (Abi-Rached et al. 2002; Castro et al. 2004; Danchin and Pontarotti 2004). A full genome assembly is not available for this species, and our analysis was therefore limited to scaffolds. To identify amphioxus orthologs of the human genes present in the *RGS1*/*RGS16* region, we used NCBI blastp to find similar sequences. The correspondence of protein domain structure was then checked for each hit, and phylogenies present in the PhylomeDB or built using FIGENIX and MEGA5 (Online Resource 3) were examined to select true orthologs. In addition to scaffolds containing only one gene of interest, seven scaffolds were found with two or more orthologs of genes from the *RGS1*/*RGS16* region. We retrieved 24 markers (30 if taking the paralogs of *RGS* and *LAMC* into account) out of the 76 genes located in the human *RGS1*/*RGS16* region (Fig. 3; Supplementary Table S2).

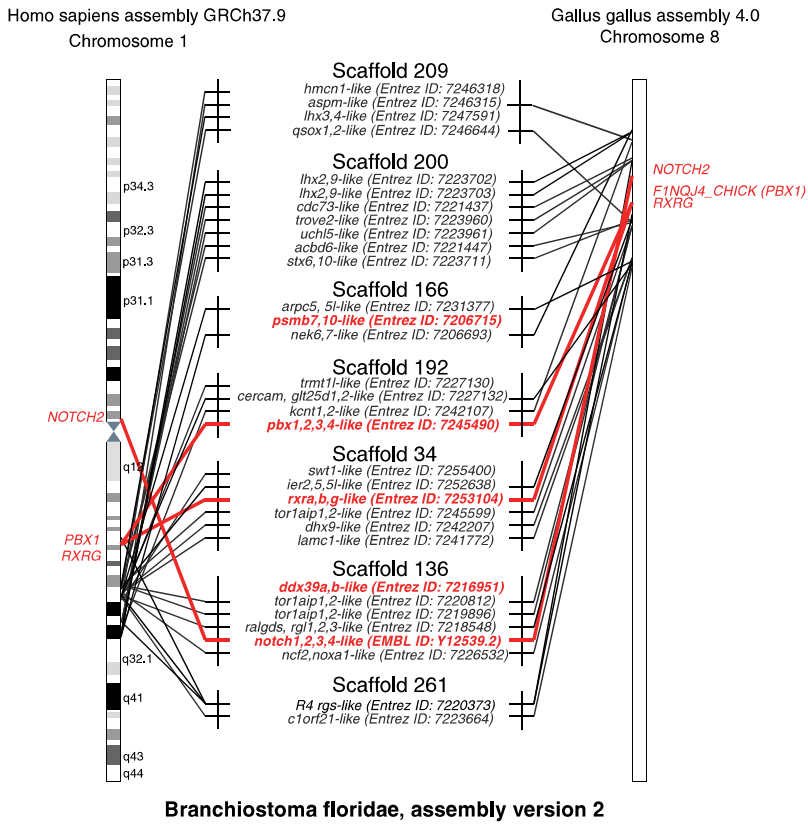


Fig. 3 The *RGS1*/*RGS16* synteny group can be traced to the cephalochordate *B. floridae* proto-MHC. Phylogenetic analysis identifies putative amphioxus orthologs for 24 (30 if taking paralogous genes into account) out of 76 genes of the human *RGS1*/*RGS16* region, located on merely seven genomic scaffolds. The figure is not in scale, and amphioxus genes are shown only when (a) phylogenetically confirmed orthologs of either human genes from the *RGS1*/*RGS16* region or of MHC marker genes and (b) located on one of these seven scaffolds. For a better overview, *lines* connect amphioxus gene names with the approximate genomic locations of their orthologs in human chromosome 1 and chicken chromosome 8. *Red lines* are used to indicate MHC marker genes. As *PSMB7,10* and *DDX39A,B* are not located on the human Chr1 or chicken Chr8, their amphioxus orthologs are mentioned but not linked to their vertebrate counterpart. Four of the seven scaffolds on the figure match those identified as a part of the amphioxus proto-MHC in Abi-Rached et al. (2002), Castro et al. (2004) and Danchin and Pontarotti (2004) and contain gene(s) with the highest similarity to the MHC marker genes *DDX39A,B*, *PBX1,2,3,4*, *NOTCH1,2,3,4*, *PSMB7,10*, or *RXRA,B,G*. In addition, one of the scaffolds contains a protein phylogenetically similar to the vertebrate R4 RGS proteins (see also Online Resource 1)

In addition, scaffolds containing amphioxus proto-MHC markers previously identified by Abi-Rached et al. (2002) and Holland et al. (2001) were retrieved from GenBank (Supplementary Table S2A and B), and four scaffolds out of seven with genes from the *RGS1*/*RGS16* region matched those with MHC markers: the scaffold 166 (GenBank ID no. NW_003101439.1) containing a *PSMB7,10*-like gene, the scaffold 192 (GenBank ID no. NW_003101413.1) containing a *PBX1,2,3,4*-like gene, the scaffold 34 (GenBank ID no. NW_003101413.1) containing a *RXRA,B,G*-like gene, and the scaffold 136 (GenBank ID no. NW_003101539.1) containing both a *DDX39A,B*-like gene and a *NOTCH1,2,3,4*-like gene (Fig. 3). Importantly, previous studies have shown that the scaffolds containing amphioxus markers of the proto-MHC are all linked together on a single chromosome (Castro et al. 2004; Danchin and Pontarotti 2004), thus indicating that the amphioxus orthologs of the *RGS1*/*RGS16* region genes are also found on this linkage group. However, the current annotation of the amphioxus genome does not always match the protein models published by Abi-Rached et al. (2002). For example, the current annotation of the amphioxus *NOTCH*-like gene predicts only a gene consisting of two EGF domains, while the mRNA of a complete amphioxus *NOTCH* with a NOTCH domain and other additional domains has been

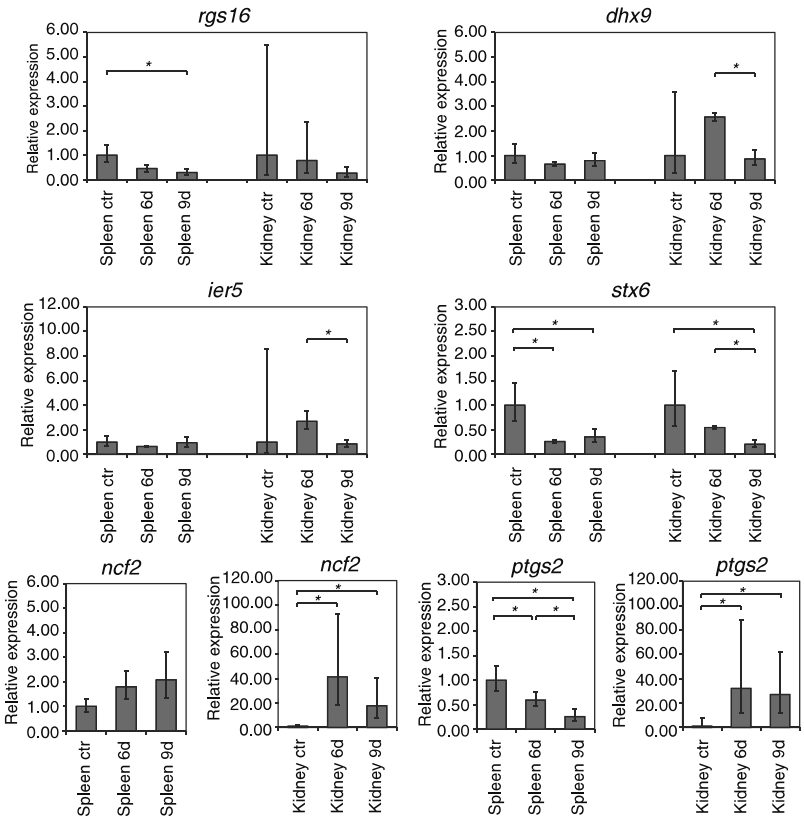


Fig. 4 Expression of *rgs1*/*rgs16* region genes in FV3-infected frogs. *6d* 6 days postinfection, *9d* 9 days postinfection, *ctr* control (noninfected) frogs. The expression of markers of interest was normalized on *gapdh* expression. *Error bars* represent the standard deviation of results obtained from three similarly treated frogs. *p* values were calculated using the unpaired two-tailed Student's *t* test. *$p<0.05$, significant changes in gene expression

already previously characterized (Holland et al. 2001). This sequence's similarity to the human *NOTCH* genes was confirmed phylogenetically (Online Resource 3) and it was also used here in the current study (Fig. 3). In fact, although the copy of *NOTCH* on human chromosome 1 is quite far from *RGS* genes in mammals, it is indeed close to *RGS* homologs in other vertebrates such as *Xenopus* and chicken and has been recognized as part of conserved vertebrate linkage (CVL) group 2, which formed a single ancestral block together with CVL block 5 in the gnathostome ancestor as proposed by Nakatani et al. (2007). Taken together, these observations strongly suggest that genes from the *RGS1*/*RGS16* region have ancestors located in the proto-MHC and could therefore be used as markers for tracking MHC evolution.

The *RGS1*/*RGS16* region contains other IFN-induced/virus-responsive genes and genes involved in immune responses to pathogens in human and in mice

To investigate whether the genes located in the *RGS1*/*RGS16* region may have functions related to immunity, as with many genes found in the MHC, we performed a systematic in silico functional analysis, mining human microarray data available about expression changes imposed by viral infections in the EMBL/EBI Gene Expression Atlas 2.0.15.[1] Many genes in the region were found to be involved in immune responses, especially to virus/host interactions (Online Resource 4). These observations suggest that the *RGS1*/*RGS16* region may be enriched in genes involved in immunity, which is consistent with its MHC-linked origin. A complete demonstration of such enrichment would require a comparison with global transcriptome data in a number of infectious contexts.

As a first step to determine if the immunity-related function of selected markers of the *RGS1*/*RGS16* region is a conserved feature, we studied their expression profile in a frog (*X. tropicalis*) infected by the *Ranavirus* FV3. The genes *ier5*, *mr1*, *ptgs2*, *ncf2*, and *stx6* were chosen, for

[1] http://www.ebi.ac.uk/gxa/qrs?specie_0=homo+sapiens&gprop_0=&gval_0=mr1%2C+ptgs2%2C+ier5%2C+rgs16%2C+ncf2%2C+stx6%2C&fact_1=&fexp_1=UP_DOWN&fmex_1=1&fval_1=IFN+poly%28I%3AC%29+%22CpG+ODN%22+interferon%2C+virus%2C+ifna%2C+ifnb%2C+ifng%2C+influenza%2C+hepatitis

which human microarray data were available about expression changes imposed by viral infections in the EMBL/EBI Gene Expression Atlas 2.0.15[2] (Online Resource 4). The induction of *rnaseL*, an important gene for antiviral immunity located close to *RGS*, could not be tested since it was absent from the frog sequence. The *Xenopus* host immune response against FV3 peaks at day 6 (e.g., height of T cell proliferation in the spleen and T cell infiltration in the kidney), whereas viral clearance is notable at day 9 (Robert et al. 2005). Since FV3 targets mainly the kidney in both *Xenopus laevis* (Robert et al. 2005) and *X. tropicalis* (Chen and Robert 2011), the expression was measured in the kidney and spleen on days 6 and 9 postinfection, in comparison with noninfected controls. As expected, the greatest changes in expression of *RGS1*/*RGS16*-related genes were seen in the kidneys, the main site of infection. The upregulation of *ncf2* (neutrophil cytosolic factor 2) and *ptgs2* (cyclooxygenase 2) in the kidneys of infected frogs at the peak of the anti-FV3 immune response (Fig. 4) was consistent with our previous study reporting the infiltration of FV3-infected *X. laevis* kidneys by CD8 T cells and activated MHC class II positive leukocytes between 3 and 6 days postinfection (Morales and Robert 2007). In naïve frog, most kidney leukocytes are either blood-derived lymphocytes from the blood vessels or resident macrophages. Thus, the increased expression of *ncf2* and *ptgs2* likely results from both an increased number of infiltrating leukocytes and upregulation of gene expression at the site of infection. Accordingly, *ncf2* and *ptgs2* were more upregulated in the kidneys than in the spleen. These results provide further evidence of the involvement of monocytic cells in anti-FV3 immune response and suggest that genes of the *rgs1*/*rgs16* region are involved in the response of myeloid cells.

The increased amounts of *dhx9* and *ier5* transcripts at 6 days postinfection may suggest a local response of the kidney epithelium. Similarly, the decrease of *rgs16* expression in the kidneys is likely due to the effects of type I interferon: in mammals, *RGS16* has been shown to be upregulated by interferon-inducing agents (Timmusk et al. 2009) or downregulated by type I interferon itself (Giorelli et al. 2002), suggesting a complex mechanism of regulation. As a matter of fact, the increased expression of type I IFN-inducible myxovirus resistance1 (*mx1*) gene at 6 days postinfection in the *Xenopus* kidney (De Jesús Andino et al. 2012; Grayfer et al. 2012) is indicative of a strong type IFN response at the site of FV3 infection. In addition, *STX6* was found significantly downregulated in both spleens and kidneys of the infected frogs, as seen in infected human.

[2] http://www.ebi.ac.uk/gxa/qrs?specie_0=homo+sapiens&gprop_0=&gval_0=mr1%2C+ptgs2%2C+ier5%2C+rgs16%2C+ncf2%2C+stx6%2C&fact_1=&fexp_1=UP_DOWN&fmex_1=1&fval_1=IFN+poly%28I%3AC%29+%22CpG+ODN%22+interferon%2C+virus%2C+ifna%2C+ifnb%2C+ifng%2C+influenza%2C+hepatitis

Taken together, our results suggest that at least some genes of the *X. tropicalis RGS1*/*RGS16* region are modulated during a viral infection similar to human genes, providing hints of putative functional basis of the evolutionary conservation of the extended MHC and MHC paralogons. However, there is still no direct evidence that the genomic configuration of the MHC tetrad of paralogs has been maintained by selection pressures related to functional issues.

Conclusion

The *RGS1*/*RGS16* region appears as a conserved synteny group with paralogs in the neighborhood of MHC and other paralogons. This study supports the conclusions of Nakatani et al. (2007), who have previously shown that CVL block 5, together with all the blocks that correspond to MHC paralogons on human chromosomes 1, 5, 6, 9, 15, and 19, have originated from a single ancestral chromosome that existed in the common vertebrate ancestor. The *RGS1*/*RGS16* region contains genes involved in immunity at least in human and frog, which suggests a possibility that they may participate in a conserved functional unit connected to the MHC. Finally, markers from the *RGS1*/*RGS16* region can be useful to track the proto-MHC in non-vertebrates.

Acknowledgments We thank Pierre Pontarotti and Louis du Pasquier for the helpful discussions and Muriel Perron for providing critical reagents. This work was supported by Tallinn University of Technology and the Institut National Recherche Agronomique, by Estonian Research Council Targeted Financing Project SF0140066s09, and by grants ETF8914 from Estonian Science Foundation, IOB-0923772 from the National Science Foundation, and R24-AI-059830 from the National Institutes of Health.

Conflict of interest The authors declare that they have no conflict of interest.

References

Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H (2002) Evidence of en bloc duplication in vertebrate genomes. Nat Genet 31:100–105

Anantharaman V, Abhiman S, de Souza RF, Aravind L (2011) Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system. Gene 475:63–78

Bansal G, Druey KM, Xie Z (2007) R4 RGS proteins: regulation of G-protein signaling and beyond. Pharmacol Ther 116:473–495

Beck S, Geraghty D, Inoko H, Rowen L (1999) Complete sequence and gene map of a human major histocompatibility complex. Nature 401:921–923

Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL (2012) Domain enhanced lookup time accelerated BLAST. Biol Direct 7:12

Burt DW (2002) Origin and evolution of avian microchromosomes. Cytogenet Genome Res 96:97–112

Cahir-McFarland ED, Carter K, Rosenwald A, Giltnane JM, Henrickson SE, Staudt LM, Kieff E (2004) Role of NF-kappa B in cell survival and transcription of latent membrane protein 1-expressing or Epstein–Barr virus latency III-infected cells. J Virol 78:4108–4119

Castro LF, Furlong RF, Holland PW (2004) An antecedent of the MHC-linked genomic region in amphioxus. Immunogenetics 55:782–784

Chen G, Robert J (2011) Antiviral immunity in amphibians. Viruses 3:2065–2086

Chen G, Ward BM, Yu KH, Chinchar VG, Robert J (2011) Improved knockout methodology reveals that frog virus 3 mutants lacking either the 18K immediate-early gene or the truncated vIF-2alpha gene are defective for replication and growth in vivo. J Virol 85:11131–11138

Danchin EG, Pontarotti P (2004) Towards the reconstruction of the bilaterian ancestral pre-MHC region. Trends Genet 20:587–591

Danchin E, Vitiello V, Vienne A, Richard O, Gouret P, McDermott MF, Pontarotti P (2004) The major histocompatibility complex origin. Immunol Rev 198:216–232

De Jesús Andino F, Chen G, Li Z, Grayfer L, Robert J (2012) Susceptibility of *Xenopus laevis* tadpoles to infection by the ranavirus Frog-Virus 3 correlates with a reduced and delayed innate immune response in comparison with adult frogs. Virology 432:435–443

Engström PG, Ho Sui SJ, Drivenes O, Becker TS, Lenhard B (2007) Genomic regulatory blocks underlie extensive microsynteny conservation in insects. Genome Res 17:1898–1908

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Flajnik MF, Du Pasquier L (2008) Evolution of the immune system. In: Paul WE (ed) Fundamental immunology, 6th edn. Wolters Kluwer/Lippincott Williams & Wilkins, Philadelphia, pp 56–124

Flajnik MF, Kasahara M (2001) Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. Immunity 15:351–362

Flajnik MF, Kasahara M (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet 11:47–59

Flajnik MF, Tlapakova T, Criscitiello MF, Krylov V, Ohta Y (2012) Evolution of the B7 family: co-evolution of B7H6 and NKp30, identification of a new B7 family member, B7H7, and of B7's historical relationship with the MHC. Immunogenetics 64:571–590

Giorelli M, Livrea P, Defazio G, Iacovelli L, Capobianco L, Picascia A, Sallese M, Martino D, Aniello MS, Trojano M, De Blasi A (2002) Interferon beta-1a counteracts effects of activation on the expression of G-protein-coupled receptor kinases 2 and 3, beta-arrestin-1, and regulators of G-protein signalling 2 and 16 in human mononuclear leukocytes. Cell Signal 14:673–678

Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin EG (2005) FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinforma 6:198

Grayfer L, Andino Fde J, Chen G, Chinchar GV, Robert J (2012) Immune evasion strategies of ranaviruses and innate immune responses to these emerging pathogens. Viruses 4:1075–1092

Hallböök F, Wilson K, Thorndyke M, Olinski RP (2006) Formation and evolution of the chordate neurotrophin and Trk receptor genes. Brain Behav Evol 68:133–144

Holland LZ, Abi-Rached L, Tamme R, Holland ND, Kortschak D, Inoko H, Shiina T, Burgtorf C, Lardelli M (2001) Characterization and developmental expression of the amphioxus homolog of Notch (AmphiNotch): evolutionary conservation of multiple expression domains in amphioxus and vertebrates. Dev Biol 232:293–507

Huerta-Cepas J, Capella-Gutierrez S, Pryszcz LP, Denisov I, Kormes D, Marcet-Houben M, Gabaldón T (2011) PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. Nucleic Acid Res 39:D556–D560

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

Kasahara M (1999) The chromosomal duplication model of the major histocompatibility complex. Immunol Rev 167:17–32

Kasahara M, Hayashi M, Tanaka K, Inoko H, Sugaya K, Ikemura T, Ishibashi T (1996) Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. Proc Natl Acad Sci U S A 93:9096–9101

Kasahara M, Suzuki T, Du Pasquier L (2004) On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. Trends Immunol 25:105–111

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, Jindo T, Kobayashi D, Shimada A, Toyoda A, Kuroki Y, Fujiyama A, Sasaki T, Shimizu A, Asakawa S, Shimizu N, Hashimoto S, Yang J, Lee Y, Matsushima K, Sugano S, Sakaizumi M, Narita T, Ohishi K, Haga S, Ohta F, Nomoto H, Nogata K, Morishita T, Endo T, Shin-I T, Takeda H, Morishita S, Kohara Y (2007) The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714–719

Katsanis N, Fitzgibbon J, Fisher EM (1996) Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. Genomics 35:101–108

Kim DH, Lim JJ, Lee JJ, Kim DG, Lee HJ, Min W, Kim KD, Chang HH, Endale M, Rhee MH, Watarai M, Kim S (2012) RGS2-mediated intracellular Ca^{2+} level plays a key role in the intracellular replication of *Brucella abortus* within phagocytes. J Infect Dis 205:445–452

Klein J (1986) Natural history of the major histocompatibility complex. Wiley, New York

Lee HK, Yeo S, Kim JS, Lee JG, Bae YS, Lee C, Baek SH (2010) Protein kinase C-eta and phospholipase D2 pathway regulates foam cell formation via regulator of G protein signaling 2. Mol Pharmacol 78:478–485

Morales HD, Robert J (2007) Characterization of primary and memory CD8 T-cell responses against ranavirus (FV3) in *Xenopus laevis*. J Virol 81:2240–2248

Muffato M, Louis A, Poisnel CE, Roest Crollius H (2010) Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. Bioinformatics 26:1119–1121

Nakatani Y, Takeda H, Kohara Y, Morishita S (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res 17:1254–1265

Ohashi K, Takizawa F, Tokumaru N, Nakayasu C, Toda H, Fischer U, Moritomo T, Hashimoto K, Nakanishi T, Dijkstra JM (2010) A molecule in teleost fish, related with human MHC-encoded G6F, has a cytoplasmic tail with ITAM and marks the surface of thrombocytes and in some fishes also of erythrocytes. Immunogenetics 62:543–559

Ohno S (1970) Evolution by gene duplication. Springer, New York

Paganini J, Gouret P (2012) Reliable phylogenetic trees building: a new web interface for FIGENIX. Evol Bioinform Online 8:417–421

Patten M, Bunemann J, Thoma B, Kramer E, Thoenes M, Stube S, Mittmann C, Wieland T (2002) Endotoxin induces desensitization of cardiac endothelin-1 receptor signaling by increased expression of RGS4 and RGS16. Cardiovasc Res 53:156–164

Pertseva MN, Shpakov AO (2009) The prokaryotic origin and evolution of eukaryotic chemosignaling systems. Neurosci Behav Physiol 39:793–804

Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-

Gutiérrez EL, Dubchak I, Garcia-Fernàndez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin-I T, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064–1071

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R (2005) InterProScan: protein domains identifier. Nucleic Acids Res 33:W116–W120

Riekenberg S, Farhat K, Debarry J, Heine H, Jung G, Wiesmuller KH, Ulmer AJ (2009) Regulators of G-protein signalling are modulated by bacterial lipopeptides and lipopolysaccharide. FEBS J 276:649–659

Robert J, Morales H, Buck W, Cohen N, Marr S, Gantress J (2005) Adaptive immunity and histopathology in frog virus 3-infected *Xenopus*. Virology 332:667–675

Shi GX, Harrison K, Han SB, Moratz C, Kehrl JH (2004) Toll-like receptor signaling alters the expression of regulator of G protein signaling proteins in dendritic cells: implications for G protein-coupled receptor signaling. J Immunol 172:5175–5184

Siderovski DP, Willard FS (2005) The GAPs, GEFs, and GDIs of heterotrimeric G-protein alpha subunits. Int J Biol Sci 1:51–66

Sierra DA, Gilbert DJ, Householder D, Grishin NV, Yu K, Ukidwe P, Barker SA, He W, Wensel TG, Otero G, Brown G, Copeland NG, Jenkins NA, Wilkie TM (2002) Evolution of the regulators of G-protein signaling multigene family in mouse and human. Genomics 79:177–185

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Timmusk S, Merlot E, Lovgren T, Jarvekulg L, Berg M, Fossum C (2009) Regulator of G protein signalling 16 is a target for a porcine circovirus type 2 protein. J Gen Virol 90:2425–2436

Tran T, Paz P, Velichko S, Cifrese J, Belur P, Yamaguchi KD, Ku K, Mirshahpanah P, Reder AT, Croze E (2010) Interferon beta-1b induces the expression of RGS1 a negative regulator of G-protein signaling. Int J Cell Biol 2010:529376

Trowsdale J (2001) Genetic and functional relationships between MHC and NK receptor genes. Immunity 15:363–374

Vienne A, Shiina T, Abi-Rached L, Danchin E, Vitiello V, Cartault F, Inoko H, Pontarotti P (2003) Evolution of the proto-MHC ancestral region: more evidence for the plesiomorphic organisation of human chromosome 9q34 region. Immunogenetics 55:429–436

PUBLICATION III

The Proto-MHC of Placozoans, a Region Specialized in Cellular Stress and Ubiquitination/Proteasome Pathways

Suurväli, J., Jouneau, L., Thépot, D., Grusea, S., Pontarotti, P., Du Pasquier, L., Rüütel Boudinot, S., Boudinot, P.

The Journal of Immunology (2014), 193(6):2891-901.
DOI: 10.4049/jimmunol.140117

The Proto-MHC of Placozoans, a Region Specialized in Cellular Stress and Ubiquitination/Proteasome Pathways

Jaanus Suurväli,* Luc Jouneau,† Dominique Thépot,‡ Simona Grusea,§ Pierre Pontarotti,¶ Louis Du Pasquier,‖ Sirje Rüütel Boudinot,* and Pierre Boudinot†

The MHC is a large genetic region controlling Ag processing and recognition by T lymphocytes in vertebrates. Approximately 40% of its genes are implicated in innate or adaptive immunity. A putative proto-MHC exists in the chordate amphioxus and in the fruit fly, indicating that a core MHC region predated the emergence of the adaptive immune system in vertebrates. In this study, we identify a putative proto-MHC with archetypal markers in the most basal branch of Metazoans—the placozoan *Trichoplax adhaerens,* indicating that the proto-MHC is much older than previously believed—and present in the common ancestor of bilaterians (contains vertebrates) and placozoans. Our evidence for a *T. adhaerens* proto-MHC was based on macrosynteny and phylogenetic analyses revealing approximately one third of the multiple marker sets within the human MHC-related paralogy groups have unique counterparts in *T. adhaerens*, consistent with two successive whole genome duplications during early vertebrate evolution. A genetic ontologic analysis of the proto-MHC markers in *T. adhaerens* was consistent with its involvement in defense, showing proteins implicated in antiviral immunity, stress response, and ubiquitination/proteasome pathway. Proteasome genes *psma*, *psmb*, and *psmd* are present, whereas the typical markers of adaptive immunity, such as MHC class I and II, are absent. Our results suggest that the proto-MHC was involved in intracellular intrinsic immunity and provide insight into the primordial architecture and functional landscape of this region that later in evolution became associated with numerous genes critical for adaptive immunity in vertebrates. *The Journal of Immunology*, **2014, 193: 2891–2901.**

The human MHC of jawed vertebrates is defined as a large genetic region of ~4 megabases (Mb) encoding more than 100 genes, approximately half of which are implicated in immunity (1). It is divided into three major regions. Class I and class II regions encode the polymorphic Ag-presenting molecules class I, IIα, and β, factors such as B30.2 proteins, and genes involved in Ag processing pathways such as proteasome genes and *TAP.* The B30.2 domain (2) mediates defense and other functions in several families of proteins, such as butyrophilins and tripartite motif proteins (TRIM) (2–7). The gene-dense class III region encodes several complement components and other genes involved in inflammation (8). The architecture of vertebrate MHCs vary, from the comparatively small MHC of the chicken to the teleost fish, in which class I and II genes are not linked (1, 9–11). However, the various genes of the complex and their basic functions have been generally conserved, and the elements of the human MHC represent the archetypal MHC genes found across jawed vertebrates (8).

The origin of the MHC is incompletely understood, but the MHC backbone is considered ancient and linked to innate immunity. Class I and class II MHC genes have been suggested to originate from families of molecules present in what is called a "proto-MHC region" for convenience but is involved in innate immunity (12, 13). This ancient backbone could be inherited from the invertebrate ancestors of deuterostomians and protostomians (14–17) (Fig. 1A), from a proto-MHC region. It has been proposed that *cis*-duplications and translocations of this single ancestral region occurred during evolution, leading to three original complexes on three different chromosome segments: a primordial MHC that gave rise to the set of genes involved in Ag presentation; a neurotrophin complex that led to genomic regions comprising neurotrophin receptors and the Leukocyte Receptor Complex; and a third genetic complex, the tunicate MHC-related complex JAM-NECTIN (JN) that is at the origin of paralogous regions containing many Ig superfamily members. These proteins have cell adhesion properties (JAM, NECTINs, poliovirus receptors' family members) and are often implicated in the biology of lymphocytes, NK cells, and other leukocytes (CD96, cd155, cd112, JAM B, and CTX family members) (15, 16, 18, 19). Pairs of structurally related molecules can act as receptor–ligand systems and mediate interactions between leukocytes and endothelia, linking this third proto-MHC derived complex to immunity.

According to the hypothesis originally proposed by Susumu Ohno (20), these three genetic regions were duplicated twice at an early stage of vertebrate evolution, leading to three independent tetrads of paralogous regions (paralogons; see Fig. 1B for definitions) with some overlaps (e.g., B7 receptors were spread over those three regions) (21) and breaks. Ohno's hypothesis of two

*Department of Gene Technology, Tallinn University of Technology, Tallinn 12618, Estonia; †Institut National de la Recherche Agronomique, Virologie et Immunologie Moléculaires, F-78350 Jouy-en-Josas, France; ‡Institut National de la Recherche Agronomique, Biologie du Développement et Reproduction, F-78350 Jouy-en-Josas, France; §Institut de Mathématiques de Toulouse, Université Paul Sabatier, F-31000 Toulouse, France; ¶Université Aix Marseille, Centre National de la Recherche Scientifique, Centrale Marseille, F-13331 Marseille, France; and ‖Zoology and Evolutionary Biology Department, University of Basel, CH-4051 Basel, Switzerland

Received for publication May 8, 2014. Accepted for publication July 15, 2014.

This work was supported by the Institut National de la Recherche Agronomique, the University of Basel, and the Tallinn University of Technology.

Address correspondence and reprint requests to Dr. Pierre Boudinot or Dr. Sirje Rüütel Boudinot, Fish Infection and Immunity, Institut National de la Recherche Agronomique, Virologie et Immunologie Moléculaires, 78350 Jouy-en-Josas, France (P.B.) or Department of Gene Technology, Tallinn University of Technology, Tallinn 12618, Estonia (S.R.B.). E-mail addresses: pierre.boudinot@jouy.inra.fr (P.B.) or sirje.boudinot@gmail.com (S.R.B.)

The online version of this article contains supplemental material.

Abbreviations used in this article: ACLG, ancestral chordate linkage group; GO, gene ontology; GRB, genomic regulatory block; JN, JAM-NECTIN; Mb, megabase; RBBH, reciprocal best blast hit; TRIM, tripartite motif proteins.

www.jimmunol.org/cgi/doi/10.4049/jimmunol.1401177

cycles of whole genome duplications in ancestral vertebrates is supported by the study of *hox* complexes, and it applies to the MHC-related complexes as well as other regions (22–30). In the human genome, there are three main paralogons of the MHC on chromosomes 1 (1q21-q25/1p11-[^{32}P]), 9 (9q32-q34), and 19 (19p13.1-p13.3) (22–24). In addition, there are smaller fragments, such as those located on 15q13-q26 and 5q11-q23, translocated from the MHC paralogons on chromosomes 6 and 9, respectively (25). Four paralogons of neurotrophin are on chromosomes 1 (1q32-44, 1p13), 14 (14q11-q32), and 19 (19q13-q14) that contains the leukocyte receptor complex, and 11 (11q12-q13, 11q23-q24, 11p12-p15) (18). Additional fragments of the neurotrophin group are on chromosomes 2 (2p12-p23), 20 (20p11-p12), and 12 (12q22-q24, 12p11-p13) with the NK complex. Finally, linkage groups corresponding to the JN complex are on chromosomes 1 (1q22-24), 3 (3q13-21), 11 (11q23-25), 21 (21q21-22), and 19q (again with some leukocyte receptor complex elements) (15, 16, 31) (see Fig. 1C for a map of all human MHC related paralogons).

The concentration of genes involved in immunity linked to class I and II MHC regions might confer a selective advantage (32). It is unknown whether the conservation of macrosynteny across vertebrates might be partly explained by inheritance independent of constraints imposed by selective advantages of natural selection. However, the conservation of linked MHC/neurotrophin/JN complexes in invertebrates suggests a significant role of such constraints. In chordates, there is a proto-MHC in the amphioxus *Branchiostoma floridae* (33, 34) (Fig. 1B), and a genetic complex related to JN exists in the tunicate *Ciona intestinalis* (14–16), whereas a clearly MHC-related region was not identified (14). There is conserved synteny between human MHC paralogons and a genomic region of the protostome *Drosophila melanogaster* (17). There is scant information on MHC-related regions in other groups of invertebrates, especially in the most ancient ones.

Placozoans constitute one of the most basal branches of the evolutionary tree of Metazoans (35) (Fig. 1A); it is generally accepted that they are a sister group to bilaterians and cnidarians, such as corals and jellyfish (36, 37), and there is increasing evidence they are the most basal branches of metazoans (38–40).

Since segments of large-scale genome organization are similar in vertebrates and the placozoan *Trichoplax adhaerens* (35), this species is a useful model to investigate the origin of the MHC and MHC-related genetic complexes in ancestral metazoans. *T. adhaerens* is the only recognized species of the phylum Placozoa, although a greater degree of diversity has been proposed for this phylum (41–44). *T. adhaerens* is a small (100–200 μm), disc-shaped marine organism consisting of two epithelial layers enclosing a layer of multinucleate fiber cells. Only four cell types have been identified; it lacks most of the elements associated with multicellular organisms, such as complex organ systems and extracellular matrix (43, 45). The genome of *T. adhaerens* is 100 megabases long with twelve pairs of chromosomes and ~11,500 predicted protein-coding genes (35), of which 6516 have been found expressed in the first comprehensive proteome analysis (46). The homologs of these genes are often associated with complex developmental and signaling pathways and are highly conserved in metazoans (35, 47). Importantly, *T. adhaerens* scaffolds have been mapped on the reconstruction of ancestral chordate linkage groups (ACLGs) and extensive phylogenetic analyses performed for conserved regions (35). The ACLGs are defined from conserved synteny between the amphioxus and vertebrates genomes, and they correspond to the gene content of a putative chromosome of their last common ancestor (48). Hence, the ACLGs are a key generic reference for studying the conservation of blocks of synteny among chordates, vertebrates (including humans), and other species.

In this study, we examined *T. adhaerens* for the presence of genomic regions enriched in archetypal markers of proto-MHC and proto-neurotrophin complexes. Identified markers of the proto-MHC region were evaluated by gene ontology analysis to test whether such markers contained genes suggestive of immune-related functions, such as stress response and ubiquitination/proteasome pathways.

Materials and Methods

Identification and analysis of T. adhaerens *counterparts of human genes located in MHC-, neurotrophin-, and JN-related paralogons*

Human protein-coding genes located within the genomic regions described as MHC-, neurotrophin-, and JN paralogons in (21, 49), and within the MHC-related region defined by the *R4 RGS* genes (50) were extracted from Ensembl Genome Browser release 70 (human genome assembly GRCh37; 7957 genes). The sequence of the longest protein isoform encoded by each gene was extracted using Ensembl Biomart. Protein sequences (human queries) were blasted using Δ-BLAST (51) against the proteome of *T. adhaerens* (assembly ASM15027v1 from release 19 of Ensembl Genomes). The e-value cutoff was set at 1e-10, and 5,865 human sequences retrieved 2540 unique best blast hits in *T. adhaerens* scaffolds. To target counterparts of the MHC-related regions, we selected *T. adhaerens* scaffolds with highest density of best blast hits—that is, scaffolds in which the proportion of hits of at least one human gene was higher than a given threshold. A density threshold was set from the distribution at 23% to optimize the contrast between the groups, selecting the scaffolds 1, 2, 3, 6, 7, 9, 10, 12, 14, 15, 22, 28, 31, 34, 35, 40, and 42 for further analysis. Protein sequences corresponding to the 6721 genes located within these scaffolds were extracted using Ensembl Biomart and blasted back against the human proteome using Δ-BLAST with an e-value cutoff of 1e-10. This second blast analysis (using *T. adhaerens* queries) identified 1273 human genes for which the *T. adhaerens* best blast hit matched the initial human query sequence when "back" blasted against the human proteome. Such human and *T. adhaerens* entries identified by the reciprocal blast analyses were designated as human and *T. adhaerens* reciprocal best blast hits (RBBH).

Identification of a putative T. adhaerens *proto-MHC*

To analyze the distribution of RBBH across *T. adhaerens* scaffolds, the number of RBBH per total number of genes (RBBH density) was calculated within 1 Mb blocks across scaffolds (or per scaffold for scaffolds shorter than 1 Mb). A threshold was set from the distribution at 15% to optimize the contrast between the groups, and identified a set of 1-Mb regions with a high density of RBBHs. Regions connecting blocks with high density of RBBH were included in this set. These RBBH-enriched regions were further analyzed as to which of the three sets of paralogons (MHC-related, neurotrophin-related, or JN) of the RBBH were related. For each block, the percentage of neurotrophin-related, MHC-related, and JN-related among all RBBH was calculated in two ways: 1) all RBBH with human counterparts on the respective paralogons were included, including regions where different paralogons overlap in the human genome; genes located in these overlapping regions were designated ambiguous and were considered for the percentage calculations for both overlapping paralogons; 2) RBBHs with a human counterpart in the ambiguous regions were omitted, and the percentages were calculated based on the number of genes with human counterpart located in nonambiguous neurotrophin-related, MHC-related, and JN-related regions. If the percentage of the dominant set among all RBBH exceeded 75% in a block, this block was considered MHC-related, neurotrophin-related, or JN-related. The putative *T. adhaerens* proto-MHC was identified as a collection of all RBBH-enriched regions (located on scaffolds 2, 3, 7, 9, 10, and 15) where MHC-related RBBH were dominant. Out the 1198 genes in these regions, 307 were found to be the *T. adhaerens* counterparts of genes located in the human MHC-related paralogons.

Statistical analysis

The significance of marker clustering in *T. adhaerens* proto-MHC region was tested using a statistical test developed in (52). This test is adapted to approaches in which one starts with a fixed, reference genomic region in the genome of a certain species A (e.g., human) and searches for orthologous regions in the genome of another species B (e.g., *T. adhaerens*). The test analyzes the probability of finding a number of genes clustered in a given region rather than randomly distributed in the whole genome B,

given the size of the region and the size of the whole genome B. Importantly, the size is computed as the number of genes not bp. Qualitatively, the statistical significance increases when the size of the region in which markers are clustered decreases and when the size of the whole genome investigated increases. Notably, the total numbers of genes in humans and *T. adhaerens* differ only by a factor of 2 (~20,000 versus 10,000, respectively). The test is based on a compound Poisson approximation for computing the *p* value of an orthologous gene cluster under the null hypothesis of random gene order. A critical feature of the method is that it accounts for the existence of multigene families—that is, the existence of multiple counterparts (co-orthologs) in the genome B for genes of the reference region (e.g., the human MHC paralogons); to do so, the co-orthologs of the target genome B are weighted in inverse proportion to the size of the multigene family to which they belong. If a gene from the starting region has a unique ortholog in the target genome, this ortholog has a weight of 1. If a gene from the starting region has *k* counterparts in the target genome, then each of these co-orthologs will be given a weight of 1/*k*. The weight of a given region of the genome B (e.g., the *T. adhaerens* proto-MHC region) is defined as the sum of the weights of the orthologs belonging to it. The *p* value of this region is the probability, under the null hypothesis of random gene order, of finding somewhere in the genome B a region of higher weight/smaller length, and is computed using a compound Poisson approximation, as explained previously (52).

Gene ontology enrichment analysis

BiNGO, a plugin of Cytoscape (53), was used to look for local enrichment of gene ontology (GO) terms relating to biological processes among human counterparts of the 307 MHC-related genes located in the *T. adhaerens* proto-MHC. The list of all 3259 genes located within the human MHC paralogons according to Ensembl 70 was used as a reference set. As every GO term was subjected to statistical test for enrichment, Benjamini and Hochberg false discovery rate *p* value adjustment was applied to correct for multiple testing (54, 55) and performed by BiNGO.

Microsynteny analysis

Conserved microsynteny between *T. adhaerens* proto-MHC and human MHC paralogons were defined as sets of RBBH genes that are located in close proximity to each other in both humans and *T. adhaerens* (i.e., separated by 30 genes or fewer). A linkage was considered a conserved microsynteny only when three or more such genes were linked in such a way in both humans and in *T. adhaerens*. These microsynteny gene sets were then used to test for the presence of conserved DNA-binding transcription factors using a list of 1988 such genes from humans (56).

Analysis of noncoding regions of the T. adhaerens *proto-MHC*

To identify putative regulatory regions, we looked for noncoding sequences conserved between *T. adhaerens* proto-MHC and the human genome. Using the masked sequence of *Trichoplax* genome (from release 19 of Ensembl Genomes), we first masked all the exons by stretches of N. The masked sequence of the proto-MHC regions was scanned for stretches of poly(N) and all stretches of more than 10 Ns were deleted, leading to sequences corresponding to introns and intergenic DNA. Fragments of <30 bp were discarded, and Blastn and tBlastX analyses were performed for the remaining 18,860 fragments.

Analysis of the Monosiga brevicollis *genome*

A publicly available list of *Mus musculus* best blast hits for the genes of *M. brevicollis* was obtained from the JGI Genome Browser. Human MHC paralogon genes corresponding to these mouse genes were determined using Ensembl Biomart. The list of 395 human-*Monosiga* gene pairs was compared with our set of human–*T. adhaerens* RBBH, including those with *T. adhaerens* counterpart located in the proto-MHC. Human genes with a counterpart in either one or both organisms were then analyzed for gene ontology enrichment as described above. The complete list of *M. brevicollis* genes with human counterparts located on MHC paralogons was further tested for microsynteny with humans as described above for *T. adhaerens*.

Results

Identification of putative proto-MHC and proto-neurotrophin genomic regions in T. adhaerens

To determine whether the most basal group of Metazoans possessed MHC-related paralogons, we examined the genome of *T. adhaerens* for conserved blocks of synteny corresponding to such genomic segments (Fig. 1).

We recently found that the *RGS1/RGS16* region, located next to the MHC paralogon on human chromosome 1, provides useful markers to investigate the origins and the evolution of the MHC in invertebrates (50). The best blast hits of most of these markers were clustered in one 9.7-Mb scaffold (scaffold 2) of *T. adhaerens*, which also contained homologs of typical MHC markers. We therefore undertook a systematic survey of *T. adhaerens* homologs of all 7957 human genes located within the 19 genomic regions corresponding to MHC-, neurotrophin-, and JN- paralogons (21, 25, 49, 50). Sequences of the longest protein isoforms encoded by each of these genes were extracted from the human genome assembly and were used as query for reciprocal Δ-BLAST (51) searches against the predicted proteome of *T. adhaerens*. Forward and reverse blast analyses identified 1273 pairs of markers for which the *T. adhaerens* gene was the best hit of the forward analysis, and the human sequence used as an initial query in the forward blast was retrieved as the best blast hit of the reverse analysis. Human and *T. adhaerens* entries of such pairs identified by reciprocal blast will be designated below as human and *T. adhaerens* RBBH, respectively.

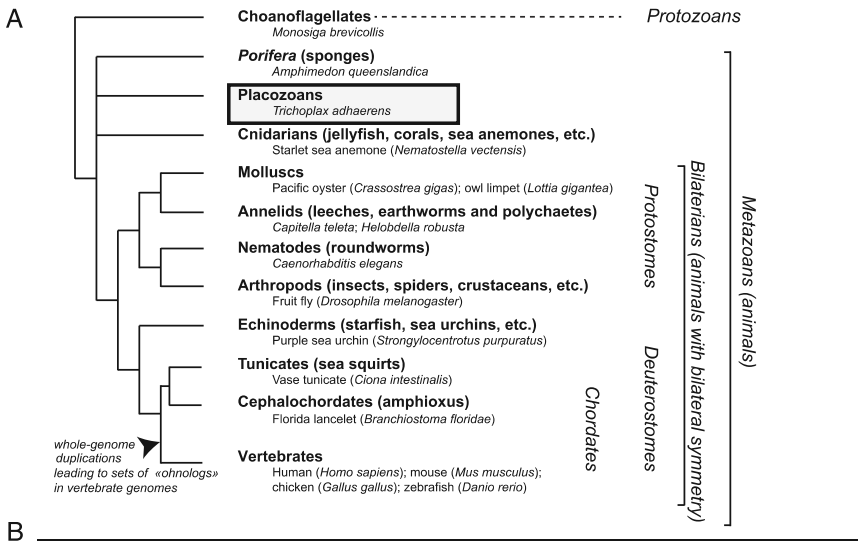
To analyze the distribution of *T. adhaerens* RBBH across genomic scaffolds, the density (ratio between the number of RBBH and the total number of genes) was calculated within 1 Mb blocks across scaffolds, or per scaffold for scaffolds shorter than 1 Mb. Among blocks containing more than 15% RHHB, we identified those specifically enriched in counterparts of human MHC, neurotrophin, and JN tetrads (Fig. 2), for which we hypothesized these regions represented *T. adhaerens* counterparts (Fig. 2). Many human genes from the MHC-related paralogons had a counterpart on the *T. adhaerens* scaffold 2 that we identified in our initial screen. However, a high density of such genes was also found on parts of scaffolds 3, 7, 9, 10, and 15 (Fig. 2). Human genes from the neurotrophin-related paralogons had counterparts in other regions of the *T. adhaerens* genome mostly scattered across scaffold 1, but also located on scaffolds 6, 9, 12, 22, 31, 34, 40, and 42. In contrast, we did not find any region enriched mostly in markers from the third set of paralogons (JN) identified by Du Pasquier et al. (15, 16), with the exception of the very small scaffold 35. This observation suggested that this tetrad might have been produced by a later duplication during the evolution of bilaterians. However, alternative scenarios cannot be excluded; these could not be properly tested in this study because our approach would miss markers that evolve very fast or do not fall into enriched regions.

To test the validity of our approach, we applied it to search for the counterparts of MHC, neurotrophin, and JN paralogons in a vertebrate species. A similar procedure as described above for *T. adhaerens* was followed for chicken (*Gallus gallus*), and it successfully identified MHC paralogons on chromosomes 8, 10, 17, 25, 28, and Z and putative neurotrophin and JN regions (Supplemental Fig. 1).

These data show that a method that successfully maps MHC paralogons in the chicken also identified candidate counterparts of proto-MHC and proto-neurotrophin regions in the *T. adhaerens* genome.

Confirmation of a proto-MHC in T. adhaerens *by macrosynteny, phylogenetic analyses, and a statistical test of gene distribution*

The *T. adhaerens* genome shows extensive large-scale conservation in genomic organization between placozoans (*T. adhaerens)* and chordates (35). One of the most prominent conserved segments is located on scaffold 2, matching the region identified in this study as the main counterpart of the MHC and MHC ohnologs. For further analysis, we therefore selected each ACLG corresponding to the putative *T. adhaerens* proto-MHC regions from reference (35). The segments of human chromosomes matching



B

Paralogues (paralogous genes.)	Homologous genes within a single species that were produced by duplication and belong to the same gene family
Paralogons (paralogous regions)	Set of four genomic regions that contain closely linked sets of paralogs (ohnologs, see above) as a result of tw o rounds of whole-genome duplication early in the vertebrate
Orthologues (orthologous genes)	Homologous genes that diverged by speciation events. Co-orthologues refer to paralogues (within a given species), that are both orthologous to a gene in another species
Ohnologues	Paralogs, named after Susumu Ohno, that emerged close to the origin of vertebrates by two rounds of whole genome duplication. Many such paralogues are not distributed at random in vertebrate genomes, but are located in clusters corresponding to paralogous regions, on four (or secundary more) chromosomes.
Plesiomorphic	Close to the ancestral (primitive) state, (present both in the outgroups and in the ancestors in a phylogenetic tree)
Synteny	Macro-synteny refers to conservation of chromosome-scale linkage, micro-synteny applies to only a few genes at a time. We do not consider the notion of gene order (colinearity) in our use of the term "synteny".

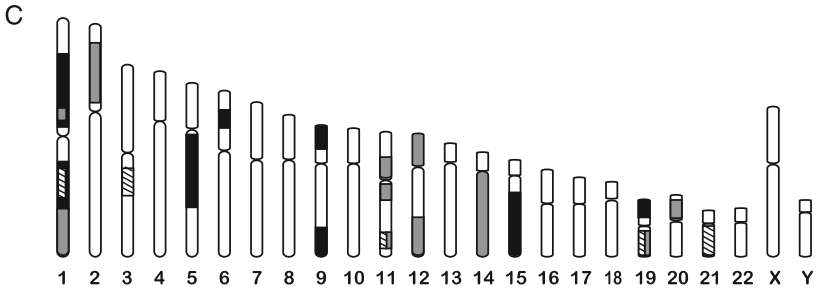


FIGURE 1. Metazoans evolutionary relationships, definitions, and human MHC-related regions. (**A**) Phylogeny of metazoans based on GIGA Community of Scientists (37). (**B**) Definitions of phylogenetic terms used in the text. (**C**) Schematic representation of MHC-related regions belonging to the major histocompatibility, neurotrophin, and JN complexes of the human genome.

these ACLG are shown in Table I. Our candidate regions matched ACLG 8, 10, and 11, which contain essentially all segments of human MHC or MHC ohnologs as described in (49), namely 1p21.1-34.2, 1q23.3-32.1, 5p15.33-q13.2, 5q13.2-31.1, 6p21.2-22.2, 9p22.3-24.3, 9q22.31-34.3, 15q15.3-26.3, 19p13.1-13.2 (Table I). Hence, the common origin of the proto-MHC region identified in this study, and of the MHC set of paralogons described in vertebrates, is confirmed by the conserved macrosynteny mapping and the extensive phylogenetic analysis reported by Srivastava et al. (35).

In addition, we independently assessed the validity of the correspondence between *T. adhaerens* proto-MHC and human MHC paralogons using a statistical test for the genomic distribution of

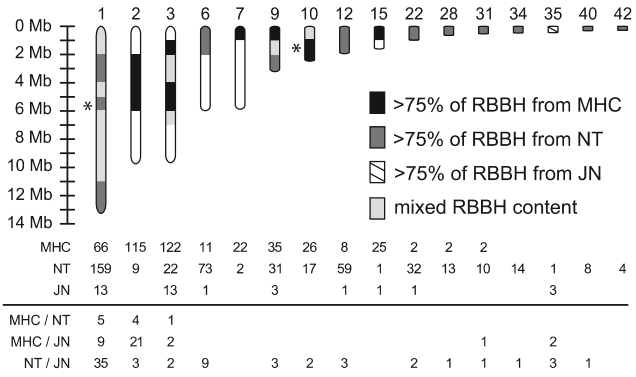


FIGURE 2. *T. adhaerens* RBBH distribution on genomic scaffolds identifies putative counterparts of MHC and MHC-related human paralogons. Scaffolds containing RBBH genes are divided into 1-Mb blocks. Blocks where >15% of all genes were RBBH are highlighted. Different tones or motifs (consistent with Fig. 1) distinguish blocks with best affinity to MHC, neurotrophin, or JN sets. Comparable results were obtained by excluding the human genes located in regions were paralogons overlap (ambiguous regions). Asterisks indicate differences observed when ambiguous regions are taken into account. The number of RBBH genes corresponding to the three sets of paralogons is indicated below each scaffold. NT, neurotrophin.

RBBH (16). Based on a compound Poisson approximation for computing its p value under the null hypothesis of random gene order, this statistical evaluation tested the clustering of the *T. adhaerens* RBBH for MHC and MHC ohnologs within selected regions from scaffolds 2, 3, 7, 9, 10, and 15. Our results showed that the enrichment in the *T. adhaerens* proto-MHC is highly significant ($p < 10^{-50}$). The distributions of conserved markers on *T. adhaerens* scaffolds were also assessed, including the genes that putatively belong to MHC or neurotrophin (MHC or JN, respectively) sets of ohnologs (Supplemental Table I), and similar statistical results were found.

These results demonstrated that the candidate *T. adhaerens* regions identified by reciprocal blast analysis from the human MHC paralogons constitute an ancestral proto-MHC. The conservation of macrosynteny between human and *T. adhaerens* MHC-related regions were supported by orthology relationships, and statistical testing unambiguously rejected a random enrichment of MHC related markers in the *T. adhaerens* proto-MHC region.

Many human MHC markers with paralogous versions on different ohnologs have a unique counterpart in the T. adhaerens *proto-MHC region*

Ohno's theory of whole genome duplications (20) predicts that human sets of paralogs represented in two, three, or four MHC paralogons would generally have a unique counterpart in the genome of a prototypic nonpolyploid invertebrate located in a proto-MHC region. We evaluated whether our results were consistent with this prediction by listing *T. adhaerens* RBBH of at least one gene from human sets of ohnologs with representatives in three or four paralogons. There were three unique *T. adhaerens* genes corresponding to sets of paralogs represented on four paralogons among the 33 such sets present in the human genome and 10 corresponding to sets of paralogs present on three paralogons, among the 106 such sets found in human (Fig. 3). These markers were mainly located in *T. adhaerens* scaffold 2 and include the homologs of several sets of MHC markers such as *RXR*, *PBX*, and *VAV* (25, 29, 49, 57, 58). Among 443 sets of paralogs with representatives on at least two of the MHC paralogons in human, 129 (~30%) had a unique counterpart in the conserved regions of *T. adhaerens.* Thirteen additional sets of paralogs had multiple targets in *T. adhaerens* due to secondary duplications. Overall, the observations indicated that approximately one third of the multiple marker sets have a unique counterpart in *T. adhaerens*, as predicted by Ohno's model of whole genome duplications during early vertebrate evolution.

Table I. Phylogenetic analyses of the *T. adhaerens* proto-MHC support its relationship with human MHC and MHC paralogons

Regions of the Putative *T. adhaerens* Proto-MHC (Scaffold:Location in Mb)	Corresponding ACLGs (11)	Location of the ACLG Blocks in *Homo Sapiens* (11)
2:2,000,000–2,500,000	10 (1.43E-19)[a,b]	1p22; 1q23-32; 6p21-22; 9q32-34[a]
2:3,000,000–3,500,000	10 (1.24E-10)[a]	1p22; 1q23-32; 6p21-22; 9q32-34[a]
2:3,500,000–4,000,000	10 (4.04E-17)[a]	1p22; 1q23-32; 6p21-22; 9q32-34[a]
2:4,000,000–4,500,000	10 (7.55E-22)[a]	1p22; 1q23-32; 6p21-22; 9q32-34[a]
2:4,500,000–5,000,000	10 (9.20E-17)[a]	1p22; 1q23-32; 6p21-22; 9q32-34[a]
2:5,000,000–5,500,000	16 (3.59E-08)	2q13; 3p21; 7p11; 7q11; 7q36; 10p13; 12q13; 17q12; 17q23
2:5,500,000–6,000,000	5 (2.49E-04)	7p21; 16p11; 17p11; 17q21; 17q24; 22q12
3:1,500,000–2,000,000	16 (1.70E-05)	2q13; 3p21; 7p11; 7q11; 7q36; 10p13; 12q13; 17q12; 17q23
3:4,000,000–4,500,000	11 (1.76E-04)[a]	1p12-q23; 5q13-31; 9p22-q22; 15q15-26; 19p13[a]
3:4,500,000–5,000,000	11 (8.43E-21)[a]	1p12-q23; 5q13-31; 9p22-q22; 15q15-26; 19p13[a]
3:5,000,000–5,500,000	11 (3.75E-08)[a]	1p12-q23; 5q13-31; 9p22-q22; 15q15-26; 19p13[a]
7:1–500,000	8 (6.42E-11)[a]	1p21-34; 5p15-q13; 9p22-24; 9q22-32; 19p13[a]
7:500,000–1,000,000	8 (1.51E-08)[a]	1p21-34; 5p15-q13; 9p22-24; 9q22-32; 19p13[a]
9:1–500,000	8 (8.16E-12)[a]	1p21-34; 5p15-q13; 9p22-24; 9q22-32; 19p13[a]
9:500,000–1,000,000	8 (4.11E-14)[a]	1p21-34; 5p15-q13; 9p22-24; 9q22-32; 19p13[a]
10:1,000,000–1,500,000	3 (1.27E-06)	4q12; 4q35; 5q31; 10q11
10:1,500,000–2,000,000	8 (4.85E-08)[a]	1p21-34; 5p15-q13; 9p22-24; 9q22-32; 19p13[a]
10:2,000,000–2,435,506	8 (3.07E-08)[a]	1p21-34; 5p15-q13; 9p22-24; 9q22-32; 19p13[a]
15:1–500,000	11 (8.71E-11)[a]	1p12-q23; 15q15-26; 5q13-31; 9p22-q22; 19p13[a]
15:500,000–1,000,000	16 (4.06E-05)	2q13; 3p21; 7p11; 7q11; 7q36; 10p13; 12q13; 17q12; 17q23

[a]Linkage groups corresponding to MHC paralogons in human.
[b]The p values for significant numbers of ancestral genes are shared between the segments of *T. adhaerens* putative proto-MHC and corresponding ACLGs (11).

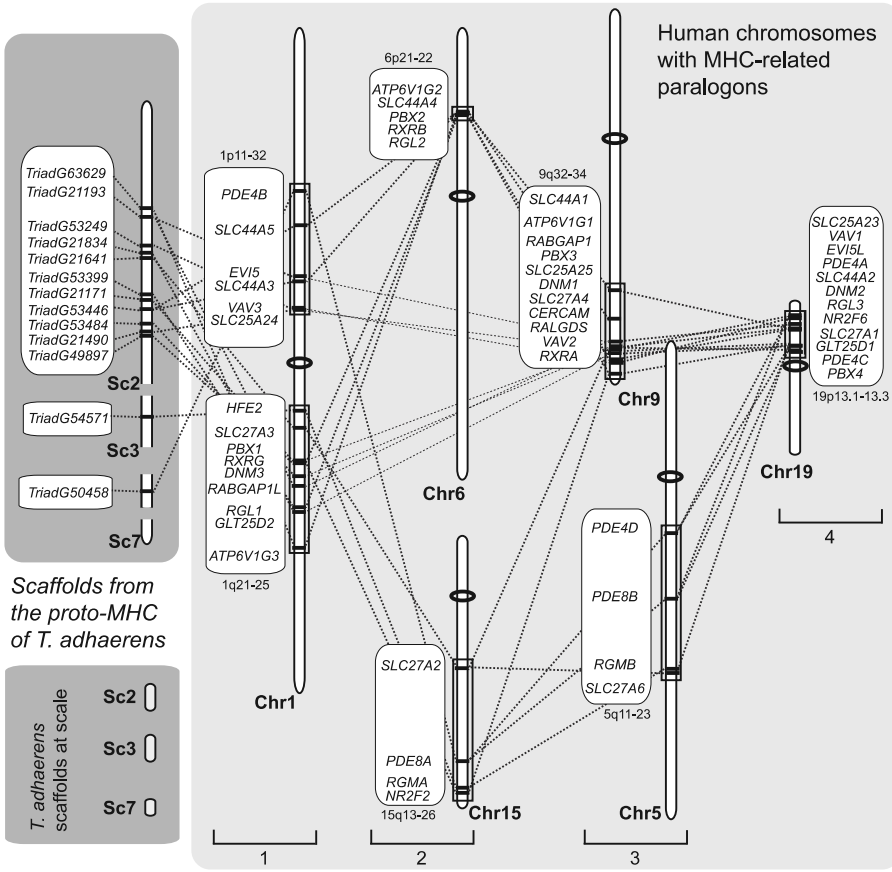


FIGURE 3. Distribution of paralogous genes present in three or four human MHC paralogons and their counterparts within *T. adhaerens* proto-MHC. *T. adhaerens* scaffolds containing relevant markers are represented at scale relative to human chromosomes (*bottom left*) and a zoom of proto-MHC regions is shown (*top left*). Note that scale relates to the chromosome length, and our statistical test for clustering is based on gene numbers. The four MHC paralogons are indicated by brackets numbered 1–4. Rings represent centromeres.

Microsynteny groups conserved between the human MHC ohnologs and the T. adhaerens *proto-MHC do not systematically colocalize with key developmental master genes*

In addition to large-scale synteny conservation between human MHC-related paralogons and *T. adhaerens* proto-MHC, there were nineteen conserved microsynteny gene sets implicating three or more genes with five having five or more genes. Marker genes involved in such sets were located in the same region, but the colinearity between humans and *T. adhaerens* was generally not conserved. A striking example was found on human chromosome 9q32-q34.3 region where 65 markers had a counterpart in *T. adhaerens* scaffold 2 (between 1.9 and 5.2 Mb, see Supplemental Table II), consistent with the MHC paralogon on human chromosome 9 having retained most of the ancestral configuration (i.e., the plesiomorphic organization) (33, 59). In addition, scaffold 2 has four microsyntenies with the RGS1/RGS16 region located on human chromosome 1 (50).

Large gene deserts with enhancers acting over long distances are mostly found in vertebrate genomes, whereas invertebrate metazoans generally have local regulatory controls of expression (60). However, it has been proposed that locked genomic regulatory blocks (GRBs) defined by key developmental transcription factors and their distal enhancers provide an explanation for the maintenance of long-range conserved synteny across vertebrate and invertebrate genomes (61, 62). It is proposed that bystander genes be trapped in the GRBs and thus form conserved syntenic blocks of genes (61, 62). These bystander genes are unrelated to the developmental transcription factor gene defining the GRBs in terms of function, regulation and phylogenetics. According to this hypothesis, observed syntenic regions of the genome are expected containing key developmental transcription factor genes. To test this hypothesis, we looked for such genes in the microsyntenic regions we identified and found no significant enrichment in transcription factors in the microsyntenies of *T. adhaerens*. Six transcription factor genes (*PBX3* [pre–B-cell leukemia homeobox 3], *RXRA* [retinoid X receptor α], *PRDM12* [PR domain-containing protein 12], *GTF3C5* [general transcription factor 3C polypeptide 5], *EDF1* [endothelial differentiation-related factor 1], and *COBRA1* [cofactor of *BRCA1*]) were found in the large conserved gene set (60 genes) from chromosome 9. However, among the other 18 gene sets from conserved microsyntenies, only 3 encompass a transcription factor (*DMRTA2* = *DMRT5* [double sex– and mab-3–related transcription factor 5], *GTF2B* [general transcription factor 2B],

and *CHAF1A* [chromatin assembly factor 1, subunit A]). With the exception of the large microsynteny group from chromosome 9q34, none of the five longest gene sets (five genes or more) encoded transcription factors.

We also looked for noncoding sequences conserved between the *T. adhaerens* proto-MHC and the human genome to identify putative regulatory regions. We did not find significant conserved noncoding sequence with obvious regulatory potential. Only two of the 94 sequences found with Blastn and tblastX showed significant conservation in at least one of selected genomes from representative metazoans: *Amphimedon queenslandica* (Porifera), *Nematostella vectensis* (Cnidaria), *Capitella teleta* and *Helobdella robusta* (Annelida), *Crassostrea gigas* and *Lottia gigantea* (Mollusca), and *Strongylocentrotus purpuratus* (Echinodermata). The first one, *PCNCS53*, a 68 bp sequence from *T. adhaerens* scaffold 7, matched an exon of the human *WDR65* gene located on human chromosome 1 in an MHC paralogon (*e* value = 8 × 1e-10 with exon 12 of transcript *ENST00000372492*) and the WDR65 gene in *A. queenslandica* (*Aqu1.205499; Aqu1.217579*), *N. vectensis* (*NEMVEDRAFT_v1g213283*), *C. teleta* (*CapteG108742*), *L. gigantea* (*LotgiG118162*), and *S. purpuratus* (*SPU_010131*). A detailed analysis revealed that this motif represents a nonannotated exon flanking the gene *TriadG27693*, one of the two *WDR65* genes found in *T. adhaerens* close to each other on Scaffold7. The second hit was *PCNCS89*, a 38-bp sequence from *T. adhaerens* scaffold 3 that was found in *H. robusta*, *C. gigas*, and *L. gigantea*. This sequence appeared to be located in a nonannotated exon at the 5′ end of a well-conserved gene, the sodium channel *scn*.

Taken together these results argue against a predominant role of transcription factor genes in the maintenance of the microsynteny gene sets conserved between humans and *T. adhaerens*.

The gene content of the T. adhaerens *proto-MHC reveals that this region was involved in immunity in the placozoan-bilaterian ancestor*

A tentative set of primordial MHC markers was inferred based on gene conservation between the proto-MHC of *T. adhaerens* and one or several human MHC paralogons. A functional ontology analysis of this gene set was then performed using BiNGO, a plugin of Cytoscape that maps the predominant functional themes of a gene set on the GO hierarchy (53), and computes enrichment calculations in comparison with a reference list. BiNGO identified the key GO terms that were overrepresented in the *T. adhaerens* proto-MHC, in reference to the whole set of genes present in all human MHC-related paralogons. Two groups of GO terms potentially linked to defense processes were found (Table II) (1): proteasome and ubiquitination (15 genes out of 307) and (2) stress response (25 genes of 307). This analysis also identified several enriched GO terms related to metabolic processes, especially RNA metabolism and gene expression control. We then analyzed the list of human genes present in MHC-related paralogons without counterpart within the *T. adhaerens* proto-MHC region. In contrast to those with counterpart, this list was significantly enriched in GO terms with function in adaptive immunity, indicating that many MHC (and MHC paralogon) genes important for immune pathways appeared late in the evolution of metazoans (e.g., MHC class I and class II). This list was also enriched in representatives related to receptors and cell-to-cell recognition, consistent with their absence in *T. adhaerens* where only four cell types have been identified (45).

The results of our GO analysis led us to look for genes encoding proteins with B30.2 domains in *T. adhaerens*. The B30.2 domain, a fusion of a PRY and a SPRY motif (2, 3), is often found in key proteins of immunity encoded in the MHC and MHC related regions of vertebrates and its presence in genes of *T. adhaerens* MHC-related regions would represent another link to immunity. In *T. adhaerens,* 12 proteins with a B30.2 domain were found (Fig. 4). Seven proteins containing a B30.2 domain were located within the proto-MHC or neurotrophin regions, including a TRIM protein with a Ring-B.Box-CC-FN-B30.2 domain structure similar to the one reported in *Nematostella vectensis* (63). This gene is located in the proto-neurotrophin region on scaffold 22, 200 kb and 400 kb from proteasome genes *psmc1* and *psmd8,* respectively. Most genes with a B30.2 domain possess at least one putative human ortholog within MHC or neurotrophin paralogons. Fig. 4 shows that *T. adhaerens* B30.2 sequences cluster with their respective human orthologs with high bootstrap values, indicating that they were already specialized/differentiated into subfamily-specific domains in the common ancestor of placozoans and vertebrates.

Absence of a genomic region enriched in proteasome and stress response genes in the choanoflagellate Monosiga brevicollis

To investigate whether proto-MHC region was a metazoan innovation, we searched for counterparts of human MHC-related markers in the choanoflagellate *Monosiga brevicollis* (genome size 41.6 Mb, 9200 genes) (64). Choanoflagellates are highly

Table II. GO analysis identifies overrepresentation of terms related to proteasome/ubiquitination and stress response within genes conserved between *T. adhaerens* proto-MHC and human MHC paralogons

Pathway	GO Category Accession	GO Category Description	Genes with a Counterpart in the *Trichoplax* Proto-MHC, Mapped within These GO Categories
Proteasome / ubiquitination	6511	Ubiquitin-dependent protein catabolic process	*ARIH1, CDC34, CLPX, EDEM3, FAF1, FZR1, HSPA5, LONP1, PSMA4, PSMB7, PSMD5, RNF11, TOPORS, USP3, USP33*
	10498	Proteasomal protein catabolic process	
	19941[a]	Modification-dependent protein catabolic process	
	43161	Proteasomal ubiquitin-dependent protein catabolic process	
	43632[a]	Modification-dependent macromolecule catabolic process	
	51603[a]	Proteolysis involved in cellular protein catabolic process	
Stress response	6281	DNA repair	*ARNT, ATF6, ATG10, BLM, CCNH, CDK7, CHAF1A, COL4A3BP, DCLRE1B, DHX9, FAN1, FZR1, GNL1, HSPA5, INTS3, LONP1, MAP2K7, MORF4L1, POLG, PRDX6, RXRA, TOPORS, UPF1, USP3, XAB2*
	33554[a]	Cellular response to stress	

[a]GO categories that are also relatively underrepresented among MHC-related human markers without a counterpart in *T. adhaerens* proto-MHC.

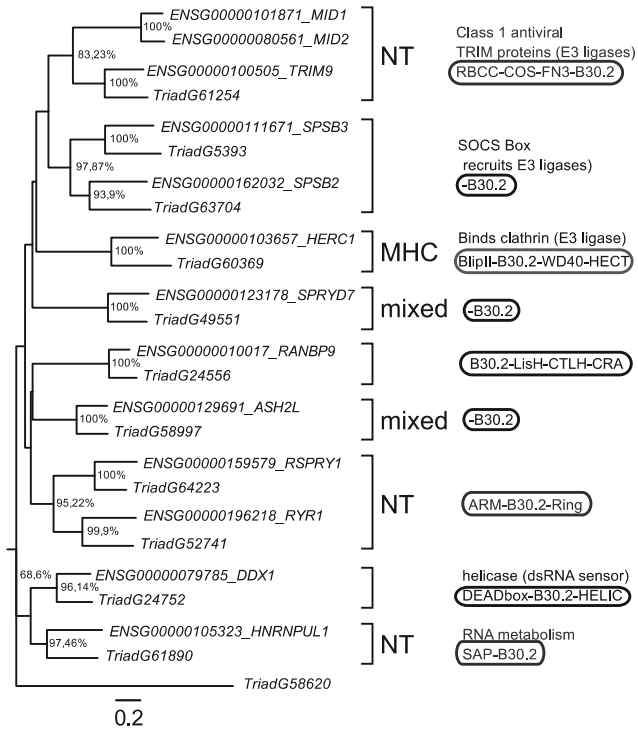


FIGURE 4. Distance tree of B30.2 domains found in *T. adhaerens*. Protein sequences were aligned using Clustal W, and a distance tree was computed using Mega5 (neighbor joining, pairwise deletion, bootstrap value = 1000).

similar to the choanocytes of sponges, and are considered the closest known relatives of metazoans (65–67). Among MHC-related markers with RBBH in *T. adhaerens*, approximately one third had a counterpart in *M. brevicollis* (Fig. 5A). Using BiNGO as above, we looked for enriched GO terms within the different gene subsets and found only markers for cellular response to stress for RBBH present in *T. adhaerens*, whereas markers for proteasome and ubiquitination were enriched both in *T. adhaerens* and *M. brevicollis.* We then examined whether these genes involved in proteasome and ubiquitination might be grouped in a single genomic region within *M. brevicollis*. Fig. 5B shows that these genes are largely dispersed among *M. brevicollis* scaffolds, resulting in the absence of a clustering of genes involved in proteolysis and ubiquitination in this species. However, a subset of MHC-related genes present on the human chromosome region 9q32-34 (59)—the most conserved region among MHC paralogons across vertebrates—appears to be located in the same synteny group in humans, *T. adhaerens* and *M. brevicollis* (Fig. 5C). These markers include two markers of human MHC tetrads, *VAV* and *RABGAP1*, but are not obviously related to immunity. A more extensive analysis of the genome of *M. brevicollis* will be required for a comprehensive picture of the distribution of the markers linked to the genes of metazoan MHC and proto-MHC related markers. However, our observations provided no evidence for a proto-MHC region containing genes involved in innate immunity, proteasome function, and stress responses in this species.

Discussion

The MHC of vertebrates is a large genetic region that determines Ag recognition by T lymphocytes, graft compatibility, and contains genes encoding receptors, cytokines, and effectors of innate immunity. One approach to understand the functional significance of the components of the vertebrate MHC is to reconstruct its evolutionary history. We report a proto-MHC with archetypal markers in one representative of the most basal branch of metazoans, the placozoan *T. adhaerens.* The presence of a proto-MHC exists in the common ancestor of deuterostomes and protostomes as revealed by studies in *Drosophila*. Our results show that it also exists in more primitive branches of the animal kingdom. Placozoan proto-MHC markers include good homologs of genes key to antiviral immunity, stress response, and ubiquitination/proteasome, suggesting that the appearance of class I and II molecules and Ag presentation pathway in vertebrates took advantage of the molecules encoded in this region. In contrast, we did not find evidence of a proto-MHC in the genome of a choanoflagellate, one of the closest known relatives of metazoans among unicellular organisms; it is therefore tempting to speculate that the proto-MHC as a genetic region is an innovation of metazoans, like other key features and pathways such as the TBCEL/coel-1 dependent microtubule function during development and neuronal differentiation (68). However, because the presence of a proto-MHC has not been evaluated in other opisthokonts (e.g., fungi), the apparent lack of proto-MHC in *M. brevicollis* could be due to the loss of the primordial linkage in this species and does not constitute definitive evidence for the absence of proto-MHC in unicellular organisms. A comprehensive survey of protozoan genomes will be necessary to clarify this issue.

The hypothesis of the presence of proteasomes—ancient components involved in the cellular stress response—in the primordial MHC (23) is consistent with our genetic analysis of *T. adhaerens*. The proteasome genes located in *T. adhaerens* are not the functional counterparts of the specialized immunoproteasomes found in vertebrates. Nonetheless, we speculate they are coregulated with genes involved in stress response, pathogen binding, and

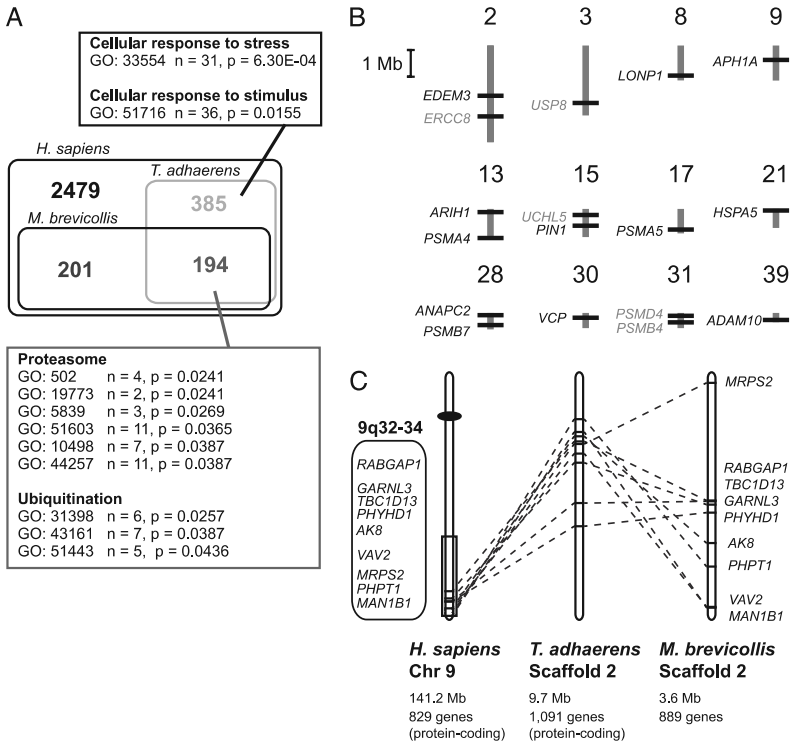


FIGURE 5. MHC-related genes in *Homo sapiens, T. adhaerens*, and *M. brevicollis*. (**A**) Venn diagrams of *T. adhaerens* and *M. brevicollis* MHC-related RBBH. Selected GO terms enriched in different compartments identified by BiNGO analysis are indicated with the number of genes involved (*n*) and the associated *p* value. (**B**) Distribution of the orthologs of markers related to proteasome and ubiquitination in the genomic assembly of *M. brevicollis*. Light gray indicates genes lacking a *T. adhaerens* RBBH counterpart. (**C**) Markers from human 9q32 region involved in conserved synteny with *T. adhaerens* scaffold 2 and *M. brevicollis* scaffold 2. Chromosome and scaffolds are not in scale, and their respective size and protein-coding gene content are indicated.

ubiquitination, and their presence in the region offers the possibility they have been co-opted and selected in the bona fide MHC during vertebrate evolution. In fact, the *T. adhaerens psmb*-like corresponds to the subset from which immunoproteasomes were derived. In addition, the early peptide presentation system may have lacked specialized and inducible proteasomes, as constitutive proteasomes can generate peptides that are presented by MHC I, although with lower efficiency (69, 70).

The association of genes of stress response, ubiquitination, and protein catabolism (proteasome) within the proto-MHC—possibly as a coregulated unit—in *T. adhaerens* is consistent with an ancient functional link between these pathways; stress response induces ubiquitination of pathogens or cellular proteins that are either redirected to new compartments or to degradation by the proteasome. Such a stress/ubiquitination/proteasome cascade has been described in *C. elegans*; DNA damage to germ cells induces a response that elicits resistance to stress as well as activation of the ubiquitination–proteasome system in somatic cells in various tissues (71). This fundamental inflammatory response can be involved in several processes, including aging, adaptation, and defense against pathogens. A specific implication of the proto-MHC–neurotrophin in immunity is further supported by the presence of several B30.2 proteins in these regions. We previously proposed that the association of B30.2 domains with key proteins of immunity (e.g., TRIM, butyrophilin) found in the MHC and MHC-related regions of vertebrates could have an ancient origin (63). The B30.2 domain structurally resembles a β-barrel (2, 3) and allows specific recognition of ligands via the loops at the top of the domain. Of note, several B30.2 domains found in *T. adhaerens* are associated with domains that possess E3 ubiquitin ligase activity, such as the RING in TRIMs. In particular, the *T. adhaerens* TRIM has multiple human co-orthologs within paralogons of the MHC, neurotrophin or JN sets, and all of them share the same domain structure (72). Among them are TRIM1 and TRIM9, two key modulators of the IFN pathway that strongly inhibit viral growth (73). In the absence of IFN, an antiviral activity of the *T. adhaerens* TRIM might proceed via direct binding (and ubiquitination) of viral proteins as for TRIM5 in primates, or via modulation of expression of other antiviral factors. Although TRIM genes and B30.2 domains are conservatively associated with vertebrate MHCs (and paralogons) (7, 63), it is striking that they are also present in the MHC/neurotrophin regions in the ancestor of placozoans and bilaterians. More generally, our data show a preferential location of B30.2 proteins within *T. adhaerens* proto-MHC and proto-neurotrophin, supporting an ancestral and strongly conservative association of this domain associated with viral sensing or defense with the MHC–neurotrophin region as suggested previously (63).

In addition to the stress/ubiquitination/proteasome system and B30.2 domains, we also found the *T. adhaerens* proto-MHC has several genes whose human counterparts are located on chromosome 1 in the "*RGS1/RGS16* region," including a typical *rgs-r4*–like gene. In humans, comparative and responsiveness quantitative trait loci analyses show that this region is critical for antiviral defense (50, 74). The markers from the human *RGS1/RGS16* region that are conserved in the *T. adhaerens* proto-MHC might be components of an ancient antiviral pathway.

Our observations do not reveal the evolutionary mechanisms responsible for the linkage conservation of genes involved in immunity within MHC related regions during evolution. However, it is interesting to compare the evolution of the genes found conserved in the proto-MHC linkage to the one of *Hox/paraHox*

genes. *Hox /paraHox* are clustered in most metazoans because they derive from old successive duplications and their sequential expression is highly regulated (75). In contrast, the constituents of the proto-MHC linkage are multiple, do not derive from the same ancestor gene, and are not found in the same order in different species, suggesting that they are subjected to different constraints.

In conclusion, *T. adhaerens* has retained a simple and primitive organization at both genetic (genomic) and organism levels, and its proto-MHC may reflect the primordial architecture and the functional landscape of this region, which later in evolution became associated with a large number of genes critical for the adaptive immunity in vertebrates.

Acknowledgments

We thank Drs. Franck Bourrat, Daniel Chourrout, Simon Fillatreau, Maria Leptin, and Jean-Pierre Levraud for helpful comments and critical reading of the manuscript, and Dr. Howard Etlinger for improving the English usage.

Disclosures

The authors have no financial conflicts of interest.

References

1. The MHC Sequencing Consortium. 1999. Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401: 921–923.
2. Henry, J., M. T. Ribouchon, C. Offer, and P. Pontarotti. 1997. B30.2-like domain proteins: a growing family. *Biochem. Biophys. Res. Commun.* 235: 162–165.
3. Rhodes, D. A., B. de Bono, and J. Trowsdale. 2005. Relationship between SPRY and B30.2 protein domains. Evolution of a component of immune defence? *Immunology* 116: 411–417.
4. Sawyer, S. L., L. I. Wu, M. Emerman, and H. S. Malik. 2005. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. USA* 102: 2832–2837.
5. Yap, M. W., S. Nisole, and J. P. Stoye. 2005. A single amino acid change in the SPRY domain of human Trim5alpha leads to HIV-1 restriction. *Curr. Biol.* 15: 73–78.
6. Nisole, S., J. P. Stoye, and A. Saïb. 2005. TRIM family proteins: retroviral restriction and antiviral defence. *Nat. Rev. Microbiol.* 3: 799–808.
7. Afrache, H., P. Gouret, S. Ainouche, P. Pontarotti, and D. Olive. 2012. The butyrophilin (BTN) gene family: from milk fat to the regulation of the immune response. *Immunogenetics* 64: 781–794.
8. Trowsdale, J., and J. C. Knight. 2013. Major histocompatibility complex genomics and human disease. *Annu. Rev. Genomics Hum. Genet.* 14: 301–323.
9. Kaufman, J., S. Milne, T. W. Göbel, B. A. Walker, J. P. Jacob, C. Auffray, R. Zoorob, and S. Beck. 1999. The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401: 923–925.
10. Bingulac-Popovic, J., F. Figueroa, A. Sato, W. S. Talbot, S. L. Johnson, M. Gates, J. H. Postlethwait, and J. Klein. 1997. Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. *Immunogenetics* 46: 129–134.
11. Hansen, J. D., P. Strassburger, G. H. Thorgaard, W. P. Young, and L. Du Pasquier. 1999. Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, Oncorhynchus mykiss. *J. Immunol.* 163: 774–786.
12. Du Pasquier, L. 2004. Innate immunity in early chordates and the appearance of adaptive immunity. *C. R. Biol.* 327: 591–601.
13. Levasseur, A., and P. Pontarotti. 2010. Was the ancestral MHC involved in innate immunity? *Eur. J. Immunol.* 40: 2682–2685.
14. Azumi, K., R. De Santis, A. De Tomaso, I. Rigoutsos, F. Yoshizaki, M. R. Pinto, R. Marino, K. Shida, M. Ikeda, M. Ikeda, et al. 2003. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". *Immunogenetics* 55: 570–581.
15. Du Pasquier, L., I. Zucchetti, and R. De Santis. 2004. Immunoglobulin superfamily receptors in protochordates: before RAG time. *Immunol. Rev.* 198: 233–248.
16. Zucchetti, I., R. De Santis, S. Grusea, P. Pontarotti, and L. Du Pasquier. 2009. Origin and evolution of the vertebrate leukocyte receptors: the lesson from tunicates. *Immunogenetics* 61: 463–481.
17. Danchin, E. G., L. Abi-Rached, A. Gilles, and P. Pontarotti. 2003. Conservation of the MHC-like region throughout evolution. *Immunogenetics* 55: 141–148.
18. Olinski, R. P., L. G. Lundin, and F. Hallböök. 2006. Conserved synteny between the Ciona genome and human paralogons identifies large duplication events in the molecular evolution of the insulin-relaxin gene family. *Mol. Biol. Evol.* 23: 10–22.
19. Xu, Z., and B. Jin. 2010. A novel interface consisting of homologous immunoglobulin superfamily members with multiple functions. *Cell. Mol. Immunol.* 7: 11–19.
20. Ohno, S. 1970. *Evolution by Gene Duplication.* Allen & Unwin, London.
21. Flajnik, M. F., T. Tlapakova, M. F. Criscitiello, V. Krylov, and Y. Ohta. 2012. Evolution of the B7 family: co-evolution of B7H6 and NKp30, identification of a new B7 family member, B7H7, and of B7's historical relationship with the MHC. *Immunogenetics* 64: 571–590.
22. Lundin, L. G. 1993. Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* 16: 1–19.
23. Kasahara, M., M. Hayashi, K. Tanaka, H. Inoko, K. Sugaya, T. Ikemura, and T. Ishibashi. 1996. Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA* 93: 9096–9101.
24. Katsanis, N., J. Fitzgibbon, and E. M. Fisher. 1996. Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. *Genomics* 35: 101–108.
25. Flajnik, M. F., and M. Kasahara. 2001. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* 15: 351–362.
26. Larhammar, D., L. G. Lundin, and F. Hallböök. 2002. The human Hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. *Genome Res.* 12: 1910–1920.
27. Holland, P. W. 2003. More genes in vertebrates? *J. Struct. Funct. Genomics* 3: 75–84.
28. Danchin, E. G., and P. Pontarotti. 2004. Statistical evidence for a more than 800-million-year-old evolutionarily conserved genomic region in our genome. *J. Mol. Evol.* 59: 587–597.
29. Danchin, E. G., and P. Pontarotti. 2004. Towards the reconstruction of the bilaterian ancestral pre-MHC region. *Trends Genet.* 20: 587–591.
30. Storz, J. F., J. C. Opazo, and F. G. Hoffmann. 2013. Gene duplication, genome duplication, and the functional diversification of vertebrate globins. *Mol. Phylogenet. Evol.* 66: 469–478.
31. Du Pasquier, L. 2004. Speculations on the origin of the vertebrate immune system. *Immunol. Lett.* 92: 3–9.
32. Gruen, J. R., and S. M. Weissman. 1997. Evolving views of the major histocompatibility complex. *Blood* 90: 4252–4265.
33. Abi-Rached, L., A. Gilles, T. Shiina, P. Pontarotti, and H. Inoko. 2002. Evidence of en bloc duplication in vertebrate genomes. *Nat. Genet.* 31: 100–105.
34. Castro, L. F., R. F. Furlong, and P. W. Holland. 2004. An antecedent of the MHC-linked genomic region in amphioxus. *Immunogenetics* 55: 782–784.
35. Srivastava, M., E. Begovic, J. Chapman, N. H. Putnam, U. Hellsten, T. Kawashima, A. Kuo, T. Mitros, A. Salamov, M. L. Carpenter, et al. 2008. The Trichoplax genome and the nature of placozoans. *Nature* 454: 955–960.
36. Collins, A. G. 1998. Evaluating multiple alternative hypotheses for the origin of Bilateria: an analysis of 18S rRNA molecular evidence. *Proc. Natl. Acad. Sci. USA* 95: 15458–15463.
37. GIGA Community of Scientists. 2014. The Global Invertebrate Genomics Alliance (GIGA): Developing community resources to study diverse invertebrate genomes. *J. Hered.* 105: 1–18.
38. Dellaporta, S. L., A. Xu, S. Sagasser, W. Jakob, M. A. Moreno, L. W. Buss, and B. Schierwater. 2006. Mitochondrial genome of Trichoplax adhaerens supports placozoa as the basal lower metazoan phylum. *Proc. Natl. Acad. Sci. USA* 103: 8751–8756.
39. Osigus, H. J., M. Eitel, M. Bernt, A. Donath, and B. Schierwater. 2013. Mitogenomics at the base of Metazoa. *Mol. Phylogenet. Evol.* 69: 339–351.
40. Philippe, H., R. Derelle, P. Lopez, K. Pick, C. Borchiellini, N. Boury-Esnault, J. Vacelet, E. Renard, E. Houliston, E. Quéinnec, et al. 2009. Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* 19: 706–712.
41. Schulze, F. E. 1883. Trichoplax adhaerens, nov. gen., nov. spec. *Zool. Anz.* 6: 92–97.
42. Monticelli, F. S. 1893. Treptoplax reptans n.g., n.sp. *Atti dell' Accademia dei Lincei. Rendiconti* II: 39–40.
43. Voigt, O., A. G. Collins, V. B. Pearse, J. S. Pearse, A. Ender, H. Hadrys, and B. Schierwater. 2004. Placozoa — no longer a phylum of one. *Curr. Biol.* 14: R944–R945.
44. Eitel, M., H. J. Osigus, R. DeSalle, and B. Schierwater. 2013. Global diversity of the Placozoa. *PLoS ONE* 8: e57131.
45. Schierwater, B. 2005. My favorite animal, Trichoplax adhaerens. *BioEssays* 27: 1294–1302.
46. Ringrose, J. H., H. W. van den Toorn, M. Eitel, H. Post, P. Neerincx, B. Schierwater, A. F. Altelaar, and A. J. Heck. 2013. Deep proteome profiling of Trichoplax adhaerens reveals remarkable features at the origin of metazoan multicellularity. *Nat. Commun.* 4: 1408.
47. Jakob, W., S. Sagasser, S. Dellaporta, P. Holland, K. Kuhn, and B. Schierwater. 2004. The Trox-2 Hox/ParaHox gene of Trichoplax (Placozoa) marks an epithelial boundary. *Dev. Genes Evol.* 214: 170–175.
48. Putnam, N. H., T. Butts, D. E. Ferrier, R. F. Furlong, U. Hellsten, T. Kawashima, M. Robinson-Rechavi, E. Shoguchi, A. Terry, J. K. Yu, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453: 1064–1071.
49. Flajnik, M. F., and M. Kasahara. 2010. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.* 11: 47–59.
50. Suurväli, J., J. Robert, P. Boudinot, and S. Rüütel Boudinot. 2013. R4 regulators of G protein signaling (RGS) identify an ancient MHC-linked synteny group. *Immunogenetics* 65: 145–156.
51. Boratyn, G. M., A. A. Schäffer, R. Agarwala, S. F. Altschul, D. J. Lipman, and T. L. Madden. 2012. Domain enhanced lookup time accelerated BLAST. *Biol. Direct* 7: 12.
52. Grusea, S., E. Pardoux, O. Chabrol, and P. Pontarotti. 2011. Compound Poisson approximation and testing for gene clusters with multigene families. *J. Comput. Biol.* 18: 579–594.

53. Maere, S., K. Heymans, and M. Kuiper. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448–3449.
54. Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57: 289–300.
55. Benjamini, Y., and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* 29: 1165–1188.
56. Ravasi, T., H. Suzuki, C. V. Cannistraci, S. Katayama, V. B. Bajic, K. Tan, A. Akalin, S. Schmeier, M. Kanamori-Katayama, N. Bertin, et al. 2010. An atlas of combinatorial transcriptional regulation in mouse and man. [Published erratum appears in 2010 *Cell* 141: 369.] *Cell* 140: 744–752.
57. Kasahara, M. 1999. The chromosomal duplication model of the major histocompatibility complex. *Immunol. Rev.* 167: 17–32.
58. Danchin, E., V. Vitiello, A. Vienne, O. Richard, P. Gouret, M. F. McDermott, and P. Pontarotti. 2004. The major histocompatibility complex origin. *Immunol. Rev.* 198: 216–232.
59. Vienne, A., T. Shiina, L. Abi-Rached, E. Danchin, V. Vitiello, F. Cartault, H. Inoko, and P. Pontarotti. 2003. Evolution of the proto-MHC ancestral region: more evidence for the plesiomorphic organisation of human chromosome 9q34 region. *Immunogenetics* 55: 429–436.
60. de Laat, W., and D. Duboule. 2013. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502: 499–506.
61. Kikuta, H., M. Laplante, P. Navratilova, A. Z. Komisarczuk, P. G. Engström, D. Fredman, A. Akalin, M. Caccamo, I. Sealy, K. Howe, et al. 2007. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.* 17: 545–555.
62. Irimia, M., J. J. Tena, M. S. Alexis, A. Fernandez-Miñan, I. Maeso, O. Bogdanovic, E. de la Calle-Mustienes, S. W. Roy, J. L. Gómez-Skarmeta, and H. B. Fraser. 2012. Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res.* 22: 2356–2367.
63. Du Pasquier, L. 2009. Fish 'n' TRIMs. *J. Biol.* 8: 50.
64. King, N., M. J. Westbrook, S. L. Young, A. Kuo, M. Abedin, J. Chapman, S. Fairclough, U. Hellsten, Y. Isogai, I. Letunic, et al. 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* 451: 783–788.
65. Lang, B. F., C. O'Kelly, T. Nerad, M. W. Gray, and G. Burger. 2002. The closest unicellular relatives of animals. *Curr. Biol.* 12: 1773–1778.
66. Burger, G., L. Forget, Y. Zhu, M. W. Gray, and B. F. Lang. 2003. Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc. Natl. Acad. Sci. USA* 100: 892–897.
67. King, N., C. T. Hittinger, and S. B. Carroll. 2003. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* 301: 361–363.
68. Frédéric, M. Y., V. F. Lundin, M. D. Whiteside, J. G. Cueva, D. K. Tu, S. Y. Kang, H. Singh, D. L. Baillie, H. Hutter, M. B. Goodman, et al. 2013. Identification of 526 conserved metazoan genetic innovations exposes a new role for cofactor E-like in neuronal microtubule homeostasis. *PLoS Genet.* 9: e1003804.
69. Yewdell, J., C. Lapham, I. Bacik, T. Spies, and J. Bennink. 1994. MHC-encoded proteasome subunits LMP2 and LMP7 are not required for efficient antigen presentation. *J. Immunol.* 152: 1163–1170.
70. Van den Eynde, B. J., and S. Morel. 2001. Differential processing of class-I-restricted epitopes by the standard proteasome and the immunoproteasome. *Curr. Opin. Immunol.* 13: 147–153.
71. Ermolaeva, M. A., A. Segref, A. Dakhovnik, H. L. Ou, J. I. Schneider, O. Utermöhlen, T. Hoppe, and B. Schumacher. 2013. DNA damage in germ cells induces an innate immune response that triggers systemic stress resistance. *Nature* 501: 416–420.
72. Short, K. M., and T. C. Cox. 2006. Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. *J. Biol. Chem.* 281: 8970–8980.
73. Versteeg, G. A., R. Rajsbaum, M. T. Sánchez-Aparicio, A. M. Maestre, J. Valdiviezo, M. Shi, K. S. Inn, A. Fernandez-Sesma, J. Jung, and A. García-Sastre. 2013. The E3-ligase TRIM family of proteins regulates signaling pathways triggered by innate immune pattern-recognition receptors. *Immunity* 38: 384–398.
74. Gat-Viks, I., N. Chevrier, R. Wilentzik, T. Eisenhaure, R. Raychowdhury, Y. Steuerman, A. K. Shalek, N. Hacohen, I. Amit, and A. Regev. 2013. Deciphering molecular circuits from genetic variation underlying transcriptional responsiveness to stimuli. *Nat. Biotechnol.* 31: 342–349.
75. Garstang, M., and D. E. Ferrier. 2013. Time is of the essence for ParaHox homeobox gene clustering. *BMC Biol.* 11: 72.

DISSERTATIONS DEFENDED AT TALLINN UNIVERSITY OF TECHNOLOGY ON *NATURAL AND EXACT SCIENCES*

1. **Olav Kongas**. Nonlinear Dynamics in Modeling Cardiac Arrhytmias. 1998.
2. **Kalju Vanatalu**. Optimization of Processes of Microbial Biosynthesis of Isotopically Labeled Biomolecules and Their Complexes. 1999.
3. **Ahto Buldas**. An Algebraic Approach to the Structure of Graphs. 1999.
4. **Monika Drews**. A Metabolic Study of Insect Cells in Batch and Continuous Culture: Application of Chemostat and Turbidostat to the Production of Recombinant Proteins. 1999.
5. **Eola Valdre**. Endothelial-Specific Regulation of Vessel Formation: Role of Receptor Tyrosine Kinases. 2000.
6. **Kalju Lott**. Doping and Defect Thermodynamic Equilibrium in ZnS. 2000.
7. **Reet Koljak**. Novel Fatty Acid Dioxygenases from the Corals *Plexaura homomalla* and *Gersemia fruticosa.* 2001.
8. **Anne Paju**. Asymmetric oxidation of Prochiral and Racemic Ketones by Using Sharpless Catalyst. 2001.
9. **Marko Vendelin**. Cardiac Mechanoenergetics *in silico*. 2001.
10. **Pearu Peterson**. Multi-Soliton Interactions and the Inverse Problem of Wave Crest. 2001.
11. **Anne Menert**. Microcalorimetry of Anaerobic Digestion. 2001.
12. **Toomas Tiivel**. The Role of the Mitochondrial Outer Membrane in *in vivo* Regulation of Respiration in Normal Heart and Skeletal Muscle Cell. 2002.
13. **Olle Hints**. Ordovician Scolecodonts of Estonia and Neighbouring Areas: Taxonomy, Distribution, Palaeoecology, and Application. 2002.
14. **Jaak Nõlvak**. Chitinozoan Biostratigrapy in the Ordovician of Baltoscandia. 2002.
15. **Liivi Kluge**. On Algebraic Structure of Pre-Operad. 2002.
16. **Jaanus Lass**. Biosignal Interpretation: Study of Cardiac Arrhytmias and Electromagnetic Field Effects on Human Nervous System. 2002.
17. **Janek Peterson**. Synthesis, Structural Characterization and Modification of PAMAM Dendrimers. 2002.
18. **Merike Vaher**. Room Temperature Ionic Liquids as Background Electrolyte Additives in Capillary Electrophoresis. 2002.
19. **Valdek Mikli**. Electron Microscopy and Image Analysis Study of Powdered Hardmetal Materials and Optoelectronic Thin Films. 2003.
20. **Mart Viljus**. The Microstructure and Properties of Fine-Grained Cermets. 2003.
21. **Signe Kask**. Identification and Characterization of Dairy-Related *Lactobacillus.* 2003
22. **Tiiu-Mai Laht**. Influence of Microstructure of the Curd on Enzymatic and Microbiological Processes in Swiss-Type Cheese. 2003.
23. **Anne Kuusksalu**. 2–5A Synthetase in the Marine Sponge *Geodia cydonium*. 2003.
24. **Sergei Bereznev**. Solar Cells Based on Polycristalline Copper-Indium Chalcogenides and Conductive Polymers. 2003.

25. **Kadri Kriis**. Asymmetric Synthesis of C_2-Symmetric Bimorpholines and Their Application as Chiral Ligands in the Transfer Hydrogenation of Aromatic Ketones. 2004.
26. **Jekaterina Reut**. Polypyrrole Coatings on Conducting and Insulating Substracts. 2004.
27. **Sven Nõmm**. Realization and Identification of Discrete-Time Nonlinear Systems. 2004.
28. **Olga Kijatkina**. Deposition of Copper Indium Disulphide Films by Chemical Spray Pyrolysis. 2004.
29. **Gert Tamberg**. On Sampling Operators Defined by Rogosinski, Hann and Blackman Windows. 2004.
30. **Monika Übner**. Interaction of Humic Substances with Metal Cations. 2004.
31. **Kaarel Adamberg**. Growth Characteristics of Non-Starter Lactic Acid Bacteria from Cheese. 2004.
32. **Imre Vallikivi**. Lipase-Catalysed Reactions of Prostaglandins. 2004.
33. **Merike Peld**. Substituted Apatites as Sorbents for Heavy Metals. 2005.
34. **Vitali Syritski**. Study of Synthesis and Redox Switching of Polypyrrole and Poly(3,4-ethylenedioxythiophene) by Using *in-situ* Techniques. 2004.
35. **Lee Põllumaa**. Evaluation of Ecotoxicological Effects Related to Oil Shale Industry. 2004.
36. **Riina Aav**. Synthesis of 9,11-Secosterols Intermediates. 2005.
37. **Andres Braunbrück**. Wave Interaction in Weakly Inhomogeneous Materials. 2005.
38. **Robert Kitt**. Generalised Scale-Invariance in Financial Time Series. 2005.
39. **Juss Pavelson**. Mesoscale Physical Processes and the Related Impact on the Summer Nutrient Fields and Phytoplankton Blooms in the Western Gulf of Finland. 2005.
40. **Olari Ilison**. Solitons and Solitary Waves in Media with Higher Order Dispersive and Nonlinear Effects. 2005.
41. **Maksim Säkki**. Intermittency and Long-Range Structurization of Heart Rate. 2005.
42. **Enli Kiipli**. Modelling Seawater Chemistry of the East Baltic Basin in the Late Ordovician–Early Silurian. 2005.
43. **Igor Golovtsov**. Modification of Conductive Properties and Processability of Polyparaphenylene, Polypyrrole and polyaniline. 2005.
44. **Katrin Laos**. Interaction Between Furcellaran and the Globular Proteins (Bovine Serum Albumin β-Lactoglobulin). 2005.
45. **Arvo Mere**. Structural and Electrical Properties of Spray Deposited Copper Indium Disulphide Films for Solar Cells. 2006.
46. **Sille Ehala**. Development and Application of Various On- and Off-Line Analytical Methods for the Analysis of Bioactive Compounds. 2006.
47. **Maria Kulp**. Capillary Electrophoretic Monitoring of Biochemical Reaction Kinetics. 2006.
48. **Anu Aaspõllu.** Proteinases from *Vipera lebetina* Snake Venom Affecting Hemostasis. 2006.

49. **Lyudmila Chekulayeva**. Photosensitized Inactivation of Tumor Cells by Porphyrins and Chlorins. 2006.
50. **Merle Uudsemaa**. Quantum-Chemical Modeling of Solvated First Row Transition Metal Ions. 2006.
51. **Tagli Pitsi**. Nutrition Situation of Pre-School Children in Estonia from 1995 to 2004. 2006.
52. **Angela Ivask**. Luminescent Recombinant Sensor Bacteria for the Analysis of Bioavailable Heavy Metals. 2006.
53. **Tiina Lõugas**. Study on Physico-Chemical Properties and Some Bioactive Compounds of Sea Buckthorn (*Hippophae rhamnoides* L.). 2006.
54. **Kaja Kasemets**. Effect of Changing Environmental Conditions on the Fermentative Growth of *Saccharomyces cerevisae* S288C: Auxo-accelerostat Study. 2006.
55. **Ildar Nisamedtinov**. Application of ^{13}C and Fluorescence Labeling in Metabolic Studies of *Saccharomyces* spp. 2006.
56. **Alar Leibak**. On Additive Generalisation of Voronoï's Theory of Perfect Forms over Algebraic Number Fields. 2006.
57. **Andri Jagomägi**. Photoluminescence of Chalcopyrite Tellurides. 2006.
58. **Tõnu Martma**. Application of Carbon Isotopes to the Study of the Ordovician and Silurian of the Baltic. 2006.
59. **Marit Kauk**. Chemical Composition of $CuInSe_2$ Monograin Powders for Solar Cell Application. 2006.
60. **Julia Kois**. Electrochemical Deposition of $CuInSe_2$ Thin Films for Photovoltaic Applications. 2006.
61. **Ilona Oja Açik**. Sol-Gel Deposition of Titanium Dioxide Films. 2007.
62. **Tiia Anmann**. Integrated and Organized Cellular Bioenergetic Systems in Heart and Brain. 2007.
63. **Katrin Trummal**. Purification, Characterization and Specificity Studies of Metalloproteinases from *Vipera lebetina* Snake Venom. 2007.
64. **Gennadi Lessin**. Biochemical Definition of Coastal Zone Using Numerical Modeling and Measurement Data. 2007.
65. **Enno Pais**. Inverse problems to determine non-homogeneous degenerate memory kernels in heat flow. 2007.
66. **Maria Borissova**. Capillary Electrophoresis on Alkylimidazolium Salts. 2007.
67. **Karin Valmsen**. Prostaglandin Synthesis in the Coral *Plexaura homomalla*: Control of Prostaglandin Stereochemistry at Carbon 15 by Cyclooxygenases. 2007.
68. **Kristjan Piirimäe**. Long-Term Changes of Nutrient Fluxes in the Drainage Basin of the Gulf of Finland – Application of the PolFlow Model. 2007.
69. **Tatjana Dedova**. Chemical Spray Pyrolysis Deposition of Zinc Sulfide Thin Films and Zinc Oxide Nanostructured Layers. 2007.
70. **Katrin Tomson**. Production of Labelled Recombinant Proteins in Fed-Batch Systems in *Escherichia coli*. 2007.
71. **Cecilia Sarmiento**. Suppressors of RNA Silencing in Plants. 2008.
72. **Vilja Mardla**. Inhibition of Platelet Aggregation with Combination of Antiplatelet Agents. 2008.

73. **Maie Bachmann**. Effect of Modulated Microwave Radiation on Human Resting Electroencephalographic Signal. 2008.
74. **Dan Hüvonen**. Terahertz Spectroscopy of Low-Dimensional Spin Systems. 2008.
75. **Ly Villo**. Stereoselective Chemoenzymatic Synthesis of Deoxy Sugar Esters Involving *Candida antarctica* Lipase B. 2008.
76. **Johan Anton**. Technology of Integrated Photoelasticity for Residual Stress Measurement in Glass Articles of Axisymmetric Shape. 2008.
77. **Olga Volobujeva**. SEM Study of Selenization of Different Thin Metallic Films. 2008.
78. **Artur Jõgi**. Synthesis of 4'-Substituted 2,3'-dideoxynucleoside Analogues. 2008.
79. **Mario Kadastik**. Doubly Charged Higgs Boson Decays and Implications on Neutrino Physics. 2008.
80. **Fernando Pérez-Caballero**. Carbon Aerogels from 5-Methylresorcinol-Formaldehyde Gels. 2008.
81. **Sirje Vaask**. The Comparability, Reproducibility and Validity of Estonian Food Consumption Surveys. 2008.
82. **Anna Menaker**. Electrosynthesized Conducting Polymers, Polypyrrole and Poly(3,4-ethylenedioxythiophene), for Molecular Imprinting. 2009.
83. **Lauri Ilison**. Solitons and Solitary Waves in Hierarchical Korteweg-de Vries Type Systems. 2009.
84. **Kaia Ernits**. Study of In_2S_3 and ZnS Thin Films Deposited by Ultrasonic Spray Pyrolysis and Chemical Deposition. 2009.
85. **Veljo Sinivee**. Portable Spectrometer for Ionizing Radiation "Gammamapper". 2009.
86. **Jüri Virkepu**. On Lagrange Formalism for Lie Theory and Operadic Harmonic Oscillator in Low Dimensions. 2009.
87. **Marko Piirsoo**. Deciphering Molecular Basis of Schwann Cell Development. 2009.
88. **Kati Helmja**. Determination of Phenolic Compounds and Their Antioxidative Capability in Plant Extracts. 2010.
89. **Merike Sõmera**. Sobemoviruses: Genomic Organization, Potential for Recombination and Necessity of P1 in Systemic Infection. 2010.
90. **Kristjan Laes**. Preparation and Impedance Spectroscopy of Hybrid Structures Based on $CuIn_3Se_5$ Photoabsorber. 2010.
91. **Kristin Lippur**. Asymmetric Synthesis of 2,2'-Bimorpholine and its 5,5'-Substituted Derivatives. 2010.
92. **Merike Luman**. Dialysis Dose and Nutrition Assessment by an Optical Method. 2010.
93. **Mihhail Berezovski**. Numerical Simulation of Wave Propagation in Heterogeneous and Microstructured Materials. 2010.
94. **Tamara Aid-Pavlidis**. Structure and Regulation of BDNF Gene. 2010.
95. **Olga Bragina**. The Role of Sonic Hedgehog Pathway in Neuro- and Tumorigenesis. 2010.

96. **Merle Randrüüt**. Wave Propagation in Microstructured Solids: Solitary and Periodic Waves. 2010.
97. **Marju Laars**. Asymmetric Organocatalytic Michael and Aldol Reactions Mediated by Cyclic Amines. 2010.
98. **Maarja Grossberg**. Optical Properties of Multinary Semiconductor Compounds for Photovoltaic Applications. 2010.
99. **Alla Maloverjan**. Vertebrate Homologues of Drosophila Fused Kinase and Their Role in Sonic Hedgehog Signalling Pathway. 2010.
100. **Priit Pruunsild**. Neuronal Activity-Dependent Transcription Factors and Regulation of Human *BDNF* Gene. 2010.
101. **Tatjana Knjazeva**. New Approaches in Capillary Electrophoresis for Separation and Study of Proteins. 2011.
102. **Atanas Katerski**. Chemical Composition of Sprayed Copper Indium Disulfide Films for Nanostructured Solar Cells. 2011.
103. **Kristi Timmo.** Formation of Properties of $CuInSe_2$ and $Cu_2ZnSn(S,Se)_4$ Monograin Powders Synthesized in Molten KI. 2011.
104. **Kert Tamm**. Wave Propagation and Interaction in Mindlin-Type Microstructured Solids: Numerical Simulation. 2011.
105. **Adrian Popp**. Ordovician Proetid Trilobites in Baltoscandia and Germany. 2011.
106. **Ove Pärn**. Sea Ice Deformation Events in the Gulf of Finland and This Impact on Shipping. 2011.
107. **Germo Väli**. Numerical Experiments on Matter Transport in the Baltic Sea. 2011.
108. **Andrus Seiman**. Point-of-Care Analyser Based on Capillary Electrophoresis. 2011.
109. **Olga Katargina**. Tick-Borne Pathogens Circulating in Estonia (Tick-Borne Encephalitis Virus, *Anaplasma phagocytophilum*, *Babesia* Species): Their Prevalence and Genetic Characterization. 2011.
110. **Ingrid Sumeri**. The Study of Probiotic Bacteria in Human Gastrointestinal Tract Simulator. 2011.
111. **Kairit Zovo**. Functional Characterization of Cellular Copper Proteome. 2011.
112. **Natalja Makarytsheva**. Analysis of Organic Species in Sediments and Soil by High Performance Separation Methods. 2011.
113. **Monika Mortimer**. Evaluation of the Biological Effects of Engineered Nanoparticles on Unicellular Pro- and Eukaryotic Organisms. 2011.
114. **Kersti Tepp**. Molecular System Bioenergetics of Cardiac Cells: Quantitative Analysis of Structure-Function Relationship. 2011.
115. **Anna-Liisa Peikolainen**. Organic Aerogels Based on 5-Methylresorcinol. 2011.
116. **Leeli Amon**. Palaeoecological Reconstruction of Late-Glacial Vegetation Dynamics in Eastern Baltic Area: A View Based on Plant Macrofossil Analysis. 2011.
117. **Tanel Peets**. Dispersion Analysis of Wave Motion in Microstructured Solids. 2011.

118. **Liina Kaupmees**. Selenization of Molybdenum as Contact Material in Solar Cells. 2011.
119. **Allan Olspert**. Properties of VPg and Coat Protein of Sobemoviruses. 2011.
120. **Kadri Koppel**. Food Category Appraisal Using Sensory Methods. 2011.
121. **Jelena Gorbatšova**. Development of Methods for CE Analysis of Plant Phenolics and Vitamins. 2011.
122. **Karin Viipsi**. Impact of EDTA and Humic Substances on the Removal of Cd and Zn from Aqueous Solutions by Apatite. 2012.
123. **David Schryer**. Metabolic Flux Analysis of Compartmentalized Systems Using Dynamic Isotopologue Modeling. 2012.
124. **Ardo Illaste**. Analysis of Molecular Movements in Cardiac Myocytes. 2012.
125. **Indrek Reile**. 3-Alkylcyclopentane-1,2-Diones in Asymmetric Oxidation and Alkylation Reactions. 2012.
126. **Tatjana Tamberg**. Some Classes of Finite 2-Groups and Their Endomorphism Semigroups. 2012.
127. **Taavi Liblik**. Variability of Thermohaline Structure in the Gulf of Finland in Summer. 2012.
128. **Priidik Lagemaa**. Operational Forecasting in Estonian Marine Waters. 2012.
129. **Andrei Errapart**. Photoelastic Tomography in Linear and Non-linear Approximation. 2012.
130. **Külliki Krabbi**. Biochemical Diagnosis of Classical Galactosemia and Mucopolysaccharidoses in Estonia. 2012.
131. **Kristel Kaseleht**. Identification of Aroma Compounds in Food using SPME-GC/MS and GC-Olfactometry. 2012.
132. **Kristel Kodar**. Immunoglobulin G Glycosylation Profiling in Patients with Gastric Cancer. 2012.
133. **Kai Rosin**. Solar Radiation and Wind as Agents of the Formation of the Radiation Regime in Water Bodies. 2012.
134. **Ann Tiiman**. Interactions of Alzheimer's Amyloid-Beta Peptides with Zn(II) and Cu(II) Ions. 2012.
135. **Olga Gavrilova**. Application and Elaboration of Accounting Approaches for Sustainable Development. 2012.
136. **Olesja Bondarenko**. Development of Bacterial Biosensors and Human Stem Cell-Based *In Vitro* Assays for the Toxicological Profiling of Synthetic Nanoparticles. 2012.
137. **Katri Muska**. Study of Composition and Thermal Treatments of Quaternary Compounds for Monograin Layer Solar Cells. 2012.
138. **Ranno Nahku**. Validation of Critical Factors for the Quantitative Characterization of Bacterial Physiology in Accelerostat Cultures. 2012.
139. **Petri-Jaan Lahtvee**. Quantitative Omics-level Analysis of Growth Rate Dependent Energy Metabolism in *Lactococcus lactis*. 2012.
140. **Kerti Orumets**. Molecular Mechanisms Controlling Intracellular Glutathione Levels in Baker's Yeast *Saccharomyces cerevisiae* and its Random Mutagenized Glutathione Over-Accumulating Isolate. 2012.
141. **Loreida Timberg**. Spice-Cured Sprats Ripening, Sensory Parameters Development, and Quality Indicators. 2012.

142. **Anna Mihhalevski**. Rye Sourdough Fermentation and Bread Stability. 2012.
143. **Liisa Arike**. Quantitative Proteomics of *Escherichia coli*: From Relative to Absolute Scale. 2012.
144. **Kairi Otto**. Deposition of In_2S_3 Thin Films by Chemical Spray Pyrolysis. 2012.
145. **Mari Sepp**. Functions of the Basic Helix-Loop-Helix Transcription Factor TCF4 in Health and Disease. 2012.
146. **Anna Suhhova**. Detection of the Effect of Weak Stressors on Human Resting Electroencephalographic Signal. 2012.
147. **Aram Kazarjan**. Development and Production of Extruded Food and Feed Products Containing Probiotic Microorganisms. 2012.
148. **Rivo Uiboupin**. Application of Remote Sensing Methods for the Investigation of Spatio-Temporal Variability of Sea Surface Temperature and Chlorophyll Fields in the Gulf of Finland. 2013.
149. **Tiina Kriščiunaite**. A Study of Milk Coagulability. 2013.
150. **Tuuli Levandi**. Comparative Study of Cereal Varieties by Analytical Separation Methods and Chemometrics. 2013.
151. **Natalja Kabanova**. Development of a Microcalorimetric Method for the Study of Fermentation Processes. 2013.
152. **Himani Khanduri**. Magnetic Properties of Functional Oxides. 2013.
153. **Julia Smirnova**. Investigation of Properties and Reaction Mechanisms of Redox-Active Proteins by ESI MS. 2013.
154. **Mervi Sepp**. Estimation of Diffusion Restrictions in Cardiomyocytes Using Kinetic Measurements. 2013.
155. **Kersti Jääger**. Differentiation and Heterogeneity of Mesenchymal Stem Cells. 2013.
156. **Victor Alari**. Multi-Scale Wind Wave Modeling in the Baltic Sea. 2013.
157. **Taavi Päll**. Studies of CD44 Hyaluronan Binding Domain as Novel Angiogenesis Inhibitor. 2013.
158. **Allan Niidu**. Synthesis of Cyclopentane and Tetrahydrofuran Derivatives. 2013.
159. **Julia Geller**. Detection and Genetic Characterization of *Borrelia* Species Circulating in Tick Population in Estonia. 2013.
160. **Irina Stulova**. The Effects of Milk Composition and Treatment on the Growth of Lactic Acid Bacteria. 2013.
161. **Jana Holmar**. Optical Method for Uric Acid Removal Assessment During Dialysis. 2013.
162. **Kerti Ausmees**. Synthesis of Heterobicyclo[3.2.0]heptane Derivatives *via* Multicomponent Cascade Reaction. 2013.
163. **Minna Varikmaa**. Structural and Functional Studies of Mitochondrial Respiration Regulation in Muscle Cells. 2013.
164. **Indrek Koppel**. Transcriptional Mechanisms of BDNF Gene Regulation. 2014.
165. **Kristjan Pilt**. Optical Pulse Wave Signal Analysis for Determination of Early Arterial Ageing in Diabetic Patients. 2014.
166. **Andres Anier**. Estimation of the Complexity of the Electroencephalogram for Brain Monitoring in Intensive Care. 2014.

167. **Toivo Kallaste**. Pyroclastic Sanidine in the Lower Palaeozoic Bentonites – A Tool for Regional Geological Correlations. 2014.
168. **Erki Kärber**. Properties of ZnO-nanorod/In_2S_3/$CuInS_2$ Solar Cell and the Constituent Layers Deposited by Chemical Spray Method. 2014.
169. **Julia Lehner**. Formation of Cu_2ZnSnS_4 and $Cu_2ZnSnSe_4$ by Chalcogenisation of Electrochemically Deposited Precursor Layers. 2014.
170. **Peep Pitk**. Protein- and Lipid-rich Solid Slaughterhouse Waste Anaerobic Co-digestion: Resource Analysis and Process Optimization. 2014.
171. **Kaspar Valgepea**. Absolute Quantitative Multi-omics Characterization of Specific Growth Rate-dependent Metabolism of *Escherichia coli*. 2014.
172. **Artur Noole**. Asymmetric Organocatalytic Synthesis of 3,3'-Disubstituted Oxindoles. 2014.
173. **Robert Tsanev**. Identification and Structure-Functional Characterisation of the Gene Transcriptional Repressor Domain of Human Gli Proteins. 2014.
174. **Dmitri Kartofelev**. Nonlinear Sound Generation Mechanisms in Musical Acoustic. 2014.
175. **Sigrid Hade**. GIS Applications in the Studies of the Palaeozoic Graptolite Argillite and Landscape Change. 2014.