Preisverleihung 2014

STIFTUNG PROFESSOR DR. MAX CLOËTTA

Heft Nr. 42

Prof. Dr. Marc Y. Donath

«Targeting Inflammation in the Treatment of Type 2 Diabetes: Time to Start»

Prof. Dr. Henrik Kaessmann

«The Evolution of Mammalian Gene Expression: Dynamics and Phenotypic Impact»

THE EVOLUTION OF MAMMALIAN GENE EXPRESSION: DYNAMICS AND PHENOTYPIC IMPACT

Henrik Kaessmann

Center for Integrative Genomics, University of Lausanne,
1015 Lausanne, Switzerland.
Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.

Correspondence: henrik.kaessmann@unil.ch

Summary

Evolutionary changes in gene expression, due to regulatory mutations, probably explain many or even most phenotypic differences between species. However, large-scale comparative gene expression analyses have only recently become possible, thanks to the advent of high-throughput RNA sequencing technologies, which allow accurate and sensitive assessments of transcript sequences and expression levels. Here I present an overview of the many new insights that the first generation of comparative transcriptome studies has provided into the functional evolution of mammalian genomes. Together with various large-scale genomic and epigenomic data, these studies have uncovered the dynamics of gene expression evolution across various biological dimensions, which include the different coding and non-coding gene types, mammalian lineages, organs, developmental stages, chromosomes and sexes. This work has also provided fascinating new evidence with respect to the regulatory foundation and phenotypic relevance of gene expression changes during mammalian evolution.

Introduction

A major goal in biology is to understand the molecular basis of phenotypic evolution, in particular that of humans and other mammals. Shared mammalian traits include lactation, hair, and relatively large brains with unique structures (e.g., the neocortex). In addition, individual species or lineages evolved distinct anatomical, physiological and behavioral characteristics, relating to differences in reproduction, life span, cognitive abilities and disease susceptibility. For example, humans have evolved particularly large and complex brains, even when compared to their closest evolutionary relatives, the chimpanzees, and also exhibit profound differences in disease susceptibility (e.g, chimpanzees do not develop AIDS upon HIV infection), life span (the maximum life span of a human is approximately twice that of a chimpanzee), and various, more obvious, anatomical differences relative to chimpanzees and other primates[1]. Examples of distinct (reproductive) features of the major mammalian lineages include the placenta of placental mammals, the elaborate lactation system of marsupials, and the intriguing combination of egg yolk nourishment and lactation of the young in monotreme mammals such as the platypus.

What are the genomic changes and molecular mechanisms underlying all of these phenotypic innovations? When and how did they occur? What is the nature of the associated selective pressures? Answers to these questions, which motivate our research, can be sought in the analysis of several types of genomic alterations. In fact, mutations that may underlie phenotypic innovation can be grouped into only two major classes. The first class consists of mutations that change the sequence and, consequently, the function of the final gene product (i.e., the protein or RNA). The second class encompasses mutations in regulatory sequences (e.g. in promoter regions) that affect transcription, post-transcriptional processing, translation, or transcript/protein degradation. But it should be noted that certain gene product sequence alterations that change the function of the protein (e.g., mutations in transcription factors[2]) may also affect gene regulation.

How do these mutations drive phenotypic evolution? Generally, both classes of mutations may contribute to the evolution of distinct tissue mor-

phologies if they affect developmental programs. The specific morphologies may, in turn, provide the basis for species- or lineage-specific physiology and behavior. However, in addition, mutations may also directly shape phenotypic properties of adult organs, without influencing developmental trajectories.

Thus, the question arises: which of the two classes of mutations contributed more to phenotypic evolution? This question has been debated ever since the late 1960ies[3-7]. The current hypothesis in the field is that regulatory mutations (i.e., mutations of the second class) should underlie most phenotypic innovations, because protein-coding mutations (i.e., mutations belonging to the first class) are expected to more often have deleterious consequences[6,7], given that they necessarily affect all tissues and developmental stages in which a gene is expressed (given that the genome sequence is the same across tissues), whereas regulatory changes can be spatially and temporally restricted. Also, coding sequence mutations may be too scarce to explain the sometimes profound phenotypic differences even between closely species[3] (e.g., humans and chimpanzees).

However, only with the rise of the genomic era in the past decade it has become possible to start to systematically explore the molecular basis of phenotypic evolution and thus to empirically address the hypothesis of a dominant role of regulatory mutations in phenotypic evolution. Direct comparisons of mammalian genome sequences, which are accumulating at an accelerating pace, have uncovered many amino acid substitutions[8] that were apparently driven by positive selection (i.e., natural selection that increases the rate of fixation of new favorable mutations in the population) and may thus be of phenotypic relevance. Sophisticated analyses have even revealed potentially adaptive promoter sequence alterations[9]. Importantly, the concomitant development of large-scale genomics technologies has generally not only facilitated genome-wide analyses of expression patterns and changes but also of underlying regulatory mechanisms. Thus, initial microarray-based expression comparisons between closely related mammals identified potential phenotypically relevant expression level changes and initial clues initial to the general evolutionary rules that dictate the evolution of tissue transcriptomes[10,11]. However, given that microarrays rely on hybridization of RNA/DNA to species-

specific probes, they do not lend themselves well for between-species comparisons. Consequently, the recent development of high-throughput sequencing technologies[12], which allow essentially unbiased qualitative and quantitative comparisons of gene expression and regulatory mechanisms between divergent species on a genome-wide scale, constituted a quantum leap in the evolutionary genomics field. Thus, gene expression evolution and its regulatory foundation can now be conveniently studied[12] using, in particular, approaches based on high-throughput RNA sequencing (RNA-seq) and chromatin immunoprecipitation coupled with sequencing (ChIP-seq).

In this review, I will discuss, in the context of earlier work, the intriguing observations and insights that these novel approaches already have provided into the patterns, dynamics and phenotypic implications of gene expression change in mammals and vertebrate outgroup species. I will round off my review by outlining potentially fruitful and/or important future research directions.

Different aspects of the transcriptome evolve at different rates

Protein-coding genes: evolution of overall expression levels. One major observation of evolutionary transcriptome comparisons is that overall expression levels of protein-coding genes evolve more slowly than most other transcribed parts of the genome studied so far. For example, protein-coding gene expression levels were shown to be overall similar in primates[11,13,14], likely due to negative (purifying) natural selection (i.e., selection that eliminates deleterious mutations in the population). Thus, expression differences between livers from human and chimpanzees are, on average, only 20 % higher than differences observed between livers from pairs of individuals from the same species[14]. A recent RNA-seq-based study for a range of organs from representative mammals and amniote outgroup species confirmed that gene expression level evolution is generally constrained. Thus, protein-coding gene expression divergence does not increase linearly with time, as would be expected under neutral evolution[15], but eventually saturates between divergent species such as human and chicken. Notably, even for these two species, expression levels are highly correlated (with correlation coefficients above 0.7 for so-

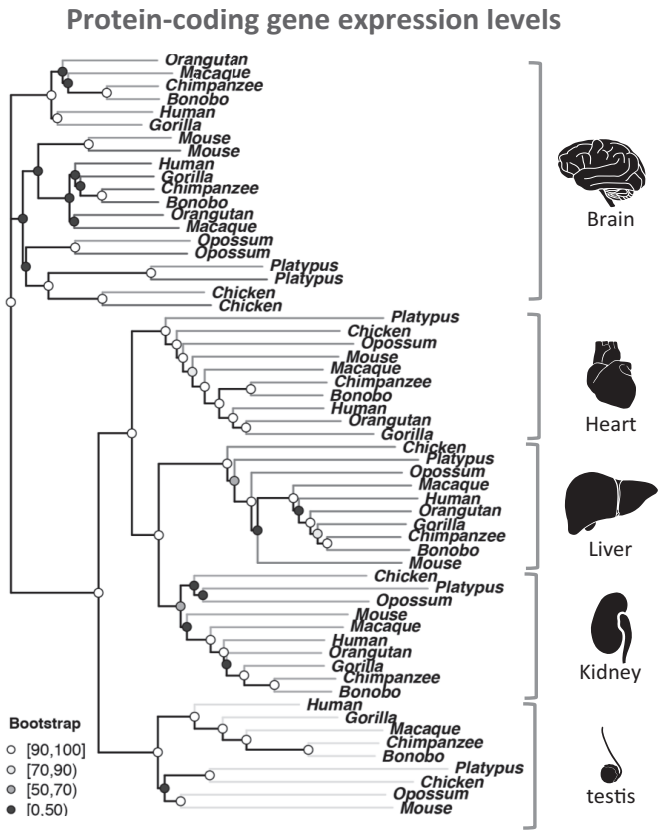

Figure 1. *The evolution of protein-coding gene expression levels in mammals. A mammalian gene expression phylogeny based on expression data[16] (constitutive exons) for six tissues from 9 representative mammals and one bird (chicken), used as an evolutionary outgroup. The Neighbour-joining tree displayed is based on pairwise expression level distance matrices (distances are based on 1 – rho, Spearman's correlation coefficient). The tree illustrates the organ-dominated clustering of gene expression level data from different species (see main text for details).*

matic organs). These observations suggest that conservation of core organ functions limits the divergence of protein-coding genes[16]. That is, similar expression levels for the same key subsets of genes in a given organ are required to sustain physiological organ functions across all mammals/amniotes. Thus, overall, different organs are more divergent than different amniotes (even the most distantly related) for the same organ in terms of protein-coding gene expression levels. Consequently, in phylogenies that are directly based on gene expression levels in the different organs and species, samples cluster together according to the organ from which they are derived[16] (Figure 1). Within each organ cluster, species generally group according to their phylogenetic relationships, reflecting the fact that regulatory changes accumulate over evolutionary time (until they eventually saturate – see above), such that more closely related species have more similar expression levels[16] (Figure 1). Altogether, these findings are consistent with the fact that mammalian/amniote organs originated in vertebrate ancestors long before the emergence and subsequent divergence of amniote lineages. Thus, a grouping of amniote species by

organs, which have common vertebrate origins and functions, in such expression analyses could, in principle, have been expected.

Protein-coding genes: evolution of alternative splicing. However, it turns out that divergence rates are much higher for other portions of the transcriptome than for protein-coding gene expression levels, leading to very different expression divergence patterns. The most rapidly evolving parts of the transcriptome investigated so far are those that result from alternative splicing events. Indeed, recent work revealed that alternative splicing evolves extremely rapidly, such that alternative splicing shows highly species-specific patterns[17,18]. The rapid evolution of alternative splicing is illustrated by a species-dominated pattern of clustering: exon skipping frequencies are more similar between different organs from the same species than between different species for a given organ[17,18] (Figure 2). Thus, for example, a human brain is more similar to a human liver than to a chimpanzee brain in terms of alternative splicing, in spite of the fact that these species are very closely related and brains and livers very likely have overall similar functions in both species (i.e., a human brain is functionally certainly more diverged from a human liver than from a chimpanzee brain). This pattern is thus exactly the inverse of that observed for protein-coding gene expression levels (Figure 1). Notably, the high species-specificity of alternative splicing was convincingly demonstrated to be driven by changes in *cis*-acting regulatory elements, based on the finding that a mouse model carrying a copy of human chromosome 21 shows human-specific exon skipping frequencies on the inserted chromosome (mouse-specific patterns would be expected if trans-acting regulatory factors were predominant[18]).

Noncoding genes. Assessments of the expression evolution of noncoding portions of the transcriptome have revealed both slowly and rapidly evolving categories of genes. MicroRNAs (miRNAs) seem to overall evolve slowly in terms of both DNA sequence[19] and expression[20]. Expression analyses thus reveal an organ-dominated clustering pattern that is reminiscent to that of protein-coding genes. Long noncoding RNAs (lncRNAs), however, show a rather rapid mode of expression evolution. LncRNA repertoires experience rapid turnover[20-22], and lncRNA sequences and expression levels tend to evolve rapidly[20,21] compared to those of protein-

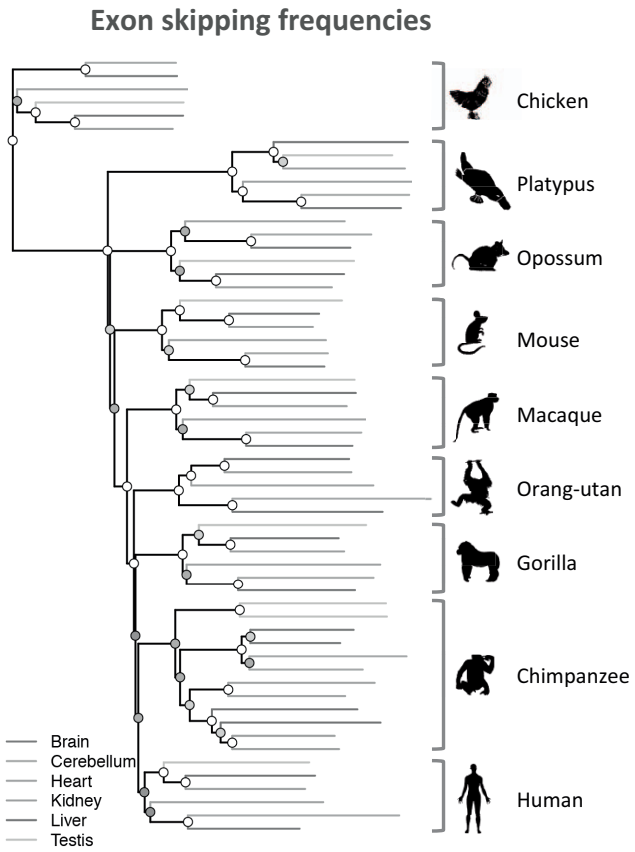

Figure 2. *The evolution of alternative splicing in mammals. A mammalian phylogeny based on alternative splicing (i.e., exon skipping) frequencies assessed from expression data[16] (constitutive exons) for six tissues from 9 representative mammals and one bird (chicken). The Neighbour-joining tree displayed is based on pairwise euclidean distances computed on the inclusion frequencies of orthologous skipped exons, for all organs. The tree illustrates the species-dominated clustering of alternative splicing data and thus constitutes the inverse pattern of that observed for global gene expression level data, as displayed in Figure 1 (see main text for details).*

coding genes. Nevertheless, hundreds of lncRNAs that are highly conserved at one or both levels have been identified[20,23]. It should be noted that the usually low abundance of lncRNA transcripts, which adds noise to expression measurements, may lead to overestimation of expression divergences[20]. Overall, comparisons of lncRNA expression levels reveal a mixed species/organ clustering, which likely reflects the presence of both fast-evolving and conserved organ-specific genes, but probably also the fact that lncRNA expression estimates are noisier than those for the more highly expressed protein-coding genes.

Rates of expression evolution are different in different mammalian lineages

Comparative transcriptome analyses have revealed that rates of expression change not only vary across the coding and noncoding transcriptome but also across mammalian lineages[10,11,16]. For example, gene expression levels evolve faster in primates (in particular in the human lineage) than

in rodents[16] (Figure 3). Similarly, novel alternatively spliced isoforms also accumulate at a higher rate in primates than in other mammalian lineages[17]. Thus, curiously, transcriptome evolution in primates is generally accelerated, in spite of the low mutation rate in this mammalian lineage[24] (e.g., rodents have approximately two to three times higher mutations rates), which implies a relatively small number of regulatory mutations in each generation.

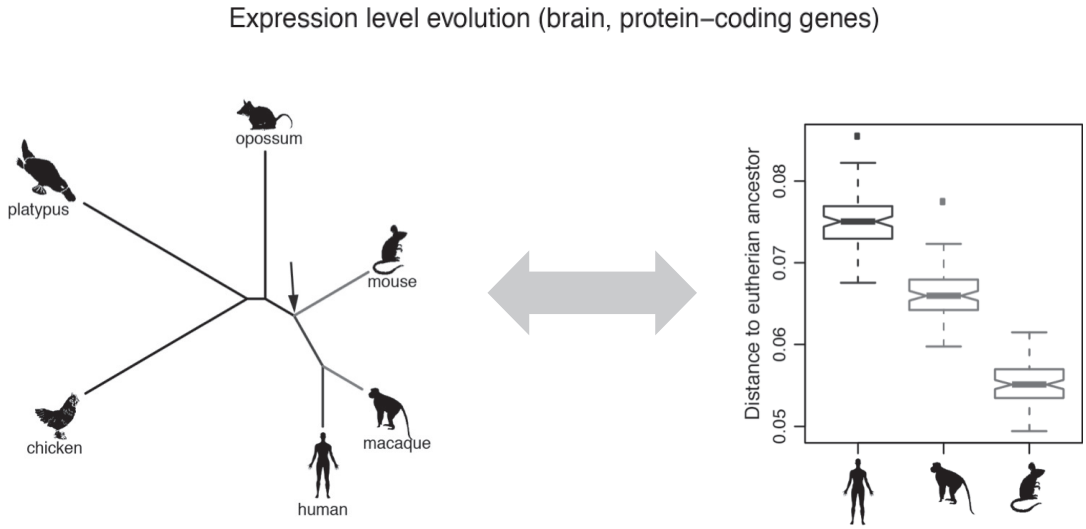

Figure 3. *Rates of gene expression evolution in different mammalian lineages. To the left: Neighbor-joining tree of protein-coding gene expression levels in the amniote brain. The tree was constructed based on pairwise distances between samples, estimated as 1-rho, where rho is Spearman's correlation coefficient. The arrow indicates the position of the ancestor of the three species. Expression data was taken from ref.[16]. To the right: Distance between each of the three placental mammal species and their common ancestor, in the expression tree displayed on the left. The boxplot represents the distribution observed with 100 bootstrap replicates, in which the same numbers of genes are resampled with replacement. The figures illustrate that gene expression evolution proceeds at a higher rate in primates and particular in humans compared to the rodent lineage due to the reduced efficiency of natural selection in primates (see main text for details).*

Given that the acceleration was observed for all organs studied so far[16], it is unlikely that it is explained by primate-specific adaptations. Furthermore, other lineages (e.g., monotremes) also show significantly higher rates of expression evolution than rodents[16]. It turns out that substantial differences in long-term effective population sizes among mammalian lineages explain the observed rate differences. For example, the long-

term effective population size of rodents[25] has been estimated to be at least ten times higher than that of humans[26,27] or chimpanzees[28]. Similarly, monotremes have had smaller long-term effective population sizes than rodents[29]. Importantly, lower effective population sizes imply less efficient natural selection. Thus, the accelerated transcriptome evolution in primates and monotremes relative to rodents is likely due to the accumulation of larger numbers of (slightly) deleterious regulatory mutations that cannot be purged by natural selection during evolution due to its inefficiency[16]. Evolutionary rate differences among mammalian lineages are thus consistent with a nearly neutral model of expression evolution, in which most mutations affecting gene expression are (slightly) deleterious[30,31]. Gene expression evolution is thus also consistent with the general notion that genetic drift is a primary force in molecular evolution, in particular in taxa with small effective population sizes[32].

Rates of expression change in different organs

In addition to cross-transcriptome and cross-lineage comparisons, expression rates have also been compared across a third dimension: organs. This work uncovered substantial rate differences among organs, likely reflecting the different selective pressures to which they are subjected. Thus, at one end of the evolutionary rate spectrum neural tissues stand out, with the overall lowest rates of gene expression divergence for the tissue examined so far[11,16,20,33] (Figure 4). This pattern reflects a genome-wide trend, as it holds true for both protein-coding[11,16,33] and lncRNA genes[20]. Notably, previous work also revealed that protein sequence divergence of brain-expressed genes also proceeds at a low rate[11,34]. These observations are surprising in view of the substantial remodeling of the size, structure, and cellular composition of the brain during evolution[16]. They are also notable when considering the difficulty in sampling and comparing corresponding brain regions across species[16], as these likely exaggerated actual biological expression differences. It was thus surmised that the brain has particularly sensitive and fine-tuned regulatory networks that are highly constrained during evolution. In this view, rapid remodeling of expression network components, for example through (developmental) expression changes of only a few key regulators[35] (e.g., miRNAs or transcription factors), which leave overall (adult)

transcriptomes unaltered, may have provided the basis for major phenotypic brain innovations – which undoubtedly took place (e.g., the sudden and pronounced increase of brain size and complexity in the human lineage[35]).

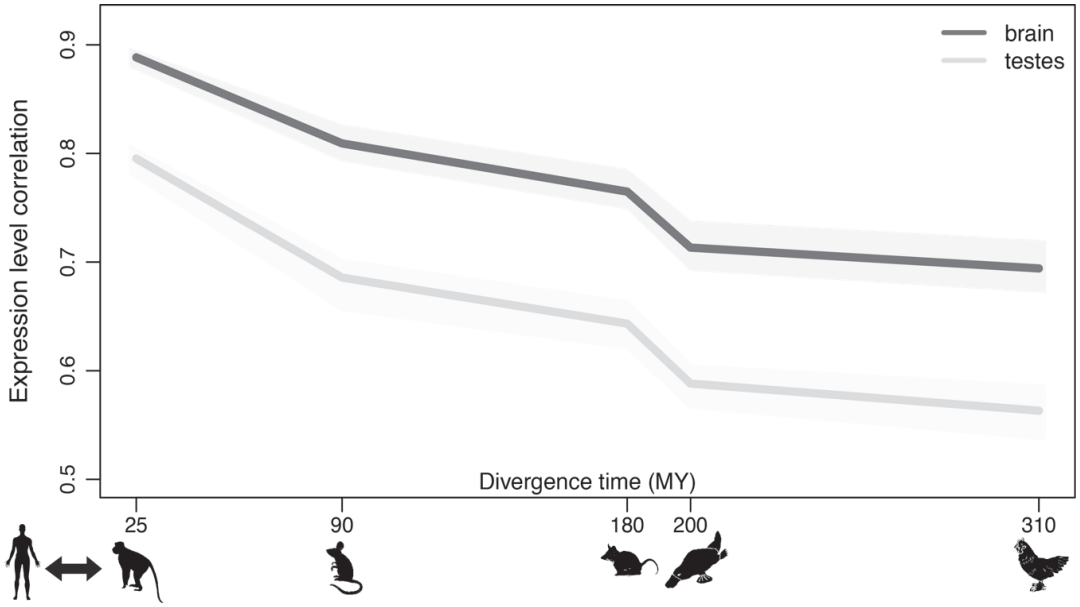

Figure 4. *Different rates of gene expression evolution in major mammalian organs. Pairwise Spearman's correlation coefficients between human and other species, as a function of the divergence time (in millions of years) between the two species, for brain and testis protein-coding gene expression levels. Expression data was taken from ref.[16]. The figure thus illustrates that expression levels evolve substantially more slowly in neural tissues than in the testes (see main text for details).*

While other somatic organs than the brain show intermediate rates of transcriptome change, the testis stands out at the other end of the spectrum, with an especially high rate of expression evolution[11,16,20] (Figure 4), consistent with rapid coding sequence evolution of testis genes[10,65]. There are likely two major reasons for this observation. First, the testis has been subject to strong positive selection throughout its evolution, which is likely associated with specific sex-related evolutionary pressures[36], in particular that of the co-called «sperm competition»; that is, the competitive process between spermatozoa of different males to fertilize the female egg. Second, purifying selection may generally be relaxed in the testis at the gene regulatory level. Indeed, transcription seems to overall be less tightly regulated and even partly «promiscuous» due to an overall per-

missive chromatin conformation (in particular during and just after meiosis) resulting from extensive repackaging of DNA during spermatogenesis[37].

Sex-related transcriptome evolution

Sex chromosomes represent the genomic hotspots of sex-related expression evolution. Recent work revealed that sex chromosomes emerged twice independently in mammals[38,39]. The original placental and marsupial (therian) sex chromosomes originated in the therian ancestor approximately 180 million years ago[39]. At about the same time, monotreme sex chromosomes[39] emerged (Figure 5). Notably, these two sex chromosomes are derived from different ancestral pairs of ordinary autosomes. The differentiation process of these sex chromosomes encompassed dramatic chromosome-wide remodeling of gene repertoires and expression patterns due to the emerging sex-related selective forces (Figure 5).

First, the specialization of the Y on sex determination and male functions, together with recombination arrests that weakened selection efficiency, entailed drastic losses of Y genes during mammalian evolution[39] (Figure 5). Several remaining Y genes evolved testis-specific expression patterns[39], consistent with the male-limited transmission of the Y and associated specialization on male functions. Strikingly, however, most Y genes apparently perform general regulatory functions (e.g., at the transcriptional and translation level) and maintained the ancestral ubiquitous expression patterns, although their expression levels overall decreased[39]. Altogether, these observations and other considerations (e.g., that regulatory genes are commonly dosage sensitive) led to the conclusion that most Y genes were initially preserved because of dosage constraints, in order to maintain (together with their X counterparts) the ancestral gene copy number. Thus, these Y genes contribute to proper functioning of male somatic tissues[39,40]. Notably, the functions in females are expected to be carried out by the corresponding two X copies in females. Consistent, with this dosage hypothesis, these genes escape X inactivation in females[39].

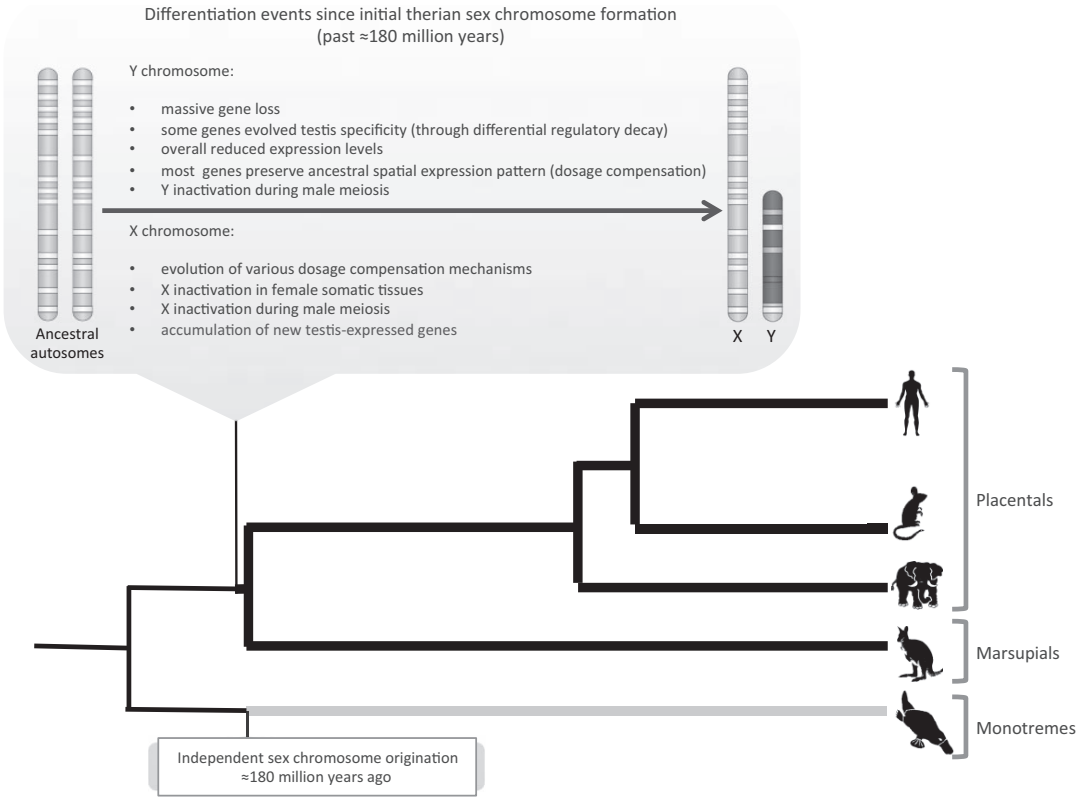

Figure 5. *Origins and functional evolution of mammalian sex chromosomes. The independent origins of sex chromosomes in the ancestor of therians (placentals and marsupials) and monotremes, and an overview of major gene content/expression remodeling events during therian sex chromosome evolution (see main text for details).*

However, this reasoning also raises the question of how the massive loss of all other genes from the Y was compensated. It is unlikely that among the hundreds of genes lost from the Y during sex chromosome differentiation there were no other dosage sensitive genes. The traditional and still prevailing hypothesis postulates that a two-fold upregulation of the single remaining X gene copies in males generally restored ancestral expression outputs[41,42]. It was further postulated that this upregulation was not specific to males and also affected females, which then would have twice too much expression output from the X. Common theory posits that this overexpression was, in turn, compensated by the by the well known process of female X chromosome inactivation. However, surprisingly, it was recently shown that the overall expression output of X-linked genes is only half that inferred for proto-sex chromosomes, at least in placental mammals[43,44] (marsupials do show signs of global upregulation of the

active X in several tissues[43,44]). Notably, it remains possible that individual X genes may nevertheless have become actively upregulated in placentals through specific mechanisms[43,45], although it has become clear that other mechanisms also contributed to X dosage compensation[43]. Furthermore, a number of genes may have been insensitive to dosage alterations altogether[43]. However, the reason underlying the evolution of chromosome-wide X inactivation in females remains unclear in view of the lack of an overall two-fold upregulation of the X in placentals.

Phenotypically relevant gene expression change

So far I have discussed global patterns of gene expression change and associated selective pressures across various biological dimensions. However, the ultimate goal in the evolutionary gene expression field is the identification of phenotypically relevant expression alterations. Specific approaches, developed to assess the selective pressures that have shaped gene expression patterns[13,14,16,46-48], have uncovered a number of lineage-specific expression changes of individual genes that may have been driven by positive selection. For example, comparisons among primates revealed an over-representation of transcription factors among potentially selectively-driven gene expression increases in the human lineage[13,49]. Other significant expression shifts in the human lineage include genes involved in metabolic pathways, consistent with the hypothesis that dietary changes may be associated with human-specific adaptations[49,50] and with previous reports of positive selection in the promoters of nutrition-related genes[9]. In most cases, the observed lineage-specific expression shifts are organ-specific[16,49]. Generally, across mammals, few lineage-specific expression shifts have been detected for the brain. The testis, on the other end, showed numerous expression shifts. Overall, these observations are consistent with highly differential selective pressures on gene expression in different organs[16] (see also above).

Given that genes commonly exert their functions together with other genes, concerted expression changes of distinct sets of functionally related genes during evolution are likely to be phenotypically relevant[16] (i.e., they are less likely to have been caused by genetic drift than expression changes of individual genes). Thus, a powerful approach to detect ex-

pression changes with a phenotypic impact may be the joint analysis of gene modules; that is, groups of genes with coherent expression patterns across a subset of samples, potentially including different species/organ/condition combinations[51]. This approach has revealed numerous modules of genes with highly conserved, organ- or tissue-specific expression[16,52,53] but also simultaneous lineage-specific expression shifts[16,54].

In a related approach, expression patterns are analyzed across multiple conditions (tissues, developmental stages, treatments etc.) and species to reconstruct an evolutionarily co-expression network[55]. Two genes are connected in this network if they show correlated expression levels in multiple species. Co-expression connections are enriched for functionally related genes[55] and thus can predict functions for previously uncharacterized genes, including lncRNAs[20]. Thus, this approach unveiled general functional properties of lncRNAs, such as that most of them likely act as negative regulators of gene expression, that many of them may have key roles in regulatory networks, and that lncRNAs are involved in key functional processes in various tissues[20] (e.g., spermatogenesis in the testis or synaptic transmission in the brain). More detailed scrutiny of individual genes in these networks uncovered intriguing individual gene expression shifts[20]. For example, a potential functional «partner» for the *H19* lncRNA was discovered. The spatial expression pattern of this lncRNA, termed *H19X,* a placenta-enriched X-linked lncRNA that acts as a miRNA precursor[20], significantly changed during evolution, from ancestral testis-specificity to placenta-predominant expression in placental mammals. This suggests that *H19X* contributed to the emergence and functional evolution of the placenta during mammalian evolution[20].

Regulatory basis of transcriptome evolution

Various insights into the underlying biological mechanisms underlying transcriptome evolution have recently emerged from intra- and inter-specific analyses of genome, epigenome and transcriptome variation. For example, genome-wide association studies in human populations have identified DNA sequence variants potentially underlying various molecular phenotypes, including gene expression, alternative splicing patterns, histone modifications or transcription factor binding patterns[56-58]. Interes-

ting examples for between-species regulatory changes that were detected based on genome sequence comparisons include human-specific alterations of promoter sequences, potentially driven by positive selection, for opioid hormone[59] genes and genes involved in neural development[9].

However, recent technological developments now enable more direct assessments of gene regulatory evolution. For example, cross-species ChIP-seq analyses have revealed that binding of the liver-specific transcription factors CEBPA and HNF4A is highly species-specific[60], contrary to the strong evolutionary conservation observed for the more general factor CTCF[61]. Comparative analyses of DNA methylation[62,63] and histone modification patterns[64] in primates also identified numerous lineage-specific changes of these epigenetic marks.

Outlook

The first wave of the new generation of transcriptome studies in mammals, facilitated by next generation sequencing technologies (in particular RNA-seq) have revealed intriguing commonalities and differences in evolutionary patterns for different parts of the transcriptome (protein-coding and noncoding gene expression levels and alternative splicing), mammalian lineages, chromosome types, cell types, organs and sexes.

However, the exploration of mammalian transcriptome evolution has only begun. First, little is known about the evolution of certain aspects of the transcriptome, such as untranslated regions of protein-coding genes, noncoding PIWI-interacting RNAs (piRNAs), or the recently discovered circular noncoding RNAs[65,66]. Second, with a few exceptions[35,67-69], most comparative transcriptome investigation have so far have centered around the analysis of adult organs. Thus, these studies have provided initial clues to the relevance of gene expression changes for the evolution of species- or lineage-specific organ physiologies. However, adult organ anatomies, which define physiological functions, are rooted in development. Hence, the bulk of phenotypically relevant expression change may occur during development. Future studies should thus seek to include developmental samples. Third, the use of samples representative of whole organs in hitherto existing studies is also sometimes potentially problematic due to

between-species differences in the cellular composition of tissues. Recent technological developments, such as the advent of sensitive single-cell RNA sequencing methods[70], will allow to resolve this limitation in future studies. Fourth, while considerable advances have been made regarding the investigation of regulatory changes underlying gene expression change, only few attempts have been made to study regulatory mechanisms and gene expression patterns at the same time. In particular, it will be important to investigate how changes in transcription factor binding patterns affect expression evolution, especially given the recent observations that most interactions between transcription factors and chromatin do not result in significant changes in the expression of putative target genes[71]. Notably, such studies face the complex task of identifying accurate regulator-target gene relationships, which have so far often relied on the simplifying notion that the closest gene to regulatory sequence is the target of regulation[72]. Future work will thus benefit from chromatin conformation capture techniques to more rigorously identify regulatory element-gene relationships[73]. Fifth, an important limitation of current gene expression comparisons is that they have mostly relied on comparisons at the transcriptome level. Given that expression regulation for protein-coding genes may also occur in the several layers that succeed transcription (e.g., translation, protein degradation), future studies should seek to also compare protein synthesis and actual protein abundance levels, in order to fully understand gene expression evolution and its phenotypic impact.

Finally, gene expression comparison can only identify candidates for potentially adaptive expression changes. The next and in many ways more challenging step is to perform detailed functional characterizations of detected expression shifts (e.g., through knockout or transgene experiments[74]), with the aim to actually demonstrate their phenotypic implications.

REFERENCES

1 Varki, A. A chimpanzee genome project is a biomedical imperative. *Genome Res* **10**, 1065–70 (2000).

2 Lynch, V.J., May, G. & Wagner, G.P. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature* **480,** 383–6 (2011).

3 King, M.C. & Wilson, A.C. Evolution at two levels in humans and chimpanzees. *Science* **188,** 107–16 (1975).

4 Britten, R.J. & Davidson, E.H. Gene regulation for higher cells: a theory. *Science* **165,** 349–57 (1969).

5 Hoekstra, H.E. & Coyne, J.A. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61,** 995–1016 (2007).

6 Carroll, S.B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134,** 25-36 (2008).

7 Wray, G.A. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* **8,** 206–16 (2007).

8 Ponting, C.P. The functional repertoires of metazoan genomes. *Nat Rev Genet* **9,** 689–98 (2008).

9 Haygood, R., Fedrigo, O., Hanson, B., Yokoyama, K.D. & Wray, G.A. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat Genet* **39,** 1140–4 (2007).

10 Enard, W. *et al.* Intra- and interspecific variation in primate gene expression patterns. *Science* **296,** 340–3 (2002).

11 Khaitovich, P. *et al.* Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309,** 1850–4 (2005).

12 Hawkins, R.D., Hon, G.C. & Ren, B. Next-generation genomics: an integrative approach. *Nat Rev Genet* **11,** 476–86 (2010).

13 Gilad, Y., Oshlack, A., Smyth, G.K., Speed, T.P. & White, K.P. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* **440,** 242–5 (2006).

14 Blekhman, R., Marioni, J.C., Zumbo, P., Stephens, M. & Gilad, Y. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res* **20,** 180–9 (2010).

15 Jordan, I.K., Marino-Ramirez, L. & Koonin, E.V. Evolutionary significance of gene expression divergence. *Gene* **345,** 119–26 (2005).

16 Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478,** 343–8 (2011).

17 Merkin, J., Russell, C., Chen, P. & Burge, C.B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338,** 1593–9 (2012).

18 Barbosa-Morais, N.L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338,** 1587–93 (2012).

19 Meunier, J. *et al.* Birth and expression evolution of mammalian microRNA genes. *Genome Res* **23,** 34–45 (2013).

20 Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505,** 635–40 (2014).

21 Kutter, C. *et al.* Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet* **8,** e1002841 (2012).

22 Washietl, S., Kellis, M. & Garber, M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res* **24,** 616–28 (2014).

23 Chodroff, R.A. *et al.* Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol* **11,** R72 (2010).

24 Li, W.H., Ellsworth, D.L., Krushkal, J., Chang, B.H. & Hewett-Emmett, D. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol Phylogenet Evol* **5,** 182–7 (1996).

25 Keightley, P.D., Lercher, M.J. & Eyre-Walker, A. Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol* **3,** e42 (2005).

26 Kaessmann, H., Heissig, F., von Haeseler, A. & Paabo, S. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat Genet* **22,** 78–81 (1999).

27 Yu, N., Fu, Y.X. & Li, W.H. DNA polymorphism in a worldwide sample of human X chromosomes. *Mol Biol Evol* **19,** 2131–41 (2002).

28 Kaessmann, H., Wiebe, V. & Paabo, S. Extensive nuclear DNA sequence diversity among chimpanzees. *Science* **286,** 1159–62 (1999).

29 Warren, W.C. *et al.* Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453,** 175–83 (2008).

30 Khaitovich, P. *et al.* A neutral model of transcriptome evolution. *PLoS Biol* **2,** E132 (2004).

31 Khaitovich, P., Enard, W., Lachmann, M. & Paabo, S. Evolution of primate gene expression. *Nat Rev Genet* **7,** 693–702 (2006).

32 Lynch, M. *The origins of genome architecture.,* (Sinauer Associates., Sunderland, USA, 2007).

33 Chan, E.T. *et al.* Conservation of core gene expression in vertebrate tissues. *J Biol* **8,** 33 (2009).

34 Warnefors, M. & Kaessmann, H. Evolution of the correlation between expression divergence and protein divergence in mammals. *Genome Biol Evol* **5,** 1324–35 (2013).

35 Somel, M. *et al.* MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biol* **9,** e1001214 (2011).

36 Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol* **3,** e170 (2005).

37 Soumillon, M. *et al.* Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep* **3,** 2179–90 (2013).

38 Veyrunes, F. *et al.* Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res* **18,** 965–73 (2008).

39 Cortez, D. *et al.* Origins and functional evolution of Y chromosomes across mammals. *Nature* **508,** 488–93 (2014).

40 Bellott, D.W. *et al.* Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508,** 494–9 (2014).

41 Charlesworth, B. Model for evolution of Y chromosomes and dosage compensation. *Proc Natl Acad Sci U S A* **75,** 5618–22 (1978).

42 Ohno, S. *Sex chromosomes and sex-linked genes,* (1967).

43 Julien, P. *et al.* Mechanisms and evolutionary patterns of Mammalian and avian dosage compensation. *PLoS Biol* **10,** e1001328 (2012).

44 Lin, F., Xing, K., Zhang, J. & He, X. Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation. *Proc Natl Acad Sci U S A* **109,** 11752–7 (2012).

45 Deng, X., Berletch, J.B., Nguyen, D.K. & Disteche, C.M. X chromosome regulation: diverse patterns in development, tissues and disease. *Nat Rev Genet* **15,** 367–78 (2014).

46 Bedford, T. & Hartl, D.L. Optimization of gene expression by natural selection. *Proc Natl Acad Sci U S A* **106,** 1133–8 (2009).

47 Warnefors, M. & Eyre-Walker, A. A selection index for gene expression evolution and its application to the divergence between humans and chimpanzees. *PLoS One* **7,** e34935 (2012).

48 Rohlfs, R.V., Harrigan, P. & Nielsen, R. Modeling gene expression evolution with an extended ornstein-uhlenbeck process accounting for within-species variation. *Mol Biol Evol* **31,** 201–11 (2014).

49 Blekhman, R., Oshlack, A., Chabot, A.E., Smyth, G.K. & Gilad, Y. Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS Genet* **4,** e1000271 (2008).

50 Babbitt, C.C., Warner, L.R., Fedrigo, O., Wall, C.E. & Wray, G.A. Genomic signatures of diet-related shifts during human origins. *Proc Biol Sci* **278,** 961–9 (2011).

51 Bergmann, S., Ihmels, J. & Barkai, N. Iterative signature algorithm for the analysis of large-scale gene expression data. *Phys Rev E Stat Nonlin Soft Matter Phys* **67,** 031902 (2003).

52 Piasecka, B., Kutalik, Z., Roux, J., Bergmann, S. & Robinson-Rechavi, M. Comparative modular analysis of gene expression in vertebrate organs. *BMC Genomics* **13,** 124 (2012).

53 Oldham, M.C., Horvath, S. & Geschwind, D.H. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci USA* **103,** 17973–8 (2006).

54 Konopka, G. *et al.* Human-specific transcriptional networks in the brain. *Neuron* **75,** 601–17 (2012).

55 Stuart, J.M., Segal, E., Koller, D. & Kim, S.K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302,** 249–55 (2003).

56 Bryois, J. *et al. Cis* and *trans* effects of human genomic variants on gene expression. *PLoS Genet* **10,** e1004461 (2014).

57 Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501,** 506–11 (2013).

58 Kilpinen, H. *et al.* Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* **342,** 744–7 (2013).

59 Rockman, M.V. *et al.* Ancient and recent positive selection transformed opioid cis-regulation in humans. *PLoS Biol* **3,** e387 (2005).

60 Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328,** 1036–40 (2010).

61 Schmidt, D. *et al.* Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148,** 335–48 (2012).

62 Hernando-Herraez, I. *et al.* Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet* **9,** e1003763 (2013).

63 Pai, A.A., Bell, J.T., Marioni, J.C., Pritchard, J.K. & Gilad, Y. A Genome-Wide Study of DNA Methylation Patterns and Gene Expression Levels in Multiple Human and Chimpanzee Tissues. *PLoS Genetics* **7,** e1001316 (2011).

64 Shulha, H.P. *et al.* Human-specific histone methylation signatures at transcription start sites in prefrontal neurons. *PLoS Biol* **10,** e1001427 (2012).

65 Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495,** 333–8 (2013).

66 Hansen, T.B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495,** 384–8 (2013).

67 Somel, M. *et al.* Transcriptional neoteny in the human brain. *Proc Natl Acad Sci USA* **106,** 5743–8 (2009).

68 Somel, M. *et al.* MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain. *Genome Res* **20,** 1207–18 (2010).

69 Irie, N. & Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat Commun* **2,** 248 (2011).

70 Sandberg, R. Entering the era of single-cell transcriptomics in biology and medicine. *Nature Methods* **11,** 22–24 (2013).

71 Cusanovich, D.A., Pavlovic, B., Pritchard, J.K. & Gilad, Y. The functional consequences of variation in transcription factor binding. *PLoS Genet* **10,** e1004226 (2014).

72 Gerstein, M.B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489,** 91–100 (2012).

73 Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–80 (2012).

74 Romero, I.G., Ruvinsky, I. & Gilad, Y. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet* **13,** 505-16 (2012).

75 Wang, E.T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456,** 470–6 (2008).

76 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24,** 1586-91 (2007).

77 Vilella, A.J. *et al.* EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19,** 327–35 (2009).

78 Cain, C.E., Blekhman, R., Marioni, J.C. & Gilad, Y. Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics* **187,** 1225–34 (2011).

79 Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res* **42,** D749–55 (2014).

Acknowledgements

I would first and foremost like to thank the Cloëtta Foundation for honoring me with this award – it represents the culmination point of the career I have had so far. I would never have received this award without the outstanding technicians, graduate students, postdoctoral fellows, and administrative assistants I have had the great pleasure and privilege to work with over the years. They have all not only contributed to the success of the science of my lab but have also made my job as a scientist an extremely pleasant and rewarding one. I indeed gain most pleasure during my workday from the interactions with my collaborators, and the pleasure of doing science is greatly amplified by making novel discoveries together – as a team. I am also extremely grateful for the invaluable support received from the various supervisors and mentors I have had during my career. I would not be the scientist I am today without them. I would also like to thank the numerous colleagues and collaborators around the world that have allowed us to carry out exciting and often unusual projects. Research is only possible with substantial funding, and I would thus like to thank the European Research Council (ERC), the European Union, the European Molecular Biology Organization (EMBO), the Federation of European Biochemical Societies (FEBS), the Human Frontier Science Program (HFSP), Novartis, Roche, the Swiss National Science Foundation, and the University of Lausanne for financial support. Finally, I would like to thank my wife, Stephanie, and my three children, Leon, Anouk, and Nuno, as well as my entire family, for their constant support and for making my life worth living.