UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1916

The revolutionary partnership of computation and biology

SALVADOR DANIEL RIVAS-CARRILLO

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2023

ISSN 1651-6206
ISBN 978-91-513-1516-4
URN urn:nbn:se:uu:diva-473354

Dissertation presented at Uppsala University to be publicly examined in Room A1:107a, BMC, Husargatan 3, Uppsala, Wednesday, 26 April 2023 at 09:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Associate Professor Jessica Lindvall (National Bioinformatics Infrastructure Sweden).

Abstract
Rivas-Carrillo, S. D. 2023. The revolutionary partnership of computation and biology. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1916. 51 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1516-4.

The organization of living beings is complex. Science uses modeling in order to gain a deeper understanding, and to be able to manipulate the processes of living organisms. To this purpose, I used and developed computational tools to investigate and model different relevant biological phenomena.

In paper I, I utilized whole-genome data from wild and domesticated European rabbit (Oryctolagus cuniculus sp.) populations to identify segregating insertions of endogenous retroviruses and compare their variation along the host phylogeny and domestication history. The results from this study highlight the importance of genomic modeling beyond reference organisms and reference individuals, and provide deep insights regarding strategies for variant analyses in host population comparative genomics. In paper IV, I studied the process of exaptation of foreign genetic elements at broad-scale by observing the presence and characteristics of retroviral env gene, syncytin, across vertebrates. I searched a library of more than 150 chromosome-length assemblies covering 17 taxonomical orders for syncytin homologs, where I identified and syntenically aligned over 300 loci insertions, including not previously known insertions. Additionally, three-dimensional structures of the recovered sequences were predicted using AlphaFold2. Phylogenomics analyses suggest a complex dynamic of multiple retroviral insertions at different time points with sequence conservation specific to clades that share a similar histo-physiological placental type.

In paper II, I expanded the scope to encompass translational medicine by developing an unsupervised machine learning methodology for detecting anomalies in biomedical signals, MindReader, which I applied primarily to electroencephalogram. In paper III, I developed a hidden Markov model implementation that includes a hypothesis generator for stream time-domain signals, which is used as a dependency for paper II. The work in this thesis substantiates that a combination of biological knowledge, cutting-edge technology, and robust algorithmic design constitute the primordial factors for scientific advancement.

Salvador Daniel Rivas-Carrillo, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

ISSN 1651-6206
ISBN 978-91-513-1516-4
URN urn:nbn:se:uu:diva-473354 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-473354)

All of those who by their courageous actions and incorruptible integrity have inspired us to carry on

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I **Salvador Daniel Rivas-Carrillo**, Mats E. Pettersson, Carl-Johan Rubin, Patric Jern (2018). "Whole-genome comparison of endogenous retrovirus segregation across wild and domestic host species populations". Proceedings of the National Academy of Sciences 115 (43): 11012-11017. doi:10.1073/pnas.1815056115.

II **Salvador Daniel Rivas-Carrillo**, Evgeny E. Akkuratov, Hector Valdez Ruvalcaba, Angel Vargas-Sanchez, Jan Komorowski, Daniel San-Juan, Manfred G. Grabherr (2023). "*MindReader*: unsupervised electroencephalographic reader". Sensors 2023, 23, 2971. https://doi.org/10.3390/s23062971.

III **Salvador Daniel Rivas-Carrillo**, Manfred G. Grabherr (2023). "*HiddenMarkovModelReaders*: A Julia implementation of a Hidden Markov Model and unsupervised hypothesis generation for signal processing". (submitted).

IV **Salvador Daniel Rivas-Carrillo**, Wensi Zhu, Olga Dudchenko, Erez Lieberman Aiden, Arne Elofsson, Tanel Punga, Manfred G. Grabherr, Parwinder Kaur (2023). "Broad-scale *in silico* assessment of retroviral exaptated gene: *syncytin*". (manuscript).

Reprints were made with permission from the publishers.

Contents

Part I:
Introduction

Revolution times: Age of information

It is our duty to make this world a better place.

Conditions are always changing, such is the nature of reality. The development of human civilization has been largely diverse and erratic. Nevertheless, a common pattern, regardless of the age, is the paroxysmal emergence of technological advancements. Such events have rippled on most, if not all aspects, of society. For example, the printing press invented by Johannes Gutenberg in 1436 allowed literacy to reach unprecedented levels; the industrial revolution radically transformed the manufacturing process and mass production of goods; or, the global system of interconnected computer networks revolutionized information spread. Likewise, scientific and technological advancements of our era promise to erradicate poverty and hunger, to improve public health and education, to widen the availability of clean and affordable energy, and in general, to provide better conditions for all. Thus, I believe that to change our world for the better we must work, create and innovate on scientific and technological breakthroughs that will facilite such goals.

Moreover, it seems as if last decades of human history have witnessed a dramatic acceleration in how these changes occur. Probably a substantial facilitator of these circumstances are high-throughput technologies and network connectivity. The fact that we, as a collective society, produce and share more information than ever before has directly and indirectly impacted all aspects of our lives. This revolution extends from our daily lives all the way to industrial and research environments, such as health sciences, media production, etc.

This wealth of information conceals valueable knowledge, which demands analytical algorithms to be unraveled. These methods exceed human capacity and stipulate the urge for computational assistance.

About this thesis

During this thesis I will elaborate on the journey I have engaged to partake into this technological and scientific revolution of our time. This exploration has led me to develop computational skills, and expand knowledge and understanding for applying and building tools to address a variety of problems.

I tackled different problems by computational modeling and data analysis from different perpespectives. That is from genetic code to phenotypes, where I studied evolutionary relationship among organisms and populations on paper I and paper IV; and from phenotypic signals to their importance at a molecular level, where I studied the electrical recording of the brain activity to predict physiological and pathological states on paper II and paper III.

Therefore, I divided this text into three sections:

First, I ponder on and present my reflections, amidst the current reproductibility crisis among scientific research, on why it is of paramount importance to build and use reliable and robust software, which profits from the advancements of other fields of science such as computer science, mathematics and physics.

Next, I provide a brief and non-exhaustive overview of the specific scientific background on the problems I studied. Thus, I address key points on genetics and evolution where I present a general vision on how information transforms in the biological context and how it relates to the factors are involved. Then, I proceed to further elaborate on transposable elements and structural variants, with special focus on endogenous retroviruses, not for a prominent reason in the biological sense, but only because these elements were the subject of a large part of these studies. I continue to discuss on the phenomena of exaptation, particularly the retroviral *env* gene, and how it has impacted evolution of the most dominant group of animals on the planet today, the underlying histo-physiological mechanisms and repercussions from an evolutionary perspective. Similarity, I elaborate on the application of computational expertise to translational medicine, particularly processing of big data and machine learning, to focus on clinical problems as well as the usefulness of data collection for both better understanding of challenging and puzzling pathologies, and improvement on health care and quality of life.

Finally, I briefly explain the highlights of the included projects in the frame of this thesis, and why they are important for scientific progress.

1. On the development of computational methods

A layer of abstraction: Computational complexity

Computer science is to biology what calculus is to physics. It's the natural mathematical technique that best maps the character of the subject.

The history of programming could probably take us back to late 40s when people input binary code directly to machines to perform basic operations: addition, subtraction, multiplication, etc. Later, during late 50s the realization that a level of abstraction could be moved up by creating a programming language closer to what we understand nowadays arose. This paradigm shift had arguably two large effects:

- First, the code was no longer tight up to a particular hardware, therefore it could potentially run on all machines and would only be written, and maintained, once.

- Second, the fact that the lower abstraction was not needed meant that specialists were not required, thus, computer power became accessible to a larger community and development accelerated.

Learning from these lessons, I wished to democratize the algorithmic abstraction of scientific computation, therefore rendering it available to people whose expertise and curiosity focusses on investigating biological phenomena. To this end, I have worked on several computational tools that have been used during the projects on this thesis that include a variety of programming languages and computational paradigms, benefit enormously from modern programming designs, and whose common thread is the interest to address biomedical issues in a cross-disciplinary manner.

During these studies, I used a variety of algorithms for data manipulation and analysis. These include methods for sequence alignment, distance measurement and phylogenetics, such as Smith-Waterman algorithm, Burrow-Wheelers transform, and Levenstein distance; and algorithms for data analysis and signal processing, such as Fourier Transform (FT), Neural Networks (NN), hidden Markov models (HMM), Principal Component Analysis (PCA),

among many others.

Algorithms for sequence alignment and sequence assembly are vital to study of genomics. Nevertheless, an important consideration is the computation time these algorithms use. Therefore, it is not uncommon to use heuristic rather than optinal solutions.

Fourier Transform (FT) is an important and widely used algorithm that transforms signals into their frequencies. The ubiquitous usability of FT in mathematical analysis is thanks to the implementation of Fast Fourier Transform (FFT). FFT is an algorithmic implementation of Discrete Fourier Transform (DFT). Unlike DFT, whose computational complexity is exponential, FFT's computational complexity is linear. This is achieved by factorizing the DFT matrix into a product of sparse factors. This process exemplifies one of the primary limitations of certain computations, as they are virtual impossible to scale up if their complexity is logarithmic.

NNs have witnessed a spike in popularity due to their adapatability to different problems. Current deployments also make them accesible to a wider variety of users, allowing room for manipulation. NNs have been particularly successful were large amounts of annotated data is available. Different architectures have been applied for different purposes to common problems, for example, autoencoders, that is symmetrical architectures with a encoding and decoding segment, are useful for data compression and reconstruction.

HMMs have as well as a wide range of applications, from genomics to signal processing. They are used to model systems that are compatible with Markov processes, that is a sequence of possible events in which the likehood of occurence at a certain point depends only on the previous state.

PCA is technique for analizing large datasets consisting of high number of dimensions features. PCA performs a linear transformation of the data to a coordinate system where the variability can be explained in a few dimensions. Thus, essentially performing dimensionality reduction. This features renders PCA ideal for data visualization.

2. On the topics of study

Que nadie suba a la tribuna sin motivo justo, y que nadie baje de ella sin el sentido de la dignidad cumplida.

2.1 Genetic information

Life is an intriguing phenom. The levels of complexity and the diversity it has created over generations is astonishing. From an evolutionary perspective, living organisms act together as information entities that in the case of animals, compose a community with their reproductively available peers.

To be able to self-preserve, information must replicate itself. However, to accommodate for random adaptability, variation must be introduced into this cycle. One possible method to achieve such objective is by permitting or tolerating errors to occur during replication process. An alternative route is to acquire and integrate foreign, fully- or semi-functional, information. Both of these methods convey risks and benefits. During these studies, I have explored both of these aspects in depth, as detailed below. This informational paradigm is a key component of living organisms as they carry encoded data into their DNA, where all necessary elements to produce a viable creature are stored. Likewise, this record can be copied and replicated by the cellular machinery in order to be conserved. And this record can also be altered within the boundaries mentioned above to produce variation, which permits adaptation. However, in complex living organisms, dependent on thousands of physiological and morphological building blocks, variation created by single mutations is, mostly, not sufficient to accommodate the change needed for adaptation to the environment. In such cases, alternative routes that involve relatively low cost and low risk to deal with the problem of limited variation arise. For instance, sexual reproduction, which functions by combining two randomized encoded information sets into a new version.

Genetic variation

Genetic information is a key aspect of preservation. By using a biochemical stable encrypted system, it is possible for living organisms to perpetuate the

information necessary to create basic biochemical building blocks, proteins. Using this condition, living organisms pass on their genetic information to their descendants.

Natural selection, in the context of biological systems, is an iterative process to create rewards for situational advantageous traits. In this case, each iteration represents a generation, and the reward is a reflection of the preparedness of the organism. In such manner, variation and natural selection make a combined driving force that shapes living organisms.

A key principle of variation is DNA sequence alterations. These alterations can range from single nucleotide to large-scale variations. Genomic studies indicate that such structural variations are common occurrences, but as is the case for most single nucleotide polymorphisms, the vast majority are selectively neutral [1].

Adaptive change

Exaptation refers to the shift on function, and thus emergence of a novel function, that enhances the fitness of individuals [2]. However, the line to define exaptation or cooption is blurred. Traditionally, these concepts are drawn as a sophisticated version of genetic repurposing. However, the fundamental problem with repurposing strives on that a precondition to not use the current version of the trait is required.

This precondition can be met in two situations:

- When the feature is duplicated, thus one copy can fulfil the canonical role whereas the second copy can be released from selective pressure, then be free to change.

- When the current feature is no longer needed or deprecated as a residual or remnant feature.

In either of these alternatives, the cost of keeping an unfunctional information copy or overhead on the overall fitness must be low enough in the context of the system in order to be tolerated.

Transposable elements match perfectly in either of these preconditions, or possibly even both at a certain moment. Contrary to this, it is not unexpected that transposable elements provide hosts with abundance of material for cooption. Probably the most characteristic exemplification of such phenomena is the *syncytin* gene. The *syncytin* gene has been exapted sev-

eral times during vertebrate evolution, which provides the ability for cellular membrane fusion, and presumably concomitant immunosuppressive properties [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13].

Questions naturally arise on the reasons why this particular retroviral gene, and no other foreign or own genes developed as *syncytin* did. Arguments in favor of this point could be their ubiquitous nature in terms of host range, germ line presence, and consequently genome presence through generations. Moreover, probably the cost of coopting an already available feature is significantly lower, and more likely than *de novo* development, or even no similar protein structure to derive it from existed. Finally, the benefit for coopting a feature to be used during an early stage of gestation would efficiently select viable versions, and facilitate long term retention since such features would significantly improve the reproductive capabilities of the individual by merely survival, thus converting this particular case into a zero-none game strategy. This latter argument also advocates for repeated cooptions.

On the contrary, *syncytin* cooption required the perfect combination of events, i.e., the usefulness of such feature, the tolerance of non-functional elements that could provide the feature, the compatibility of this potential feature with current features, i.e., developed immune system and amniote gestational development. However, as unlikely as this combination of events is at a given single time point in the evolution of life on this planet, once these requirements are met, reimplementing different variants of this feature becomes more probable, thus, reducing the cost and increasing the likelihood.

2.1.1 Eukaryotic transposable elements

In the early days of 2000s, Lander and collaborators have published the **Initial sequencing analysis of the human genome** where they described, among other points, the presence of repeat sequences in roughly 50% of the human genome, namely mobile elements [14]. Others, such as de Koning and collaborators, have estimated based on computer algorithms that highly mutated transposable elements account for as much as 70% of the human genome [15]. In both cases, these observations suggested the importance of repeat sequences as paleontological records, a track to the past [16].

Importance of mobile elements

Mobile elements constitute a significant genome fraction of many species, mainly due to their prolific self-replicating behavior [17]. Mobile elements are also responsible for genetic alterations during and independent of cell replica-

tion that sometimes leads to alterations and serious pathological conditions, such as neoplasias [18, 19, 20]. They composed a heterogeneous group whose common trait is the capability of self-replication and insertion into the host genome. They used different intermediate amplification mechanisms, either by RNA, transposable elements class I, or by DNA, transposable elements class II. Retrotransposons are further sub classified according to the presence of flanking sequences called long terminal repeats (LTRs) [21, 22, 23].

Enormous variation has been observed across species in abundance of DNA transposons as well as RNA transposons [24]. For instance, DNA transposons seem completely absent in budding and fission yeasts [25] whereas they contribute to about 65% of genome expansion in *Trichomonas vaginalis* [26].

Transposable elements are an important source of genomic alterations and structural variation [1]. Moreover, transposable elements also play different roles on other biological levels, for example:

- **Abundance**: Some of the largest genomes studied to this day have reached such sizes at the expense of transposable elements. Examples of such are copious among plants [27], e.g. LTR retrotransposons represent three quarters of the maize (*Zea mays*) genome [28]; pea (*Pisum sativum*) genome has grown most likely due to a recent expansion of retrotransposons [29]. The same pattern has been observed in some vertebrates, for instance Mexican axolotl (*Ambystoma mexicanum*), a representative of salamanders and important regeneration model, harbors approximately 60% of LTR retrotransposons of more than 10 kb in length, indicative of an evolutionary recent expansion burst [30, 31]. Contrary to this, other large scale studies have presented evidence of negative correlation between body size and LTR retrotransposons in mammals [32, 33, 34]. In medium-ground finch (*Geospiza fortis*) and zebra finch (*Taeniopygia guttata*), an unusual proportion of LTR-endogenous retroviruses has been observed in comparison with other birds (approximately 4 to 10%) [35].

- **Gene regulation**: Transposable elements insertions in primates, including humans, are a common occurrence. *Alu* elements, a type of retrotransposon particularly abundant in primates, have been observed to influence RNA processing [36]. In the cultivated octoploid strawberry (*Fragaria x ananassa*), where transposable elements make up approximately 36% of the genome, subgenome dominance is negatively correlated with the presence of transposable elements. Moreover, transposable elements abundance and expression predict gene expression dominance and gene loss at individual homolog level [37, 38]. Identification of specific transposable elements silencing systems, such as KRAB-

ZFPs, emphasizes the biological importance of transposable elements [17, 19].

- **Recombination**: Since transposable elements arise from a few sets of copies, they constitute repetitive portions of genomes, which greatly facilities recombination [39, 40]. In addition to this, studies have demonstrated transposable elements playing a fundamental role in genome recombination.

- **Protein innovation**: Probably the most emblematic example of protein innovation mediated by transposable elements are the *syncytin* genes that serve a key function in placentation [41]. These genes are the result of multiple exaptation events of the retroviral envelope (*env*) gene on several eukaryote linages, mostly eutherian mammals, such as primates [3], ruminants [7], rodents [13, 42], lagomorphs [4], but also marsupials [9] and even Mabuya lizards [10]. Moreover, some cases are strongly correlated with the emergence of specific taxonomic groups, e.g., *Hyaenidae* [11].

2.1.2 Retroviruses

Retroviruses are a widely spread group of cellular parasites whose common characteristic is the use of a reverse transcriptase to copy their RNA genome to the hosts DNA genome as a provirus. This process is essential for their replication cycle, which involves invading the genetic material of the host. Since these entities have been interacting with vertebrate genomes for millions of years they have repeatedly entered hosts germ line, embedding themselves in the genome and passing on to the progeny, thus becoming endogenous retroviruses [24, 23, 43]. In fact, they have invaded their hosts genomes so successfully that it blurs the perception of invader and host. Contrary to this, a combination of recent availability of a large amount of eukaryotic genomes with enhanced bioinformatics tools to detect and date endogenous retroviruses has yielded the emerging field of paleovirology [44, 45, 46, 47, 48].

Endogenous retroviruses

LTR autonomous elements are commonly known to as endogenous retroviruses (ERVs) since they are considered relics of ancient retroviruses that once invaded the germ line. They are part of a larger heterogeneous group of mobile elements whose common trait is the capability of self-replication and insertion into hosts genome [23, 49].

Retroviral taxonomy

Retroviruses are taxonomically classified by the International Committeee on the Taxonomy of Viruses (ICTV) into the family *retroviridae* and further subdivided into subfamilies *orthovirinae* and *spumavirinae* that together composed seven genera, i.e., *alpharetrovirus*, *betaretrovirus*, *gammaretrovirus*, *deltaretrovirus*, *epsilonretrovirus*, *lentivirus* and *spumavirus*. To this day, millions of endogenous retroviruses have been identified in a variety of hosts, representing every genera of the *retroviridae* family. Nevertheless, only 56 species of retroviruses are actively infecting hosts. A recent and comprehensive nomenclature was proposed by experts in the field [49, 50, 21, 51], in order to facilitate their annotation.

Proviral retrovirus structure

A complete retrovirus proviral sequence is approximately 7 to 11 kb, including flanking LTR sequences. They consist of structural genes (*gag*, *pro*, *pol* and *env*), which order is conserved among all retroviruses. In addition, arrangement of their major cleavage products is conserved, too, mainly because virion proteins should be expressed in the proper amount, interact in specific order and be guided into correct position. Furthermore, some endogenous retroviruses have acquired additional genes, such as *vif*, *vpr*, *nef* or *vpu* [52, 53, 54].

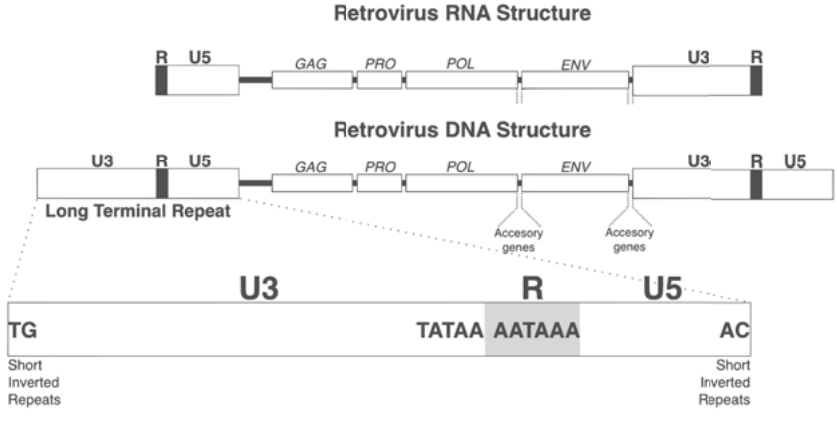

Figure 2.1. Schematic representation of retroviral structure in RNA and DNA space with zoom into long terminal repeat (LTR).

Importance of LTR

LTRs are the hallmark of endogenous retrovirus replication. They are generated simultaneously during reverse transcription at both ends of the genome.

Each LTR is composed of unique U3 and U5 regions separated by a repeated segment (R). The U3 region varies in length and contains binding sites for different cellular transcription factors for proviral enhancement and promotion. These sequences vary in length, but most of them are around 300 to 1200 base pairs each [45, 48].

Genome invasions

When retroviral infections occur in the germ line of the host, these events lead to retrovirus perpetuating as structural variants in the genome of the host as endogenous retroviruses. Germ line invasion events occur at a low probability. However, for retroviruses, as a whole group, these events occur in an extremely large, virtually infinite, time dimension. As a particular subset, however, evidence indicates that these time periods are restricted. Furthermore, genome integrations are points of no return since endogenous retroviruses cannot be completely eliminated. However, observations are constrained by the power of the observations (approximately 300 million years), selection, host fate [55, 45]. Taken together, these characteristics translate into observing endogenous retroviruses in all organisms analyzed to this date, and there is no reason to believe that this will not be the case in future studies. Moreover, endogenous retroviruses have special properties in relation with other transposable elements, and they are not static unlike other structural variants. It has been observed that gene density and local recombination rates determine fixation and full length persistence of endogenous retroviruses [55]. Endogenous retroviruses can remain active, as is the case of mouse, *Mus musculus*, retrovirus, or become completely inactivated by mutations [34].

Retroviral proliferation mechanisms

Once the have become endogenous, retroviruses can proliferate by two mechanisms: by budding of exogenous particles and infecting neighboring cells, which requires the presence and full functionality of all genes, including *env*; or by intracellular retrotransposition within the germ line, either in trans, where defective retroviruses are complemented by active proteins from other viruses but requires an intact promoter within LTR and other motifs for expression and packing of viral RNA, or in *cis*, where the retrovirus itself supplies all necessary elements for replication [45].

Endogenous retrovirus distribution

Their spread through a variety of hosts and particularly from mammalian Class is just as vast. In brief, two groups dominate ERV abundance: *Gamma*-like

endogenous retroviruses and the most abundant *Beta*-like endogenous retroviruses. Contrary to this, *Epsilon*-like endogenous retroviruses make up almost the entire landscape in fish genomes whereas *Alpha*-like endogenous retroviruses are recovered almost entirely from birds. In addition, it is clear that *Lenti* and *Spuma*-like endogenous retroviruses either invade the germ line much more rarely than other endogenous retroviruses or do not persist long as endogenous retroviruses. Despite this, *Lentivirus*-like endogenous retroviruses have been described in rabbit [56], lemurs [57], and colugo [57]. Estimations about the time of integration can be made either by LTR nucleotide identity or by average pairwise nucleotide identity across the complete sequence. These estimations date the time of integration for endogenous *Lentiviruses* to > 7 million years ago.

2.1.3 On the repurposing of foreign elements

Exaptation

In order to survive to an ever-changing environment, organisms must change or adopt new characteristics. Exaptation, and the related term co-option, describe a shift in the function of a trait during evolution [2]. For example, a particular trait might serve a specific function, but can be repurposed to fulfil another: feather wings initially adapted for thermal regulation. Exaptations are common in both anatomy and behavior. This situation is not unique of biological systems, it is also a common occurrence in science and technology with serendipitous expansion of products into new domains.

Placentation

The placenta is vital organ for the establishment of the materno-fetal interface as it is also the first organ to form in mammals. The formation of a transport interface is the *raison d'etre* of the placenta [58, 59]. The fetus and the placenta are metabolically and physically interconnected through the vascular system [60]. They are both dependent on the maternal circulation. Most amino acids in fetal circulation are at higher concentrations than in maternal circulation, indicating active uptake and synthesis of these nutrients by the placenta or by the fetus. Moreover, both the placenta and some fetal organs are capable of producing and metabolizing various nutrients, which impacts their levels in fetal-placental circulation largely independently of placental transport mechanisms [61]. Beyond vascular supply, the placenta represents a histological complex that fulfills various functions: attachment, invasion, and vascular remodeling to cell fusion, hormone production, and nutrient transport [60]. Interestingly, despite being of fetal origin, placenta does not trigger a foreign-body response by the maternal tissues, suggestive of complex and evolutionarily conserved

immunotolerance mechanisms [59].

This situation represents an alternative to other amniotes where embryos are deposited inside a shell to prevent desiccation and protect against enviromental fluctuations, i.e., reptiles and birds [59, 62]. Fundamental to such design are *syncytins*, a group of coopted retroviral genes with fusogenic and immunomodulatory properties [3], which in contrast to other gene families, did not derive from an ancestral version. Vertebrate lineages, such as primates [3], lagomorphs [4], rodents [63], ruminants [7], carnivores [6, 11], marsupials [64, 9] and even some lizards [10, 12], have exapted retroviral genes on independent occasions. Not all these events corresponded with the exact same genetic profile, but rather a cluster of distinct physiological mechanisms involving retroviral *env*, and thus resulting in morphological and physiological diverse placental types [59].

The fusiogenic function in *syncytin*, originally responsible for joining the cellular membrane with viral membrane during virus entry, serves a pivotal role during the formation of the materno-fetal interface for metabolite exchange. Aside from membrane fusion, retroviral *env* gene also harbors immunomodulatory domains that serve during cell invasion. These domains are believed to play an important role on the early tolerance of the syncytiotrophoblast by the mother [65]. Thus, such mechanisms might play an important role on placental and obstetric pathologies [65].

The architecture of the primordial placenta might be difficult to determine, however it has been proposed that such structure was hemochorial, discoid, and labyrinthine [66], resembling current murine placentas. It is also hypothesized that a new retroviral insertion that offers advantageous traits gradually replaces a former version [67].

2.2 Translational medicine

Translational medicine refers to the application of basic science into clinical medicine [68]. At its core, translational medicine reflects the combination of disciplines, resources, expertise, and techniques. Thus, translational medicine initiative is fundamental for the rapid utilization of scientific knwoledge and emerging technologies to improve public health, in constrast to the more traditional academic health science research.

2.2.1 Epilepsy

Epilepsy, a chronic central neurological disease affects individuals of all ages and has a worldwide distribution. It is characterized by abnormal or synchronous electrocortical discharges, clinically identified as epileptic seizures [69]. The range of pattern varies as a function of the state and conditions of the patient as well as the topographical localization of the ethological focus and variant of the disease. Abnormal patterns, such as difusse slowing, focal slowing, focal attenuation, of the normal patters are considered pathological [70, 71, 72].

Moreover, other morbidities other primary epilepsy can create epileptiform discharges, for examples: dementias, ischemic and hemorragic stroke, subdural hematoma, metabolic disorders, coma, *status epilepticus*, brain death, and even neurotropic drugs [73, 74].

2.2.2 Electroencephalography

Biomedical signal measurement is a pivotal resource for assessment of the patient well-being. Increase availability in computational resources and advances in automation and artificial intelligence hold enormous promises for improvement of the patient well-being [75, 76, 77, 78]. Electroencephalogram (EEG) is a cornerstone in the assessment, treatment and prognosis of neurological conditions. EEG recording, is a graphic portrayal of the difference in voltage between two different cerebral regions plotted over time [79].

Electroencephalogram captures the electrical activity occurring at the cerebral cortex. This electrical activity is measured in microvolts and further amplified by a factor of 1 million in order to be displayed. During this amplification, the difference between two recorded points is measured. Large part of the measurement originates from neural physiology, possibly action potentials, post-synaptic potentials and chronic neuronal depolarization. However, action potentials are remarkably short, occurring in the range of tens of miliseconds or less, in contrast to post-synaptic potentials, which are much longer at 50-200 miliseconds. This feature makes post-synaptic potentials

larger contributors to the signals capture on electroencephalographic recordings. Thus, electroencephalogram is essentially measuring voltage change in the extracellular matrix. Furthermore, excitatory post-synaptic potentials are characterized by depolarization, whereas inhibitory post-synaptic potentials create hyperpolarization. Topographically, the two large contributors to electroencephalographic rhythm are the interaction between the thalamus and the cortex, and the functional properties of large neuronal networks in the cortex. However, the fact that electroencephalogram is essentially measured by amplification of miniscule electrical currents into a graphical representation that can be interpreted, introduces naturally a substantial amount of methodological noise. Besides, sources of electrical interference other than electrocortical potentials are also amplified and displayed in the output. Importantly, any signal that affects both sources equally will be cancelled or severely reduced, i.e., in-phase cancellation [79, 80].

Electrode placement is standardized to the 10-20 International System of Electrode Placement and 10-10 system, where electrodes are placed using distinctive anatomical landmarks on three planes - sagital, coronal, and horizontal - to create an accurate topographical map of the cerebral cortex. Electrodes with odd numbers are placed on the left and electrodes with even numbers are on the right In this manner, electrical spikes propagate through the cerebral cortex as ripples most often in eliptical shape [79].

Electrocortical patterns are diverse since biological variation is substantial, which also depends on intrinsic factors from the patient, such age, neurological activity, state, comorbidities, among others. Electrical patterns that are considered physiological are named: alpha wave, beta wave, theta wave, delta wave. Moreover, due to the fact that capturing pathological events on recordings, activation procedures are routinely performed during: hyperventilation, photostimulation, and sleep deprivation [81, 82].

Part II:
Results

When we push the limits to the edge, it is then, when marvelous things can happen.

2.3 Paper I

Results

Retroviruses integrate into the genome of their hosts as part of their life cycle. When such event takes place in the germ line, retroviruses can be passed on to the host offspring Mendelian genes, known as endogenous retroviruses. Advances in genomic sequencing technologies have permitted to obtain better sampling in terms of number of individuals as well as representation of each haplotype. However, most of this information comes in the form of short-read whole-genome sequencing. Thus, identification of structural variants remains challenging.

To address this problem I designed a methodology that could identify endogenous retroviral insertions from short-read sequencing data, as illustrated in **Figure 2.2**. We used pooled whole-genome sequencing data from wild and domestic European rabbit populations. We constructed a library of endogenous retroviruses based on identification from the reference genome of *Oryctulagus cuniculus*, **oryCun2**, and enriched such library with other endogenous retroviral sequences. We aligned the whole-genome sequenced short-reads to our library of endogenous retroviruses to obtain sequence similarity to en-

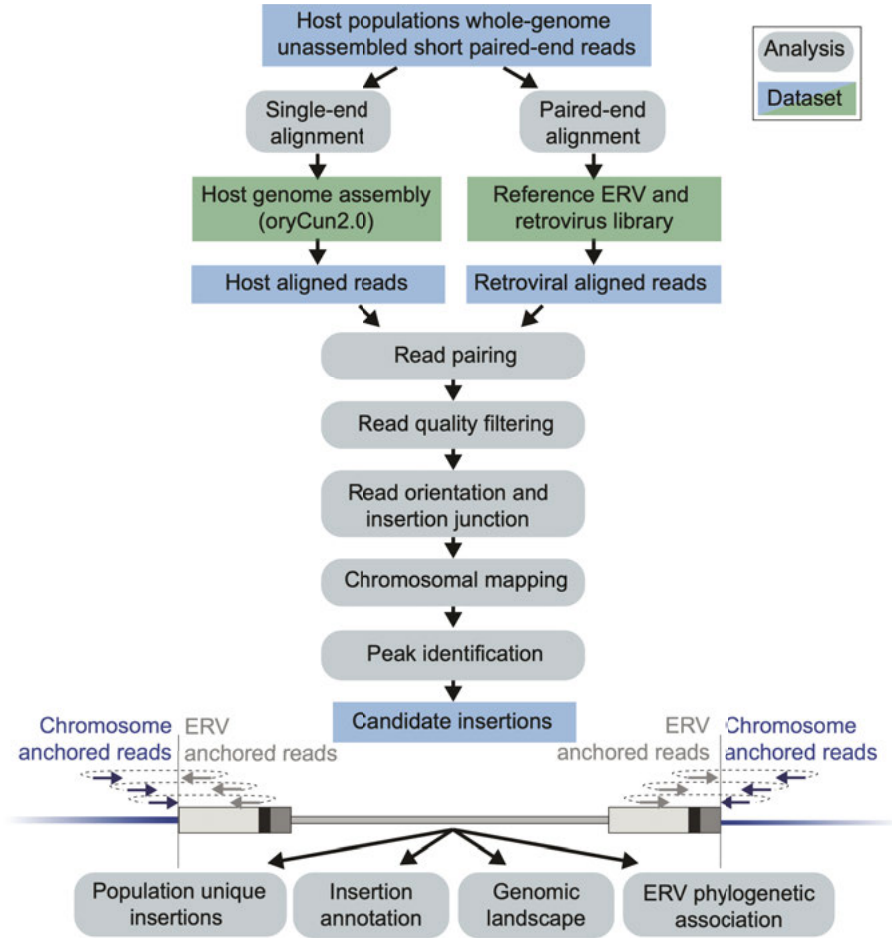

Figure 2.2. ERV identification in unassembled paired-end short-read sequence libraries. Computational strategy for detecting reference and non reference assembly endogenous retroviruses utilizing unassembled paired-end short-read sequence libraries. The output presents a list of candidate insertions (blue data box) anchored to chromosomal positions along host DNA to which downstream analyses can be applied.

dogenous retroviral potential insertions. In parallel, I aligned whole-genome sequenced short-reads to the reference genome, oryCun2, to obtain chromosomal positions. We used the alignment information, i.e. directionality and orientation, of both read pairs to target retroviral long terminal repeat sequences and insertion points along the genome. Next, assuming that sequencing reads would be Poisson distributed along the genome, I selected a threshold based on the likelyhood of finding reads whose pair had anchored to a retroviral long terminal repeat and sharing directionality and orientation in a genomic neighborhood. With this method I identified segregation of endogenous retroviruses

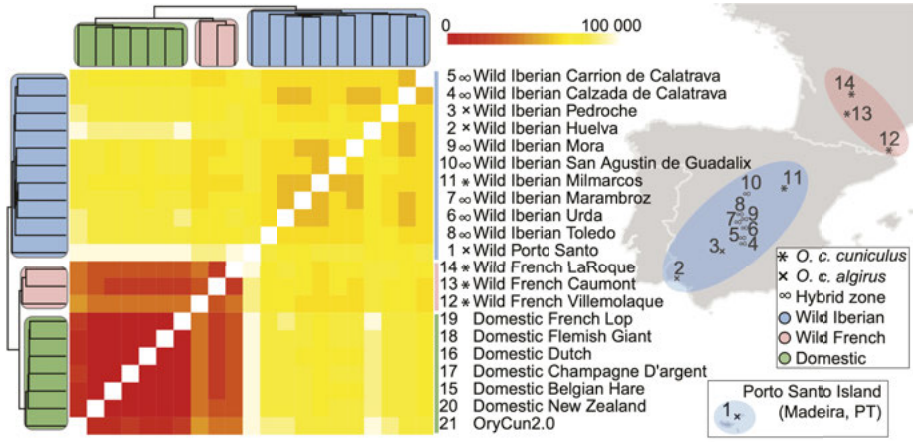

Figure 2.3. Oryctulagus cuniculus sp. geographical distribution and endogenous retroviral segregation profiles. Hierarchical clustering and heatmap illustrates presence or absence of endogenous retroviral insertion junctions among *O. cuniculus sp.*, from wild populations across the Iberian Peninsula, southern France, and the Porto Santo Island, as well as from domestic rabbits. The heatmap and cladogram show pairwise differences among individual groups: domestic (green), wild French (red), and wild Iberian (blue). The two subspecies *O. c. cuniculus* and *O. c. algirus* are indicated, as well as rabbit populations sampled from a hybrid zone on the Iberian peninsula.

on wild and domestic populations of European rabbit, as well as identified different levels of allele frequencies on retroviral insertions, even not present in the reference genome, as shown in **Figure 2.3** and **Figure 2.4**. I used phylogenetic association from endogenous retroviral sequences on our library to make inferences about insertions identified in wild and domestic European rabbit populations, as displayed in **Figure 2.5**.

In this study, I showed that genetic structural variation extends beyond the reference genome. Moreover, the study addresses pertinent issues regarding genome sampling, interpretation of genetic diversity beyond single nucleotide variants and population comparative genomics.

Contribution

For this project, I constructed the computational analysis pipeline. I also analyzed all the genomic samples, innovated on the data interpretation and methods for efficient computation. I created the all figures and wrote the initial draft.

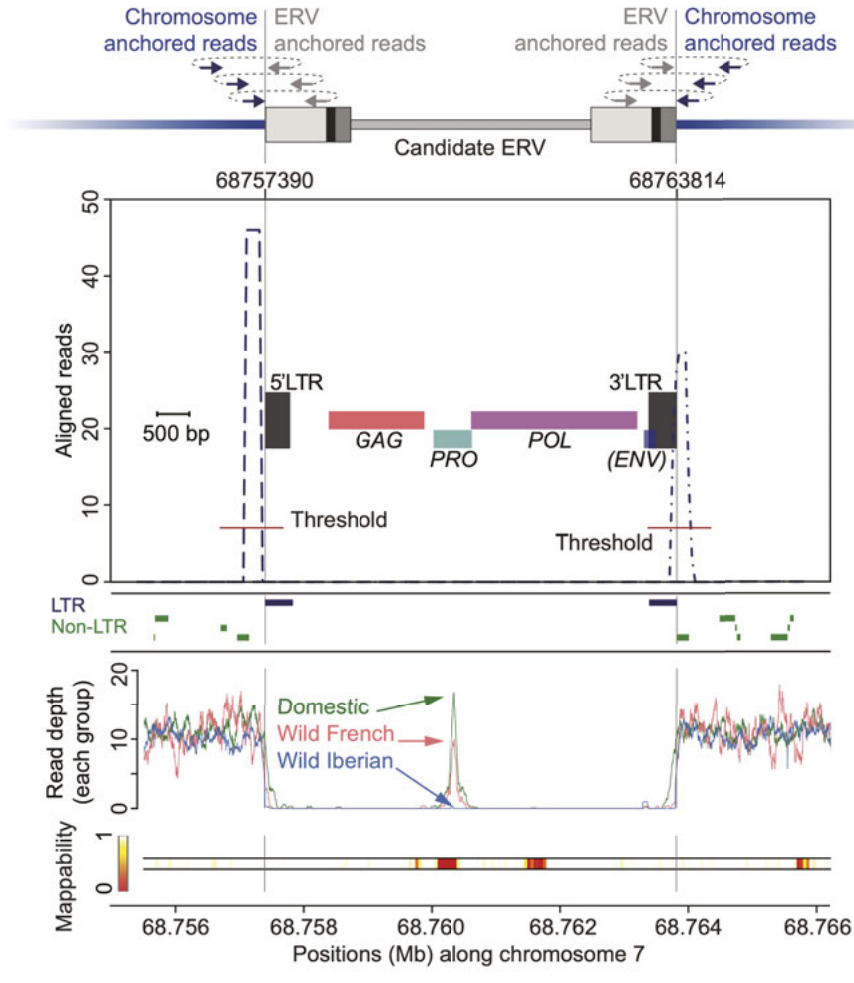

Figure 2.4. Candidate ERV. (Top) A schematic shows identification strategy and chromosomal anchoring of endogenous retrovirus-associated sequence read pairs, which demonstrates marked peaks in sliding window counts with their independent dynamic thresholds immediately adjacent to the endogenous retrovirus insertion junctions. The present locus was also identified in the reference rabbit genome assembly (**oryCun2.0**) using RetroTector, which shows the predicted gene structure and confirms a truncated endogenous retrovirus. The RepeatMasker track for LTRs (blue) and non-LTRs (green) lines up with our identification as well as RetroTector prediction. (Bottom) Average read depth along the host DNA as well as the mappability are summarized. Decreased sequence read depth across the locus is because of poor locus-specific mapping due to similar endogenous retroviruss elsewhere in the genome, whereas the peak in the middle of the locus indicates a short segment permitting locus-specific sequence read pair mapping, which in this case also allow estimation of segregation at this locus among the domestic, wild French, and wild Iberian rabbit populations (indicated by arrows).

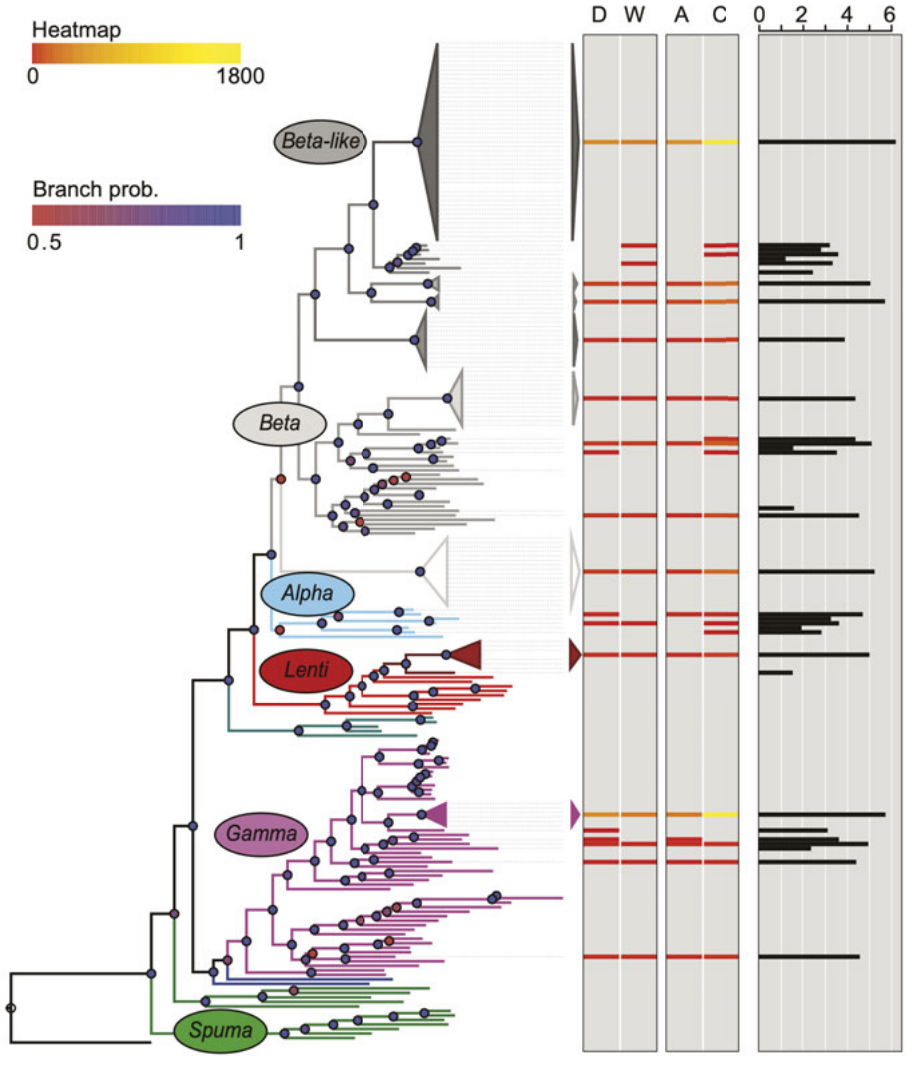

Figure 2.5. Endogenous retrovirus phylogenetic association and spread among *O. cuniculus sp.*. Retroviral *gag*- and *pol*-based tree derived from reference retroviruses and endogenous retroviruses where colors indicate retrovirus-like nomenclature as previously described. Dashed lines indicate endogenous retroviral sequences identified in the rabbit reference genome assembly, **oryCun2.0**. The heatmap shows number of endogenous retroviral insertions junctions in pairwise comparisons between domestic rabbit populations (D) and wild French rabbit populations (W), as well as between wild *O. c. algirus* populations (A) and wild *O. c. cuniculus* populations (C). The overall presence of candidate endogenous retroviral insertion junctions across all *O. cuniculus sp.* populations is represented by the black histogram using log 10 scale.

2.4 Paper II and III

Results

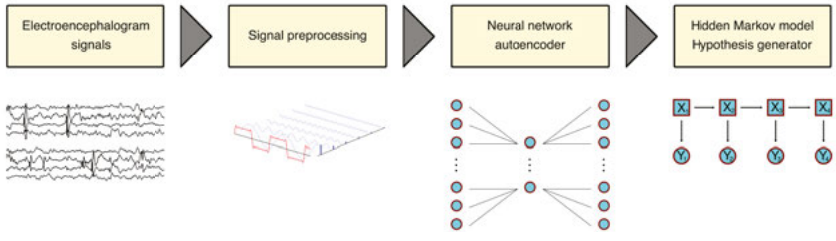

Figure 2.6. MindReader algorithm. From left to right, electroencephalographic (EEG) signals are loaded from standard European Data Format (EDF) files, then pre-processed by Fast Fourier Transform (FFT) and binned on overlapping windows. Next, signals are input to a neural network autoencoder where the autoencoder error is calculated, that is the difference between the input and the autoencoder model post training is obtained, and considered anomaly. Finally, the signal is input to a Hidden Markov Model and Hypothesis Generator where states are assumed and labels are assigned. Importantly, the entire process is unsupervised and each channel is processed independently.

Assement of patient status is crucial for medical attention. Thus, measurement of biomedical signals is critical. However, interpretation of some biomedical signals, i.e., electroencephalogram, requires highly specialized training. In addition, the rate of capturing events renders the interpretation time-consuming, resource-hungry and, overall, an expensive process in terms of human resources. These characteristics mean that interpretation is only available at specialized centers, and continuum interpretation is virtually unfeasible. In contrast, automatic detection offers potential to improve quality of patient care by shortening the assessment time, thus reducing time to diagnosis, better management of big and confidential data and optimization on the allocation of human resources.

We developed **MindReader**, an unsupervised method for biomedical signal interpretation, whose methodological approach is displayed in **Figure 2.6**. **MindReader** preprocess the stream of data by using binning and Fast Fourier Transform (FFT). Then, a neural network autoencoder is built where each neuron represents a time point in the bin. The hidden layer is scale down in relation to the input and expanded to the original size of the output on the

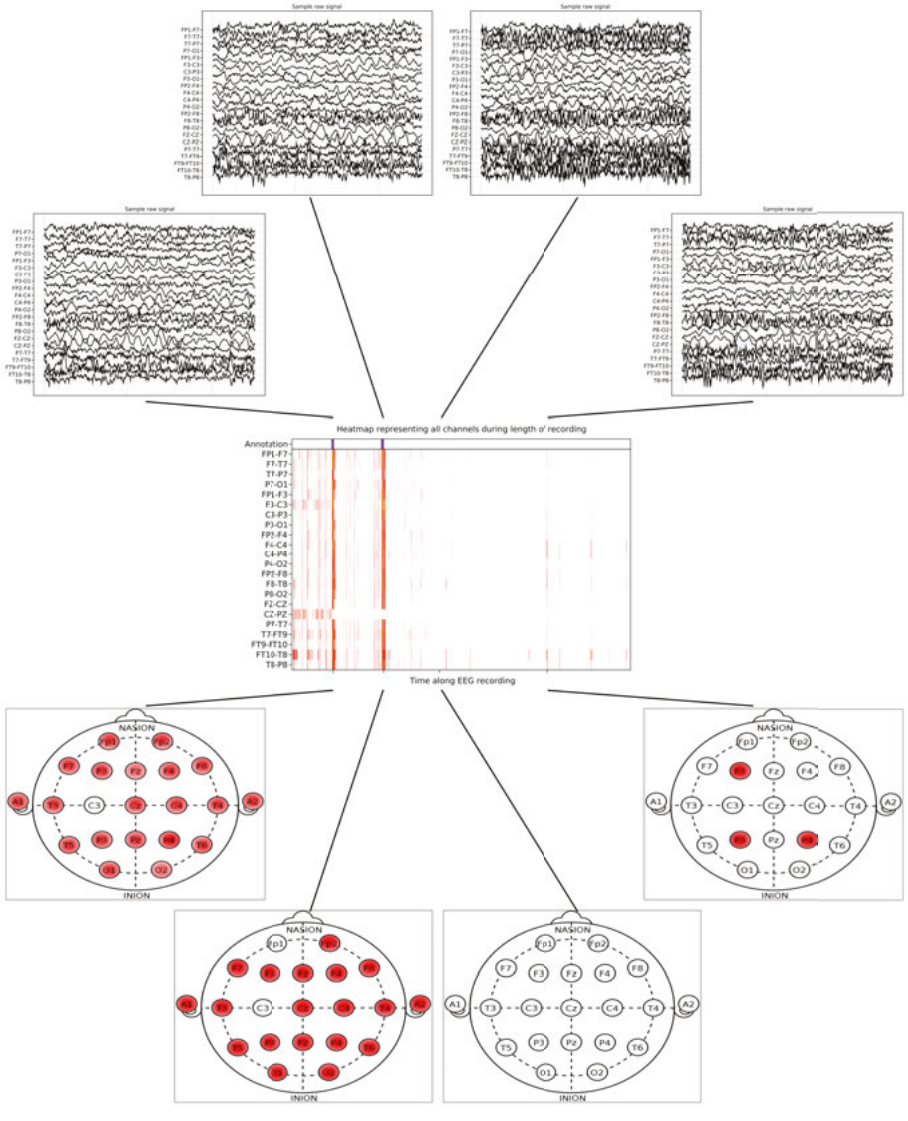

Figure 2.7. MindReader sample recording interpretation on subject chb04 (male 22 years old) record 28. Top plots illustrate MindReaderś output on EEG montage at four different time point during the recording. Middle heatmap shows MindReaderś hypothesized states along the recording per channel. Original Physionet manual annotation is indicated on top. Bottom plots display original EEG signals from same time points as interpretations.

subsequent layer. Once the neural network is trained, the difference between the input and the output (input by the model) is calculated, i.e., the autoencoder error. Assuming that a neural network autoencoder would compress the input information, any other data present outside this domain could be

considered an anomaly. Next, **MindReader** uses a dependency library, **HiddenMarkovModelReaders**. This dependency consists of a hidden Markov model to stream the autoencoder error and generate hypothesis when a distinctive states / patterns are identified. Importantly, each channel / electrode is process independently and the generation of hypothesis is unsupervised. The output from **MindReader** is illustrated in **Figure 2.7**

Contribution

During these two projects I participated in the conceptualization and design of the study, and established the collaboration with Instituto Nacional de Neurologia y Neurocirugia (INNN), in Mexico city, Mexico. I developed and deployed the computational methods. I also tested and verified the computational methods' internal validity, and performed the all analyses. I created the all figures and wrote the initial draft. I edited subsequent draft versions and corresponded with the journal editors.

2.5 Paper IV

Results

Syncytin is a fossil protein exapted from retroviruses that fulfills a pivotal role during trophoblast implantation and placental metabolite exchange. Profiting from technological advances on genomic sequencing and assembly protocols, we searched for *syncytin* insertions across vertebrates. We compiled a library of known *syncytin* sequences. Then, we searched a library of more than 150 chromosome-lenght assemblies across 17 taxonomical orders for *syncytin* homologues. We identified and syntenically aligned over 300 loci insertions, including not previously described, as displayed in **Figure 2.8**. Additionally, we predicted the protein tridimensional structures of database *syncytin* sequences, as illustrated in **Figure 2.9**.

Contribution

For this project I conceived the initial idea, and successfully applied for the Matariki fellowship funding. This helped cement the international collaboration with the DNA zoo consortium, an international conglomerate purposed with assembling genomes. I reached out to experts on the protein structure. I performed most analyses, but protein prediction, and created most figures as well as wrote the initial draft.

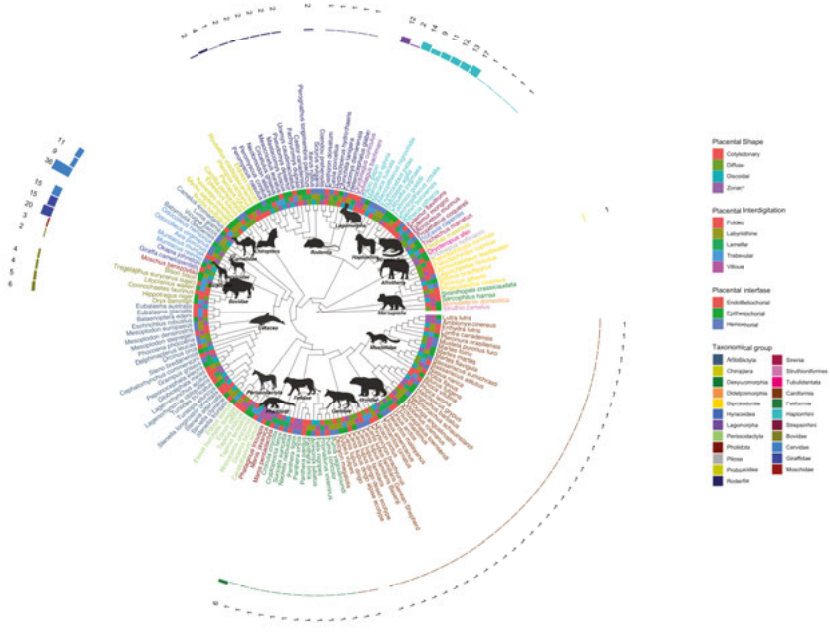

Figure 2.8. Broad screening for *syncytin* sequences. Inner circle depicts phylogenetic tree of genome assemblies searched comprising 17 taxonomical orders. Outer tiles indicate characteristics of the placenta per species: from inner to outer, gross placental shape, materno-fetal interdigitation and histological placental interfase. Outer annotations indicate number of loci found on each genome assembly.

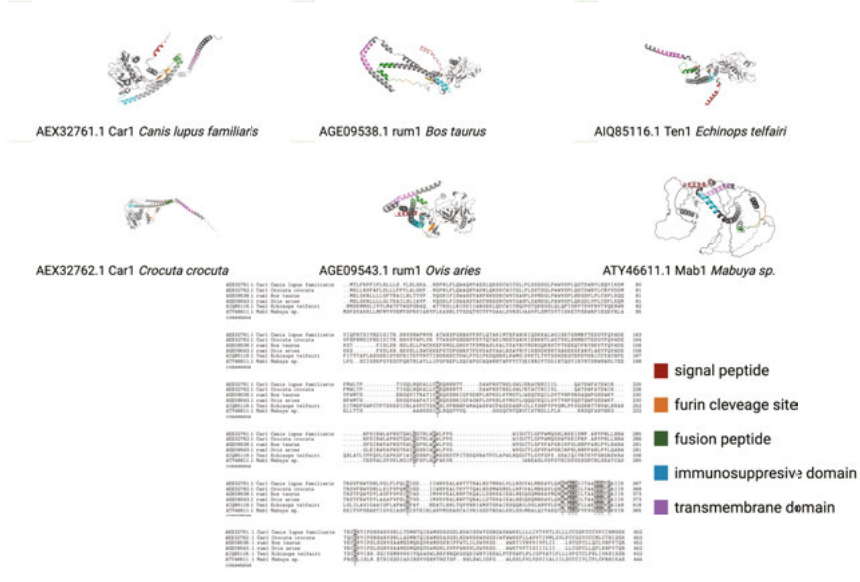

Figure 2.9. Tridimensional protein prediction of reference *syncytin* sequences. Upper panel illustrates protein prediction of six reference *syncytin* sequences, whereas lower panel show multiple sequence alignments. Protein domains are annotated by colors.

Part III:
Conclusions

Evolutionary insights

Genomes are not simply static catalogues of genes. Instead, they are ever changing, genetically dynamic entities.

Evolution is a perpetual race for entities to adapt to their environment or ultimately perish. Dismantling the intricacies of how this process takes place is of major importance. One vital aspect of evolution in biological systems is the insertion, tolerance, and adoption of foreign elements. Genetic transposable elements represent a substantial proportion of eukaryotic genomes, where they can disrupt or enhance gene expression on the host. Although advances in sequencing and computational analyses have facilitated their characterization, further studies are necessary for deciphering the evolution dynamics among foreign agents and their hosts. Moreover, sampling effects during genetic studies from different host populations present major challenges.

During this thesis I briefly addressed two aspects related to evolution:

- The insertion of transposable elements, specifically endogenous retroviruses, into host genomes and their segregation pattern in populations as structural variants.

- The adoption and repurposing of a foreign information element into a host genome and its variational pattern across vertebrates.

I believe that transposable elements as a group present a fascinating set of features in the context of constrained systems; where constrains in different species are defined by their genomic dynamics. Thus, the evolutionary benefit against cost of harboring or purging foreign elements must be higher. In others words, tolerating additional and different information set in terms of proteins, gene expression controllers and tridimensional modelers. If a group of foreign elements was successful to insert early in the evolutionary history of an also successful species, the host, the possibility for said element group to propagate and persist increases exponentially given that small differences during initial phases translate to galloping differences over time. However, if a feature or design is best adapted for a certain condition, such feature will probably converge again. And clearly, much is yet to be uncovered and investigated in this field.

Scientific pragmatism

Every discovery, every innovation, everything we have in our lives rests on the shoulders of great people.

The scientific method is the key for generation of new knowledge. However, as interesting as diving into the whys and hows of a phenom we feel passionate about is, society demands practical applications to address problems. With this concept always present, I worked during these studies to participate in the generation of knowledge that could target better health care conditions and improved quality of life. A road, probably out of many, to achieve such objective is to profit from automation and data collection, in other words, to use data-driven machine learning to aid health care workers in different areas. For this purpose, I created a set of computational tools that integrate into a machine learning unsupervised methodology for pattern recognition of biomedical signals with temporal dimension. I centered this methodology on electroencephalographic signals, but given its containerized and portable design it is potential extendable and applicable to other similar areas where accurate and rapid, even live, identification is required.

Future research

The function and duty of a quality human being is the sincere and honest development of one's potential.

The importance and potential power of cross-disciplinary projects cannot be undervalued. During these studies I used techniques from different domains and collaborated with a wide sprectrum of scientists. This has led me to understand that the future of scientific advancement greatly lies on collaborative efforts. Moreover, a holistic approach in terms of knowledge is just as necessary. Thus, integration of different biological organisation levels could render possible the future goals of precision medicine.

In the case of structural variation and transposable element composition of a single individual, characterization of the genomic profile beyond single-nucleotide variants would make possible the identification of genomic rearrangements. This is important in several clincal scenarios considering that it is vital for treatment and prognosis:

- In newborns and oncological patients, to assess their carotype for genomic translocations, inversions, duplications, etc., and their penetration to certains tissues or cell populations.
- In the context of cancer, to identify transposable elements, considering that certain therapies focus on modification of methylation profiles, which could allow transposable elements to reactivate.

In the case of the study of placentation, deeper understanding of the physiology of immunotolerance of two immunologically different organisms could lead to, for example:

- Applications on xenotransplantation, such as bio-implants that would benefit from a placental-like metabolic exchange systems.
- Direct application to prevention and treatment of gineco-obstetric diseases, such as preclampsia / eclampsia, a life treating pathology for the mother and the product.

In the case of biomedical signals, potential expansion to other domains where automation and integration to other measurements would greatly benefit the quality of care. For example:

- Automatic non-invasive and continuous measurement of cardic electrical signals to predict seric metabolic profile.
- Continous evaluation of muscle and nervous performance for rehabilitation therapies.

Acknowledgments

Farthest from your mind is the thought of falling back, in fact, it isn't there at all. And so you dig your hole carefully and deep, and wait...

First and foremost, I would like to acknowledge all people whose hard work finances the contribution to public funding agencies because I believe it is them I work for.

Second, I would like to dedicate this work to people who do not have the fortune to access basic human rights, particularly public education. I believe that it is thanks to them that I have the privilege to be able to do what I do, subsequently I owe an obligation to them to do my work to the best of my capabilities.

I would also like to acknowledge the administration personel at our Uppsala University who do their best to keep everything logisticaly functional.

Likewise, I would like to thank all my supervisors, advisors and collaborators who at some point or another were involved in some of these projects. Especially, my now good friends Manfred and Jan, whose support and guidance has been fundamental not only during this academic journey, but also in life.

Finally, and on a personal note, I would like to acknowledge my family and friends for their role in this endeavor.

References

[1] Jinchuan Xing, Yuhua Zhang, Kyudong Han, Abdel Halim Salem, Shurjo K. Sen, Chad D. Huff, Qiong Zhou, Ewen F. Kirkness, Samuel Levy, Mark A. Batzer, and Lynn B. Jorde. Mobile elements create structural variation: Analysis of a complete human genome. *Genome Research*, 19(9):1516–1526, may 2009.

[2] Pierre Capy. Taming, domestication and exaptation: Trajectories of transposable elements in genomes. *Cells*, 10(12), dec 2021.

[3] Sha Mi, Xinhua Lee, Xiang ping Li, Geertruida M. Veldman, Heather Finnerty, Lisa Racie, Edward LaVallie, Xiang-Yang Tang, Philippe Edouard, Steve Howes, James C. Keith, and John M. McCoy. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature*, 403(6771):785–789, feb 2000.

[4] Odile Heidmann, Cécile Vernochet, Anne Dupressoir, and Thierry Heidmann. Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new "syncytin" in a third order of mammals. *Retrovirology*, 6(1), nov 2009.

[5] Christian Lavialle, Guillaume Cornelis, Anne Dupressoir, Cécile Esnault, Odile Heidmann, Cécile Vernochet, and Thierry Heidmann. Paleovirology of syncytins, retroviral env genes exapted for a role in placentation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1626):20120507, 2013.

[6] Guillaume Cornelis, Odile Heidmann, Sibylle Bernard-Stoecklin, Karine Reynaud, Géraldine Véron, Baptiste Mulot, Anne Dupressoir, and Thierry Heidmann. Ancestral capture of syncytin-car1, a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in carnivora. *Proceedings of the National Academy of Sciences*, 109(7), jan 2012.

[7] Guillaume Cornelis, Odile Heidmann, Séverine A. Degrelle, Cécile Vernochet, Christian Lavialle, Claire Letzelter, Sibylle Bernard-Stoecklin, Alexandre Hassanin, Baptiste Mulot, Michel Guillomot, Isabelle Hue, Thierry Heidmann, and Anne Dupressoir. Captured retroviral envelope syncytin gene associated with the unique placental structure of higher ruminants. *Proceedings of the National Academy of Sciences*, 110(9), feb 2013.

[8] Guillaume Cornelis, Cécile Vernochet, Sébastien Malicorne, Sylvie Souquere, Athanasia C. Tzika, Steven M. Goodman, François Catzeflis, Terence J. Robinson, Michel C. Milinkovitch, Gérard Pierron, Odile Heidmann, Anne Dupressoir, and Thierry Heidmann. Retroviral envelope syncytin capture in an ancestrally diverged mammalian clade for placentation in the primitive afrotherian tenrecs. *Proceedings of the National Academy of Sciences*, 111(41), sep 2014.

[9] Guillaume Cornelis, Cécile Vernochet, Quentin Carradec, Sylvie Souquere, Baptiste Mulot, François Catzeflis, Maria A. Nilsson, Brandon R. Menzies,

Marilyn B. Renfree, Gérard Pierron, Ulrich Zeller, Odile Heidmann, Anne Dupressoir, and Thierry Heidmann. Retroviral envelope gene captures and syncytin exaptation for placentation in marsupials. *Proceedings of the National Academy of Sciences*, 112(5), jan 2015.

[10] Guillaume Cornelis, Mathis Funk, Cécile Vernochet, Francisca Leal, Oscar Alejandro Tarazona, Guillaume Meurice, Odile Heidmann, Anne Dupressoir, Aurélien Miralles, Martha Patricia Ramirez-Pinilla, and Thierry Heidmann. An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental mabuya lizard. *Proceedings of the National Academy of Sciences*, 114(51), nov 2017.

[11] Mathis Funk, Guillaume Cornelis, Cécile Vernochet, Odile Heidmann, Anne Dupressoir, Alan Conley, Stephen Glickman, and Thierry Heidmann. Capture of a hyena-specific retroviral envelope gene with placental expression associated in evolution with the unique emergence among carnivorans of hemochorial placentation in hyaenidae. *Journal of Virology*, 93(4), feb 2019.

[12] Hans Recknagel, Madeleine Carruthers, Andrey A. Yurchenko, Mohsen Nokhbatolfoghahai, Nicholas A. Kamenos, Maureen M. Bain, and Kathryn R. Elmer. The functional genetic architecture of egg-laying and live-bearing reproduction in common lizards. *Nature Ecology & Evolution*, 5(11):1546–1556, oct 2021.

[13] Anne Dupressoir, Geoffroy Marceau, Cécile Vernochet, Laurence Bénit, Colette Kanellopoulos, Vincent Sapin, and Thierry Heidmann. Syncytin-a and syncytin-b, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in muridae. *Proceedings of the National Academy of Sciences*, 102(3):725–730, jan 2005.

[14] Eric S. Lander, Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William FitzHugh, Roel Funke, Diane Gage, Katrina Harris, Andrew Heaford, John Howland, Lisa Kann, Jessica Lehoczky, Rosie LeVine, Paul McEwan, Kevin McKernan, James Meldrim, Jill P. Mesirov, Cher Miranda, William Morris, Jerome Naylor, Christina Raymond, Mark Rosetti, Ralph Santos, Andrew Sheridan, Carrie Sougnez, Nicole Stange-Thomann, Nikola Stojanovic, Aravind Subramanian, Dudley Wyman, Jane Rogers, John Sulston, Rachael Ainscough, Stephan Beck, David Bentley, John Burton, Christopher Clee, Nigel Carter, Alan Coulson, Rebecca Deadman, Panos Deloukas, Andrew Dunham, Ian Dunham, Richard Durbin, Lisa French, Darren Grafham, Simon Gregory, Tim Hubbard, Sean Humphray, Adrienne Hunt, Matthew Jones, Christine Lloyd, Amanda McMurray, Lucy Matthews, Simon Mercer, Sarah Milne, James C. Mullikin, Andrew Mungall, Robert Plumb, Mark Ross, Ratna Shownkeen, Sarah Sims, Robert H. Waterston, Richard K. Wilson, LaDeana W. Hillier, John D. McPherson, Marco A. Marra, Elaine R. Mardis, Lucinda A. Fulton, Asif T. Chinwalla, Kymberlie H. Pepin, Warren R. Gish, Stephanie L. Chissoe, Michael C. Wendl, Kim D. Delehaunty, Tracie L. Miner, Andrew Delehaunty, Jason B. Kramer, Lisa L. Cook, Robert S. Fulton, Douglas L. Johnson, Patrick J. Minx, Sandra W. Clifton, Trevor Hawkins, Elbert Branscomb, Paul Predki, Paul Richardson, Sarah Wenning, Tom Slezak, Norman Doggett, Jan-Fang Cheng, Anne Olsen, Susan Lucas, Christopher Elkin, Edward Uberbacher, Marvin

Frazier, Richard A. Gibbs, Donna M. Muzny, Steven E. Scherer, John B. Bouck, Erica J. Sodergren, Kim C. Worley, Catherine M. Rives, James H. Gorrell, Michael L. Metzker, Susan L. Naylor, Raju S. Kucherlapati, David L. Nelson, George M. Weinstock, Yoshiyuki Sakaki, Asao Fujiyama, Masahira Hattori, Tetsushi Yada, Atsushi Toyoda, Takehiko Itoh, Chiharu Kawagoe, Hidemi Watanabe, Yasushi Totoki, Todd Taylor, Jean Weissenbach, Roland Heilig, William Saurin, Francois Artiguenave, Philippe Brottier, Thomas Bruls, Eric Pelletier, Catherine Robert, Patrick Wincker, André Rosenthal, Matthias Platzer, Gerald Nyakatura, Stefan Taudien, Andreas Rump, Douglas R. Smith, Lynn Doucette-Stamm, Marc Rubenfield, Keith Weinstock, Hong Mei Lee, JoAnn Dubois, Huanming Yang, Jun Yu, Jian Wang, Guyang Huang, Jun Gu, Leroy Hood, Lee Rowen, Anup Madan, Shizen Qin, Ronald W. Davis, Nancy A. Federspiel, A. Pia Abola, Michael J. Proctor, Bruce A. Roe, Feng Chen, Huaqin Pan, Juliane Ramser, Hans Lehrach, Richard Reinhardt, W. Richard McCombie, Melissa de la Bastide, Neilay Dedhia, Helmut BlÃ¶cker, Klaus Hornischer, Gabriele Nordsiek, Richa Agarwala, L. Aravind, Jeffrey A. Bailey, Alex Bateman, Serafim Batzoglou, Ewan Birney, Peer Bork, Daniel G. Brown, Christopher B. Burge, Lorenzo Cerutti, Hsiu-Chuan Chen, Deanna Church, Michele Clamp, Richard R. Copley, Tobias Doerks, Sean R. Eddy, Evan E. Eichler, Terrence S. Furey, James Galagan, James G. R. Gilbert, Cyrus Harmon, Yoshihide Hayashizaki, David Haussler, Henning Hermjakob, Karsten Hokamp, Wonhee Jang, L. Steven Johnson, Thomas A. Jones, Simon Kasif, Arek Kaspryzk, Scot Kennedy, W. James Kent, Paul Kitts, Eugene V. Koonin, Ian Korf, David Kulp, Doron Lancet, Todd M. Lowe, Aoife McLysaght, Tarjei Mikkelsen, John V. Moran, Nicola Mulder, Victor J. Pollara, Chris P. Ponting, Greg Schuler, JÃ¶rg Schultz, Guy Slater, Arian F. A. Smit, Elia Stupka, Joseph Szustakowki, Danielle Thierry-Mieg, Jean Thierry-Mieg, Lukas Wagner, John Wallis, Raymond Wheeler, Alan Williams, Yuri I. Wolf, Kenneth H. Wolfe, Shiaw-Pyng Yang, Ru-Fang Yeh, Francis Collins, Mark S. Guyer, Jane Peterson, Adam Felsenfeld, Kris A. Wetterstrand, Richard M. Myers, Jeremy Schmutz, Mark Dickson, Jane Grimwood, David R. Cox, Maynard V. Olson, Rajinder Kaul, Christopher Raymond, Nobuyoshi Shimizu, Kazuhiko Kawasaki, Shinsei Minoshima, Glen A. Evans, Maria Athanasiou, Roger Schultz, Aristides Patrinos, and Michael J. Morgan. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, feb 2001.

[15] A. P. Jason de Koning, Wanjun Gu, Todd A. Castoe, Mark A. Batzer, and David D. Pollock. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genetics*, 7(12):e1002384, dec 2011.

[16] Sandra R. Richardson, Aurélien J. Doucet, Huira C. Kopera, John B. Moldovan, José Luis Garcia-Perez, and John V. Moran. The influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Microbiology Spectrum*, 3(2), apr 2015.

[17] Rachel L Cosby, Ni-Chen Chang, and Cédric Feschotte. Host–transposon interactions: conflict, cooperation, and cooption. *Genes & Development*, 33(17-18):1098–1116, sep 2019.

[18] Susanne N. GrÃ¶bner, , Barbara C. Worst, Joachim Weischenfeldt, Ivo Buchhalter, Kortine Kleinheinz, Vasilisa A. Rudneva, Pascal D. Johann,

Gnana Prakash Balasubramanian, Maia Segura-Wang, Sebastian Brabetz, Sebastian Bender, Barbara Hutter, Dominik Sturm, Elke Pfaff, Daniel HÃ¼bschmann, Gideon Zipprich, Michael Heinold, JÃ¼rgen Eils, Christian Lawerenz, Serap Erkek, Sander Lambo, Sebastian Waszak, Claudia Blattmann, Arndt Borkhardt, Michaela Kuhlen, Angelika Eggert, Simone Fulda, Manfred Gessler, Jenny Wegert, Roland Kappler, Daniel Baumhoer, Stefan Burdach, Renate Kirschner-Schwabe, Udo Kontny, Andreas E. Kulozik, Dietmar Lohmann, Simone Hettmer, Cornelia Eckert, Stefan Bielack, Michaela Nathrath, Charlotte Niemeyer, GÃ¼nther H. Richter, Johannes Schulte, Reiner Siebert, Frank Westermann, Jan J. Molenaar, Gilles Vassal, Hendrik Witt, Birgit Burkhardt, Christian P. Kratz, Olaf Witt, Cornelis M. van Tilburg, Christof M. Kramm, Gudrun Fleischhack, Uta Dirksen, Stefan Rutkowski, Michael FrÃ¼hwald, Katja von Hoff, Stephan Wolf, Thomas Klingebiel, Ewa Koscielniak, Pablo Landgraf, Jan Koster, Adam C. Resnick, Jinghui Zhang, Yanling Liu, Xin Zhou, Angela J. Waanders, Danny A. Zwijnenburg, Pichai Raman, Benedikt Brors, Ursula D. Weber, Paul A. Northcott, Kristian W. Pajtler, Marcel Kool, Rosario M. Piro, Jan O. Korbel, Matthias Schlesner, Roland Eils, David T. W. Jones, Peter Lichter, Lukas Chavez, Marc Zapatka, and Stefan M. Pfister and. The landscape of genomic alterations across childhood cancers. *Nature*, 555(7696):321–327, feb 2018.

[19] Sophia Groh and Gunnar Schotta. Silencing of endogenous retroviruses by heterochromatin. *Cellular and Molecular Life Sciences*, 74(11):2055–2065, feb 2017.

[20] Melissa A. Wilson Sayres, Chris Venditti, Mark Pagel, and Kateryna D. Makova. Do variations in substitution rates and male mutation bias correlate with life-history traits? a study of 32 mammalian genomes. *Evolution*, 65(10):2800–2815, jun 2011.

[21] Jonas Blomberg, Farid Benachenhou, Vidar Blikstad, GÃ¶ran Sperber, and Jens Mayer. Classification and nomenclature of endogenous retroviral sequences (ERVs). *Gene*, 448(2):115–123, dec 2009.

[22] Reyad A. Elbarbary, Bronwyn A. Lucas, and Lynne E. Maquat. Retrotransposons as regulators of gene expression. *Science*, 351(6274), feb 2016.

[23] Marie Dewannieux and Thierry Heidmann. Endogenous retroviruses: acquisition, amplification and taming of genome invaders. *Current Opinion in Virology*, 3(6):646–656, dec 2013.

[24] George Kassiotis and Jonathan P. Stoye. Immune responses to endogenous retroelements: taking the bad with the good. *Nature Reviews Immunology*, 16(4):207–219, mar 2016.

[25] Cédric Feschotte and Ellen J. Pritham. DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, 41(1):331–368, dec 2007.

[26] Jane M. Carlton, Robert P. Hirt, Joana C. Silva, Arthur L. Delcher, Michael Schatz, Qi Zhao, Jennifer R. Wortman, Shelby L. Bidwell, U. Cecilia M. Alsmark, Sebastien Besteiro, Thomas Sicheritz-Ponten, Christophe J. Noel, Joel B. Dacks, Peter G. Foster, Cedric Simillion, Yves Van de Peer, Diego Miranda-Saavedra, Geoffrey J. Barton, Gareth D. Westrop, Sylke Muller, Daniele Dessi, Pier Luigi Fiori, Qinghu Ren, Ian Paulsen, Hanbang Zhang,

Felix D. Bastida-Corcuera, Augusto Simoes-Barbosa, Mark T. Brown, Richard D. Hayes, Mandira Mukherjee, Cheryl Y. Okumura, Rachel Schneider, Alias J. Smith, Stepanka Vanacova, Maria Villalvazo, Brian J. Haas, Mihaela Pertea, Tamara V. Feldblyum, Terry R. Utterback, Chung-Li Shu, Kazutoyo Osoegawa, Pieter J. de Jong, Ivan Hrdy, Lenka Horvathova, Zuzana Zubacova, Pavel Dolezal, Shehre-Banoo Malik, John M. Logsdon, Katrin Henze, Arti Gupta, Ching C. Wang, Rebecca L. Dunne, Jacqueline A. Upcroft, Peter Upcroft, Owen White, Steven L. Salzberg, Petrus Tang, Cheng-Hsun Chiu, Ying-Shiung Lee, T. Martin Embley, Graham H. Coombs, Jeremy C. Mottram, Jan Tachezy, Claire M. Fraser-Liggett, and Patricia J. Johnson. Draft genome sequence of the sexually transmitted pathogen trichomonas vaginalis. *Science*, 315(5809):207–212, jan 2007.

[27] Meixia Zhao and Jianxin Ma. Co-evolution of plant LTR-retrotransposons and their host genomes. *Protein & Cell*, 4(7):493–501, jun 2013.

[28] Regina S. Baucom, James C. Estill, Cristian Chaparro, Naadira Upshaw, Ansuya Jogi, Jean-Marc Deragon, Richard P. Westerman, Phillip J. SanMiguel, and Jeffrey L. Bennetzen. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the b73 maize genome. *PLoS Genetics*, 5(11):e1000732, nov 2009.

[29] Jonathan Kreplak, Mohammed-Amin Madoui, Petr Cápal, Petr Novák, Karine Labadie, Grégoire Aubert, Philipp E. Bayer, Krishna K. Gali, Robert A. Syme, Dorrie Main, Anthony Klein, Aurélie Bérard, Iva Vrbová, Cyril Fournier, Leo d'Agata, Caroline Belser, Wahiba Berrabah, Helena Toegelová, Zbyněk Milec, Jan Vrána, HueyTyng Lee, Ayité Kougbeadjo, Morgane Térézol, Cécile Huneau, Chala J. Turo, Nacer Mohellibi, Pavel Neumann, Matthieu Falque, Karine Gallardo, Rebecca McGee, Bunyamin Tar'an, Abdelhafid Bendahmane, Jean-Marc Aury, Jacqueline Batley, Marie-Christine Le Paslier, Noel Ellis, Thomas D. Warkentin, Clarice J. Coyne, Jérome Salse, David Edwards, Judith Lichtenzveig, Jiří Macas, Jaroslav Doležel, Patrick Wincker, and Judith Burstin. A reference genome for pea provides insight into legume genome evolution. *Nature Genetics*, 51(9):1411–1422, sep 2019.

[30] Sergej Nowoshilow, Siegfried Schloissnig, Ji-Feng Fei, Andreas Dahl, Andy W. C. Pang, Martin Pippel, Sylke Winkler, Alex R. Hastie, George Young, Juliana G. Roscito, Francisco Falcon, Dunja Knapp, Sean Powell, Alfredo Cruz, Han Cao, Bianca Habermann, Michael Hiller, Elly M. Tanaka, and Eugene W. Myers. The axolotl genome and the evolution of key tissue formation regulators. *Nature*, 554(7690):50–55, jan 2018.

[31] Jeramiah J. Smith, Nataliya Timoshevskaya, Vladimir A. Timoshevskiy, Melissa C. Keinath, Drew Hardy, and S. Randal Voss. A chromosome-scale assembly of the axolotl genome. *Genome Research*, 29(2):317–324, jan 2019.

[32] Aris Katzourakis, Gkikas Magiorkinis, Aaron G. Lim, Sunetra Gupta, Robert Belshaw, and Robert Gifford. Larger mammalian body size leads to lower retroviral activity. *PLoS Pathogens*, 10(7):e1004214, jul 2014.

[33] Arian F.A. Smit. Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Research*, 21(8):1863–1872, 1993.

[34] Eugene M McCarthy and John F McDonald. Long terminal repeat retrotransposons of mus musculus. *Genome Biology*, 5(3):R14, 2004.

[35] Guojie Zhang, Cai Li, Qiye Li, Bo Li, Denis M. Larkin, Chul Lee, Jay F. Storz, Agostinho Antunes, Matthew J. Greenwold, Robert W. Meredith, Anders Odeen, Jie Cui, Qi Zhou, Luohao Xu, Hailin Pan, Zongji Wang, Lijun Jin, Pei Zhang, Haofu Hu, Wei Yang, Jiang Hu, Jin Xiao, Zhikai Yang, Yang Liu, Qiaolin Xie, Hao Yu, Jinmin Lian, Ping Wen, Fang Zhang, Hui Li, Yongli Zeng, Zijun Xiong, Shiping Liu, Long Zhou, Zhiyong Huang, Na An, Jie Wang, Qiumei Zheng, Yingqi Xiong, Guangbiao Wang, Bo Wang, Jingjing Wang, Yu Fan, Rute R. da Fonseca, Alonzo Alfaro-Núñez, Mikkel Schubert, Ludovic Orlando, Tobias Mourier, Jason T. Howard, Ganeshkumar Ganapathy, Andreas Pfenning, Osceola Whitney, Miriam V. Rivas, Erina Hara, Julia Smith, Marta Farré, Jitendra Narayan, Gancho Slavov, Michael N Romanov, Rui Borges, João Paulo Machado, Imran Khan, Mark S. Springer, John Gatesy, Federico G. Hoffmann, Juan C. Opazo, Olle Håstad, Roger H. Sawyer, Heebal Kim, Kyu-Won Kim, Hyeon Jeong Kim, Seoae Cho, Ning Li, Yinhua Huang, Michael W. Bruford, Xiangjiang Zhan, Andrew Dixon, Mads F. Bertelsen, Elizabeth Derryberry, Wesley Warren, Richard K Wilson, Shengbin Li, David A. Ray, Richard E. Green, Stephen J. O'Brien, Darren Griffin, Warren E. Johnson, David Haussler, Oliver A. Ryder, Eske Willerslev, Gary R. Graves, Per AlstrÃ¶m, Jon Fjeldså, David P. Mindell, Scott V. Edwards, Edward L. Braun, Carsten Rahbek, David W. Burt, Peter Houde, Yong Zhang, Huanming Yang, Jian Wang, Erich D. Jarvis, M. Thomas P. Gilbert, Jun Wang, Chen Ye, Shaoguang Liang, Zengli Yan, M. Lisandra Zepeda, Paula F. Campos, Amhed Missael Vargas Velazquez, José Alfredo Samaniego, María Avila-Arcos, Michael D. Martin, Ross Barnett, Angela M. Ribeiro, Claudio V. Mello, Peter V. Lovell, Daniela Almeida, Emanuel Maldonado, Joana Pereira, Kartik Sunagar, Siby Philip, Maria Gloria Dominguez-Bello, Michael Bunce, David Lambert, Robb T. Brumfield, Frederick H. Sheldon, Edward C. Holmes, Paul P. Gardner, Tammy E. Steeves, Peter F. Stadler, Sarah W. Burge, Eric Lyons, Jacqueline Smith, Fiona McCarthy, Frederique Pitel, Douglas Rhoads, and David P. Froman and. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346(6215):1311–1320, dec 2014.

[36] Chammiran Daniel, Mikaela Behm, and Marie Ohman. The role of alu elements in the cis-regulation of RNA processing. *Cellular and Molecular Life Sciences*, 72(21):4063–4076, jul 2015.

[37] Patrick P. Edger, Thomas J. Poorten, Robert VanBuren, Michael A. Hardigan, Marivi Colle, Michael R. McKain, Ronald D. Smith, Scott J. Teresi, Andrew D. L. Nelson, Ching Man Wai, Elizabeth I. Alger, Kevin A. Bird, Alan E. Yocca, Nathan Pumplin, Shujun Ou, Gil Ben-Zvi, Avital Brodt, Kobi Baruch, Thomas Swale, Lily Shiue, Charlotte B. Acharya, Glenn S. Cole, Jeffrey P. Mower, Kevin L. Childs, Ning Jiang, Eric Lyons, Michael Freeling, Joshua R. Puzey, and Steven J. Knapp. Origin and evolution of the octoploid strawberry genome. *Nature Genetics*, 51(3):541–547, feb 2019.

[38] Vance M. Whitaker, Steven J. Knapp, Michael A. Hardigan, Patrick P. Edger, Janet P. Slovin, Nahla V. Bassil, Timo HytÃ¶nen, Kathryn K. Mackenzie, Seonghee Lee, Sook Jung, Dorrie Main, Christopher R. Barbey, and Sujeet Verma. A roadmap for research in octoploid strawberry. *Horticulture Research*, 7(1), mar 2020.

[39] Ana Teixeira-Silva, Raquel M. Silva, João Carneiro, António Amorim, and Luísa Azevedo. The role of recombination in the origin and evolution of alu subfamilies. *PLoS ONE*, 8(6):e64884, jun 2013.
[40] David J Witherspoon, W Scott Watkins, Yuhua Zhang, Jinchuan Xing, Whitney L Tolpinrud, Dale J Hedges, Mark A Batzer, and Lynn B Jorde. Alu repeats increase local recombination rates. *BMC Genomics*, 10(1):530, 2009.
[41] Anne Dupressoir, Cécile Vernochet, Olivia Bawa, Francis Harper, Gérard Pierron, Paule Opolon, and Thierry Heidmann. Syncytin-a knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proceedings of the National Academy of Sciences*, 106(29):12127–12132, jul 2009.
[42] Amélie E. Coudert, François Redelsperger, Yasmine Chabbi-Achengli, Cécile Vernochet, Caroline Marty, Xavier Decrouy, Thierry Heidmann, Marie-Christine de Vernejoul, and Anne Dupressoir. Role of the captured retroviral envelope syncytin-b gene in the fusion of osteoclast and giant cell precursors and in bone resorption, analyzed ex vivo and in vivo in syncytin-b knockout mice. *Bone Reports*, 11:100214, dec 2019.
[43] Hannah E Volkman and Daniel B Stetson. The enemy within: endogenous retroelements and autoimmune disease. *Nature Immunology*, 15(5):415–422, apr 2014.
[44] Maulik R Patel, Michael Emerman, and Harmit S Malik. Paleovirology—ghosts and gifts of viruses past. *Current Opinion in Virology*, 1(4):304–309, oct 2011.
[45] Aris Katzourakis, Andrew Rambaut, and Oliver G. Pybus. The evolutionary dynamics of endogenous retroviruses. *Trends in Microbiology*, 13(10):463–468, jan 2005.
[46] Aris Katzourakis. Paleovirology: inferring viral evolution from host genome sequence data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1626):20120493, sep 2013.
[47] Amr Aswad and Aris Katzourakis. Paleovirology and virally derived immunity. *Trends in Ecology & Evolution*, 27(11):627–636, nov 2012.
[48] Jennifer F. Hughes and John M. Coffin. Human endogenous retrovirus k solo-LTR formation and insertional polymorphisms: Implications for human and viral evolution. *Proceedings of the National Academy of Sciences*, 101(6):1668–1672, feb 2004.
[49] Robert J. Gifford, Jonas Blomberg, John M. Coffin, Hung Fan, Thierry Heidmann, Jens Mayer, Jonathan Stoye, Michael Tristem, and Welkin E. Johnson. Nomenclature for endogenous retrovirus (ERV) loci. *Retrovirology*, 15(1), aug 2018.
[50] Arifa S. Khan, Jochen Bodem, Florence Buseyne, Antoine Gessain, Welkin Johnson, Jens H. Kuhn, Jacek Kuzmak, Dirk Lindemann, Maxine L. Linial, Martin LÃ¶chelt, Magdalena Materniak-Kornas, Marcelo A. Soares, and William M. Switzer. Spumaretroviruses: Updated taxonomy and nomenclature. *Virology*, 516:158–164, mar 2018.
[51] Rafael Sanjuán. From molecular genetics to phylodynamics: Evolutionary relevance of mutation rates across viruses. *PLoS Pathogens*, 8(5):e1002685, may 2012.
[52] Jay Lubow and Kathleen L. Collins. Vpr is a VIP: HIV vpr and infected

macrophages promote viral pathogenesis. *Viruses*, 12(8):809, jul 2020.
[53] Ryan P. Staudt, John J. Alvarado, Lori A. Emert-Sedlak, Haibin Shi, Sherry T. Shu, Thomas E. Wales, John R. Engen, and Thomas E. Smithgall. Structure, function, and inhibitor targeting of HIV-1 nef-effector kinase complexes. *Journal of Biological Chemistry*, 295(44):15158–15171, oct 2020.
[54] Nabab Khan and Jonathan D. Geiger. Role of viral protein u (vpu) in HIV-1 infection and pathogenesis. *Viruses*, 13(8):1466, jul 2021.
[55] Aris Katzourakis, Vini Pereira, and Michael Tristem. Effects of recombination rate on human endogenous retrovirus fixation and persistence. *Journal of Virology*, 81(19):10712–10717, oct 2007.
[56] Aris Katzourakis, Michael Tristem, Oliver G. Pybus, and Robert J. Gifford. Discovery and analysis of the first endogenous lentivirus. *Proceedings of the National Academy of Sciences*, 104(15):6261–6265, apr 2007.
[57] Robert J. Gifford, Aris Katzourakis, Michael Tristem, Oliver G. Pybus, Mark Winters, and Robert W. Shafer. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proceedings of the National Academy of Sciences*, 105(51):20362–20367, dec 2008.
[58] M.R. Dilworth and C.P. Sibley. Review: Transport across the placenta of mice and women. *Placenta*, 34:S34–S39, mar 2013.
[59] Emin Maltepe and Susan J. Fisher. Placenta: The forgotten organ. *Annual Review of Cell and Developmental Biology*, 31(1):523–552, nov 2015.
[60] Emin Maltepe, Anna I. Bakardjiev, and Susan J. Fisher. The placenta: transcriptional, epigenetic, and physiological integration during development. *Journal of Clinical Investigation*, 120(4):1016–1025, apr 2010.
[61] Irene Cetin. Amino acid interconversions in the fetal-placental unit: The animal model and human studies in vivo. *Pediatric Research*, 49(2):148–154, feb 2001.
[62] Xuzhe Zhang and Louis J. Muglia. Baby's best foe-riend: Endogenous retroviruses and the evolution of eutherian reproduction. *Placenta*, 113:1–7, sep 2021.
[63] Anne Dupressoir, Cécile Vernochet, Francis Harper, Justine Guégan, Philippe Dessen, Gérard Pierron, and Thierry Heidmann. A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. *Proceedings of the National Academy of Sciences*, 108(46), oct 2011.
[64] Guillaume Cornelis, Sylvie Souquere, Cécile Vernochet, Thierry Heidmann, and Gérard Pierron. Functional conservation of the lncRNA NEAT1 in the ancestrally diverged marsupial lineage: Evidence for NEAT1 expression and associated paraspeckle assembly during late gestation in the opossum monodelphis domestica. *RNA Biology*, 13(9):826–836, jul 2016.
[65] Julian Buchrieser, Séverine A. Degrelle, Thérèse Couderc, Quentin Nevers, Olivier Disson, Caroline Manet, Daniel A. Donahue, Françoise Porrot, Kenzo-Hugo Hillion, Emeline Perthame, Marlene V. Arroyo, Sylvie Souquere, Katinka Ruigrok, Anne Dupressoir, Thierry Heidmann, Xavier Montagutelli, Thierry Fournier, Marc Lecuit, and Olivier Schwartz. IFITM proteins inhibit placental syncytiotrophoblast formation and promote fetal demise. *Science*, 365(6449):176–180, jul 2019.

[66] Derek E. Wildman, Caoyi Chen, Offer Erez, Lawrence I. Grossman, Morris Goodman, and Roberto Romero. Evolution of the mammalian placenta revealed by phylogenetic analysis. *Proceedings of the National Academy of Sciences*, 103(9):3203–3208, feb 2006.

[67] Kazuhiko Imakawa, So Nakagawa, and Takayuki Miyazawa. Baton pass hypothesis: successive incorporation of unconserved endogenous retroviral genes for placentation during mammalian evolution. *Genes to Cells*, 20(10):771–788, sep 2015.

[68] Randall J. Cohrs, Tyler Martin, Parviz Ghahramani, Luc Bidaut, Paul J. Higgins, and Aamir Shahzad. Translational medicine definition by the european society for translational medicine. *New Horizons in Translational Medicine*, 2(3):86–88, 2015.

[69] Ettore Beghi, Giorgia Giussani, Emma Nichols, Foad Abd-Allah, Jemal Abdela, Ahmed Abdelalim, Haftom Niguse Abraha, Mina G. Adib, Sutapa Agrawal, Fares Alahdab, Ashish Awasthi, Yohanes Ayele, Miguel A Barboza, Abate Bekele Belachew, Belete Biadgo, Ali Bijani, Helen Bitew, Félix Carvalho, Yazan Chaiah, Ahmad Daryani, Huyen Phuc Do, Manisha Dubey, Aman Yesuf Yesuf Endries, Sharareh Eskandarieh, Andre Faro, Farshad Farzadfar, Seyed-Mohammad Fereshtehnejad, Eduarda Fernandes, Daniel Obadare Fijabi, Irina Filip, Florian Fischer, Abadi Kahsu Gebre, Afewerki Gebremeskel Tsadik, Teklu Gebrehiwo Gebremichael, Kebede Embaye Gezae, Maryam Ghasemi-Kasman, Kidu Gidey Weldegwergs, Meaza Girma Degefa, Elena V. Gnedovskaya, Tekleberhan B Hagos, Arvin Haj-Mirzaian, Arya Haj-Mirzaian, Hamid Yimam Hassen, Simon I Hay, Mihajlo Jakovljevic, Amir Kasaeian, Tesfaye Dessale Kassa, Yousef Saleh Khader, Ibrahim Khalil, Ejaz Ahmad Khan, Jagdish Khubchandani, Adnan Kisa, Kristopher J Krohn, Chanda Kulkarni, Yirga Legesse Nirayo, Mark T Mackay, Marek Majdan, Azeem Majeed, Treh Manhertz, Man Mohan Mehndiratta, Tesfa Mekonen, Hagazi Gebre Meles, Getnet Mengistu, Shafiu Mohammed, Mohsen Naghavi, Ali H Mokdad, Ghulam Mustafa, Seyed Sina Naghibi Irvani, Long Hoang Nguyen, Molly R Nixon, Felix Akpojene Ogbo, Andrew T Olagunju, Tinuke O Olagunju, Mayowa Ojo Owolabi, Michael R Phillips, Gabriel David Pinilla-Monsalve, Mostafa Qorbani, Amir Radfar, Anwar Rafay, Vafa Rahimi-Movaghar, Nickolas Reinig, Perminder S Sachdev, Hosein Safari, Saeed Safari, Saeid Safiri, Mohammad Ali Sahraian, Abdallah M. Samy, Shahabeddin Sarvi, Monika Sawhney, Masood A Shaikh, Mehdi Sharif, Gagandeep Singh, Mari Smith, Cassandra E I Szoeke, Rafael Tabarés-Seisdedos, Mohamad-Hani Temsah, Omar Temsah, Miguel Tortajada-Girbés, Bach Xuan Tran, Amanuel Amanuel Tesfay Tsegay, Irfan Ullah, Narayanaswamy Venketasubramanian, Ronny Westerman, Andrea Sylvia Winkler, Ebrahim M Yimer, Naohiro Yonemoto, Valery L. Feigin, Theo Vos, and Christopher J L Murray. Global, regional, and national burden of epilepsy, 1990–2016: a systematic analysis for the global burden of disease study 2016. *The Lancet Neurology*, 18(4):357–375, apr 2019.

[70] John J. Wittman and Lawrence J. Hirsch. Continuous electroencephalogram monitoring in the critically ill. *Neurocritical Care*, 2(3):330–341, 2005.

[71] Chiara Cirelli and Giulio Tononi. Cortical development, electroencephalogram

rhythms, and the sleep/wake cycle. *Biological Psychiatry*, 77(12):1071–1078, jun 2015.

[72] Mahmoud Al-Kadi, Mamun Reaz, and Mohd Ali. Evolution of electroencephalogram signal analysis techniques during anesthesia. *Sensors*, 13(5):6605–6635, may 2013.

[73] Shaurya Taran, Wael Ahmed, Esther Bui, Lara Prisco, Cecil D. Hahn, and Victoria A. McCredie. Educational initiatives and implementation of electroencephalography into the acute care environment: a protocol of a systematic review. *Systematic Reviews*, 9(1), aug 2020.

[74] Aura Silva and Luis Antunes. Electroencephalogram-based anaesthetic depth monitoring in laboratory animals. *Laboratory Animals*, 46(2):85–94, apr 2012.

[75] Mohammad Khubeb Siddiqui, Ruben Morales-Menendez, Xiaodi Huang, and Nasir Hussain. A review of epileptic seizure detection using machine learning classifiers. *Brain Informatics*, 7(1), may 2020.

[76] Bardia Abbasi and Daniel M. Goldenholz. Machine learning applications in epilepsy. *Epilepsia*, 60(10):2037–2047, sep 2019.

[77] Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of Neural Engineering*, 16(3):031001, apr 2019.

[78] Awni Y. Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H. Tison, Codie Bourn, Mintu P. Turakhia, and Andrew Y. Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1):65–69, jan 2019.

[79] Piotr Olejniczak. Neurophysiologic basis of EEG. *Journal of Clinical Neurophysiology*, 23(3):186–189, jun 2006.

[80] Patrick L. Purdon, Aaron Sampson, Kara J. Pavone, and Emery N. Brown. Clinical electroencephalography for anesthesiologists. *Anesthesiology*, 123(4):937–960, oct 2015.

[81] Elisa Baldin, W. Allen Hauser, Jeffrey R. Buchhalter, Dale C. Hesdorffer, and Ruth Ottman. Yield of epileptiform electroencephalogram abnormalities in incident unprovoked seizures: A population-based study. *Epilepsia*, 55(9):1389–1398, jul 2014.

[82] A. Schreiner and B. Pohlmann-Eden. Value of the early electroencephalogram after a first unprovoked seizure. *Clinical Electroencephalography*, 34(3):140–144, jul 2003.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1916

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-473354

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2023