Master Thesis

Mobile Genetic Elements

Transposons - Beneficial Jumpers of Evolution or Pathogenic Parasites of a Viral Past?

eingereicht von

MMag. Mohamed Abdel Keream

zur Erlangung des akademischen Grades

Master of Science

(MSc)

Medizinische Universität Graz
Universitätslehrgang Medizinische Genetik 2019
Institut für Humangenetik

unter der Anleitung von

Prof. DDr. Erwin Petek

Wien, August, 2021

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Wien, August, 2021

Table of contents

I. Abbreviations

5′ LT-RMSCV: 5′ long terminal repeat of the murine stem cell virus
AA: Amino acid
ACE2: angiotensin-converting enzyme 2
ADH: antidiuretic hormone
ADH: atypical ductal hyperplasia
AID: activation induced deaminase
AngI/II: angiotensin I/II
APC: Adenomatous polyposis coli
AQAMA: absolute quantitative assessment of methylated alleles
ARB: angiotensin II receptor blocker
ARDS: acute respiratory distress syndrome sepsis
BER: base excision repair
CCG: candidate cancer (driver) genes
cfDNA: cell free DNA
CLL: chronic lymphocytic leukemia
CML: chronic myeloid leukemia
CNV: copy number variations
Cobra: combined bisulphite restriction analysis (COBRA)
CpG: cytosine nucleotide followed by a guanine nucleotide
CRC: colorectal carcinomas
CRISPR: clustered regularly interspaced short palindromic repeats
CRTL: control group
CSR: Class switch recombination
DCIS: ductal carcinoma in situ
DDE motif: Aspartate-97, Aspartate-188 and Glutamate-326
DNMT3a/b: DNA methyltransferases
EFV: efavirenz
EGFR: Epidermal growth factor receptor
EMT: Epithelial–mesenchymal transition
EN: deoxyribonuclease
ENV: envelope protein

ER: endoplasmic reticulum (ER

FEA: flat epithelial atypia

GoF: Gain of Function

H(D)R: homologous directed recombination

IBC: invasive breast cancer

IgG: Immunoglobulin G

IL1-6: interleukins

IRRDRL: inverted repeat sequences left end

IRRDRR: right inverted repeat sequence right end

JGA: juxtaglomerular apparatus

LINE: long interspersed elements

LoF: Loss of Function

LTR: long terminal repeats

MGE: mobile genetic elements

MHC: major histocompatibility complex (

MIRA: methylated-CpG island recovery assay

miRNA: microRNAs

MM: multiple myeloma

MOF: multisystem organ failure

MSP: methylation-specific PCR

N: nucleocapsid

NHEJ: non-homologous end joining (NHEJ)

NVR: nevirapine

ORF: open reading frame

PBS: primer binding site (PBS)

PG: piggyBac

piRNA: PIWI-interacting RNAs

PLE: Penelope-like elements

POL: polypeptide/polyprotein gene

PSA: prostate specific antigen (PSA)

RAAS: renin-angiotensin-aldosterone system

RAG1/ 2: recombination-activating genes

RNase: Ribonuclease

RT: reverse transcriptase

S: spike protein

SA: splice site acceptor

SB: Sleeping Beauty

SCF: synaptic complex

SD: splice site donor

shRNA: short hairpin RNA

SINE: short interspersed elements

siRNA: small interfering RNA

SMT: somatic mutation theory (SMT)

ss-cDNA: double stranded complementary DNA

T-ALL: T-cell acute lymphoblastic leukemia

TE: transposable elements

Tnp: transposase

TPRT: target primed reverse transcription

TSD: target-site duplication

TST: target-site truncation

UNG: Uracil-N-glycosylase

UTR: Untranslated region

VNTR: variable number tandem repeat

II. List of Figures

III. Zusammenfassung

Das menschliche Genom umfasst im diploiden Zustand über 6 Millionen Basenpaare, von denen nur 1,5 % für die 20.000-25.000 bekannten Gene kodieren - der Rest, über 90 % des Genoms, war bisher - und wird immer noch als "Junk"-DNA bezeichnet. Dies impliziert, dass diese Abschnitte keine (proteinogene) Funktion haben und führt dazu, dass sie in der Regel nicht in den Fokus der Forschung geraten.

Bei näherer Betrachtung zeigt sich, dass es sich bei diesen Abschnitten hauptsächlich um Transposons, also mobile genetische Elemente, auch "springende Gene" genannt, handelt. Sie wurden bereits in den 1940er Jahren von Nobelpreisträger Barbara McClintock in der Maispflanze entdeckt.

Die bekanntesten Transposons sind einerseits die Alu-Elemente, die fast eine Million Mal im Genom kodiert sind, und anderseits die Line-1-Retrotransposons, die 20 % unseres Genoms ausmachen und von denen noch etwa 100 Kopien transponierfähig sind, was zum genetischen Polymorphismus bei uns Menschen führt.

Transposons kodieren und rekrutieren spezielle Enzyme für die Transposition. Retrotransposons (Klasse I) benötigen dazu reverse Transkriptase, während DNA-Transposons (Klasse II) auf eine Transposase angewiesen sind. Während Transposons in somatischen Zellen durch eine Reihe epigenetischer Mechanismen unterdrückt werden, insbesondere durch die DNA-Methylierung von CpG-Inseln in der Promotorregion und die Unterdrückung durch piwiRNA, können sie in bestimmten Geweben und insbesondere in bestimmten Stadien der Embryogenese aktiv sein und eine wichtige Rolle bei der Regulierung der Expression von Plazenta-Genen spielen.

Transposons wurden über verschiedene evolutionäre Wege in das menschliche Genom eingebaut und danach entweder reprimiert oder domestiziert. Eines der bekanntesten Beispiele für eine solche erfolgreiche Zähmung ist die modifizierte Transposase, die für die somatische Rekombination von Immunglobulingenen verantwortlich ist (RAG 1 und RAG 2) - ebenso wie die ursprüngliche Reverse

Transkriptase, die als Telomerase für die chromosomale Integrität umfunktioniert wurde.

Transposons und ihre Derivate haben also eine wichtige Funktion auf der genetischen und genomischen Ebene, aber auch auf der Ebene des Transkriptoms oder Proteoms. Ihre Transposonaktivität sowie Repression durch das Wirtsgenom befinden sich in einer fragilen Homöostase, die durch exogene, aber auch endogene Faktoren gestört werden kann. Zu den exogenen Faktoren gehört vor allem evolutionärer Stress. Endogene Faktoren, die die Beziehung zwischen Transposon und Wirtsgenom stören, sind in erster Linie genetischer und epigenetischer Natur und führen unweigerlich zu genetisch bedingten bösartigen Erkrankungen, vor allem zu Tumoren.

In der Tat kann die Retrotransposition der zuvor erwähnten Line-1-Retroelemente sowohl eine "driver mutation", d. h. ursächlich für die Krebsentstehung, als auch eine "passanger mutation" für bestimmte Tumorentitäten sein. Daher eignen sich Line-1-Elemente auch als Tumormarker für das gesamte Genom: Alle Tumoren sind durch die Hypomethylierung von Retroelementen gekennzeichnet. Das Methylierungsmuster, aber auch die Insertionsmutagene sowie die erzeugten Proteinprodukte, d.h. die Reverse-Transkriptase-Aktivität, sind eindeutige Merkmale von Krebs und liefern diagnostische, prognostische und prädiktive Informationen, die praktikabel über Liquid Biopsy überprüft werden können. Gleichzeitig dienen Transposons als therapeutische Targets in der personalisierten Onkogenetik: Die Inhibition der Reversen Transkriptase hat in verschiedenen klinischen Studien bei Brustkrebs, Darmkrebs, Lungenkrebs und Prostatakrebs gezeigt, dass sie das Tumorwachstum hemmt.

Transposons sind nicht nur ein therapeutisches Ziel, sondern bieten sich auch als Werkzeug für eine mögliche Gentherapie an. Besonders vielversprechend ist das Transposon Sleeping Beauty, das viele Vorteile gegenüber viralen Vektoren hat und in Kombination mit inaktiven CRISPR/Cas-Systemen zur gezielten Zellbehandlung eingesetzt werden kann.

IV. Abstract

The human genome in the diploid mode comprises over 6 million base pairs, of which only 1.5% encode the 20,000-25,000 known genes - the rest, over 90% of the genome, were previously - and are still referred to as "junk" DNA. This implies that these Segments would have no (proteinogenic) function and leads to the fact that they usually do not get in the focus of research.

A closer look reveals that these regions consist mainly of transposons, videlicet mobile genetic elements, also known as "jumping genes". They were already discovered in the 1940s by Nobel Prize Laureate Barbara McClintock in the maize plant.

Transposons can be divided into Class I retrotransposons with RNA intermediates ("copy and paste") and Class II DNA transposons ("cut and paste").The best known transposons are Alu elements, which are encoded almost a million times in the genome, and Line-1 retrotransposons, which make up 20% of our genome and of which about 100 copies are still capable of transposition, leading to the genetic polymorphism of us humans.

Transposons encode and recruit special enzymes for transposition. Retrotransposons (Class I) require reverse transcriptase for this, while DNA transposons (Class II) rely on a transposase. While transposons are repressed in somatic cells by a number of epigenetic mechanisms, most notably DNA methylation of CpG islands in the promoter region and repression by piwiRNA, they can be active in certain tissues and especially at certain stages of embryogenesis, playing an important role in regulating the expression of placental genes.

Transposons have been incorporated into the human genome through several evolutionary pathways, thereafter, either being silenced or domesticated. One of the best-known examples of such successful taming is the modified transposase responsible for somatic recombination of immunoglobulin genes- as well as the original reverse transcriptase, which has been repurposed as a telomerase for chromosomal integrity.

Thus, transposons and their derivatives have an important function at the genetic and genomic level, but also at the level transcriptome or proteome. Their transposon activity and repression by the host genome are in a fragile homeostasis that can be disrupted by exogenous but also endogenous factors. Exogenous factors primarily include evolutionary stress. Endogenous factors that disrupt the trasposon-host relationship are primarily genetic and epigenetic in nature and inevitably induce genetic malignancies-most notably tumors.

In fact, retrotransposition of the previously mentioned Line-1 retroelements can be both "driver mutation", i.e. causative for carcinogenesis, and "passenger mutation" of certain cancers. Therefore, Line-1 elements are also suitable as whole-genome tumor markers: all tumors are characterized by the hypomethylation of retroelements. The methylation pattern, but also the insertional mutagens as well as the generated protein products, i.e. the reverse transcriptase activity are clear characteristics of cancer and provide diagnostic, prognostic and predictive information. As tumor markers they can feasibly be monitored via Liquid Biopsy At the same time, transposons serve as therapeutic targets in personalized oncogenetics: inhibition of reverse transcriptase has been shown to inhibit tumor growth in various clinical trials of breast cancer, colorectal carcinoma, lung cancer, and prostate cancer.

In addition to being a therapeutic target, transposons also offer themselves as a tool for potential gene therapy. Particularly promising is the transposon Sleeping Beauty, which has many advantages over viral vectors and can be used in combination with inactive CRISPR/Cas systems to target cells.

1. Introduction

In the genome of eukaryotes, we find a large number of repetitive sequences (Waring & Britten, 1966). It is now well established, that most of these repetitive nucleotide sequences belong to mobile genetic elements, i.e., transposable elements (TE) (Makalowski, 2001). Although the repetitive portion of a genome varies significantly between different species, it is quite high in Mammalia and humans at a percentage of over 50% (Makalowski, 2001). In plants, where Nobel laureate Barbara McClintock first discovered transposons in 1944, they account for up to 80% of the genome (McClintock, 1950). Therefore, it is not surprising that transposons have a significant impact on the organization and evolution of the genome. However, in the scientific community, this influence is approached differently: On the one hand, transposons are praised as helpful, even necessary, engines of evolution, emphasizing their role in the diversification of biological diversity in fauna and flora (Carroll et al., 2001; Roy-Engel et al., 2002; Casacuberta & González, 2013). On the other hand, there is increasing research on the relationship between transposable elements and genetic diseases (Reilly et al., 2013). Biémont (2010) puts this dichotomous perspective on transposons as follows:

> *Since the radical suggestion [...] that some genes might move along chromosomes, our knowledge of transposable elements (TEs) has vastly increased. TEs are no longer seen as "junk" and "selfish" pieces of DNA-the predominant view from the 1960s through the 1990s-but as major components of genomes that have played a significant role in evolution [...]. The history of these genomic elements provides one of the best examples of how scientific concepts in biology emerge and then evolve into new concepts. It is a salutary lesson for researchers, both young and old, to be tolerant of striking new ideas when they appear and not to dismiss them simply because they conflict with current theories and knowledge. (Biémont, 2010: 1085)*

This rethinking not only challenges axioms or dogmas of genetics, but also opens up new perspectives on human genetic research desiderata. A deeper understanding of transposons will help to better elucidate our own evolution and understand the (epi-) genetic causes of genetic malignancies and establish new tools for diagnosis and therapy.

2. Research Questions

As previously stated, transposons make up at least 50% of our human genome. Interestingly, however, they find no more than a brief mention and classification in very few textbooks and introductions to human genetics. The preceding quote from Biémont (2010) emphasizes that the existence of mobile sequences counteracts the previous operating concept of a static – almost linear- genome and repeatedly challenges the central dogma of molecular biology. Moreover, the tentative and belated engagement of genetic research with transposons owes much to the fact that both constructive and destructive implications accompany their activity and genetic impact. The aim of this review is, in a first approach, to collect previous research results on classification, structure and assembly, transposition mechanisms, and interaction with the genome.

The second approach will use the knowledge gained to address the overarching question: are **transposons "beneficial jumpers of evolution or pathogenic parasites of a viral past? "**

Specifically, we aim to shed light on the constructive and destructive implications of transposons' (retro-) transpositional activity in the human genome. What is the importance of transposable elements (TEs) for phylogenesis? What is the impact of their presence or absence on gene expression? How is their activity regulated, and what are the effects of de-regulation of transposon activity? What is the relationship between transposons and malignant genetic diseases- preeminently tumor diseases?

In the final part, opportunities for operationalizing TEs for oncogenomic research are elucidated. The guiding question here is also: *What opportunities for better diagnosis of genetically determined pathogeneses, but also for prevention and therapy of these mentioned malignancies are opened up by TEs?*

3. Theoretical Foundation

For the theoretical foundation, the introductory volume ***Basiswissen Humangenetik*** by Schaaf and Zschocke (2018) and the study book ***Concepts of Genetics*** by Klug et al. (2015) were used as main references. For the results section, primary and secondary literature was searched via databases (PubMed, Ovid, Medline, Embase, Scopus) and cross-referenced.

3.1 Discovery

The history of discovery and research on transposons reveal their dichotomous perception in diachrony (Ravindran, 2012). When botanist Barbara McClintock challenged traditional, rigid concepts of genetics in the late 1940s with her discovery of mobile nucleotide sequences in the genome of the maize plant, her publications met with minor approval or even disdain. In the 1970s, significant advances in molecular biology led to the discovery of transposons in other organisms, beginning with viruses and bacteria. We now know that transposons make up as much as half of our genome and about 85% of the maize genome (Ravindran, 2012).

The finding that transposons were equally common among eukaryotes eventually led to a broader recognition of their original discovery. Barbara McClintock received a number of prestigious awards, including the 1970 National Medal of Science, climaxing in 1983 in an unshared Nobel Prize in Physiology or Medicine.

In the press release about the Nobel Prize, McClintock stated: ***"It might seem unfair to reward a person for having so much pleasure, over the years, asking the maize plant to solve specific problems and then watching its responses"*** (McGrayne & Byers, 1994).

3.2 Evolution

In their review *Ten things you should know about transposable elements*, Bourque et al. (2018) bring together all the unique features of transposons. These include the fact that TEs are conserved in the mammalian genome over long distances, emphasizing their evolutionary importance (Buckley et al., 2017):

> *Thus, the success and diversity of TEs in a genome are shaped both by properties intrinsic to the elements as well as evolutionary forces acting at the level of the host species. A solid comprehension of how these forces act together is paramount to understanding the impact of TEs on organismal biology. (Bourque et al. 2018:1)*

3.2.1 Origin of Transposons

After their discovery in the last 50 years, TEs have been detected in all eukaryotes, occurring in different compositions and classifications (Quesneville et al., 2005). But also, in prokaryotes, they play a crucial role in the activation and expression of genes by upstream insertions and in DNA rearrangements (Nagy & Chandler, 2004).

Their similarity to viruses especially parallels between retrotransposons and retroviruses in their Ability to integrate into host genomes using an enzymatic toolkit, suggest a phylogenetic relationship between transposons and viruses (Mustafin, 2018). Did viruses possibly evolve from primordial transposons? Have transposons incorporated into prokaryotic and eukaryotic organisms via viral vectors? Do transposons form a missing link between viral and cellular life forms?

Pourrajab and Hekmatimoghaddam (2021) explore these questions in their review Article subtitled *"from an arms race to a source of raw materials*." They locate the origin of TEs as a linker between "*the hypothetical primordial RNA world*" and viral and cellular organisms that emerged after that.

3.2.2 Role in Evolution and Diversification

Miousse et al. (2015) argue that the transposition activity of mobile genetic elements is a response to evolutionary stress. Stress is defined by Hoffmann and Parson (1997) as *"any environmental change that drastically reduces the fitness of an organism."* Bijlsma and Loeschcke (1997) also distinguished between:

a. Individual stressors: stressors evoke a physiological response, as in inflammation, oxidative stress, or ageing processes
b. evolutionary stressors: stressors evoke a change in phenotype or genotype with genetic implications.

Capy et al. (2000) take up evolutionary stress factors in their considerations and explain this using the example of antibiotic resistance in bacteria: Resistance genes are transferred horizontally with the help of flanking insertion sequences. In this context, the research group speaks of a "*domestication of TEs by the host genome*" (Capy et al., 2000). Coevolution of mobile genetic elements and host genome in the sense of symbiosis is also conceivable. Capy et al. summarize the relationship between transposition activity MGE and the evolution of host genome based on the studies of McClintock (1984), Hall (1997;1998; 2000) as follows:

1. Induction of transposition activity of MGE by stressors has been studied only in a few species.

2. Stress factors mainly affect the transposition activity of Class I transposons, which replicate in the genome with the help of an RNA mediator and reverse transcriptase.

3. The influence of transposition dynamics in somatic cells on the germline remains the subject of intensive research: "*Whether this somatic induction has effects on the next generation, both on somatic and germlines, remains uncertain. To find out, the dynamics of TEs within a species will have to be considered.*" This has to do with the fact that the activity of transposons can be passed on from one generation to the next, e.g., via the transcription factors involved for TE and the enzymes involved in the

cytosol of the oocyte: "*Therefore, a relationship may exist between the somatic activity of an element in a given generation and its germline activity in the following generations*." (Capy et al., 2000).

4. Stress factors in plants evoke local genetic changes in the respective organs (for example, on the sector of a leaf blade); in animals, the entire organism is apparently affected.

5. Stressors can cause short-term activity of MGE, but they can also have a lasting effect and remain active transgenerationally, as mentioned earlier.

6. The stress factors do not directly affect the genetic elements, but rather a stress situation leads to a shift in the activity of transcription factors. Since TE also have sequence motifs to which transcription factors can dock and which are similar to those regulatory sequences of genes involved in stress responses, retro/transposition is activated. In other words, a stress factor, such as an inflammatory response or oxidative stress, triggers proteolytic cascade, which results in the hyper- but also hypo-activity of activating or repressing transcription factors acting on DNA segments with corresponding signal sequences, including TE. This happens on an individual level for now but still has implications for population genetics.

3.3 Classification

The first hierarchical classification of transposons was proposed by Finnegan in 1989 and distinguished two classes of TEs. They are characterized by their transposition intermediate: RNA (class I or retrotransposons) and DNA (class II or DNA transposons). While the retrotransposition mechanism of class I is commonly referred to as "copy and paste", class II transposition is described as as "cut and paste" (Finnegan, 1989). This classification has been further refined and, especially to the class of retrotransposons, further subclasses have been distinguished depending on structure, retrotransposition mechanisms and gene regulation (Wicker et al., 2007).

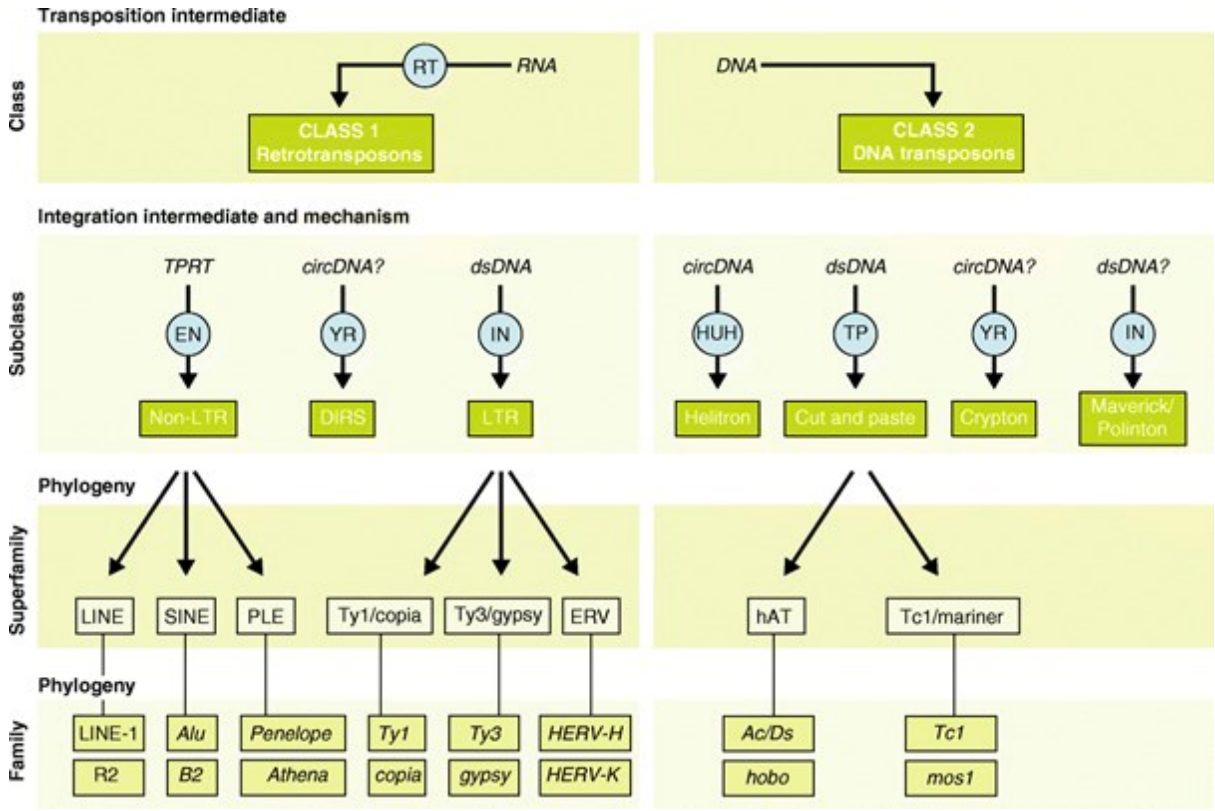


Fig.1: Classification of eukaryotic transposable elements (TE). Schematic and examples show key features and relationships between TE classes, subclasses, superfamilies, and families. Blue circles represent TE-encoded enzymes. *(Bourque et al., 2018)*

As mentioned above (Fig. 1), class I TEs transpose via an RNA intermediate. The RNA intermediate is transcribed from the genomic DNA. Afterwards, it is reverse

transcribed into DNA by a TE-encoded reverse transcriptase (RT); the cDNA is then reintegrated into the genome. Each replication cycle generates a new copy, and as a result, class I elements contribute most to the repetitive fraction in large genomes. These are divided into five orders (Makałowski, 2001: 181):

a. LTR retrotransposons (long terminal repeats),
b. DIRS-like elements,
c. Penelope-Like Elements (PLEs),
d. LINEs (long interspersed elements)
e. SINEs (short interspersed elements).

This classification is based on the mechanistic features, organization, and phylogeny of the reverse transcriptase of these retroelements. Inadvertently, the reverse transcriptase (RT) encoded by an autonomous TE can reverse transcribe another RNA present in the cell, e.g., mRNA, and generate a retrocopy of it, resulting in a pseudogene in most cases. LTR retrotransposons are named after the long terminal repeats (LTRs) ranging from several hundred to several thousand base pairs; this is reminiscent of (exogenous) retroviruses: "*Retroviruses and long terminal repeat (LTR) retrotransposons shared a common evolutionary ancestry, and therefore show several structural similarities*" (Naville et al., 2016: 313).

Both exogenous retroviruses and LTR retrotransposons include a *GAG gene* encoding a viral particle envelope and a *POL gene* encoding a reverse transcriptase, ribonuclease H, and an integrase that deliver the enzymatic machinery for reverse transcription and integration into the host genome (see Fig. 2). In contrast to LTR retrotransposons, exogenous retroviruses contain an *ENV gene that* encodes an envelope that facilitates their migration to other cells. Some LTR retrotransposons may contain rudimentary nonfunctional remnants of an *ENV gene* (Kazazian, 2004). This would tend to indicate that they are derived from exogenous retroviruses by losing the *ENV gene*. Kazazian (2004) suggest that LTR retrotransposons target their reinsertion to specific genomic sites, often around genes, with putative important functional implications for a host gene

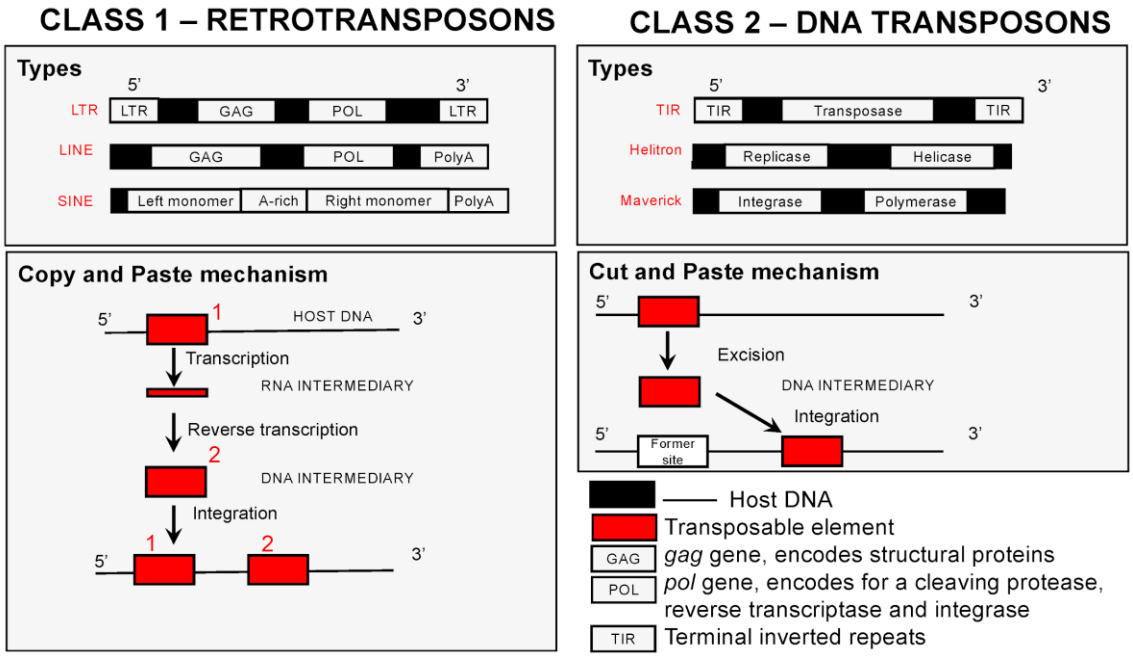


Fig.2: Graphical classification of Transposons: The left panel demonstrates Class 1 retrotransposons, and the right panel Class 2 DNA transposons. The upper panels summarize three examples of each of these two classes of elements. The lower panels summarize (retro-) transposition mechanism) of each class. *(Serrato-Capuchina & Matute, 2018)*

Lander et al. projected that 450,000 LTR copies make up 8% of our genome (Lander et al., 2001). LINEs do not have LTRs (Singer, 1982); however, they have a poly-A tail at the 3' end. They are flanked by target site duplications (TSDs), which are repetitive sequences (<10bp) at the insertion sites of TEs. They are thought to occur due to the filling in of sticky ends caused by transposase displaced editing. They flank transposable elements and can be used to find their loci in the genome. (Linheiro & Bergman, 2012; Gladyshev & Arkhipova, 2009). LINEs constitute about 21% of the human genome and among them L1 is the most abundant and best described LINE family with about 850,000 copies. L1 is the only LINE retrotransposon still active in the human genome (Mills et al., 2007). The majority of L1 in the human genome is inactive.

However, almost 80-100 retrotransposons have retained the ability to retrotranspose, with considerable variation between individuals (Ostertag & Kazazian, 2001; Erwin et al., 2014). These active L1s can disrupt the genome through insertions, deletions, rearrangements, and copy number variations (CNVs) (Kazazian & Moran, 2017). On the one hand, L1 activity has contributed to the instability but thus also to the evolution of the human genome. This research will likewise demonstrate that the transposition competencies of L1 in the germline are tightly regulated by epigenetic mechanisms for instance DNA methylation, histone modifications and RNA interference (Wang, 2017).

SINEs evolve from RNA genes such as those for 7SL RNA or t-RNA (Singer, 1982). By definition, they are short (max. 1kb) and do not encode their own retrotranscription tools; accordingly, they are considered non-autonomous TEs. For their transposition, they resort to the L1 machinery (Kajikawa & Okada, 2002). The most prominent example of this retrotransposon class in the human genome is definitely the Alu sequences (Schmid & Deininger, 1975), whose name derives from Ar*throbacter luteus*, a Gram-positive bacterium in which the Alu restriction enzyme was first described (Houck et al., 1979). Makalowski (2001) highlights its importance for evolutionary genetics as follows:

> *With over a million copies in the human genome, Alu is probably the most successful transposon in the history of life. Primate-specific Alu and its rodent relative B1 have limited phylogenetic distribution suggesting their relatively recent origins. The mammalian-wide interspersed repeats, by contrast, spread before eutherian radiation, and their copies can be found in different mammalian groups including marsupials and monotremes. (Makalowski, 2001: 185)*

Of particular note in this context are those TEs in hominids known as SVA elements, which are unique to the primate order because of their composite structure. They are named after their major components: SINE, VNTR (variable number tandem repeat) and Alu (Hancks & Kazazian, 2010):

> *SINE-VNTR-Alus (SVA) are non-autonomous hominid specific retrotransposons that are associated with disease in humans. SVAs are evolutionarily young and presumably mobilized by the LINE-1 reverse transcriptase in trans. SVAs are currently active and may impact the host through a variety of mechanisms including insertional mutagenesis, exon shuffling, alternative splicing, and the generation of differentially methylated regions (DMR). (Hancks & Kazazian, 2010: 234)*

Figure 3. summarizes those transposons that are found in mammals:

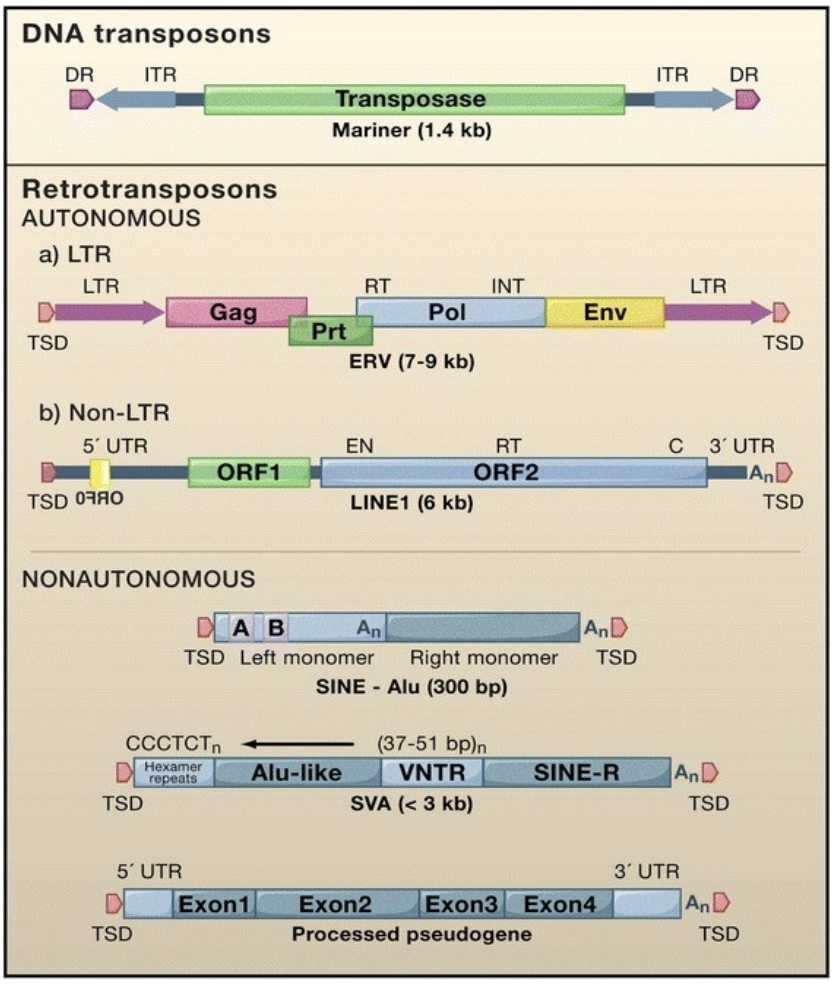


Fig. 3: Types of transposable elements in mammals. *(Goodier, 2016)*

3.4 Transposition Mechanisms

In the next sections, the two main catalytic proteins, reverse transcriptase (RT) and transposase are introduced.

3.4.1 Reverse Transcriptase (RT)

As mentioned earlier, **copy-and-paste** mechanisms of genomic retroelements require an RNA intermediate that is transcribed from the original immobile RNA transposon, then transcribed into DNA and reinserted at another site. This mechanism is catalyzed by a specific enzyme, reverse transcriptase (RT).

It was first discovered and isolated in viruses by Nobel laureates Howard Temin and David Baltimore in 1970 (Baltimore, 1970; Temin & Mizutani, 1970). Subsequently, RT has been particularly well studied in retroviruses (HIV-1, M-MLV, AMV). This is because both exogenous and endogenous viruses rely precisely on this molecule, the RT, for the integration and replication of their genetic material in a host genome (Coffin & Fan, 2016); this is therefore also encoded by RNA viruses.

However, the discovery of RT in the 1970s have countered the Central Dogma of biology, which posits a unidirectional hierarchy of DNA → RNA (as an unstable but functional intermediate) → proteins. A new perspective on the origin of life and the organizational form of the RNA world was opened. Mustafin and Khusnutdinova (2019) postulate that retroelements were at the universal beginning of life, whose activity was catalyzed by RT. The authors highlight the mutational capacity and insertion capacity of retroelements and take them as the basis for the emergence of complex DNA structures; according to them, RNA-mediated translation, which connects the nucleic acid world with that of the protein world, is a modification of the ribozyme activity of RT. Accordingly, reverse transcription is the link of the prehistoric RNA world and the present organization of life via DNA. According to Hughes (2015), RT, except for telomerase, is no longer active as an enzyme in eukaryotes and is associated exclusively with retroelements (retroviruses and retrotransposons) and hepadnaviruses. In the coming sections, the molecular structure of the RT protein and its genetic coding in retroelements will now be

presented, and the process of reverse transcription will be explained in more detail.

Molecular structure

The molecular structure of RT is best described using the corresponding enzyme of HIV-1, as this has been very well studied biochemically, genetically, and structurally in the context of research for an antiviral therapeutic as a therapy target.

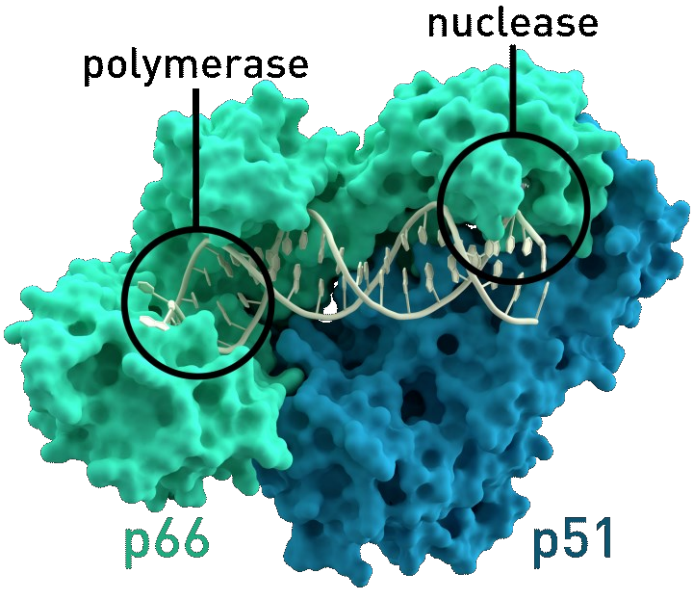


Fig. 4: Surface representation of the crystal structure of wild-type HIV-1 Reverse Transcriptase, based on PDB 3KLF. The active sites of polymerase and RNase are highlighted. *(Splettstoesser, 2014)*

The RT of HIV-1 consists of two subunits, each 560 AA (66k Das) and 440 AA (51k Das) long, respectively, also referred to as p66 and p51 according to their molecular weight. As can be seen in Figure 4, two functional domains in the p66 subunit are important: while RNA-dependent DNA polymerase produces a complementary cDNA single strand, RNA H-nuclease degrades the RNA template strand down to a few ribonucleotides, which are used as primers by DNA-dependent DNAP to complete the cDNA double strand.

The active site of the polymerase is composed of a triad of aspartic acids (D110, D185 and D186) in the p66 subunit, these three amino acids are associated with two Mg cations that catalyze the polymerization of nucleotides. The active site of

ribonuclease consists of another four acidic amino acids (D443, E478, D498 and D549), these amino acids with negative carboxyl group in the side chain are also associated with two Mg cations that catalyze the degradation of ribonucleotides.

Thus, while p66 encompasses both catalytic domains, p51 appears to have a structural function, according to Hughes (2015). On top of that, the author describes the HIV-1 RT as a "*small molecular machine*": with each step of polymerization, the molecule undergoes a conformational change. In addition, the synthesized DNA molecule is also actively rotated by the RT polymerase. For a better overview, researchers use the comparison of the molecular structure of the polymerase of the HIV-1 RT with a human hand: Fingers, Thumb, Palm and Connection, as shown in color in Figure 5 below.

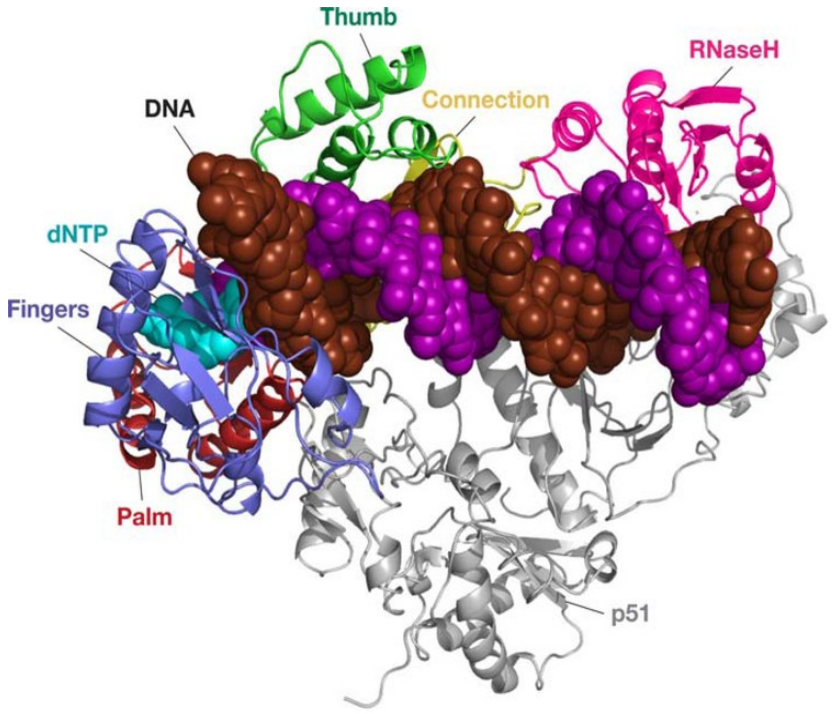


Fig 5: HIV-1 RT in complex with dsDNA and an incoming dNTP. The RNase H domain is shown in pink, and the four subdomains to the polymerase domain are shown in different colors: Finger, blue; thumb, green; palm, red; and junction, yellow. The DNA template strand (the strand that is copied) is dark red and the primer strand (the strand that is extended) is purple. The incoming dNTP is light blue. *(Hughes, 2015)*

Finally, some remarks on priming: after all, RT also needs a 3'-OH to bind new nucleotides. Here, DNA polymerase δ of the host plays a crucial role: this mistakes the RNA for a primer and constructs a double-stranded DNA similar to primer removal in the Okazaki fragment.

Reverse transcriptase in non-viral retroelements.

Also, in the case of non-viral retroelements, the reverse transcriptase is synthesized from a polyprotein that is post-translationally processed with the help of proteases. In most cases, the RT complex, together with two other enzymes, forms an ORF with a start and stop codon, which in the case of LTR transposons is also summarized as a POL gene (polypeptide/polyprotein). Within this polyprotein there are three functional sections:

(1) Proteases serve the enzymatic processing of the translated paraprotein
(2) Reverse transcriptase catalyzes the transcription of RNA into DNA and requires three motifs: a. RNA-dependent polymerase (which requires a t-RNA for priming), b. RNA H-nuclease to degrade the RNA template after the assembly of ss-cDNA, c. DNA-dependent polymerase to assemble the complementary DNA strand to dsDNA.
(3) integrase to insert the copied retroelement at a new site, it also catalyzes the double-strand break with sticky ends.

Figures 6 and 7 graphically illustrate the structural differences between LTR and non-LTR retrotransposons:

(1) LTR Retrotransposons (Fig. 6)

Fig. 6: Structure of an LTR retrotransposon. Typically, LTR retrotransposons contain a gag (group-specific antigen) and a pol gene. The pol gene translates into a protein that has activity for a reverse transcriptase, for an RNase H, for a protease, and for an integrase. *(Kubiak & Makałowska, 2017)*

(2) Non LTR Retrotransposons (Fig. 7)

Fig. 7: Structure of a non LTR rterotransposon, LINE elements: 5'UTR, ORF 1 (encodes RNA binding protein and linker sequence), ORF2 (encodes RT and an endonuclease, 3'UTR, polyadenyl tail. (Kubiak & Makałowska, 2017)

Various representatives of the LTR retrotransposons, such as the gypsy family, also possess a defective gene for an envelope protein (env for envelop). Thus, LTR retrotransposons comprise of the same elements as retroviruses, whereby the envelope proteins are either defective or deleted. They can be regarded as close related to retroviruses.

Because of the repetitive sequences, LTR elements can be easily excised from the genome, propagated, and reintegrated at arbitrary locations in the genome. Thus, the insertion of an LTR element can increase or decrease the activity of genes in the environment, alter tissue specificity, or generate new gene products through alternative splicing. Thus, they contribute significantly to genetic variability and thus to the evolution of organisms. 20 human genes are known to be controlled by viral LTRs. In total, at least 600,000 LTR elements could be found in the human genome. (Donner et al., 1999)

For instance, there are five amylase genes in the human genome. Two gens are active in the pancreas and three genes in the salivary glands. Although many mammals do not have an active amylase in their saliva, they still have the genes for amylase – but inactivated in a tissue specific manner- in the cells of the salivary glands. By integrating an LTR sequence of a retrovirus, it was possible to activate the amylase genes in humans in the salivary glands of the mouth as well, which improved the digestion of starchy food.

Retrotransposition by reverse transcriptase

Now, the process of retrotransposition using RT will be briefly explained, combining the considerations of Hughes (2015) on HIV-1 reverse transcription with those of Schorn and Martienssen (2018):

1. Prior to transposition, retroelement is transcribed: for this purpose, transcription factors bind to the promoter region of the retroelement (mostly in the 5'LTR) and RNA polymerase II forms an mRNA that encodes either both ORFs or only one ORF for a polyprotein that is later cleaved into the subunits after translation. The mRNA now serves as an RNA template for reverse transcription, comparable to the viral ssRNA.
2. Initiation: Lysyl-tRNA acts as a primer and hybridizes with the complementary primer binding site (PBS) downstream to the 5'LTR in the retroelement. The polymerase now synthesizes upstream from the 3' end of the t-RNA and forms a complementary cDNA copy of the 5'LTR.
3. Transfer: The newly synthesized minus strand of the 5'LTR is now transferred together with the t-RNA to the 3'end - here the two complementary LTR sequences hybridize, the 5'-section transferred at that time now offers a free 3'-OH as primer.
4. Minus strand synthesis: the RNA-dependent DNA polymerase now synthesizes the entire minus strand in the course of reverse transcription.
5. RNA degradation: In parallel, the RNA H-nuclease degrades the original mRNA (or viral RNA strand) - what remains is a nuclease-resistant polypurine tract just downstream of the 3'LTR. The PPP tract then serves as a primer for the assembly of the plus strand.
6. Plus, strand synthesis: the second strand is synthesized similarly with an intermediate step, in which the DNA-dependent DNA polymerase uses the PPP tract as a primer and now synthesizes the 3' and LTR region, thereby also building up the section that is complementary to the PBS. Now the t-RNA leaves the plus strand, and a transfer occurs again: the recently synthesized PBS of the plus strand hybridizes with the complementary PBS of the minus strand.

7. Bidirectional cDNA synthesis: while the polymerase is now completing the strand starting from +PBS, another polymerase is building the complementary LTR or UTR region at the 3' end of the minus strand starting from -PBS.
8. Transport: At the end of reverse transcription, two polymerases flank a complete cDNA double strand and transport it into the cell nucleus with the help of integrase (pre-integration complex). From HIV-1 RT research, we know that integrase eventually removes a few nucleotides at each of the 5' and 3' ends, creating sticky ends (Ellison et al., 1995).
9. Integration: Integration into the genome requires a palindromic signal sequence with a CA dinucleotide. A double-strand break with sticky ends catalyzed by the integrase occurs, where the transposon now fits in. If there is a gap, this is closed with conventional DNA repair mechanisms.

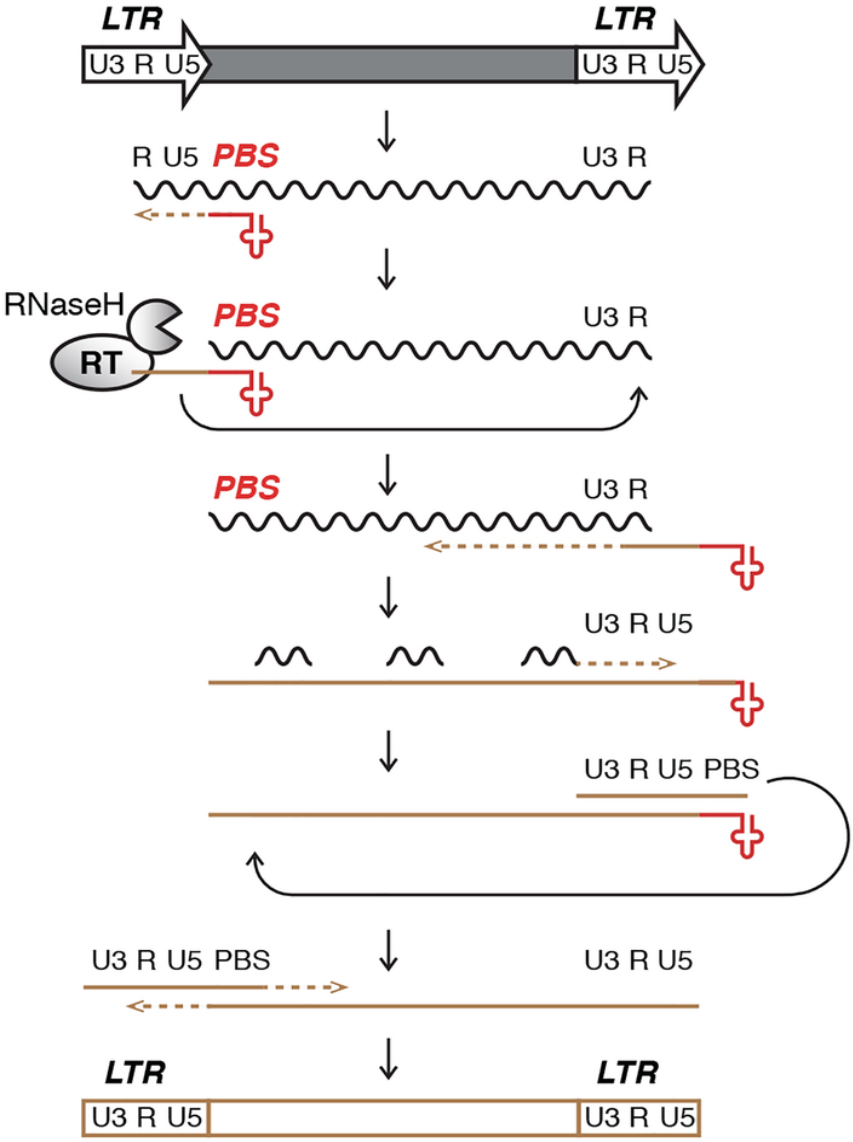


Fig. 8: Reverse transcription of retroelements: LTR encode promoter termination signals. The (m)RNA transcript contains a repeated region (R) at both ends and U3' and U5' respectively. The 3' end of cellular tRNAs (red) initiates reverse transcription by hybridizing to PBS. As this segment is copied into the (-) ssDNA (brown), RNase H activity of RT degrades the template RNA. The elongated cDNA is transferred to the 3' end of the retrotransposon transcript by hybridization with the R region. The remaining RNA is partially degraded by RNase H, leaving primers for second-strand cDNA synthesis. After another transfer event, first and second strand synthesis are completed to yield a full-length double-stranded cDNA that is integrated into the genome. *(Schorn & Martienssen, 2018)*

3.4.2 Transposase

The transposition mechanism of DNA transposons ("cut and paste") is catalyzed by the transposase (Tnp). Reznikoff (2002; 2003; 2008) describes the structure and function of this molecule using the Tn5 transposon as an example. The transposase (Tnp) Tn5 is found, for example, in E. coli, in prokaryotes DNA transposons play a role in antibiotic resistance, as they sometimes (co-)encode and mobilize resistance genes (Reznikoff, 1998). Reznikoff (2003) describes the Tn5 transposase, which is 476 amino acids long, as rather inactive; he reasons that this is due to the folding of the molecule, as the N- and C-termini are close to each other and tend to inhibit each other. However, mutations such as L372P, in which a leucine is replaced by a proline at position 372, change the alpha-helix structure at the C-terminal domain, which is now separated from the N-terminus enough to increase the activity of the transposase.

Structure of the transposase

The central catalytic motif for a number of transposases is the DDE motif that catalyzes transposition of the transposon. These Tnp are then grouped together to form DD[E/D]: Aspartate-97, Aspartate-188 and Glutamate-326 forming a triad of acidic residues and thus the active site. The DDE motif (blue in Fig. 9) is thought to coordinate divalent metal ions, mostly magnesium and manganese, which are important for the catalytic reaction.

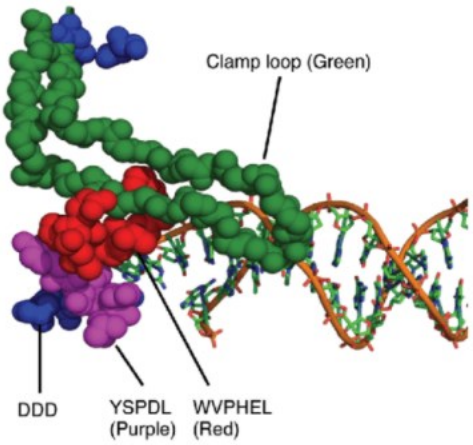


Fig 9: Structure of transposase exemplified by Mariner transposase, which has a DDD motif (blue). (Bouuaert et al., 2014)

Transposition

According to Reznikoff (2003), transposition (Fig. 10) by transposase proceeds as follows:

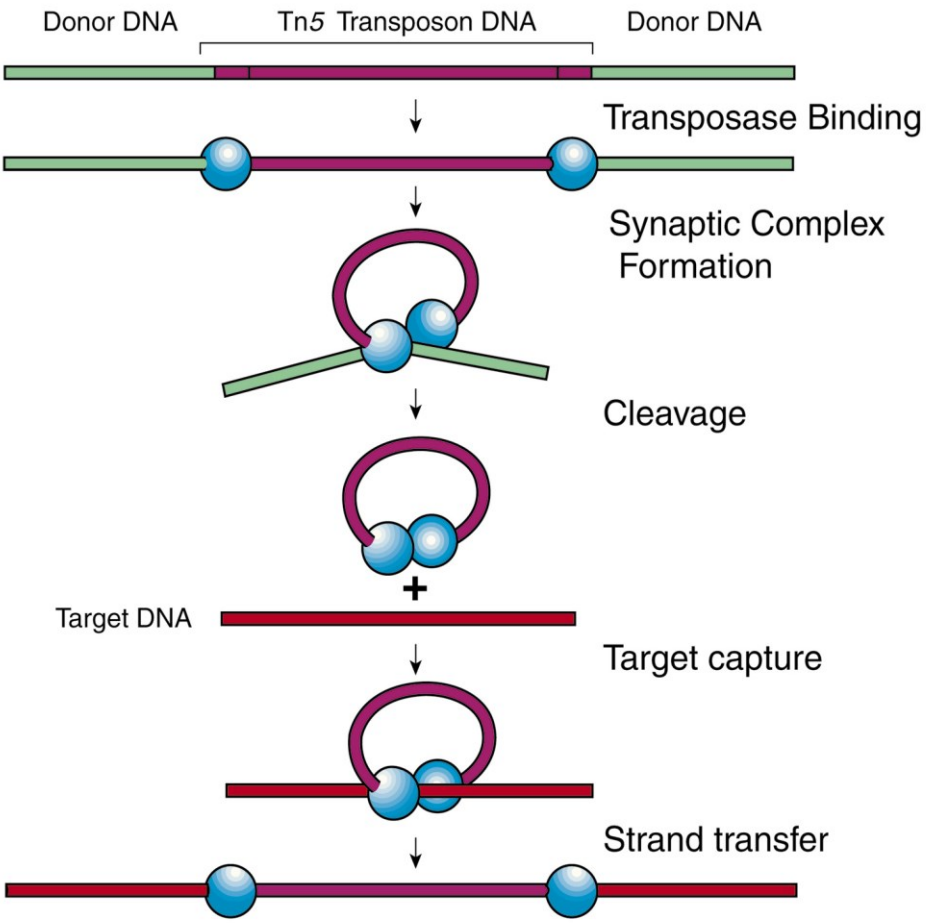


Fig. 10: Transposition (Image credit: **© 2000 Science**)

1. **Transposase binding:** formation of the synaptic complex in which two Tnps bind to the 19bp sequences. Next, amino acids 26-65 form specific contacts with DNA terminal positions 6-17 (Reznikoff, 2002).
2. **Synaptic Complex Formation:** Dimerization of the Tnp-end DNA complex to form the synaptic complex. The two C-terminal α-helices cross and form a protein-protein contact. Positions 1-9 of the terminal DNA molecule are inserted into the active site of the partner monomer, forming multiple

'trans' Tnp-DNA contacts with residues near the active site containing the critical DDE residues (Reznikoff, 2002).

3. **DNA cleavage (cleavage):** three catalytic steps are involved in DNA cleavage: a. 3′-strand-nicking, b. hairpin formation, and c. Hairpin cleavage (Reznikoff, 2002). The DDE residues in the active site of Tnp coordinate two Mg++ ions, these are necessary to cleave oxygen atoms in the P-O bond between nucleotides, (Reznikoff, 2002). A hairpin structure is then formed, which is accompanied by terminal denaturation of the base pairs in the interface.
4. **Target capture and strand transfer:** the synaptic complex is after the, to bind to the target DNA by target capture. The 3′-hydroxyl group of the transposon end attacks the phosphodiester backbone of the target DNA during strand transfer. Due to the staggered strand transfer reactions and subsequent DNA repair by the host, a 9-bp duplication occurs in the target.
5. Disengagement of the strand transfer complex: After strand transfer, the polymerase of the host sequence fills the two 9 bp gaps at both ends of the integrated gaps.

Although according to Nesmelova and Hackett (2010) the human genome encodes over 300,000 DNA transposons, none of the transposase molecules is active. According to Reznikoff (2003), this is mainly due to cumulative inactivating mutations that alter the catalytically important centers of the enzyme.

However, four changes to the primary sequence of the inactive transposase could restore its former activity:

1. L372P: as earlier mentioned, this amino acid exchange leads to a conformation change in which the C- and N-termini no longer inhibit each other. The C-terminal dimerization domain tends to be separated from the DNA binding domain at the N-terminus.
2. E54K: enhances the initial binding of Tnp to the 19 bp DNA. Other mutations near position 54 also enhance end DNA binding.
3. E110K /E345K: The residues E110 and E345 interact strongly with the metal ions. A change from the rather acidic amino acid GLU (E) to the basic amino acid LYS (K) improves the catalytic effect of the metal ions on DNA.

4. P242A/ P242G: Proline 242 has a much stiffer backbone than alanine or glycine, amino acid replacement at this position increases the flexibility of the molecule.

These changes to the primary sequence of the transposase can be used to artificially generate transposable molecules. These can then be used in genetic engineering, as explained later using the example of library preparation by transposase tagmentation for next generation sequencing. The study of the transposase Tn5 is of general importance because of its similarities to HIV-1 and other retroviruses. By studying Tn5, much can also be learned about other transposases and their activities.

Finally, it should be noted that DDE transposases are among the most important catalytic tools for insertion sequences (IS), i.e., prokaryotic TEs (Nagy & Chandler, 2004).

3.4.3 Effects of (Retro-) Transposition

Zeng et al. (2018) cite several potential effects of transposition or retrotransposition of mobile genetic elements on the human genome:

1. Transposition into coding exonic regions: lead to **nonsense mutations or** to truncation or loss of function of the protein.
2. Transposition in which parts of coding exonic regions are transposed: Lead to inactivation or **knock-out of** the affected gene.
3. Transposition into regulatory sequences in the promoter region or UTR: alteration of **gene regulation** leads to different gene expression patterns.
4. Transpositions in splice sites: **Exon skipping** or aberrant splicing, or **splice mutations.**
5. Insertion of transposons that transpose exons of other genes: **Exon shuffling** and addition of new exons.
6. Homologous recombination between sequence similar transposons on homologous chromosomes: non-equivalent crossing over leads to **duplication of** the region between transposons in one chromosome and deletion of the section on the other chromosome.
7. Homologous recombination between two sequence similar transposons of the same orientation: crossing over between the direct repeats leads to looping and **deletion of** the genetic segment between the transposons in the form of circular DNA.
8. Homologous recombination transposons of opposite direction: hairpin structure and **inversion of** the region between transposons.
9. Excision of transposons with strand breakage and defective DNA repair: **chromosome breakage in** which segments of chromosomes may be lost or translocated to other sites (inversely), i.e., **chromosome rearrangement**.
10. Transposition of promoters upstream of the gene or of sequences upstream of termination signal: **emergence of new genes.**

Figure 11 provides a representation of the aforementioned effects of transposition on the genome:

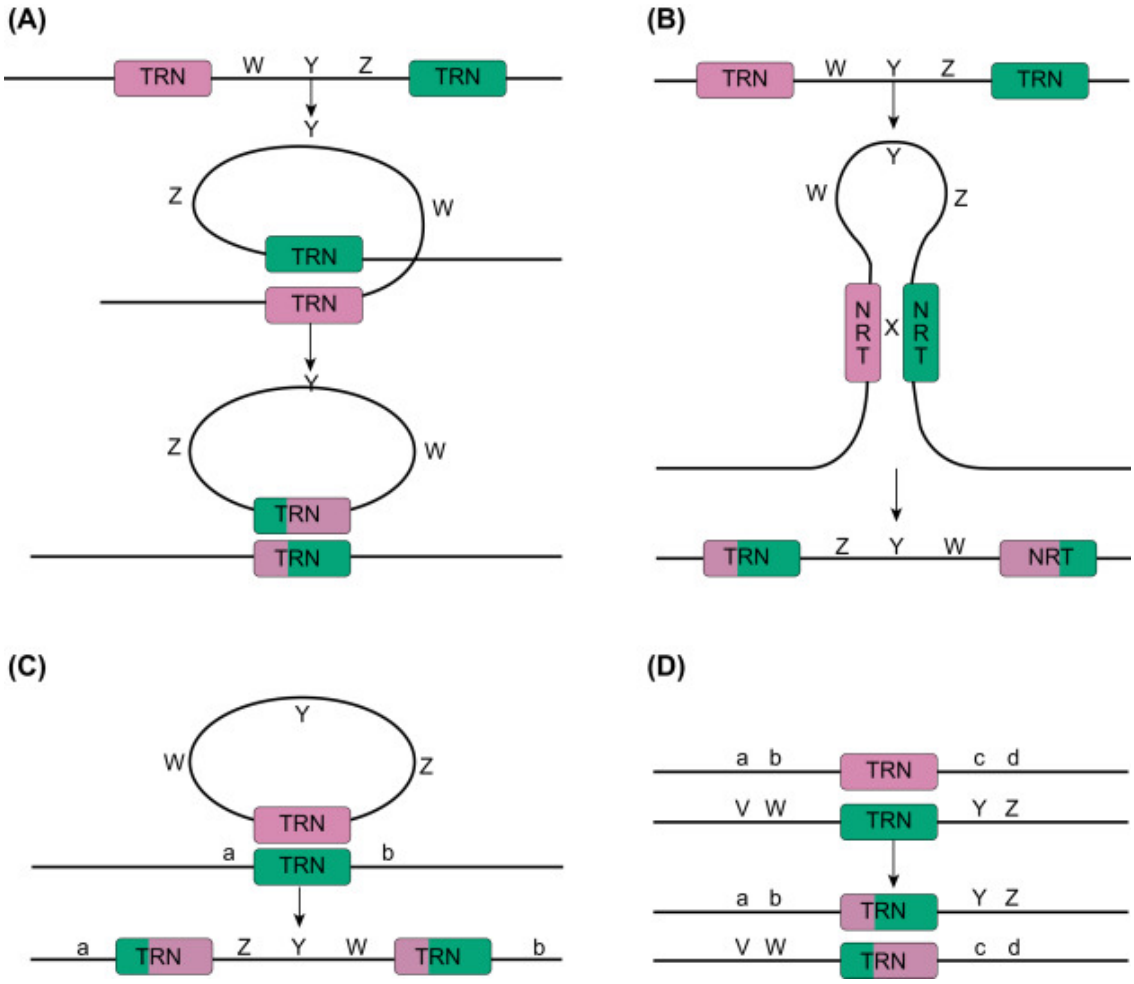


Fig. 11: The figures show how transposition and recombination lead to chromosome rearrangements. (A) shows a chromosome loop resulting from the meeting of two transposons and subsequent recombination. The recombination results in a deletion. (B) shows recombination between two inverted transposons and following recombination, that results into an inversion in the chromosome sequence. (C) shows a ring chromosome with a transposon and a linear chromosome with a transposon. Recombination between the transposons results in an insertion within the linear chromosome. (D) shows two non-homologous chromosomes (abcd and vwyz). Recombination between the TRN sequences results in a chromosome translocation where a and b are now linked to y and z, and v and w are now linked to c and d, respectively. *(Herrera et al., 2016)*

3.4.4 Epigenetic Regulation of Transposition

The effects of retro/transposition can thus enable diversification under evolutionary stress, as will be shown later in the example of RAG1/2 activity in somatic recombination of immunoglobulins. On the other hand, retro/transposition can also be detrimental to the integrity of the (host) genome, leading to genetic aberrations and even tumorigenesis. Consequently, the effects of retro/transposition described above must be well regulated by the organism and repressed especially in pluripotent stem cells and in the germline.

The central control mechanism is the epigenetic modification of the TE or possible transcription and translation products. Three mechanisms (Fig. 12) will be briefly explained in the following:

1. Hypermethylation of the CpG-rich signal sequences of TE.
2. Histone-based repression by modification of the chromatin structure
3. RNAi silencing by piRNA and PIWI proteins.

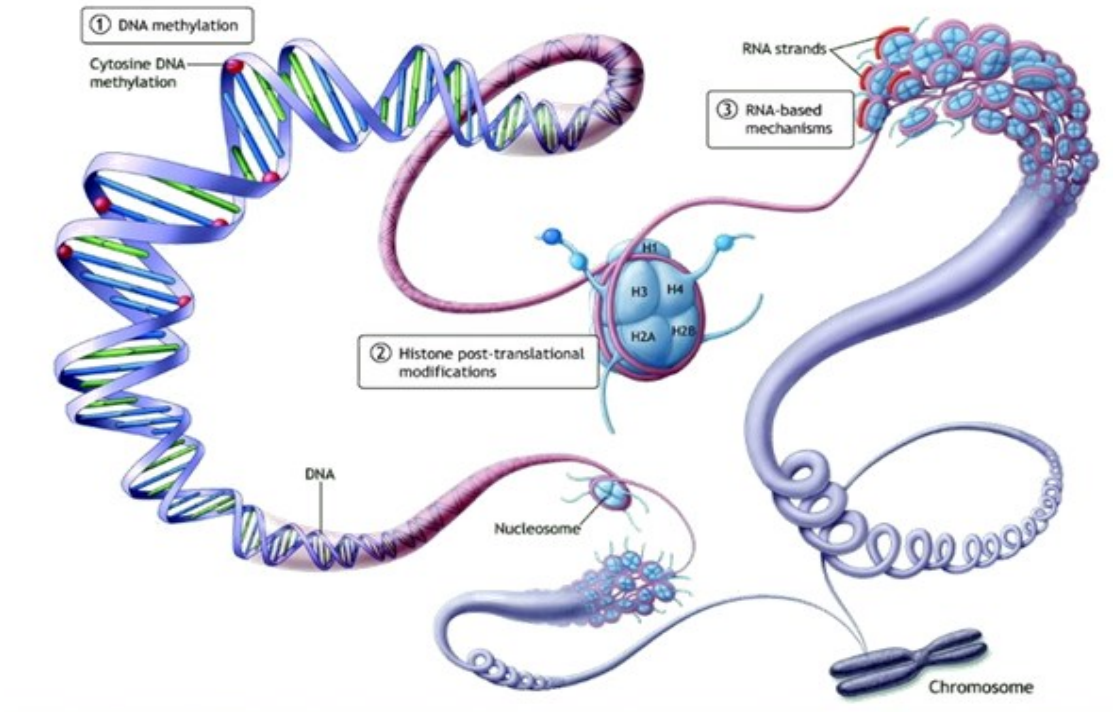


Fig. 12: Overview of epigenetic regulatory mechanisms (Matouk & Marsden, 2008).

(1) DNA methylation of CpG motifs of TE can be explicated at retrotransposons, respectively LINE-1 elements and Alu sequences. According to Baba et al. (2013), up to 30 CpG islands can be identified in the 5'-UTR region of LINE-1 elements. These are not equally hypermethylated in each tissue type. In tumor tissue, LINE-1 is clearly hypomethylated. Here, the authors compared the different tumor entities and conclude that although the methylation level of LINE-1 correlates with tumorigenesis, there is still a great deal of heterogeneity, with the consequence that the methylation pattern of TE alone is not a meaningful tumor marker. Baba et al. (2014) describe the implications of hypomethylation using colorectal cancer as an example:

> *The relationship between LINE-1 hypomethylation and early-onset colorectal cancer is intriguing. Early-onset colorectal cancer presents a clinically distinct colorectal cancer phenotype and is often associated with an unfavorable prognosis. LINE-1 extreme hypomethylators (methylation < 40 %) appear significantly more frequently in younger than in older patients. (Baba et al., 2013: 1809)*

In addition to LINE-1 elements, Alu-Sequences are also methylated in the genome. Malousi and Kouao (2012) conclude that their methylation depends on their role in gene expression or in the splicing process after transcription. Alujb, AluSx, and AluJo (Sorek, 2002; 2004) in particular play critical roles in alternative splicing, so it is not surprising that their CpG motifs are often hypermethylated and thus prevented from possible retrotransposition by epigenetic silencing.

Changes in the methylation pattern may be indicative of pathological neoplasia. Cho et al. (2007) explicate the correlation between hypomethylation of CpG inserts in repetitive noncoding sequences and tumorigenesis. Using methylation-specific PCR (MSP) and combined bisulphite restriction analysis (COBRA), they investigate CpG islands at 22 candidate genes for prostate cancer. They focused LINE-1 and Alu sequences and amplified them specifically. Indeed, they found a correlation between the methylation level of mobile genetic elements and the concentration of the prostate specific antigen (PSA) marker: *"Alu and LINE-1 methylation levels were significantly lower in the high PSA group than in the low PSA group"* (Cho et al., 2007:272).

In summary, genome-wide hypomethylation can be considered as one of the hallmarks of tumorigenesis. Repetitive sequences, such as retrotransposons, are also affected. Their deregulation leads to increased transposition, resulting in chromosomal breaks and rearrangements. Hypomethylation of SINE and LINE results in whole-genome instability. In their conclusion, Ogino et al. (2008) therefore suggest that the degree of methylation of LINE-1 sequences should also be used as a prognostic marker in the therapy of carcinomas. Kurkjian, Kummar, and Murgo (2008) go one step further and recommend CpG motifs of transposons, similar to methylation of promoters of hyper-expressed oncogenes, also be investigated as therapeutic targets and with the help of re-methylation their transposition activity in colorectal carcinomas be suppressed. Baba et al. (2014) also suggest the same for gastrointestinal cancer:

> *In contrast to irreversible genetic changes, epigenetic changes may provide potentially reversible molecular targets for both cancer therapy and chemoprevention. Further investigations in this field would provide deeper insights into the pathogenesis of GI cancer and assist in the development of new therapeutic strategies against these cancers. (Baba et al., 2013: 1813)*

As shown above, the methylation of CpG dinucleotides in repetitive sequences is one of the epigenetic silencing methods. Under certain conditions, post replicative methylation of cytosine into a 5'methylcytosine by DNA methyltransferase is followed by deamination of cytosine to deoxyuridine. Upon repair, the transition to deoxythymidine subsequently occurs. The C-to-T transition leads to the loss of CpG islands and thus to the loss of signal sequences in promoter regions necessary for transcription and retrotransposition (Zhou et al., 2020). At the same time, TEs take over the function as regulatory sequences for other proteinogenic segments. In the Zhou et al. (2014) model, they act mainly as enhancers; the authors also argue that methylation and deamination of CpG motifs in TE "*allows for genome expansion and also leads to new opportunities for gene control by TE-based regulatory sites*" (Zhou et al., 2020: 19359). They also use this to explain the C-value paradox in eukaryotes: after all, the enormous genome size in humans is also due to the previously intact retro/transposition activity of MGE, which is repressed with the help of DNA methylation, for example.

However, not all CpG islands are deaminated in the course of methylation, nor are all remaining CpG islands silenced by methylation at all times.

The examples of DNA methylation of TE mentioned above primarily affect somatic cells and, if deregulated, lead to genetic instability that ends in apoptosis or neoplasia, but which is not passed on to the next generation because germline cells are not affected. Only changes in germline cells, such as replication and insertion of retrotransposons, can be passed on to the next generation. The consequence of DNA demethylation of TE in the germline are more severe: since there is a genome-wide loss of DNA methylation in early embryonic development, other mechanisms of epigenetic control of TE have to be taken.

Indeed, loss of the epigenetic methylation pattern occurs during embryogenesis, but also in stem cells. Studies and experiments by Walter et al. (2016) showed that as the loss of DNA methylation progresses, transposons are initially reactivated but later silenced by other mechanisms like the disappearance of DNA methylation by stimulating histones around transposons to condense, preventing TEs from transposing. In this context, Walter et al. speak of "*histone-based repression strategies*":

> *This was accompanied by a reconfiguration of the repressive chromatin landscape: while H3K9me3 was stable, H3K9me2 globally disappeared and H3K27me3 accumulated at transposons. Interestingly, we observed that H3K9me3 and H3K27me3 occupy different transposon families or different territories within the same family, defining three functional categories of adaptive chromatin responses to DNA methylation loss. Our work highlights that H3K9me3 and, most importantly, polycomb-mediated H3K27me3 chromatin pathways can secure the control of a large spectrum of transposons in periods of intense DNA methylation change, ensuring longstanding genome stability. (Walter et al., 2016: 1)*

(2) Thus, **histone modification** is the second form of epigenetic control of TE. Histone methyltransferase (SETDB1) suppresses transposons via methylation of H3K9me3 (Castro-Diaz et al., 2014). In addition to H3K9me3, other modes of epigenetic suppression of TEs include methylation of H4K20me3, H3K27me3, and H4R3me2; biotinylation and sumoylation of H2A, H3, and H4 histones; and deposition of the histone variant H3.3 (Leeb et al., 2010).

He et al. (2019) show via Western blot studies that TEs can be marked by chromatin modifications. They found evidence not only for repressive labeling but also for widespread labeling of TEs by activating labels, including bivalent labeling of TEs by repressive H3K9me3 and activating H3K27ac modifications. Loss of the catalytic factors such as SETDB1 mentioned at the beginning of this article led to changes in the expression of TEs and a corresponding change in chromatin accessibility.

The methylation pattern of DNA but also the regulatory histone code is reprogrammed in the germline during gametogenesis as well as in early embryogenesis. To control the activity of TE in the genome even under these circumstances, we rely on an RNAi mechanism: the PIWI/piRNA pathway (Russell & LaMarre, 2018).

(3) RNA interference (RNAi) is the third way to efficiently suppress TE mobility. RNAi is an RNA-dependent gene silencing process controlled by the RNA-induced silencing complex (RISC), initiated by short (double-stranded) RNA molecules in the cytoplasm of a cell, where they form a complex with the catalytic RISC proteins, the Argonauts, and address the target RNA (Carthew & Sontheimer, 2009). This can then subsequently be enzymatically cleaved, or prevented from translating, by blocking the UTR regions, making interaction with ribosomes impossible (Roberts, 2015). The short RNAs involved can be divided into three classes: small interfering RNAs (siRNAs), microRNAs (miRNAs), and PIWI-interacting RNAs (piRNAs). While the first two are quite well studied, many detailed questions about piRNA have only been clarified in recent years (Fig. 13).

Studies of small RNAs in the germline of Drosophila melanogaster were crucial in the elucidation of this phenomenon. As early as the 1980s, these studies revealed that mutations in a specific gene segment activate transposons, i.e., make them "*dance*": the gene segment was subsequently named Flamenco loci, and the transposons Gypsy (Guida et al., 2016). Another example of transposon regulation in Drosophila melanogaster is the P element. Basically, P elements are DNA transposons (3kb) that encode a transposase (from four exons) in wild type (Majumdar & Rio, 2015).

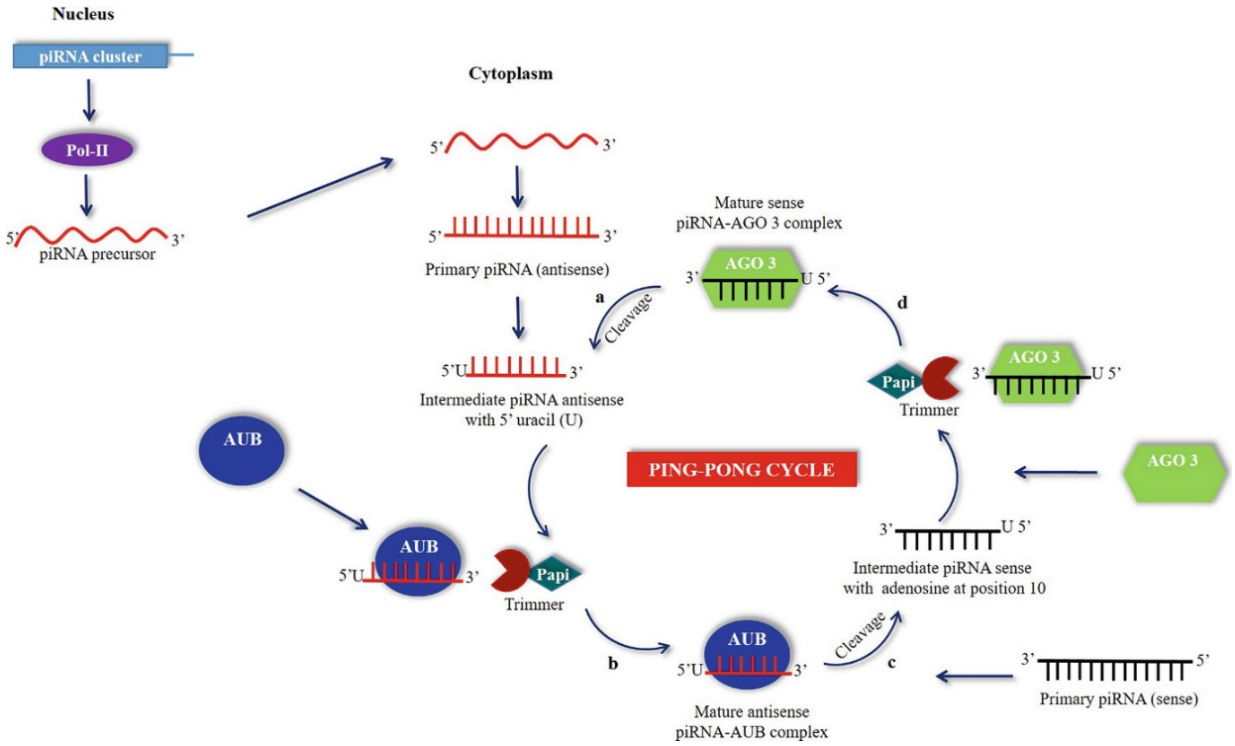


Fig. 13: piwi-interacting RNAs (piRNAs) are a class of RNAs discovered in 2006 that are slightly longer than miRNAs and siRNAs (26-31 nt). As implied by their name, these RNAs bind to PIWI proteins and are found mainly in germ cells (germline), where they are essential for spermatogenesis. Among other things, they are involved in gene silencing of retrotransposons. (Calcagno et al., 2019)

4. Methods

This Master thesis is designed as a literature review, summarizing the current state of understanding, surveying previously published primary and secondary scientific literature on the topic of mobile genetic elements. Thus, this review roughly consists of two sections: After the introduction and presentation of the already mentioned research questions, a localization of these questions in the most current research on transposons would follow. The focus was on the different subclasses of TEs (Chapter 3). For this purpose, not only the history of research was emblematically reviewed, but also findings on questions of transposition mechanisms, as well as a recent classification and overview of regulatory mechanisms, especially of epigenetic nature, emanating from retrotransposons to the genome and vice versa were presented. The main catalytic players of RNA transposons, reverse transcriptase (3.4.1) and the transposase (3.4.2) of DNA transposons were expounded.

Chapter 5. will now present the arguments for a constructive but also destructive implication of transposons on the human genome briefly addressed earlier. In this chapter, the dichotomous role of transposons is to be rescored in order to be discussed in more detail in the following sections.

Chapter 6. is devoted to those evolutionary implications of the mobilization of genetic elements and explains the genetic diversification of immunoglobulins due to transposon-derived enzyme complexes. Chapter 7 deals with the relationship between transposons and gene expression, this is mainly discussed in chapter 8 in relation to tumorigenesis. Chapter 9 gives examples of the utilization of Transposons in oncogenomic research.

5. Results

After the classification of transposons, their structure and the different (retro-) transposition mechanisms have been presented in chapter 3, the overall research question, which has already been posed in the subtitle of this review, should now guide the further considerations. **Are transposons beneficial jumpers of evolution or are they junk DNA pathogenic parasites of a viral past?**

The answer to this question is not simple and impossible for two reasons. First, these attributions of "beneficial" or "parasitic", but also the designation "junk DNA" or "selfish" are primarily normative, i.e., evaluative and moralizing, ascriptions. They originate from anthropological or anthropomorphic metaphors and are widely used in biology. There, they cause more harm than supposed benefit - which would be the illustration of complex interrelationships. For evolution is not good or bad. Genes are neither sick nor healthy - they are primarily causative for traits. Second, the opening question is binary - it implies an “either-or”. However, the examples in the next chapters show that (retro-)transposition and repression of TEs in the human genome are in a fragile homeostasis and this dynamic equilibrium is very specific for each differentiation stage and each tissue.

It would be better to ask the question in a more global and open-ended way: ***What role did transposons play in the phylogenesis of homo sapiens, what genetic and proteinic structures derive from mobile genetic elements? How does their presence or absence affect chromosomal arrangement and gene expression? What happens when the fragile balance between transposition and repression is disturbed? What triggers this de-regulation and what are the implications for the affected tissue?***

Having already outlined the possible origins of TEs and their contributions to the evolution of viral and cellular organisms in the theory section, we will now briefly review the implications of TEs specifically in the human genome. To do this, it makes sense to look at different levels of interaction between transposons and host organisms. For although some TEs, such as Alu elements are no longer autonomically transposable, LINE-1 Retroelements are still capable of retrotransposition.

At the **genetic level**, transposon activity allowed existing genes to be overexpressed by the addition of new regulatory elements (such as promoters and enhancers). Transpositions also result in duplication of genes, which can lead to the formation of pseudogenes (Pavlicek & Jurka, 2021). They also affect the internal architecture of the transcribed mRNA of proteins - Ule (2013) suggests that intronic mutations acting on Alu elements lead to the formation of new exons, a process called exonization (Sela et al., 2010).

At the **genomic level**, i.e. at the chromosomal structural level, which encompasses more than just coding regions, transposons have strongly influenced the organization of chromosomes, mainly through their repetitive homologous sequences (Bousios et al., 2020). Especially "topologically associated domains" (TADs), i.e. three-dimensional chromosome structures that self-interact with each other, are separated from each other by transposons (Dixon et al., 2006). Furthermore, other chromosomal structures such as telomeres (Casacuberta, 2017) but also many centromeric regions (Talbert & Henikoff, 2020) arise from transposons and still carry typical TE features in their interaction with enzymes or RNAi.

At the **transcriptome level**, transposons have mainly led to the diversification of RNAi: Pourrajab and Hekmatimoghaddam (2021) suggest that retrotransposons have mainly contributed to the creation of lncRNA and mi RNA. DNA transposons, on the other hand, are the source of dsRNA, siRNA, piRNA.

At the **proteome level**, certain transposon-derived structures can also be found in human cells (Koch, 2021). Thus, there are obvious parallels in structure and function between the reverse transcriptase of retroelements and telomerase: both can convert and incorporate DNA via an RNA intermediate. But also, from the transposase new enzymatic structures have emerged in eukaryotes and prokaryotes- excitingly in some analogy, maybe even homology. In both, enzymatic catalysts of the adaptive immune system trace back to original transposases: the CRISPR/Cas system in prokaryotes (Klompe et al., 2019) appears to have evolved from mobile genetic elements called casposons (Krupovic, Béguin and Koonin, 2017). In eukaryotes, the RAG (Recombination Activating Genes 1 and 2) enzyme complexes, which are responsible for the diversification of

immunoglobulins, trace back to a common transposon ancestor that likely formed the transib transposon in all deuterostomes (Fugmann, 2010).

However, even their sole presence in the genome, whether or not they are autonomously transposable, has enormous effects on the expression of other genes. Häsler and Strub (2006) were able to show that Alu elements, which are found over a million times in the human genome, can play a crucial role in post-transcriptional processing of mRNA and thus also play a decisive role in splicing.

Thus, it is clear that transposons have played - and continue to play - a key role in evolution: TEs are highly active in embryogenesis due to the fact that DNA methylation is suppressed and also have a crucial function for pluripotent stem cells. It is becoming increasingly clear that their transposition capacity, which could threaten genomic stability, is in a **fragile balance** with the capacity of the (human) genome to control them. If this equilibrium displaces, it causes genetic instabilities (Burns, 2020).

Two factors can induce this shifting: environmental, exogenous factors such as evolutionary stress - but also endogenous genetic factors, they include for example (somatic) mutations that activate transposons or epigenetic dysregulations that reawaken transposition capability (Callinan and Batzer, 2006). The result of this newly awakened activity of transposons can be fatal and almost always leads to genetic malignancies - respectively cancer.

However, this is where a better understanding of TEs and transposition assists - TEs can be used as diagnostic, prognostic, pharmacogenetic and predictive biomarkers. In oncology, they can even be specifically addressed as therapeutic targets and, for example, their proteins, respectfully reverse transcriptase, can be inhibited. But transposons can do more than that: they can also be used as therapeutic tools for non-transposon-derived neoplasms, such as hereditary tumor syndromes caused by loss-of-function in tumor repressor genes. In combination with CRISPR/Cas, transposons, especially the Sleeping Beauty transposon, are a powerful weapon against cancer and an innovative tool in cancer research.

6. Transposons in Diversification of Immunoglobulins.

Miller et al. (1997) already speak of "molecular domestication" of transposons and refer on the one hand to cis-regulatory regions that arise from the insertions of TEs, and on the other hand to trans-acting factors. This domestication process is particularly impressive in the example of TE-derived proteins (Jangam, Feschotte & Betrán, 2017).

The most vivid example is the molecular diversity of immunoglobulins generated by B and T cells. Brandt and Roth (2007) refer to this (almost) endless diversity as "**G.O.D.'s Holy Grail**," referring to the "generation of diversity"- because immunoglobulins are encoded on only three gene loci: the h-chains on chromosome 14, the κ-chain on chromosome 2, and the λ-chain on chromosome 22- but each of these gene loci comprises only one to a few megabases (Matsuda et al., 1998). However, through specialized recombination, these gene segments can form 10^{11} or more immunoglobulins (Janeway et al., 2014). The extent of the G.O.D.-question is especially apparent when compared with another protein system specialized for the recognition of chemical substances: the olfactory receptors. This is because, unlike the limited three gene loci that together make up all the myriad immunoglobulins, each olfactory receptor is encoded by its own gene. In the mammalian genome, with over a thousand genes, they can account for up to 3% of the genome (in mice, there are over a thousand, and in *Ratus novergicus*, there are as many as 1200, according to Niimura (2009)). In humans, about 400 genes encode olfactory receptors, while 600 are pseudo genes. However, unlike the immune system, where the fit to antigens must be very accurate, diversity in olfactory receptors is increased via two additional features: 1. receptors can recognize different odors and are not limited to only one chemical substance, 2. conversely, odors can stimulate more than one type of receptor (Araneda et al., 2004). Immunoglobulins, on the other hand, must respond in a precise matching manner to the antigen in order to provide appropriate protection: they achieve this via in situ somatic recombination, which will be explained in more detail in the following chapter.

But first things first: in 1987, Susumu Tonegawa was awarded the Nobel Prize for his discovery of "the genetic principle for generation of antibody diversity" (Hozumi & Tonegawa, 1976). Together with his team, Tonegawa was already able to prove in the 1970s that rearrangements of the DNA occur during the formation of antibodies by B cells. The genes for the individual segments are rearranged in the process of somatic recombination (in T and B cells), similar to a combination lock. However, it was David Schatz who discovered the enzymatic background of this somatic recombination a year later, in 1988: the recombination activating genes encode two proteins RAG 1and RAG2 that catalyze V(D)J recombination in cells of the immune system (Schatz, Oettinger & Baltimore, 1989).

But the story does not end here, a few years later Gellert and Mizuuchi realized that the interaction between the RAG enzyme complexes, and the immunoglobulin gene loci strongly resemble the hydrolysis and transesterification reactions catalyzed by DNA transposons-but also retroviral integrases (van Gent, Mizuuchi and Gellert, 1996).

This has been sensational insofar as transposons were considered a few decades ago - under the collective term endogenous viruses - together with viruses as parasites. The fact that precisely that organ system, that fights against pathogens, whether endogenous or exogenous, is derived from a domesticated mobile genetic element, should not be surprising (Hiom, Melek and Gellert, 1998). The same is indeed true for prokaryotes - as described in more detail later in Chapter 9, the Cas protein responsible for the adaptive immune response of bacteria to bacteriophages, together with CRSIPR RNA, also traces back to domesticated transposases: Casposons (Koonin & Makarova, 2019) .

In the literature, this host-transposon relationship is often referred to as the "evolutionary arm race" (Ozata et al., 2018):

> *Transposable elements (TEs) are mobile DNA sequences that colonize genomes and threaten genome integrity. [...]the evolutionary success of TEs cannot be explained solely by evasion from host control mechanisms. Rather, some TEs have evolved commensal and even mutualistic strategies that mitigate the cost of their propagation. These coevolutionary processes promote the emergence of complex cellular activities, which in turn pave the way for cooption of TE sequences for organismal function.(Cosby, Chang and Feschotte, 2019)*

6.1 Adaptive immune system

The adaptive immune system includes B and T lymphocytes and the effector molecules they produce. Two essential properties characterize the adaptive immune system: The cells possess clonal receptors, and they form the immunological memory. Congenital or acquired immunodeficiencies in which essential elements of the adaptive immune system are absent or dysfunctional are fatal if left untreated.

Our survival depends on our lymphocytes being able to recognize and eliminate every conceivable pathogen. At the same time, the immune system must not damage our own organism. The adaptive immune system must therefore fulfill two contradictory conditions: maximum diversity and avoidance of autoreactivity. The basic molecular mechanisms for generating diversity of the B and T cell receptor repertoires are the same.

B- and T-cell receptors are heterodimers consisting of a heavy and a light chain. The light chains consist of three segments called constant, joining, and variable regions or abbreviated as C, J and V regions. The heavy chains have one more segment, diversity region (D), so they consist of the four elements C, J, D and V. The heavy chains have one more segment, diversity region (D).

B- and T-lymphocyte receptors are formed by random recombination of the large number of genes that can code for the different regions. Thus, there are 70 different V gene elements, 61J gene elements and one C gene element. To produce the light chain of the, one V-gene element and one J-gene element are randomly selected from this construction kit and combined with the C-gene element. For the heavy chain of the 52 V-, 2 D-, 13 - and 2 C-gene elements are available for selection; consequently, there are more than 12 million different combinations.

6.2 Clonal Expansion and Clonal Selection

All lymphocytes differ from each other in their receptor specificity. Different receptors can bind to an antigen, but only a limited number. Now, to produce enough effector cells to fight an infection, the body needs a mechanism to produce the required lymphocytes. This is done by **clonal expansion** or **clonal selection.** The activated lymphocyte multiplies and differentiates into the effector cells. This process occurs within the lymph nodes (secondary lymphoid organs) in a microenvironment called the germinal center (Adams et al., 2020).

Clonal selection occurs for both types of lymphocytes, B and T cells, but in different ways. In the case of B cells, mutation is added to proliferation, which also occurs in this phase, whereas in the case of T cells it does not. Accordingly, they also act differently: the B cells produce antibodies, the T cells act by either secreting lymphokines or directly as TK cells. Figure 14 illustrates the clonal selection principle. The body contains antigens, which can be either foreign or endogenous (Cohn et al., 2007).

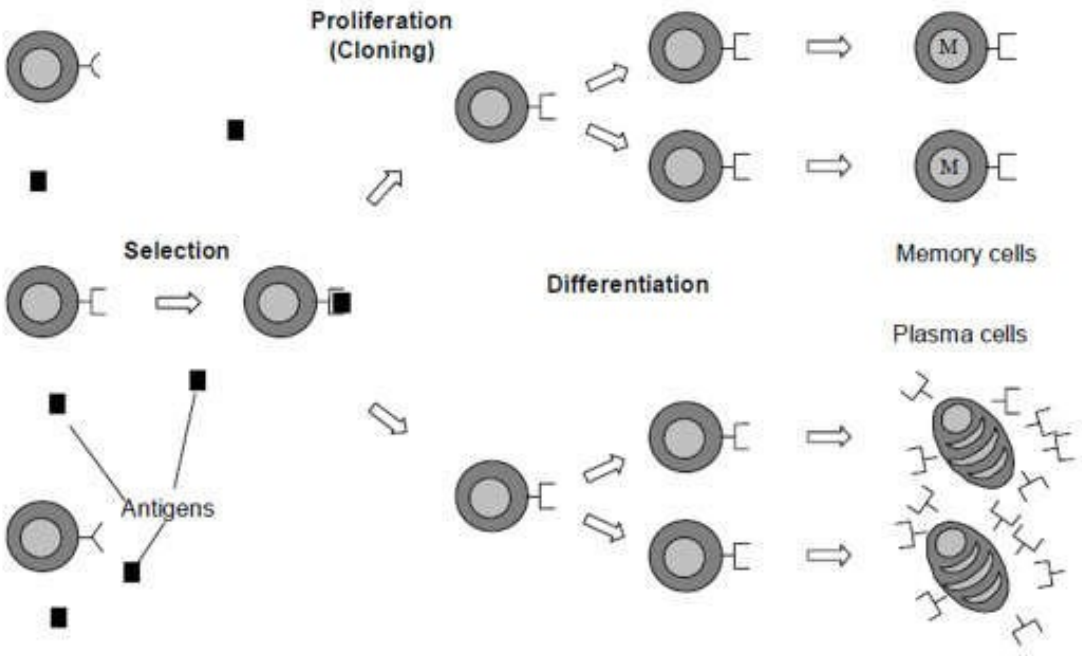


Fig. 14: Clonal selection for the immune system (Hongtao & Fengju, 2016).

First, a costimulatory signal[1] must be present for an antibody to couple to an antigen on a B cell. This is emitted by helper cells, e.g., TH cells (Fontana and Vance, 2011). If the costimulatory signal is absent, then the antigen present is an endogenous one. In this case, negative selection takes place: the B cell dies. If the signal is present, then the antigen is recognized as foreign and the B cell that has coupled to an antigen is activated. In the second step, the B cell begins to proliferate until finally plasma cells are formed, which no longer divide but secrete antibodies (III). B cells can also secrete antibodies, but much less than plasma cells. The B cells can also develop into memory cells instead of plasma cells (IV). These no longer divide, nor do they produce antibodies, but they remain in the blood, lymph, and tissues for a long time. When they later come into contact with the same or a similar antigen, they very rapidly transform into plasma cells that secrete antibodies specific for the antigen. This enables a rapid specific immune response.

According to Starr et al. (2002), a distinction is made between positive and negative selection: **Positive selection** is used to increase the concentration of beneficial mutants in the proliferation of lymphocytes. This is done by preserving them from cell death, while the others are eliminated by cell death, thus slowing down their proliferation.

T-cells can only recognize antigens presented in the form of MHC/peptide complexes. First, therefore, those immature T cells, called thymocytes, whose receptors can recognize and bind self-MHC molecules must be segregated. Positive selection, so to speak, allows the immune system to recognize the self. Consequently, the mature T cells that enter the bloodstream can usually only be activated by foreign antigens presented by Self-MHC/peptide complexes. Thus, these T-cells act in a Self-MHC-restricted adaptive immune response. The process

[1] Co-stimulation or costimulatory signaling refers to molecules that must be activated as a second receptor in addition to antigen-specific signaling for an adaptive immune response to occur (Azuma, 2019).

occurs in the thymus, but the recognition events lack the level of affinity, intensity, and efficacy that would be required for clonal expansion.

The receptors of B cells consist of the antigen receptor chain produced during the formation of the cells, and mature B cells exhibit these receptors. The receptors are connected to signal-transmitting molecules. If binding to a corresponding site (epitope) occurs, then cell death is prevented by transfer of the signaling substances and a further stage of maturation is reached, in which the rearrangement of the immunoglobulin light chain is suppressed. As a result of binding to an antigen and in the presence of a costimulatory signal from the T cells, the B cells proliferate and undergo extensive somatic hypermutation of their receptor V domain. Many of the mutants lose binding ability and die cell death. For the cells whose mutant receptors bind best to the antigen, the cell death mechanism is turned off, as is the mutation process. Some of them remain as long-living memory cells.

Positive selection is thus quite similar in its basic mechanisms for B and T cells. They differ only in their binding capacity and in their ability to mutate.

Negative selection, on the other hand, is used to eliminate cells with anti-self-receptors. Such receptors can arise during the maturation process in T cells and during mutation in B cells. The interaction between a lymphocyte and a self-antigen leads to the death of the cell. These processes occur in the lymphoid organs. In the primary lymphoid organs, foreign antigens are largely kept away, and endogenous antigens are retained, whereas in the secondary lymphoid organs, foreign material is filtered out and concentrated, and costimulatory intercellular immune responses are initiated.

Negative selection of T cells occurs in the thymus or outside the thymus. In the thymus, it functions as follows: The thymus contains large amounts of MHC I- and MHC II-bearing APCs, including macrophages, dendritic cells, and specialized epithelial cells. Since the thymus is protected by a barrier against the bloodstream (blood-thymic barrier), these APCs present mainly self-MHC/peptide complexes to the maturing T cells. The interaction of the immature T cells with these complexes results in negative selection, and the cells die. T cells that do not show

significant interaction with the self-MHC/peptide complexes are secreted by positive selection and enter the bloodstream. How quickly and how extensively the self-antigen T cells are eliminated depends on their affinity for the antigen; T cells with higher affinity die more rapidly than those with lower affinity.

However, negative selection is not perfect, some self-reactive T cells escape to the outside as immunocompetent cells. They can trigger autoimmune disease in the host organism. However, activation of a T cell outside the thymus requires more than just contact with an MHC/peptide complex. Several accompanying processes must occur, such as adhesion of various molecules to the T cell. If these costimulatory signals are absent, then contact with an MHC/peptide complex can lead to cell extinction. Signals of this type are mainly sent out by the innate immune system.

Immature B cells in the bone marrow readily respond to tolerance signals. However, mature B cells can also become tolerant, namely when they have contact with an antigen in the absence of a costimulatory signal. The decision for activation or tolerance in this case depends on the number, strength, and timing of arrival of the costimulatory signals. Rapid and sudden association of the receptor with the antigen, as typically occurs with foreign antigens, generally induces a clonal response. On the other hand, constant and relatively weak stimulation, as typically occurs with endogenous antigens, leads to tolerance, i.e., the clonal response is absent and cell death occurs later. If the antigens occur in high concentration this can also lead to tolerance.

6.3 Immunoglobulins (V/C)

Immunoglobulins are heterodimeric immune proteins consisting of two heavy (H) and two light (L) chains. Functionally, a distinction is made between variable (V) domains, which bind antigens, and constant (C) domains (Fig 15). While constant (C) domains are responsible for complement activation or binding to membrane receptors of myeloid cells of the immune system (Hayes et al., 2014), variable (V) domains form the antigen formation site (Schroeder & Cavacini, 2010).

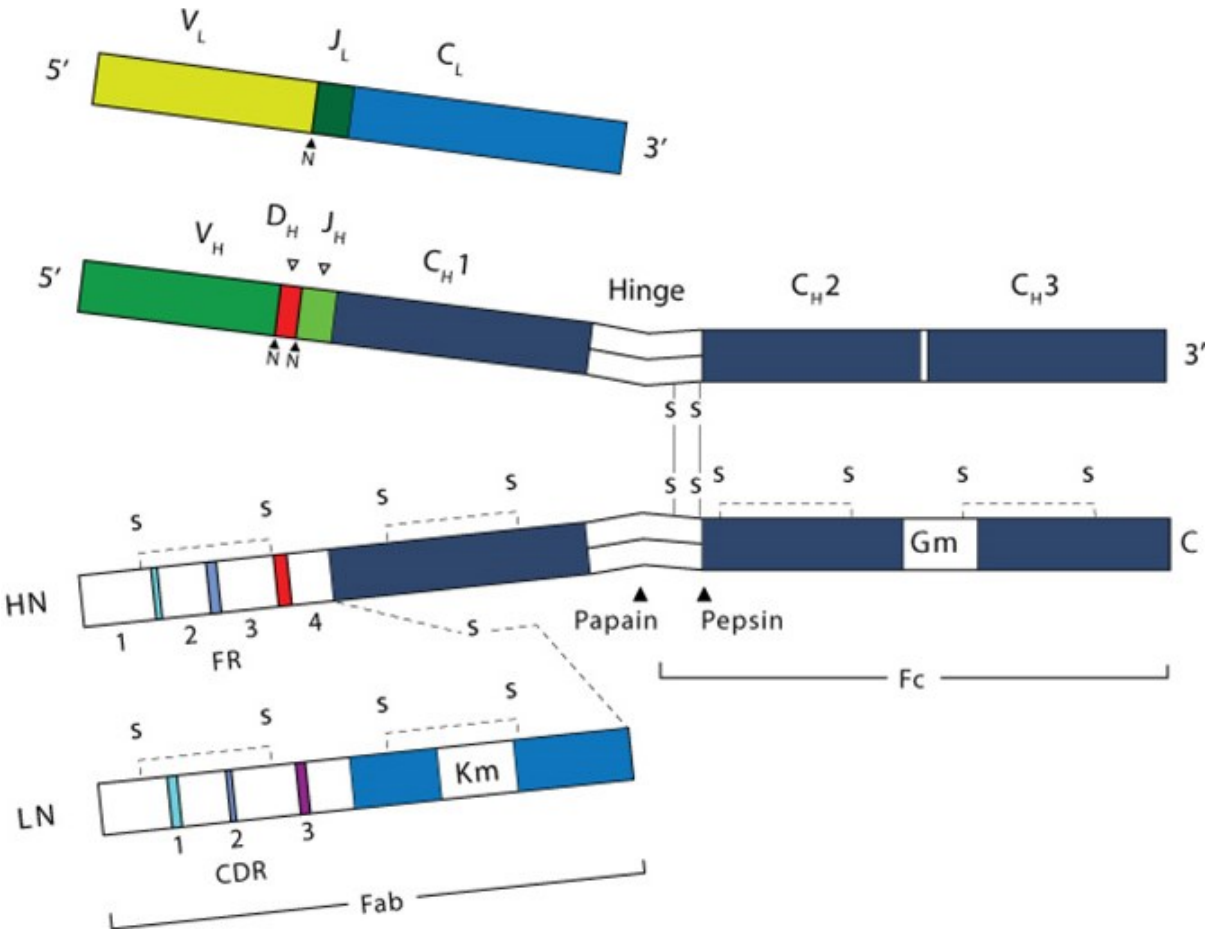


Fig. 15: Structure of IgG molecule – H= Heavy chain, L) Light Chain, N= Amino Terminus, C= Carboxy Terminus, s-s= Disulfide bridge, Gm/ Km= Allotypes as genetic markers (Schroeder and Cavacini, 2010)

The variable domains arise through a complex series of gene rearrangement events, which will be explained in detail below. They can then undergo somatic hypermutation after exposure to antigen and undergo "affinity maturation" (Schroeder & Cavacini, 2010). Each V-domain can be divided into three regions of variability and plural mutuality an, referred to as Complementarity Determining Regions (CDRs), and four regions of relatively constant sequence, referred to as Framework Regions (FRs). The three H-chain CDRs are paired with the three L-chain CDRs to form the antigen binding site as classically defined. Five major classes of heavy chain C domains can be distinguished. Each of those classes identifies the isotypes IgM, IgG, IgA, IgD, and IgE.

IgG can be split into four subclasses: IgG1, IgG2, IgG3, and IgG4, each with its own biological properties; and IgA can be similarly divided into IgA1 and IgA2. The constant domains of the H-chain can be "switched" to allow altered function while retaining antigen specificity.

Immunoglobulins exhibit high diversity due to genetic mechanisms. Lucas (2003) postulates that "*immunoglobulin gene systems and T-cell receptors appear to have evolved from a primordial gene encoding a single immunoglobulin-like domain*" (Lucas, 2003:8). Inherent mechanisms such as duplication, retention of segmentation, and somatic recombination appear to be a legacy of this primordial gene. Molecules of the major histocompatibility complex and other cell surface receptor systems appear to have evolved from this primordial immunoglobulin gene and are classified in a group of related structures called the immunoglobulin gene superfamily.

6.3.1 Class Switch Recombination (CSR)

In immunology, class switch recombination or isotype switching (CSR) refers to an isotype switch in the B cells of the immune system. In the course of an immune response, various isotypes of immunoglobulins are needed on the B cells. Class switching allows B cells to change their antibody isotype. In the VDJ heavy chain sequence, there is a switch from one C region to another, downstream. Class switching occurs primarily in the germinal centers of lymph nodes (Xu et al., 2012). At the immunoglobulin heavy chain gene locus, there are constant sections for IgM (Cμ), IgD (Cδ), IgG (Cγ), IgE (Cε), and IgA (Cα). At the 5' end of these genes is an intron, the so-called switch region. An exception is IgD, which does not have a switch region. At the 3' end of these genes is the I-exon (initiator-of-transcription exon). The germline transcripts containing the sequences from the I-exon, the switch region and the C-region of the selected isotype start from the I-exon. Transcription of the germline transcript permits enzymes of recombination to interact with DNA. Now, the open single-stranded loop is called the R-loop (Fig. 16).

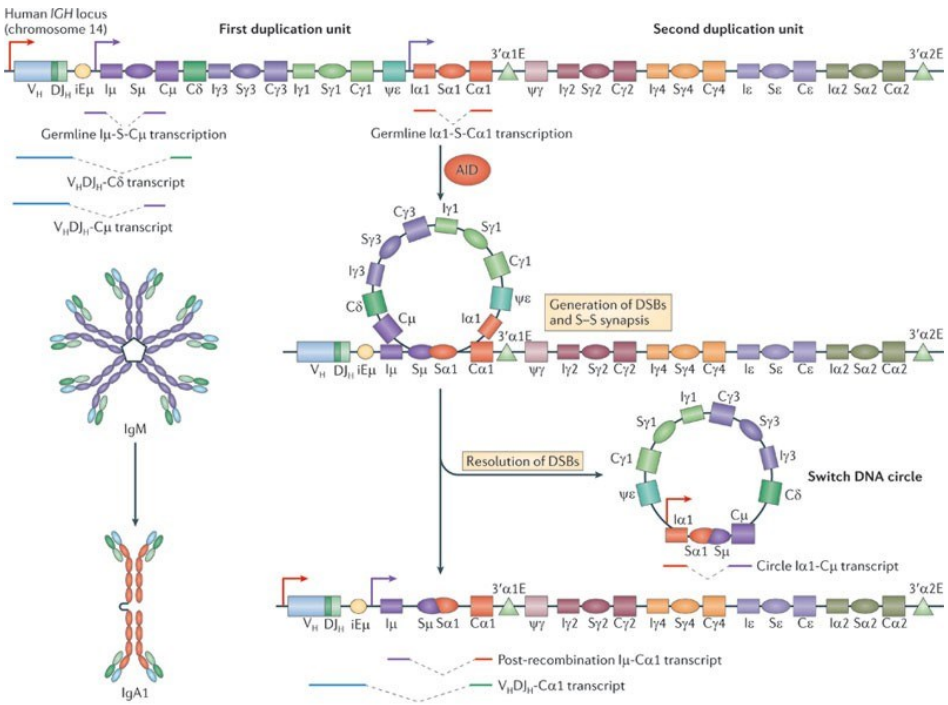


Fig.16: In CSR, double-strand breaks, and subsequent repairs occur - leading to rearrangement of re-arranged VHDJH DNA segments (Xu et al., 2012).

The switch regions of the isotype and the isotype to be formed are brought into proximity. The enzyme AID (activation induced deaminase) deaminates cytidine to uracil, which is an interfering factor in the DNA. Uracil-N-glycosylase (UNG) from the repair mechanism BER (base excision repair) removes the resulting U residues. The endonuclease Ape1 cuts at the resulting abasic sites. A double-strand break is formed. The part of the germline transcript that lies between the switch regions of the isotype used so far and the switch region of the isotype to be formed is deleted and the two strand ends are joined by non-homologous end joining.

6.3.2 V(D)J Recombination

In immunology, V(D)J recombination is a genetic rearrangement process at the DNA that enables the formation of a very large number of different antibodies and lymphocyte receptors that would never occur by coding in the human genome alone (Roth, 2014).

V(D)J recombination is a key mechanism of the adaptive immune system. It enables the recognition of a variety of different antigens from bacteria, viruses, and parasites (Chi, Li & Qiu, 2020).

Rather, complex rearrangement mechanisms occur within the receptors, which give rise to this almost unlimited number of receptors (Nguyen et al., 2016). In the following, the mechanism is explained for B cells and immunoglobulins, respectively; apart from a few details, the process also corresponds to the formation in T cell receptors (Aribi, 2020).

The antibodies, as secreted B-cell receptors, consist of a heavy and a light chain. These chains have both variable regions (V) and constant regions (C). The variable regions, as shown in Fig.17, are further subdivided into a V-gene segment and a J-gene segment, which lies between the V- and the C-region. In the heavy chain, there is also a D-gene segment.

In the light chain as well as in the heavy chain, the different segments accumulate on the DNA level. For the sake of clarity, the segments are first considered as a whole here (Fig. 17):

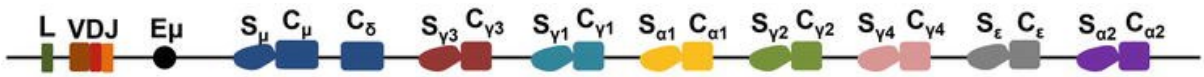


Fig.17: Organization of Immunoglobulin Genes.

At the light chain, the VL region forms by combining a larger V segment and a smaller J segment. This occurs by DNA restructuring during B cell maturation, where a random V segment is joined with a random J segment. This newly formed VJ exon is spliced together with an exon for the C region during gene expression, eventually resulting in a complete polypeptide chain (Fig- 18).

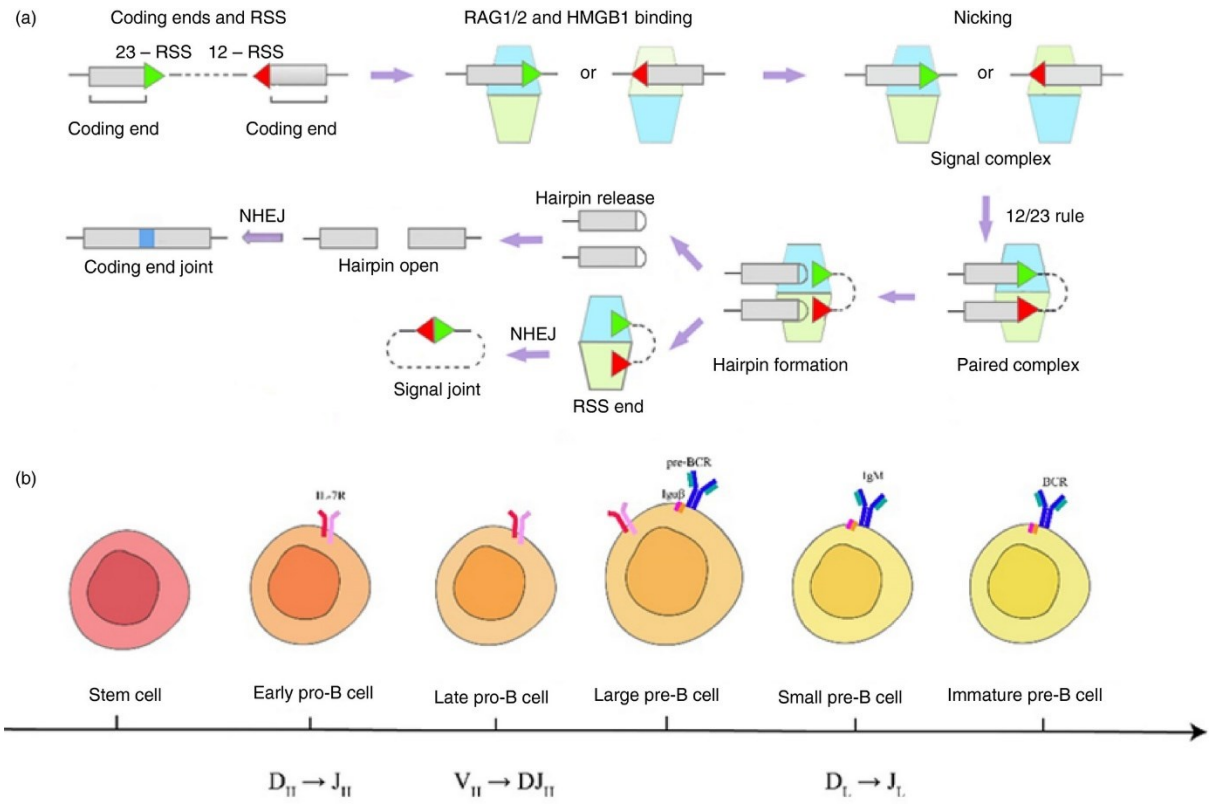


Fig. 18: Overview of V(D)J recombination. (a) RAG pairs recombination signal sequence (RSS) according to the 12/13 principle. A hairpin structure is formed, released, and rejoined by non-homologous end joining (NHEJ). (b) B cell maturation process (Chi, Li & Qiu, 2020).

The mechanism is analogous to the light chain, with the only difference that between the V-region and the J-region lies the D-region. A D-segment unites first with a J-segment, the DJ-segment formed in this way then with a V-segment. This alone cannot yet lead to this high number of different receptors. For this, a look at the gene segments themselves is necessary (Fig. 19).

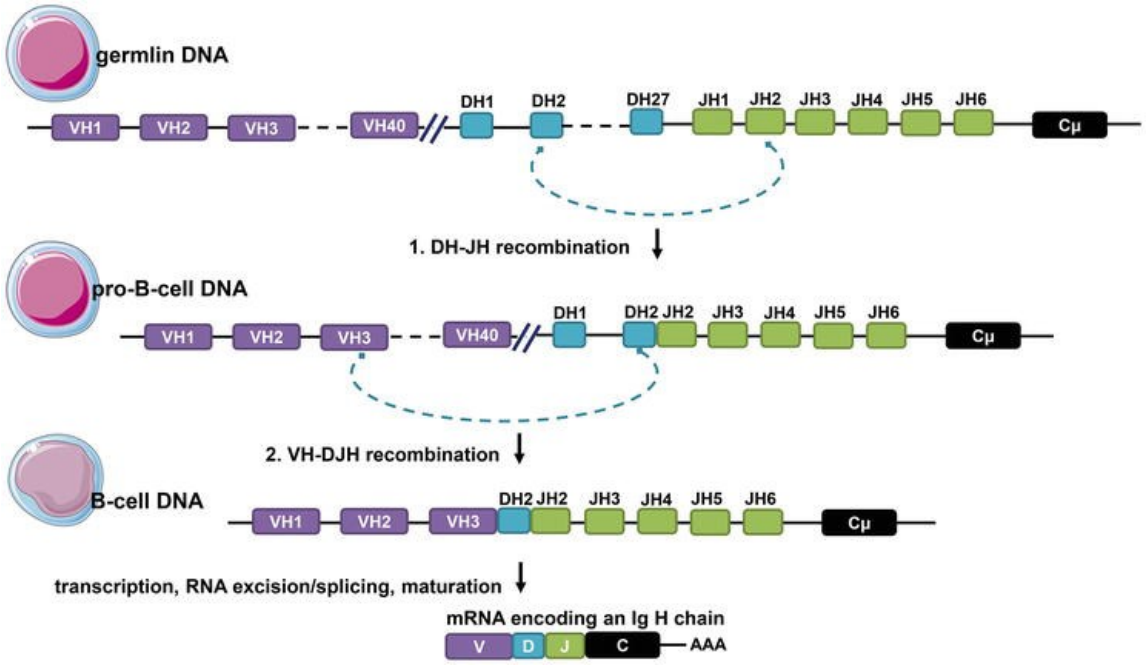


Fig. 19: Steps of the V(D)J Recombination in germline DNA (Aribi, 2020)

The rearrangement of the V, D and J segments is controlled by the so-called recombination signal sequences (RSS for short), which regulate the rearrangements via the 12/23 rule. Additional variability arises from the cutting and joining of the different segments, as this biochemical process is not precisely predetermined. In addition, random nucleotides can be incorporated. This fuzziness leads to what is known as junctional diversification.

To further increase diversification, point mutations occur in the rearranged genes, more precisely in the variable regions, which further increase the number of different receptors and ultimately ensure that a specific receptor arises in the human immune system for virtually every antigen. At the same time, somatic

hypermutation increases the affinity of the receptors for their antigens, which is also referred to as affinity maturation

6.4 Recombinases and Recombination-Activating Genes

The Recombination-Activating Genes (RAGs) encode components of a protein complex that plays an important role in the rearrangement of those genes encoding immunoglobulin and T cell receptor molecules. Among other active enzymes, V(D)J recombinase consists of two lymphoid-specific proteins, RAG1 (1040 AS) and RAG2 (527). Both cleave DNA adjacent to gene segments to be recombined; the segments are then joined by DNA repair factors. Cellular expression of the two recombination-activating genes, RAG1 and RAG2, is restricted to lymphocytes during their developmental stages. RAG-1 and RAG-2 are essential for the maturation of B and T cells and, accordingly, are important components of the adaptive immune system (Zhang et al., 2019).

To reiterate: In the vertebrate immune system, each antibody is tailored for use against a specific antigen. The human genome, however, has a maximum of 25,000 genes, and yet can generate hundreds of millions of different antibodies, enabling it to respond to the invasion of millions of different antigens. As discussed earlier, the immune system generates this diversity of antibodies by recombining the V(D)J genes. This shuffling occurs within B cells and T cells during their maturation and is catalyzed by the RAG proteins, which will be briefly introduced in the sequel (Carmona & Schatz, 2017).

6.4.1 Genomic Structure of RAG Genes

The RAGs are translation products of the RAG genes. For example, the RAG1 gene (Gene ID: 5896) is located on chromosome 11p12 (36,510,353-36,579,762) and is 70 kb long. It is transcribed in a tissue-specific manner, primarily in bone marrow where it plays a critical role in lymphocyte maturation. It is equally often expressed in thyroid tissues.

The ClinVar database contains 345 variations of RAG1 (303 germline mutations), half of which are due to missense mutations, of which about 100 are (likely)

pathogenic. Missense variants that are pathogenic lead to Severe combined immunodeficiency. For example, in the OMIM database, Omenn syndrome (MIM:603554) is associated with mutations in the RAG1 and RAG2 genes. According to Ege (2005), Omenn syndrome (OS) is "*characterized by severe combined immunodeficiency (SCID) associated with erythroderma, hepatosplenomegaly, lymphadenopathy, and alopecia. In patients with OS, B cells are mostly absent, T-cell counts are normal to elevated, and T cells are frequently activated and express a restricted T-cell receptor (TCR) repertoire. Thus far, inherited hypomorphic mutations of the recombination activating genes 1 and 2 (RAG1/2) have been described in OS.*" (Ege, 2005: 4179).

6.4.2 Molecular Structure of the RAG Proteins

Both RAG proteins are large multidomain proteins consisting of core and non-core regions. The catalytic center and DNA-binding motifs are largely contained in RAG-1. The core region of RAG1 is required for nonameric binding (nonameric-binding domain, NBD) as well as catalysis of cleavage (Yin et al., 2009). The non-core domains of RAG proteins provide important regulatory functions. Less clear is the role of RAG2, which may function to activate RAG1 for sequence-specific binding and cleavage and provide additional DNA-binding capabilities. That is, RAG-1 and RAG-2 must work synergistically to activate VDJ recombination. It has been shown in vitro that RAG-1 has limited ability to induce VDJ gene recombination activity when isolated and transfected into fibroblast samples. When RAG-1 is co-transfected with RAG-2, the recombination frequency increased 1000-fold in vitro.

RAG enzymes work as a complex with multiple subunits to induce cleavage of a single double-stranded DNA molecule (dsDNA) between the coding segment of the antigen receptor and a flanking recombination signal sequence (RSS). Efficient cleavage of a DNA substrate requires RAG1, RAG2, a divalent metal ion, and proteins of the highly mobile group (HMGB1 or HMGB2). They do this in two steps. The first step of V(D)J recombination is the binding of RAG, probably together with HMGB1 or HMGB2. This RAG complex catalyzes two reactions, nicking and hairpin formation.

First, the RAG complex binds either a 12-RSS or a 23-RSS, resulting in a 12- or 23-signal complex. They first introduce a nick into the 5' end (upstream) of the conserved RSS heptamer adjacent to the coding sequence, leaving a specific biochemical structure on this region of DNA: a 3' hydroxyl (OH) group at the coding end and a 5' phosphate group at the RSS end. In vitro, HMGB1 or HMGB2 have been shown to stimulate the activity of RAG in DNA binding, nicking, and hairpin formation, presumably by inducing RSS bending (Chi, Li & Qiu, 2020).

In the next step, these chemical groups are coupled by binding the OH group (at the coding end) to the PO4 group (which sits between the RSS and the gene segment on the opposite strand). This creates a 5'-phosphorylated double-strand break at the RSS and a hairpin structure at the coding end. The RAG proteins stay at these junctions until further enzymes (mainly TDT) repair the DNA breaks.

Notarangelo et al. (2016) also investigate the role of RAG in the maturation of pre-B and pre-T cells, for which V(D)J recombination is essential. The researchers showed that mature B cells possess two other remarkable RAG-independent phenomena of diversification of their VDJ DNA: so-called class-switch recombination and somatic hypermutation.

As with many enzymes, RAG proteins are quite large. RAG-1 in mice contains 1040 amino acids and mouse RAG-2 contains 527 amino acids. As shown in the figure: RAG-1 residues 384-1008 and RAG-2 residues 1-387 maintain most of the DNA cleavage activity. The RAG-1 core contains three key amino acids (D600, D708, and E962) in the so-called DDE motif - the major catalytic site for DNA cleavage. These residues are crucial for the nicking of the DNA strand and for the formation of the DNA hairpin. AS 384-454 of RAG-1 comprise a nonameric-binding region (NBR) that specifically binds the conserved nonomer (9 nucleotides) of the RSS, and the core domain (amino acids 528-760) of RAG-1 specifically binds to the RSS heptamer. Afterwards, the core region of RAG-2 forms a six-bladed beta-propeller structure that appears to be less specific for its target than RAG-1. Figures 20 and 21 illustrate the domains and the subdomains of RAG1 and RAG2 and show the active motifs in gene and protein.

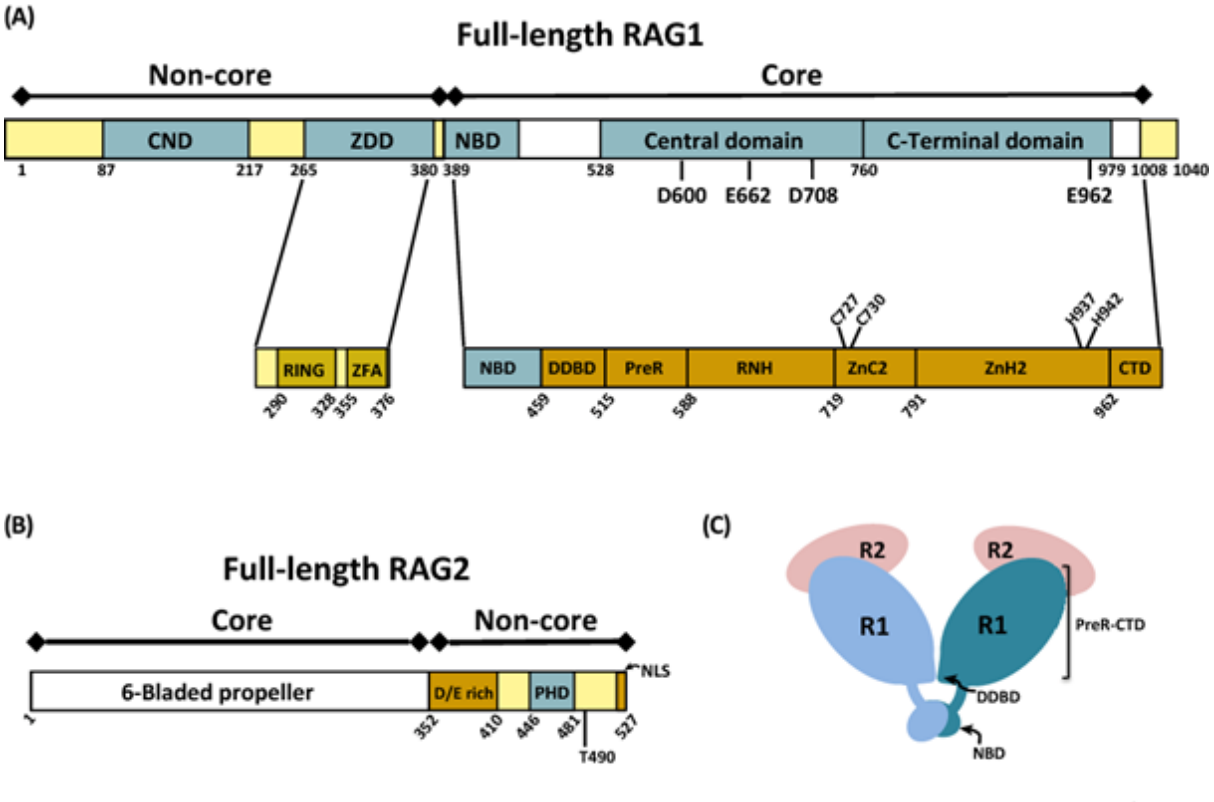


Fig. 20: Schematic representation of domains and subdomains in RAG proteins. (A) Full-length RAG1. Non-core region in RAG1 include an N-terminal region (residues 1388) and a short C-terminal region (residues 10091040). Topologically independent domains are shown as blue boxes. For RAG1 without core, these are the central non-core domain (CND) and the zinc dimerization domain (ZDD). A crystal structure of the ZDD showed two subdomains, the RING finger and the zinc finger A (ZFA). Domains in the core region of RAG1 include the nonameric binding domain (NBD), the central domain, and the C-terminal domain. The core region contains several modules, which are highlighted in the lower bar. The three DDE active site residues (D600, D708, and E962), a putative fourth active site residue (E662), and zinc-coordinating ligands (C727, C730, H937, and H942) are marked in the RAG1 bar graphs. (B) Full-length mouse RAG2. The core region is a six-leaf propeller domain. The non-core region of RAG2 contains an acidic region labeled as D/E-rich and a plant homology domain (PHD) that specifically binds the trimethylated lysine-4 residue of histone H3 (H3K4me3). The positions of residue T490 and the C-terminal localization signal (NLS) are indicated. (C) Cartoon showing the core of the RAG12RAG22 complex. The RAG1 (R1) subunits are colored in light and dark blue, and the RAG2 (R2) subunits are colored in pink. The domains (NBD and DDBD) that form the RAG1 dimer interface are highlighted. The remaining subdomains (PreR to CTD) form the catalytically active region of core RAG1. (Rodgers, 2017)

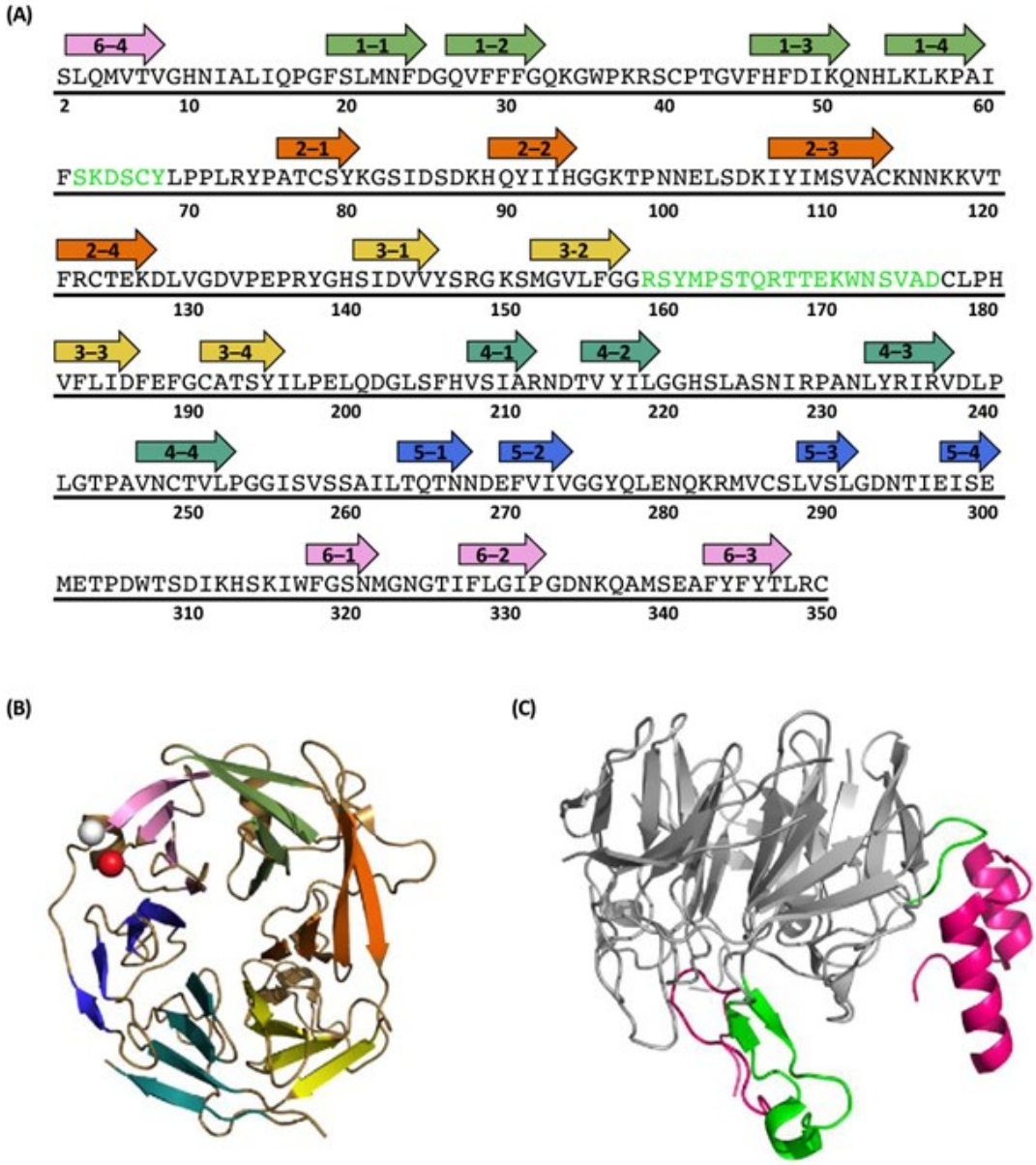


Fig. 21: Core RAG2 topology and interface with RAG1. (A) Murine core RAG2 sequence. Arrows are shown above the residues that form the beta strands in the beta propeller structure. The first number in each arrow corresponds to the leaf number (one to six). The second number refers to the strand (one through four) in each blade, with strands one and four located on the inside and outside of the blade, respectively, according to the orientation of the structure shown in (B). The residues marked in green are located in the crystal structure of RAG12RAG22 at the interface with RAG1. (B) Ribbon diagram of the murine core of RAG2 (from PDB ID: 4WWX). The N and C termini are shown as white and red spheres, respectively. Each propeller blade is color-coded according to the arrows in (A), where blades one to six are light green, orange, yellow, light blue, dark blue, and pink, respectively. (C) The core structure of RAG2 rotated 90o relative to (B). The residues at the interface with RAG1 are shown in green and correspond to the residues shown in green in (A). The murine Core RAG1 loops (residues 537 553 and 750782) that form the bulk of the interface with Core RAG2 are shown as magenta bands. (Rodgers, 2017)

6.4.3 RAG Complexes

Cryoelectronic microscopy structures of the synaptic RAG complexes appears to form a closed dimer conformation with new intermolecular interactions between two RAG1 and RAG2 monomers upon DNA binding, compared with the apo-RAG complex, which exists as a relatively open conformation. TH two RAG1 molecules in the closed dimer are involved in cooperative binding of the 12-RSS and 23-RSS intermediates with base-specific interactions in the heptamer of the signaling end. To avoid collision in the active site, the first base of the heptamer in the signal end is folded out. Each coding end of the nicked RSS intermediate is exclusively stabilized by a RAG1-RAG2 monomer with nonspecific protein-DNA interactions (Ru, Zhang & Wu, 2018). To facilitated hairpin formation through a potential two-metal ion catalytic mechanism, the coding end is highly distorted with one base flipped out of the active site DNA duplex. The 12-RSS as well as the 23-RSS intermediates are highly bent and asymmetrically bound to the synaptic RAG complex, with the dimer of the nonamer binding domain tilted toward the nonamer of the 12-RSS but tilting away from the nonamer of the 23-RSS, emphasizing the 12/23 rule (Fig. 22). Two HMGB1 molecules bind on either side of the 12-RSS as well as 23-RSS to stabilize the highly bent RSSs. These structures are necessary for the molecular mechanisms for DNA recognition, catalysis, and the unique synapsis underlying the 12/23 rule. They also provide new insights into RAG-associated human diseases and represent the most complete set of complexes in the catalytic pathways of any recombinase, transposase, or integrase in the DDE family (Ru et al., 2015).

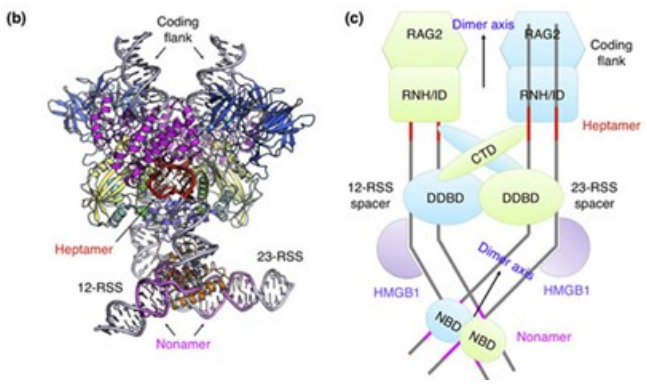


Fig. 22: Overall structure of the RAG dimer in complex with nicked 12- and 23-RSS intermediates (Ru, Zhang & Wu, 2018)

6.4.4 RAGs Homologies to DNA-Transposons

Kapitonov and Jurka (2005) postulated, after comparing those homologous sequences in the core region, that RAG1 evolved from a transposase of the Transib superfamily. Transib is a superfamily of interspersed repeats (TIR) DNA transposons. It was named after the Trans-Siberian Express (Kapitonov & Jurka, 2003). It normally encodes a single protein, the DDE transcriptase. TIRs are structurally similar to recombination signal sequences (RSS). The heptamer carries the consensus CACWRTG, whereas the nonamer is more divergent.

In this context, Jonas and Gellert (2004) even speak of the "taming" of transposons: although the transposon origins of these genes are well established, there is still no consensus on when the ancestral RAG1/2 locus appeared in the vertebrate genome. Kasahara, Suzuki, and Pasquier (2004) hypothesize that RAG1 invaded between 1001 and 590 million years ago. Kapitonov and Koonin (2015) postulate that both proteins, RAG and RAG2. originated from the same transib transposon that invaded non-vertebrate species multiple times and invaded the ancestral vertebrate genome about 500 MYA ago. It is currently believed that the invasion of RAG1/2 is the most important evolutionary event in terms of shaping the adaptive immune system of gnathostomes against the variable lymphocyte receptor system of agnates (Huang et al., 2016).

According to the NCBI dataset, 5 transcripts exist for the RAG1 gene, all of which are c2 after splicing. 6.5 - 7kb in length. These transcripts can now be used in the question of RAG protein evolution. Here, Fugmann et al. (2006) compare the vertebrate RAG1 (from mouse) with the RAG gene from the invertebrate model organism of the sea urchin *Strongylocentrotus purpuratus* (SpRag1L Like genes). Alignment of SpRag1L with mouse Rag1 shows matches in both core and non-core regions. The nonamer binding domain is highlighted in yellow in the graphic below, the conserved amino acids of the DDE are shown in red boxes, and a zinc finger mediating interaction with Rag2 (ZFB) is shaded in gray:

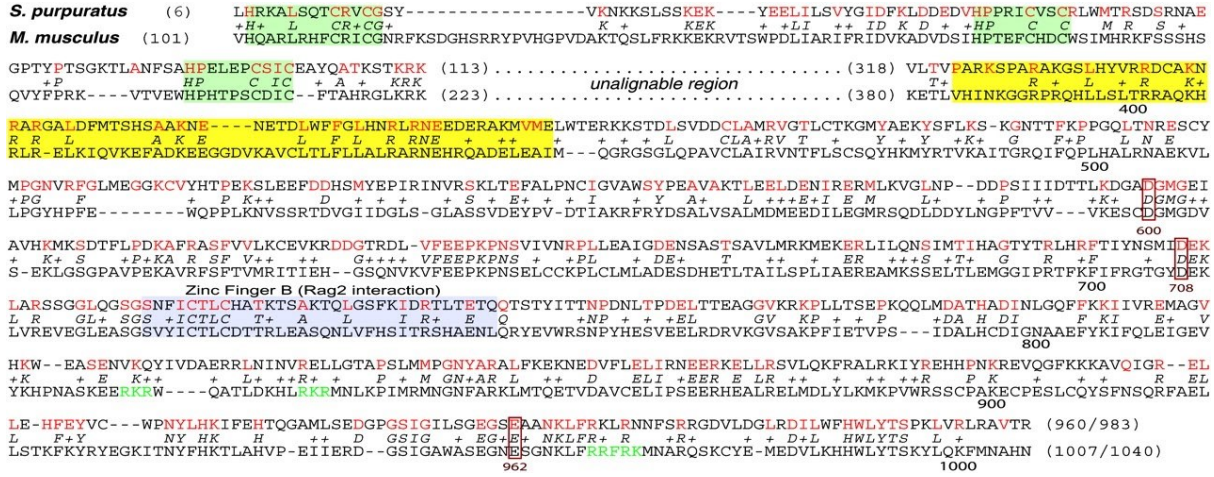


Fig. 23: Homology comparison between SpRAG1L and RAG1 of invertebrates and vertebrates indicate a common origin in Deuterostomia, although they had a different function in sea urchins. This suggests that operationalization for V(D)J recombination occurred later in vertebrates. (Fugmann et al., 2006)

Fugman et al. (2006) can confirm homology by sequence comparisons as shown in Figure 23, especially by identification of the DDE motif (RAG1: D600, D708 and E962/ SpRAG1L: D548, D658 and E914) as well as functional domains (zinc finger domain and nanomer).

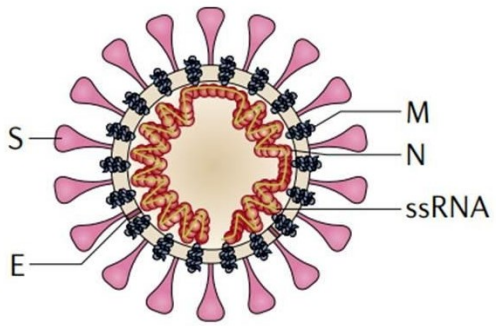
7. Transposons and Polymorphism in Gene Expression

In the following section, the preceding considerations on transposons are related to the currently all-dominating question of the COVID-19 pandemic caused by the SARS-CoV-2. Once again, the ambivalent role of transposable elements becomes apparent, which on the one hand can decisively shape gene expression patterns, but on the other hand can also influence the balance to the detriment of a disease. The focus of these considerations is the Alu element of the ACE1 gene.

7.1 Excursus: SARS-CoV-2

SARS-CoV-2 belongs to the (human) coronaviruses that have Mammalia and Aves as (intermediate) hosts in addition to humans. With SARS-CoV, MERS-CoV and finally SARS-CoV-2, highly pathogenic variants of the coronaviruses have spread in the last two decades. The three mentioned have in common bronchial epithelial cells, pneumocytes and other cells of the upper respiratory tract. The pathophysiology of COVID-19 virus is described in the next chapter. In these introductory considerations, the virus will be briefly introduced in a genetic context. The excellent review by V'kovski and colleagues (October 2020) will be used to describe the (genetic) blueprint and replicative mechanism of SARS-CoV-2.

7.1.1 Structure of the SARS-CoV-2



S (Spike) consisting of two subunits

M (membrane) Main component of the envelope

E (envelope) integrated in the double lipid layer

S/M/N anchored in double bilipid layer.

N (nucleocapsid) arginine/lysine-rich phosphoprotein.

+ssRNA positive-sense single-stranded 30 kb RNA

7.1.2 Genetics of SARS-CoV-2

At 30kb, the +ssRNA is one of the largest compared to other RNA viruses. The genome is organized as follows (Fig.24/ Fig.25):

5'-leader/cap- UTR- ORF 1a -ORF 1b-S-E-M-16 non-structural proteins- UTR-PolyA-3'

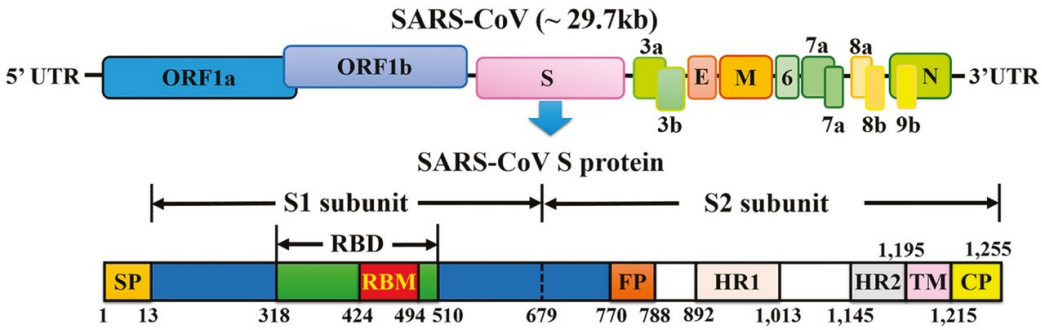


Fig. 24: Viral genome of Sars-CoV Virus and the translated protein.

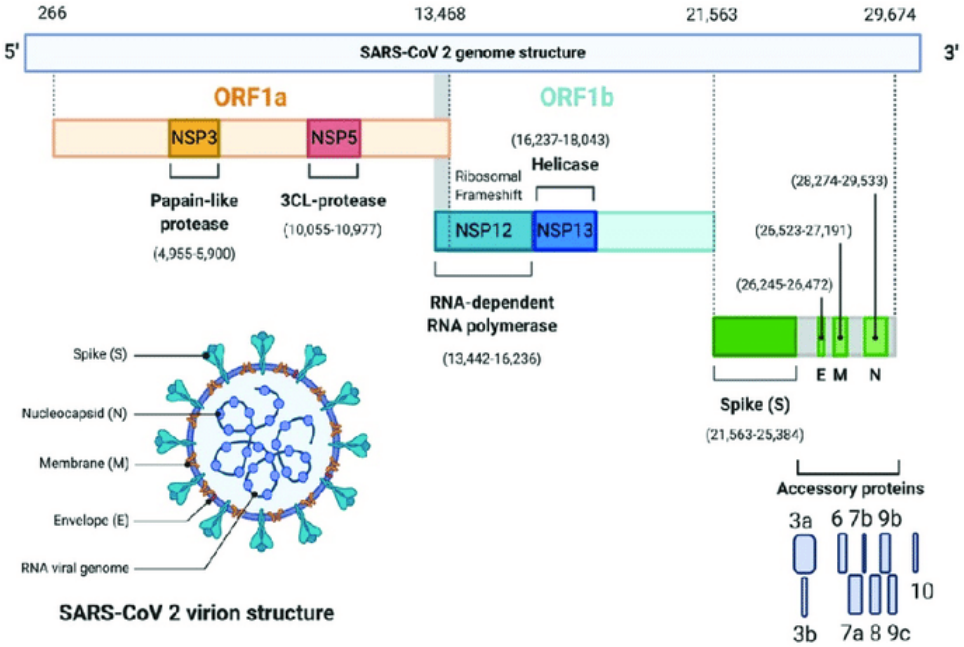


Fig. 25: Coding sequences of SARS-CoV-2.

7.1.3 Life cycle of the SARS-CoV-2

After the S-glycoproteins of the virion bind to the cellular receptor angiotensin-converting enzyme 2 (ACE2), the virion enters the cell via endosomal or non-endosomal pathways. After the virus enter the host cell, viral RNA is released into the cytoplasm. ORF1a and ORF1ab are translated to produce the polyproteins pp1a and pp1ab, which are cleaved by the proteases of the RTC.

ORF1a and ORF 1b encode two polyproteins (pp1a, pp1ab). Different combinations of the translation products result in four important proteins: Two proteases, one helicase and one RNA-dependent RNA polymerase. Together they form the replicase-transcriptase unit.

The latter now reads viral RNA from the antisense direction and has three functions:

- **Replication of** viral RNA- replication with proofreading function by exoribonuclease.
- **Transcription of** the 16 non-structural protein genes (accessory proteins) as well as the sub genomic structural genes (N, M, E, S), which can now be translated
- Recombination

SARS-CoV-2 structural proteins are assembled into a nucleocapsid in the cytoplasm, followed by budding into the lumen of the endoplasmic reticulum (ER)-Golgi apparatus. The virions are then released from the infected cell by exocytosis.

7.1.4 Pathophysiology of SARS-CoV-2 infection

To understand the interaction between ACE2- receptors and the genetic polymorphism of ACE1 in the context of COVID-19 infection, it is necessary to study the pathophysiology of SARS-CoV-2 infection in more detail.

The SARS-CoV-2 virus can enter the human body via different transmission routes. For successful replication of the genetic material and the proteinogenic components, the virus must enter the corresponding cells via an ACE2 receptor and occupy the translation machinery there. ACE2 receptors are found, among other places, on the surface of epithelial cells in the alveoli.

Alveoli, as blind ends of the respiratory tract, are a structural element of the lung; their primary function is gas exchange between the inhaled alveolar air and the blood in the alveolar capillaries. The following three cell types can be distinguished:

1. Alveolar macrophages, specialized immune cells
2. Type I pneumocytes, flat epithelial cells for gas exchange
3. Type II pneumocytes, cubic epithelial cells produce surfactant

The ACE2- receptor, which is important for the virus, is located on the surface of the type II cells. The basic function of these cells is to secrete surfactant, i.e., surface-active surfactants, in order to reduce the surface tension between humid alveolar air and alveolar epithelium and thus prevent collapse of the alveoli.

The spike protein (S protein) mediates the entry of SARS-CoV-2 into type II cells by binding to ACE2. According to Hofmann and Pöhlmann (2004), this can happen via two pathways:

(1) Non-endosomal: Binding of a viral S2 subunit glycoprotein to the ACE2 receptor can induce conformational changes that facilitate viral and plasma membrane fusion. This requires prior enzymatic separation of the S protein: After binding of virus to ACE2, the S protein is proteolytically

cleaved into a 1 and an S2 subunits by TMPRSS2 serine protease.[2] Only this leads to a fusion of the viral membrane with the cell membrane and hence releasing the viral nucleocapsid into the cytoplasm of the host cell.

(2) Endosomal: Alternatively, binding to the receptor can be followed by uptake of the virion into an endosomal compartment. This requires the absence of other enzymes and an acidic ph. Proton influx into the endosome can then trigger membrane fusion activity in glycoproteins. The decisive factor here is the enzyme cathepsin L.

Entry (Fig. 26) via the non-endosomal pathway is more efficient than the endosomal pathway, indicating the greater relevance of protease-dependent cleavage. In addition to proteolysis by TMPRSS2, the sheddase ADAM17, an enzyme belonging to the ADAM family of proteins, i.e., a disintegrin, also plays a crucial role (Heurich et al., 2013). This enzyme has a proteolytic effect on ACE2. This process is called shedding, hence the name sheddase.

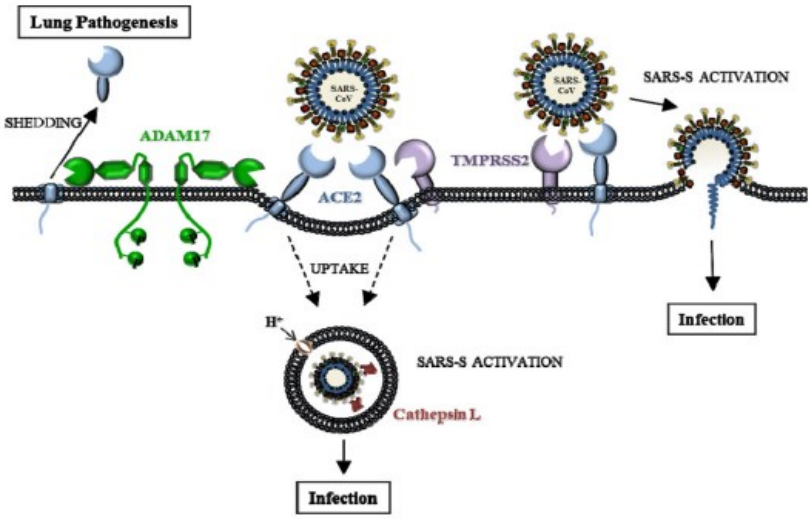


Fig. 26: Entry of SARS-CoV into the host cell. (left) If no SARS-S-activating proteases are expressed at the cell surface, then virions are taken up into endosomes. SARS-CoV is cleaved and activated by the pH-dependent cysteine protease cathepsin L. (right) When the SARS-S-activating protease TMPRSS2 is expressed together with ACE2 on the surface of target cells, proteolysis by TMPRSS2 allows fusion at the cell surface. (Heurich et al, 2013)

[2] The S protein consists of two subunits: S1 unit is the outer segment that contains the receptor-binding domain and S2, which is the transmembrane segment associated with functional components such as glycoproteins for fusion.

After entry, the viral positive single stranded RNA (+ssRNA) is released. This is now translated from 5' to 3' by cellular ribosomes. The result is an RNA polymerase (viral RNA-dependent RNA polymerase), which has two tasks: First, it replicates the viral RNA; second, it transcribes it, but this time from 3' to 5' to smaller mRNAs. These sub genomic RNAs code for other proteinogenic viral particles, including the S protein. The individual components are assembled and leave the type II pneumocytes by exocytosis (Fig. 27)

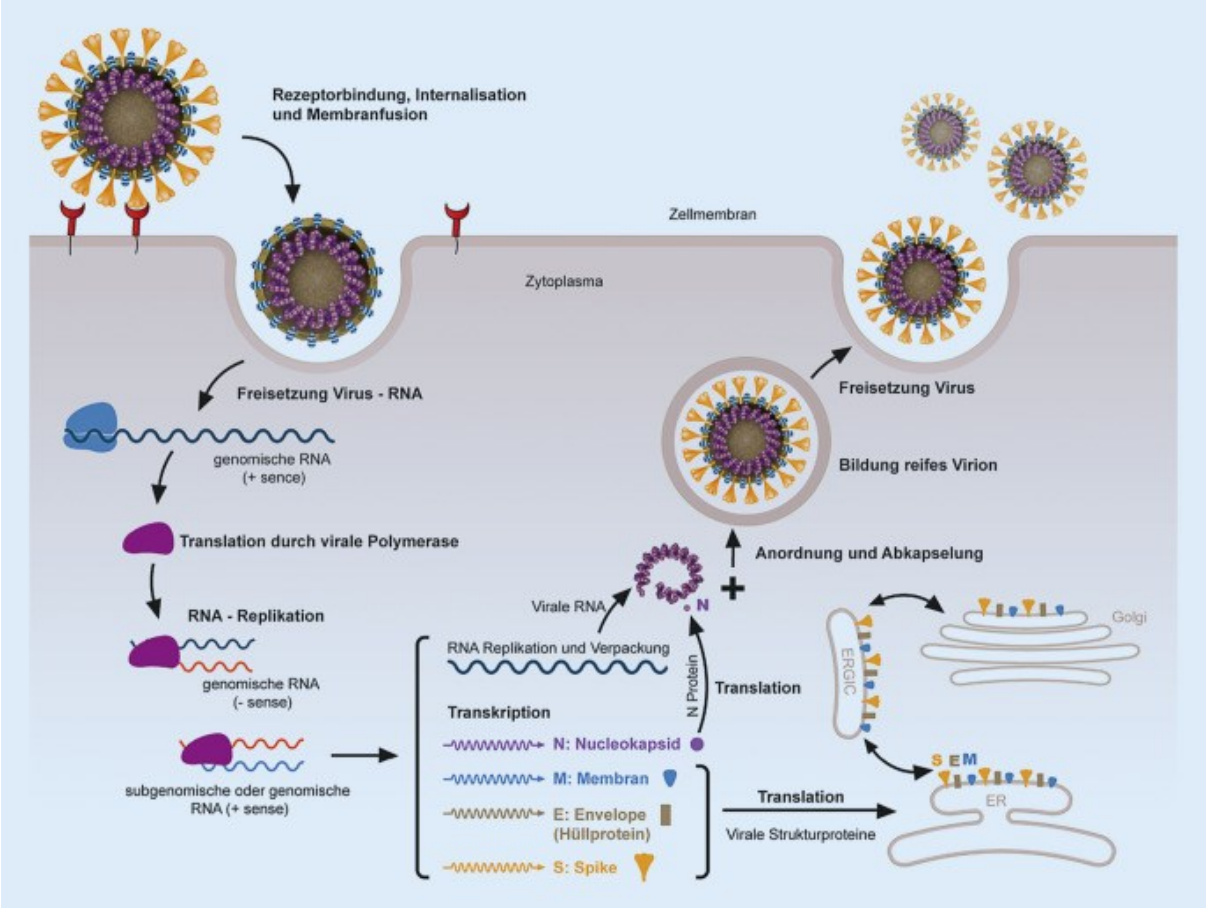


Fig. 27: Pathophysiology of SARS-CoV-19 (Ueffing et al., 2020).

The resulting virions are returned to the alveolar lumen and can now be exhaled. In the course of the takeover of the protein biosynthesis apparatus of the host cell, the latter perishes and is affected by necrosis. As a result, pro-inflammatory signal substances, cytokines, are released into the environment. Among other things, they activate alveolar macrophages, which now initiate an inflammatory response via interleukins (IL1, IL6). The following can be observed (Bohn et al., 2020):

(1) Increased permeability due to vasodilation of the walls of the alveolar capillaries favor the leakage of plasma into the interstitium. In some cases, plasma also enters the alveolar lumen via the epithelium. The resulting edema impedes gas exchange.
(2) The severely throttled gas exchange leads to hypoxia and the deficient supply of oxygen. The body reacts with shortness of breath.
(3) Due to the transfer of cytokines into the blood, there is a risk that the initially local inflammation develops into a systemic inflammation, which can also lead to sepsis with the risk of multi-organ failure.
(4) Interleukins also activate further neutrophils, which flood into the focus of inflammation and attempt to destroy the pathogens. This is also referred to as a cytokine storm. In the alveoli the gas exchange becomes more and more difficult. This can also lead to an autoimmune reaction. In addition, fibrosis by fibroblasts can now occur, a type of scarring in the epithelial tissue.
(5) Due to the death of the type II pneumocytes, their function, i.e. the secretion of surfactant, is now also missing. In addition, the accumulated plasma in the interstitial space now exerts pressure on the alveoli. The increased surface tension now leads to the collapse of alveoli.
(6) IL1 and IL6 now reach the hypothalamus via the bloodstream. Here, the body temperature is regulated, which is increased in the course of the inflammatory reaction and results in fever (Fig. 28).

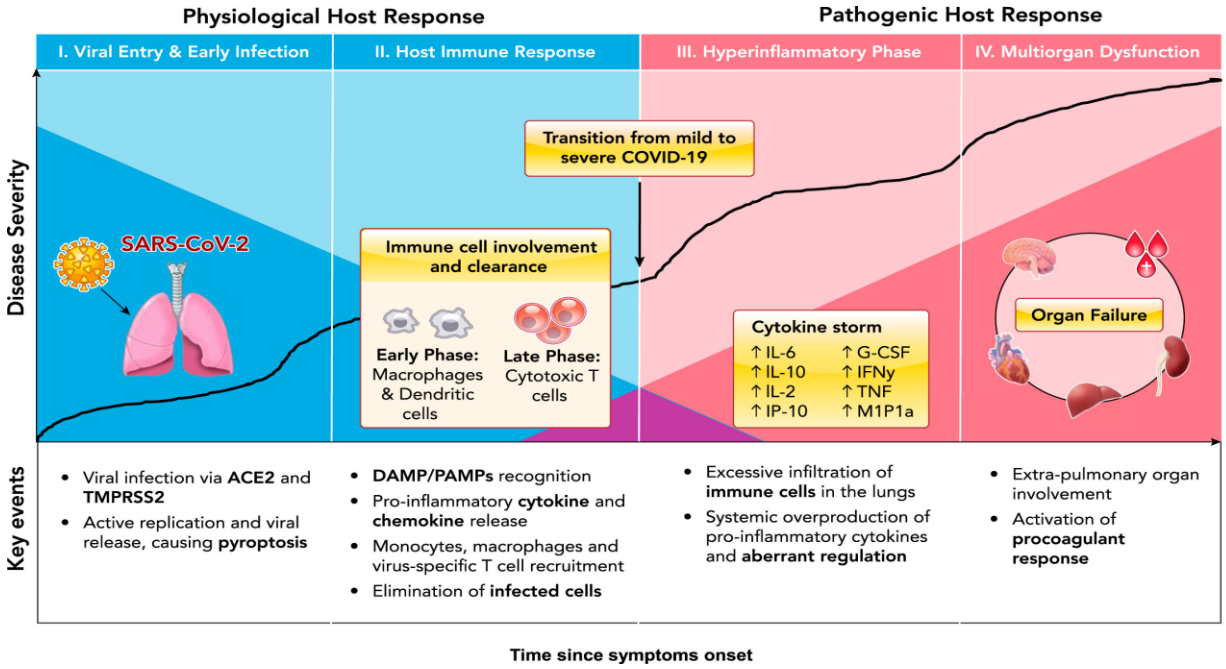


Fig. 28: Pathophysiology of COVID-19 infection: Blue shows the physiological response. Red shows the pathogenic hyperinflammatory immune response. (Bohn et al., 2020)

Necrosis of the alveoli eventually leads to acute respiratory distress syndrome (ARDS). In summary, Bohn et al. (2020) state:

[...]The large majority of confirmed SARS-CoV-2 cases are mild (81%), with ~14% progressing to severe pneumonia and 5% developing acute respiratory distress syndrome (ARDS), sepsis, and/or multisystem organ failure (MOF).

7.1.5 Renin-angiotensin-aldosterone system

SARS-CoV can enter cells via the ACE2 receptor. The transmural metalloprotein ACE2, angiotensin converting enzyme-2, is an element of the renin-angiotensin-aldosterone system (RAAS) and is found on the surface of cells in the lungs, digestive tract, kidneys, and vasculature of the blood. Accordingly, the concentration of ACE2 and its expression pattern also determines the course and severity of infection. The function of ACE2 can be clearly illustrated by the functioning of RAAS (Fig. 29). RAAS is a tight-knit hormonal feedback system that regulates renal blood flow (Ghazi and Drawz, 2017; Moon, 2011):

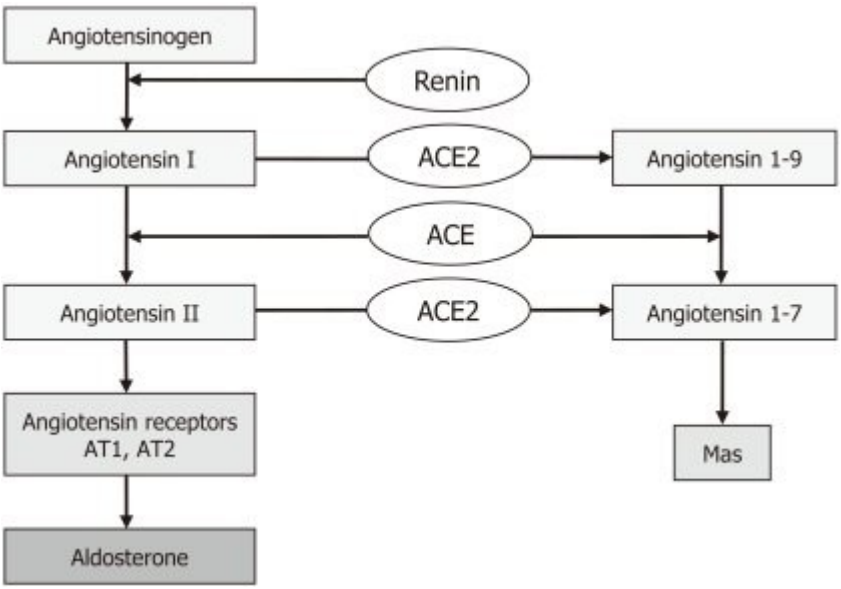


Fig. 29: Renin-angiotensin-aldosterone system. ACE2: angiotensin-converting enzyme-related carboxypeptidase 2; ACE, angiotensin-converting enzyme; AT1, angiotensin II type 1; AT2, angiotensin II type 2. (Moon, 2011).

- The first step of the RAAS is the production of the enzyme renin in the kidneys, which require a constant blood flow for the filtration and excretion of nitrogen metabolites. Renin is thus released from granular cells of the renal juxtaglomerular apparatus (JGA) in response to one of three factors: (1) Reduced sodium delivery to the distal convoluted tubule, detected by macula densa- cells. (2) Reduced perfusion pressure in the

kidney, demonstrated by baroreceptors in the afferent arteriole. (3) Sympathetic stimulation of the JGA via β 1 -adrenoreceptors.

- Angiotensinogen is a precursor protein that is produced in the liver and that is cleaved by reninin order to produce angiotensin I (AngI).
- Angiotensin I is converted to angiotensin II (AngII) by angiotensin-converting enzyme (ACE). This conversion occurs mainly in the lung, where ACE is produced by vascular endothelial cells, although ACE is also produced in smaller amounts in the renal endothelium.
- Angiotensin II exerts its effects by binding to various receptors throughout the body. It binds to one of two G-protein-coupled receptors, the AT1 and AT2 receptors. Most actions occur through the AT1 receptor: in the arterioles, AngII causes vasoconstriction; in the kidneys, it stimulates sodium reabsorption; in the adrenal gland, it stimulates the release of aldosterone; in the hypothalamus, it stimulates the secretion of antidiuretic hormone (ADH) and increases the feeling of thirst; finally, it causes the release of norepinephrine by the sympathetic nervous system.
- Aldosterone is a mineralocorticoid steroid hormone that acts on the cells of the collecting ducts in the nephron. It enhances the expression of apical epithelial Na^+ channels (ENaC) to reabsorb sodium in urine. Furthermore, the activity of basolateral Na + / K + / ATPase is increased. This increases blood pressure and blood volume.
- The antagonist of ACE is the ACE-2 angiotensin 1-7 axis. ACE-2 catalyzes the hydrolysis of AngII.
- Angiotensin 1-7 (Ang1-7) is on the one hand anti-inflammatory and antioxidant; on the other hand, it acts as a vasodilator reciprocally to Ang II by binding and activating the MAS receptor.

7.1.6 ACE1 polymorphism

The angiotensin system plays a key role in the pathogenesis of COVID-19. First, ACE2 is the cellular receptor for SARS-CoV-2, and the expression of the ACE2 gene may regulate the individual's susceptibility to the infection. In addition, the balance between ACE1 and ACE2 activity may be involved in the pathogenesis of respiratory disease and have significance in the severity of COVID-19. Functional ACE1/ACE2 gene polymorphisms have been associated with the risk of cardiovascular and pulmonary disease and thus could also contribute to the outcome of COVID-19.

Angiotensin-converting enzyme (ACE) is an endopeptidase consisting of two catalytic domains and is normally expressed by endothelial, epithelial, and neuronal cells. It can be found in both membrane-bound (ACE) and soluble (sACE) forms, the latter formed by the action of an as yet unidentified zinc metalloprotease ("ACE secretase") that cleaves mature, membrane-bound ACE at a juxtamembranous extracellular domain and releases the large extracellular portion of the enzyme. The function of ACE is closely associated with the renin-angiotensin system (RAAS) pathway, within which ACE catalyzes the formation of the vasoconstrictor octapeptide angiotensin II (Ang II) from its nonvasoactive precursor angiotensin I (Ang I) and is also responsible for cleavage and inactivation of the vasodilator bradykinin. The end result is vasopressor activity that can be blocked by ACE blockers - a standard treatment for hypertension.

The ACE gene, ACE, encodes two isoenzymes. The somatic isoenzyme is expressed in many tissues, mainly in the lung, including vascular endothelial cells, epithelial kidney cells, and testicular Leydig cells, while the germinal one is expressed only in sperm.

The angiotensin converting enzyme gene is located on chromosome 17 (17q23.3) and has more than 160 described polymorphisms. A polymorphism in the gene encoding angiotensin I-converting enzyme (ACE) was identified in 1990 by Rigat et al. The polymorphism is due to a 287 bp fragment in the ACE gene in intron 16 of the gene (rs1799752) on chromosome 17. Homozygote II and DD and heterozygote DI Genotype accounts for approximately half of the variance in

circulating ACE levels, and from the II genotype to the DD genotype, the presence of each D allele is associated with an additive effect on ACE activity (50% higher in the DD genotype compared with the II genotype).

The DD genotype is associated with high plasma levels of the protein (Fig. 30), the DI genotype with intermediate levels, and the II genotype with low plasma protein levels, suggesting that this polymorphism may affect the renin-angiotensin system and its abnormal function may be associated with various diseases. (Cintra et al., 2018)

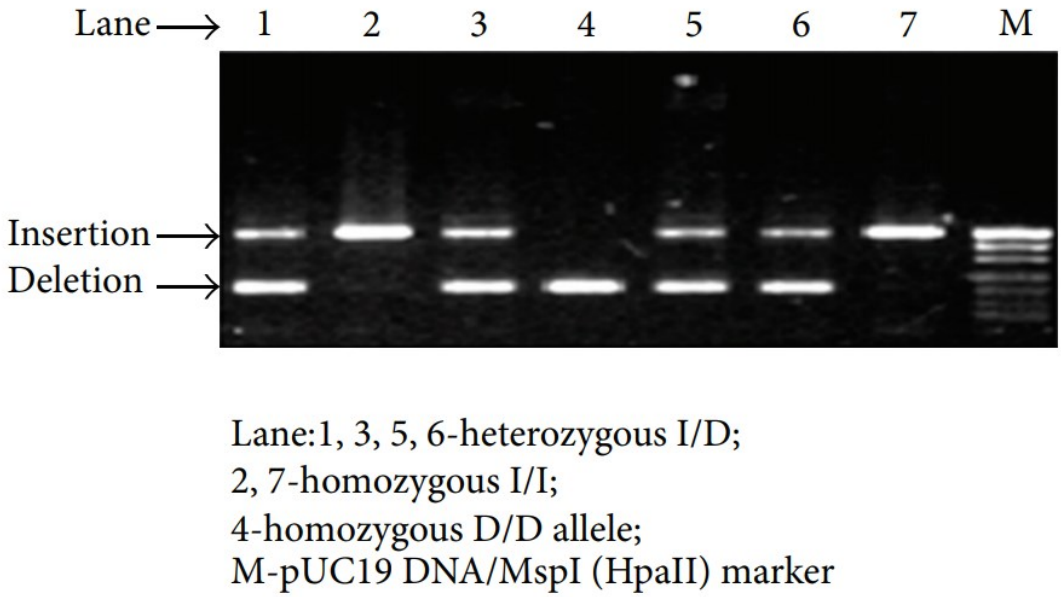


Fig. 30: Polymorphism Representative band patterns of the ACE I/D polymorphism were analyzed by agarose gel electrophoresis. The amplification products of the two alleles D and I were identified at 319-bp and 597-bp amplicons, respectively. (Prabhakar et al., 2014)

The intronic sequence is an ALU element whose insertion appears to reduce ACE expression. It was also found that DD (homozygotes) and ID (heterozygotes) have more circulating ACE than II (homozygotes).

Alu elements are short DNA sequences (300bp) originally characterized by the action of the restriction endonuclease of *Arthrobacter luteus* (Alu) (Schmid & Deininger, 1975). Alu elements are the most abundant TE with over one million copies distributed throughout the human genome, accounting for up to 10% of it.

They are particularly abundant in the higher-than-average gene rich R bands. Alu sequences are transcribed by RNA polymerase III, but not translated; RNAs are also formed, but these are not translated into proteins. They are replicated like any other DNA sequence, but rely on LINE retrotransposons to generate new elements. They are originally derived from the small cytoplasmic 7SL RNA.

Alu elements belong to the SINE (Short Interspersed Nuclear Elements) family. They were formed 55 million years ago in primates by fusion of the 5' and 3' ends of the 7SL RNA gene. Latter encodes the RNA unit of the signal recognition particle (SRP). This fusion gave rise to the first fossil Alu monomers (FAMs). They were about 160 bp long and rather rare to find in the human genome. According to the current model, modern Alu- elements arose from a head-tail fusion of two different FAMs, resulting in a dimeric structure consisting of two similar but distinct monomers (left and right arms) joined by an A-rich linker. Modern Alu-elements have a length of 300 bp and are classified into subfamilies according to their relative age. Alu elements in the form of dimers are unique to primates. They replicated throughout the primate genome via RNA intermediates by a mechanism of retrotransposition, but their amplification was dependent on the transposition machinery of other retrotransposons because they do not encode protein. Dewannieux, Esnault and Heidmann (2003) confirmed in their research that Alu elements can use LINE-1 elements for this purpose.

Alu elements (Fig. 31) are responsible for the regulation of tissue-specific genes and are involved in the transcription of neighboring genes and can sometimes alter the way a gene is expressed. Alu elements are a common source of mutations in humans; however, such mutations are often restricted to noncoding regions of pre-mRNA (introns), which is why Alu elements have been proposed to affect gene expression and have been found to contain functional promoter regions for steroid hormone receptors. Because of the high content of CpG dinucleotides found in Alu elements, these regions serve as sites of methylation and contribute to up to 30% of the methylation sites in the human genome. (Schmid, 1998)

Fig. 31: Structure of an Alu element consisting of a left and right monomer, a polyadenylate tail, and two promoter boxes for RNA polymerase III (5' A box with the consensus TGGCTCACGCC/3' B box with the consensus GWTCGAGAC) and a tandem site duplication at both ends.

To investigate the sequence of this polymorphism and the associated phenotypes as well as the classification, interpretation and allele frequency distribution, databases are used:

1. HGNC (Gene ID 1636): Provides a summary and refers to the Alu element and the connection with various diseases. Attention is also drawn to the relationship to SARS-CoV: *Regulation of the homologous ACE2 gene may be involved in progression of disease caused by several human coronaviruses, including SARS-CoV and SARS-CoV-2. Alternative splicing results in multiple transcript variants encoding both somatic (sACE) and male-specific testicular (tACE) isoforms.*

2. **UCSC Browser:** If one enters the ACE gen in the Genome Browser in the Human Assembly GRCh38/hg38, one can now click on the intron 16/25 and display the sequence - the intron 16 would therefore be ENST00000290866.10_32, with the position chr17:63488792-63488940. Here it can be seen that this is the allele that deleted the Aku sequence, as it is only 149 bp:

gtgagtccctgctgccaacatcactggcacttgggtcccttcattttcctca
aagaggtgctgtgaaaccccaagcctaggaaaaggtagatccctggaggagg
caggtaatgtgagagcctggctgtcccctctgtag

Neither the ClinVar (Phenotype and Literature) nor dbSNP 153 (Variation) tracks show polymorphism.

3. dbSNP **(rs1799752):** intron variant is reported as insertion of 50 nucleotides at position 17:63,488,530: GRCh38.p12 chr 17NC_000017.11:g.63488543_63488544 insTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCATACAGTCACTTTT. You can see the polyadenylate tail here (reverse, compliment), but 50 base pairs are too short for an Alu element.

4. ClinVar (ID: 18061): 228 variations of the ACE gene appear in ClinVar, when limiting the variation type to insertion, 6 variants now appear, including the variant of the Ins/Del polymorphism last clinically evaluated in 2009. Nomenclature is NM_000789.3(ACE):c.2306-117_2306-116 insAF118569.1:g.14094_14382. This shows that an insertion occurs at cDNA position 2.306.116. If one continues to follow the addition AF118569.1, one arrives at the gene bank. Both here and in dbSNP an allele frequency is missing.

5. gene bank (ID: AF118569): As already evident in the nomenclature from ClinVar, this is an insertion on genomic level: g.14094_14382, i.e. 288 bp:

14041 ctgctggaga ccactcccat cctttctccc atttctctag acctgctgcc tatacagtca

14101 cttttttt ttttttgaga cggagtctcg ctctgtcgcc caggctggag tgcagtggcg

14161 ggatctcggc tcactgcaag ctccgcctcc cgggttcacg ccattctcct gcctcagcct

14221 cccaagtagc tgggaccaca ggcgcccgcc actacgcccg gctaattttt tgtatttta

14281 gtagagacgg ggtttcaccg ttttagccgg gatggtctcg atctcctgac ctcgtgatcc

14341 gcccgcctcg gcctcccaaa gtgctgggat tacaggcgtg atacagtcac ttttatgtgg

14401 tttcgccaat tttattccag ctctgaaatt ctctgagctc cccttacaag cagaggtgag

The poly adenyl tail can be assumed, to verify the sequence the 288 bp are now complementary directed and reverse set up using the http://arep.med.harvard.edu tool, as the Alu element on the opposite strand is mirror coded:

ATCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACGA
GGTCAGGA**GATCGAGAC**CATCCCGGCTAAAACGGTGAAACCCCGTCTCTACT
AAAAATACAAAAAATTAGCCGGGCGTAGTGGCGGGCGCCTGTGGTCCCAGCT
ACTTGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCGGAGCTTGC
AGTGAGCCGAGATCCCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACT
CCGTCTCAAAAAAAAAAAAAAAGTGACTGT

For the evaluation of this sequence, we refer to the Figure 32 of (Cantarella et al., 2019):

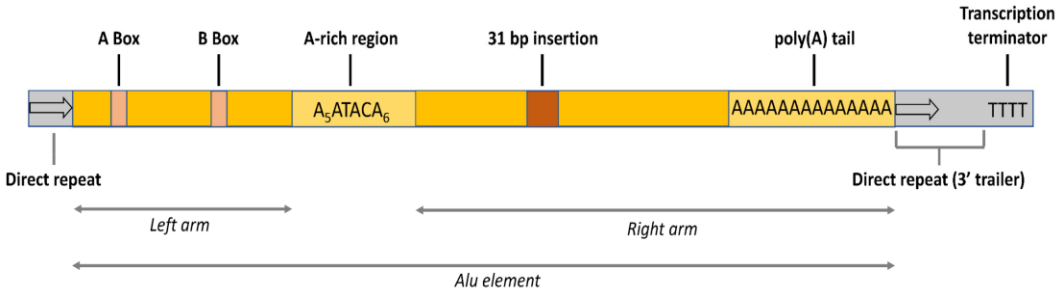


Fig. 32: Structure of Alu Element as reference. (Cantarella et al., 2019)

Now all important components of the transposon are recognizable. At the 3'-end is the poly(A)-tail, also the linker sequence (A5ATACA6) is clearly recognizable and separates the 288 bp long Alu-element into a right and left monomer (arm). In the left monomer, both conserved boxes, which are actually intended for RNA polymerase III, are also recognizable.

In this context, the possible hairpin or cloverleaf structure of transcribed Alu elements is also exciting, since they are formed from two building blocks 7SL RNA (Ullu & Tschudi, 1984). Häsler and Strub (2006) therefore propose the following secondary structure for AluRNA by transcription of RNA Pol. III: Using the Emboss Needle tool of EMBL-EBI Hinxton (https://www.ebi.ac.uk/), this secondary structure can also be detected by mirroring the individual monomers with themselves. In addition, the insertion of an additional 31 bp in the right monomer is clearly visible, resulting in a longer monomer (Fig. 33).

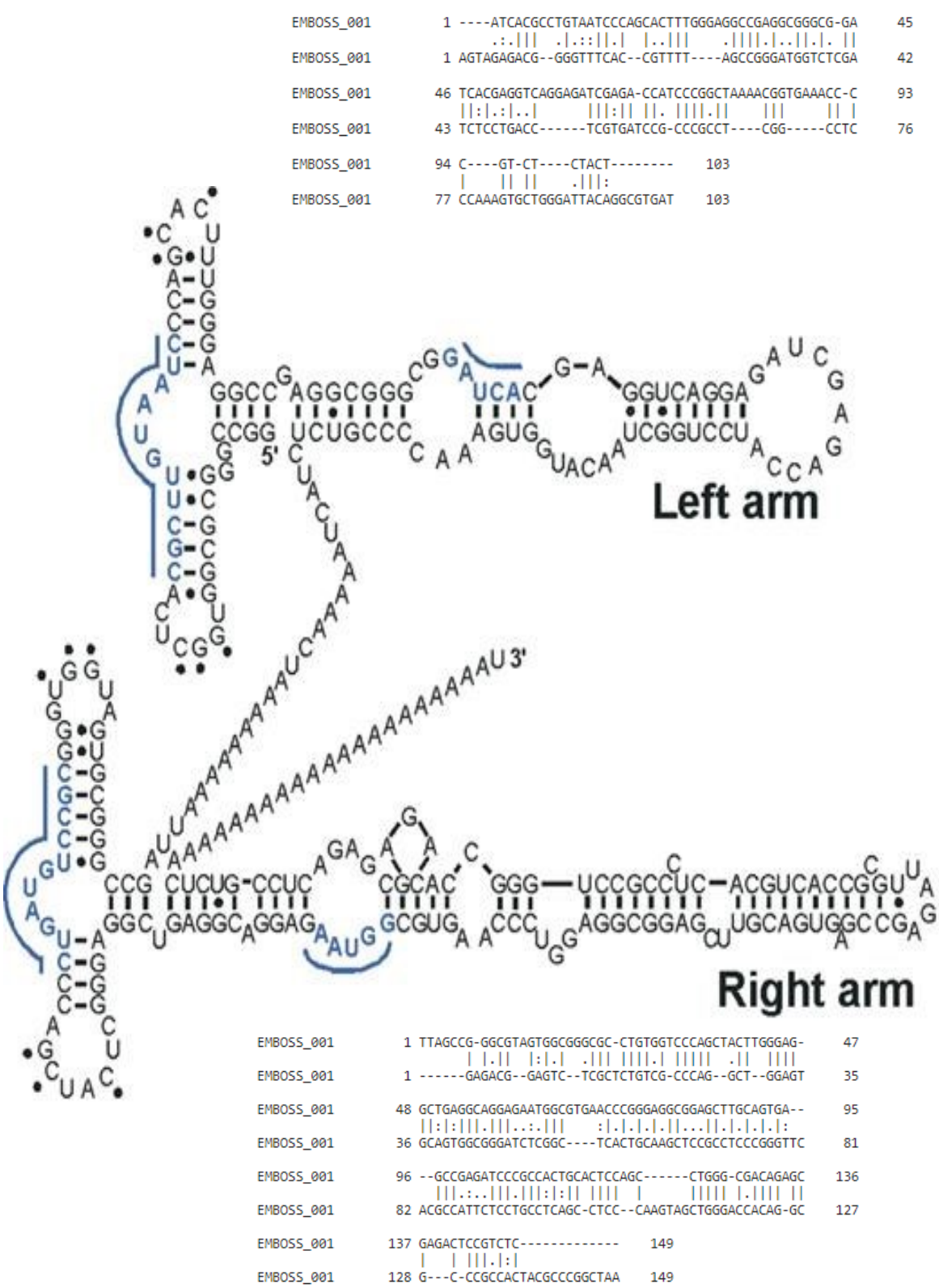


Fig. 33: AluRNA secondary structure (Häsler & Strub, 2006).

The B-box from base 60 to 68 also corresponds to the descriptions in the literature. The A-box, however, does not completely correspond to the information in the literature. In fact, some bases seem to be missing. To complete this for the sake of completeness, reference was made to known studies in which the Alu element that is part of intron 16 of the ACE gene has already been detected by PCR.

For this purpose, the study by Bolli et al. (2010) "Improvement in ACE I/D polymorphism detection" is recapitulated. To address typing issues and enable rapid and accurate analysis, they performed a stepdown PCR reaction followed by detection using Nanogen technology. We compared these results with those obtained using conventional genotyping methods such as classical and confirmatory PCR. In the process, they designed new primers: *"[...] We designed two novel primers, different from those reported by Rigat (Figure 1) and also used in the stepdown protocol from Chiang (Table 1). Additionally, the insertion sequence (Alu) was 289 bp long, following Villar,7 and not 287 bp, as Rigat described. "*

Table I. Traditional and new primers.

Traditional Primers	Sequence	Primer size	CG content	Amplicon size
Forward	5'-CTGGAGACCACTCCCATCCTTTCT-3'	24	54%	482
Reverse	5'-GATGTGGCCATCACATTCGTCAGAT-3'	25	48%	
New Primers	**Sequence**	**Primer size**	**CG content**	**Amplicon size**
Forward	5'-CTGGAGA**G**CCACTCCCATCCTTTCT-3'	25	56%	483
Reverse	5'-GA**Y**GTGGCCATCACATTCGTCAGAT-3'	25	48%	

*CTGGAGAGCCACTCCCATCCTTTCT*CCCATTTCTCTAGACCTGCTGCCTATACAG
TCACTTTT[**TTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGG**
CTGGAGTGCAGTGGCGGGATCTCGGCTCACTGCAAGCTCCGCCTC
CCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGA
CCACAGGCGCCCGCCACTACGCCCGGCTAATTTTTTGTATTTTTAG
TAGAGACGGGGTTTCACCGTTTTAGCCGGGATGGTCTCGATCTCCT
GACCTCGTGATCCGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTA
CAGGCGTGATACAGTCACTTTT]ATGTGGTTTCGCCAATTTTATTCCAGCT
CTGAAATTCTCTGAGCTCCCCTTACAAGCAGAGGTGAGCTAAGGGCTGGAGC
TCAAGSCATTCAAMCCCCTACCAG*ATCTGACGAATGTGATGGCCACRTC*

Figure 1. Corrected *ACE* gene sequence (in bold insertion sequence (Alu), in italics and underlined primer sequence).

With the help of this information, the ALU element could now be completed:

AAAAGTGACTGTATCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATC
ACGAGGTCAGGAGATCGAGACCATCCCGGCTAAAACGGTGAAACCCCGTCTCTACTAAA
AATACAAAAAATTAGCCGGGCGTAGTGGCGGGCGCCTGTGGTCCCAGCTACTTGGGAG
GCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCGGAGCTTGCAGTGAGCCGAGATCC
CGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAA
GTGACTGT

Tandem Repeats A Block B Block Left Arm Left Right Arm Poly-A Tandem Repeats

Now the sequence could be entered with the reverse/complement sequence in the Blast Alignment (BLAST®), indeed the whole chromosome 17 appeared, under Alignments you get to the Sequence ID: AP023477.1 and then to this evaluation:

```
 Homo sapiens DNA, chromosome 17, nearly complete genome
 GenBank: AP023477.1
 Go to:
 LOCUS      AP023477    301 bpDNA   linear PRI 26-SEP-2020
 DEFINITION Homo sapiens DNA, chromosome 17, nearly complete genome.
 ACCESSION AP023477 REGION: 60940689..60940989
 VERSION    AP023477.1
 DBLINK     BioProject: PRJDB10452
            BioSample: SAMD00243993
 KEYWORDS   .
 SOURCE     Homo sapiens (human)
   ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates;
Haplorrhini;
            Catarrhini; Hominidae; Homo.
 REFERENCE 1
   AUTHORS Takayama,J. , Kinoshita,K., Yamamoto,M. and Tamiya,G.
          TITLE   Construction and Integration of Three De Novo
Japanese Human Genomes.
             Assemblies toward a Population-Specific Reference
   JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 301)
     AUTHORS Tamiya,G. and Takayama,J.
          TITLE   Direct Submission
   JOURNAL Submitted (25-SEP-2020) Contact:Jun Takayama Tohoku
University,
             Tohoku Medical Megabank Organization; 2-1 Seiryo-machi
Aoba-ku,
             Sendai, Miyagi 980-8573, Japan
 COMMENT    Please visit our website, jMorp
            https://jmorp.megabank.tohoku.ac.jp/
             ##Genome-Assembly-Data-START##
             Assembly Method :: Falcon v. 2017.11.02-16.04-py2.7-
                                ucs2.tar.gz; BionanoSolve v. 3.1,
v. 3.2
             Genome Coverage :     : 1167x
             Sequencing Technology :: PacBio RSII; Bionano; Illumina
HiSeq
             ##genome-assembly-data-END##
 FEATURES   Location/Qualifiers
      source 1.. 301
                     /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                     /chromosome="17"
 ORIGIN
 1    acagtcactt tttttttt   tttgagacgg agtctcgctc tgtcgcccag gctggagtgc
 61   agtggcggga tctcggctca ctgcaagctc cgcctcccgg gttcacgcca ttctcctgcc
 121  tcagcctccc aagtagctgg gaccacaggc gcccgccact acgcccggct aatttttttgt
 181  atttttagta gagacggggt ttcaccgttt tagccgggat ggtctcgatc tcctgacctc
 241  gtgatccgcc cgcctcggcc tcccaaagtg ctgggattac aggcgtgata cagtcacttt
 301  t
```

7.1.7 ACE2 as entry point for SARS-CoV-2

As a transmembrane protein, ACE2 serves as a major entry site into cells for some coronaviruses, including SARS-CoV. Binding of the spike S1 protein of SARS-CoV-2 to the catalytic domain of ACE2 on the cell surface results in endocytosis and translocation of both the virus and the enzyme into the endosomes of the cell; this entry process also requires priming of the S protein by the host serine protease TMPRSS2, inhibition of which is currently being investigated as a potential therapeutic. The disruption of S protein glycosylation significantly impairs viral entry, indicating the importance of glycan-protein interactions in this process (Fig. 34).

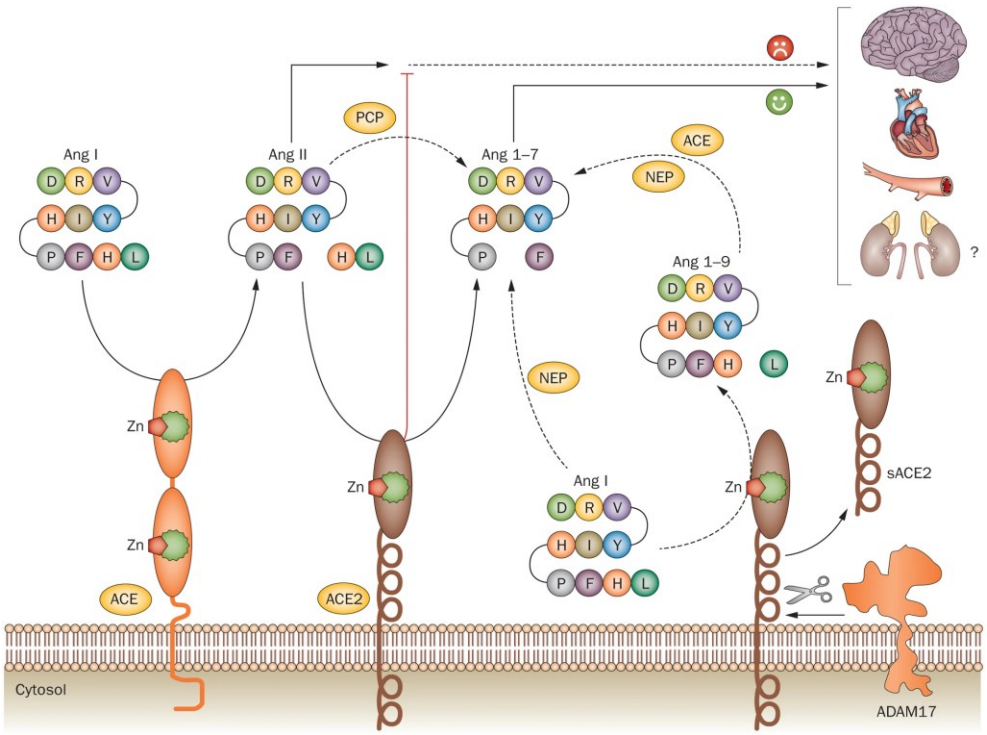


Fig. 34: ACE2 converts Ang II to Ang 1-7 as well as Ang I to Ang 1-9, which is then cleaved by NEP to give Ang 1-7. ACE2, that is membrane bound, can be cleaved by the metalloproteinase ADAM17 to form a soluble form of ACE2.. (Jiang et al., 2014)

In summary, it has the task of catalyzing the active angiotensin II (AngII) and converting it into the vasoactive angiotensin1-7. This has the reciprocal effect. Primarily, blood pressure is lowered by vasodilation. Another important characteristic of Ang (1-7) is its antioxidant, anti-inflammatory (and antiarrhyhtmogenic) effect on the organism. ACE2 is a single pass type I protein I the membrane whose enzymatically active domain is exposed on the surface of cells in the lung and other tissues (Fig. 35).

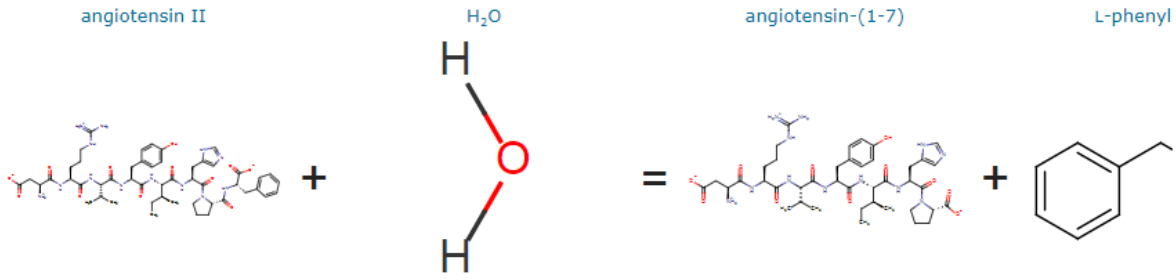


Fig 35: Catalytic activity of Angiotensin-converting enzyme 2
(https://www.uniprot.org/uniprot/Q9BYF1)

The extracellular domain of ACE2 is separated

from the transmembrane domain by another enzyme, called sheddase, and the resulting soluble protein is discharged into the bloodstream and eventually excreted in the urine.

This has led some to hypothesize that lowering ACE2 levels in cells may help fight infection. On the other hand, ACE2 has a protective effect against virus-induced lung injury by increasing the production of the vasodilator angiotensin 1-7. Moreover, interaction of coronavirus spike protein with ACE2 induces a decrease in ACE2 levels in cells through internalization but also through degradation of the protein and thus may contribute to lung injury.

Since the gene for ACE2 is located on the reverse strand of the X chromosome (X: 15,494,566-15,607,236), Gagliardi et al. (2020) assume that this explains the different mortality rates of women and men (with the same infection rate). The intersexual differences in lifestyle (smoking, diet) should not be neglected.

Moreover, the authors emphasize that other hormonal factors, such as the presence of estrogens, which enhance the immune response and inhibit viral replication more, should be included. Nevertheless, the location on the X chromosome could be of crucial importance:

> *Furthermore, the gene encoding ACE2 is located on the X chromosome, in sites commonly escaping the inactivation of one X chromosome in mammalian XX cells (XCI), a mechanism that determines the X chromosome transcriptional silencing and avoids redundant gene expression in female cells. However, the silencing is not complete but about 10% of the genes escape the inactivation; as a consequence XX, cells over-express genes located in XCI sites, like ACE2. (Gagliardi et al., 2020)*

Consequently, there might be an overexpression of ACE2 in women. However, other studies, such as that of Sama et al. (2020), assume that the concentration of soluble ACE2 receptors in the blood is much higher in men, which probably has to do with the fact that ACE2 is expressed just as strongly in the testis. These studies seem to contradict each other only at first sight, because while Gagliardi et al. investigate membrane-bound ACE2 receptors, Sama et al. focus on ACE2 concentration in blood, i.e., soluble ACE2 receptors that have been "shed" from the membrane. Key role plays hereafter the Sheddase ADAM17, Sama and Voors (2020) explain this contradiction as follows:

> *The ACE2 gene is located on the X chromosome and is expressed in various tissues, including the heart, kidneys, and testes.1 Endogenous soluble ACE2 (found in the circulation) is shed from the cell membrane-bound form and the enzyme responsible for this shedding is ADAM17, which is also membrane anchored. We recently postulated that the co-expression of ACE2 and ADAM17 in the testes might partially explain why plasma ACE2 concentrations are higher in men than in women. (Sama & Voors, 2020)*

Both ACE blockers and angiotensin II receptor blockers (ARBs) used to treat hypertension have been shown to upregulate ACE2 expression, potentially affecting the severity of coronavirus infections. Despite the lack of conclusive evidence, some have argued for and against discontinuing ACE blocker or ARB treatment in COVID-19 patients with hypertension. However, several professional

societies and regulatory agencies have recommended continuation of standard therapy with ACE inhibitors and ARBs.

7.2 Alu elements

Alu elements, with more than one million copies, are the most abundant repetitive sequences that could be counted as transposons, accounting for 10% of the total genomic mass. As already vividly explained in the ACE1 gene, their presence, even in non-proteinogenic intronic sequences, is not without significance for the regulation gene expression. Häsler and Strub (2006) explicate three levels of regulation of Alu elements:

1. They influence the splicing process
2. They are involved in RNA editing
3. They act at the post-transcriptional level in translational regulation.

7.2.1 Role of Alu Elements in alternative splicing

Alternative splicing allows the creation of a variety of translational isoforms after transcription of a pre-mRNA. This also explains the proteome diversity, since 30-60% of all human genes produce alternative exons. Exonication is what Häsler and Strub (2006) call the process by which intronic sequences become coding regions. This is also how thousands of Alu elements end up in the coding regions of mature mRNAs. Due to the fact that several 5'/3' splice sites are found in the Alu consensus sequence (especially in antisense orientation like in intron 16 of the ACE1 gene), the whole Alu element does not have to be taken over during exonication. This can also lead to diseases, as in the example of the CTDP1 gene, where an alternatively spliced Alu exon in intron 6 is exonified: in this case to the, CCFDN syndrome (Congenital Cataracts Facial Dysmorphism Neuropathy) (Varon et al., 2003).

For our considerations, exonication possibilities in the ACE1 gene are of particular interest. Here, antisense orientation in particular plays a decisive role, as this reinforces exonication (Lei, 2005).

7.2.2 Role of Alu Elements in RNA Editing

RNA editing involves co- or post-transcriptional manipulation of RNA molecules. "*The best-characterized base conversions are hydrolytic deamination reactions by which cytosines are converted to uracil and adenosine (A) to inosine (I).*" (Häsler & Strub, 2006). Therefore, the researchers also speak of A-I editing and emphasize the role of Alu elements, since adenosines 27, 28, 136 and 162 of the Alu consensus sequence are converted into iosines by hydrolytic deamination reactions by adenosine deaminases (ADAR). The preferential editing of Alu- sequences within mRNAs could initially be due to the secondary structure of Alu- RNA, which contains long double-stranded regions. RNA editing affects gene expression in several steps: on the one hand, inosines do not pair with uracil but with cytosine, which affects the stability of the RNA molecules; on top of that, if splicing/translation occurs, the Alu inosines are recognized as guanosines, which results in alternative splicing or amino acid substitution.

7.2.3 Role of Alu elements in protein biosynthesis

Alu elements are translated in the pre-transcript of the mRNA and can also enter the cytosol in small amounts after the splicing process. Häsler & Strub (2006) assume 10,000-100,000 molecules per cell. Interestingly, this concentration is not constant: if there is a stress response (viral infection, heat shock...) the expression of Alu elements is abruptly stimulated. The researchers therefore assume a physiological function in stress management. One possible explanation is that Alu RNA transcripts interact with RNA-dependent protein kinases (PKR). On the one hand, they inhibit their autophosphorylation and thus their inhibitory effect on the initiation of translation of proteins. However, they can also have a direct activating or repressive effect on translation. Yulug and Fisher (1995) prove already in the 90 years that Alu elements can be detected in the UTR sequences of genes. Accordingly, they are located between the promoter region and the translation start and are transcribed by RNA Pol. II. Smith et al. (1996) describe this vividly using the BRCA1 gene: this exists in two isoforms that differ in their expression pattern. In fact, they differed in the length of the 3'-UTR region. In the longer UTR region, there is actually an insertion of an Alu sequence, this affects

expression as the gene with longer UTR is less efficiently translated. This is because the Alu element that is not fully inserted (60 bases of the left monomer are missing), or the right monomer, forms a stable secondary structure through double bonds that partially inhibits translation initiation. Consequently, the concentration of BRCA1 proteins decreases, their function as repair molecules is less pronounced, accumulation of defects in DNA occurs. The high mutation burden eventually leads to the onset of tumor disease (in the sense of the two-hit hypothesis). This example clearly shows that in such a case a whole-exome analysis of BRCA1 alleles would not have detected trait-related variants. De Brakeleer, De Grève and Teugels (2020) therefore require:

> *Moreover, the analysis of a near exonic Alu insertion in PTPN14 (a mediator of P53 tumor suppressor activity) revealed that this gene was imprinted and that the presence of an intronic Alu element can lead to loss of imprinting. Our data underline the relevance of incorporating the search for uncommon retrotransposition events in Next Generation Sequencing pipelines when analyzing patients with a suspected genetic disease.(De Brakeleer, De Grève and Teugels, 2020).*

7.3 ACE (I/D) polymorphism and COVID-19

Two recent studies in 2020 address the ACE1 (I/D) polymorphism and COVID-19 infection incidence in regional populations. Delanghe, Speeckaert and De Buyzere (2020) conduct a meta-study using published results from the Johns Hopkins Institute (March 2020) and conclude that in the 33 countries studied, including Austria, the prevalence of COVID-19 correlated negatively with the allele frequency of the D polymorphism of the ACE1 gene. "*Our findings suggest that the ACE1 D/I polymorphism may be regarded as a confounder in the spread of COVID19 and the outcome of the infection.*"

The researchers observe a global east-west gradient, which also applies regionally to Europe:

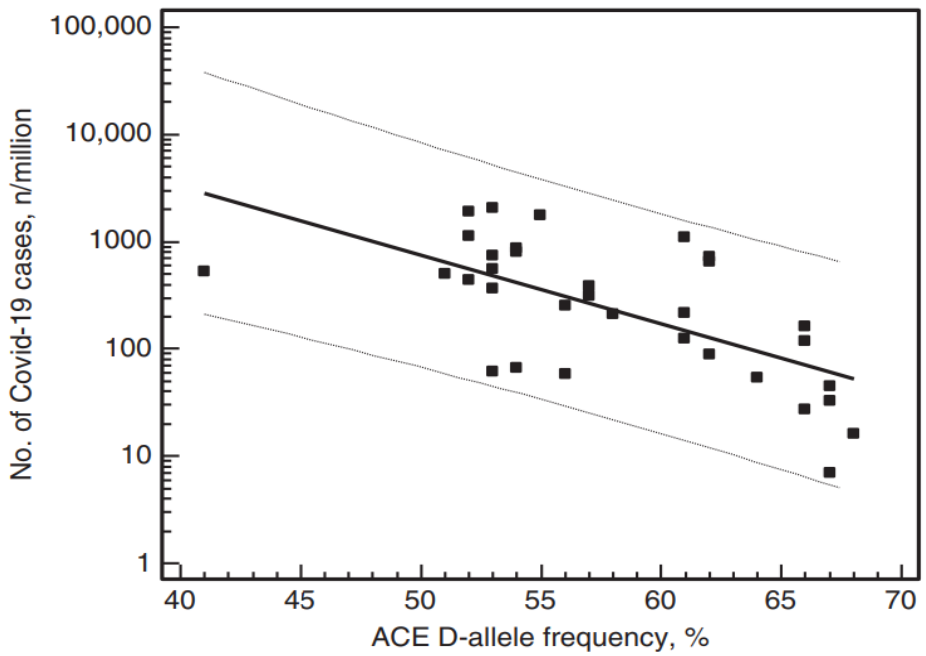


Figure 1: Prevalence of COVID-19 in 33 countries (on April 1, 2020) vs. ACE1 D-allele frequency (%): log (prevalence; no. of cases/10^6 inhabitants) = 6.074 – 0.064 (D-allele frequency, %), $r^2 = 0.410$; $p = 0.0001$.

Deletion of the Alu element from intron 16 of the ACE1 gene seems to have a direct effect on the expression of the ACE2 gene, which is reduced in ACE1-D alleles. On the one hand, this would reduce the entry possibilities of SARS-CoV - which inhibits the replication of the virus, on the other hand, a deficiency of ACE2 enzymes at the surface of epithelial cells leads to an overactivity of AngII, thus repeatedly demonstrating the ambiguous role of gene expression patterns of involved proteins of different pathways.

The study by Hatami et al. (June 2020) is somewhat more recent. Again, the researchers collected data from over 50,000 female subjects in over 30 countries and evaluated them in a meta-analysis for the correlation of two variables: Association between ACE1 allele frequency (I/D) and recovery rate after COVID-19 infection. For this purpose, they analyze 116 studies (out of a total of 1400), among them also case numbers from Austria.

In your conclusion, a specific geographical distribution of ACE1 (I/D) alleles is repeatedly shown. While in Europe (except Denmark) the D alleles without Alu element are predominant, in Asia (especially in China and Japan) significantly more alleles with insertions are found. Especially homozygote II, which are a minority in Europe, are widespread in China. This is lowest throughout Europe in Italy, which the researchers associate with the dramatic escalations in the spring of 2020.

> *European countries like Italy, Spain, and the UK have been widely affected by COVID-19. It is well known that ACE I allele frequency in Europe is lower than in East Asia as seen in our study, which indicated the average ACE I/D allele ratio in Europe as 0.55, while it was 0.93 in the whole of Asia. (Hatami et al., 2020)*

These results are not only interesting in the light of the COVID-19 pandemic but play a crucial role in the analysis of any disease in which the RAAS is affected. However, the researchers also clearly point out that other known and unknown factors play a role (demographic characteristics such as gender, age, underlying diseases), as well as the collection, accessibility, and presentation of COVID-19 data in each country.

Allele frequency data are also available for Austria from 2007, collected by Ay et al. (2007) at the Medical University of Vienna. They collected ACE1 (I/D) polymorphism by genotyping and ACE serum levels by enzymatic assays in 100 high-risk patients with

confirmed recurrent venous thromboembolism. Excitingly in this context, data on pulmonary embolism patients were also collected. 125 Age- and sex-matched healthy subjects served as controls. ACE genotype frequencies differed between the patients (DD: 26.0%, ID: 52.0%, II: 22.0%) and the control group (DD: 29.6%, ID: 44.8%, II: 25.6%; p = 0.56). Although the distribution is basically similar (ID>DD>II), more people in the control group are homozygous II. Serum ACE levels (U/) did not differ between the patients (median = 25.25, 25th-75th percentile: 20.20-33.70) and the control group (24.20, 17.85-34.50, p = 0.49). In the overall population involved in the study, the ACE-DD genotype (n = 63: 36.00 [26.40-43.00]) was linked to a hiher ACE levels than the ACE-ID genotype (n = 108: 24.10 [19.80-31.48], p <0.001)) and the ACE II genotype (n = 54: 19.35 [15.00-22.95], p <0.001). In this context, the significant correlation between polymorphism and serum ACE level is interesting. The association with thrombosis or embolism could be neglected. Studies relating the ACE1 polymorphism with simultaneous consideration of other factors in the sense of a correlation analysis to the incidence of infection in Austria are pending. However, it is already possible to predict the ambiguous role of ACE1 and ACE2. Especially ACE2 has an ambivalent function here, because as a portal of entry for the virus it could be a target for therapeutic approaches, but as an anti-inflammatory, antioxidant antagonist in the RAAS it cannot be a primary target of inhibitory therapy. More promising might be the manipulation of TMPRSS2 as well as of sheddase ADAM17.

8. Transposons in tumorigenesis

The development of malignant tumors is currently regarded as a kind of "genetic accident", whereby direct DNA damage occurs as a result of external influences. Mutations thus created lead to activation of oncogenes and inactivation of tumor suppressor genes (Weinberg, 1994). The resulting disruption of gene expression leads to multiple changes such as abnormal cell proliferation, resistance of cells to apoptotic stimuli, induction of neovascularization and metastasis. This model is also known as somatic mutation theory "*According to the somatic mutation theory (SMT), cancer begins with a genetic change in a single cell that passes it on to its progeny, thereby generating a clone of malignant cells*" (Vaux, 2011).

However, there are a number of criticisms of this theory (Brücher &Jamall, 2016). For example, one would expect a mutation to have an immediate phenotypic effect. However, tumorigenesis is a long process both experimentally and in a person's life. Furthermore, multiple mutations are thought to be required for tumor induction.

> *The hysteron proteron of the somatic mutation theory (SMT) appears because the first event (mutations), in fact, occur later in the process, i.e., only after the cell has been transformed from a normal cell to a cancer cell via a process termed carcinogenesis. Mutations have increasingly been perceived as the causal event in the origin of the vast majority of cancers even as clinical data show little support for this theory when compared against the metrics of patient outcomes. (Brücher and Jamall, 2016)*

However, the spontaneous mutation rate of a normal cell is relatively low, and the same cell being hit many times by mutations is extremely unlikely. Therefore, the existence of a mutator phenotype for cells hit by a mutation has been hypothesized. Such mutator phenotypes are described to arise from defects in the DNA repair system and the cell cycle control system, such as the p53 gene, which inhibits proliferation of cells with abnormal DNA content. But neither does the loss of p53 necessarily lead to the development of cancer, nor does a functional p53 safely protect against it.

Another criticism of the somatic mutation hypothesis is that no clear correlation between mutagenicity and carcinogenicity of specific agents has been demonstrated. Also, there are no tumor-specific mutations. Genomic instability is common in malignant neoplasms and is thought to be an early step in carcinogenesis. The complex phenotype of a cancer cell's genome cannot be explained by mutations alone. Another conception of the origin of malignancies views carcinogenesis as an evolutionary process in terms of aberrant adaptation.

Activation of transposable elements as part of adaptive responses to cellular stress has been described for various organisms. Shapiro (2014) paraphrases the genomic changes resulting from this activation as active DNA engineering. In the human genome, the L1 retrotransposons are of particular importance in this context, as they are the only independently active mobile genetic elements here. Activation of the L1 elements in the context of carcinogenesis would thus be assumed. This could also explain the development of the unstable karyotype of tumor cells. In the following section, the interrelationship between L1 transposons and tumorigenesis will therefore be discussed in more detail.

8.1 Structure of the LINE-1 retrotransposons

As indicated in several places, all autonomous and active retrotransposons belong to the group of LINEs (long interspaced nuclear elements), which can be divided into LINE-1, LINE-2, and LINE-3. Of the named, LINE-1, which account for 17% of the human genome, are the most abundant and effective mobile elements (Hancks and Kazazian, 2016). They replicate in the genome via the "copy-and-paste" retrotranscription mechanism explained previously. Of the 500,00 copies still in our genome according to Kazazian and Moran (2017), only 100 L1 copies are still capable of transposition in the human genome (Brouha et al., 2003). These retrotransposons, which are still intact, are approximately 6kb long and are transcribed starting from an internal promoter and later translated into two large protein complexes: ORF1p and ORF2p (Sultana et al., 2019). The ribonucleoprotein particle (RNP) from the protein products in combination with the L1 RNA in the cytoplasm also contains two important catalytic centers: on the one hand, as an endonuclease, it catalyzes the cleavage of DNA at the sequence-

specific target site; on the other hand, the reverse transcriptase is responsible for the transcription of L1 mRNA into DNA to be integrated (Viollet, Monot and Cristofari, 2014). Accordingly, ORF2p combines an endonuclease (EN) and a reverse transcriptase (RT). Figure 36 shows the structure of the individual building blocks of the L1 element:

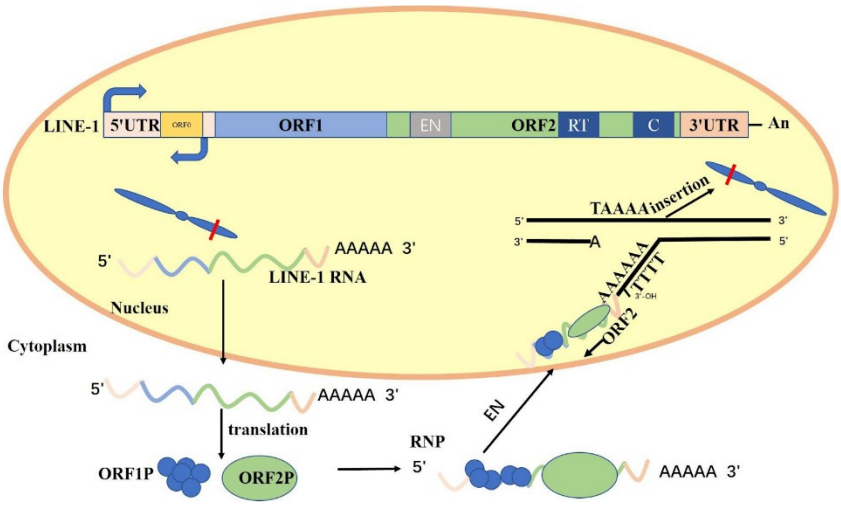


Fig. 36: Structure of the individual building blocks of the L1 element, (Zhang, Zhang and Yu, 2020).

At the 5' end there is an untranslated region (5'UTR). The promoter of the element is located here. This is followed by two open reading frames: ORF1 and ORF2. The untranslated region at the 3' end (3'UTR) ends with a polyadenylation signal. Flanking an L1 element are sequence duplications, so-called target site duplications (TSD). In the figure, one can also see the life cycle of an L1 element. (Zhang, Zhang & Yu, 2020).

In addition to ORF1 and ORF2, Figure X also shows primate specific ORF0, which is still found in the 5'UTR region and is oriented in the antisense direction. Denli et al. (2015) found ORF0 transcripts in primate induced pluripotent stem (iPS) cells and concluded that ORF0 forces L1 mobility and thus also plays an important role in retrotransposon-associated genome instability or diversity.

The "life cycle" of an L1 element, as described by Rodić and Burns (2013), includes the following steps:

> *The L1 life cycle entails three steps. The first step is the transcription of a genomic L1 into RNA, which is mediated by RNA polymerase II from an internal L1 promoter. Transcription from an internal antisense L1 promoter may occur concurrently. In the second step, the RNA is translated into two L1-encoded proteins: ORF1p, an RNA-binding protein, and ORF2p, a protein with reverse transcriptase and endonuclease activities. These proteins associate with the L1 transcript and the resulting ribonucleoprotein (RNP) complexes are then transferred to the nucleus. The third step is termed target-primed reverse transcription (TPRT). In the course of TPRT, ORF2p cleaves the target DNA (often at a 5′-TTTTAA-3′ consensus sequence) and uses the 3′ hydroxyl group to prime the reverse transcription reaction. Synthesis of the second strand and resolution of the structure is poorly understood. Because the L1 life cycle generates DNA breaks, cell host proteins that mediate DNA repair are likely involved." (Rodić and Burns, 2013:1)*

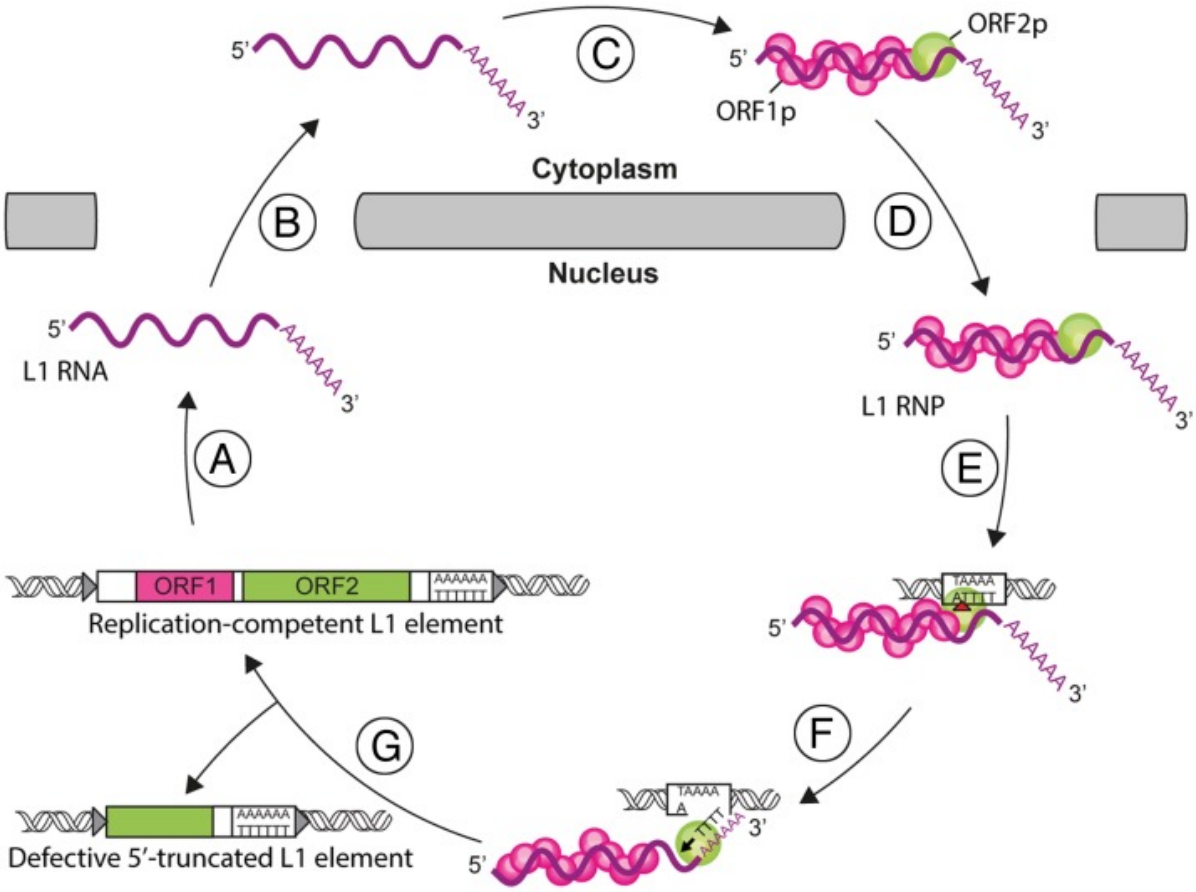


Fig 37: Summary of the life cycle of an L1 retrotransposon according to Viollet, Monot and Cristofari (2014).

Viollet, Monot and Cristofari (2014) identify a total of seven steps (Fig. 37)in the life cycle of an inactivated L1 retro transposon, although precisely "*The mechanisms involved in the final steps of this process and the resolution of the integration are unresolved yet*.":

Transcription: The individual factors required for transcription and which RNA polymerase is used have not yet been fully clarified for human cells. Polymerase II-mediated transcription has been shown for the retrotransposon "jockey" in Drosophila melanogaster. It is considered certain that the YY1 and RUNX3 transcription factors in the human genome are involved. The promoter region is located in the 5'UTR. The transcription ends at the poly-A sequence, sometimes only at a stronger polyadenylation signal located further downstream. This can lead to the entrainment of neighboring sequences (transduction) during retrotransposition. Evolutionarily important sequences may have been shifted by this mechanism (Yang, 2003). After transcription, the L1 RNA is exported from the nucleus. Since no introns are contained, splicing is omitted,

Translation: The first open reading frame (ORF1) encodes a 40 kD protein (p40 protein). It is a nucleic acid binding protein that has a chaperone function for L1 RNA and is essential for retrotransposition. ORF2 encodes a 150 kD protein that contains an endonuclease domain, a reverse transcriptase domain, and a zinc finger domain. The ORF2 protein is responsible for reverse transcription and for insertion of the transposed element into the genome. In the cytoplasm, the protein products of the two ORFs are assembled and together with L1 RNA now form the RNP (Wang and Jordan, 2018).

Entry into the nucleus: How the ORF2 proteins, or the ribonucleoprotein particles, enter the nucleus is also still a matter of debate since proteins heavier than about 60 kD cannot enter the nucleus by passive diffusion through nuclear pores. One explanatory model is that the ORF2 protein, or the entire RNP complex, enters the nucleus only during mitosis or meiosis, i.e., when the

integrity of the nuclear membrane is lost. Mita et al. (2018) tested this hypothesis using functional and biochemical imaging, respectively, and were able to prove that L1 is start mobile, especially in the S phase of mitosis. However, these results mainly refer to ORF1, which could be detected using antibodies; whether this also applies to ORF2 to the same extent is the subject of research.

Endonuclease-meditated **DNA nick: According** to current thinking, after a DNA strand is cut by the endonuclease, the 3'-OH end of the genomic DNA is used as a primer for reverse transcription. The process is therefore also referred to as target primed reverse transcription (TPRT). In this process, ORF2-endonuclase detects the consensus sequence 3-′AA/TTTT-5′ and cleaves the DNA at this site. The free 3'-OH end resulting from the cleavage can now be used as a template for reverse transcription in the next step. Viollet, Monot and Cristofari (2014) have found in vitro that ORFsp dimerizes and thus catalyzes DNA cleavage.

Reverse transcription: reverse transcription exposes a 3'-OH overhang as initiation. In research, this step is often compared to that in which telomerase: the reverse **transcriptase** enzyme is primed by the L1-poly adenyl tail, which now hybridizes to the 5′-TTTTA-3′ at the "sticky end" (Richardson et al., 2014). After reverse transcription of the strand, the DNA RNA is interrupted, and now the complimentary DNA strand can be replicated.

Integration: TPRT results in sequence duplication at the site of integration of the element (**target** site duplication, TSD) (Gilbert et al., 2005).

Truncation: Most L1 elements are truncated at the 5` end. This can be explained by premature dissociation of reverse transcriptase from RNA. Partial degradation of L1 RNA by RNAses before reverse transcription is complete is also conceivable. Furthermore, a process known as twin priming is discussed. Here, the second DNA strand is also cleaved during integration and the 3'-OH end functions as a second primer for reverse transcription.

8.2 Life cycle of an L1 element

Ichiyanagi and Okada (2006) use the target analysis of nested transposons (TANT) to further investigate the pathway of target-primed reverse transcription (TRPT) in the vertebrate model organism zebrafish.

Figure 38 explains the L1 retrotransposition in more detail:

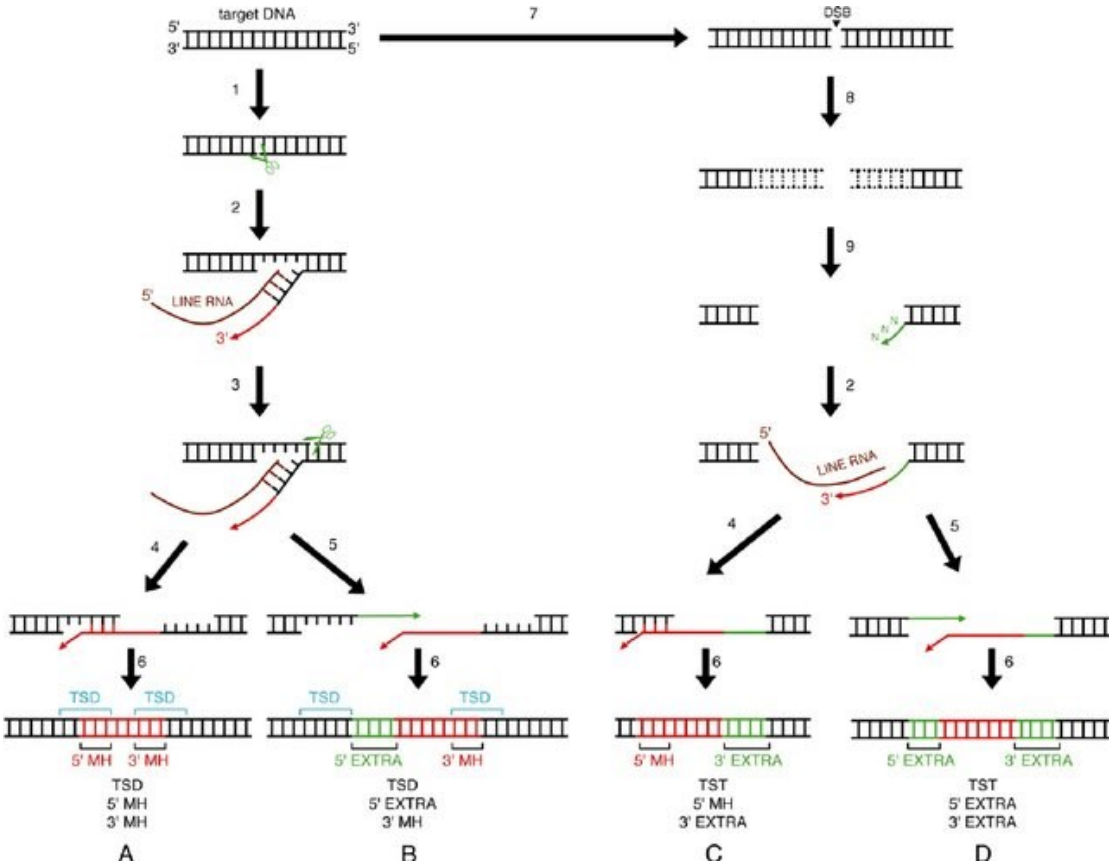


Fig. 38: The life cycle of an L1 retrotransposon. Target DNA (black), L1 RNA (brown), newly synthesized L1 cDNA (red), additional Nt (green). DSB (double-strand break); EN (endodeoxyribonuclease); MH (microhomology); RT (reverse transcriptase); TANT (target analysis ofnested transposons); TPRT(target-primed reverse transcription); TSD (target-site duplication); TST(target-site truncation). Adapted from Ichiyanagi and Okada (2006).

1. Single stranded break of target DNA by an L1-encoded endonuclease on the antisense strand (3′→5′).
2. Reverse transcription of L1 RNA, reverse transcriptase now adds one deoxyribonucleotide complementary to the RNA at each 3'-OH, which is now read from 3' to 5'.
3. DNA sense strand (5′→3') is now also cleaved.
4. Fully synthesized L1 cDNA hybridized with target DNA
5. Addition of new nucleotides at the 3' end of the DNA sense strand.
6. Synthesis of the sense strand (see 4) after annealing of the L1 cDNA and ligation to form target-site duplications (TSDs) that are approximately 11-19 bp long.
7. Truncations, on the other hand, are the result of a double-strand break (DSB) at the target DNA.
8. This leads to digestion of nucleotides by exonucleases.
9. Nucleotides are now added at the 3' ends, which on the one hand serve as primers for the reverse transcriptase at the antisense strand and on the other hand form an overhang at the sense strand. Not all of the L1 RNA is transcribed into DNA by RT, but there is a truncation of about 11 to 666 bp at the 5' end. This is referred to as target-site truncation (TST).

At this point, it is important to note that there are significant differences between the L1 dynamics of mammals and other vertebrates. For example, in the zebrafish study Ichiyanagi and Okada (2006), the L1 reverse transcriptase is only capable of rewriting its own L1 RNA and L1 cDNA. In the human genome, we know that L1 transcriptase can transcribe any RNA substrate provided it has a poly-adenyl sequence. This also explains the particular success of L1 elements in primates and the interaction with other mobile elements such as the ALU sequences.

8.3 Consequences of L1 retrotransposon activity.

The activity of L1 elements, i.e., the occurrence of retrotransposition, can have both destructive and constructive effects on the genome of the corresponding cells. The destructive processes include, above all, insertion (so-called insertion mutagenesis by integration of a LINE1 into coding regions of a gene). For example, insertion mutagenesis has been described for the factor VIII gene, resulting in hemophilia A. Proteins encoded by L1 elements not only interact with their own RNA (cis), but are also responsible for mobilizing nonautonomous retrotransposons, e.g., Alu elements (trans). Homologous recombination between L1 or Alu elements can lead to deletions or duplications of DNA segments. More extensive genomic rearrangements such as large deletions and chromosomal inversions have also been described. These processes can significantly affect genome stability. Other effects of L1 activity are considered to be constructive. For example, together with the L1 element at the 3' end, adjacent sequences can be carried along during retrotransposition (transduction). In this way, exons (exon shuffling) and regulatory sequences can be moved to other locations in the genome. L1 elements can influence the expression of neighboring genes via an antisense promoter in the 5'UTR region. Protein-coding sequences may also have originated in retrotransposons: This has been described, for example, for telomerase, the enzyme maintaining the length of telomeres, the ends of chromosomes. Telomerase is structurally closely related to the L1 RT. An evolutionary benefit is attributed to these constructive mechanisms.

In the following subchapters, the entire range of consequences in which LINE-1 retrotransposition is driver or at least co-regulator will be presented. Insights into evolutionary but also pathological consequences of the activity of L1 retrotransposons will be presented. The focus is on benign and malignant diseases in the fields of hematopathology, immunology, neurology and finally oncology, the latter will be dealt with in a separate chapter.

8.4 Role of Retrotransposons in Hematology

The initial choice of hematopathology as a field of investigation for L1 retrotransposition activity is mainly due to the fact that this is the first described L1-associated disease in the history of human medicine. In 1988, Kazazian and colleagues at the Johns Hopkins University School of Medicine described two cases in which a defective factor VIII gene, an essential blood clotting factor, leads to hemophilia A. The first case was in the United States.

Hemophilia A is an X-linked hereditary disease in which blood clotting is disturbed. Due to the inheritance, primarily male offspring are affected. However, in the said two cases it is a de novo mutation that was absent in the parents, both also have in common that an insertion of 2-4 kb in exon 14 is characteristic. Kazian et al (1988) examined both insertions and found the following: it is not a disruption due to expansion of short repeats, both insertions have an adenosine-rich sequence (77 nucleotides) in close proximity to the 3' breakpoints on the antisense strand. This provided the first clue that this is probably an insertional mutation with an RNA intermediate reverse transcribed into exon 14. Also conspicuous were 12-13 nucleotides that flanked the insertion and were later identified as target sites duplications.

With the aid of a homology comparison of the sequence studied with L1 sequences published by Skowronski and Singer in 1986, the insertional mutation could be attributed to the retrotransposition activity of an L1 element: " *These L1 insertions are the first large non-viral insertions described in man which are not due to expansion of short repeats by unequal crossing over events*" (Kazazian et al., 1988: 164).

This is the first time that the relationship between transposable activities of mobile genetic elements and diseases has been empirically established. Kazazian et al. (1988) also relate these results to the research of Fanning and Singer (1987), who concluded a year earlier in their research on RNA of human teratocarcinoma that a small number of intact L1 transposons are transcribed and translated in the germline: "*Because we do not know when these L1 insertions events occur, whether in the sperm or ovum, after fertilization or during early stages of*

embryogenesis, the proportion of such insertions that are heritable is unknown" (Kazazian et al., 1988: 166). In addition to retrotransposon-induced coagulopathies, the pathological influence of retrotransposons on the production of red blood cells has also been demonstrated - as early as 1987, Nicholls and colleagues demonstrated that ß-thalassemia lead to deletions in the α-globin gene due to an exonic L1 insertion during *Alu element-mediated* recombination with two Alu breakpoints.

A decade later, Schneider et al. (2009) present in their review article retrotransposon-associated hematologic diseases, because besides the first described hemophilia and thalassemia, polymorphic intronic insertions of L1 transposons trigger mainly neoplastic haematopathologies. These are primarily leukemias, malignant lymphomas, and myelomas. The majority of these are sporadic de novo mutations resulting from chromosomal translocations or other genetic instabilities in the somatic genome. Schneider et al. group the effects of retrotransposition in haemato-oncology at three levels: (1) epigenetic interference, (2) alteration of gene expression, (3) insertional mutations in coding regions:

1. **Epigenetics:** transposition capacity is primarily epigenetically regulated and here mainly repressed by the methylation of CpG islands in mobile elements. Indeed, a covalently bound methyl residue at the 5-C of cytosine is found in the CG dinucleotides of mobile genetic elements in the genome of non-cancerous somatic and differentiated cells, which are already transferred in embryogenesis by DNA methyltransferases (DNMT3a/DNMT3b). Zhu et al. (2006) demonstrate another protein, LSH, plays an important role in maintenance methylation by DNMT1 and de novo methylation by DNMT3a/3b. It is a helicase that is equally involved in chromatin modification (Han et al., 2020). Low activity of LHS leads to less efficient methylation of DNA and histones and eventually to increased activity of transposons due to insufficient epigenetic repression. This could be demonstrated for chronic lymphocytic leukemia (CLL), multiple myeloma (MM) as well as chronic myeloid leukemia (CML). The relationship between hypomethylation of L1

retrotransposons and CML has been particularly well demonstrated by Roman-Gomez et al. (2005): using methylation-specific PCR, they demonstrated that in 75% of cases of blast crisis in CML, hypomethylation of L1 could be detected - as well as empirically demonstrating that in the case of hypomethylation, a poorer prognosis for CML can also be expected due to poorer response to therapy. Therefore, Schneider and colleagues suggest that epigenetic therapies should also be considered for haemato-oncological neoplasms - and refer to the successful implementation of this in the treatment of myelodysplastic syndrome (MDS): by inhibiting DNA methyltransferases and histone deacetylases, an "epigenetic therapy" was developed that provided better results than classical chemotherapies (Griffiths & Gore, 2008).

2. **Gene expression:** active retrotransposons can also strongly influence gene expression by intronic insertion. This leads, among other things, to deregulation of (a) transcription (insertions in the promoter region, or alteration of non-transcribable cis elements such as enhancers or silencers), of (b) splicing by insertions in splice sites or branches, to (c) exon skipping or (d) exonization, and can also influence the length of the mRNA. For leukemias, in addition to mobile L1 transposons, mobile Alu elements are particularly dangerous.
3. **Insertional mutagens:** Clearly, insertion into coding sequences is not without consequences. It has already been demonstrated in animal models that reverse transcription of L1 RNA into coding segments of the Notch1 gene is contributory or potentiating for T-cell lymphoblastic leukemia. This is a gene encoding the transmembrane receptor, which is important for normal T-cell development. Activation and transposition of (hypomethylated) retrotransposons in this gene are associated with T-cell acute lymphoblastic leukemia (T-ALL), as Howard et al. (2007) could prove in a mouse model.

8.5 Retrotransposons in tumorigenesis.

The link between tumorigenesis and retrotransposition is obvious, as both have been identified in the context of genetic instability. This was already established in 1992 by Miki et al. using the example of a somatic L1 insertion in the tumor suppressor gene APC, whose loss of function is responsible for 85% of colorectal carcinomas (CRC). The researchers detected a 750 bp insertion in exon 16 by Southern blotting using cDNA clones from the APC gene as probes. Notably, there was a polyadenylate tract (nearly 180 bp) and an 8-bp flanking target site duplication.

Similar studies in other tumors, show the relative abundance of inserted retroelements in tumor suppressor genes or oncogenes. Scott et al. (2016) recently addressed the question of whether this studied L1 insertional mutagenesis in the APC gene is a passenger mutation, i.e., a mere concomitant due to the genetic instability of cancerous genomes and only become active during the more advanced stages of tumor progression and metastasis - or whether L1 insertional mutagenesis is causative for the onset of cancer and thus initiates tumorigenesis. To test this, they perform a series of tests on a CRC patient who has an L1 insertion in the previously described exon 16 of the APC gene. They scan the genome of the tumor cells for other L1 insertions, examining all CRC candidate genes to reconstruct the process of tumorigenesis. It should be prefaced, however, that APC as a tumor suppressor gene must have function-impairing mutations on both alleles; the mutational status of the second allele was not elicited in Miki et al. (1992). In addition, it is important to consider that two subtypes are distinguished in CRC: whereas the more common MSS (microsatellite stability) phenotype arises from the causative APC mutation, the MSI (microsatellite instability) phenotype is initiated by hypermutation leading to faulty DNA repair (Nojadeh et al., 2018). To rule out that L- retrotransposition was activated by genomic instability, researchers work with CRC of the MSS phenotype.

To illuminate the role of L1 insertional mutagenesis in CRC tumorigenesis, Scott et al. (2016):

a. All CRC samples are scanned using an L1-Seq assay. In one patient, an L1 insertion is found in codon 1396 of exon 16 of the APC gene. Using Sanger sequencing, the L1 insertion was determined to be 1378 bp in length. It ends with a ~27-bp long polyadenylate tract and is flanked by two 14-bp target side duplications (TSD). The insertion is truncated. Further PCR examination of the DNA of the adjacent tissue confirms that the insertion is found only in cancerous tissue.
b. Using NGS whole genome sequencing, the entire genome of the patient is now examined. The genome of non-cancerous cells serves as a reference genome, against which the DNA of the tumor cells is mapped. Comparison of non-cancerous and cancerous DNA reveals an additional 26 L1 insertions distributed across the tumor DNA. All of them show both a polyadenylate tract and flanking TSD, clearly indicating an RNA intermediate and target primed reverse transcription (TPRT).
c. Using the WGS, a stop codon was also identified in the APC gene just 160 bp downstream from the somatic L1 insertion. However, this is located on the other allele. Now it is clear: while one allele is interrupted by the L1 insertion, the other allele fails due to a truncating nonsense mutation.
d. However, WGS identified additional pathogenic variants in CRC-associated genes: in the PIK3CA gene (a point mutation leading to an amino acid substitution), in the KRAS gene (a tandem duplication) and in ACVR18 (a deletion leads to a frameshift mutation). All three mutations were identified in different tumor locations but were absent in the adventitious tissue. This suggests an intimate role in early tumorigenesis but not necessarily primary initiation.
e. Using the Mobile Element Locator Tool (MELT) developed for the 1000 Genomes Project (Sudmant et al., 2015), a total of 308 L1 elements were identified in the patient's genome and subsequently sequenced. By comparing the APC-L1 insertion with the identified L1 retrotransposons, the L1 source element could now be identified: it is a retrotransposon on chromosome 17, which has two intact ORFs and thus also appears to be active.

f. Using strand-specific RNA-seq, it was possible to prove that the source retroelement of chromosome 17 is active and transcribed in all somatic cells of the patient, whereas this is not the case in unaffected healthy individuals. Sequencing of the retroelement revealed a point mutation in one of the four CpG islands in the promoter region (G61A).
g. bisulfite sequencing revealed that the promoter is hypomethylated due to the loss of a CpG island, allowing the retroelement to escape epigenetic repression.
h. by means of allele frequency comparison, it could be determined that this point mutation (G61A) in the L1 element is specific to Central African populations or African-American descendants. Thus, it is an inherited mutation.

As shown in figure 39, this seems to have been the starting point for tumorigenesis. Due to the nonsense mutation at the other APC allele, the tumorsurpressor gene loses its function and further mutations like this one at PIK3CA or KRAs accumulate and lead to polyp formation, eventually resulting in adenocarcinoma:

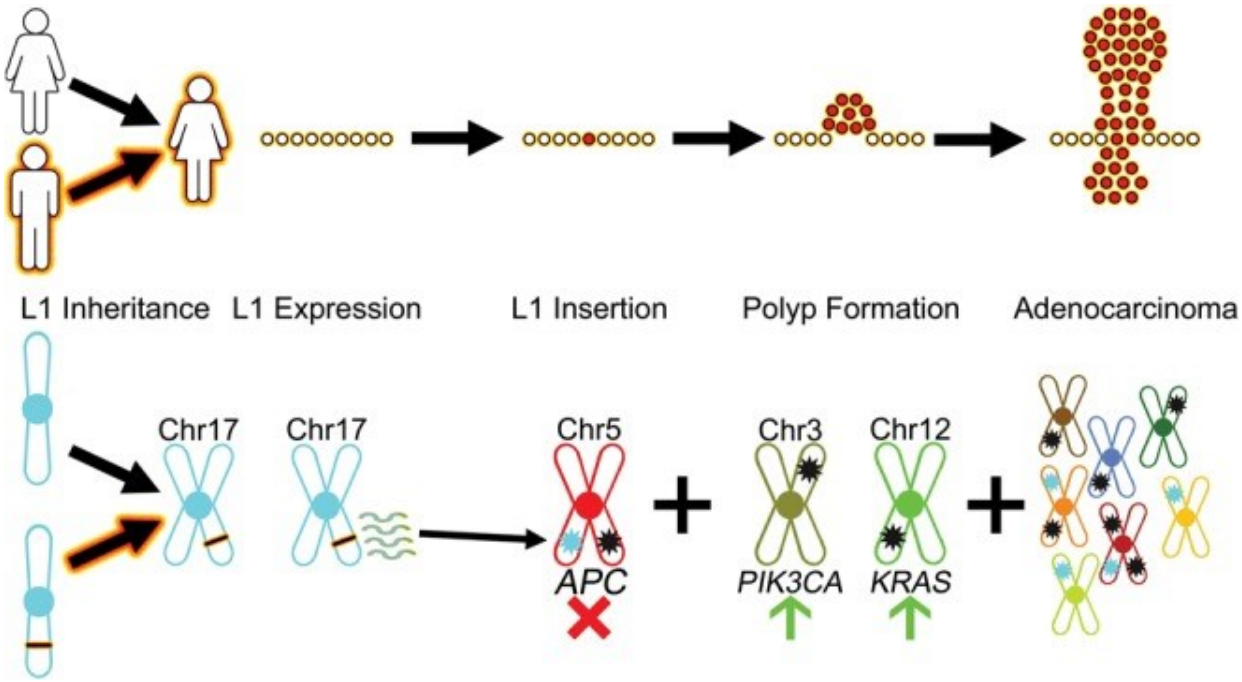


Fig. 39: Figure shows the role of the L1 insertion in the APC gene and other driver mutations in the PIK3A and KRAS genes that subsequently lead to adenocarcinoma. (Scott et al., 2016)

Cajuso et al. (2019) now apply the results of Scott et al. (2016) to another 200 CRC patients to better understand the role of retroelements in addition to the progression of CRC tumorigenesis. In doing so, they conclude that:

> *In approximately 1% of the cases we identify insertions in APC, likely to be tumor-initiating events. Insertions are positively associated with the CpG island methylator phenotype and the genomic fraction of allelic imbalance. Clinically, high number of insertions is independently associated with poor disease-specific survival. (Cajuso et al., 2019:1)*

Both Cajuso et al. (2019) and Scott et al. (2016) refer to the "Cancer Genome Landscapes" by Vogelstein et al. (2013) when interpreting the observed data. Vogelstein and colleagues describe different driver gene mutations, which stand at the beginning of certain pathways and are responsible for three central cellular processes. At the beginning, the tumor always requires an initial mutation, usually of a gatekeeping gene, in order to have a selection advantage with regard to cell growth and cell division compared to normal epithelial cells. For CRC, it is clear that the APC gene is mutated and initially leads to the formation of benign lesions, such as small adenoma. Further mutations like in KRAS stimulate tumorigenesis; this is called clonal expansion. Mutations in PIK3CA, SMAD4 and TP53 generate a malignant tumor, which now also forms metastases. Vogelstein therefore makes a clear distinction between "driver" and "passenger" mutations:

> *Although it is easy to define a "driver gene mutation" in physiologic terms (as one conferring a selective growth advantage), it is more difficult to identify which somatic mutations are drivers and which are passengers. Moreover, it is important to point out that there is a fundamental difference between a driver gene and a driver gene mutation. A driver gene is one that contains driver gene mutations. But driver genes may also contain passenger gene mutations. For example, APC is a large driver gene, but only those mutations that truncate the encoded protein within its N-terminal 1600 amino acids are driver gene mutations. Missense mutations throughout the gene, as well as protein-truncating mutations in the C-terminal 1200 amino acids, are passenger gene mutations.*

While the LoF of tumor suppressor genes could be explained by L1 insertion mutations using the APC gene as an example, retroelements play a different role

in oncogenes. Here, they primarily serve as promoter donors that lead to a hyperexpression of oncogenes - this is referred to as onco-exaptation.

8.6 Retrotransposons derived Onco-exaptation

As previously noted, epigenetic modes play a central role not only in embryogenesis but also in the totipotency of somatic cells. According to Vidal et al. (2017), epigenetic de-regulations lead to genetic malignancies and eventually to cancer.

Lynch-Sutherland et al. (2020) hypothesize that onco-exaptation, as previously described by Babaian et al. (2015), drives oncogene expression. In this process, regulatory motifs within retrotransoposons, particularly promoters, are "recruited, which consequently forced the expression of oncogenes. In Fig. X, one sees that in non-cancerous tissue, transposons are silenced by epigenetic repressors. However, if this repression is lost during onco-exaptation (for example, due to hypomethylation), transcription factors can now use the accessible cis-promoter genes to express oncogenes (Fig. 40).

proliferate more vigorously.

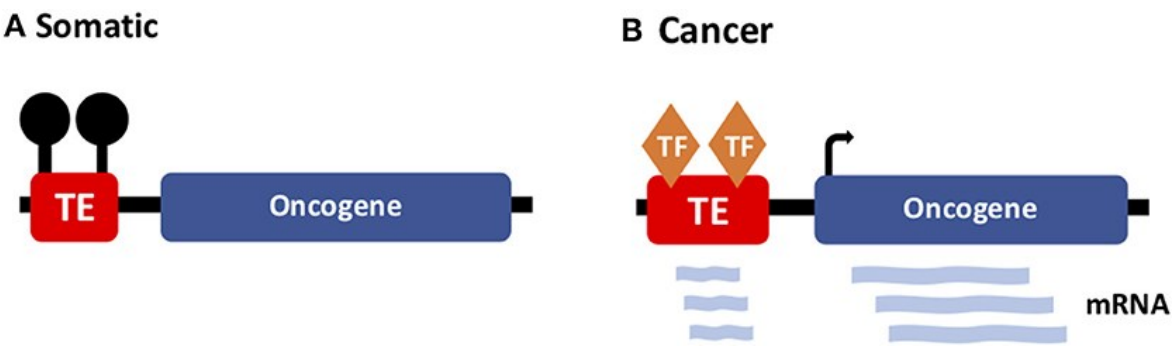


Fig 40: Transposon-driven onco-exaptation leads to overexpression of oncogenes (Lynch-Sutherland et al., 2020).

In the literature, two models were distinguished with respect to the relationship between transposons and oncogenes:

(1) The **de-repression model** (Lamprecht et al., 2010) proposes that activation of TEs are a consequence of molecular changes during (earlier) oncogenesis.

(2) The **epigenetic evolution model**, on the other hand, as previously described by Babaian and Mager (2016), assumes that at a given time in certain cells some TE loci are unetylated. They therefore have the potential to provide cis-regulatory elements for oncogenes. This is not per se triggered by epigenetic pathogenicity, nor does it automatically lead to neoplasia. If this cis-regulatory element does become active, then overexpression of the affected oncogene will occur, in some cases conferring a selective advantage to the cell. If tumor formation now occurs, then all clones of this cell have a selective advantage over other tumor cells and

8.7 LINE-1 elements as biomarkers

Numerous empirical studies have demonstrated the link between LINE-1 retrotransposition activity and tumorigenesis. These include (1) hypomethylation of the LINE-1 promoter, which was discovered as early as 1993 to be causative for colorectal carcinoma (Thayer et al., 1933). However, this was also proven for other tumor entities in the coming years, such as lung cancer (Saito et al., 2010), hepatocellular carcinoma (Honda and Rahman, 2019) and esophageal cancer (Iwagami et al., 2013, Kawano et al., 2014).

In addition to the methylation status of the retrotransposons, the activity of LINE-1 at the transcriptional and translational levels has also been demonstrated. For example, Harris et al. (2010) previously demonstrated that ORF1 of LINE-1 was expressed at high levels in the nucleus of breast cancer cells. Rodic et al. (2014) provide similar evidence for ovarian carcinoma and also show that hypoexpression of ORF1 of retroelements correlates with loss of TP53. According to Hosseinnejad et al. (2018), ELISA measurements of ORF1 in serum allow conclusions about prostate cancer.

Therefore, Lavasanifar et al. (2019) propose to use indicators of LINE-1 retrotransposon activity as biomarkers for cancer and to use them for diagnosis but also as therapeutic targets. Biomarkers, according to the U.S. Food and Drug Administration, are an umbrella term for all measurable characteristics of a tumor

(FDA, 2017). Their determination provides additional information on the disease situation, the likely course of the disease, or the effectiveness of treatments. Tumor markers are a subset of biomarkers. They are endogenous substances that indicate cancer. They are increasingly produced by the tumor cells themselves or by healthy body cells in response to a tumor. The "classic" tumor markers are usually proteins or protein-containing compounds. They can be detected in blood, urine or other body fluids.

The National Cancer Institute distinguishes between:

(a) diagnostic tumor markers,
(b) prognostic tumor markers,
(c) predictive tumor markers,
(d) pharmacokinetic and pharmacogenetic tumor markers,
(e) therapeutic tumor markers.

While solid tumors can only be diagnosed by examining the tissue via invasive sampling by means of biopsy, non-invasive methods are used to try to determine the response to certain therapy approaches or the respective therapy course via circulating tumor cells or cell-free circulating tumor DNA in the blood, primarily in a predictive manner (Sawyers, 2008).

In a review, Rodic (2018) collects over a thousand articles on PubMed that focus the role of LINE-1 retrotransposons on tumorigenesis - only 35 of which have an experimental approach that illuminates the complex relationship between retrotransposition and oncogenesis in an evidence-based manner:

> *Small portion, or 9 out of 35 articles, link LINE-1 expression in cancer to a potpourri collection of unique empirical observations, such as: micronuclei formation, LINE-1 as pre-diagnostic biomarker within circulating tumor cells/tumor DNA, gene expression profiles linked to LINE-1 expression, putative mechanism of LINE-1-mediated cellular transformation, as well as induction of apoptosis and proliferation following LINE-1 expression in cancer. To my mind, these nine studies represent preliminary discovery-based work that may prove to be of interest in the future. (Rodic, 2018: 1680)*

In the following sections, the putative role of Line-1 transposons at the genome, transcriptome and proteome level as tumor markers will be discussed in more detail. The focus is not only on the opportunities for improved diagnosis and new

therapeutic approaches, but also on the limitations with which research with transposons is confronted.

8.7.1 L1 retroelement as diagnostic tumor marker

Tumor markers or biomarkers are biological substances that can be highly elevated in body fluids due to malignant tumor disease and can thus be considered as diagnostic tools. Due to a lack of specificity and sensitivity, tumor markers cannot usually be used as search parameters but only as follow-up parameters. With a few exceptions, tumor markers are therefore rarely used in the initial diagnosis.

The diagnostic practicability of a biomarker depends on how specifically it can be operationalized for a specific tumor entity. Since mobile genetic elements, as part of non-coding DNA, have only been studied more intensively in recent years, the relationship between (retro-) transposition and tumorigenesis has not yet been definitively spelled out. In addition, transposons in general, but also transposon-capable retrotransposons in particular, are very *"polymorphic for insertion presence/absence in different human populations at particular chromosomal locations"* (Konkel et al., 2007).

However, one of the first cancer entities in which the role of retrotransposons could also be causatively demonstrated is CRC (Miki et al., 1992; Scott et al. 2016). Cajuso et al. (2019) studied over 200 colorectal tumors and found that in 1% of all cases L1-driven insertions into the APC gene are crucial for tumor initiation. As described by Scott et al. (2016), the majority of these insertions are due to de-repression of DNA methylation at the CpG islands of the promoter region of the still inactive L1 transposons.

Approximately 30,000-45,000 GpG islands are distributed over the entire human genome: up to 70% of them are located in regulatory sections, i.e., in the promoter region. There they are mostly methylated. Non-regulatory sections of the genome are therefore low in CpG, since the 5-methyl cytosine is spontaneously deaminated to thymine (Lander et al., 2001). DNA methylation not only affects the organization of coding regions (gene expression, transcription regulation, X-

chromosome inactivation, imprinting, gene repression) but is the key mechanism for transposon repression (Yoder et al., 1997).

Epigenetic modification of DNA is an important indicator of pathological processes, especially tumorigenesis. For example, defective methylation of promoters can lead to overexpression of oncogenes, or repression of tumor suppressor gene expression. Burns (2017) highlights that certain hypo- but also hypermethylation patterns of transposons are specific to certain cancers. Thus, L-retroelements would be a useful diagnostic and prognostic biomarker-regardless of whether the retrotransposition is the driver mutation that initiates tumorigenesis or a passenger mutation that goes hand-in-hand with increased genomic instability.

This can now also be applied in the diagnosis of CRC. To date, the gold standard for the diagnosis of CRC is still colonoscopy (Pox et al., 2012). However, since it is an invasive method -that is also associated with morbid complications- it also has its limitations. One possibility would be to examine occult blood in the stool - but this has less sensitivity or specificity compared to colonoscopy. Another possibility is liquid biopsy, i.e., the examination of cell-free DNA (cfDNA) in blood. In CRC, carcinoembryonic antigen (CEA) is usually examined - but this is also only of limited significance.

Therefore, Ponomaryova et al. (2020) propose to examine L1 retroelements for the degree of methylation of their CpG islands and thus provide an indication of the genome-wide degree of methylation. The advantage over studying individual malignant aberrations in genes is the quantity of L1 retroelements. Considering that more than half of our genome consists of mobile genetic elements, better statistical conclusions can be made about the whole genome. What remains unanswered is to what extent this is a causative driver mutation (e.g., point mutations in CpG islands leading to hypomethylation).

How could L1 methylation status now be studied in a resource-efficient manner? The following techniques are proposed by Ponomaryova et al. (2020) (see graph X):

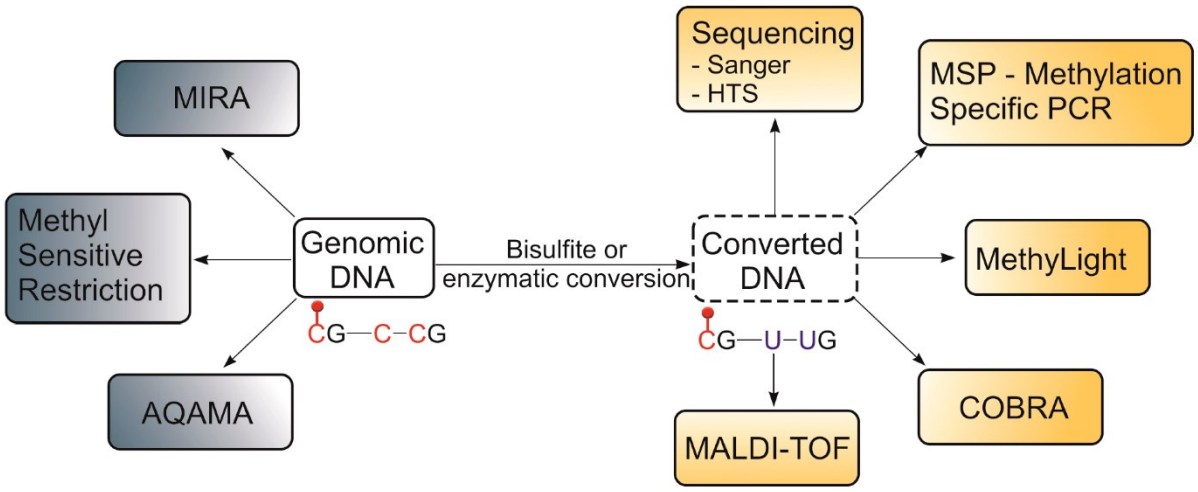


Fig. 41: Techniques to analyze the methylation status of CpG islands of L1 retroelements. On the right (in yellow) methods involving prior conversion of unmethylated C to U - on the left (gray) methods that can be applied directly to genomic DNA. (Ponomaryova et al., 2020)

(1) on genomic DNA, for example, MIRA (methylated-CpG island recovery assay (Rauch et al., 2006) or AQAMA (absolute quantitative assessment of methylated alleles) (de Maat et al., 2007).

(2) after chemical or enzymatic conversion, in which unmethylated cytosines are converted to uracils, and only 5-methyl-cytosines remain, methylation pattern can be performed by either methylation-specific PCR (MSP) or combined bisulfite restriction analysis (COBRA) (Herman et al., 1996; (Xiong & Laird, 1997).

Although the methylation level of CpG islands of transposons is a promising marker for liquid biopsy simply because they are present in higher concentration in cfDNA, Ponomaryova et al. (2020) mention three limitations devaluing L1 retrotransposons as diagnostic markers and concludes that *"Therefore, L1 methylation in cfDNA should be used as an additional marker indicating the increased probability of a malignant disease."*

(1) L1 methylation patterns in cfDNA are poorly studied and there are data comparing methylation in cfDNA and in tumor tissue DNA for only four tumor diseases.

(2) even if hyper- or hypomethylations of CpG islands of retroelements are detected in cfDNA, the tissue of origin remains enigmatic.

(3) Inter-patient heterogeneity and polymorphism of retroelements makes it difficult to assign LINE-1 methylation profiles to a specific type or stage of tumor.

8.7.2 L1 retroelement as a prognostic tumor marker

Lee et al. (2019) suggest that LINE-1 elements should not be used as a diagnostic marker in liquid biopsy, but that the methylation mute should be used as a prognostic tool after tumor diagnosis. Again, the basis is the difference in methylation of 5'-UTR in retroelements in normal and cancerous tissues. Here, they are studying in particular samples from patients who have already been diagnosed and staged for breast cancer. The following studies will be performed:

1. Isolation of cfDNA from approximately 4ml EDTA blood and subsequent digestion by restriction enzymes.
2. Bisulfite conversion PCR (BSC-PCR) followed by sequencing to determine the methylation status of CpG islands.
3. Methylation-sensitive restriction enzyme digestion (MSRED) followed by rtPCR as a control for BSC-PCR.

The results (Fig. 42) clearly show that first, the concentration of cfDNA was significantly higher in breast cancer patients than in unaffected ones (represented by the cell line MCF10a, these are healthy cells). Second, the level of methylation was significantly lower in affected individuals. However, CpG islands were not equally methylated in all affected individuals. In Fig. X, it can be seen that the cancer cell line MCF7 has significantly lower methylation levels than MDA-MB-231.

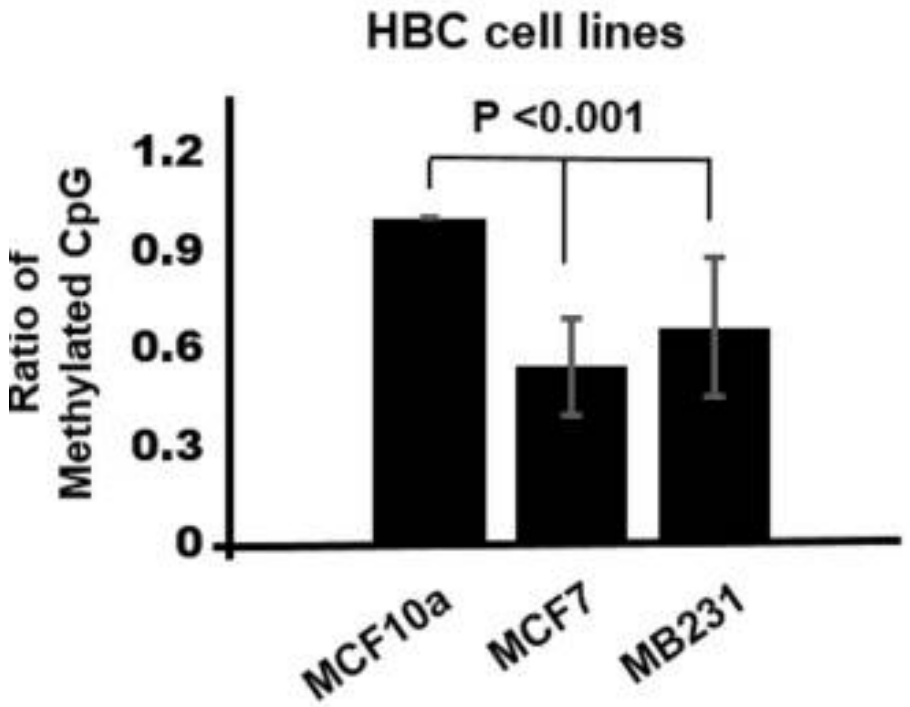


Fig. 42: Although the methylation level of cfDNA is significantly lower in breast cancer cell lines - a clear difference is seen between hormone receptor positive and triple negative cell lines. (Lee et al. 2019)

Both target lines studied were related to luminal and mesenchymal-like breast cancers, respectively, but MCF-7 cells are primary tumors, and are characterized by overexpression of hormone receptors in the simultaneous absence of HER2. They are mainly used as a model for hormone therapy. MDA-MB-231 cell lines, on the other hand, are triple negative and are typically metastatic. They are a model for chemotherapy.

Thus, the methylation level of retroelements could be used to distinguish between malignant and benign breast cancer entities after a successful diagnosis. The novel aspect of the study by Lee et al. (2019) was the comparative approach: they compared breast cancer in humans and in the model organism dog. A few years earlier, however, Park et al. (2014) had proposed using retroelements in prognosis. Prognostic biomarkers provide clues to the expected individual course of the disease.

Park et al. (2014) investigated the methylation pattern of retroelements directly in the cancerous breast tissue of about 180 invasive breast cancer samples.

Previously, these were subdivided into different subtypes by immunohistochemistry. For this purpose, primarily hormone receptors (ER, PR), HER2, Ki-67, cytokeratins and growth factor receptors were labeled. Subsequently, the methylation level of retrotransposons in the tissue was analyzed (Fig. 43).

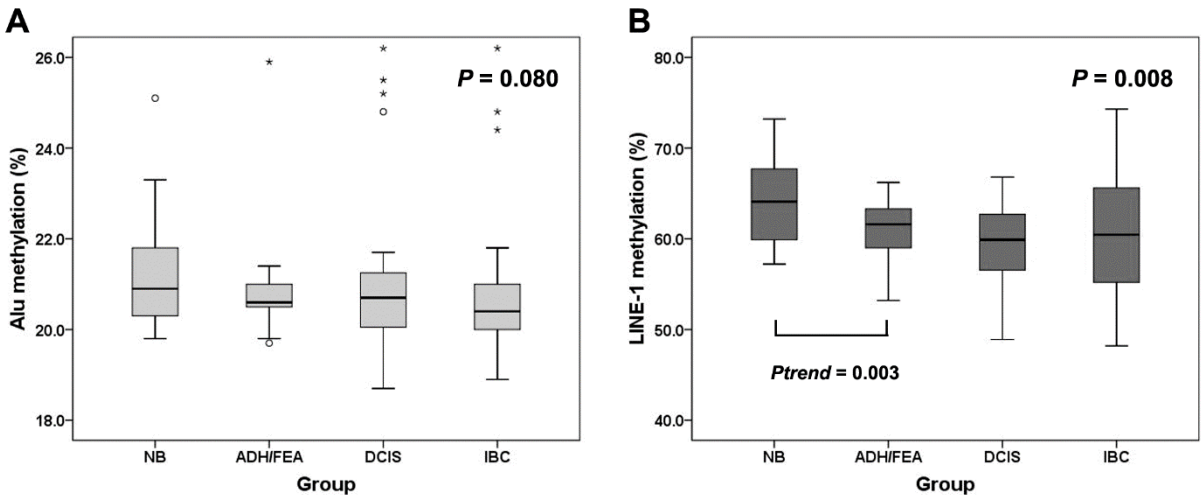


Fig. 43: Metyhlation level of Alu elements and LINE-1 in normal breast tissue (NB), atypical ductal hyperplasia (ADH)/flat epithelial atypia (FEA), ductal carcinoma in situ (DCIS) and invasive breast cancer (IBC). While Alu elements are relatively uniformly methylated, there are significant differences (p=0.008) between the four groups of breast lesions: Metyhlation level decreases significantly in progression. (Park et al., 2014)

Thus, Park et al. (2014) clearly demonstrated that hypomethylation is indeed related to cancer progression. For breast cancer, the lower the methylation index, the worse the prognosis, especially for younger breast cancer patients (van Hoesel et al., 2012).

However, this is not true for all tumor diseases. While Park et al (2014) showed a continuous decrease in methylation status from normal to invasive tumor tissue, this is not the case for CRC, for example. According to Bae et al (2012), although the methylation index decreases significantly during the transition from normal tissue to adenoma, it is not evident during progression from adenoma to carcinoma. On the one hand, this could indicate that hypomethylation is an early event in CRC tumorigenesis, and on the other hand, that it is initial and thus causative for it.

For breast cancer, the low methylation seems to indicate a subtype: hormone receptor negative but HER2 positive. This indicates a higher number of chromosomal alterations and genomic instability (Ellsworth et al., 2008).

The correlation between methylation status of retroelements and tumor progression is not only crucial for the prediction of disease progression and the determination of tumor stage but also for the choice of therapeutic agent. Thus, retroelements offer themselves not only as prognostic markers but also as predictive markers for evidence-based therapy decision and monitoring.

8.7.3 L1 retroelement as a predictive tumor marker

Predictive biomarkers identify patients who are more likely to experience a more favorable or unfavorable effect from exposure to a therapeutic agent than similar individuals without that biomarker. For example, predictive biomarkers include the expression of RAS family genes in the tumor tissue of colorectal cancer patients. Before therapy with an EGFR antibody, for example, the KRAS gene is examined. Only if this gene shows no alteration (which is the case in about 60 percent of patients) can therapy with an antibody directed against the EGF receptor be successful. Hypomethylation of the promoters of retroelements can lead either to knock-out of tumor suppressor genes or to overexpression of oncogenes. Central to this process is the transcription of the retroelement into an mRNA followed by reverse transcription by the enzyme specifically encoded for this purpose.

Zhang et al (2019) identify the derivative FGGY gene as a potential predictive tumor marker in lung squamous cell carcinoma (LUSC). FGGY is a hydrocarbon kinase and also an important tumor suppressor. This gene is most commonly affected by retrotransposition through LINE-1 elements in LUSC: In this process, insertional mutagenesis disrupts the start of exon 13 of the gene. This results in expression of the L1-FGGY, which subsequently also strongly inhibited the expression of FGGY genes of the other allele (Fig. 44).

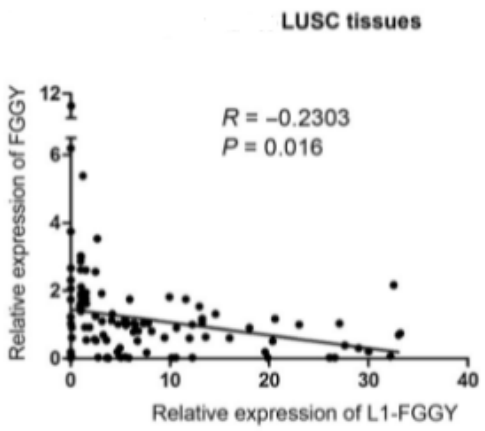


Fig. 44: Spearman rank correlation between the expression of L1-FGGY and FGGY: the higher the concentration of derivative gene, the more the expression of the normal gene was inhibited. (Zhang et al. 2019)

The consequence of the overexpression of derivative L1-FGGY is that many other pathways are deregulated, in particular those that regulate the lipid metabolism of the cell. In addition, local immune evasion occurs, i.e., in LUSC tumor tissue that has particularly high levels of derivative L1-FGGY, CD4+ T cells are significantly downregulated. Zhang et al (2019) attribute this to the fact that with the dysregulation of the FGGY gene, cytokines (particularly IFNL4, TNFRSF11A, TNFSF12, IL17RD, IL34, and IL27RA) are also downregulated, but these are important for promoting de T cells. In addition, however, those cytokines that regulate immunosuppression in normal cells (IFNγ, IL17, and IL27) are upregulated in L1-FGGY-positive tumor tissue, enhancing immune evasion.

To illuminate the precise relationship between retrotransposition and oncogenic immune evasion, Zhang et al. (2019) develop an in vitro model: using a recombinant lentivirus as a vector, the derivative L1-FGGY was introduced into an LUSC cell line and overexpressed, and at the same time shRNA was used to suppress the expression of FGGY. Thus, they confirmed that retrotransposition of a LINE-1 into a FGGY significantly stimulated cell proliferation and repressed cell apoptosis. In addition, cell invasion and EMT were promoted, facilitating carcinogenesis. The next step is now to inhibit transposition by reverse transcriptase inhibitors. More on this in the next subsection.

8.7.4 L1 retroelements as therapeutic tumor markers

To conclude this chapter on the potential of retroelements as tumor markers, the previously presented research by Zhang et al. (2019) on lung cancer now presents the use of reverse transcriptase inhibitors. The mentioned L1-FGGY-positive LUSC tumor cell line produced using recombinant lentivirus and shRNA was now subcutaneously implanted into mice and treated with either nevirapine (NVR) or efavirenz (EFV) for 22 days. NVR and EVF are both reverse transcriptase inhibitors that block the enzymatic activity of endogenous RT (Patnala et al., 2013). On the one hand, the experiments demonstrated that the RT inhibitors can effectively inhibit tumor growth after 22 days (Fig. 45).

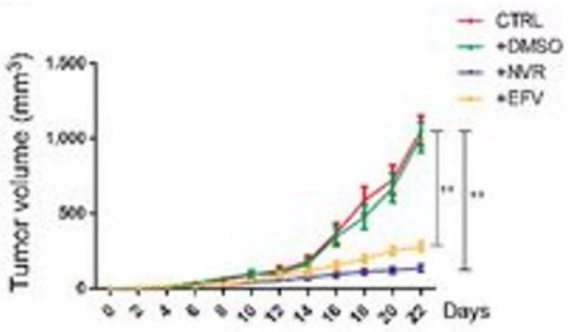


Fig. 45: RT inhibitors effectively inhibit tumor growth (CRTL: control group, DMSO-treated). (Zhang et al. 1019)

Another important therapeutic benefit is the inhibition of EMT and thus the likelihood of metastasis (Fig 46).

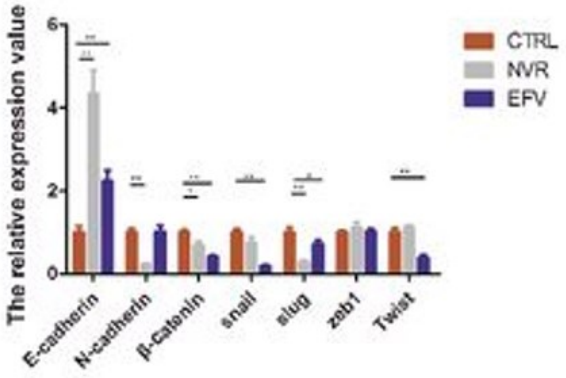


Fig. 46: qPCR analysis of EMT gene markers upon treatment with RTI: epithelial cell marker E-cadherin increased, whereas mesenchymal cell markers e.g., N-cadherin decreased. (Zhang et al. 2019)

8.8 Reverse transcriptase as therapeutic target in oncology

Regardless of whether transposons are drivers or passengers of tumor genes through insertional mutagenesis in tumor suppressors or promoter donors for oncogenes, they remain a visible hallmark of cancer. In recent citations, LINE-1 have been [3]addressed primarily as mobile genetic elements in theragnostic. However, in recent years, reverse transcriptase (RT), the gene product of ORF2, is gaining particular importance. Sciamanna et al. (2016) refer to L1-RT as *an "underestimated player in cancer"*.

Research by Mangiacasale et al. (2003), Gualtieri et al. (2013) Patnala et al. (2013) demonstrate high RT activity in tumor cells, which also correlates with tumor progression. Inhibitors, as already presented in the last chapter, can already reduce tumor cell proliferation, and antagonize de-differentiation through pharmacological inhibition of RT (Miousse & Koturbash, 2015).

What is special about RT as a therapeutic target in the "war on cancer", as Hanahan (2014) describes the role of RT, is that it is no longer a genetic target but a whole-genomic one and thus can be used independently of cancer entity with the great advantage of being an early diagnostic cancer marker.

However, this would require a better understanding of the molecular basis for RT-driven tumorigenesis. In particular, the interdependence between retrotransposition and miRNA needs to be elucidated Sciamanna et al. (2013).

[3] According to Wiesing (2019), theragnostic describes the interdependence of diagnostics and therapy. The goal of theragnostic is to provide the right therapy for the right patient at the right time. The main elements of diagnosis in theragnostic include determining genetic predisposition, characterizing the stage of the disease, and monitoring the progress of healing. Theragnostic is used, among other things, to determine the suitability and efficacy of a drug for a particular disease in an individual patient.

9. Transposons as tools for research

Transpons-derived enzymes are already an integral part of human genetics research. One would be tagmentation and reverse trancriptase for the preparation of cDNA libraries.

Tagmentation (Fig. 47) is a library preparation procedure in which the amplicon is fragmented and labeled in parallel (Lu et al., 2015). This requires very little starting material and still gets relatively uniform coverage of the genome (Picelli et al., 2014). In this method, genomic DNA is fragmented by a Tn5 transposase, and sequencing adaptors are attached in the same step (Hennig et al., 2018).

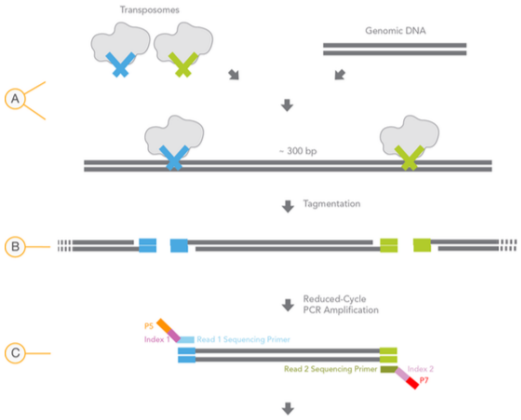


Fig. 47: lirbrary prepartion via tn5 tagmentation (Nextera).

Retrotransposon-derived reverse transcriptase (Fig. 48), on the other hand, is used to convert RNA to cDNA, such as in reverse transcription PCR (Bachman, 2013):

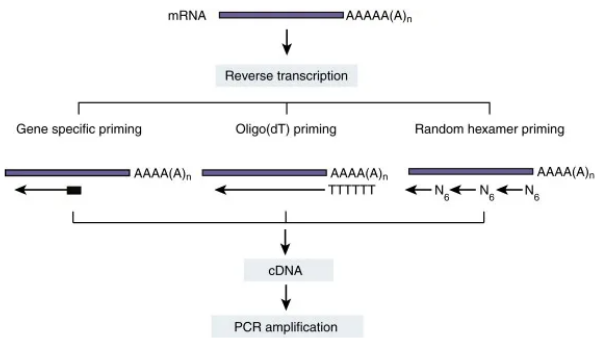


Fig. 48: reverse transcription PCR

9.1 Genome modification using Sleeping Beauty Transposon

The following considerations go back to the publications of Izsvák and Ivics (2004, 2015), who have dealt in detail with the use of SB transposons in gene therapy, in their 2004 article the authors describe strengths but also limitations of transposon technology, in their article a decade later they go a step further and try to predict the consensus sequence using careful sequence analysis to reconstruct an archetypal but active transposon within vertebrates. As a genetic engineering tool, transposons have long been limited to invertebrate research. The SB transposon revives and reactivates a transposon fossil in the fish genome. It is a DNA transposon that transposes using the cut-and-paste method. Transposition into the target genome forms the basis for long-term or possibly permanent expression of transgenes in transgenic cells and organisms. This is a prerequisite for a gene therapy approach, which aims to ensure that the tools used are 1. safe, 2. easy to use, 3. cost-effective, and 4. industrially producible. These criteria are particularly true for operationalized vector systems based on viruses[4]. Non-viral technologies [5]transfer nucleotides either unpackaged or using liposomes, polypeptides, or synthetic polymers.

In this juxtaposition (viral vs. non-viral), transposons could be seen as a link: they are dependent on the resources of the host cell, but unlike viruses, they are firstly non-infectious and secondly only active in one cell. The basic idea of the SB transposon is to recognize that transposons are precisely not *"fossilized molecular parasites"* in the genome but exist in a peculiar symbiosis with the genome and have been a driving force of evolution in the past. For example, Lander et al. (2001) highlight that up to 50 genes have emerged from transposons, the most significant being those genes that play an important role in the somatic

[4] Viral gene transfer vehicles used include fall adenovirus, retroviruses, herpesvirus, and adeno-associated virus (AAV). Viral vectors are efficient are used in 70% of studies. and approximately 70% of clinical trials are conducted with viral vectors (Schigetaka et al., 2002). Two key risks here are the risk of insertional mutagenesis of endogenous genes upon repeated integration into the human genome, or an undesirable immune response.

[5] Non-viral gene transfers are safer but usually more inefficient (Herweijer & Wolff, 2003) and therefore have little application in clinical practice.

recombination of the V(D)J mechanism in the genetic diversity of immunoglobulins. Likewise, the authors emphasize that -even though transposons constitute 45% of the human genome- there are very few transposable elements that still possess the competence to transpose. The latter belong to the retrotransposons (class 1), while DNA transposons (class 2), to which the SB transposon system also belongs, have long since ceased to be active in humans.

This is, on the one hand, the result of inactivating random mutations in transposon sequences, on the other hand, the result of efficient epigenetic repression. Nevertheless, to operationalize DNA transposons for research, DNA transposons from invertebrates have been used. These include transposons TC3 in nematodes (Babity et al., 1990) and mariner in Drosophila (Jacobson et al., 1986).

In 1997, the researchers Ivic and Izsvak succeeded in reviving a DNA transposon that was active in the genome of salmon fish about 10 million years ago. Through systematic mutation of the transposase protein and specific targeted modification of the transposon DNA sequences, Zoltán Ivics and his colleagues were able to reawaken the DNA transposon fossil from its proverbial slumber. These are again DNA transposons of the TC1/ Mariner type, but once active, they show significantly higher activity in the target genome of vertebrates than the previously tested transposons from invertebrates (Fischer, Wienholds & Plasterk, 2001).

9.2 Structure of the Sleeping Beauty Transposon (SB)

The SB transposon has a simple structure (Fig. 49). In its original form, it consists of a single gene encoding the transposase polypeptide, the enzymatic factor of transposition. This gene is flanked by terminal inverted repeats (IRs), which are also binding sites for the transposase. Researchers can now detach the transposase gene from the IRs and replace it with other DNA sequences. This is because the transposase can mobilize transposons by trans-acting[6] as long as they retain the IRs. In other words, the transposase gene does not have to be on the same DNA molecule as the transposon; it can be included either as an untranscribed gene, or in the form of transcribed mRNA or as a translated protein.

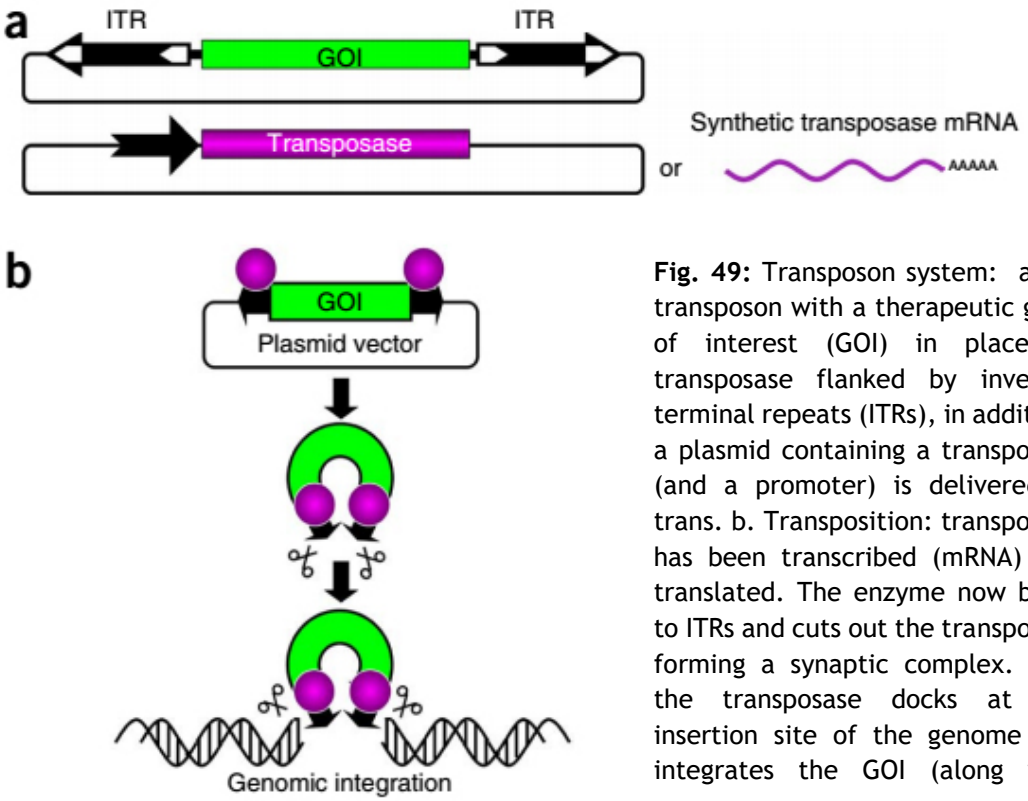


Fig. 49: Transposon system: a. SB transposon with a therapeutic gene of interest (GOI) in place of transposase flanked by inverted terminal repeats (ITRs), in addition, a plasmid containing a transposase (and a promoter) is delivered in trans. b. Transposition: transposase has been transcribed (mRNA) and translated. The enzyme now binds to ITRs and cuts out the transposon, forming a synaptic complex. Now the transposase docks at the insertion site of the genome and integrates the GOI (along with

[6] In the context of transcriptional regulation, a transacting factor is a regulatory protein that binds to DNA and alters expression. The transactive gene may be located on a different chromosome than the target gene, but activity occurs via the intermediate protein or RNA it encodes. Cis-acting elements, on the other hand, do not code for protein or RNA; they are often located near the gene to be regulated. Both the trans-acting gene and the protein / RNA it encodes should act on the target gene "in trans".

Like all other Tc1 / Mariner- type transposase molecules, the SB transposase inserts a transposon into a TA dinucleotide base pair in a target sequence. In vertebrates, including humans, there are approximately 200 million TA sites. The TA insertion site is duplicated during the transposon integration process. This duplication of the TA sequence is a hallmark of transposition and is used to establish the mechanism in some experiments. However, a recent study showed that SB is also integrated into non-TA dinucleotides with low frequency (Guo, Zhang & Hu, 2018).

9.3 Reconstruction of the active transposase

In the artificial construction of an active transposase, inactive transposon sequences of two DNA transposons from salmon (*Salmo salar*) and rainbow trout (Oncorhynchus mykiss) were combined. Subsequently, functional domains in the transposase gene were corrected based on the consensus sequence identified in other fish species apart from salmonids. The stepwise modification of the salmonid transposase is shown in Fig. 50:

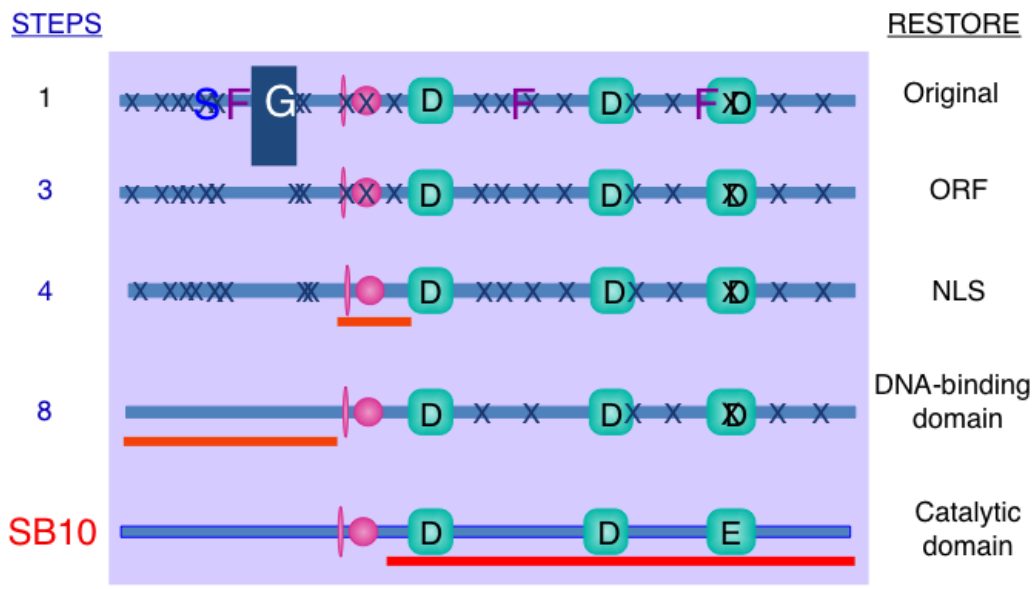


Fig. 50: Reconstruction of the SB transposase: Step 1: Scheme of fossil Tc1 / marine-like transposons in modern salmonid genomes; X: missense mutations; S: stop codons; F: frameshift mutations; G: major gaps/ missing amino acids. Step 3: Elimination of gap (G) and termination and frameshift mutations. Step 4: Reconstruction of the two-membered NLS sequence (orange underlining). Steps 5-8: Reconstruction of the N-terminal DNA binding domain (orange underline). Steps 9-10: Reconstruction of the catalytic domain (orange underline) including the DDE signature residues (green boxes).

Steps 1-3 aimed to restore a complete protein by filling gaps in the sequences using the reference sequence and removing and reversing stop codons that prematurely aborted the synthesis of the peptide (from 360 amino acids). In step 4, mutations in the nuclear localization signal (NLS[7]) were now reversed so that the transposase, which is synthesized in the cytoplasm, can now be imported into the nucleus. In steps 5-8, the N-terminus of the transposase, which contains the binding domain for the ITRs motif, is now reorganized. in the last two steps, those catalytic domains that catalyze the insertion in the target gene are now corrected. In doing so, the researchers are guided by conserved amino acids in integrase and recombinases: In sequence, the conserved amino acids aspartic acid (D) and glutamic acid (E) are reconstituted in steps 9 and 10. The result is now a transposase (SB10) that contains all the domains and motifs necessary for transposition. The final result was SB10, which contains all the motifs required for function (Ivics et al., 1997). The now intact transposase of 360 amino acids has the following structure (Fig. 51):

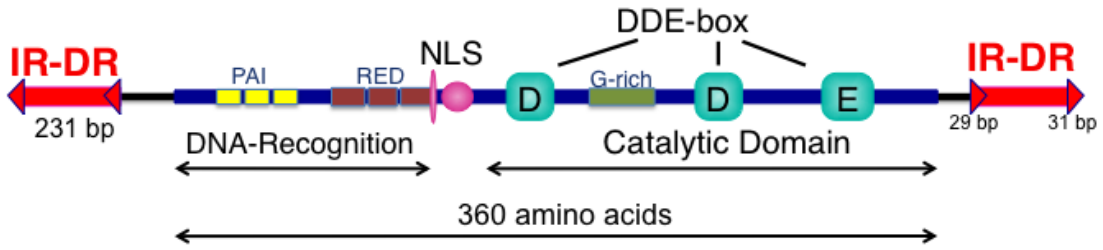


Fig. 51: Transposase has three main subdomains: the N-terminal recognition motif for binding to the ITRs sequences, a nuclear localization sequence (NLS), and a DDE domain that catalyzes the cut-and-paste process. The DNA recognition domain features two paired box sequences that can bind to DNA and are related to different motifs found on some transcription factors; the two paired boxes are labeled PAI and RED. The catalytic domain has the typical DDE amino acids (sometimes DDD) found in many transposase and recombinase enzymes. In addition, there is a region that is highly enriched in glycine (G) -amino acids.

[7] A nuclear localization sequence (NLS) is an amino acid sequence that marks a protein for import into the nucleus. Typically, this signal consists of one or more short sequences of positively charged lysines or arginines exposed on the protein surface. Different nuclear-localized proteins may share the same NLS. An NLS has the opposite function of a nuclear export signal (NES), which targets proteins outside the nucleus.

9.4 Mechanism of SB transposition

The SB transposition is a cut-and-paste process in which the transposable element is cut out of its original position by the transposase and integrated in a new position. The transposition process can be arbitrarily divided into at least four main steps, which will be briefly outlined below:

(1) Binding of transposase to ITRs motifs.
(2) Formation of a synaptic complex (SCF).
(3) Excision from donor molecule
(4) Insertion/reintegration at a target site

(1) Since potentially mobilizable transposons are found in every genome, the specificity of transposition is an important issue. The human genome also inactivated transposons, but these are only distantly related to SB and they are unlikely to be mobilized by the SB transposase. This is because the recognition and transposition of transposons (by SB transposase) appears to be highly specific. According to Izsvák et al, (2002), the specificity of transposition is regulated at several levels: On the one hand, the transposase recognizes the associated substrate very well and can distinguish it from similar sequences that differ in only three base pairs, and on the other hand, the specificity of the transposase is controlled within the process of synaptic complex formation (SCF).

(2) SCF is a process in which the two ends of the transposon are excised and held together. Successful excision and formation of a synaptic complex requires that the ITRs sequences in the transposase and transposon match. Specifically, these are the ITR motif, a tetramer of the transposase protein, and a cellular host factor, HMGB1[8]. The SCF process appears to be a regulatory "check point" of

[8] HMGB1 is a non-histone protein associated with eukaryotic chromatin. Together with the closely related HMGB2 protein, HMGB1 regulates a number of cellular processes, including gene regulation, DNA replication, and recombination, including immunoglobulin V(D)J recombination. In addition, HMGA1 proteins are required for retroviral cDNA integration. SB transposase was found to interact with HMGB1 in vivo, suggesting that transposase can actively recruit HMGB1 to transposon DNA. HMGB1 enhances the preferential binding of the SB transposase to the inner transposase binding sites within the transposon

transposition, where certain molecular requirements must be met for it to proceed. The regulated formation of synaptic SB complexes serves as a quality control of transposition, filtering out aberrant or nonspecific excision products.

(3) Excision results in the formation of sticky ends: the cuts at the ends of the SB transposon are offset inward by 3 nucleotides, creating 3′ overhangs and leaving the terminal nucleotides of the transposon at the donor site. Repair of the offset double-strand break now creates a three-base pair-long transposon "footprint"(Luo et al., 1998). Both excision and insertion are catalyzed by the same catalytic domain of the transposase. Interestingly, however, flanking sequences at the excision and insertion sites have different meanings for transposition. While excision frequencies do not appear to significantly affect transposition, certain sequences are more desirable for transposition integration than others. This also explains the target selection of SB transposons.

(4) Most transposons do not randomly integrate into the target DNA; in fact, they exhibit some specificity in recognizing the target sequence. There is a wide spectrum in target site selection, ranging from highly specific to essentially random. In some elements, accessory proteins play a role in localizing a potential target. In other systems, likely including SB, target selection is primarily determined by the transposase itself. SB addresses into a TA dinucleotide that is duplicated upon transposition and flanks the integrated element (Vigdal et al., 2002). Consider again that there are approximately 200 million TA sites in the vertebrate genome, as in the human genome. The TA insertion site is duplicated during the transposon integration process. Recent studies such as this one by Guo et al. (2018), suggest that the SB transposase can also address other insertion sites away from the TA dinucleotides. This duplication of the TA sequence is a hallmark of transposition and is used to establish the mechanism in some experiments. In general, the SB transposon is not autonomous, it does not encode its own

---

ITRs and promotes the bending of DNA fragments comprising the transposon inverted repeats]. HMGB1 is likely required for the efficient formation and/or stability of synaptic complexes during SB transposition, thereby directing the subsequent catalytic steps of transposition. (Zayed, 2003)

transposase, the latter being supplied by another source. This is useful for genetic engineering applications in that one can regulate the process using the transposase.

Vigdal et al. (2002) found that insertion of the SB transposon occurs randomly at the genomic level. Approximately 35% of insertions occurred in transcribed regions. However, most insertions that occur in genes are in introns. This is due to base composition and accordingly length, which makes introns more attractive targets for the transposon than exons or promoters. Surprisingly, only eight insertion motifs could be recovered from repetitive sequences. Considering that 45% of the human genome consists of repetitive elements, this number seems exceptionally low. Possible explanations for this phenomenon are base composition and chromatin structure. Gu et al. (2019) comment on this as follows:

> *The standard model of SB transposase-mediated transposition includes symmetrical cleavages at both ends of the transposon for excision and re-integration into another DNA sequence. In our analysis of excised transposon fragments (ETFs), we found evidence for the requirement of certain flanking sequences for efficient cleavage and a significant rate of asymmetrical cleavage during the excision process that generates multiple ETFs. Our results suggest that the cleavage step by SB transposase is not as precise as indicated in most models. Repair of the donor ends can produce eight footprint sequences (TACTGTA, TACAGTA, TACATA, TACGTA, TATGTA, TACTA, TAGTA and TATA). [...] These findings may be important in considerations of efficiency of SB transposon remobilization, selection of TA integration sites and detection of SB excision and integration loci, all of which may be important in human gene therapy. (Gu et al., 2019).*

Important in this context is the comparison with viral vectors, namely these tend to integrate more strongly in coding regions, furthermore AAV integration sites are often associated with large deletions as well as chromosomal rearrangements. Therefore, it seems that SB transplantation may be a safer alternative for stable genomic integration than viruses.

9.5 Transfer and transgenesis of SB tools

Gene transfer is a necessary step in molecular therapy to introduce a new gene or silencing a (possibly pathogenic) gene. There are many different methods of gene transfer for different types of cells and tissues (Nayerossadat et al., 2012).

For gene therapy, transfection by vectors of viral or non-viral origin plays a more important role than physical or chemical methods. The latter methods are usually accompanied by damage to cellular structures. Therefore, vectors are used for artificial transfection in eukaryotes. Vectors include a wide variety of transporter vehicles for foreign DNA into the eukaryote cell, including phages and phagemids, cosmids, and even germ cells. For the topic discussed here, however, plasmids and viral vectors are of particular importance, as SB transposition tools represent a combination of these two tools, thus optimally exploiting their advantages and minimizing their disadvantages. Viral vectors are introduced into the host cell by transduction, and the desired sequence to be integrated (or expressed) is incorporated into the viral genome.

In figure 52, the transfer of the SB transposon system into the intended cell is described in detail: The components (plasmid with genes of interest flanked by ITR, plasmid with blueprint for transposase or mRNA[9] of the enzyme) can be delivered either via: 1. transfection, 2. nucleofection using electrotransfer directly into the nucleus, 3. liposome packaging (or 4. through viral hybride vectors). After entering the cell, transcription/translation of the transposase gene occurs in the cytoplasm and an enzyme is formed. This binds now to the flanking ITRs, excises, and integrates the gene of interest into chromosomal DNA of the genome.

[9] The use of transposase mRNA is preferable to a dsDNA version in the format of a plasmid for two reasons (Bire et al., 2013): 1. On the one hand, it prevents the DNA transposase gene from integrating into the human genome, and on the other hand, it reduces toxicity during elctrotransfer.

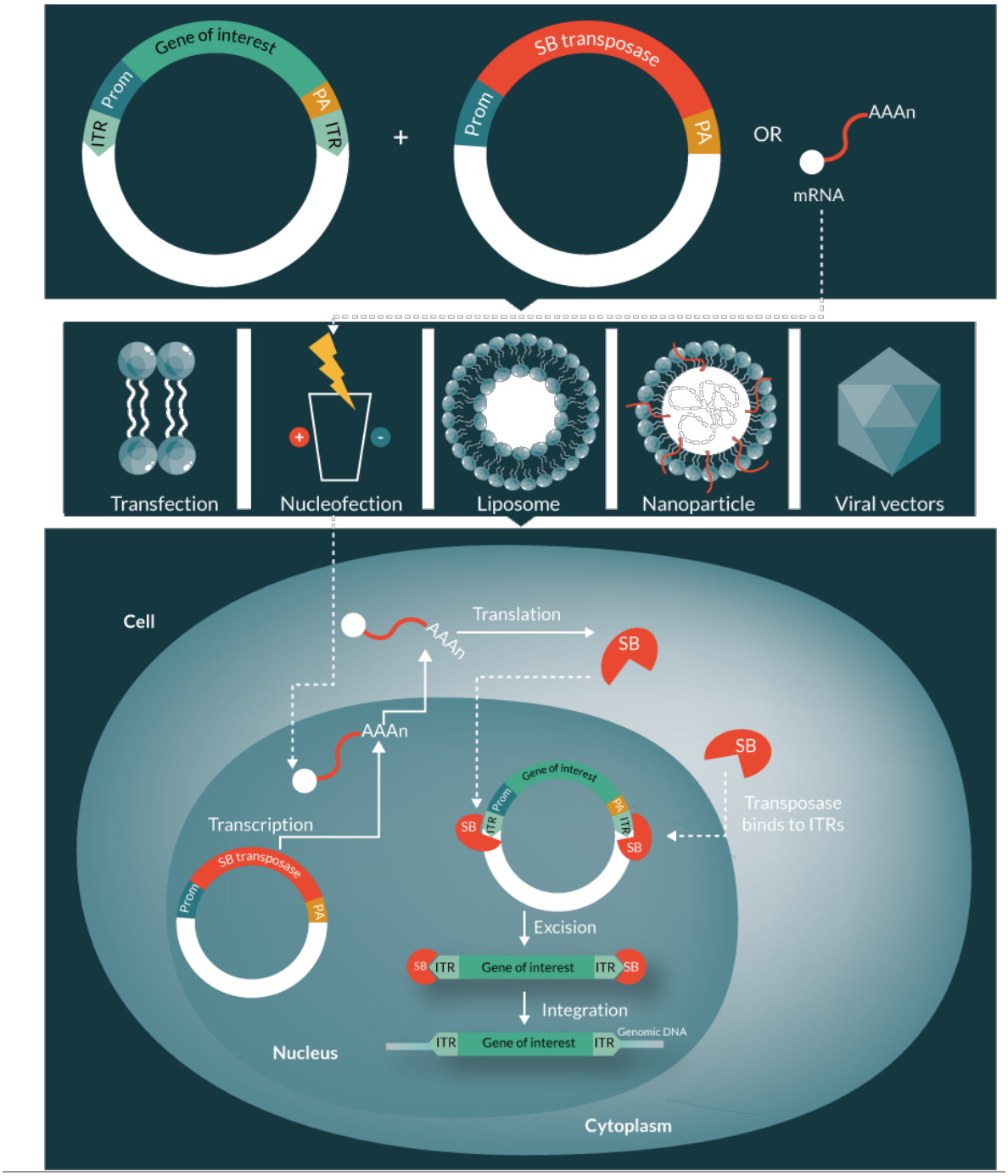


Fig. 52: Insertion of the components of the SB stems into target cells (Sandoval-Villegas et al., 2021)

9.6 SB Transposons in Genetic Engineering

SB technology is becoming increasingly important in human therapeutic research but also in the field of gene transfer and gene discovery. This is due to the fact that it is an excellent non-viral transfer tool of which the transposition mechanism has already been well studied and characterized (Narayanavari, 2017). The following ten reasons are equally crucial for the use of SB technologies in genetic engineering:

1. **Resource-efficient and economical** (Hudecek et al., 2017): compared to viral vectors, the production and use of plasmids is much cheaper and less labor-intensive.

2. **Stable transgene expression** (Wilber et al., 2007): in the long term, the gene products introduced by the SB vector will be produced in different cell types, even if SB vectors are not entirely resistant to silencing.

3. **Broad spectrum of target cells** (Izsvák, Ivics & Plasterk, 2000): a variety of cell types, including those primary cells that are critical for therapy, can be addressed using SB technologies.

4. **No restriction to cells capable of division** (Huang et al., 2006): primary cells that are no longer capable of division can also be targeted.

5. **Structure of the transgene remains intact** (Voigt et al., 2016): even complex transgenes can be faithfully incorporated and expressed using the SB vector.

6. **Transfer of large sequences** (Rostovskaya et al., 2012): the cloning capacity of SB vectors can exceed 10 kb, with combinations with bacterial artificial chromosomes (BAC) allowing even transgenes up to 0.1 Mb.

7. **Benign promoter activity** (Walisko et al., 2008): compared to Moloney murine leukemia virus (MoMLV), the enhancer or promoter activity of the ITRs of SB transposons is a hundredfold lower, making unintentional hyperexpression or exaptation of genes less likely.

8. **Low immunogenicity** (Tipanee et al., 2017): due to the separation of transposase and transposon vector, the enzyme is only transiently present in the target cell and does not trigger an immune response, which can be observed in the case of viral vectors, for example.

9. **No cross-mobilization in the human genome** (Ivics, 2016): unlike piggyBac, the SB transposon is not mobilized by any of the human genes or proteins.

10. **Random transposition in the genome** (Vigdal et al., 2002): incorporation of with SB transposons occurs almost completely randomly, as the transposase integrates into TA dinucleotides, resulting in more frequent transposition of recombinante sequences into non coding DNA segments than is the case with viral vectors, which preferentially integrate into coding sequences.

Precisely the last point mentioned, i.e., the random transposition of the SB transposon at TA dinucleotides[10] in the human genome, distinguishes SB transposons from CRISPR/Cas tools (Sandoval-Villegas et al., 2021). After all, there are over 200 million TA dinucleotides in the mammalian genome (Hackett et al., 2013) and thus a vast number of possible insertions (Jung et al., 2016). This makes the interactions between transposon and target genome difficult to predict, which is better achieved with targeted nucleases such as the CRISPR/Cas9, as a specific DNA target sequence is specifically recognized via a guide RNA and subsequently cleaved via the nuclease (Doudna & Charpentier, 2014).

In Crispr/Cas systems, repair occurs after double strand break (DSB) - either via non-homologous end joining (NHEJ) or homologous directed recombination (HDR). However, Kovač et al.(2020) emphasize that knock-out of genes (for example, via frameshifts in the NHEJ pathway) is much more efficient than knock-in of new sequences, which is achieved via HR (Fig. 53).

[10] The insertion into the target sequence TA is not as stringent as previously thought. Guo, Zhang and Hu (2017) further identified 28 thousand non-TA SB insertion sites using next generation sequencing.

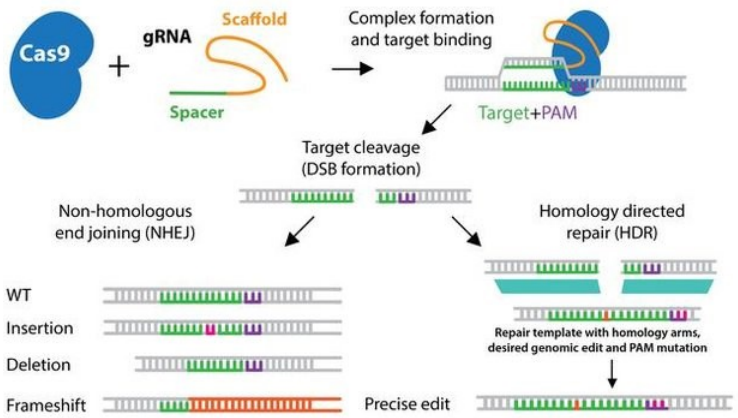


Fig. 53: Outline of CRISPR/CAS9 DSB and repair via NHEJ/ HDR. (Iqbal et al., 2020)

As can be seen in Fig. 53, both tools can now be combined (Kovač et al., 2020):

a. **Single guide (sg) RNA**, complementary to the target DNA, in this case one has targeted repetitive Alu sequences in the human genome.

b. **Fusion protein of a one SB100X and an inactivated Cas9** (dCas9) connected by a **linker** (Fig. 54 right)

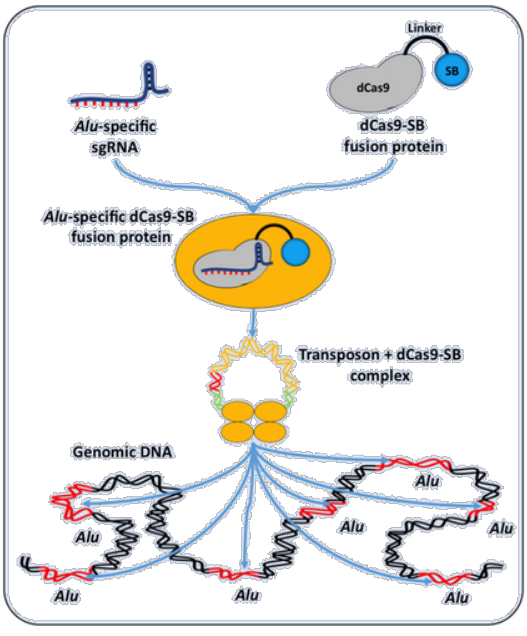


Fig.54: sgRNA guided transposition, a combination of the targeting capacity of Crispr-Cas9 systems and the insertion capacity of SB transposons (Amberger & Ivics, 2020).

9.7 SB Transposons in Oncogenomics

Transposons, whether PiggyBack (PB) or Sleeping Beauty (SB) have also been applied to oncology in the last decade (Rad et al., 2010; Starr et al., 2009; March et al., 2011). In particular, insertional mutagenesis of transposons has been used to: 1. describe tumor evolution in further detail, 2. understand networks of cooperating genes in a tumor entity, 3. differentiate between driver and passenger mutations in a specific genetic tumor profile, 4. understand the process of metastasis and 5. identify any resistance against therapeutic approaches.

Thus, SB-mediated insertional mutagenesis became an important tool to provide prognostic and predictive information for tumors. In combination with comparative oncogenomics, in which candidate cancer (driver) genes (CCGs) are compared between human and mouse models, the succession of cancerous mutations and thus the origin and progression of tumorigenesis can be better understood (Morris et al., 2016).

In the past, retroviral mutagenesis was the main tool used to identify CCGs. However, as described in the last chapter, viral screening tools are not recommended because of one drawback: They do not interact as randomly as transposons with the targeted genome and consequently have a higher promoter activity (Kool & Berns, 2009). Therefore, there has been a shift to using SB transposons for the purpose of CCG determination (Takeda, Jenkins & Copeland, 2021).

Abb. 55: Structure of the screening SB systems (SB10), that can lead to prematurely terminating expression of tumor suppressor gene or overexpression of oncogenes (Takeda, Jenkins & Copeland, 2021).

Figure 55 provides an example of the structure of the screening SB systems. From left to right we find the following structures with the corresponding functions:

- **IRRDRL:** unique inverted repeat sequences, which is recognized by transposase.
- **SA:** splice site acceptor, can lead to loss of function by exon skipping in the case of a tumor suppressor gene and thus to truncation of the protein product.
- **5′ LT-RMSCV:** 5′ long terminal repeat of the murine stem cell virus containing powerful promoter and enhancer elements leading to an overexpression of oncogenes.
- **SD:** splice site donor
- **IRRDRR:** right inverted repeat sequence.

In comparative oncogenomics, transgenic mouse models that already carry the SB10 transposase by knock-in in the genome are used. In addition, the lox-STOP-lox[11] sequence was also introduced in mouse genomes, and now the SB system can be controlled patio temporally by tamoxifen-inducible tissue-specific Cre recombinases.

To identify CCGs for colorectal cancer, for example, several strains of transgenic mice already carrying a mutant APC allele are treated with SB transposons. As early as 1992, Su et al. proposed that CRC develops via multi-step acquisition of mutations. For FAP, LoF of APC is initial, followed by overactivation of the KRAS gene and LoF of SMAD4, which may eventually lead to inactivation of TP53.

[11] The lox DNA sequence originates from bacteriophage and consists of 34 base pairs, including an asymmetric sequence of eight base pairs (the central two of which are conserved) flanked on both sides by palindromic thirteen bases. This is a recombination system. It allows the targeted removal of DNA sequences in living organisms. This technique can be used, for example, to genetically modify individual specific cell or tissue types, while leaving other tissues unaffected (Piovan et al., 2014).

Takeda, Jenkins and Copeland (2021) generate mouse models with corresponding mutations and show that APC plays the role of a gate-keeping gene in all mutants, but surprisingly different selection dynamics occur in SMAD4 knock-out mice. The researchers summarize:

> *[...] the cancer-causing mutation that occurs first, influences the nature of the mutation that is selected next, which in turn, influences the mutation that is selected third, and so on down the line. To put this another way, the order in which mutations accumulate in tumor cells over time is not random. Mutations occur randomly but the selection of mutations does not. This helps to explain why, in human CRC, the mutations largely occur in a non-random order, with APC mutations occurring first, KRAS mutations second, followed by SMAD4 and then TP53 mutations. (Takeda, Jenkins and Copeland 2021: 2094)*

Hou et al. (2014) propose the application of SB transposons not only as a screening tool as in the previous example, but also in gene therapy of cancer. For example, Chimeric Antigen Receptor (CAR) T-cell produced using SB transoposons have already been successfully used for immunotherapy of leukemia (Magnani et al., 2020). Another approach is the insertion of therapeutic transgenes into the genome of treated cells, so that a previously inactivated or truncated protein can now be permanently expressed, or genes whose expression leads to tumor cell death or impedes angiogenesis can be introduced.

Chang and Mou (2019) succeed in the latter, using a Sleeping Beauty transposon-mediated asparaginase to arrest cancer cell growth. As a semi-essential amino acid, asparagine plays a key role in tumor metabolism. Krall et al. (2016) show that apoptosis is initiated in tumors when the synthesis of asparagine is suppressed, this is typical for cancer cells as asparaginase synthetase is only produced in low amounts in leukemias in the cancer cell, among others - suggesting it as a therapeutic target. If the synthesis of this amino acid is now further inhibited via overexpression of asparaginase, tumor growth can be controlled.

10 Conclusion and Prospects

As already anticipated in the section on results: Transposons are integral components of viral and chromosomal genomes and, together with the circular ssRNA viroids, they provide the link to the primordial RNA world. Tranpsons themselves, like the Alu elements in the primate genome, but also their derivatives are important components of the human genome.

But why should we pay more attention to transposons? After having presented transposons as biomarkers in personalized medicine, as therapeutic targets in oncology, and as effective tools in oncogenomics in the preceding chapters, four further meta-conceptual reasons will now be given why transposons-related research desiderata should be brought more into focus:

1. Transposons make up a large proportion of the non-coding segments of DNA. The reality is that in the past, due to technical and practical limitations, the focus has been on coding regions. Sanger sequencing has primarily been used to sequence coding genes, usually specific exons. However, with the high-throughput techniques of next generation sequencing, these limitations have become obsolete. Nevertheless, the gold standard is still exome sequencing. This has not only to do with the fact that sequencing the exome or a panel of exons of candidate genes is cheaper - they are, above all, easier to interpret. This is because coding regions are subject to greater conservation pressure which makes aberrations, i.e. variants, easier to interpret. Non-coding regions, on the other hand, exhibit a polymorphism that has only been recorded to a limited extent to date, and deviations from this are difficult to measure - because this would require a reference genome that more or less provides a norm. And yet, with the use of whole genome sequencing, we will not be spared the opportunity to take non-coding regions into greater consideration, for example, in order to find causative genetic changes for tumor diseases**. Random and sproadic mutagenic events in non-coding regions happen at least fifty times more frequently than in the coding exome.** We recall: a maximum of 2% of DNA is coding. While promoter regions and splice sites have recently been included in variant

annotation, intronic variations are often left out. Non-coding variants outside the intron receive little attention.

Transposons offer a practical starting point for the study of these non-coding regions: transposons have structural features that could help us to stratify and order these exome regions. Another suggestion that could eventually find its way into the clinical research would be not to map the tumor genome against a reference genome but to use the DNA of non-cancerous cells as a reference and compare them.

2. Research around transposons could represent a step forward in **theory building**. Indeed, we have no shortage of empirical data today - through Open Access and Open Data options, researchers have an incredible amount of empirical data at their disposal. However, as mentioned in the first point - our interpretive scope is limited by the polymorphism of non-coding domains and by the lack of knowledge about the implication of different non-coding mutations. We need theoretical concepts that relate these vast amounts of empirical data. As well as new hypotheses, we need to have the courage to challenge old concepts such as the somatic mutation theory. In this context, we should also take the research of Barbara McClintock as a good example. About a decade before the discovery of the helical structure of the DNA, she had already hypothesized the existence of "jumping genes" on the basis of her empirical observations. This should not, however, be done with rigid concepts and dogmatic doctrines as in the past, but rather by promoting a healthy interrelationship between theoretical and empirical data.

3. Similarly, the **relationship between embryogenesis and tumorigenesis** should be researched more intensively. Transposons, which play a key role in both developments, could be the starting point for comparative genomics to not only understand the origin and progression of cancers (in the context of aging processes) but to find a therapeutic approach that offers chances of success for all tumor entities and equips personalized medicine with powerful therapeutic targets.

4. Fourth and finally, the study of transposons is a cross-sectional topic that is not only relevant for oncogenomic questions but also connects evolutionary, phylogenetic, ontogenetic and agig-related research fields. The **transdisciplinary approach** has allowed the Sleeping Beauty transposon to be awakened from the fish genome for research on the mammalian genomes - and the transdisciplinary approach has also promoted the combination of transposons as powerful vectors together with deactivated RNA-directed CRISPR-CAS components for more accurate work with target genomes. The collaboration of genetics, medicine, evolutionary biology, immunohistochemistry and other disciplines will open new perspectives in the future.

Finally, as in the introduction, I would like to cite the concluding quote from **Baraba McClintock** in her 1983 Nobel Lecture:

> *"The mobility of these activated elements [transposons] allows them to enter different gene loci and to take over control of action of the gene wherever one may enter. Because the broken end of a chromosome entering a telophase nucleus can initiate activations of a number of different potentially transposable elements, the modifications these elements induce in the genome may be explored readily. In addition to modifying gene action, these elements can restructure the genome at various levels, from small changes involving a few nucleotides, to gross modifications involving large segments of chromosomes, such as duplications, deficiencies, inversions, and other more complex reorganizations. It was these various effects of an initial traumatic event that alerted me to anticipate unusual responses of a genome to various shocks it might receive, either produced by accidents occurring within the cell itself, or imposed from without, such as virus infections, species crosses, poisons of various sorts, or even altered environments such as those imposed by tissue culture. [...] In the future attention undoubtedly will be centered on the genome, and with greater appreciation of its significance as a highly sensitive organ of the cell, monitoring genomic activities and correcting common errors, sensing the unusual and unexpected events, and responding to them, often by restructuring the genome."* (McClintock, 1983)

11. Bibliography

1. Adams, N.M., Grassmann, S. and Sun, J.C. (2020). Clonal expansion of innate and adaptive lymphocytes. Nature Reviews Immunology, [online] 20(11), pp.694–707. Available at: https://pubmed.ncbi.nlm.nih.gov/32424244/ [Accessed 25 Jul. 2021].

2. Aiewsakun, P. and Katzourakis, A. (2017). Marine origin of retroviruses in the early Palaeozoic Era. Nature Communications, [online] 8(1). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5512871/ [Accessed 13 Jan. 2021].

3. Alanagreh, L., Alzoughool, F. and Atoum, M. (2020). The Human Coronavirus Disease COVID-19: Its Origin, Characteristics, and Insights into Potential Drugs and Its Mechanisms. Pathogens, [online] 9(5), p.331. Available at: https://pubmed.ncbi.nlm.nih.gov/32365466/ [Accessed 30 Dec. 2020].

4. Amoretti, M., Amsler, C., Bonomi, G., Bouchta, A., Bowe, P., Carraro, C., Cesar, C.L., Charlton, M., Collier, M.J.T., Doser, M., Filippini, V., Fine, K.S., Fontana, A., Fujiwara, M.C., Funakoshi, R., Genova, P., Hangst, J.S., Hayano, R.S., Holzscheiter, M.H. and Jørgensen, L.V. (2002). Production and detection of cold antihydrogen atoms. Nature, [online] 419(6906), pp.456–459. Available at: https://doi.org/10.1038/.

5. Anisimova Editor, M. (n.d.). Evolutionary Genomics Statistical and Computational Methods Second Edition Methods in Molecular Biology 1910. [online] . Available at: https://link.springer.com/content/pdf/10.1007%2F978-1-4939-9074-0.pdf.

6. Anwar, S., Wulaningsih, W. and Lehmann, U. (2017). Transposable Elements in Human Cancer: Causes and Consequences of Deregulation. International Journal of Molecular Sciences, [online] 18(5), p.974. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5454887/ [Accessed 22 Mar. 2020].

7. Araneda, R.C., Peterlin, Z., Zhang, X., Chesler, A. and Firestein, S. (2004). A pharmacological profile of the aldehyde receptor repertoire in rat olfactory epithelium. The Journal of Physiology, [online] 555(3), pp.743–756. Available at: https://pubmed.ncbi.nlm.nih.gov/14724183/ [Accessed 16 Aug. 2021].

8. Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M. and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Current Biology, [online] 11(13), pp.1017–1027. Available at: https://pubmed.ncbi.nlm.nih.gov/11470406/ [Accessed 3 Feb. 2021].

9. Arias, J.F., Koyama, T., Kinomoto, M. and Tokunaga, K. (2012). Retroelements versus APOBEC3 family members: No great escape from the magnificent seven. Frontiers in Microbiology, [online] 3. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2012.00275/full [Accessed 1 Feb. 2021].

10. Aribi, M. (2020). Immunogenetic Aspect of B-Cell Antigen Receptor Diversity Generation. Normal and Malignant B-Cell. [online] Available at: https://www.intechopen.com/chapters/70842 [Accessed 26 Jul. 2021].

11. Ay C;Bencur P;Vormittag R;Sailer T;Jungbauer C;Vukovich T;Mannhalter C;Pabinger I (2018). The angiotensin-converting enzyme insertion/deletion polymorphism and serum levels of angiotensin-converting enzyme in venous thromboembolism. Data from a case control study. Thrombosis and haemostasis, [online] 98(4). Available at: https://pubmed.ncbi.nlm.nih.gov/17938801/ [Accessed 30 Dec. 2020].

12. Azuma, M. (2019). Co-signal Molecules in T-Cell Activation. Co-signal Molecules in T Cell Activation, [online] pp.3–23. Available at: https://pubmed.ncbi.nlm.nih.gov/31758529/ [Accessed 29 Jul. 2021].

13. Baba, Y., Murata, A., Watanabe, M. and Baba, H. (2013). Clinical implications of the LINE-1 methylation levels in patients with gastrointestinal cancer. Surgery Today, [online] 44(10), pp.1807–1816. Available at: https://link.springer.com/article/10.1007/s00595-013-0763-6 [Accessed 23 Jan. 2021].

14. Babaian, A. and Mager, D.L. (2016). Endogenous retroviral promoter exaptation in human cancer. Mobile DNA, [online] 7(1). Available at: https://mobilednajournal.biomedcentral.com/articles/10.1186/s13100-016-0080-x [Accessed 17 Aug. 2021].

15. Babaian, A., Romanish, M.T., Gagnier, L., Kuo, L.Y., Karimi, M.M., Steidl, C. and Mager, D.L. (2015a). Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. Oncogene, [online] 35(19), pp.2542–2546. Available at: https://www.nature.com/articles/onc2015308 [Accessed 17 Aug. 2021].

16. Bachman, J. (2013). Reverse-Transcription PCR (RT-PCR). Laboratory Methods in Enzymology: RNA, [online] pp.67–74. Available at: https://pubmed.ncbi.nlm.nih.gov/24034314/ [Accessed 17 Aug. 2021].

17. Bae, J.M., Shin, S.-H., Kwon, H.-J., Park, S.-Y., Kook, M.C., Kim, Y.-W., Cho, N.-Y., Kim, N., Kim, T.-Y., Kim, D. and Kang, G.H. (2012). ALU and LINE-1 hypomethylations in multistep gastric carcinogenesis and their prognostic implications. International Journal of Cancer, [online] 131(6), pp.1323–1331. Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.27369 [Accessed 20 Jul. 2021].

18. Baltimore, D. (1970). Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. Nature, [online] 226(5252), pp.1209–1211. Available at: https://www.nature.com/articles/2261209a0 [Accessed 12 Jan. 2021].

19. Belfort, M., Curcio, M.J. and Lue, N.F. (2011). Telomerase and retrotransposons: Reverse transcriptases that shaped genomes. Proceedings of the National Academy of Sciences, [online] 108(51), pp.20304–20310. Available at: https://www.pnas.org/content/108/51/20304 [Accessed 31 Jul. 2021].

20. Berger-Bächi, B. (1983). Insertional inactivation of staphylococcal methicillin resistance by Tn551. Journal of bacteriology, [online] 154(1), pp.479–87. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC217482/ [Accessed 22 Mar. 2020].

21. Biémont, C. (2010). A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution: Figure 1.—. Genetics, [online] 186(4), pp.1085–1093. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2998295/ [Accessed 21 Mar. 2020].

22. Bijlsma, R. and Loeschcke, V. (1997). Environmentam Stress Adaptation and Evolution, Birkhäuser. Berlin.

23. Bire, S., Ley, D., Casteret, S., Mermod, N., Bigot, Y. and Rouleux-Bonnin, F. (2013). Optimization of the piggyBac Transposon Using mRNA and Insulators: Toward a More Reliable Gene Delivery System. PLoS ONE, [online] 8(12), p.e82559. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3849487/ [Accessed 15 Aug. 2021].

24. Bohn, M.K., Hall, A., Sepiashvili, L., Jung, B., Steele, S. and Adeli, K. (2020). Pathophysiology of COVID-19: Mechanisms Underlying Disease Severity and Progression. Physiology, [online] 35(5), pp.288–301. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7426542/.

25. Bolli, P., Sticchi, E., Giusti, B., Saracini, C., Abbate, R. and Fatini, C. (2010). Improvement inACEI/D polymorphism detection. Journal of the Renin-Angiotensin-Aldosterone System, 12(3), pp.129–132.

26. Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., Mager, D.L. and Feschotte, C. (2018). Ten things you should know about transposable elements. Genome Biology, [online] 19(1). Available at: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1577-z [Accessed 16 Aug. 2021].

27. Bousios, A., Nützmann, H.-W., Buck, D. and Michieletto, D. (2020a). Integrating transposable elements in the 3D genome. Mobile DNA, [online] 11(1). Available at: https://mobilednajournal.biomedcentral.com/articles/10.1186/s13100-020-0202-3 [Accessed 16 Aug. 2021].

28. Bouuaert,C. , Tellier, M. and Chalmers, R. (2014). One to rule them all. [online] ResearchGate. Available at:

https://www.researchgate.net/publication/262194183_One_to_rule_them_all [Accessed 15 Jul. 2021].

29. Brandt, V.L. and Roth, D.B. (2007). G.O.D.'s Holy Grail: Discovery of the RAG Proteins. The Journal of Immunology, [online] 180(1), pp.3–4. Available at: https://www.jimmunol.org/content/180/1/3 [Accessed 16 Aug. 2021].

30. Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V. and Kazazian, H.H. (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proceedings of the National Academy of Sciences, [online] 100(9), pp.5280–5285. Available at: https://pubmed.ncbi.nlm.nih.gov/12682288/ [Accessed 22 May 2021].

31. Brücher, B.L.D.M. and Jamall, I.S. (2016). Somatic Mutation Theory - Why it's Wrong for Most Cancers. Cellular Physiology and Biochemistry, [online] 38(5), pp.1663–1680. Available at: https://www.karger.com/Article/Fulltext/443106 [Accessed 16 Aug. 2021].

32. Burns, K.H. (2017). Transposable elements in cancer. Nature Reviews Cancer, [online] 17(7), pp.415–424. Available at: https://www.nature.com/articles/nrc.2017.35 [Accessed 20 Jul. 2021].

33. Burns, K.H. (2020). Our Conflict with Transposable Elements and Its Implications for Human Disease. Annual Review of Pathology: Mechanisms of Disease, [online] 15(1), pp.51–70. Available at: https://pubmed.ncbi.nlm.nih.gov/31977294/ [Accessed 16 Aug. 2021].

34. Cajuso, T., Sulo, P., Tanskanen, T., Katainen, R., Taira, A., Hänninen, U. A., Kondelin, J., Forsström, L., Välimäki, N., Aavikko, M., Kaasinen, E., Ristimäki, A., Koskensalo, S., Lepistö, A., Renkonen-Sinisalo, L., Seppälä, T., Kuopio, T., Böhm, J., Mecklin, J.-P., & Kilpivaara, O.

(2019). Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-11770-0

35. Calcagno, D.Q., Mota, E.R. da S., Moreira, F.C., de Sousa, S.B.M., Burbano, R.R. and Assumpção, P.P. (2019). Role of PIWI-Interacting RNA (piRNA) as Epigenetic Regulation. Handbook of Nutrition, Diet, and Epigenetics, [online] pp.187–209. Available at: https://link.springer.com/referenceworkentry/10.1007%2F978-3-319-55530-0_77 [Accessed 15 Jul. 2021].

36. Callinan, P.A. and Batzer, M.A. (2006). Retrotransposable Elements and Human Disease. Genome and Disease, [online] pp.104–115. Available at: https://pubmed.ncbi.nlm.nih.gov/18724056/ [Accessed 16 Aug. 2021].

37. Cantarella, S., Carnevali, D., Morselli, M., Conti, A., Pellegrini, M., Montanini, B. and Dieci, G. (2019). Alu RNA Modulates the Expression of Cell Cycle Genes in Human Fibroblasts. International Journal of Molecular Sciences, [online] 20(13), p.3315. Available at: https://www.mdpi.com/1422-0067/20/13/3315 [Accessed 29 Dec. 2020].

38. Capy, P., Gasperi, G., Biémont, C. and Bazin, C. (2000). Stress and transposable elements: co-evolution or useful parasites? Heredity, [online] 85(2), pp.101–106. Available at: https://www.nature.com/articles/6887510 [Accessed 20 Jan. 2021].

39. Carmona, L.M. and Schatz, D.G. (2017). New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. The FEBS Journal, [online] 284(11), pp.1590–1605. Available at: https://pubmed.ncbi.nlm.nih.gov/27973733/ [Accessed 26 Jul. 2021].

40. Carroll, M.L., Roy-Engel, A.M., Nguyen, S.V., Salem, A.-H., Vogel, E., Vincent, B., Myers, J., Ahmad, Z., Nguyen, L., Sammarco, M., Watkins, W.S., Henke, J., Makalowski, W., Jorde, L.B., Deininger, P.L. and Batzer, M.A. (2001). Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. Journal of Molecular Biology, 311(1), pp.17–40.

41. Carthew, R.W. and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell, [online] 136(4), pp.642–655. Available at: https://pubmed.ncbi.nlm.nih.gov/19239886/ [Accessed 3 Feb. 2021].

42. Casacuberta, E. (2017). Drosophila: Retrotransposons Making up Telomeres. Viruses, [online] 9(7), p.192. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5537684/ [Accessed 16 Aug. 2021].

43. Casacuberta, E. and González, J. (2013). The impact of transposable elements in environmental adaptation. Molecular Ecology, 22(6), pp.1503–1517.

44. Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S.M., Turelli, P. and Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. Genes & Development, [online] 28(13), pp.1397–1409. Available at: http://genesdev.cshlp.org/content/28/13/1397 [Accessed 24 Jan. 2021].

45. Chang, J.-H., Mou, K.Y. and Mou, C.-Y. (2019). Sleeping Beauty Transposon-Mediated Asparaginase Gene Delivery by a Nanoparticle Platform. Scientific Reports, [online] 9(1). Available at: https://www.nature.com/articles/s41598-019-47927-6 [Accessed 15 Aug. 2021].

46. Chi, X., Li, Y. and Qiu, X. (2020). V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. Immunology, [online] 160(3), pp.233–247. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/imm.13176 [Accessed 6 Feb. 2021].

47. Cho, N-Y., Kim, B-H., Choi, M., Yoo, E., Moon, K., Cho, Y-M., Kim, D. and Kang, G. (2007). Hypermethylation of CpG island loci and hypomethylation of LINE-1 and Alu repeats in prostate adenocarcinoma and their relationship to clinicopathological features. The Journal of Pathology, [online] 211(3), pp.269–277. Available at: https://onlinelibrary.wiley.com/doi/epdf/10.1002/path.2106?saml_referrer [Accessed 23 Jan. 2021].

48. Cintra, M.T.R., Balarin, M.A.S., Tanaka, S.C.S.V., Silva, V.I.M. da, Marqui, A.B.T. de, Resende, E.A.M.R. de, Lima, M.F.P. and Gomes, M.K.O. (2018). Polycystic ovarian syndrome: rs1799752 polymorphism of ACE gene. Revista da Associação Médica Brasileira, [online] 64(11), pp.1017–1022. Available at: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-42302018001101017&lng=en&tlng=en [Accessed 29 Dec. 2020].

49. Coffin, J.M. and Fan, H. (2016). The Discovery of Reverse Transcriptase. Annual Review of Virology, [online] 3(1), pp.29–51. Available at: https://pubmed.ncbi.nlm.nih.gov/27482900/ [Accessed 12 Jan. 2021].

50. Cosby, R.L., Chang, N.-C. and Feschotte, C. (2019). Host–transposon interactions: conflict, cooperation, and cooption. Genes & Development, [online] 33(17-18), pp.1098–1116. Available at: http://genesdev.cshlp.org/content/33/17-18/1098.full.html [Accessed 16 Aug. 2021].

51. Cotton, J. (2001). Retroviruses from retrotransposons. Genome Biology, [online] 2(2). Available at: https://genomebiology.biomedcentral.com/articles/10.1186/gb-2001-2-2-reports0006 [Accessed 13 Jan. 2021].

52. Crichton, J.H., Dunican, D.S., Maclennan, M., Meehan, R.R. and Adams, I.R. (2014). Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. Cellular and molecular life sciences : CMLS, [online] 71(9), pp.1581–605. Available at: https://www.ncbi.nlm.nih.gov/pubmed/24045705/ [Accessed 22 Mar. 2020].

53. Davies, D., Goryshin, I.Y., Reznikoff, W.S. and Rayment, I. (2000). Three-Dimensional Structure of the Tn5 Synaptic Complex Transposition Intermediate. [online] ResearchGate. Available at: https://www.researchgate.net/publication/12435343_Three-Dimensional_Structure_of_the_Tn5_Synaptic_Complex_Transposition_Intermediate [Accessed 15 Jul. 2021].

54. Day, D.S., Luquette, L.J., Park, P.J. and Kharchenko, P.V. (2010). Estimating enrichment of repetitive elements from high-throughput sequence data. Genome Biology, [online] 11(6), p.R69. Available at: https://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-6-r69 [Accessed 24 Jan. 2021].

55. De Brakeleer, S., De Grève, J. and Teugels, E. (2020). A systematic screen of breast cancer patients' exomes for retrotransposon insertions reveals disease associated genes. [online] Available at: https://www.biorxiv.org/content/10.1101/2020.06.04.123240v1.full [Accessed 30 Dec. 2020].

56. De Maat, M.F.G., Umetani, N., Sunami, E., Turner, R.R. and Hoon, D.S.B. (2007). Assessment of Methylation Events during Colorectal

Tumor Progression by Absolute Quantitative Analysis of Methylated Alleles. Molecular Cancer Research, [online] 5(5), pp.461–471. Available at: https://pubmed.ncbi.nlm.nih.gov/17510312/ [Accessed 20 Jul. 2021].

57. Delanghe, J.R., Speeckaert, M.M. and De Buyzere, M.L. (2020). COVID-19 infections are also affected by human ACE1 D/I polymorphism. Clinical Chemistry and Laboratory Medicine (CCLM), [online] 58(7), pp.1125–1126. Available at: https://www.degruyter.com/view/journals/cclm/58/7/article-p1125.xml?language=en [Accessed 30 Dec. 2020].

58. Denli, Ahmet M., Narvaiza, I., Kerman, Bilal E., Pena, M., Benner, C., Marchetto, Maria C.N., Diedrich, Jolene K., Aslanian, A., Ma, J., Moresco, James J., Moore, L., Hunter, T., Saghatelian, A. and Gage, Fred H. (2015). Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. Cell, [online] 163(3), pp.583–593. Available at: https://pubmed.ncbi.nlm.nih.gov/26496605/ [Accessed 23 May 2021].

59. Dewannieux, M., Esnault, C. and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nature Genetics, [online] 35(1), pp.41–48. Available at: https://delmbsgqukah2ms4rqeonbcu3i--pubmed-ncbi-nlm-nih-gov.translate.goog/12897783/ [Accessed 30 Dec. 2020].

60. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature, [online] 485(7398), pp.376–380. Available at: https://pubmed.ncbi.nlm.nih.gov/22495300/ [Accessed 16 Aug. 2021].

61. Donner, H., Tonjes, R.R., Bontrop, R.E., Kurth, R., Usadel, K.H. and Badenhoop, K. (1999). Intronic sequence motifs of HLA-DQB1 are shared

between humans, apes and old world monkeys, but a retroviral LTR element (DQLTR3) is human specific. Tissue Antigens, [online] 53(6), pp.551–558. Available at: https://pubmed.ncbi.nlm.nih.gov/10395105/ [Accessed 15 Jul. 2021].

62. Doudna, J.A. and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science, [online] 346(6213), p.1258096. Available at: https://science-sciencemag-org.translate.goog/content/346/6213/1258096.long?_x_tr_sl=auto&_x_tr_tl=de&_x_tr_hl=de&_x_tr_pto=ajax,se,elem [Accessed 15 Aug. 2021].

63. Ege, M. (2005). Omenn syndrome due to ARTEMIS mutations. Blood, [online] 105(11), pp.4179–4186. Available at: https://pubmed.ncbi.nlm.nih.gov/15731174/ [Accessed 8 Feb. 2021].

64. Ellison, V., Gerton, J., Vincent, K.A. and Brown, P.O. (1995). An Essential Interaction between Distinct Domains of HIV-1 Integrase Mediates Assembly of the Active Multimer. Journal of Biological Chemistry, [online] 270(7), pp.3320–3326. Available at: https://www.jbc.org/article/S0021-9258(18)82928-3/fulltext [Accessed 2 Feb. 2021].

65. Ellsworth, R.E., Ellsworth, D.L., Patney, H.L., Deyarmin, B., Love, B., Hooke, J.A. and Shriver, C.D. (2008). Amplification of HER2 is a marker for global genomic instability. BMC Cancer, [online] 8(1). Available at: https://link.springer.com/article/10.1186/1471-2407-8-297 [Accessed 20 Jul. 2021].

66. Erwin, J.A., Marchetto, M.C. and Gage, F.H. (2014). Mobile DNA elements in the generation of diversity and complexity in the brain. Nature Reviews Neuroscience, 15(8), pp.497–506.

67. Fanning, T.G. and Singer, M.F. (1987). LINE-1: A mammalian transposable element. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, [online] 910(3), pp.203–212. Available at: https://pubmed.ncbi.nlm.nih.gov/2445384/ [Accessed 29 Jun. 2021].

68. FDA (2020). FDA approves first cancer treatment for any solid tumor with a specific genetic feature. [online] U.S. Food and Drug Administration. Available at: https://www.fda.gov/news-events/press-announcements/fda-approves-first-cancer-treatment-any-solid-tumor-specific-genetic-feature [Accessed 16 Jul. 2021].

69. Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution. Trends in genetics : TIG, [online] 5(4), pp.103–7. Available at: https://www.ncbi.nlm.nih.gov/pubmed/2543105 [Accessed 21 Mar. 2020].

70. Fontana, M.F. and Vance, R.E. (2011). Two signal models in innate immunity. Immunological Reviews, [online] 243(1), pp.26–39. Available at: https://pubmed.ncbi.nlm.nih.gov/21884165/ [Accessed 25 Jul. 2021].

71. Fugmann, S.D. (2010). The origins of the Rag genes—From transposition to V(D)J recombination. Seminars in Immunology, [online] 22(1), pp.10–16. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2823946/ [Accessed 16 Aug. 2021].

72. Gagliardi, M.C., Tieri, P., Ortona, E. and Ruggieri, A. (2020). ACE2 expression and sex disparity in COVID-19. Cell Death Discovery, [online] 6(1). Available at: https://www.nature.com/articles/s41420-020-0276-1 [Accessed 28 Dec. 2020].

73. Galli, U.M., Sauter, M., Lecher, B., Maurer, S., Herbst, H., Roemer, K. and Mueller-Lantzsch, N. (2005). Human endogenous retrovirus rec interferes with germ cell development in mice and may cause carcinoma in situ, the predecessor lesion of germ cell tumors. Oncogene, [online] 24(19), pp.3223–8. Available at: https://www.ncbi.nlm.nih.gov/pubmed/15735668/ [Accessed 22 Mar. 2020].

74. Garcia-Pérez, J.L. (2016). Transposons and Retrotransposons: Methods and Protocols. [online] ResearchGate. Available at: https://www.researchgate.net/publication/321530042_Transposons_and_Retrotransposons_Methods_and_Protocols [Accessed 22 Mar. 2020].

75. Garcia-Perez, J.L., Widmann, T.J. and Adams, I.R. (2016). The impact of transposable elements on mammalian development. Development, [online] 143(22), pp.4101–4114. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5830075/ [Accessed 22 Mar. 2020].

76. Gerdes, P., Richardson, S.R., Mager, D.L. and Faulkner, G.J. (2016b). Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. Genome Biology, [online] 17(1). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4862087/ [Accessed 22 Mar. 2020].

77. Ghazi, L. and Drawz, P. (2017). Advances in understanding the renin-angiotensin-aldosterone system (RAAS) in blood pressure control and recent pivotal trials of RAAS blockade in heart failure and diabetic nephropathy. F1000Research, [online] 6, p.297. Available at: https://f1000research.com/articles/6-297/v1 [Accessed 28 Dec. 2020].

78. Gilbert, N., Lutz, S., Morrish, T.A. and Moran, J.V. (2005). Multiple Fates of L1 Retrotransposition Intermediates in Cultured Human Cells.

Molecular and Cellular Biology, [online] 25(17), pp.7780–7795. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1190285/ [Accessed 23 May 2021].

79. Gladyshev, E.A. and Arkhipova, I.R. (2009). Rotifer rDNA-specific R9 retrotransposable elements generate an exceptionally long target site duplication upon insertion. Gene, 448(2), pp.145–150.

80. Goodier, J.L. (2016). Restricting retrotransposons: a review. [online] ResearchGate. Available at: https://www.researchgate.net/publication/306077875_Restricting_retrotransposons_a_review [Accessed 15 Jul. 2021].

81. Griffiths, E. A., & Gore, S. D. (2008). DNA Methyltransferase and Histone Deacetylase Inhibitors in the Treatment of Myelodysplastic Syndromes. Seminars in Hematology, 45(1), 23–30. https://doi.org/10.1053/j.seminhematol.2007.11.007

82. Gualtieri, A., Andreola, F., Sciamanna, I., Sinibaldi-Vallebona, P., Serafino, A. and Spadafora, C. (2013). Increased expression and copy number amplification of LINE-1 and SINE B1 retrotransposable elements in murine mammary carcinoma progression. Oncotarget, [online] 4(11), pp.1882–1893. Available at: https://pubmed.ncbi.nlm.nih.gov/24231191/ [Accessed 25 Jul. 2021].

83. Guida, V., Cernilogar, F.M., Filograna, A., De Gregorio, R., Ishizu, H., Siomi, M.C., Schotta, G., Bellenchi, G.C. and Andrenacci, D. (2016). Production of Small Noncoding RNAs from the flamenco Locus Is Regulated by the gypsy Retrotransposon of Drosophila melanogaster. Genetics, [online] 204(2), pp.631–644. Available at: https://pubmed.ncbi.nlm.nih.gov/27558137/ [Accessed 3 Feb. 2021].

84. Guo, Y., Zhang, Y. and Hu, K. (2017). Sleeping Beauty transposon integrates into non-TA dinucleotides via an alternative mechanism. [online] Available at: https://www.biorxiv.org/content/10.1101/177584v1 [Accessed 15 Aug. 2021].

85. Hackett, P.B., Largaespada, D.A., Switzer, K.C. and Cooper, L.J.N. (2013). Evaluating risks of insertional mutagenesis by DNA transposons in gene therapy. Translational Research, [online] 161(4), pp.265–283. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3602164/ [Accessed 15 Aug. 2021].

86. Hall, B. G. (1998). Adaptive mutagenesis: a process that generates almost exclusively beneficial mutations. Genetica, 102–103, 109–125.

87. Hall, B. G. (2000). Mobile elements as activators of cryptic genes in. E Coli Genetica, in press.

88. Hall, R. M. (1997). Mobile gene cassettes and integrons: moving antibiotic resistance genes in gram-negative bacteria. Ciba Foundation Symposium, 207: 192–202.

89. Han, M., Li, J., Cao, Y., Huang, Y., Li, W., Zhu, H., Zhao, Q., Han, J.-D. J., Wu, Q., Li, J., Feng, J., & Wong, J. (2020). A role for LSH in facilitating DNA methylation by DNMT1 through enhancing UHRF1 chromatin association. Nucleic Acids Research, 48(21), 12116–12134. https://doi.org/10.1093/nar/gkaa1003

90. Hancks, D.C. and Kazazian, H.H. (2010). SVA retrotransposons: Evolution and genetic instability. Seminars in Cancer Biology, [online] 20(4), pp.234–245. Available at:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2945828/ [Accessed 22 Mar. 2020].

91. Hancks, D.C. and Kazazian, H.H. (2016). Roles for retrotransposon insertions in human disease. Mobile DNA, [online] 7(1). Available at: https://tkw5qgfqtbwpibfk3tfuoodgte--pubmed-ncbi-nlm-nih-gov.translate.goog/27158268/ [Accessed 15 Apr. 2021].

92. Harris, C.R., Normart, R., Yang, Q., Stevenson, E., Haffty, B.G., Ganesan, S., Cordon-Cardo, C., Levine, A.J. and Tang, L.H. (2010). Association of Nuclear Localization of a Long Interspersed Nuclear Element-1 Protein in Breast Tumors with Poor Prognostic Outcomes. Genes & Cancer, [online] 1(2), pp.115–124. Available at: https://journals.sagepub.com/doi/full/10.1177/1947601909360812 [Accessed 15 Jul. 2021].

93. Häsler, J. and Strub, K. (2006). Alu elements as regulators of gene expression. Nucleic Acids Research, [online] 34(19), pp.5491–5497. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1636486/ [Accessed 16 Aug. 2021].

94. Hatami, N., Ahi, S., Sadeghinikoo, A., Foroughian, M., Javdani, F., Kalani, N., Fereydoni, M., Keshavarz, P. and hosseini, A. (2020). Worldwide ACE (I/D) polymorphism may affect COVID-19 recovery rate: an ecological meta-regression. Endocrine, [online] 68(3), pp.479–484. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7294766/ [Accessed 30 Dec. 2020].

95. Hayes, J.M., Cosgrave, E.F.J., Struwe, W.B., Wormald, M., Davey, G.P., Jefferis, R. and Rudd, P.M. (2014). Glycosylation and Fc Receptors. Fc

Receptors, [online] pp.165–199. Available at: https://pubmed.ncbi.nlm.nih.gov/25116100/ [Accessed 30 Jul. 2021].

96. He, J., Fu, X., Zhang, M., He, F., Li, W., Abdul, M.Md., Zhou, J., Sun, L., Chang, C., Li, Y., Liu, H., Wu, K., Babarinde, I.A., Zhuang, Q., Loh, Y.-H., Chen, J., Esteban, M.A. and Hutchins, A.P. (2019). Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. Nature Communications, [online] 10(1). Available at: https://pubmed.ncbi.nlm.nih.gov/30604769/ [Accessed 24 Jan. 2021].

97. Hennig, B.P., Velten, L., Racke, I., Tu, C.S., Thoms, M., Rybin, V., Besir, H., Remans, K. and Steinmetz, L.M. (2018). Large-Scale Low-Cost NGS Library Preparation Using a Robust Tn5 Purification and Tagmentation Protocol. G3 Genes|Genomes|Genetics, [online] 8(1), pp.79–89. Available at: https://pubmed.ncbi.nlm.nih.gov/29118030/ [Accessed 17 Aug. 2021].

98. Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D. and Baylin, S.B. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proceedings of the National Academy of Sciences, [online] 93(18), pp.9821–9826. Available at: https://www.pnas.org/content/93/18/9821 [Accessed 20 Jul. 2021].

99. Herquel, B., Ouararhni, K., Martianov, I., Le Gras, S., Ye, T., Keime, C., Lerouge, T., Jost, B., Cammas, F., Losson, R. and Davidson, I. (2013). Trim24-repressed VL30 retrotransposons regulate gene expression by producing noncoding RNA. Nature structural & molecular biology, [online] 20(3), pp.339–46. Available at: https://www.ncbi.nlm.nih.gov/pubmed/23377542 [Accessed 22 Mar. 2020].

100. Herrera RJ, et al. (2016). Genomes evolution and culture: past present and future of humankind. New York, NY: Wiley-Blackwell.

101. Heurich, A., Hofmann-Winkler, H., Gierer, S., Liepold, T., Jahn, O. and Pohlmann, S. (2013). TMPRSS2 and ADAM17 Cleave ACE2 Differentially and Only Proteolysis by TMPRSS2 Augments Entry Driven by the Severe Acute Respiratory Syndrome Coronavirus Spike Protein. Journal of Virology, [online] 88(2), pp.1293–1307. Available at: https://jvi.asm.org/content/88/2/1293 [Accessed 27 Dec. 2020].

102. Hoffmann, A. A. and Parsons, P. A. (1997). Extreme Environmental Change and Evolution, Cambridge University Press.

103. Hofmann, H. and Pöhlmann, S. (2004). Cellular entry of the SARS coronavirus. Trends in Microbiology, [online] 12(10), pp.466–472. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7119031/ [Accessed 27 Dec. 2020].

104. Honda, T. and Rahman, Md. (2019). Profiling of LINE-1-Related Genes in Hepatocellular Carcinoma. International Journal of Molecular Sciences, [online] 20(3), p.645. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6387036/ [Accessed 15 Jul. 2021].

105. Hongtao, L. and Fengju, K. (2016). Adaptive chaos parallel clonal selection algorithm for objective optimization in WTA application. Optik, [online] 127(6), pp.3459–3465. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0030402615020409 [Accessed 25 Jul. 2021].

106. Hosseinnejad, K., Yin, T., Gaskins, J.T., Bailen, J.L. and Jortani, S.A. (2018). Discovery of the Long Interspersed Nuclear Element-1 activation

product [Open Reading Frame-1 (ORF1) protein] in human blood. Clinica Chimica Acta, [online] 487, pp.228–232. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0009898118305266?via%3Dihub [Accessed 16 Jul. 2021].

107. Hou, X., Du, Y., Deng, Y., Wu, J. and Cao, G. (2014). Sleeping Beautytransposon system for genetic etiological research and gene therapy of cancers. Cancer Biology & Therapy, [online] 16(1), pp.8–16. Available at: https://pubmed.ncbi.nlm.nih.gov/25455252/ [Accessed 15 Aug. 2021].

108. Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979). A ubiquitous family of repeated DNA sequences in the human genome. Journal of Molecular Biology, 132(3), pp.289–306.

109. Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., & Eden, A. (2007). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene, 27(3), 404–408. https://doi.org/10.1038/sj.onc.1210631

110. Hozumi, N. and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proceedings of the National Academy of Sciences, [online] 73(10), pp.3628–3632. Available at: https://pubmed.ncbi.nlm.nih.gov/824647/ [Accessed 16 Aug. 2021].

111. Huang, S., Tao, X., Yuan, S., Zhang, Y., Li, P., Beilinson, Helen A., Zhang, Y., Yu, W., Pontarotti, P., Escriva, H., Le Petillon, Y., Liu, X., Chen, S., Schatz, David G. and Xu, A. (2016). Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. Cell, [online] 166(1), pp.102–114. Available at: https://pubmed.ncbi.nlm.nih.gov/27293192/ [Accessed 26 Jul. 2021].

112. Huang, X., Wilber, A.C., Bao, L., Tuong, D., Tolar, J., Orchard, P.J., Levine, B.L., June, C.H., McIvor, R.S., Blazar, B.R. and Zhou, X. (2006). Stable gene transfer and expression in human primary T cells by the Sleeping Beauty transposon system. Blood, [online] 107(2), pp.483–491. Available at: https://ashpublications.org/blood/article/107/2/483/131440/Stable-gene-transfer-and-expression-in-human [Accessed 12 Aug. 2021].

113. Hudecek, M., Izsvak, Z., Johen, S., Renner, M., Thuman, G. and Ivics, Z. (2017). Going non-viral: the Sleeping Beauty transposon system breaks on through to the clinical side. [online] Critical Reviews in Biochemistry and Molecular Biology. Available at: https://www.tandfonline.com/doi/full/10.1080/10409238.2017.1304354 [Accessed 12 Aug. 2021].

114. Hughes, S.H. (2015). Reverse Transcription of Retroviruses and LTR Retrotransposons. Microbiology Spectrum, [online] 3(2). Available at: https://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.MDNA3-0027-2014 [Accessed 21 Mar. 2020].

115. Ichiyanagi, K. and Okada, N. (2006). Genomic alterations upon integration of zebrafish L1 elements revealed by the TANT method. Gene, [online] 383, pp.108–116. Available at: https://pubmed.ncbi.nlm.nih.gov/17049188/ [Accessed 26 May 2021].

116. Iqbal, A., Shah, S.R.A., Cetingul, I.S., qudoos, A. and Bayram, I. (2020). A Review-Halal Animal Nutrition Perspective to the Halal Meat Production. Malaysian Journal of Halal Research, [online] 3(1), pp.17–23. Available at: https://www.sciendo.com/article/10.2478/mjhr-2020-0002 [Accessed 15 Aug. 2021].

117. Ivics, Z. (2016). Endogenous Transposase Source in Human Cells Mobilizes piggyBac Transposons. Molecular Therapy, [online] 24(5),

pp.851–854. Available at: https://www.cell.com/molecular-therapy-family/molecular-therapy/fulltext/S1525-0016(16)30212-X?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS152500161630212X%3Fshowall%3Dtrue [Accessed 15 Aug. 2021].

118. Ivics, Z., Li, M.A., Mátés, L., Boeke, J.D., Nagy, A., Bradley, A. and Izsvák, Z. (2009). Transposon-mediated genome manipulation in vertebrates. Nature methods, [online] 6(6), pp.415–22. Available at: https://www.ncbi.nlm.nih.gov/pubmed/19478801 [Accessed 22 Mar. 2020].

119. Iwagami, S., Baba, Y., Watanabe, M., Shigaki, H., Miyake, K., Ishimoto, T., Iwatsuki, M., Sakamaki, K., Ohashi, Y. and Baba, H. (2013). LINE-1 Hypomethylation Is Associated With a Poor Prognosis Among Patients With Curatively Resected Esophageal Squamous Cell Carcinoma. Annals of Surgery, [online] 257(3), pp.449–455. Available at: https://journals.lww.com/annalsofsurgery/Abstract/2013/03000/LINE_1_Hypomethylation_Is_Associated_With_a_Poor.12.aspx [Accessed 15 Jul. 2021].

120. Izsvák, Z., Ivics, Z. and Plasterk, R.H. (2000). Sleeping Beauty , a wide host-range transposon vector for genetic transformation in vertebrates 1 1Edited by J. Karn. Journal of Molecular Biology, [online] 302(1), pp.93–102. Available at: https://www.sciencedirect.com/science/article/pii/S0022283600940476?via%3Dihub [Accessed 12 Aug. 2021].

121. Janeway, C.A., Travers, P., Walport, M. and Shlomchik, M.J. (2014). The generation of diversity in immunoglobulins. [online] Nih.gov. Available at: https://www.ncbi.nlm.nih.gov/books/NBK27140/ [Accessed 16 Aug. 2021].

122. Jangam, D., Feschotte, C. and Betrán, E. (2017). Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. Trends in Genetics, [online] 33(11), pp.817–831. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5659911/ [Accessed 4 Aug. 2021].

123. Jansson, L.I. and Stone, M.D. (2019). Single-Molecule Analysis of Reverse Transcriptase Enzymes. Cold Spring Harbor Perspectives in Biology, [online] 11(9), p.a032458. Available at: https://cshperspectives.cshlp.org/content/11/9/a032458.long#ref-83 [Accessed 12 Jan. 2021].

124. Jiang, F., Yang, J., Zhang, Y., Dong, M., Wang, S., Zhang, Q., Liu, F.F., Zhang, K. and Zhang, C. (2014). Angiotensin-converting enzyme 2 and angiotensin 1–7: novel therapeutic targets. Nature Reviews Cardiology, [online] 11(7), pp.413–426. Available at: https://www.nature.com/articles/nrcardio.2014.59/figures/1 [Accessed 28 Dec. 2020].

125. Johnson, W.E. (2019). Origins and evolutionary consequences of ancient endogenous retroviruses. Nature Reviews Microbiology, [online] 17(6), pp.355–370. Available at: https://pubmed.ncbi.nlm.nih.gov/30962577/ [Accessed 13 Jan. 2021].

126. Jones, J.M. and Gellert, M. (2004). The taming of a transposon: V(D)J recombination and the immune system. Immunological Reviews, [online] 200(1), pp.233–248. Available at: https://zenodo.org/record/1230591#.YCD-7DFKhGo [Accessed 8 Feb. 2021].

127. Jung, C.J., Ménoret, S., Brusselle, L., Tesson, L., Usal, C., Chenouard, V., Remy, S., Ouisse, L.-H., Poirier, N., Vanhove, B., de Jong, P.J. and Anegon, I. (2016). Comparative Analysis of piggyBac, CRISPR/Cas9 and

TALEN Mediated BAC Transgenesis in the Zygote for the Generation of Humanized SIRPA Rats. Scientific Reports, [online] 6(1). Available at: https://www.nature.com/articles/srep31455 [Accessed 15 Aug. 2021].

128. Kajikawa, M. and Okada, N. (2002). LINEs Mobilize SINEs in the Eel through a Shared 3′ Sequence. Cell, 111(3), pp.433–444.

129. Kapitonov, V.V. and Jurka, J. (2003). Molecular paleontology of transposable elements in the Drosophila melanogaster genome. Proceedings of the National Academy of Sciences, [online] 100(11), pp.6569–6574. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC164487/ [Accessed 8 Feb. 2021].

130. Kapitonov, V.V. and Jurka, J. (2005). RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. PLoS Biology, [online] 3(6), p.e181. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1131882/ [Accessed 8 Feb. 2021].

131. Kapitonov, V.V. and Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. Proceedings of the National Academy of Sciences, 103(12), pp.4540–4545.

132. Kapitonov, V.V. and Koonin, E.V. (2015). Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon. Biology Direct, [online] 10(1). Available at: https://pubmed.ncbi.nlm.nih.gov/25928409/ [Accessed 8 Feb. 2021].

133. Kasahara, M., Suzuki, T. and Pasquier, L.D. (2004). On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. Trends in Immunology, [online] 25(2), pp.105–111.

Available at: https://pubmed.ncbi.nlm.nih.gov/15102370/ [Accessed 8 Feb. 2021].

134. Katzourakis, A. and Gifford, R.J. (2010). Endogenous Viral Elements in Animal Genomes. PLoS Genetics, [online] 6(11), p.e1001191. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2987831/ [Accessed 13 Jan. 2021].

135. Kazazian, H.H. (2004). Mobile Elements: Drivers of Genome Evolution. Science, [online] 303(5664), pp.1626–1632. Available at: https://science.sciencemag.org/content/303/5664/1626/tab-pdf [Accessed 21 Mar. 2020].

136. Kazazian, H.H. and Moran, J.V. (2017b). Mobile DNA in Health and Disease. New England Journal of Medicine, [online] 377(4), pp.361–370. Available at: https://pubmed.ncbi.nlm.nih.gov/28745987/ [Accessed 22 May 2021].

137. Kazazian, H.H., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G. and Antonarakis, S.E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature, [online] 332(6160), pp.164–6. Available at: https://www.ncbi.nlm.nih.gov/pubmed/2831458 [Accessed 22 Mar. 2020].

138. Kelleher, E.S. (2016). Reexamining the P-Element Invasion of Drosophila melanogaster Through the Lens of piRNA Silencing. Genetics, [online] 203(4), pp.1513–1531. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4981261/ [Accessed 3 Feb. 2021].

139. Klug, Wi., Cummins, M., Spencer, C. and Palladino, M. (2015). Concepts of genetics. [online] . Available at:

https://www.worldcat.org/title/concepts-of-genetics/oclc/880404074 [Accessed 16 Aug. 2021].

140. Koch, L. (2021). Capturing transposases for new proteins. Nature Reviews Genetics, [online] 22(5), pp.266–267. Available at: https://www.nature.com/articles/s41576-021-00347-7 [Accessed 16 Aug. 2021].

141. Konkel, M.K., Wang, J., Liang, P. and Batzer, M.A. (2007). Identification and characterization of novel polymorphic LINE-1 insertions through comparison of two human genome sequence assemblies. Gene, [online] 390(1-2), pp.28–38. Available at: https://pubmed.ncbi.nlm.nih.gov/17034961/ [Accessed 16 Jul. 2021].

142. Kool, J. and Berns, A. (2009). High-throughput insertional mutagenesis screens in mice to identify oncogenic networks. Nature Reviews Cancer, [online] 9(6), pp.389–399. Available at: https://pubmed.ncbi.nlm.nih.gov/19461666/ [Accessed 15 Aug. 2021].

143. Koonin, E.V. and Makarova, K.S. (2019). Origins and evolution of CRISPR-Cas systems. Philosophical Transactions of the Royal Society B: Biological Sciences, [online] 374(1772), p.20180087. Available at: https://royalsocietypublishing.org/doi/10.1098/rstb.2018.0087 [Accessed 16 Aug. 2021].

144. Kordyukova, M., Olovnikov, I. and Kalmykova, A. (2018). Transposon control mechanisms in telomere biology. Current Opinion in Genetics & Development, [online] 49, pp.56–62. Available at: https://pubmed.ncbi.nlm.nih.gov/29571043/ [Accessed 31 Jul. 2021].

145. Krall, A.S., Xu, S., Graeber, T.G., Braas, D. and Christofk, H.R. (2016). Asparagine promotes cancer cell proliferation through use as an amino acid exchange factor. Nature Communications, [online] 7(1). Available

at: https://www.nature.com/articles/ncomms11457 [Accessed 15 Aug. 2021].

146. Krupovic, M., Béguin, P. and Koonin, E.V. (2017). Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. Current Opinion in Microbiology, [online] 38, pp.36–43. Available at: https://www.sciencedirect.com/science/article/abs/pii/S1369527416301710?via%3Dihub [Accessed 16 Aug. 2021].

147. Kubiak, M.R. and Makałowska, I. (2017). Protein-Coding Genes' Retrocopies and Their Functions. Viruses, [online] 9(4), p.80. Available at: https://www.mdpi.com/1999-4915/9/4/80/htm [Accessed 15 Jul. 2021].

148. Kumar, A. (2020). Jump around: transposons in and out of the laboratory. F1000Research, [online] 9, p.135. Available at: https://www-1ncbi-1nlm-1nih-1gov-10013b50k0c67.han.medunigraz.at/pmc/articles/PMC7043111/ [Accessed 24 Sep. 2020].

149. Kurkjian, C., Kummar, S. and Murgo, A.J. (2008). DNA Methylation: Its Role in Cancer Development and Therapy. Current Problems in Cancer, [online] 32(5), pp.187–235. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2588419/ [Accessed 23 Jan. 2021].

150. Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köchert, K., Bouhlel, M.A., Richter, J., Soler, E., Stadhouders, R., Jöhrens, K., Wurster, K.D., Callen, D.F., Harte, M.F., Giefing, M., Barlow, R., Stein, H., Anagnostopoulos, I. and Janz, M. (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nature Medicine, [online] 16(5),

pp.571–579. Available at: https://pubmed.ncbi.nlm.nih.gov/20436485/ [Accessed 17 Aug. 2021].

151. Lander, E., Linton, L., Birren, B. and Nussbaum, C. (2001a). Initial sequencing and analysis of the human genome. Nature, [online] 409(6822), pp.860–921. Available at: https://www.nature.com/articles/35057062 [Accessed 24 Jan. 2019].

152. Lavasanifar, A., Sharp, C.N., Korte, E.A., Yin, T., Hosseinnejad, K. and Jortani, S.A. (2019). Long interspersed nuclear element-1 mobilization as a target in cancer diagnostics, prognostics and therapeutics. Clinica Chimica Acta, [online] 493, pp.52–62. Available at: https://pubmed.ncbi.nlm.nih.gov/30776360/ [Accessed 16 Jul. 2021].

153. Lee, H.-E., Ayarpadikannan, S. and Kim, H.-S. (2015). Role of transposable elements in genomic rearrangement, evolution, gene regulation and epigenetics in primates. Genes & genetic systems, [online] 90(5), pp.245–57. Available at: https://www.ncbi.nlm.nih.gov/pubmed/26781081 [Accessed 21 Mar. 2020].

154. Lee, K.-H., Shin, T.-J., Kim, W.-H. and Cho, J.-Y. (2019). Methylation of LINE-1 in cell-free DNA serves as a liquid biopsy biomarker for human breast cancers and dog mammary tumors. Scientific Reports, [online] 9(1). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6336845/ [Accessed 20 Jul. 2021].

155. Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K. and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. Genes & Development, [online] 24(3), pp.265–276. Available at: http://genesdev.cshlp.org/content/24/3/265 [Accessed 24 Jan. 2021].

156. Lei, H. (2005). Exonization of AluYa5 in the human ACE gene requires mutations in both 3’ and 5’ splice sites and is facilitated by a conserved splicing enhancer. Nucleic Acids Research, [online] 33(12), pp.3897–3906. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1175817/ [Accessed 30 Dec. 2020].

157. Linheiro, R.S. and Bergman, C.M. (2012). Whole Genome Resequencing Reveals Natural Target Site Preferences of Transposable Elements in Drosophila melanogaster. PLoS ONE, 7(2), p.e30008.

158. Lu, H., Yuan, Z., Tan, T., Wang, J., Zhang, J., Luo, H.-J., Xia, Y., Ji, W. and Gao, F. (2015). Improved tagmentation-based whole-genome bisulfite sequencing for input DNA from less than 100 mammalian cells. Epigenomics, [online] 7(1), pp.47–56. Available at: https://pubmed.ncbi.nlm.nih.gov/25687465/ [Accessed 17 Aug. 2021].

159. Lucas, A.H. (2003). Immunoglobulin Gene Construction: Human. eLS. [online] Available at: https://onlinelibrary.wiley.com/doi/full/10.1038/npg.els.0001172 [Accessed 10 Feb. 2021].

160. Luft, F.C. (2010). Sleeping Beauty jumps to new heights. Journal of Molecular Medicine, [online] 88(7), pp.641–643. Available at: https://pubmed.ncbi.nlm.nih.gov/20467721/ [Accessed 22 Mar. 2020].

161. Lynch-Sutherland, C.F., Chatterjee, A., Stockwell, P.A., Eccles, M.R. and Macaulay, E.C. (2020). Reawakening the Developmental Origins of Cancer Through Transposable Elements. Frontiers in Oncology, [online] 10. Available at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00468/full [Accessed 17 Aug. 2021].

162. Magnani, C.F., Tettamanti, S., Alberti, G., Pisani, I., Biondi, A., Serafini, M. and Gaipa, G. (2020). Transposon-Based CAR T Cells in Acute Leukemias: Where Are We Going? Cells, [online] 9(6), p.1337. Available at: https://pubmed.ncbi.nlm.nih.gov/32471151/ [Accessed 15 Aug. 2021].

163. Makałowski, W. (2001). The human genome structure and organization. Acta Biochimica Polonica, 48(3), pp.587–598.

164. Malik, H.S. (2000). Poised for Contagion: Evolutionary Origins of the Infectious Abilities of Invertebrate Retroviruses. Genome Research, [online] 10(9), pp.1307–1318. Available at: https://pubmed.ncbi.nlm.nih.gov/10984449/ [Accessed 13 Jan. 2021].

165. Malki, S., van der Heijden, G.W., O'Donnell, K.A., Martin, S.L. and Bortvin, A. (2014). A role for retrotransposon LINE-1 in fetal oocyte attrition in mice. Developmental cell, [online] 29(5), pp.521–533. Available at: https://www.ncbi.nlm.nih.gov/pubmed/24882376/ [Accessed 22 Mar. 2020].

166. Malnic, B., Godfrey, P.A. and Buck, L.B. (2004). The human olfactory receptor gene family. Proceedings of the National Academy of Sciences, [online] 101(8), pp.2584–2589. Available at: https://www.pnas.org/content/101/8/2584 [Accessed 16 Aug. 2021].

167. Malousi, A. and Kouidou, S. (2012). DNA hypermethylation of alternatively spliced and repeat sequences in humans. Molecular Genetics and Genomics, [online] 287(8), pp.631–642. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3407362/ [Accessed 23 Jan. 2021].

168. Mangiacasale, R., Pittoggi, C., Sciamanna, I., Careddu, A., Mattei, E., Lorenzini, R., Travaglini, L., Landriscina, M., Barone, C., Nervi, C.,

Lavia, P. and Spadafora, C. (2003). Exposure of normal and transformed cells to nevirapine, a reverse transcriptase inhibitor, reduces cell growth and promotes differentiation. Oncogene, [online] 22(18), pp.2750–2761. Available at: https://pubmed.ncbi.nlm.nih.gov/12747369/ [Accessed 25 Jul. 2021].

169. March, H.N., Rust, A.G., Wright, N.A., ten Hoeve, J., de Ridder, J., Eldridge, M., van der Weyden, L., Berns, A., Gadiot, J., Uren, A., Kemp, R., Arends, M.J., Wessels, L.F.A., Winton, D.J. and Adams, D.J. (2011). Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. Nature Genetics, [online] 43(12), pp.1202–1209. Available at: https://pubmed.ncbi.nlm.nih.gov/22057237/ [Accessed 15 Aug. 2021].

170. Matouk, C.C. & Marsden, P.A. (2008). Epigenetic Regulation of Vascular Endothelial Gene Expression. Circulation Research, [online] 102(8), pp.873–887. Available at: https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.107.171025?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed& [Accessed 25 Jan. 2021].

171. Matsuda, F., Ishii, K., Bourvagnet, P., Kuma, K., Hayashida, H., Miyata, T. and Honjo, T. (1998). The Complete Nucleotide Sequence of the Human Immunoglobulin Heavy Chain Variable Region Locus. Journal of Experimental Medicine, [online] 188(11), pp.2151–2162. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2212390/ [Accessed 16 Aug. 2021].

172. Majumdar, S. and Rio, D. C. (2015). P Transposable Elements in Drosophila and other Eukaryotic Organisms. In: Mobile DNA III, [online] pp.727–752. Available at:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4399808/ [Accessed 3 Feb. 2021].

173. McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences, 36(6), pp.344–355.

174. McClintock, B. (1983). The significance of responses of the genome to challenge. Science, 226(4676), pp.792–801.

175. McGrayne, S.B. and Byers, N. (1994). Nobel Prize Women in Science. Physics Today, 47(7), pp.63–63.

176. Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K., Vogelstein, B., & Nakamura2, Y. (1992). Disruption of the APC Gene by a Retrotransposal Insertion of LI Sequence in a Colon Cancer1. CANCER RESEARCH, 52, 643–645. https://cancerres.aacrjournals.org/content/52/3/643.full-text.pdf

177. Miller (1997). Molecular domestication of mobile elements. Genetica, [online] 100(1-3). Available at: https://pubmed.ncbi.nlm.nih.gov/9440279/ [Accessed 4 Aug. 2021].

178. Mills, R.E., Bennett, E.A., Iskow, R.C. and Devine, S.E. (2007). Which transposable elements are active in the human genome? Trends in Genetics, [online] 23(4), pp.183–191. Available at: https://www.cell.com/trends/genetics/fulltext/S0168-9525(07)00059-5?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0168952507000595%3Fshowall%3Dtrue [Accessed 22 Mar. 2020].

179. Miousse, I.R. and Koturbash, I. (2015). The Fine LINE: Methylation Drawing the Cancer Landscape. BioMed Research International, [online] 2015, pp.1–8. Available at: https://pubmed.ncbi.nlm.nih.gov/26448926/ [Accessed 25 Jul. 2021].

180. Miousse, I.R., Chalbot, M.-C.G., Lumen, A., Ferguson, A., Kavouras, I.G. and Koturbash, I. (2015). Response of transposable elements to environmental stressors. Mutation Research/Reviews in Mutation Research, [online] 765, pp.19–39. Available at: https://pubmed.ncbi.nlm.nih.gov/26281766/ [Accessed 20 Jan. 2021].

181. Mita, P., Wudzinska, A., Sun, X., Andrade, J., Nayak, S., Kahler, D.J., Badri, S., LaCava, J., Ueberheide, B., Yun, C.Y., Fenyö, D. and Boeke, J.D. (2018). LINE-1 protein localization and functional dynamics during the cell cycle. eLife, [online] 7. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5821460/ [Accessed 23 May 2021].

182. Monot, C., Kuciak, M., Viollet, S., Mir, A.A., Gabus, C., Darlix, J.-L. and Cristofari, G. (2013). The Specificity and Flexibility of L1 Reverse Transcription Priming at Imperfect T-Tracts. PLoS Genetics, [online] 9(5), p.e1003499. Available at: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003499 [Accessed 23 May 2021].

183. Moon, J.-Y. (2011). ACE2 and Angiotensin-(1-7) in Hypertensive Renal Disease. Electrolytes & Blood Pressure, [online] 9(2), p.41. Available at: https://e-bp.org/DOIx.php?id=10.5049/EBP.2011.9.2.41 [Accessed 28 Dec. 2020].

184. Morris, S.M., Davison, J., Carter, K.T., O'Leary, R.M., Trobridge, P., Knoblaugh, S.E., Myeroff, L.L., Markowitz, S.D., Brett, B.T., Scheetz, T.E., Dupuy, A.J., Starr, T.K. and Grady, W.M. (2016a). Transposon mutagenesis identifies candidate genes that cooperate with loss of transforming growth factor-beta signaling in mouse intestinal neoplasms. International Journal of Cancer, [online] 140(4), pp.853–863. Available at: https://pubmed.ncbi.nlm.nih.gov/27790711/ [Accessed 15 Aug. 2021].

185. Muñoz-López, M. and García-Pérez, J.L. (2010). DNA transposons: nature and applications in genomics. Current genomics, [online] 11(2), pp.115–28. Available at: https://www.ncbi.nlm.nih.gov/pubmed/20885819 [Accessed 22 Mar. 2020].

186. Mustafin, R.N. (2018). Hypothesis on the Origin of Viruses from Transposons. Molecular Genetics, Microbiology and Virology, [online] 33(4), pp.223–232. Available at: https://link.springer.com/article/10.3103/S0891416818040067#:~:text=The%20emergence%20of%20viral%20particles,new%20properties%2C%20including%20adaptive%20properties. [Accessed 10 Aug. 2021].

187. Mustafin, R.N. and Khusnutdinova, E.K. (2019). The Role of Reverse Transcriptase in the Origin of Life. Biochemistry (Moscow), [online] 84(8), pp.870–883. Available at: https://link.springer.com/article/10.1134%2FS0006297919080030 [Accessed 12 Jan. 2021].

188. Nagy, Z. and Chandler, M. (2004). Regulation of transposition in bacteria. Research in Microbiology, [online] 155(5), pp.387–398. Available at: https://www.sciencedirect.com/science/article/pii/S0923250804000592 [Accessed 10 Aug. 2021].

189. Narayanavari, S.A. (2017). Sleeping Beauty transposon vectors for therapeutic applications: advances and challenges. [online] BioInsights. Available at: https://insights.bio/cell-and-gene-therapy-insights/journal/article/434/Sleeping-Beauty-transposon-vectors-for-therapeutic-applications-advances-and-challenges [Accessed 12 Aug. 2021].

190. Naville, M., Warren, I.A., Haftek-Terreau, Z., Chalopin, D., Brunet, F., Levin, P., Galiana, D. and Volff, J.-N. . (2016). Not so bad after all: retroviruses and long terminal repeat retrotransposons as a source of new genes in vertebrates. Clinical Microbiology and Infection, [online] 22(4), pp.312–323. Available at: https://www.sciencedirect.com/science/article/pii/S1198743X16001063 [Accessed 21 Mar. 2020].

191. Nesmelova, I.V. and Hackett, P.B. (2010). DDE transposases: Structural similarity and diversity. Advanced Drug Delivery Reviews, [online] 62(12), pp.1187–1195. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0169409X10001420?via%3Dihub [Accessed 25 Jan. 2021].

192. Nguyen, T.V., Pawlikowska, P., Firlej, V., Rosselli, F. and Aoufouchi, S. (2016). V(D)J recombination process and the Pre-B to immature B-cells transition are altered in Fanca−/− mice. Scientific Reports, [online] 6(1). Available at: https://www.nature.com/articles/srep36906 [Accessed 26 Jul. 2021].

193. Nicholls, R. D., Fischel-Ghodsian, N., & Higgs, D. R. (1987). Recombination at the human α-globin gene cluster: Sequence features and topological constraints. Cell, 49(3), 369–378. https://doi.org/10.1016/0092-8674(87)90289-3

194. Niimura, Y. (2009a). Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents. Human Genomics, [online] 4(2), p.107. Available at: https://humgenomics.biomedcentral.com/articles/10.1186/1479-7364-4-2-107 [Accessed 16 Aug. 2021].

195. Nojadeh, J. N., Behrouz Sharif, S., & Sakhinia, E. (2018). Microsatellite instability in colorectal cancer. EXCLI Journal, 17, 159–168. https://doi.org/10.17179/excli2017-948

196. Noorani, I., Bradley, A. and de la Rosa, J. (2020). CRISPR and transposon in vivo screens for cancer drivers and therapeutic targets. Genome Biology, [online] 21(1). Available at: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02118-9 [Accessed 15 Aug. 2021].

197. Notarangelo, L.D., Kim, M.-S., Walter, J.E. and Lee, Y.N. (2016). Human RAG mutations: biochemistry and clinical implications. Nature Reviews Immunology, [online] 16(4), pp.234–246. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5757527/ [Accessed 8 Feb. 2021].

198. Ogino, S., Nosho, K., Kirkner, G.J., Kawasaki, T., Chan, A.T., Schernhammer, E.S., Giovannucci, E.L. and Fuchs, C.S. (2008). A Cohort Study of Tumoral LINE-1 Hypomethylation and Prognosis in Colon Cancer. JNCI: Journal of the National Cancer Institute, [online] 100(23), pp.1734–1738. Available at: https://pubmed.ncbi.nlm.nih.gov/19033568/ [Accessed 23 Jan. 2021].

199. Ostertag, E.M. and Kazazian Jr, H.H. (2001). Biology of Mammalian L1 Retrotransposons. Annual Review of Genetics, 35(1), pp.501–538.

200. Ostertag, E.M., Goodier, J.L., Zhang, Y. and Kazazian, H.H. (2003). SVA Elements Are Nonautonomous Retrotransposons that Cause Disease in Humans. The American Journal of Human Genetics, [online] 73(6), pp.1444–1451. Available at: https://www.cell.com/ajhg/fulltext/S0002-9297(07)63994-9 [Accessed 22 Mar. 2020].

201. Park, S.Y., Seo, A.N., Jung, H.Y., Gwak, J.M., Jung, N., Cho, N.-Y. and Kang, G.H. (2014). Alu and LINE-1 Hypomethylation Is Associated with HER2 Enriched Subtype of Breast Cancer. PLoS ONE, [online] 9(6), p.e100429. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0100429 [Accessed 20 Jul. 2021].

202. Patnala, R., Lee, S.-H., Dahlstrom, J.E., Ohms, S., Chen, L., Dheen, S.T. and Rangasamy, D. (2013a). Inhibition of LINE-1 retrotransposon-encoded reverse transcriptase modulates the expression of cell differentiation genes in breast cancer cells. Breast Cancer Research and Treatment, [online] 143(2), pp.239–253. Available at: https://link.springer.com/article/10.1007/s10549-013-2812-7 [Accessed 23 Jul. 2021].

203. Pavlicek, A. and Jurka, J. (2021). Ancient Transposable Elements, Processed Pseudogenes, and Endogenous Retroviruses. Genomic Disorders, [online] pp.57–72. Available at: https://link.springer.com/chapter/10.1007/978-1-59745-039-3_4 [Accessed 16 Aug. 2021].

204. Philosophical Transactions of the Royal Society B. (2020). Crossroads between transposons and gene regulation | Philosophical Transactions of the Royal Society B: Biological Sciences. [online] Available at: https://royalsocietypublishing.org/doi/10.1098/rstb.2019.0330 [Accessed 26 Jul. 2021].

205. Picelli, S., Björklund, Å.K., Reinius, B., Sagasser, S., Winberg, G. and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Research, [online] 24(12), pp.2033–2040. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4248319/ [Accessed 17 Aug. 2021].

206. Piovan, C., Amari, F., Lovat, F., Chen, Q. and Coppola, V. (2014). Generation of Mouse Lines Conditionally Over-expressing MicroRNA Using the Rosa26-Lox-Stop-Lox System. Methods in Molecular Biology, [online] pp.203–224. Available at: https://pubmed.ncbi.nlm.nih.gov/25064105/ [Accessed 15 Aug. 2021].

207. Pourrajab, F. and Hekmatimoghaddam, S. (2021). Transposable elements, contributors in the evolution of organisms (from an arms race to a source of raw materials). Heliyon, [online] 7(1), p.e06029. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7829209/ [Accessed 10 Aug. 2021].

208. Pox, C.P., Altenhofen, L., Brenner, H., Theilmeier, A., Stillfried, D.V. and Schmiegel, W. (2012). Efficacy of a Nationwide Screening Colonoscopy Program for Colorectal Cancer. Gastroenterology, [online] 142(7), pp.1460-1467.e2. Available at: https://pubmed.ncbi.nlm.nih.gov/22446606/ [Accessed 20 Jul. 2021].

209. Prabhakar, P., De, T., Nagaraja, D. and Christopher, R. (2014). Angiotensin-Converting Enzyme Gene Insertion/Deletion Polymorphism and Small Vessel Cerebral Stroke in Indian Population. International Journal of Vascular Medicine, [online] 2014, pp.1–4. Available at: https://www.hindawi.com/journals/ijvm/2014/305309/ [Accessed 29 Dec. 2020].

210. Quesneville, H., Bergman, C.M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M. and Anxolabehere, D. (2005). Combined Evidence Annotation of Transposable Elements in Genome Sequences. PLoS Computational Biology, [online] 1(2), p.e22. Available at: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0010022#s1 [Accessed 1 Aug. 2021].

211. Rad, R., Rad, L., Wang, W., Cadinanos, J., Vassiliou, G., Rice, S., Campos, L.S., Yusa, K., Banerjee, R., Li, M.A., de la Rosa, J., Strong, A., Lu, D., Ellis, P., Conte, N., Yang, F.T., Liu, P. and Bradley, A. (2010). PiggyBac Transposon Mutagenesis: A Tool for Cancer Gene Discovery in Mice. Science, [online] 330(6007), pp.1104–1107. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3719098/ [Accessed 15 Aug. 2021].

212. Rauch, T., Li, H., Wu, X. and Pfeifer, G.P. (2006). MIRA-Assisted Microarray Analysis, a New Technology for the Determination of DNA Methylation Patterns, Identifies Frequent Methylation of Homeodomain-Containing Genes in Lung Cancer Cells. Cancer Research, [online] 66(16), pp.7939–7947. Available at: https://cancerres.aacrjournals.org/content/66/16/7939 [Accessed 20 Jul. 2021].

213. Ravindran, S. (2012). Barbara McClintock and the discovery of jumping genes. Proceedings of the National Academy of Sciences, [online] 109(50), pp.20198–20199. Available at: https://www.pnas.org/content/109/50/20198 [Accessed 21 Mar. 2020].

214. Reilly, M.T., Faulkner, G.J., Dubnau, J., Ponomarev, I. and Gage, F.H. (2013). The Role of Transposable Elements in Health and Diseases of the Central Nervous System. Journal of Neuroscience, [online] 33(45), pp.17577–17586. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3818539/ [Accessed 22 Mar. 2020].

215. Reznikoff, W.S. (2003). Tn5 as a model for understanding DNA transposition. Molecular Microbiology, [online] 47(5), pp.1199–1206. Available at: https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-2958.2003.03382.x [Accessed 24 Jan. 2021].

216. Reznikoff, W.S., Bhasin, A., Davies, D.R., Goryshin, I.Y., Mahnke, L.A., Naumann, T., Rayment, I., Steiniger-White, M. and Twining, S.S. (1999). Tn5: A Molecular Window on Transposition. Biochemical and Biophysical Research Communications, [online] 266(3), pp.729–734. Available at: https://www.sciencedirect.com/science/article/pii/S0006291X99918910?via%3Dihub [Accessed 25 Jan. 2021].

217. Richardson, S. R., Doucet, A. J., Kopera, H. C., Moldovan, J. B., Garcia-Perez, J. L., & Moran, J. V. (2015). The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. Microbiology spectrum, 3(2), MDNA3–2014. https://doi.org/10.1128/microbiolspec.MDNA3-0061-2014

218. Rigat, B., Hubert, C., Alhenc-Gelas, F., Cambien, F., Corvol, P. and Soubrier, F. (1990). An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. Journal of Clinical Investigation, [online] 86(4), pp.1343–1346. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC296868/ [Accessed 26 Dec. 2020].

219. Roberts, T.C. (2015). The microRNA Machinery. microRNA: Basic Science, [online] pp.15–30. Available at: https://link.springer.com/chapter/10.1007%2F978-3-319-22380-3_2 [Accessed 3 Feb. 2021].

220. Rodgers, K.K. (2017). Riches in RAGs: Revealing the V(D)J Recombinase through High-Resolution Structures. [online] undefined. Available at: https://www.semanticscholar.org/paper/Riches-in-RAGs%3A-Revealing-the-V(D)J-Recombinase-Rodgers/0c08af7431bf2b1c8428a7a67abd5bbf27079c9f [Accessed 8 Feb. 2021].

221. Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M.S., Hruban, R.H., Iacobuzio-Donahue, C.A., Maitra, A., Torbenson, M.S., Goggins, M., Shih, I.-M., Duffield, A.S., Montgomery, E.A., Gabrielson, E., Netto, G.J., Lotan, T.L., De Marzo, A.M., Westra, W. and Binder, Z.A. (2014). Long Interspersed Element-1 Protein Expression Is a Hallmark of Many Human Cancers. The American Journal of Pathology, [online] 184(5), pp.1280–1286. Available at: https://ajp.amjpathol.org/article/S0002-9440(14)00077-7/fulltext [Accessed 16 Jul. 2021].

222. Roman-Gomez, J., Jimenez-Velasco, A., Agirre, X., Cervantes, F., Sanchez, J., Garate, L., Barrios, M., Castillejo, J. A., Navarro, G., Colomer, D., Prosper, F., Heiniger, A., & Torres, A. (2005). Promoter hypomethylation of the LINE-1 retrotransposable elements activates sense/antisense transcription and marks the progression of chronic myeloid leukemia. Oncogene, 24(48), 7213–7223. https://doi.org/10.1038/sj.onc.1208866

223. Rostovskaya, M., Fu, J., Obst, M., Baer, I., Weidlich, S., Wang, H., Smith, A.J.H., Anastassiadis, K. and Stewart, A.F. (2012). Transposon-mediated BAC transgenesis in human ES cells. Nucleic Acids Research, [online] 40(19), pp.e150–e150. Available at: https://academic.oup.com/nar/article/40/19/e150/2414550 [Accessed 12 Aug. 2021].

224. Roth, D.B. (2014). V(D)J Recombination: Mechanism, Errors, and Fidelity. Microbiology Spectrum, [online] 2(6). Available at: https://pubmed.ncbi.nlm.nih.gov/26104458/ [Accessed 26 Jul. 2021].

225. Roth, D.B. (2015). V(D)J Recombination: Mechanism, Errors, and Fidelity. Mobile DNA III, [online] pp.313–324. Available at: https://pubmed.ncbi.nlm.nih.gov/26104458/ [Accessed 6 Feb. 2021].

226. Roy-Engel, A.M., Carroll, M.L., El-Sawy, M., Salem, A.-H., Garber, R.K., Nguyen, S.V., Deininger, P.L. and Batzer, M.A. (2002). Non-traditional Alu evolution and primate genomic diversity. Journal of Molecular Biology, 316(5), pp.1033–1040.

227. Ru, H., Chambers, M.G., Fu, T.-M., Tong, A.B., Liao, M. and Wu, H. (2015). Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. Cell, [online] 163(5), pp.1138–1152. Available at: https://pubmed.ncbi.nlm.nih.gov/26548953/ [Accessed 8 Feb. 2021].

228. Ru, H., Zhang, P. and Wu, H. (2018). Structural gymnastics of RAG-mediated DNA cleavage in V(D)J recombination. Current Opinion in Structural Biology, [online] 53, pp.178–186. Available at: https://pubmed.ncbi.nlm.nih.gov/30476719/ [Accessed 8 Feb. 2021].

229. Russell, S.J. and LaMarre, J. (2018). Transposons and the PIWI Pathway – Genome Defence in Gametes and Embryos. Reproduction. [online] Available at: https://rep.bioscientifica.com/view/journals/rep/156/4/REP-18-0218.xml [Accessed 24 Jan. 2021].

230. Saito, K., Kawakami, K., Matsumoto, I., Oda, M., Watanabe, G. and Minamoto, T. (2010). Long Interspersed Nuclear Element 1 Hypomethylation Is a Marker of Poor Prognosis in Stage IA Non–Small Cell Lung Cancer. Clinical Cancer Research, [online] 16(8), pp.2418–2426. Available at: https://clincancerres.aacrjournals.org/content/16/8/2418 [Accessed 15 Jul. 2021].

231. Sama, I.E. and Voors, A.A. (2020). Men more vulnerable to COVID-19: explained by ACE2 on the X chromosome? European Heart Journal, [online] 41(32), pp.3096–3096. Available at:

https://academic.oup.com/eurheartj/article/41/32/3096/5861742 [Accessed 28 Dec. 2020].

232. Sama, I.E., Ravera, A., Santema, B.T., van Goor, H., ter Maaten, J.M., Cleland, J.G.F., Rienstra, M., Friedrich, A.W., Samani, N.J., Ng, L.L., Dickstein, K., Lang, C.C., Filippatos, G., Anker, S.D., Ponikowski, P., Metra, M., van Veldhuisen, D.J. and Voors, A.A. (2020). Circulating plasma concentrations of angiotensin-converting enzyme 2 in men and women with heart failure and effects of renin–angiotensin–aldosterone inhibitors. European Heart Journal, [online] 41(19), pp.1810–1817. Available at: https://academic.oup.com/eurheartj/article/41/19/1810/5834647 [Accessed 28 Dec. 2020].

233. Sandoval-Villegas, N., Nurieva, W., Amberger, M. and Ivics, Z. (2021). Contemporary Transposon Tools: A Review and Guide through Mechanisms and Applications of Sleeping Beauty, piggyBac and Tol2 for Genome Engineering. International Journal of Molecular Sciences, [online] 22(10), p.5084. Available at: https://www.mdpi.com/1422-0067/22/10/5084 [Accessed 15 Aug. 2021].

234. Sawyers, C.L. (2008). The cancer biomarker problem. Nature, [online] 452(7187), pp.548–552. Available at: https://pubmed.ncbi.nlm.nih.gov/18385728/ [Accessed 16 Jul. 2021].

235. Schaaf, C. and Johannes Zschocke (2018). Basiswissen Humangenetik. Berlin: Berlin Springer.

236. Schatz, D.G. and Swanson, P.C. (2011). V(D)J Recombination: Mechanisms of Initiation. Annual Review of Genetics, [online] 45(1), pp.167–202. Available at: https://pubmed.ncbi.nlm.nih.gov/21854230/ [Accessed 6 Feb. 2021].

237. Schatz, D.G., Oettinger, M.A. and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. Cell, [online] 59(6), pp.1035–1048. Available at: https://pubmed.ncbi.nlm.nih.gov/2598259/ [Accessed 16 Aug. 2021].

238. Schmid, C.W. and Deininger, P.L. (1975). Sequence organization of the human genome. Cell, 6(3), pp.345–358.

239. Schneider AM;Duffield AS;Symer DE;Burns KH (2014). Roles of retrotransposons in benign and malignant hematologic disease. Cellscience, [online] 6(2). Available at: https://pubmed.ncbi.nlm.nih.gov/20200603/ [Accessed 29 Jun. 2021].

240. Schneider, A.M., Duffield, A.S., Symer, D.E. and Burns, K.H. (2009). Roles of retrotransposons in benign and malignant hematologic disease. Cellscience, [online] 6(2), pp.121–145. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2830787/ [Accessed 29 Jun. 2021].

241. Schorn, A.J. and Martienssen, R. (2018). Tie-Break: Host and Retrotransposons Play tRNA. Trends in Cell Biology, [online] 28(10), pp.793–806. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6520983/ [Accessed 2 Feb. 2021].

242. Schroeder, H.W. and Cavacini, L. (2010a). Structure and function of immunoglobulins. Journal of Allergy and Clinical Immunology, [online] 125(2), pp.S41–S52. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3670108/ [Accessed 10 Feb. 2021].

243. Sciamanna, I., De Luca, C. and Spadafora, C. (2016a). The Reverse Transcriptase Encoded by LINE-1 Retrotransposons in the Genesis,

Progression, and Therapy of Cancer. Frontiers in Chemistry, [online] 4. Available at: https://www.frontiersin.org/articles/10.3389/fchem.2016.00006/full [Accessed 18 Feb. 2021].

244. Scott, E. C., Gardner, E. J., Masood, A., Chuang, N. T., Vertino, P. M., & Devine, S. E. (2016). A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Research, 26(6), 745–755. https://doi.org/10.1101/gr.201814.115

245. Sela, N., Mersch, B., Hotz-Wagenblatt, A. and Ast, G. (2010). Characteristics of Transposable Element Exonization within Human and Mouse. PLoS ONE, [online] 5(6), p.e10907. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0010907 [Accessed 16 Aug. 2021].

246. Serrato-Capuchina, A. and Matute, D. (2018). The Role of Transposable Elements in Speciation. Genes, [online] 9(5), p.254. Available at: https://www.mdpi.com/2073-4425/9/5/254/htm [Accessed 1 Feb. 2021].

247. Shapiro, J.A. (2014). Epigenetic control of mobile DNA as an interface between experience and genome change. Frontiers in Genetics, [online] 5. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4007016/ [Accessed 16 Aug. 2021].

248. Singer, M.F. (1982). SINEs and LINEs: Highly repeated short and long interspersed sequences in mammalian genomes. Cell, 28(3), pp.433–434.

249. Skowronski, J. and Singer, M.F. (1986). The Abundant LINE-1 Family of Repeated DNA Sequences in Mammals: Genes and Pseudogenes. Cold

Spring Harbor Symposia on Quantitative Biology, [online] 51(0), pp.457–464. Available at: https://pubmed.ncbi.nlm.nih.gov/3472735/ [Accessed 29 Jun. 2021].

250. Smith, T.M., Lee, M.K., Szabo, C.I., Jerome, N., McEuen, M., Taylor, M., Hood, L. and King, M.C. (1996). Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1. Genome Research, [online] 6(11), pp.1029–1049. Available at: https://pubmed.ncbi.nlm.nih.gov/8938427/ [Accessed 30 Dec. 2020].

251. Sorek, R. (2002). Alu-Containing Exons are Alternatively Spliced. Genome Research, [online] 12(7), pp.1060–1067. Available at: https://pubmed.ncbi.nlm.nih.gov/12097342/ [Accessed 23 Jan. 2021].

252. Sorek, R., Lev-Maor, G., Reznik, M., Dagan, T., Belinky, F., Graur, D. and Ast, G. (2004). Minimal Conditions for Exonization of Intronic Sequences. Molecular Cell, [online] 14(2), pp.221–231. Available at: https://www.cell.com/molecular-cell/fulltext/S1097-2765(04)00181-9?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1097276504001819%3Fshowall%3Dtrue [Accessed 23 Jan. 2021].

253. Starr, T.K., Allaei, R., Silverstein, K.A.T., Staggs, R.A., Sarver, A.L., Bergemann, T.L., Gupta, M., O'Sullivan, M.G., Matise, I., Dupuy, A.J., Collier, L.S., Powers, S., Oberg, A.L., Asmann, Y.W., Thibodeau, S.N., Tessarollo, L., Copeland, N.G., Jenkins, N.A., Cormier, R.T. and Largaespada, D.A. (2009). A Transposon-Based Genetic Screen in Mice Identifies Genes Altered in Colorectal Cancer. Science, [online] 323(5922), pp.1747–1750. Available at: https://pubmed.ncbi.nlm.nih.gov/19251594/ [Accessed 15 Aug. 2021].

254. Starr, T.K., Jameson, S.C. and Hogquist, K.A. (2003). POSITIVE ANDNEGATIVESELECTION OFT CELLS. Annual Review of Immunology,

[online] 21(1), pp.139–176. Available at: https://pubmed.ncbi.nlm.nih.gov/12414722/ [Accessed 25 Jul. 2021].

255. Su, L., Kinzler, K., Vogelstein, B., Preisinger, A., Moser, A., Luongo, C., Gould, K. and Dove, W. (1992). Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. Science, [online] 256(5057), pp.668–670. Available at: https://pubmed.ncbi.nlm.nih.gov/1350108/ [Accessed 15 Aug. 2021].

256. Su, Y.-Q., Sugiura, K., Sun, F., Pendola, J.K., Cox, G.A., Handel, M.A., Schimenti, J.C. and Eppig, J.J. (2012). MARF1 regulates essential oogenic processes in mice. Science (New York, N.Y.), [online] 335(6075), pp.1496–9. Available at: https://www.ncbi.nlm.nih.gov/pubmed/22442484/ [Accessed 22 Mar. 2020].

257. Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Hsi-Yang Fritz, M., Konkel, M. K., Malhotra, A., Stütz, A. M., Shi, X., Paolo Casale, F., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., & Malig, M. (2015). An integrated map of structural variation in 2,504 human genomes. Nature, 526(7571), 75–81. https://doi.org/10.1038/nature15394

258. Sultana, T., van Essen, D., Siol, O., Bailly-Bechet, M., Philippe, C., Zine El Aabidine, A., Pioger, L., Nigumann, P., Saccani, S., Andrau, J.-C., Gilbert, N. and Cristofari, G. (2019). The Landscape of L1 Retrotransposons in the Human Genome Is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection. Molecular Cell, [online] 74(3), pp.555-570.e7. Available at: https://www.sciencedirect.com/science/article/pii/S1097276519301479#:~:text=L1%20retrotransposons%20are%20transposable%20elements,of%20genetic%20variation%20in%20humans.&text=At%20regional%20scales

%2C%20L1%20shows,other%20known%20mobile%20genetic%20elements. [Accessed 23 May 2021].

259. Sun, W., Samimi, H., Gamez, M., Zare, H. and Frost, B. (2018). Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. Nature Neuroscience, [online] 21(8), pp.1038–1048. Available at: https://pubmed.ncbi.nlm.nih.gov/30038280/ [Accessed 22 Mar. 2020].

260. Takeda, H., Jenkins, N.A. and Copeland, N.G. (2021). Identification of cancer driver genes using Sleeping Beauty transposon mutagenesis. Cancer Science, [online] 112(6), pp.2089–2096. Available at: https://onlinelibrary.wiley.com/doi/10.1111/cas.14901#cas14901-bib-0005 [Accessed 15 Aug. 2021].

261. Talbert, P.B. and Henikoff, S. (2020). What makes a centromere? Experimental Cell Research, [online] 389(2), p.111895. Available at: https://www.sciencedirect.com/science/article/pii/S001448272030094X [Accessed 16 Aug. 2021].

262. Temin, H.M. and Mizutani, S. (1970). Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. Nature, [online] 226(5252), pp.1211–1213. Available at: https://www.nature.com/articles/2261211a0 [Accessed 12 Jan. 2021].

263. Thayer, R.E., Singer, M.F. and Fanning, T.G. (1993). Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1 -encoded protein. Gene, [online] 133(2), pp.273–277. Available at: https://www.sciencedirect.com/science/article/abs/pii/037811199390651I?via%3Dihub [Accessed 15 Jul. 2021].

264. Tipanee, J., Chai, Y., VandenDriessche, T. and Chuah, Marinee K. (2017). Preclinical and clinical advances in transposon-based gene therapy. Bioscience Reports, [online] 37(6). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5715130/ [Accessed 12 Aug. 2021].

265. Ueffing, M., Bayyoud, T., Schindler, M. and Ziemssen, F. (2020). Grundlagen der Replikation und der Immunologie von SARS-CoV-2. Der Ophthalmologe, [online] 117(7), pp.609–614. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7328300/ [Accessed 30 Dec. 2020].

266. Ule, J. (2013). Alu elements: at the crossroads between disease and evolution. Biochemical Society Transactions, [online] 41(6), pp.1532–1535. Available at: https://pubmed.ncbi.nlm.nih.gov/24256249/ [Accessed 16 Aug. 2021].

267. Ullu, E. and Tschudi, C. (1984). Alu sequences are processed 7SL RNA genes. Nature, [online] 312(5990), pp.171–172. Available at: https://ui.adsabs.harvard.edu/abs/1984Natur.312..171U/abstract [Accessed 29 Dec. 2020].

268. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. and Thiel, V. (2020). Coronavirus biology and replication: implications for SARS-CoV-2. Nature Reviews Microbiology. [online] Available at: https://www.nature.com/articles/s41579-020-00468-6 [Accessed 30 Dec. 2020].

269. Van Gent, D.C., Mizuuchi, K. and Gellert, M. (1996a). Similarities Between Initiation of V(D)J Recombination and Retroviral Integration. Science, [online] 271(5255), pp.1592–1594. Available at: https://science.sciencemag.org/content/271/5255/1592?ijkey=d77939f

59e93daaee6a767683c52549649a646f0&keytype2=tf_ipsecsha [Accessed 16 Aug. 2021].

270. Van Hoesel, A.Q., van de Velde, C.J.H., Kuppen, P.J.K., Liefers, G.J., Putter, H., Sato, Y., Elashoff, D.A., Turner, R.R., Shamonki, J.M., de Kruijf, E.M., van Nes, J.G.H., Giuliano, A.E. and Hoon, D.S.B. (2012). Hypomethylation of LINE-1 in primary tumor has poor prognosis in young breast cancer patients: a retrospective cohort study. Breast Cancer Research and Treatment, [online] 134(3), pp.1103–1114. Available at: https://link.springer.com/article/10.1007/s10549-012-2038-0 [Accessed 20 Jul. 2021].

271. Varon, R., Gooding, R., Steglich, C., Marns, L., Tang, H., Angelicheva, D., Yong, K.K., Ambrugger, P., Reinhold, A., Morar, B., Baas, F., Kwa, M., Tournev, I., Guerguelcheva, V., Kremensky, I., Lochmüller, H., Müllner-Eidenböck, A., Merlini, L., Neumann, L., Bürger, J., Walter, M., Swoboda, K., Thomas, P.K., von Moers, A., Risch, N. and Kalaydjieva, L. (2003). Partial deficiency of the C-terminal-domain phosphatase of RNA polymerase II is associated with congenital cataracts facial dysmorphism neuropathy syndrome. Nature Genetics, [online] 35(2), pp.185–189. Available at: https://delmbsgqukah2ms4rqeonbcu3i--pubmed-ncbi-nlm-nih-gov.translate.goog/14517542/ [Accessed 30 Dec. 2020].

272. Vaux, D.L. (2011). In defense of the somatic mutation theory of cancer. BioEssays, [online] 33(5), pp.341–343. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/bies.201100022 [Accessed 16 Aug. 2021].

273. Vidal, E., Sayols, S., Moran, S., Guillaumet-Adkins, A., Schroeder, M.P., Royo, R., Orozco, M., Gut, M., Gut, I., Lopez-Bigas, N., Heyn, H. and Esteller, M. (2017). A DNA methylation map of human cancer at single base-pair resolution. Oncogene, [online] 36(40), pp.5648–5657. Available at: https://www-nature-

com.translate.goog/articles/onc2017176?error=cookies_not_supported&code=49d2198d-8d7d-4c1f-a82c-24f9c250be56&_x_tr_sl=auto&_x_tr_tl=de&_x_tr_hl=de&_x_tr_pto=ajax,se,elem [Accessed 17 Aug. 2021].

274. Vigdal, T.J., Kaufman, C.D., Izsvák, Z., Voytas, D.F. and Ivics, Z. (2002). Common Physical Properties of DNA Affecting Target Site Selection of Sleeping Beauty and other Tc1/mariner Transposable Elements. Journal of Molecular Biology, [online] 323(3), pp.441–452. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0022283602009919?via%3Dihub [Accessed 15 Aug. 2021].

275. Viollet, S., Monot, C. and Cristofari, G. (2014). L1 retrotransposition. Mobile Genetic Elements, [online] 4(2), p.e28907. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4014453/ [Accessed 22 May 2021].

276. Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., & Kinzler, K. W. (2013). Cancer Genome Landscapes. Science, 339(6127), 1546–1558. https://doi.org/10.1126/science.1235122

277. Voigt, F., Wiedemann, L., Zuliani, C., Querques, I., Sebe, A., Mátés, L., Izsvák, Z., Ivics, Z. and Barabas, O. (2016). Sleeping Beauty transposase structure allows rational design of hyperactive variants for genetic engineering. Nature Communications, [online] 7(1). Available at: https://www.nature.com/articles/ncomms11126 [Accessed 12 Aug. 2021].

278. Walisko, O., Schorn, A., Rolfs, F., Devaraj, A., Miskey, C., Izsvák, Z. and Ivics, Z. (2008). Transcriptional Activities of the Sleeping Beauty Transposon and Shielding Its Genetic Cargo With Insulators. Molecular Therapy, [online] 16(2), pp.359–369. Available at:

https://www.cell.com/molecular-therapy-family/molecular-therapy/fulltext/S1525-0016(16)31403-4 [Accessed 12 Aug. 2021].

279. Walter, M., Teissandier, A., Pérez-Palacios, R. and Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. eLife, [online] 5. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4769179/ [Accessed 28 Jul. 2020].

280. Wang, L. and Jordan, I.K. (2018). Transposable element activity, genome regulation and human health. Current Opinion in Genetics & Development, [online] 49, pp.25–33. Available at: https://pubmed.ncbi.nlm.nih.gov/29505964/ [Accessed 23 May 2021].

281. Wang, P.J. (2017). Tracking LINE1 retrotransposition in the germline. Proceedings of the National Academy of Sciences, [online] 114(28), pp.7194–7196. Available at: https://pubmed.ncbi.nlm.nih.gov/28663337/ [Accessed 22 Mar. 2020].

282. Waring, M. and Britten, R.J. (1966). Nucleotide Sequence Repetition: A Rapidly Reassociating Fraction of Mouse DNA. Science, 154(3750), pp.791–794.

283. Weinberg, R.A. (1994). Oncogenes and tumor suppressor genes. CA: A Cancer Journal for Clinicians, [online] 44(3), pp.160–170. Available at: https://pubmed.ncbi.nlm.nih.gov/7621068/ [Accessed 16 Aug. 2021].

284. Wiesing, U. (2019). Theranostics: is it really a revolution? Evaluating a new term in medicine. Medicine, Health Care and Philosophy, [online] 22(4), pp.593–597. Available at: https://pubmed.ncbi.nlm.nih.gov/30976948/ [Accessed 25 Jul. 2021].

285. Wilber, A., Linehan, J.L., Tian, X., Woll, P.S., Morris, J.K., Belur, L.R., McIvor, R.S. and Kaufman, D.S. (2007). Efficient and Stable Transgene Expression in Human Embryonic Stem Cells Using Transposon-Mediated Gene Transfer. Stem Cells, [online] 25(11), pp.2919–2927. Available at: https://pubmed.ncbi.nlm.nih.gov/17673526/ [Accessed 12 Aug. 2021].

286. Wolff, E.M., Byun, H.-M., Han, H.F., Sharma, S., Nichols, P.W., Siegmund, K.D., Yang, A.S., Jones, P.A. and Liang, G. (2010). Hypomethylation of a LINE-1 Promoter Activates an Alternate Transcript of the MET Oncogene in Bladders with Cancer. PLoS Genetics, [online] 6(4), p.e1000917. Available at: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000917 [Accessed 18 Feb. 2021].

287. Xiong, Z. and Laird, P.W. (1997a). COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Research, [online] 25(12), pp.2532–2534. Available at: https://academic.oup.com/nar/article/25/12/2532/2901842 [Accessed 20 Jul. 2021].

288. Yang, N. (2003). An important role for RUNX3 in human L1 transcription and retrotransposition. Nucleic Acids Research, [online] 31(16), pp.4929–4940. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC169909/ [Accessed 23 May 2021].

289. Yant, S.R., Wu, X., Huang, Y., Garrison, B., Burgess, S.M. and Kay, M.A. (2005). High-Resolution Genome-Wide Mapping of Transposon Integration in Mammals. Molecular and Cellular Biology, [online] 25(6), pp.2085–2094. Available at: https://journals.asm.org/doi/10.1128/MCB.25.6.2085-2094.2005 [Accessed 12 Aug. 2021].

290. Yin, F.F., Bailey, S., Innis, C.A., Ciubotaru, M., Kamtekar, S., Steitz, T.A. and Schatz, D.G. (2009). Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. Nature Structural & Molecular Biology, [online] 16(5), pp.499–508. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2715281/ [Accessed 8 Feb. 2021].

291. Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends in Genetics, [online] 13(8), pp.335–340. Available at: https://www.cell.com/trends/genetics/pdf/S0168-9525(97)01181-5.pdf?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0168952597011815%3Fshowall%3Dtrue.

292. Yulug, I.G., Yulug, A. and Fisher, E.M.C. (1995). The Frequency and Position of Alu Repeats in cDNAs, as Determined by Database Searching. Genomics, [online] 27(3), pp.544–548. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0888754385710907?via%3Dihub [Accessed 30 Dec. 2020].

293. Zeng, L., Pederson, S.M., Kortschak, R.Daniel. and Adelson, D.L. (2018). Transposable elements and gene expression during the evolution of amniotes. Mobile DNA, [online] 9(1). Available at: https://mobilednajournal.biomedcentral.com/articles/10.1186/s13100-018-0124-5 [Accessed 26 Jul. 2021].

294. Zhang, R., Zhang, F., Sun, Z., Liu, P., Zhang, X., Ye, Y., Cai, B., Walsh, M.J., Ren, X., Hao, X., Zhang, W. and Yu, J. (2019a). LINE-1 Retrotransposition Promotes the Development and Progression of Lung Squamous Cell Carcinoma by Disrupting the Tumor-Suppressor Gene FGGY. Cancer Research, [online] 79(17), pp.4453–4465. Available at: https://cancerres.aacrjournals.org/content/79/17/4453.long [Accessed 22 Jul. 2021].

295. Zhang, X., Zhang, R. and Yu, J. (2020). New Understanding of the Relevant Role of LINE-1 Retrotransposition in Human Disease and Immune Modulation. Frontiers in Cell and Developmental Biology, [online] 8. Available at: https://www.frontiersin.org/articles/10.3389/fcell.2020.00657/full [Accessed 15 Apr. 2021].

296. Zhang, Y., Cheng, T.C., Huang, G., Lu, Q., Surleac, M.D., Mandell, J.D., Pontarotti, P., Petrescu, A.J., Xu, A., Xiong, Y. and Schatz, D.G. (2019b). Transposon molecular domestication and the evolution of the RAG recombinase. Nature, [online] 569(7754), pp.79–84. Available at: https://pubmed.ncbi.nlm.nih.gov/30971819/ [Accessed 25 Jul. 2021].

297. Zheng, H. and Cao, J.J. (2020). Angiotensin-Converting Enzyme Gene Polymorphism and Severe Lung Injury in Patients with Coronavirus Disease 2019. The American Journal of Pathology, [online] 190(10), pp.2013–2017. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7387924/ [Accessed 26 Dec. 2020].

298. Zhou, W., Liang, G., Molloy, P.L. and Jones, P.A. (2020). DNA methylation enables transposable element-driven genome expansion. Proceedings of the National Academy of Sciences, [online] 117(32), pp.19359–19366. Available at: https://www.pnas.org/content/117/32/19359 [Accessed 20 Jan. 2021].

299. Zhu, H., Geiman, T. M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., Li, E., & Muegge, K. (2006). Lsh is involved in de novo methylation of DNA. The EMBO Journal, 25(2), 335–345. https://doi.org/10.1038/sj.emboj.7600925

300. Zipeto, D., Palmeira, J. da F., Argañaraz, G.A. and Argañaraz, E.R. (2020). ACE2/ADAM17/TMPRSS2 Interplay May Be the Main Risk Factor

for COVID-19. Frontiers in Immunology, [online] 11. Available at: https://www.frontiersin.org/articles/10.3389/fimmu.2020.576745/full [Accessed 27 Dec. 2020].