INTERNATIONAL DOCTORAL THESIS

UNIVERSIDAD
DE GRANADA

New methodologies based on nanotechnology to prepare libraries for massive sequencing: T cell receptor (TCR) and small RNA.

Memoria presentada por **Anaïs Redruello Romero** para optar a la mención de Doctora por la Universidad de Granada

Programa de Doctorado en Bioquímica y Biología Molecular

Director:
Ángel Carazo Gallego

Granada, 2024

Editor: Universidad de Granada. Tesis Doctorales
Autor: Anäis Redruello Romero
ISBN: 978-84-1195-923-0
URI: https://hdl.handle.net/10481/108784

Ángel Carazo Gallego, PhD from the University of Granada, senior researcher at the Biosanitary Research Institute of Granada, and member of the Nicolás Monardes Program of the Andalusian Health Service, as the director of the doctoral thesis; and **Anaïs Redruello Romero**, as the doctoral candidate, certify by signing that the work presented in this doctoral thesis entitled: **"New methodologies based on nanotechnology to prepare libraries for massive sequencing: T cell receptor (TCR) and small RNA"**, has been carried out by the doctoral candidate under the guidance of the director, to qualify for the Doctoral degree from the University of Granada. The rights of authorship have been respected in its execution by citing other results or publications.

Ángel Carazo Gallego, doctor por la Universidad de Granada, investigador sénior en el Instituto de Investigación Biosanitaria de Granada y miembro del Programa Nicolás Monardes del Servicio Andaluz de Salud, como director de la tesis doctoral; y **Anaïs Redruello Romero**, como doctoranda, certifican al firmar que el trabajo recogido en esta memoria de tesis doctoral titulada: **"New methodologies based on nanotechnology to prepare libraries for massive sequencing: T cell receptor (TCR) and small RNA"**, ha sido realizado por la doctoranda bajo la dirección del director para optar al grado de Doctora por la Universidad de Granada. Para su realización se han respetado los derechos de autoría al ser citados otros resultados o publicaciones.

Director: Doctoranda:

Ángel Carazo Gallego Anaïs Redruello Romero

The completion of this doctoral thesis has been possible thanks to the funding obtained through projects PI15/01361 and DTS22/00151 from the Carlos III Health Institute-FEDER, and project PIP-0236-2021 from the Ministry of Health and Families of the Andalusian Regional Government.

La realización de esta tesis doctoral ha sido posible gracias a la financiación obtenida a través de los proyectos PI15/01361 y DTS22/00151 del Instituto de Salud Carlos III-FEDER, y el proyecto PIP-0236-2021de la Consejería de Salud y Familias de la Junta de Andalucía.

ibs.GRANADA
INSTITUTO DE INVESTIGACIÓN BIOSANITARIA

HOSPITAL UNIVERSITARIO CLÍNICO SAN CECILIO

JUNTA DE ANDALUCIA
CONSEJERÍA DE SALUD Y FAMILIAS

MINISTERIO DE CIENCIA E INNOVACIÓN

Instituto de Salud Carlos III

UNIVERSITAS GRANATENSIS
PLUS ULTRA
1531

Index

Abbreviations

Ago: argonaute

APC: Allophycocyanin

APC: antigen-presenting cells

arm-PCR: amplicon rescued multiplex-polymerase chain reaction

BB: Brilliant Blue

BLAST: Basic Local Alignment Search Tool

bp: base pair

BV: Brilliant Violet

CC: click chemistry

CCS: circular consensus sequences

cDNA: complementary DNA

CLR: single continuous long read

CMV: cytomegalovirus

C_T: cycle threshold

DGCR8: DiGeourge critical region 8

DMSO: dimethyl sulfoxide

DNA: deoxyribonucleic acid

dNTPs: deoxynucleotide triphosphates

dsDNA: double-stranded DNA

EBV: Epstein-Barr Virus

FBS: Fetal Bovine Serum

FITC: fluorescein isothiocyanate

gDNA: genomic DNA

HIV: human immunodeficiency virus

HLA: Human Leukocyte Antigen

IE1: immediate-early protein 1

IPD: interpulse duration

IPTG: isopropil-β-D-1-tiogalactopiranósido

kb: kilobase

KO: knock-out

LRS: long-read sequencing

mCMV: murine cytomegalovirus

MHC: major histocompatibility complex

miRNA: microRNA

mRNA: messenger RNA

NCBI: National Center for Biotechnology Information

NGS: Next-Generation Sequencing

ONT: Oxford Nanopore Technologies

ORF: open reading frame

oWAT: omental white adipose tissue

PacBio: Pacific Biosciences

PAP: poly(A) polymerase

PBMCs: Peripheral Blood Mononuclear Cells

PBS: phosphate-buffered saline

PCR: polymerase chain reaction

PE: phycoerythrin

PEG: polyethylene glycol

piRNA: piwi-interacting RNA

pri-miRNA: primary miRNA transcripts

qPCR: quantitative PCR

RACE: Rapid Amplification of cDNA Ends

RAG: recombination-activating gene

RISC: RNA-induced silencing complex

RNA: ribonucleic acid

RSS: recombination signal sequence

RT: retrotranscription or reverse transcription

SBS: sequencing by synthesis

siRNA: small interfering RNA

SMS: single molecule sequencing

SNV: single-nucleotide variants

SOC: Super Optimal broth with Catabolite repression

ssDNA: single-stranded DNA

SVF: stromal vascular fraction

SVs: structural variants

sWAT: subcutaneous white adipose tissue

Tcm: central memory T cells

TCR: T cell receptor

TdT: terminal deoxynucleotidyl transferase

Tem: effector memory T cells

Temra: effector memory reexpressing CD45RA T cells

TGS: third-generation sequencing

TRBC: T cell receptor β constant segment

TRBD: T cell receptor β diverse segment

TRBJ: T cell receptor β joining segment

TRBV: T cell receptor β variable segment

Trm: tissue-resident memory T cells

TSO: template switching oligonucleotide

UMI: unique molecular identifier

UTR: untranslated region

VDJdb: Curated database of T-cell receptor (TCR) sequences with known antigen specificities

WT: wildtype

ZMW: zero mode waveguides

Useful terms

- **Alkaline wash** – Washing steps in which beads are placed on a magnet, its supernatant is discarded and NaOH is added to eliminate anything not covalently bound to the beads. Afterwards, pH is restored with TRIS-HCl washes.

- **Decreasing temperature protocol** – Protocol in which temperature declines from 70 to 50 ºC in 5 ºC decreases, with each stage lasting 5 minutes.

- **DTT (Double-Tailin Trap)** – Methodology created to prepare small-DNA NGS-libraries.

- **Elongation triplets** – Any primer sequence complementary to the last three nucleotides of the 3' end of any of the primers (including itself) present in the reaction

- **Empty library** – Libraries that do not present any miRNA between the two tails of the trap.

- **Extension PCR** – A PCR where only one DNA strand acquires the sequence of interest at its 3' end. There is no amplification of the DNA material.

- **Hybridization buffer** – Buffer containing TRIS-HCl 50 mM and NaCl 250 mM at pH 7.7.

- **miRNA-beads** – Magnetic beads that are functionalized to employ in the small-RNA NGS-library preparation with DTT methodology.

- **miRNA-beads for CC** – Magnetic beads that are functionalized to employ in the small-RNA NGS-library preparation with the evolved DTT methodology.

- **miRNA-free oligonucleotides** – oligonucleotides attached to the beads that have not captured any miRNA.

- **Mock-TCR –** Synthetic construction of a TCRβ in which TRBV segment is composed by the fusion of TRBV6-1 and TRBV27.

- **TCRβ-beads –** Magnetic beads that are functionalized to employ in the TCRβ NGS-library preparation.

Abstract

Since the discovery of DNA, researchers have devoted all their efforts to creating methodologies and techniques that would allow us to read the information contained in genes. Initially, Sanger sequencing revolutionised the field, being one of the first technologies that enabled us to determine the sequence of a DNA region. Decades later, the development of massive sequencing would allow the simultaneous generation of millions of sequences, thereby producing a large amount of genetic data.

Despite the progress presented by massive sequencing, it still exhibits limitations and sources of error that often make data non-reproducible and/or unreliable. The main sources of error occur during the library preparation or sequencing processes. Various steps in both procedures can introduce biases or errors, such as the ligation of adapters necessary for massive sequencing or the PCR amplification of genetic material. Furthermore, massive sequencing is limited to short reads, up to about 600 nucleotides. However, the development of technologies capable of obtaining reads of tens of thousands of nucleotides with high accuracy (third-generation sequencing) is very promising.

In this thesis, two methodologies have been developed to address the issues associated with library preparation for both long and short sequences. For long sequences, we developed the library preparation methodology for the β chain of the T cell receptor (TCR). For short sequences, the methodology was created for small RNA. In both cases, the use of magnetic microparticles allows us to covalently couple the DNA molecules of interest and include sequences like massive sequencing adapters without the need for any ligation reaction. Additionally, they facilitate the exchange of reagents and the removal of waste from previous reactions due to their magnetism.

The methodology for TCRβ has allowed us to conduct a pilot study to understand the role of adaptive immunity in adipose tissue (AT). AT was initially considered as a mere lipid storage, but its concept has changed over the years. Currently, AT is recognised as an endocrine organ, as well as playing an important role in non-shivering thermogenesis. In recent years, it has been demonstrated that AT plays a significant role in the immune system, both in the innate and adaptive aspects. AT contains a large number of immune cells, as well as lymphoid structures considered secondary or tertiary lymphoid organs, known as FALCs (fat-associated lymphoid clusters). Each cell type has a crucial role in one of AT's functions and contributes directly or indirectly to other functions.

The large number of tissue-resident memory T cells (Trm) in adipose tissue suggests its fundamental role in defence against viruses, bacteria, and other pathogens. However, few studies have been conducted in humans. In this study, we collected human samples from blood, liver, and white adipose tissue, both omental and subcutaneous, from obese patients who underwent bariatric surgery. The goal was to shed light on the relationship between these solid tissues and adaptive immunity and compare them with blood using massive sequencing techniques, functional assays, and flow cytometry. The main findings indicated that solid tissues had a less diverse T cell repertoire and were characterised by hyperexpanded clonality. Additionally, we predicted the presence of

TCRs specific for viral epitopes such as cytomegalovirus, which we confirmed through functional assays. These assays reported a significantly higher number of cytomegalovirus-reactive T cells in omental AT. Finally, we observed that the majority of T cells in AT had a Trm phenotype.

Furthermore, this study allowed us to observe non-canonical TCRβs. TCR formation occurs through somatic recombination of its TRVB, TRDB, and TRJB gene segments. This recombination is imprecise, removing and adding nucleotides at the segment junctions, creating a hypervariable region crucial for antigen recognition. Through complete TCRβ massive sequencing, we observed aberrant rearrangements where the TRVB region consist of a fusion between two TRVB segments in both blood and AT. Although in low frequency, this indicates a previous selection of lymphocytes with non-canonical TCRs that remain in the immunological memory. Experiments were conducted to demonstrate their veracity and ensure they were not artefacts created in the library preparation or sequencing processes. Likewise, in vitro assays were performed with TCRβ knock-out T cells by introducing various TCR constructs and verifying their assembly on the cell membrane.

In the pursuit of technological innovation in massive sequencing, nanotechnology-based methodologies were developed for library preparation. However, due to the limited time of the study, direct comparisons with commercial kits were not possible. Likewise, the validation of the methodology for small RNA with real samples remains pending. Despite these constraints, the research contributes valuable insights to the understanding of adaptive immunity, particularly in adipose tissue, as well as to the generation of TCRβ diversity, and provides a foundation for future developments in sequencing technologies.

Resumen

Desde el descubrimiento del ADN, los investigadores han centrado todos sus esfuerzos en crear metodologías y técnicas que nos permitieran leer la información que contenían los genes. En un primer lugar, la secuenciación Sanger revolucionó la investigación, siendo una de las primeras tecnologías que nos permitían saber la secuencia de una región del ADN. Décadas más tarde, el desarrollo de la secuenciación masiva permitiría obtener millones de secuencias simultáneamente, generando así gran cantidad de datos.

A pesar del avance que supone la secuenciación masiva, ésta aún presenta limitaciones y fuentes de error que hacen que, en muchos casos, los datos no sean reproducibles o fiables. Las principales fuentes de error se encuentran durante el proceso de preparación de librerías o la secuenciación. Hay diversos pasos en ambos procediminetos que pueden sesgar los resultados o inducir errores, por ejemplo, la ligación de los adaptadores necesarios para la secuenciación masiva o la amplificación por PCR del material genético. Por otro lado, la secuenciación masiva está limitada a lecturas cortas, de hasta unos 600 nucleótidos. Sin embargo, el desarrollo de tecnologías capaces de obtener lecturas de decenas de miles de nucleótidos con alta precisión (secuenciación de tercera generación) es muy prometedor.

Durante esta estudio se han desarrollado dos metodologías que intentan mitigar los problemas asociados a la preparación de librerías abordando secuencias largas y cortas. Para las secuencias largas desarrollamos la metodología de preparación de librerías la cadena β del receptor de linfocitos T (TCR, del inglés *T cell receptor*) y para las cortas, la del ARN de pequeño tamaño. En ambos casos hacemos uso de micropartículas magnéticas que nos permiten unir de manera covalente las moléculas de ADN de interés e incluirles secuencias como los adaptadores de secuenciación masiva sin necesidad de ligación. Asimismo, nos facilitan el cambio de reactivos y la eliminación de desechos de reacciones anteriores gracias a su magnetismo.

La metodología para el TCRβ nos ha permitido llevar a cabo un estudio piloto para conocer el papel de la inmunidad adaptativa en el tejido adiposo (AT, del inglés *adipose tissue*). Inicialmente, se pensaba que el AT era un mero almacén de lípidos, sin embargo, el concepto ha cambiado a lo largo de los años. Actualmente, se reconoce el AT como un órgano endocrino, así como su papel en la termogénesis adaptativa (sin temblores). En los últimos años se ha demostrado que el AT tiene un rol importante en el sistema inmunitario, tanto en el innato como en el adaptativo. El AT contiene gran cantidad de células inmunitarias, así como estructuras linfoides que se consideran órganos linfoides secundarios o terciarios, conocidas como FALCs (del inglés, *fat-associated lymphoid clusters*). Cada tipo celular tiene un papel importante en alguna de las funciones del AT y, además, contribuyen directa o indirectamente a otras.

El gran número de células T de memoria residentes en tejido (Trm, del inglés *tissue-resident memory T cells*) postulan al AT como un pilar fundamental en la defensa contra virus, bacterias y otros patógenos. Sin embargo, se han llevado a cabo pocos estudios en humanos. En este estudio, recogimos muestras humanas de sangre, hígado y tejido adiposo, tanto omental como subcutáneo, de pacientes obesos que se sometieron a

una cirugía bariátrica. El objetivo fue arrojar luz sobre la relación entre estos tejidos con la inmunidad adaptativa y compararlos con la sangre, mediante técnicas de secuenciación masiva, ensayos funcionales y citometría de flujo. Los principales hallazgos indicaron que los tejidos tenían un repertorio de linfocitos T menos diverso que la sangre, y estaban caracterizados por una clonalidad hiperexpandida. Además, predijimos la presencia de TCRs específicos para epítopos virales como el citomegalovirus, que confirmamos mediante ensayos funcionales. Estos ensayos reportaron un número significativamente mayor de linfocitos T reactivos a citomegalovirus en el AT omental. Por último, observamos que la mayoría de los linfocitos en el AT tenían un fenotipo Trm.

Por otro lado, este estudio nos permitió observar TCRβs no canónicos. La formación del TCR se lleva a cabo por la recombinación somática de sus segmentos génicos V, D y J. Esta recombinación es imprecisa, eliminando y añadiendo nucleótidos en la unión de los segmentos. De esta manera se crea la región hipervariable CDR3 (del inglés, *complementary-determining región 3*), fundamental para el reconocimiento de antígenos. Gracias a la secuenciación masiva del TCRβ completo, observamos reordenamientos aberrantes donde se producía una fusión entre dos segmentos V tanto en sangre como en AT. Aunque en baja frecuencia, esto indica una selección previa de linfocitos con TCRs no canónicos que permanecen en la memoria inmunológica. Se llevaron a cabo experimentos para demostrar su veracidad y comprobar que no fueran artefactos creados ni en la preparación de librerías ni en la secuenciación. Del mismo modo, realizamos ensayos *in vitro* con células T *knock-out* para el TCRβ introduciendo diversas construcciones de TCRs y verificando su ensamblaje en la membrana celular.

En la búsqueda de innovación tecnológica en la secuenciación masiva, se desarrollaron metodologías basadas en nanotecnología para la preparación de librerías. Sin embargo, debido al tiempo limitado del estudio, no fue posible realizar comparaciones directas con kits comerciales. Del mismo modo, la validación de la metodología para ARN pequeño con muestras reales está aún pendiente. A pesar de estas limitaciones, la investigación supone un gran avance en la comprensión de la inmunidad adaptativa, especialmente en el tejido adiposo, así como en la generación de diversidad del TCRβ, y sienta las bases para futuros progresos en tecnologías de secuenciación.

1. Introduction

1.1. Next Generation Sequencing (NGS)

In the last years, the development of next-generation sequencing (NGS) technologies (also referred to as massive sequencing, second generation or deep sequencing) has transformed the field of molecular biology, offering extraordinary insights into the intricate complexities of genetic information. This revolutionary technology has supplanted traditional sequencing methods, enabling researchers to decode vast amounts of genetic information with remarkable speed and accuracy.

1.1.1. Sanger sequencing *versus* NGS

One of the first generations of sequencing, and that is still commonly used, was Sanger sequencing in 1977 (Sanger et al., 1977). It employs the chain termination method, providing longer read lengths and high accuracy (Figure 1, left). Unfortunately, it is slower and less cost-effective because it focuses on one template at a time. In contrast, NGS is known for its rapid, high-throughput capabilities, generating massive amounts of data simultaneously. NGS instruments combine the enzymatic processes with the data acquisition in a stepwise manner, allowing the simultaneous generation of sequence data from tens of thousands to billions of templates (Figure 1, right). While Sanger sequencing is suitable for targeted analyses on pre-known genes/variants, NGS can be applied to whole-genome sequencing, exome sequencing, transcriptome analysis, etc. In the past, sequencing the first human genome as a haploid reference took nearly 10 years, whereas currently, a full diploid human genome sequence can be achieved within a matter of days (McCombie et al., 2019).

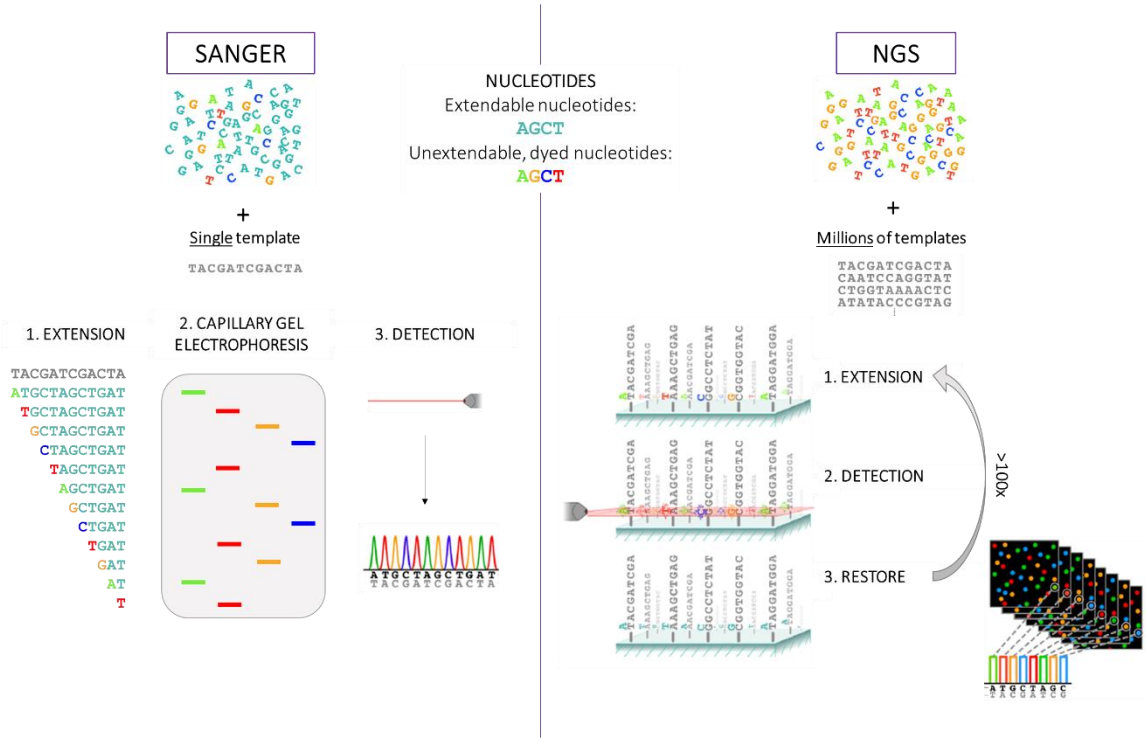

***Figure 1**. Comparison between Sanger sequencing (left) and Next-Generation Sequencing (NGS, right) processes. Modified from Muzzey et al., 2015.*

Another major difference between first- and second-generation sequencing methods is the data analyses. In Sanger sequencing, reads are commonly obtained from subcloned fragments of a larger DNA, allowing read assembly through sequence similarity. In that way, it involves simpler bioinformatics workflows, making it suitable for targeted analyses and variant identification through visual inspection of electropherograms. On the other hand, NGS methods involve shorter read lengths, making read assembly challenging, particularly when analysing a whole genome rather than cloned fragments. This technology demands advanced algorithms and computational resources for tasks like read alignment, variant calling, and coverage assessment of massive data volume. Because of these specialized processes, dedicated bioinformatics personnel are needed.

1.1.2. How does NGS work? Main differences between Illumina and Ion Torrent

Most commercial platforms of NGS are based on the concept of sequencing by synthesis (SBS) (Ronaghi et al., 1998), which permits the incorporation of nucleotides and data collection by signal detection at the same time. Other shared steps are the fragmentation of DNA/RNA; the generation of a library containing all the molecules to be sequenced, by adding platform-specific adapters to both ends; and the on-surface amplification of the templates hybridized to oligonucleotides with the complementary sequence of the adapters covalently attached to a surface. A pre-amplification is needed since NGS is not capable of sequencing a unique molecule. Thus, a clonal expansion of isolated molecules is required. This on-surface amplification fixates X-Y coordinates for each cluster of identical copies of a DNA template, ensuring that the signal can be distinguished from background noise (Goodwin et al., 2016; McCombie et al., 2019).

In the last decades, many different NGS techniques have arisen. In this study, we will focus on Illumina and Ion Torrent platforms. One of the key differences between both platforms is the mechanism by which the incorporation of a nucleotide is detected. Illumina copies Sanger sequencing, using labelled nucleotides distinguished by a specific fluorescent emission wavelength. These nucleotides are also modified to only permit one base incorporation each cycle. After the incorporation of one nucleotide into each template, the rest are removed, and the surface is imaged to identify which one was incorporated at each cluster. The blocking modification and fluorophore are then cleaved off. In contrast, Ion Torrent was the first NGS platform without optical sensing. The signal is detected by a pH change resulting from the release of protons after incorporating a nucleotide. Therefore, bases are introduced sequentially one at a time after removing the previous one (Goodwin et al., 2016; McCombie et al., 2019; Quail et al., 2012).

Another major difference between Illumina and Ion Torrent lies in their on-surface amplification approaches. In Illumina, oligonucleotides (forward and reverse), complementary to the sequences of the adapters, are covalently attached to a flat surface, either randomly or patterned. Following the hybridization of the target molecules, their amplification is performed with unlabelled nucleotides (dNTPs, deoxynucleotide

triphosphates). DNA folds over into a bridge-like structure. Consequently, fragments will become double-stranded to later be denatured, and repeat the process several times, forming clusters that enhance signal detection (Figure 2, top). It is important to control the DNA concentration so that non-overlapping clonal clusters are formed, maintaining spatial integrity. Then, the first sequencing cycle begins by adding labelled-reversible nucleotides (Goodwin et al., 2016).

Conversely, in Ion Torrent, oligonucleotides complementary to the sequences of one of the adapters are covalently attached to beads. In this case, an emulsion PCR (polymerase chain reaction) is performed to amplify the molecules to be sequenced. A water-in-oil emulsion is created by mixing an aqueous solution (containing DNA fragments, primers, and enzymes) with an oil phase. This results in the formation of tiny water droplets, each containing a single DNA fragment, within the oil matrix. Amplification occurs independently within each water droplet. As well as in the Illumina sequencing process, it is important to control DNA concentration so that each bead captures one template. If a bead captures two different templates, it will be considered as polyclonal and aberrant, and will be discarded during the first data analysis. After the PCR amplification, droplets are broken, releasing the amplified DNA on the beads. Then, only beads with templates are introduced into a chip containing millions of wells. Only when a well contains a unique bead the signal is valid (Figure 2, bottom). During the sequencing process, each of the four bases is introduced sequentially (Goodwin et al., 2016).

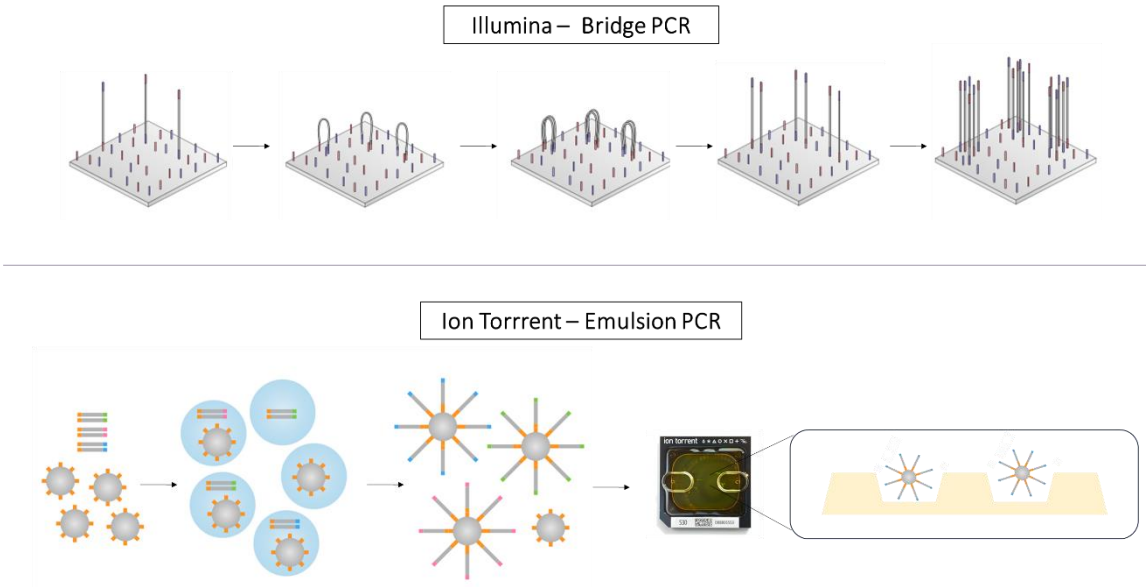

Figure 2. *Representation of on-surface amplification of two different NGS platforms, Illumina (bridge PCR, top) and Ion Torrent (emulsion PCR, bottom). Modified from (*ATDBio - Next Generation Sequencing*, n.d.).*

1.1.3. NGS limitations and sources of error

Despite the capability of reading thousands of templates simultaneously, NGS is limited and has associated biases. In this section, we will try to cover some of the

limitations and sources of error that exist in NGS sequencing, and the improvements being carried out to overcome them.

1.1.3.1. Library preparation.

Library preparation bias can arise from various steps. Generally, one of the main biases lies in the under or overrepresentation of some sequences. This can happen over the PCR amplification, but also during the random priming during the reverse transcription (RT) in RNA-seq for example, hybridization in the capture-based approaches or during the ligation reaction. All these steps' biases are based on the same premise: some sequences take precedence over others during these processes.

It is known that polymerases amplify certain sequences preferentially over others, being GC regions or repetitive elements more difficult to amplify (Goodwin et al., 2016; McCombie et al., 2019; Shi et al., 2021). Thus, optimizing PCR conditions, including additives in the process, employing high-fidelity enzymes with enhanced GC tolerance, or including unique molecular identifiers (UMIs) can help mitigate this type of bias (van Dijk et al., 2014). The amplification process holds significance as the initial quantity of genetic material is another potential source of bias. Therefore, pre-amplification must be used judiciously, applying quality controls to achieve the required quantity of material, even more when a sample is scarce.

Likewise, random priming or hybridization may exhibit bias toward certain sequences more than others. Researchers might avoid this by optimizing conditions, assessing the efficiency of the mix of random primers or probes, and validating capture performance across diverse genomic regions (Shi et al., 2021; van Dijk et al., 2014).

The ligation bias arises during the attachment of the adapters, needed for NGS, to the DNA or RNA fragments. Bias can be introduced across samples due to the adapter sequence (Alon et al., 2011; Jayaprakash et al., 2011), or within each sample due to the sequences to be ligated (Hafner et al., 2011). Certain sequences may exhibit differential ligation efficiency, resulting in an over or underrepresentation of sequences in the final library. In addition, ligases also present differences in terms of fidelity and efficiency performance (Bilotti et al., 2022). Furthermore, the ligation reaction step increases its bias when the target molecules are a few nucleotides long, such as small RNA (Song et al., 2014). Likewise, RNA secondary structures also influence the adapter-ligation bias (Baran-Gale et al., 2015). Many studies have addressed this problem and some of them suggested specific solutions. Some of the improvements include the optimization of the ligase choice, including additives in the reaction such as DMSO (dimethyl sulfoxide) or PEG (polyethylene glycol), the use of chimeric DNA/RNA adapters for RNA ligation, pre-heat the sample at 90 ºC to denature secondary RNA structures, etc. But undoubtedly, the most suggested solution was the use of randomized adapter pools (Alon et al., 2011; Bilotti et al., 2022; Hafner et al., 2011; Jayaprakash et al., 2011; Shi et al., 2021; Song et al., 2014).

Altogether, these library preparation-associated biases lead to non-reproducible data and non-reliable results. Some studies, mainly in the miRNA field, have proved that

final libraries from the same sample carried out for different platforms with their specific kits resulted in very different outcomes (Baran-Gale et al., 2015; Coenen-Stass et al., 2018; Heinicke, Zhong, Zucknick, Breidenbach, Sundaram, T. Flåm, Leithaug, Dalland, Rayner, et al., 2020). In this study, we will focus our attention on solving the ligation bias problem, without ignoring the rest of the error sources. Our main objective was to eliminate the ligation reaction when preparing NGS libraries.

1.1.3.2. Error rates.

Various steps in the sequencing process can introduce errors, from the sequencing reaction to the data analysis. Each platform relies on a specific chemistry that can introduce errors. For example, Ion Torrent is based on postreaction detection by change in pH. When multiple identical nucleotides form a homopolymer, the pH change detected by the sensor is imperfectly proportional to the actual number of nucleotides elongating in a template. It is established that a higher voltage is detected for homopolymers up to 6-8 bases. However, beyond this range, accurately determining the exact number of incorporated bases becomes challenging due to the imperfect correlation between the detected pH change and the true length of the homopolymer. On the other hand, Illumina produces errors when homopolymers are >20 bases long, particularly with GC-rich motifs (Goodwin et al., 2016; McCombie et al., 2019; Quail et al., 2012; Yohe & Thyagarajan, 2017).

Errors may depend on the instrument's performance such as signal-to-noise ratios, cycle-specific errors, algorithms used for base calling and data analysis, or low coverage of difficult regions. Different studies have tested the effectiveness of different platforms. Error rate in Illumina platforms is reported to be under 0.4%, whereas Ion Torrent is around 1.5%, with 76.45% and 15.92% error-free reads (without any mismatch or indel), respectively (Loman et al., 2012; Quail et al., 2012). Thus, accuracy is better in Illumina platforms. In contrast, Ion Torrent correctly determines 82% of the SNPs, while Illumina does in 70% of the cases. Moreover, the latest presents around 2.5% of false positives (Goodwin et al., 2016; Quail et al., 2012).

The intermediate on-surface amplification step introduces diverse artefacts. The first one relies on the error rate of polymerases, which can mask true variants if the error occurs in an early cycle. Moreover, during the sequencing of clusters, certain molecules may incorporate an erroneous base, contributing to an escalating noise with every nucleotide addition. The second one relies on the composition of the template, being those with an uneven GC composition inefficiently amplified (McCombie et al., 2019; Quail et al., 2012; Shi et al., 2021).

Therefore, single-molecule sequencing (SMS) is pursued to avoid intermediate amplification and its associated artefacts. Platforms like Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) are based on this approach. SMS reads typically have a higher error rate than SBS due to the difficulty of overcoming the signal-to-noise ratio associated with the collection of data from a single molecule. Nevertheless, both

technologies have recently improved and can achieve 87-98% (ONT) and 87-99% (PacBio) long-read accuracy (Logsdon et al., 2020).

1.1.3.3. Read length.

One notable limitation of NGS compared to Sanger sequencing is the relatively short read lengths. The above-mentioned accumulation of errors in the amplification process, or the detecting methods of nucleotide incorporation, reduces the read length. This occurs because the noise might become excessively high, impeding further accurate sequencing (McCombie et al., 2019).

Additionally, short reads challenge the mapping of the genome, particularly where repeated sequences appear, leading to ambiguities in data interpretation. Longer reads help bridge repetitive elements and enable the reconstruction of more contiguous and accurate genomic sequences. Shorter reads may result in fragmented assemblies, making it challenging to piece together the full genomic picture. This limitation is more important when identifying specific structural variations, allelic variants, or haplotypes. The ability to have all the information in a single read becomes crucial (Logsdon et al., 2020).

Regarding read length, Ion Torrent is a little bit ahead of Illumina. Both Illumina and Ion PGM work with pair-end reads, reaching around 400 base pairs (bp) (2x200 bp). Pair-end reads refer to a type of sequencing strategy in which both ends of a DNA fragment are sequenced separately. On the other hand, Ion S5 abandoned this approach to work with single-end reads and reach reads up to 600 bp. Now, Illumina (in particular, MiSeq) is capable of reading 2x300 bp. Despite reaching 600 bp, these sequences are not present in a single read but need to be overlapped. Therefore, the effective length of these sequences is not truly 600 bp.

In recent years, emerging technologies such as ONT or PacBio, have significantly improved in this field, which resulted in third-generation sequencing (TGS) or long-read sequencing (LRS). These platforms can generate much longer reads, ranging from 10 kb (kilobases) to 1500 kb in ONT and from 5 kb to 200 kb in PacBio (Logsdon et al., 2020). However, these platforms were not still suitable for short-read sequencing, which in some cases is essential. Last year, ONT presented the “short-fragment mode”, allowing reads as short as 20 bp to be included in the analysis in the MinKNOW platform (only available for DNA) (van Dijk et al., 2023).

1.1.3.4. GC content.

There is a systematic variation in the sequencing depth coverage based on the GC content of the genome. Firstly, as mentioned before, during the sequencing process, amplification is often employed to create sufficient material for later detection. The GC pair has three hydrogen bonds compared to the two in the AT, making it more stable. This stability can affect the efficiency of processes like denaturation and amplification, leading to underrepresented GC-rich regions. This uneven coverage may impact the sensitivity of variant detection, challenging the accurate analysis of some regions due to reduced sequencing depth. For this reason, researchers try to mitigate this bias by using specialized

polymerases with enhanced GC tolerance, optimizing bioinformatic tools to correct biases or improving library preparation techniques (Shi et al., 2021).

1.1.3.5. Other limitations.

Other types of limitations exist such as cost, time, sample quantity and quality, or computational complexity. Although sequencing prices have decreased significantly, the overall cost, either by doing the whole process or outsourcing part or all of it, is unfeasible for some research groups, particularly small or emerging groups.

Additionally, the time invested from sample collection to final results is huge. Despite the speed achieved in data generation, the entire NGS workflow still takes several days, limiting its application in certain time-sensitive scenarios such as clinical diagnosis (Yohe & Thyagarajan, 2017). Platforms like Ion Torrent have developed machines (Ion Chef) that prepare the library and load the chip, making it a faster and simpler platform (Goodwin et al., 2016).

A critical aspect to highlight is the quality and quantity of the DNA or RNA to be sequenced. Low-quality or limited samples may lead to biased results or failed experiments, which cannot be afforded given the cost of money and time involved. Regarding computational processes, the amount of data generated by NGS requires specialized researchers to handle and interpret it. This aspect also hampers its use in clinical diagnosis (Yohe & Thyagarajan, 2017).

Despite limitations, NGS techniques provide researchers with a comprehensive understanding of diverse biological processes. Furthermore, these limitations are deeply studied by companies, which allows the continuous improvement and development of new platforms. In fact, as this study was conducted, some aspects started to become outdated due to the rapid innovation in this field.

Regardless of remarkable advancements in the sequencing process, challenges persist in the preparation of reliable sequencing libraries, particularly concerning limitations and sources of error in existing methodologies. This thesis addresses such challenges by introducing novel methodologies for T cell receptor (TCR) and microRNA (miRNA) library preparation, aiming to avoid principal sources of biases such as the ligation reaction or minimizing the initial quantity of genetic material required.

1.2. T cell receptor (TCR)

The T cell receptor (TCR) stands as a pivotal player in the complex and intricate realm of the immune system. Found on the surface of T lymphocytes, or T cells, possess an extraordinary ability to discriminate between self- and non-self-antigens, ensuring the immune system's precision in targeting threats while avoiding attacks on the body's own tissues. The TCR's role in antigen recognition is a key step in the adaptive immune

response, orchestrating a cascade of events that ultimately lead to the activation of T cells and the initiation of an immune defence tailored to specific pathogens.

1.2.1. Brief introduction on T cells

T cells are the central players in adaptive immunity, together with B cells. T cells play a crucial role in defending the body against infections, including viruses and bacteria, as well as in recognizing and eliminating cancerous cells. Their precursor cells are produced in the bone marrow and migrate to the thymus, where they mature into T cells. There are three types of T cells: T helpers (Th) activate other immune cells; T cytotoxic (Tc) target and destroy infected cells or abnormal cells, and T regulatory (Treg) balance the immune system by suppressing excessive immune responses and preventing autoimmune reactions. They all are characterized by specific surface cell receptors (T cell receptors). Unlike B cells, T cells need to have antigens presented to them by an antigen-presenting cell (APC) via the major histocompatibility complex (MHC), referred to as human leukocyte antigen in humans (HLA).

Generally, HLA molecules engage with short peptide fragments that mirror the proteins expressed or internalized by a cell at a given moment. Consequently, T cells survey the antigenic landscape within a cell, examining the peptides presented by the HLA molecules. In essence, this interaction of HLA-peptide-TCR permits the real-time response by the immune system to possible threats (Figure 3A). The first antigen encounter produces a slower response. Afterwards, lymphocytes experience clonal expansion, generating twin cells within 5-7 days and allowing a rapid response for the next antigen-specific encounter. These twin cells will persist in the body, creating the immunological memory (Delves et al., 2017). In practical terms, this eventuality is highly improbable due to the immense diversity of receptors generated, each exhibiting exquisite specificity for a particular HLA-peptide combination. Additionally, most peptides presented on HLA molecules are typically derived from self, unless the antigen-presenting cell is infected, so the probability is further diminished.

1.2.2. Anatomy of the TCR: structure and diversity

T cell receptor is composed of a heterodimer of two different protein chains. There are two types of TCRs based on their variable chains: TCRαβ and TCRγδ. Notably, each T cell expresses one unique TCR. The disulfide-linked variable chains are anchored to the cell membrane as part of a complex with three CD3 protein dimers (CD3εγ - CD3εδ - CD3ζζ). Additionally, T cells also express CD4 or CD8 membrane proteins, acting as co-receptors in the HLA-antigen-TCR interaction. They control whether a T cell recognizes antigens presented by HLA class I, becoming Tc cells (CD8+) or HLA class II and becoming Th cells (CD4+). Likewise, the cytoplasmatic part of these proteins is pivotal as it is bound to Lck, a tyrosine kinase protein that initiates the signal transduction cascade that follows when a T-cell encounters an antigen (Figure 3B) (Attaf, Legut, et al., 2015; Delves et al., 2017).

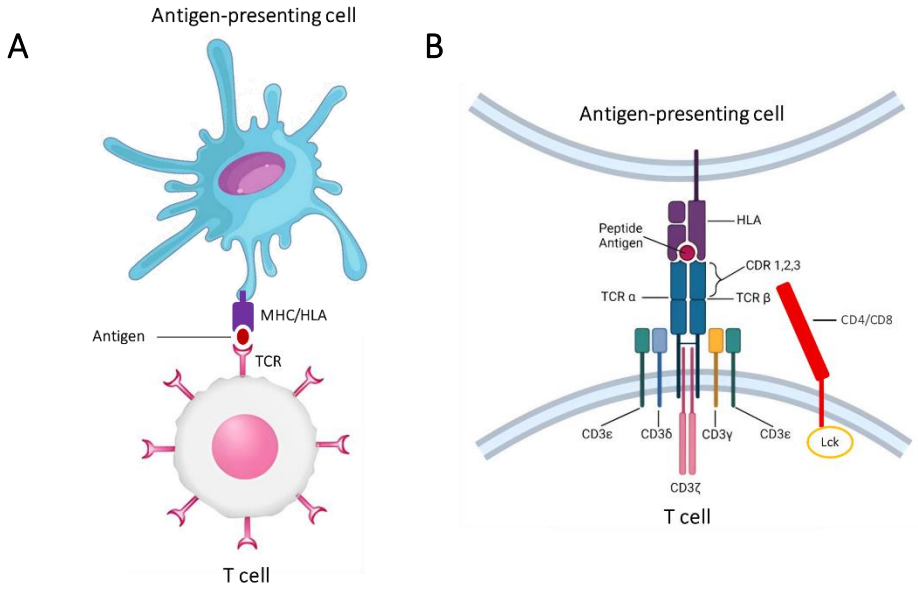

Figure 3. (A) *Interaction between an antigen-presenting cell and a T cell through the MHC/HLA-antigen-TCR complex.* ***(B)*** *Close-up of the MHC/HLA-antigen-TCR encounter where it is detailed the TCR structure with two variable chains (α and β) containing CDR1, 2 and 3 in both chains, three CD3 protein dimers (CD3εδ - CD3ζζ - CD3εγ), and the CD4 or CD8 protein bound to the Lck protein. Modified from Sun et al., 2021.*

In immunology, the term repertoire defines the collection of TCR variants in a sample. The TCR repertoire offers a measurable indicator of immune activation, evolution, or the reconstitution of T cells. To achieve a broad diversity in antigen recognition by lymphocytes, the variable chains of the TCR undergo somatic recombination of their gene segments, known as V(D)J recombination. The complexity of the repertoire is not entirely random, as certain gene segments are preferentially selected (Robins et al., 2010). Both genetic and epigenetic factors influence the pre-selection of the repertoire, as the gene segments must be available for recombination to occur (Attaf, Legut, et al., 2015). Subsequently, there is thymic selection, eliminating T cells expressing receptors against self-proteins, preventing autoimmunity, and those with improperly synthesized chains. This study will focus its attention on Tαβ lymphocytes, specifically on the β chain.

The V(D)J recombination occurs during the maturation of T lymphocytes in the thymus. It owes its name to the semi-random somatic recombination of gene segments TRVB (variable), TRDB (diversity), and TRJB (joining) by the RAG1 and RAG2 genes (recombination-activating gene). RAG 1 and 2 proteins introduce double-strand DNA breaks between segments and their flanking recombination signal sequences (RSSs). This process creates a repertoire of different T cell clonotypes, theoretically estimated to be around 10^{12} in humans, although some authors (Attaf, Legut, et al., 2015) have suggested lower variability. From the vast number of TCR clonotypes, a small proportion is referred to as "public TCRs" because they are found in different patients with the same pathologies, indicating an evolutionary "adaptive convergence" (Madi et al., 2014). However, the majority are termed "private TCRs" as they are unique to a particular individual.

VDJ gene segments are located in the 5' region of the *tcrβ* locus, while the constant segment (TRCB) is in the 3'. V(D)J recombination is a rearrangement of the locus, resulting in a high number of TCR variants. The locus begins at the 5' end with a cluster of multiple TRVB (presumable > 60) gene segments, followed by two TRDB (2) -TRJB (13) -TRCB (2) clusters towards the 3' end. The recombination process consists of three parts. Firstly, one of the two TRDB segments recombines with one TRBJ segment. Note that, due to the specific arrangement of genes, the TRDB1 has the flexibility to pair with any of the 13 TRJB segments, while TRDB2 is limited to selecting from the seven TRJB2 segments. Secondly, the TRDB-TRJB junction recombines with one of the multiple TRVB segments. Thirdly, after splicing, the corresponding TRCB segment is joined to the rest of the recombination (Figure 4) (Attaf, Legut, et al., 2015; Delves et al., 2017).

Furthermore, the number of TCR variants increases due to imprecise addition or deletion of nucleotides during segment recombination (Attaf, Legut, et al., 2015; Delves et al., 2017). This phenomenon alters the reading frame in two-thirds of cases, leading to non-functional TCRs that will be eliminated through apoptosis. When the reading frame is not altered, the region created by the junction of gene segments is known as the complementarity-determining region 3 (CDR3). This region, also referred to as the hypervariable part, gives that unique identity to the TCRβ. Moreover, TRVB segments germline encode two more CDRs (CDR1 and CDR2) (Figure 4), which combined with the V(D)J recombination and the three CDRs from the TCRα chain, create the immense diversity of Tαβ cells.

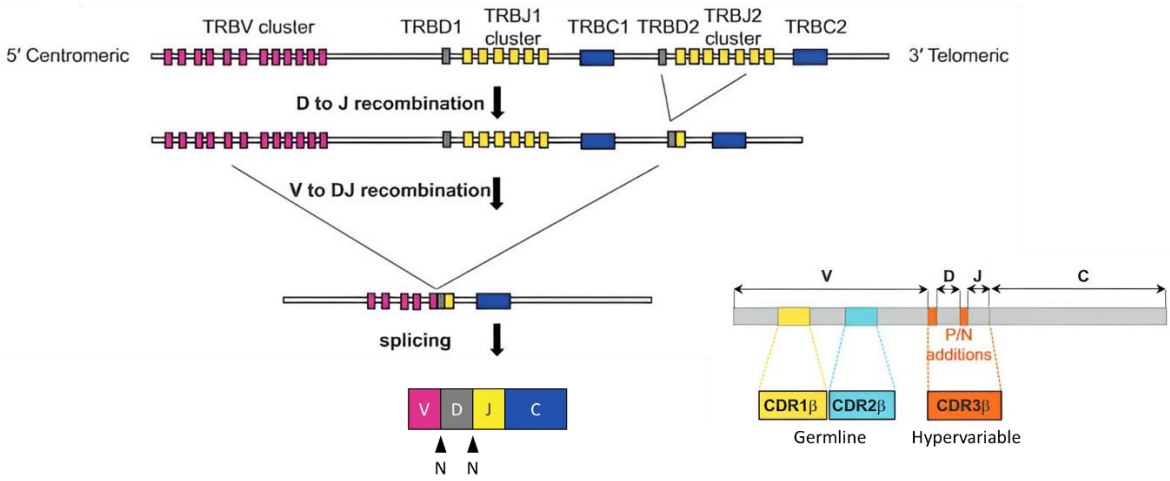

Figure 4. *V(D)J recombination of the tcrβ chain gene locus and CDRs distribution along the chain. Modified from (Attaf, Huseby, et al., 2015).*

While the precise mechanism by which the TCR interacts with the HLA-peptide complex remains not entirely elucidated, it seems that CDRs 1 and 2 play a significant role in engaging with the HLA molecule, while CDR3 establishes contact with the peptide. However, exceptions were observed where CDR1 or CDR2 interact with the peptide depending on the TRBV segment recombined (Culshaw et al., 2017; Szeto et al., 2021; Van Braeckel-Budimir et al., 2017). Moreover, subtle changes in the TCR structure were found to impact the interaction with the HLA (Sharma & Kranz, 2018). Regardless of which CDRs are responsible for contacting MHC or peptide, these regions constitute

the recognition components of the receptor. Consequently, it is in these regions that much of the variability between TCRs is observed.

It is widely believed that the rearrangement of the *tcrβ* locus is regulated in the context of allelic exclusion, although the precise mechanisms remain unclarified. Allelic exclusion is a regulatory mechanism that ensures the expression of only one allele of a given gene, thereby promoting monoallelic expression in a single cell. The TCRβ synthesis of the β chain occurs during the CD4-/CD8- (double negative) maturation phase and involves both alleles during the TRVB-TRDB-TRJB rearrangement. From this stage onward, TCR products are generally detected only at one allele. Significantly, the interallelic sequencing of TCRβ chain gene assembly is regulated in a manner that all TRVB-TRDB-TRJB rearrangements are not exclusively finished on a single allele before commencing the rearrangements on the alternative allele. However, it is worth noting that evidence suggests a non-negligible probability of removal of a TRVB-TRDB-TRJB joint using DJβ1 through a secondary Vβ–DJβ cis-rearrangement involving a DJβ2 unit (Outters et al., 2015).

The final TCRβ pairs with a constant pre-Tα chain, accompanied by the CD3 protein dimers. This pre-TCR will oversee suppressing RAG genes activity, mediating TCRβ allelic exclusion, and proliferating towards the α chain synthesis. RAG activity will be resumed for the α chain synthesis in the CD4+CD8+ phase (double positive). Conversely, during the TCRα formation, there is not allelic exclusion, sometimes resulting in dual T cells. These dual lymphocytes present a unique β chain and two different α chains (Elliott & Altmann, 1995; Schuldt & Binstadt, 2019). Lastly, cells develop into CD4+ or CD8+ (single positive) differentiated T cells (Delves et al., 2017).

1.2.3. The interplay of immunity and adipose tissue

Our comprehension of white adipose tissue (WAT) is constantly evolving. Initially perceived as a simple storage site for neutral lipids, it is now recognized as an endocrine organ crucial for regulating systemic energy homeostasis (Scheja & Heeren, 2019). Additionally, WAT contributes to non-shivering thermogenesis through the generation of specific thermogenic adipocytes (Wu et al., 2013). The latest facet attributed to WAT is its important role in the immune machinery (Trim & Lynch, 2022).

While it is commonly assumed that the majority of cells in white adipose tissue are adipocytes, these are the ones occupying more space. However, WAT includes a broad range of cell types, apart from white and thermogenic adipocytes, such as fibroblasts, endothelial cells, and various stem cell populations. Few studies addressed the importance and extensive diversity of leukocytes within WAT as we explored in García-Rubio et al., 2018. The traditional viewpoint regarding the interplay between immunity and WAT suggests that pro-inflammatory M1 macrophages are responsible for recruiting T cells due to the chronic low-level inflammation present in the tissue during obesity (Sell et al., 2012; Weisberg et al., 2003). Nevertheless, our research demonstrated that leukocytes persist in WAT even in lean individuals, constituting 45-50% of cells (García-Rubio et al., 2018).

Each of the diverse types of cells plays a role in WAT's primary functions: energy storage, endocrine regulation, non-shivering thermogenesis, and host defence. While some cells exhibit specialization in specific functions, they contribute directly or indirectly to others. For example, adipocytes, specialized in neutral lipid storage, also secrete adipokines for endocrine regulation (Unamuno et al., 2018), and upon microbial challenge, they release antimicrobial peptides for host defence (L. Zhang, Guerrero-Juarez, et al., 2015). Macrophages, for instance, crucially manage lipid distribution between adipocytes, acting as temporal lipid storage cells when faced with excessive cytotoxic free fatty acids and limiting cell hypertrophy (Guerrini & Gennaro, 2019; Jaitin et al., 2019; Prieur et al., 2011). Thus, WAT is viewed as a complex tissue with synergistic interactions between diverse cell types (Trim & Lynch, 2022).

Lymphoid structures, that resemble secondary or tertiary lymphoid organs, are present in WAT. These structures are referred to as fat-associated lymphoid clusters (FALCs) or milky spots (Bénézech et al., 2015; Moro et al., 2010; Trim & Lynch, 2022)). They are particularly notable in the peritoneal cavity, as they filter peritoneal fluid and capture soluble antigens (Meza-Perez & Randall, 2017). Furthermore, upon detecting microbial antigens, fibroblastic reticular cells in FALCs recruit peripheral monocytes, promoting the formation of new FALCs, followed by rapid activation of T and B cells (Trim & Lynch, 2022).

A substantial number of memory T cells are present in WAT. Interestingly, they reside in the tissue and do not recirculate. This accumulation of tissue-resident memory T cells (Trm) in FALCs offers two advantages. Firstly, WAT connects with all organs through lymphoid structures. During inflammation, lymph leaks from vessels reach WAT lymphatics and FALCs, synergizing with classic lymphatic drainage. Secondly, Trm's main source of energy is mitochondrial fatty acid oxidation, and upon microbial challenge, adipocytes increase lipolysis, providing fatty acids to Trm (García-Rubio et al., 2018; Han et al., 2017; Trim & Lynch, 2022). In summary, T cells in FALCs are strategically located in an environment abundant with fuel to respond to microbial encounters. Additionally, CD8+ Trm in WAT exhibit a higher proliferative potential than Trm in other locations and coexpress IFN-gamma and TNF-alfa shortly after activation (Han et al., 2017).

1.2.4. Why is it important to study the TCR landscape of WAT?

A significant portion of WAT-resident memory T cells is likely reactive to viral antigens. Presently, limited research, primarily in murine models, has explored the WAT reservoir of viral-specific memory T cells (Contreras et al., 2019; Han et al., 2017). Murine adipose tissue contains CD8 T cells specific to murine cytomegalovirus (mCMV) associated with the regulation of lifelong infection (Contreras et al., 2019). Some studies in humans have described the role of a minor subpopulation of CD4+ T cells in subcutaneous WAT of patients coinfected with HIV (human immunodeficiency virus) and CMV (cytomegalovirus) (Wanjalla et al., 2021), and the existence of CMV-reactive Trm cells in respiratory mucosa (Gordon et al., 2017).

The intricate relationship between immunity and WAT, particularly with the omental fat depot, remains relatively understudied. Despite the pivotal role that adipose tissue plays in immune responses, there is a scarcity of comprehensive studies focusing on the immune landscape within the oWAT. This adipose depot, located in the abdominal cavity, holds unique characteristics and interactions. As highlighted previously, our understanding of white adipose tissue has evolved beyond its traditional portrayal as a simple energy storage site, revealing its complexity as an immune organ. Leukocytes, both from adaptive and innate immunity, contribute significantly to WAT functions. However, the specific nuances within oWAT remain to be explored. Recognizing this research gap, our group is driven to unravel the immune interactions within omental adipose tissue through flow cytometry and high-throughput TCR sequencing. Thus, we aim to shed light on the immune dynamics in this unique fat depot and its potential implications for overall health and disease.

1.2.5. TCR and NGS

Until recently, when studying the TCR repertoire, the focus had predominantly centred on the hypervariable CDR3 region. As previously mentioned, this region of the TCR is formed through the imprecise joining of gene segments during V(D)J recombination. Importantly, it is typically the one interacting with the peptide presented by the HLA. Consequently, the diversity generated is immense, as each CDR3 is unique, aiming at encompassing maximum diversity and readiness to combat any foreign antigen or abnormality. This is why researchers find this part particularly intriguing.

However, as demonstrated in some studies, the CDR1 and 2 regions also interact with the peptide in certain cases (Culshaw et al., 2017; Szeto et al., 2021; Van Braeckel-Budimir et al., 2017), adding a layer of complexity to our understanding of TCR functionality. Similarly, small modifications throughout the TCR can influence its affinity for certain antigens (Sharma & Kranz, 2018). This is why we believe that having a comprehensive understanding of the entire TCR would be beneficial for advancing the design of new therapies related to TCR modification, such as CAR-T therapies. Manipulating TCRs holds promise for enhancing the precision and efficacy of immune-based treatments. Likewise, sequencing the whole TCR gives a better identification of the TRVB segment, since some variants differ in very few nucleotides (like segments from families TRVB5 and TRVB6). Shorter reads for the exclusive CDR3 study impede an accurate identification of TCRβ segments. This may produce the loss of sequences given the impossibility of a correct alignment with their corresponding segment (Barennes et al., 2021).

Nowadays, different companies have developed technologies for long-read TCRs, both for Illumina and Ion Torrent. To our knowledge, at the moment we were starting to develop the methodology later presented for the TCRβ library preparation, none of them existed. Table 1 summarizes commercially available technologies, including the ones that are focused on the CDR3 and those where all CDRs are sequenced (J. Zhang et al., 2019). Each of these products follows its own procedure for the preparation of TCRβ libraries,

with the most common being multiplex PCR using primers designed against TRVB and TRCB segments (iRepertoire, Illumina, and Thermo Fisher Scientific). On the other hand, Takara Bio employs the 5'RACE (Rapid Amplification of cDNA Ends) technology, and Qiagen performs end-repair and attaches one of the adapters through a ligation reaction.

Table 1. *Commercial technologies for the TCRβ library preparation. Information on the company, product, sequencing platform, CDR regions covered, and RNA input recommended, and methodology used for the ones covering three CDRs is shown. Modified Zhang et al., 2019.*

Company	**Product**	**Sequencing platform**	**CDR regions covered**	**RNA input**	**Methodology*** (Figure A 1)
Adaptive technologies	ImmunoSEQ® Human T-cell Assay	Illumina	CDR3	NA	NA
iRepertoire	Shor-Read iR-Profile	Illumina	CDR3	NA	NA
Illumina	AmpliSeq™ for Illumina TCR Beta-SR Assay	Illumina	CDR3	NA	NA
Themo Fisher Scientific	Oncomine™ TCR Beta-SR Assay	Ion GeneStudio S5	CDR3	NA	NA
BGI	Immune Repertoire Sequencing	Illumina	CDR3	NA	NA
BGI	Immune Repertoire Sequencing (5'RACE)	Illumina	CDR3	NA	NA
iRepertoire	Long-Read iR-Profile	Illumina/ Roche 454	CDR1, CDR2, CDR3	50-1800 ng	arm-PCR**
Illumina	AmpliSeq™ Immune Repertoire Plus TCR Beta	Illumina	CDR1, CDR2, CDR3	10-1000 ng	Multiplex PCR, 2 adapters ligation, PCR
Themo Ficher Scientific	Oncomine™ TCR Beta-LR Assay	Ion GeneStudio S5	CDR1, CDR2, CDR3	Min. 25 ng	Multiplex PCR, 2 adapters ligation
Takara Bio	SMARTer® Human TCR a/b Profiling Kit	Illumina	CDR1, CDR2, CDR3	10 ng-3 µg	RT with TSO*** (5'RACE), semi-nested PCRs to add adapters
QIAGEN	QIAseq Targeted RNA-seq Panel for T-cell Receptor	Illumina	CDR1, CDR2, CDR3	200 pg-1000 ng	End-repair with poly(A), 1 adapter ligation, 2 PCRs

*All explained methodologies start with a reverse transcription and have a last step of library purification.
**arm-PCR (amplicon rescued multiplex-PCR): RT and PCR1 at the same time with nested primers, chemical rescue of the product, and PCR2 amplification to add adapters.
***TSO: template switching template

1.3. Small RNA

Eukaryotic cells (and some viruses) produce small non-coding RNA molecules (19-25 bases) essential in genetic regulation. They act as selective expression repressors by inhibiting translation frequency and reducing the messenger RNA's (mRNA) half-life (Gebert & MacRae, 2019). Importantly, their deregulation is associated with most chronic pathologies processes. Small regulatory RNAs consist of three families: microRNA (miRNA), small interfering RNA (siRNA), and piwi-interacting RNA (piRNA). miRNAs oversee the regulation of endogenous genes, while siRNAs act as guardians of genome integrity, responding to external threats by foreign or invasive nucleic acids, such as viruses, transposons, and transgenes. Additionally, piRNAs play a crucial role in silencing transposons specifically within the germ line (Aalto & Pasquinelli, 2012; Carthew & Sontheimer, 2009; Ozata et al., 2019) Each of these RNA classes contributes distinctively to the intricate regulatory network that safeguards genomic stability and functionality. In the present text, we will focus solely on the predominant (and best-known) group: microRNAs.

1.3.1. MicroRNAs

The biogenesis of miRNAs, a process of a certain complexity, is subject to strict spatial and temporal control. It starts with the transcription of primary miRNA transcripts (pri-miRNAs) by RNA polymerase II or III. These pri-miRNAs undergo folding into hairpin structures with a terminal loop and two single-stranded flanking regions, forming pre-miRNAs. The imperfectly base-paired stem is critical for the cleavage by DGCR8 (DiGeourge critical region 8) and RNase III enzyme Drosha. Subsequently, exportin-5 facilitates the transport of pre-miRNAs from the nucleus to the cytoplasm, where the Dicer enzyme typically cleaves them, generating miRNA-miRNA duplexes. The functionally relevant mature miRNA strand is selectively retained, while the other strand is typically degraded. Following this, the mature miRNA is integrated into the RNA-induced silencing complex (RISC), which targets specific genes, initiating posttranscriptional gene silencing. Notably, the pivotal protein effector in these processes is the Argonaute (Ago) protein family (Figure 5) (Kim et al., 2009). This family constitute a diverse group of proteins that uses single-stranded small nucleic acids as guides, enabling them to identify and bind to complementary sequences in RNA or DNA. Specifically, miRNAs play a vital role in gene silencing by directing Ago to specific sites within the 3' UTR of mRNAs (Gebert & MacRae, 2019; Wu et al., 2020). Furthermore, miRNAs have been observed to interact with gene promoters, 5' UTRs, or coding sequences, yielding diverse outcomes. In specific circumstances, these interactions can result in gene expression activation rather than repression (O'Brien et al., 2018; Xiao et al., 2016).

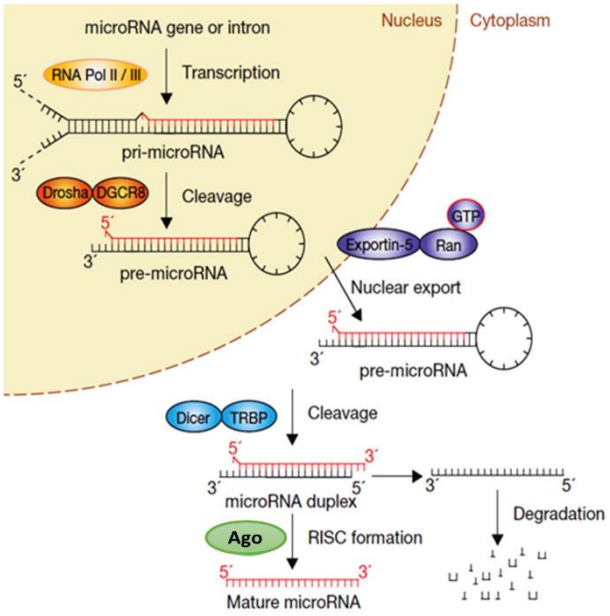

Figure 5. *miRNA biogenesis. Modified from Winter et al., 2009.*

One specific messenger RNA can be regulated by different miRNAs. Simultaneously, it is common for a specific miRNA variant to regulate several messenger RNAs. This phenomenon is known as a signalling axis. Frequently, each signalling axis is regulated by multiple miRNAs in synergetic cooperation. This complex network of interactions controls relevant processes such as cellular division, apoptosis, cellular differentiation, and energy metabolism (Bracken et al., 2016).

The majority of miRNAs act exclusively at an intracellular level. This swift-response mechanism involves direct interactions with cytoplasmic mRNA molecules. Notably, research has unveiled that gene-silencing effects mediated by miRNAs and siRNAs extend to the membranes of the endoplasmic reticulum, enabling the regulation of actively translated mRNAs (Barman & Bhattacharyya, 2015). Additionally, complexes involving Ago and miRNA have been identified not only in the cell nucleus but also in mitochondria and endosomes (Antoniou et al., 2014; Paramasivam & Vijayashree Priyadharsini, 2020; Wei et al., 2014). These complexes serve as versatile tools, exerting regulatory control over a diverse array of intracellular processes through gene repression or even activation.

However, there is also an intercellular level of functioning. Cells release communication vesicles referred to as exosomes, when derived from the endosomal compartment, or microvesicles, when shed from the plasma membrane. Generally, miRNAs are primarily packaged within exosomes. These vesicles contain miRNA, messenger RNA, metabolites, and surface protein complexes. Their content reflects the current state of the cell. Some mechanisms that regulate selective loading are known, and specific receptor-mediated interactions between exosomes and target cells have been described (Iraci et al., 2016; O'Brien et al., 2018).

Communication vesicles often have very short half-lives, predisposing them to function at a paracrine and autocrine level within the tissue environment. In fact, unless

they display "don't eat me" surface markers (such as CD47), they are avidly phagocytosed by macrophages and monocytes (Kamerkar et al., 2017). Nevertheless, there have been described exosomes that transmit information between cell populations located in different organs (Cheng et al., 2019; L. Zhang, Zhang, et al., 2015). However, most of the data come from pathological contexts, and it is unknown whether this is an ordinary physiological mechanism.

Thousands of miRNA variants have been described in different species, which are annotated in various databases (the most known being: miRBase). In humans, it has been estimated that the overall number of miRNA variants exceeds 2300 (after removing annotations of dubious reliability) (Alles et al., 2019). However, the actual number, likely higher, is still unknown.

1.3.2. miRNAs as relevant biomarkers

Approximately 20% of miRNA variants can be found in blood. A fraction of the circulating miRNAs is integrated into exosomes, while a high proportion of the rest is associated with lipoproteins. According to the current consensus, the circulating miRNAs that are free from exosomes are derived from passive release due to cell death (Turchinovich et al., 2013). miRNAs present in serum and plasma exhibit high stability against RNase digestion (Mitchell et al., 2008), mainly due to their small size and their inclusion in exosomes, Ago proteins and lipoproteins (Vickers et al., 2011). The quantity and composition of circulating miRNAs vary significantly depending on the physiopathological context. This, combined with their stability, allows the identification of markers for the presence, severity, and prognosis of diseases. Thus, miRNAs are considered nowadays as important biomarkers. This relevant field of research has been particularly developed in cancer studies (Hirschberger et al., 2018). Moreover, it is worth noting that miRNAs can also be found in other biological fluids such as urine or saliva, which opens new possibilities for non-invasive research and diagnosis.

Deregulation of microRNAs (miRNAs) is intricately linked to most chronic pathological processes in humans, including cancer, type 2 diabetes, neurodegenerative diseases, and immune disorders (Alles et al., 2019; Bracken et al., 2016; Fransquet & Ryan, 2018; Guay & Regazzi, 2013; Kalla et al., 2015; Pauley et al., 2009; Peng & Croce, 2016). The utility of circulating miRNAs as biomarkers was initially demonstrated in the serum samples of patients with diffuse large B-cell lymphoma (Lawrie et al., 2007). Similarly, miRNA expression profiles have proven valuable in identifying the primary tissue of cancers with unknown origins (Rosenfeld et al., 2008). In the context of different human cancers, the analysis of hundreds of miRNAs revealed a distinct reflection of the tumours' developmental lineage and differentiation status. Notably, there was a general downregulation of miRNAs in tumours compared to normal tissues, underscoring the potential of miRNA profiling in cancer diagnosis (Lu et al., 2005). Another example where miRNAs can be used as a biomarker is in type 2 diabetes. A specific set of five miRNAs showed characteristic deregulation in individuals with either prediabetes or type 2 diabetes. Importantly, alterations in the levels of these miRNAs were observed 5–10

years before the beginning of the disease, suggesting the potential of circulating miRNAs as early indicators of disease (Guay & Regazzi, 2013).

1.3.3. The importance of miRNAs

To sum up, here are some of the reasons highlighting the significance of miRNA research:

- Gene Regulation. miRNAs play a crucial role in post-transcriptional gene regulation, influencing the expression of target genes. Understanding the mechanisms through which miRNAs control gene expression provides insights into fundamental biological processes and cellular functions (Gebert & MacRae, 2019).

- Disease Association. Deregulation of miRNAs is implicated in various human diseases (Alles et al., 2019; Bracken et al., 2016; Fransquet & Ryan, 2018; Guay & Regazzi, 2013; Kalla et al., 2015; Pauley et al., 2009; Peng & Croce, 2016). miRNA expression profiles have shown promise as biomarkers, offering valuable diagnostic, prognostic, and treatment response information. Specific miRNA signatures provide tools for the early detection and monitoring of pathologies.

- Therapeutic Targets. Unravelling the complexities of miRNA biology can lead to the development of innovative tools and strategies for precision medicine. Likewise, manipulating miRNA expression levels or activity holds therapeutic potential. Researchers are exploring miRNA-based therapeutic strategies for conditions where miRNA deregulation plays a critical role. This includes the development of miRNA mimics or inhibitors for targeted interventions (He et al., 2020; Ho et al., 2022).

- Cellular Processes. miRNAs are involved in the regulation of diverse cellular processes such as proliferation, differentiation, apoptosis, and immune response. Investigating miRNA function enhances our understanding of these intricate cellular pathways. For example, immune cells frequently communicate through exosomes and microvesicles too. There have been described the transfer of surface co-activating molecules and antigen-presenting complexes via microvesicles (Pauley et al., 2009) and the exchange of immuno-miRNAs via exosomes (Hirschberger et al., 2018). These are mechanisms of rapid coordination between immune cells, and their significance is starting to be glimpsed. Over the last decade, numerous miRNAs have been identified to be involved in key regulating processes of the immune system (Hirschberger et al., 2018): differentiation from hematopoietic stem cells, phenotypic modulation in mature cells, cytokine response and synthesis, regulation of activation threshold, anergy, apoptosis, survival, etc.

- Biotechnological Applications. The knowledge gained from miRNA research has practical applications in biotechnology, such as the development of RNA-based technologies and gene editing tools.

1.3.4. Existing methodologies in miRNA analysis

The chemical nature of miRNAs (small size and similarity among some variants), and their low concentration in some samples, pose a methodological challenge at all levels (extraction, amplification, sequencing). The analysis of miRNAs is addressed using three groups of techniques:

- RT-qPCR. Well-established technique but limited to the study of already known variants of miRNAs. Moreover, primer design poses a challenge due to the limited size of miRNAs. Additionally, miRNA variants that differ in 1 or 2 nucleotides are difficult to identify by RT-qPCR.

- Microarray. Discrimination by hybridizations on an array of probes fixed on solid support (glass or nylon membrane) allows the simultaneous analysis of numerous miRNAs but it is also limited to already-known variants. While there are technological modalities with high specificity, the microarray has limited sensitivity (as it does not involve previous PCR amplification) (Garcia-Elias et al., 2017).

- miRNA-seq. Quantification of the group of miRNAs in a biological sample through NGS platforms (Illumina being the current gold standard). Unlike previous techniques, it allows the identification of new and unknown variants of miRNA. Likewise, it identifies other small RNA-regulating families (siRNA or piRNA). In contrast with RT-qPCR, it efficiently distinguishes between variants at a single nucleotide resolution as well as variations such as alternative transcripts (Byron et al., 2016).

Among the previously discussed methodologies for miRNA analysis, miRNA-seq stands out due to its notable features, including high sensitivity, specificity, accurate quantification, and the ability to identify previously unknown variants. Despite the high-quality standards of NGS, challenges such as sampling and ligation bias arise when applying it to the unique characteristics of miRNA (Fuchs et al., 2015; Hafner et al., 2011; Raabe et al., 2014; Tian et al., 2010).

1.3.5. Methodological barriers in miRNA-seq

In the same way that we discussed in section 1.1.3. NGS limitations and sources of error, when constructing libraries for massive miRNA sequencing, the same limitations and sources of error exist. Moreover, all of these are exacerbated when adapting methodologies to the peculiarities of miRNAs (small size, low concentration, similarity among variants, etc). In miRNA-seq protocols, nearly every step introduces a potential for bias, necessitating careful consideration and optimization to ensure accurate and reliable results.

Regarding the miRNA-library preparation, relevant sources of biases are concentrated at this point, but the most important one from our point of view is the ligation bias. To recap, sequencing libraries are constructed by adding specific adapters to the 3’

and 5' ends of the population of molecules to be sequenced. Current methods of miRNA-seq are based on ligation reactions (of RNA and/or DNA) to add the sequencing adapters. Importantly, there are several aspects to consider when performing a ligation reaction with miRNAs. Firstly, the probability of joining two nucleic acid molecules through a ligation reaction depends on the sequences of their ends. Secondly, ligation bias increases as the size of the molecules to be joined decreases. It is a known and accepted issue in conventional NGS, but in the case of miRNA-seq, it becomes exceptionally significant. Thirdly, there are ligase-resistant variants that are challenging to detect using miRNA-seq (Coenen-Stass et al., 2018; Dard-Dascot et al., 2018; Raabe et al., 2014). As a result, some variants are overestimated, while others are underestimated.

Currently, several commercial methods are available to perform NGS of small RNAs (Table 2). However, significant differences have been described in terms of reproducibility, specificity, and sensitivity. Additionally, discrepancies in the performance ranking were observed among studies (Coenen-Stass et al., 2018; Dard-Dascot et al., 2018; Heinicke, Zhong, Zucknick, Breidenbach, Sundaram, T. Flåm, Leithaug, Dalland, Rayner, et al., 2020; Herbert et al., 2020). QIAseq miRNA library kit is the gold standard at the moment, being the one with the best results in the examined metrics of diverse studies (Coenen-Stass et al., 2018; Heinicke, Zhong, Zucknick, Breidenbach, Sundaram, T. Flåm, Leithaug, Dalland, Farmer, et al., 2020). In terms of biases, protocols that do not include ligation reactions have less bias but also produce large amounts of side products (Dard-Dascot et al., 2018). However, none of the kits demonstrated remarkable performance in accurately representing the relative input levels of all miRNAs of the samples examined (Herbert et al., 2020). There is room for improvement in the development of enhanced methods or solutions to precisely quantify miRNA levels, and this is the intention of this study.

Similarly, other sources of error appear when preparing miRNA libraries that indirectly affect the ligation reaction as well. miRNA sequences may exhibit different efficiencies during the ligation process, but also during reverse transcription or PCR amplification. This is closely related to the GC content, where GC-rich or GC-poor sequences may show varying efficiencies. Additionally, secondary miRNA and adapters' structures may impact the performance of diverse processes (Dard-Dascot et al., 2018; Heinicke, Zhong, Zucknick, Breidenbach, Sundaram, T. Flåm, Leithaug, Dalland, Farmer, et al., 2020). In addition, primary and secondary structures also influence poly(A) polymerases (PAP) (mainly derived from *E. coli*) in protocols where polyadenylation is needed (Yehudai-Resheff & Schuster, 2000).

Another problem is the contamination with small RNA fragments from the degradation of larger molecules (especially ribosomal RNA). Again, there is methodological inconsistency. For the same type of sample and extraction method, the proportion of contaminating RNA varies between 35% and 80%, depending on the library preparation method (Coenen-Stass et al., 2018). Thus, sample quantity and quality are pivotal.

Researchers working with miRNA-seq data need to be aware of these potential biases and implement appropriate quality control measures and normalization techniques to mitigate their impact. Careful experimental design and optimization of library preparation protocols are essential to minimize biases and ensure accurate interpretation of miRNA-seq results.

Table 2. *miRNA-library preparation kits commercially available. Information on the company, product, total RNA and/or small RNA input recommended by the manufacturer, and method used in the protocol is shown.*

Company	**Product**	**Platform**	**RNA input**	**Small RNA input**	**Methodology*** (Figure A 2)
QIAGEN	QIAseq miRNA Library Kit	Ion Torrent/ Illumina	1-500 ng	NA	2 sequential adapters ligation, RT, purification, PCR
New England Biolabs	NEBNext® Small RNA Library Prep	Illumina	100 ng - 1µg	NA	2 sequential adapters ligation, RT, PCR
Illumina	TruSeq Small RNA Library Prep Kit	Illumina	1 µg	10-50 ng	2 sequential adapters ligation, RT, PCR
TriLink BioTechnologies	CleanTag™ Small RNA Library Prep Kit	Ion Torrent/ Illumina	1-1000 ng	NA	2 sequential modified-adapters ligation, RT, PCR
Bioo Scientific	NEXTflex® Small RNA-Seq	Illumina	1-2 µg	1-10 µg	1st adapters ligation, purification and adapter inactivation, 2nd adapter ligation, RT, clean-up, PCR
Lexogen	Small RNA-Seq Library Prep Kit	Illumina	0.05-1000 ng	NA	1st adapters ligation, purification, 2nd adapter ligation, RT, PCR
SeqMatic	TailorMix microRNA Sample Preparation Kit	Illumina	1-1000 ng	NA	1st adapters ligation, purification, 2nd adapter ligation, RT, PCR
Takara Bio	SMARTer® miRNA-seq kit	Illumina	100-1000 ng	2-200 ng	Polyadenylation, RT with TSO (5'RACE), PCR
Diagenode	D-Plex Small RNA-seq Kit (Previously CATS)	Illumina/ MGI	Max. 100 ng	Min. 10 pg	Polyadenylation, RT with TSO (5'RACE), PCR

*All methodologies have a last step of library purification.

2. Hypothesis and objectives

Next-generation sequencing (NGS) technologies have revolutionized molecular biology, providing rapid and accurate decoding of genetic information. In contrast to traditional Sanger sequencing, NGS offers high-throughput capabilities, generating massive data simultaneously. Library preparation biases, errors in sequencing processes, read length limitations, and challenges in GC content coverage are important considerations in NGS, among others. Despite these limitations, NGS techniques offer comprehensive insights into biological processes, continuously advancing with innovations. Despite challenges, ongoing advancements and innovations in NGS technologies address limitations. This thesis specifically addresses challenges in T cell receptor (TCR) and microRNA (miRNA) library preparation.

The traditional focus in studying the TCR repertoire has been on the hypervariable CDR3 region, known for its diversity in combating foreign antigens. However, recent studies highlight the importance of CDR1 and 2 regions and small TCR modifications in peptide interaction, adding complexity to understanding TCR functionality. A comprehensive understanding of the entire TCR is deemed crucial for advancing therapies, offering potential precision and efficacy in immune-based treatments.

The T cell receptor (TCR) plays a crucial role in the immune system by enabling T cells to distinguish between self and non-self antigens, orchestrating adaptive immune responses. Notably, they require antigen presentation via HLA-peptide complexes. The TCR repertoire, shaped by V(D)J recombination and thymic selection, contributes to immune diversity. This vast diversity, arising from imprecise junctions between gene segments during the recombination process, and the nucleotide addition or deletion, enhances TCR variability creating the CDR3. CDR1, CDR2, and hypervariable CDR3 regions form the recognition components of TCR, crucial for engaging with HLA-peptide complexes.

Hypothesis: Developing a methodology to prepare TCRβ libraries capable of sequencing the whole TCRβ rearrangement, with the primary goal of achieving a comprehensive identification of the TRVB segment and without ligation reactions, is feasible.

Objectives:

- Design a methodology for TCRβ-library preparation based on nanotechnology to capture full-length TCRβ sequences in a sample (i.e., TCRβ repertoire).
- Design a methodology that does not require ligation reactions to add the adapters.

White adipose tissue (WAT) is now recognized as an organ with several crucial roles in energy homeostasis, endocrinology, thermogenesis, and immunity. Beyond adipocytes, WAT comprises diverse cell types, including leukocytes, fibroblasts, and endothelial cells. Likewise, lymphoid structures, such as FALCs (fat-associated lymphoid clusters), are present in WAT. Leukocytes, constituting 45-50% of cells, persist in lean individuals, challenging the traditional view of their association with obesity-induced inflammation. Remarkably, studies have revealed the presence of tissue-resident memory T cells (Trm) in WAT, where they have advantages in energy utilization during inflammation and exhibit higher proliferative potential and cytokine expression upon activation.

Studying the T-cell receptor repertoire of WAT is crucial, as a significant portion of WAT-resident memory T cells may be reactive to viral antigens. Limited research in murine models has explored viral-specific memory T cells in adipose tissue. The omental fat depot, located in the abdominal cavity, remains understudied (due to the difficulty in obtaining samples) despite its unique characteristics and interactions.

Hypothesis: White adipose tissue (WAT), particularly omental WAT, is a major reservoir of tissue-resident memory T cells against viral infections.

Objectives:

- Analyse the adaptive immunity landscape of the WAT reservoir via massive sequencing of the TCRβ and compare it with the TCRβ repertoire of blood.
- Analyse T cells' reactivity against viral epitopes through ELISpot assays.
- Analyse the phenotype of T cells within the white adipose tissue and blood using flow cytometry.

Changing the focus to small RNA molecules, such as miRNAs, existing methodologies for miRNA analysis include RT-qPCR, microarrays, and miRNA-seq. The latest stands out due to its ability to identify new variants. However, methodological barriers in miRNA-seq, such as ligation bias, need careful consideration. Likewise, one of the main problems in miRNA analysis is the quantity and quality of samples. Commercial kits show varying performances, emphasizing the need for improved methods.

miRNAs play a crucial role in genetic regulation, including gene silencing, and participating in intricate cellular processes. They have intracellular and intercellular functions, packaged in communication vesicles like exosomes and microvesicles. Importantly, certain miRNA variants are associated with chronic human pathologies, such as cancer or neurodegenerative diseases. Circulating miRNAs, stable in blood, are postulated as important biomarkers due to their significant variations in different diseases. The importance of miRNA research lies in gene regulation, disease association, therapeutic targets, understanding cellular processes, and biotechnological applications. Thus, an accurate and reliable analysis is needed.

Hypothesis: The elimination of ligation reactions, a principal source of bias in the preparation of small RNA libraries, does not hinder the development of a new methodology that minimizes sensitivity.

Objectives:

- Design a methodology for small RNA-library preparation based on nanotechnology that does not need ligation reactions to add the adapters.
- Design a methodology capable of capturing small RNA at low concentrations and improving commercial kits' sensitivity.

3. Materials and Methods

3.1. Cohort of patients

In this thesis, samples come from obese patients who underwent bariatric surgery (gastric bypass or gastric sleeve). All participants signed written informed consent and the hospital's ethics committee (Andalucia's Biomedical Research Ethics Committee of Granada) approved the study. Patients were excluded from the cohort if they presented any autoimmune disorder.

The biochemical analyses were conducted within 24 hours using standard procedures at the Clinical Analysis Laboratory of San Cecilio University Hospital in Granada, Spain. HLA genotyping, CMV, and EBV serology were performed using established methods at the Clinical Analyses and Immunology Unit of Virgen de las Nieves University Hospital in Granada, Spain.

From the total cohort, 11 patients are part of the hereafter subcohort 1 and 24 to the subcohort 2. Samples from subcohort 1 were used in Chapter II, and samples from subcohort 2 in Chapter III.

3.2. Sample processing

Four different types of samples were collected for this study: blood, liver, omental white adipose tissue (oWAT), and subcutaneous white adipose tissue (sWAT). Blood samples were collected after 10 hours of fasting, before surgery. Liver and adipose tissue biopsies were obtained during surgery and preserved in PBS (phosphate-buffered saline) at 4 ºC until further processing. oWAT biopsies were obtained close to the stomach, from the greater omentum, whereas sWAT biopsies were obtained near the surgical incision. Approximately, 3 g of oWAT and sWAT is collected. Liver biopsies, of ~10 mg, were obtained using a percutaneous liver biopsy procedure by inserting a surgical needle through the liver tissue.

3.2.1. Sample processing for RNA extraction

Adipose tissue biopsies were examined to remove any small lymph nodes and blood vessels that might be present in the tissue. Liver biopsies were exclusively used for Chapter II. To that end, they were preserved in 500 µl of Qiagen lysis buffer at -20 ºC until total RNA extraction, as well as 300 mg of adipose tissue biopsies and peripheral blood mononuclear cells (PBMCs) isolated by ficoll gradient from 3 ml of blood (3.3. RNA extraction and retrotranscription).

3.2.2. Sample processing for flow cytometry

The remaining portion of each adipose tissue biopsy (around 2 g) was digested in 10 ml of RPMI 1640 medium (Gibco, Billings, MT, USA) supplemented with 2 mg/ml type I collagenase (Sigma, San Luis, MO, USA), 1 mg/ml hyaluronidase (Sigma, San Luis, MO, USA), and 5 mM CaCl2 for 2 hours at 37ºC. The digestion was subsequently diluted with PBS and filtered through a 1 mm sieve. The resulting mixture was then centrifuged at 900 x g for 10 minutes. The resulting pellet was resuspended in PBS,

filtered through a 100 µm filter, and centrifuged once more at 900 x g for 10 minutes. This final pellet contained the stromal vascular fraction (SVF), which was divided for flow cytometry and ELISpot.

For flow cytometry, one-fourth of the SVF was resuspended in 200 µl of antibody staining buffer (PBS, 2% FBS (fetal bovine serum), 0.09% albumin, and 0.05% sodium azide) and combined with an internal standard known as CountBright Absolute Counting Beads (Thermo Fisher, Waltham, MA, USA) (3.17. Flow cytometry). The internal standard comprises autofluorescent beads with diverse emission spectra, offering different size and complexity values than any cell population, thus facilitating differentiation between them.

3.2.3. Sample processing for ELISpot assays

For ELISpot, the remaining portion of the SVF was resuspended in 1 ml of RPMI 1640 medium (Gibco, Billings, MT, USA) to later isolate the mononuclear fraction by ficoll gradient. Similarly, PBMCs were isolated from 3 ml of blood samples, diluted 1:1 with PBS, and then separated using ficoll gradient. The interphase obtained from the SVF and blood was washed with PBS. The resulting pellet from the SVF was used for ELISpot assays (3.11. ELISpot assays).

3.3. RNA extraction and retrotranscription (RT)

Before RNA extraction of tissues (oWAT, sWAT, and liver), these, in lysis buffer, were homogenized manually and centrifuged at maximum speed for 10 minutes. Later, we kept the interphase, where the genetic material is located. Standard silica-membrane columns from the RNeasy Mini Kit by Qiagen (Hilden, Germany) were used for total RNA extraction from isolated PBMCs from blood and the interphase obtained from tissues. RNA quantification was carried out as described in 3.9. RNA and DNA quantification. Then, total RNA, with a minimum concentration of 20 ng/µl, was retrotranscribed to cDNA using the iScript™ cDNA Synthesis Kit provided by Bio-Rad (Hercules, CA, USA).

3.4. Magnetic particles

Magnetic particles need to meet different requirements to be useful for this study: they must exhibit stability under high temperatures and extreme pHs, possess a low sedimentation coefficient, minimize non-specific nucleic acid absorption, and cannot inhibit molecular biology reactions. Two different types of magnetic beads were used during the fabrication of NGS libraries that met those requirements, both from Cytiva (Marlborough, MA, USA). We used Sera-Mag™ Magnetic Carboxylate-Modified Microparticles (MG-CM) for TCRβ libraries (Chapter I) and Sera-Mag™ SpeedBead Carboxylated-Modified [E3] Magnetic Particles for miRNA libraries (Chapter IV). These magnetic beads are ~1µm diameter, uniform, colloidally stable, monodispersed, non-porous, and magnetic. The core of these microparticles contains 40% magnetite (Fe3O4)

and is encapsulated by proprietary polymers. They have carboxylic acid groups on their surface that allow for the covalent amide bond formation using carbodiimide. We can easily couple amino-modified oligonucleotides or molecules with an amino end to the beads. The main difference between the two types of beads lies in the double magnetite layer of Sera-Mag™ SpeedBeads and the composition of the coating material.

3.5. Library preparation

During the preparation of NGS libraries, an important source of bias is found in the ligation reaction, often used to join adaptors (Potapov et al., 2018). Our methodologies are designed to avoid this kind of bias. To achieve this, we leverage the properties that magnetic beads offer us. We use beads to capture our molecules of interest and incorporate de NGS adaptors, effectively eliminating ligation-related biases. Likewise, we based our new technologies on PCRs of different types (multiplex, nested, “elongation”, and usual PCRs) in conjunction with beads.

3.5.1. TCRβ-repertoire libraries

Our methodology can prepare long-read TCRβ-repertoire barcoded libraries without any ligation reaction to avoid this source of bias. In contrast to current techniques that rely on short reads, focusing on the CDR3 region and sequencing only the terminal end of the TRVB segment, our long-read approach covered the entire TRVB segment. In this way, we improved the identification within highly conserved V families, such as V5 and V6 families. Reads also included not only TRVB, TRDB and TRJB segments, but also a portion of the TRCB segment, the CDR-3 loop (known as the functional part of the hypervariable region) and CDR-1 and CDR-2 loops, both provided by individual TRVB segments.

For a complete understanding of the TCRβ library preparation methodology, refer to Chapter I. This method involves an enrichment multiplex PCR, where all TCRs were simultaneously amplified with a set of primers capable of hybridizing with a conserved region from the 5’ UTR of the TRVB segment (near the start codon). Subsequently, the molecules of interest were selectively isolated through a magnetic particle-based process often referred to as "fishing”. After that, the sequences of the Ion S5 adaptors (A and P1) were incorporated by extending the DNA strands covalently bound to the magnetic particles. This step was pivotal in avoiding ligation bias. Lastly, a final PCR was performed to release the DNA from the microparticles. The resulting product was size checked by agarose gel electrophoresis (3.7. Electrophoresis) and purified as explained in 3.8. DNA purification.

3.5.2. Small RNA libraries

This technology receives the name of Double-Tailing Trap (DTT). It can prepare libraries of small RNA, in which miRNA is included. Similar to the previous methodology explained, DTT eliminates any ligation step, in contrast to the majority of the commonly used technologies. It is worth highlighting that the bias associated with the

ligation reaction becomes greater when dealing with small molecules, as is the case with miRNAs.

For a complete comprehension of the miRNA library preparation methodology, refer to Chapter IV. In this case, as the name implies, the technology relies on two tailings of nucleotides to avoid the ligation reaction. The first one is an adenine tailing at the 3' end of the miRNAs, which serves as a binding site to selectively capture the miRNAs with magnetic beads. The second one is a guanine tailing at the 3' end of the retrotranscribed miRNA, priming the molecules for the subsequent elongation of the rest of the sequence. The final step is a PCR to release the DNA from the beads. The resulting product was size checked by agarose gel electrophoresis (3.7. Electrophoresis) and purified as explained in 3.8. DNA purification.

3.6. Standard curves

Standard curves were carried out to assess the efficacy of the enrichment primers' collection from the TCRβ library preparation. The enrichment primers can amplify every V segment simultaneously. We used concentrations of 1000, 200, 40, and 8 fg/µl of initial DNA in the qPCRs. This DNA was obtained from previous PCRs products, which were purified, and concentrations were quantified (3.9. RNA and DNA quantification). The preceding PCRs were performed with cDNA from blood lymphocytes and two primers: a common reverse primer and a forward primer for each V segment. The C_T (cycle threshold) for each initial DNA concentration was calculated in duplicate. To prevent potential re-amplification of waste products, a semi-nested PCR strategy was employed, shifting the position of the reverse primer towards 5' in the qPCR. C_T values and initial DNA were plotted on a semilogarithmic scale base 2. The slope of the obtained line was compared with the ideal theoretical slope (-1) to calculate the amplification efficiency.

3.7. Electrophoresis

Agarose electrophoresis gels were primarily used to assess the size of the libraries. Specifically, to verify TCR libraries (~500 bp long), we used gels at 1.5% agarose with the 100 bp DNA Ladder (Thermo Fisher, Waltham, MA, USA). On the other hand, to verify miRNA libraries (~200 bp long), we used gels at 2.2% agarose with the Low Molecular Weight DNA Ladder (New England BioLabs, Ipswich, MA, USA).

3.8. DNA purification

After diverse techniques through this thesis, we will need a step of DNA purification, which will be mentioned when needed. ProNex® Size-Selective Purification System kit by Promega (Madison, WI, USA) was used with different conditions depending on the purpose. When purifying TCR sequencing libraries (Chapter I) of about 500 bp long, 1.15x of ProNex® Chemistry magnetic particles were used. When purifying miRNA sequencing libraries (Chapter IV) of about 200 bp long, 2.2x of ProNex®

Chemistry magnetic particles were used. DNA quantification was carried out as described in 3.9. RNA and DNA quantification.

3.9. RNA and DNA quantification

We conducted quantification of both RNA and DNA using the Qubit 3 fluorometer by Thermo Fisher (Waltham, MA, USA). The choice of assay kits for quantification depended on the specific nature of the molecule being measured. We used Qubit™ RNA HS Assay Kit (Thermo Fisher, Waltham, MA, USA) for RNA quantification, Qubit™ microRNA Assay Kit (Thermo Fisher, Waltham, MA, USA) for miRNA quantification and Qubit™ dsDNA HS Assay Kit (Thermo Fisher, Waltham, MA, USA) for DNA quantification.

3.10. Next Generation Sequencing (NGS)

NGS was performed on Ion Torrent S5 by Thermo Fisher (Waltham, MA, USA) for TCRβ libraries using Ion-530 chips and the extended version, which can obtain reads up to 600 bp. Sequencing was carried out in the Ion S5 platform of our laboratory.

3.11. ELISpot assays

ELISpot analysis was performed using the ELISpot Path PRO: Human IFN-γ (CMV) kit by Mabtech (Nacka Strand, Sweden). The ELISpot plate, precoated with the 1-D1K monoclonal antibody, was washed four times with sterile PBS and incubated with RPMI 1640 medium (Gibco, Billings, MT, USA) supplemented with 10% FBS (Gibco, Billings, MT, USA) and 100 U/ml of penicillin + 100 µg/ml of streptomycin (Gibco, Billings, MT, USA) for 2 hours.

Following the manufacturer's guidelines, duplicate samples of 250,000 cells (when possible) from the mononuclear fraction of both SVF and PBMCs were seeded into each well. These cells were suspended in 100 µl of RPMI 1640 medium (Gibco, Billings, MT, USA) supplemented with 10% FBS (Gibco, Billings, MT, USA) and 100 U/ml of penicillin + 100 µg/ml of streptomycin (Gibco, Billings, MT, USA) containing 2 µg/ml CMV peptide pool for human CD4 and CD8 T cells. Positive control cells were resuspended in media with a 1:1000 dilution of anti-CD3. Additionally, a negative control was included. However, positive and negative controls were only performed with PBMCs due to a lack of cells in oWAT and sWAT.

The ELISpot plates were then incubated for 16 - 24 hours at 37 °C in a 5% CO2 incubator. Afterwards, plates were washed five times with PBS and later, 100 µl of a diluted 7-B6-1-ALP conjugate (1 µg/ml) was added to each well and incubated for 2 hours at room temperature. After five more PBS washes, 100 µl of filtered BCIP/NBT-plus solution were added to each well and left until spots emerged.

Finally, plates were washed with tap water to stop the colour reaction. Once the plate had dried, photographs were taken using a FujiFilm X-T4 camera with a Canon

macro lens of 100 mm. Subsequently, the spots were counted, and their area was measured using ImageJ software.

3.12. Transformation of competent cells, plasmid purification and Sanger sequencing

Throughout this study, multiple checkpoints were essential to ensure the accuracy and error-free nature of the sequences. Due to the nature of this thesis, following the final PCR in the preparation of NGS libraries, we obtained different molecules that needed to be isolated for subsequent Sanger sequencing. To do so, DNA is ligated into a pGEM®-T Easy vector (depicted in Figure 6) from Promega (Madison, WI, USA) after the PCR product purification. This ligation is easily carried out using the pGEM®-T Easy Vector Systems kit, taking advantage of the two free thymines at the 5' ends of the vector and the adenines left by the *Taq* DNA polymerase at the 3' ends of the DNA molecules.

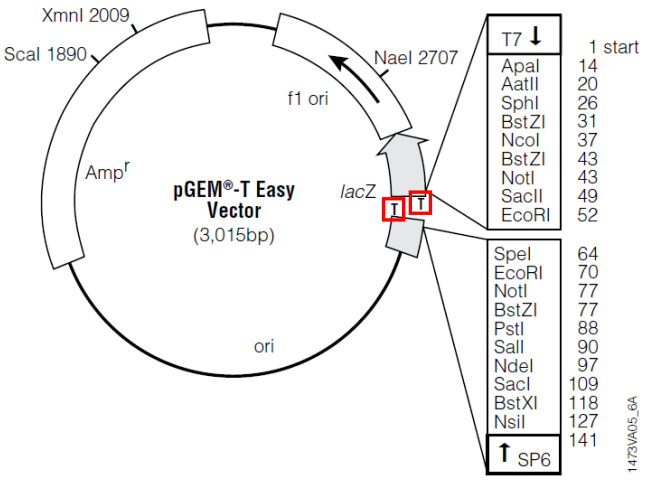

Figure 6. *pGEM®-T Easy vector map. The thymines used to ligate the DNA inserts are marked with red squares. The operon lactose is represented in grey. Image modified from Promega.*

3.12.1. Competent cells' transformation

Following the ligation reaction, the transformation of competent cells is performed. The product of the ligation reaction is suspended in 40 µl of JM109 competent cells (Promega, Madison, WI, USA), which have been previously slow-thawed on ice. This suspension is immediately placed on ice for 30 minutes. To facilitate the insertion of the plasmid into the competent cells, a brief heat shock of 40 seconds at 42 ºC is applied to create small pores in the cell membrane. Tubes are immediately placed again on ice for 5 minutes. Later, a resuspension with room-temperature 500 µl of SOC medium (Condalab, Madrid, Spain) and an incubation of 1h at 37 ºC with gentle shaking will recover the competent cells. Lastly, 200 µl of the transformation reaction are plated on the antibiotic plate. Each antibiotic plate consists of 30 ml of LB-Agar medium (Condalab, Madrid, Spain), supplemented with 80 µg/ml of X-Gal (Promega, Madison, WI, USA), 30 µl of IPTG 0,5M (Promega, Madison, WI, USA) and 30 µl of ampicillin

at a concentration of 100 mg/ml (Sigma, San Luis, MO, USA). Antibiotic plates are then incubated overnight at 37 ºC.

After overnight incubation, positive colonies are selected. In the absence of the insert, the lacZ operon is reconstituted and is capable of metabolizing X-Gal, producing a blue-colored precipitate (negative colonies). In the presence of the insert, the lacZ operon remains inactive, resulting in white colonies (positive colonies). Each isolated colony is cultured overnight at 37 ºC in 5 or 50 ml of SOC medium (Condalab, Madrid, Spain) supplemented with 5 or 50 µl of ampicillin at a concentration of 100 mg/ml (Sigma, San Luis, MO, USA), respectively, depending on the specific requirements.

3.12.2. Plasmids purification

To extract the plasmid from the competent cells, we performed either a miniprep or midiprep procedure (depending on the purpose). Normally, we will perform a miniprep in the first place to check through Sanger sequencing our sequence of interest. Minipreps are carried out using 3 ml of the cultivated colony with Wizard® Plus SV Minipreps DNA Purification Systems by Promega (Madison, WI, USA). Once the sequence of interest is verified, we proceed to scale up the colony to a volume of 50 ml, facilitating a subsequent midiprep with Wizard® Plus Midipreps DNA Purification System by Promega (Madison, WI, USA). DNA quantification was carried out as described in 3.9. RNA and DNA quantification.

3.12.3. Sanger sequencing

Sanger sequencing was carried out at the Genomics Unit from *Instituto de Parasitología y Biomedicina López-Neyra*. Samples were prepared as follows: a minimum of 400 ng of DNA and 6.4 pmol of the primer in a total volume of 12 µl. Later, Sanger sequencing results were displayed and examined with Geneious Prime (New Zealand).

3.13. Synthetic TCRβs

Two synthetic constructions of TCRβ were purchased from Integrated DNA Technologies (IDT, Coralville, IA, USA). One of them resembled a typical TCRβ structure with a variable segment (TRBV6-1), a joining segment (TRJB2-7), a hypervariable part (HVR), and a constant segment (TRCB2) (hereafter referred to as "Canonical TCRβ"). The second construction was identical but had a peculiarity in its V segment (hereafter referred to as "Non-Canonical TCRβ"). Specifically, this V segment constituted a fusion between two different TRBV segments (TRBV6-1 and TRBV27), maintaining the reading frame and approximate length of a conventional TCRβ. Additionally, both possessed the elements described below:

- Firstly, a restriction site for the EcoRI (gaattc) enzyme with a small extension (aat) at the beginning of the sequence.
- Secondly, the Kozak sequence (in purple) which contained the initial ATG (in bold).
- Thirdly, the TCRβ sequence *per se* (TRBV-HVR-TRJB-TRCB).
- Finally, a stop codon (TAG, in red) and a restriction site for the NotI (gcggccgc) enzyme with a small terminal extension (atgaaggct) at the end of the constant region, TRCB. It is worth noting that the Kozak sequence facilitates the recognition of the initiation codon ATG by the ribosome, stressing its importance in the sequence design of the constructs.

aatgaattcgccgccacc**atg**-TRVB-HVR-TRJB-TRCB-taggcggccgcatgaaggct

It is worth mentioning that at be beginning of the TRCB segment exists a restriction site for the EcoRI enzyme. For this reason, we introduced a silent mutation in the synthetic TCRβs to eliminate it (without altering the amino acid sequence).

Likewise, we created a "Non-Productive TCRβ" introducing the stop codon and the NotI restriction site after the TRJB segment of the "Non-Canonical TCRβ". To do so, we performed a PCR with the forward primer (tcaatgaattcgccgccaccatgggccccca) and the reserve primer (gggccgggcaccaggctcacggtcacatag**gcggccgc**atgaaggct), which included the stop codon (in red) and the NotI restriction site (underlined).

To ensure all constructions were accurate, we followed the protocol in section 3.12. Transformation of competent cells, plasmid purification and Sanger sequencing. In this occasion, constructs were integrated into the pCI expression vector (depicted in Figure 7) using the EcoRI and NotI restriction sites during the ligation reaction. Previously, constructs underwent a digestion reaction with EcoRI and NotI restriction enzymes from Promega (Madison, WI, USA) following manufacturer's instructions. Once their sequence was verified through Sanger sequencing, a midiprep procedure was performed and the resulting plasmids were used for transfection into the Jurkat-KOβ cell line (3.16. Cells transfection).

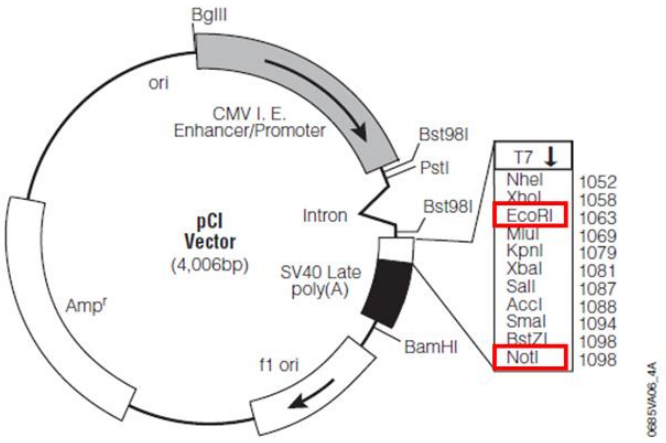

Figure 7. *pCI expression vector map. The EcoRI and NotI restriction sites used to ligate the DNA inserts are marked with red squares. Image modified from Promega.*

3.14. Cell culture

Jurkat cells are a well-known and widely used immortalized cell line of human T lymphocyte cancer cells. They have played a crucial role in various areas of biomedical research, particularly in immunology, cancer biology, and molecular biology. Jurkat cells were derived in the late 1970s from the peripheral blood of a 14-year-old male with T-cell leukemia.

In this study, Jurkat cells (hereafter, Jurkat-WT – wildtype) and one derivative from those (hereafter, Jurkat-KOβ – knock-out) were used. Jurkat-WT were purchased from the *Centro de Instrumentación Científica* of the University of Granada. On the other hand, Jurkat-KOβ were obtained from the knock-out by a stop codon in the *tcrβ* locus (near the start codon) from the Jurkat-WT cell culture. This knock-out was performed by DNA Breaks Laboratory (Lisbon, Portugal). Both cell cultures were cultivated with RPMI 1640 medium (Gibco, Billings, MT, USA) supplemented with 10% FBS (Gibco, Billings, MT, USA) and 100 U/ml of penicillin + 100 µg/ml of streptomycin (Gibco, Billings, MT, USA).

3.15. Confocal microscopy

To characterize the cell cultures and verify that the knockout was correct, confocal microscopy was performed. For image acquisition, we used the ZEISS LSM900 Airyscan 2 microscope (Oberkochen, Germany) with the Zen Blue 3.0 software. Characteristics of fluorochromes and lasers are shown in Table A 1. Specifically, we used the LD LCI Plan-Achromat 40x/1.2 imm korr DIC M27 objective. Images were obtained performing a Z-Stack of 2 µm intervals. The total length of the images was 20 µm.

For this purpose, 125,000 cells were counted and centrifuged for 10 minutes at 3,500 rpm. The obtained pellet was resuspended in 100 µl of PBS with the corresponding antibodies and incubated for 20 minutes at room temperature. In this case, the antibodies used were Hoescht 33342 (Santa Cruz Biotechnology, Dallas, TX, USA) to stain the cell nucleus, anti-CD45 Tx-Red (BioLegend, San Diego, CA, USA) to stain the cell membrane, and anti-CD3 FITC (BioLegend, San Diego, CA, USA) to label the CD3-TCR complex. Then, cells were washed twice with 700 microliters of diluted 1:10 BD FACS™ Lysing Solution 10X Concentrate (BD Biosciences, Franklin Lakes, NJ, USA) to eliminate as many dead cells and debris as possible. After each wash, cells were centrifugated for 5 minutes at 4,000 rpm. Following the second wash, the pellet was resuspended in 200 microliters of PBS and placed in the well of the µ-Slide 8 Well high Glass Bottom from Ibidi (Gräfelfing, Germany). To observe these cells under the microscope, a waiting time of 5 minutes was allowed for them to sediment.

3.16. Cells transfection

We carried out the transfection of Jurkat-KOβ cells using the Gene Pulser Xcell system from Bio-Rad (Hercules, CA, USA). The transfection followed a protocol of one

pulse with the following parameters: 250 V of voltage, 350 µF of capacitance, and 1000 Ω of resistance. Additionally, Jurkat-WT cells were also subjected to the same transfection procedure, serving as a control group.

Following diverse experiments, we determined the optimal conditions for electroporation, which involved 750000 cells and 5µg of pCI plasmids containing one of the synthetic constructions ("Canonical TCRβ", "Non-Canonical TCRβ", or "Non-Productive TCRβ"). Transfection was conducted in a total volume of 400 µl of RPMI 1640 medium (Gibco, Billings, MT, USA) in the 0.4 cm Gene Pulser/MicroPulser Electroporation cuvettes by Bio-Rad (Hercules, CA, USA).

After electroporation, transfected cells were resuspended in 5 ml of RPMI 1640 medium (Gibco, Billings, MT, USA) supplemented with 10% FBS (Gibco, Billings, MT, USA) and 100 U/ml of penicillin + 100 µg/ml of streptomycin (Gibco, Billings, MT, USA). These cells were then cultured in a T25 flask. Following a 24 hours-incubation, we subjected the cells to flow cytometry analysis (3.17.2. Flow cytometry of Jurkat cells) to assess the transfection efficiency.

3.17. Flow cytometry

Flow cytometry was performed using BD FACS ARIA IIIu equipment (BD, Franklin Lakes, NJ, USA) with the FACS Diva 8 software, and Cytek Northern Lights (Cytek, Fremont, CA, USA) with the Spectroflo software. Characteristics of fluorochromes and lasers of each flow cytometer are shown in Table A 2 and Table A 3. The first one was used to characterize cell populations from patients whereas the second one was used to check the efficiency of Jurkat transfection. The acquired data were represented on a logarithmic scale.

3.17.1. Flow cytometry of patients' samples

Patients' SVF and fresh blood were labelled with either 2 ml of controls or fluorophore-conjugated antibodies at room temperature for 20 minutes. Controls (blanks, without staining and isotype controls) were used to determine sytometric parameters. Subsequently, cells were fixed, and the erythrocytes were lysed using 1 ml of BD FACS Lysing Solution for 30 minutes. Later, samples were centrifuged for 10 minutes at 3500 rpm, and the resulting pellets were resuspended in 200 µl of PBS. Then, samples were stored at 4 ºC overnight.

An internal standard (CountBright Absolute Counting Beads de Themo Fisher, Waltham, MA, USA) was employed to calculate the number of cells/mg of tissue in the FACS ARIA instrument. The internal standard consists in a known number of autofluorescent beads that are added to a certain quantity of processed tissue. Two different panels with fluorescent-conjugated antibodies were used for the analysis (Table 3). Panel 1 discriminates subpopulations of T cells regarding their naïve or memory phenotype, while Panel 2 discriminates subpopulations of T cells regarding the expression of memory markers associated with tissue.

3.17.2. Flow cytometry of Jurkat cells

Cells were labelled with either 200 µl of controls (blank or isotypes) or fluorophore-conjugated antibodies at room temperature for 20 minutes. Subsequently, cells were washed with 1.5 ml of PBS and centrifuged for 10 minutes at 3500 rpm. The resulting pellet underwent a second washing using the same procedure. Finally, pellets were resuspended in 200 µl of PBS. For this experiment, cells were stained with anti-CD3 BV510 (clone OKT3, BioLegend, San Diego, CA, USA) and Zombie Red Fixable Viability Kit (BioLegend, San Diego, CA, USA).

Table 3. Antibodies used for flow cytometry of patients in Panel 1 and Panel 2.

Panel 1: Naïve *versus* memory subpopulations			
Target	**Fluorochrome**	**Clone**	**Manufacturer**
CD3	BB515	HIT3a	BD Biosciences
CD4	APC	RPA-T4	BD Biosciences
CD8	BV711	SK1	BioLegend
CCR7	PE-Cy7	G043H7	BioLegend
CD45RA	BV421	HI100	BioLegend
Panel 2: Tissue memory subpopulations			
Target	**Fluorochrome**	**Clone**	**Manufacturer**
CD3	BV510	OKT3	BioLegend
CD4	FITC	RPA-T4	BD Biosciences
CD8	BV711	SK1	BioLegend
CD49a	APC	TS2/7	BioLegend
CD69	PE-Cy7	FN50	BD Biosciences
CD103	PE	Ber-ACT8	BioLegend

3.18. Statistical analysis

Statistical analyses were carried out using or GraphPad 8 (La Jolla, CA, USA). The normality of the data was evaluated using both Normality and Lognormality tests, leading to the decision to perform non-parametric tests in Chapter I and parametric tests in Chapter III. Multiple group comparisons for Chapter I were conducted using the Kruskal-Wallis test, and subsequent pairwise comparisons were performed using the two-sided Wilcoxon test. In Chapter III, multiple group comparisons were carried out using the ANOVA test, and pair comparisons were conducted using the *t*-test.

3.19. Bioinformatic analyses

Bioinformatic analyses were carried out in collaboration with different researchers with expertise in this field. Analyses from Chapter II were performed by María S. Benítez Cantos from the Department of Biochemistry and Molecular Biology III and Immunology of the University of Granada. Analyses from Chapter III were performed by A. Jesús Muñoz Pajares from the Department of Genetics of the University of Granada.

4. Results

Chapter I:

T cell receptor β (TCRβ) NGS-library preparation

4.1. Chapter I: TCRβ-repertoire NGS-library preparation

The complete process of preparing TCRβ libraries for NGS (Figure A 3) is detailed in this chapter. It must be noted that all the steps were carried out with the *Pfu* DNA polymerase (Promega, Madison, WI, USA) due to its high thermal stability and its low error rate (1-2 × 10^{-6}, McInerney et al., 2014). It is also important to highlight that this technology is designed for Ion S5 sequencing platform. Oligonucleotides and templates used for this procedure were purchased from IDT (Coralville, IA, USA) and they are shown in Table A 4. Additionally, PCR conditions are summarized in Table A 5.

4.1.1. Functionalization of TCRβ-beads

For this procedure, diverse beads were tested, both commercial and non-industrial crafted by collaborators. The best results were obtained by the Sera-Mag™ Magnetic Carboxylate-Modified Microparticles (MG-CM) (Cytiva, Marlborough, MA, USA). Hereafter, these microparticles, once functionalized, will be referred to as "TCRβ-beads". Other beads tested exhibited either incompatibility with the procedure or lower efficiency than the chosen ones. The predominant issue was the instability during the alkaline washes, leading to a structural breakdown of the particles.

Sera-Mag™ Magnetic Carboxylate-Modified Microparticles are spheres of ~1 µm diameter that have carboxylic acid groups on their surface. The presence of carboxylic groups on their surface, negatively charged, ensures their stability, and facilitates the covalent coupling of amino groups. These beads are well-suited for our technology due to their elevated magnetism, enhanced mobility, low sedimentation rate, and compatibility with high temperatures and extreme pH conditions.

Microparticles need to be functionalized to proceed with the technology. This is the first step of the preparation of NGS libraries of TCRβ. We coupled oligonucleotides (named A-BCX-R, Table A 4) that include the sequence of adapter A, one barcode (BC) and a complementary sequence to the TRCB segment at its 3' end (Figure 8). The latest will be the part "fishing" the molecules of interest (4.1.3. "Fishing" PCR). It is worth noting that TRCB1 and TRCB2 share this same sequence so that both hybridize with the oligonucleotide.

To prevent the oligonucleotide's nitrogen bases from interacting with the carboxylic groups on the beads' surface, it is necessary to hybridize it with its complementary. For this aim, we prepared a stock of hybridised oligonucleotides incubating 20 µl each of the forward and reverse oligonucleotides (10 nmol/µl) with 10 µl of PBS and 10 µl of 1.5 M NaCl during the "decreasing temperature protocol" (temperature declines from 70 to 50 ºC in 5 ºC decreases, with each stage lasting 5 minutes).

Each hybridised oligonucleotide with its corresponding BC (1 nmol per mg of particles) is coupled via a 5' amino (NH2) modification in the reverse primer to the beads (10 mg/ml). The reaction medium was 1M NaCl + 100 mM MES (2-(N-morpholino) ethanesulfonic acid) at pH 5 supplemented with 1.25 M of EDAC (1-ethyl-3-(3-

dimethylaminopropyl) carbodiimide hydrochloride). EDAC establishes the amide bond between the 5' amino end of the oligonucleotides by activating the carboxylic groups on the surface of the microparticles. The reaction needs to be vortex for 10 min and then incubated for 4 hours with continuous rotation.

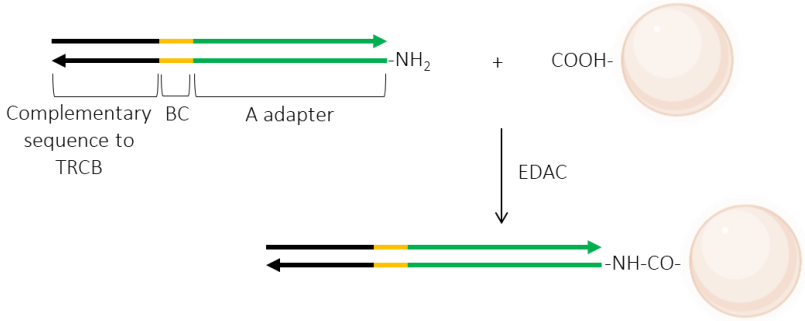

Figure 8. *Diagram of a functionalized TCRβ-bead. Arrowhead simulates the 3' end of the oligonucleotide. It must be noted this is just a scheme, beads are covered all over with the same oligonucleotide. BC: barcode.*

After the incubation, we performed an alkaline wash with NaOH 200 mM to remove any DNA strands not covalently bound to the beads and reagents excess (Figure 9). To do so, we placed the beads on a magnet, discard the supernatant and resuspend them with NaOH. Likewise, we restored the pH with two additional washes in TRIS-HCl 200 mM and 50 mM pH 7.4. Lastly, particles were resuspended in TRIS-HCl 50 mM pH 7.4.

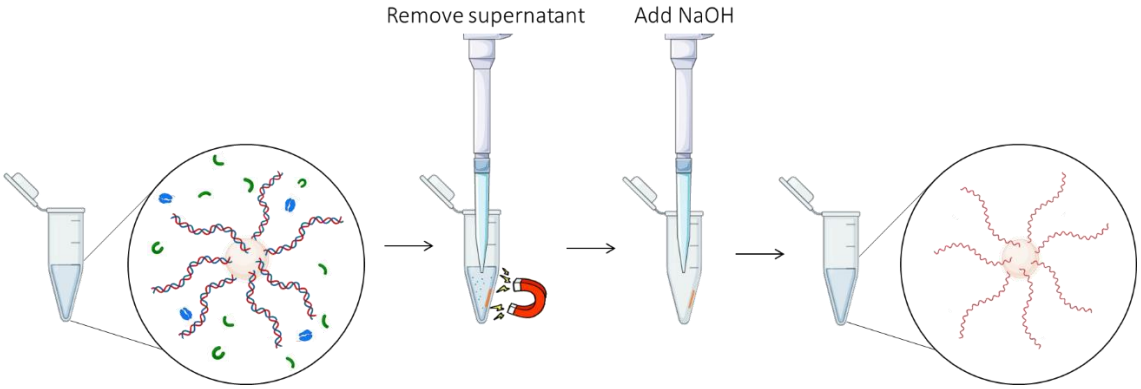

Figure 9. *Diagram of an alkaline wash.*

The alkaline washing step is critical. It ensures that only single-stranded DNA molecules, attached to the TCRβ-beads, remain for the following reaction. Hence, these single-stranded DNA molecules are prepared to acquire any new sequence through their 3' end. Moreover, washing steps allow us to switch reaction reagents easily, facilitating the continuous work on the particles' surface.

4.1.2. Enrichment PCR

The first step of the TCRβ library preparation *per se* is an enrichment multiplex PCR (Figure 10). To amplify every TRVB segment simultaneously, we designed a

collection of 48 forward primers. On the other hand, we only used 1 reverse primer with a common sequence for the two TRCB segments. The enrichment PCR was carried out following the manufacturer's instructions with 75 nM of the forward collection primers, 150 nM of the reverse primer, 1.5 mM of $MgCl_2$, 0.2 mM of dNTPs, and 5µl of cDNA in a total volume of 20 µl. The protocol consisted of a pre-denaturation of 2 min at 95 ºC, followed by 15 cycles of denaturation at 95 °C for 20 s, annealing at 62 °C for 30 s, and extension at 74 °C for 2 min, and a final extra extension of 5 min at 74 ºC.

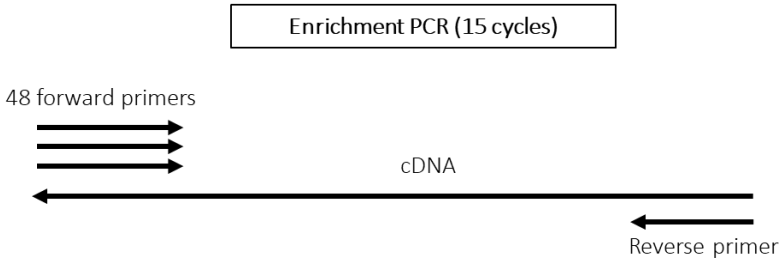

Figure 10. *Diagram of the enrichment PCR. Arrowheads simulate the 3' ends.*

4.1.2.1. Designing criteria for the enrichment primers' collection

The main goal when designing this collection of primers was to 1) amplify all functional V segments with the smallest number of primers, and 2) minimize the probability of dimerization. For this purpose, the collection was designed over a highly conserved region in the 5'-UTR of TRVB segments, near the start codon, the Kozak sequence. In the end, this collection amplifies the 54 functional TRVB segments, but also 11 TRVB segments classified as pseudogenes.

We considered four important features to design the enrichment primers' collection to achieve the goals previously described:

- Relatively small size, between 18 and 21 nucleotides (19.10 nt ± 0.66).
- Few differences in nucleotide composition. Each primer hybridizes with a similar number of H-bonds, between 48 and 50 (49.19 H-bonds ± 1.02).
- Elimination of all "elongation triplets" in the primer sequences.
- Preferential location in conserved regions, near the start codon in the 5' UTR.

A productive hybridization only occurs when there is a 3' end extendable by a DNA polymerase. To avoid the primer polymerization phenomenon, it is critical to control the sequence at the 3' end. In this study, we defined the concept of "elongation triplet" as any primer sequence complementary to the last three nucleotides of the 3' end of any of the primers (including itself) present in the reaction.

The use of a highly conserved region when designing most forward primers not only allowed the reduction of the number of primers in the collection due to its similarity, but it also facilitated the control of the 3' ends (Table 4). Thus, reducing the probability of finding "elongation triplets", hybridizing and producing an effective elongation, resulting in undesired artifacts of PCR. In fact, the enrichment primer's collection is

composed of only 4 endings: "cat" (39 primers), "tat" (4 primers), "aaa" (3 primers), and "cgg" (2 primers).

Table 4. *Information on the number of TRVB segments amplified by each primer, their 3' end sequence and the position of the first 3'end nucleotide relative to the adenine of the "star" codon. Adenine on the ATG start codon represents position 0. Positive numbers are displacements towards 3' and negative numbers towards 5'.*

Primer	Number of V segments	3' end sequence	Position relative to the start codon
1.1-F	1	cat	+1
1.2-F	1		
1.3-F	2		
1.4-F	3		
1.5-F	7		
1.6-F	1		
1.7-F	1		
1.8-F	1		
1.9-F	1		
1.10-F	1		
2.1-F	4		
2.2-F	1		
2.3-F	1		
2.4-F	1		
3.1-F	1		
3.2-F	2		
3.3-F	1		
3.4-F	1	tat	+24
4.1-F	1	cat	+1
4.2-F	1		
5-F	1		
6-F	1	cat	-4
7-F	1	cat	+1
8-F	3	cgg	-19
9.1-F	2	aaa	-1
9.2-F	1		
10.1-F	1	cat	+1
10.2-F	1		
11.1-F	1	cat	-10
11.2-F	1		
12-F	1	cat	+1
13-F	1	cat	+54
14-F	1	cat	+28
15-F	1	cat	+1
16-F	1	cat	-14
17-F	1	tat	+39
18-F	1	aaa	-12

PG1.1-F	1	cat	+32
PG3.2-F	1	cat	+1
PG5.2-F	1		
PG7.5-F	1		
PG8.2-F	1		
PG12.1-F	1		
PG12.2-F	1		
PG16-F	1		
PG21.1-F	1	cat	+31
PG22.1-F	1	tat	+23
PG26-F	1	cat	+31

However, point mutations were introduced in the sequences of four primers to avoid elongation triplets. Additionally, one primer was degenerated due to the strong homology between sequences of different TRVB segments, allowing the amplification of all those segments with a single primer. Both primers, original and mutated, were tested to ensure their efficiencies remained consistent (Table 5).

Table 5. *Original and mutated/degenerated sequences of primers of the enrichment collection and their CTs ± standard deviation of 3 replicates. In red, point mutations and the degenerated base (B=G +T + C).*

Primer	Original sequence	C_T	Mutated/degenerated sequence	C_T
2.4-F	gcctgccccttttgtgccat	24.28 ± 0.03	gcctgcccc**a**ttgtgccat	24.36 ± 0.62
4.2-F	cgtcatgcagcatctgccat	25.39 ± 0.32	cgtc**t**tgcagcatctgccat	25.45 ± 0.17
7-F	ttctttcttcaaagcagccat	30.39 ± 0.14	ttc**c**ttcttcaaagcagccat	29.66 ± 0.38
17-F	aaggcccccttttgcactat	25.12 ± 0.52	aaggcccccattgcactat	24.40 ± 0.13
1.5-F	-		cctg**B**cctgaccctgccat	25.09 ± 0.38

4.1.2.2. Standard curves of the enrichment primers' collection

It is important to highlight that standard curves were performed exclusively for the primers that amplified functional TRVB segments. Primers amplifying pseudogenes were included in the collection after segment TRVB23-1 (which primer is 14-F) was classified as a pseudogene at NCBI (National Center for Biotechnology Information), and as an ORF (open reading frame) "because ngt replaced by nat (gat) in donor-splice" at IMGT (the international ImMunoGeneTics information system, Giudicelli, 2006). Although its classification indicated that TRBV23-1 is not a functional segment, we could find patients amplifying this segment. For this reason, we decided to design primers for the rest of the segments classified as pseudogenes.

The 37 primers from the collection, that amplified functional TRVB segments, were individually analysed to confirm their proper functionality. Verification tests were conducted using qPCR, with cDNA obtained from circulating lymphocytes of the same individual and employing only two primers: the forward primer under analysis and a reverse primer, common to all tests, located in the conserved region of segment TRCB.

Firstly, it was tested that each primer amplifies separately, yielding positive results. The quality and size of the PCR product were verified through 1.5% agarose gel electrophoresis. Secondly, the amplification efficiency was determined through standard curves (Table 6). Primers designed for the non-functional TRVB segments were not individually analysed since not all individuals presented those segments.

Table 6. *Individual efficiency of forward primers from the enrichment collection was calculated through standard curves. The average value is 0.96 ± 0.05.*

Primer	Efficiency	Primer	Efficiency	Primer	Efficiency	Primer	Efficiency
1.1-F	1.00	**2.1-F**	0.98	**5-F**	1.00	**12-F**	0.91
1.2-F	0.99	**2.2-F**	0.91	**6-F**	0.89	**13-F**	0.98
1.3-F	1.00	**2.3-F**	1.00	**7-F**	1.00	**14-F**	1.00
1.4-F	0.90	**2.4-F**	1.00	**8-F**	0.98	**15-F**	0.98
1.5-F	0.92	**3.1-F**	1.00	**9.1-F**	0.99	**16-F**	0.94
1.6-F	0.97	**3.2-F**	0.89	**9.2-F**	0.99	**17-F**	0.89
1.7-F	0.89	**3.3-F**	0.97	**10.1-F**	0.97	**18-F**	1.00
1.8-F	0.91	**3.4-F**	0.96	**10.2-F**	0.87		
1.9-F	0.89	**4.1-F**	1.00	**11.1-F**	1.00		
1.10-F	1.00	**4.2-F**	0.90	**11.2-F**	0.95		

Furthermore, the entire collection was assessed collectively to verify their optimal hybridization temperature. A temperature gradient was conducted with temperatures of 70 ºC, 67 ºC, 62 ºC and 57ºC. Optimal amplification efficiency was observed at 62ºC and 57 ºC, with low amplification at 67 ºC and no amplification detected at 70ºC (Figure 11). The lowest temperature tested yielded satisfactory results but with a higher occurrence of primer dimers. Consequently, we selected 62ºC as the preferred temperature due to its balanced performance.

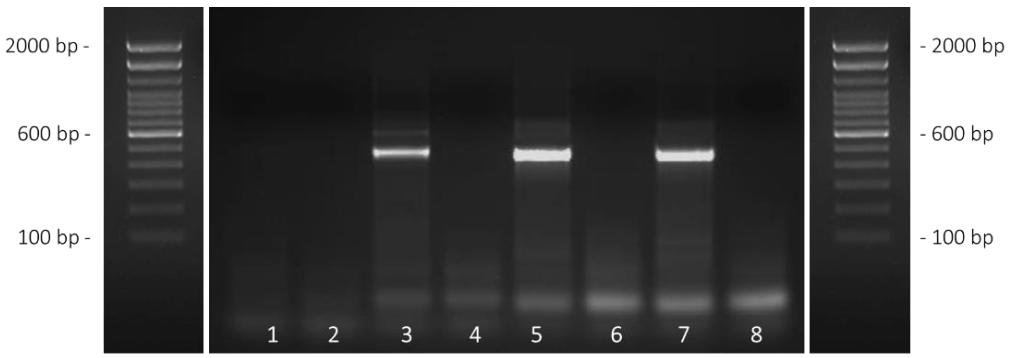

Figure 11. *Evaluation of the optimal hybridization temperature for the enrichment collection. 1.5% agarose. Modified image. Lanes 1-2: 70 °C – 70 ºC blank. Lanes 3-4: 67.7 °C – 67.7 ºC blank. Lanes 5-6: 62 °C – 62 ºC blank. Lanes 7-8: 57 °C – 57ºC blank.*

4.1.3. "Fishing" PCR

After the enrichment PCR, 12µl of functionalized particles were added to its product. Three additional PCR cycles were performed, with a pre-denaturation of 2 min at 95 ºC and a final extra extension of 5 min at 74 ºC. Cycles consisted of denaturation at 95 °C for 20 s, annealing at 68 °C for 1 min, and extension at 74 °C for 2 min. The oligonucleotides in the TCRβ-beads acted as nested reverse primers, elongating, and acquiring the reverse sequence of the TCRβ population from the enrichment PCR (Figure 12). Thus, leaving behind all previous waste PCR products. After the "fishing" PCR, another alkaline wash was performed (Figure 9).

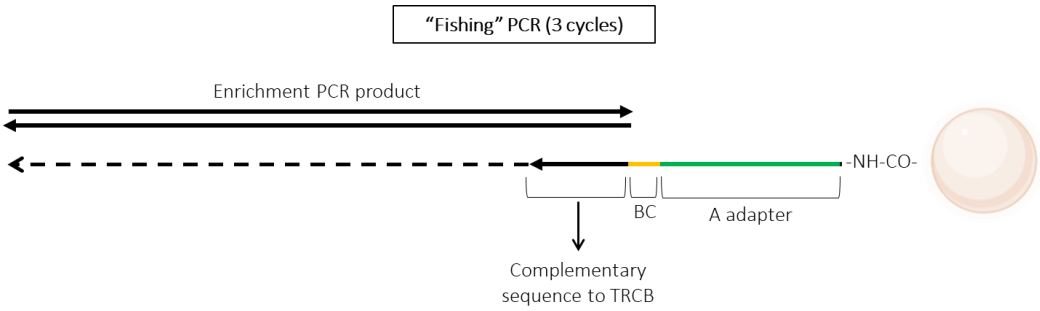

Figure 12. *Diagram of the "fishing" PCR. Arrowheads simulate the 3' ends. It must be noted this is just a scheme, beads are covered all over with DNA molecules. TRCB: C segment from the TCRβ, BC: barcode.*

4.1.4. Including the P1 adapter into the libraries

The next step of the library preparation is the elongation of the second adapter for Ion S5, which is P1. For that purpose, we designed the second collection of 48 oligonucleotides (Table A 4). In this case, they served as templates. The sequence on their 3' end hybridized with each primer from the enrichment collection, thus with all the amplified TRVB segments. The collection provided the P1 adapter sequence on their 5' end, which will be the template for the DNA strands covalently bound to the TCRβ-beads. Furthermore, to avoid the elongation of the templates, an inverted thymine at the 3' end was included (Figure 13).

The process to include any sequence of interest through a simple "extension PCR" on the surface of the magnetic bead was patented by our group (patent number: P201830701). This step replaced the ligation reaction commonly used to add the NGS adapters. By an "extension PCR" of 3 cycles, the single-stranded DNA completed the P1 adapter sequence. TCRβ-beads were resuspended in the "extension PCR" reaction volume (100 µl), which followed the manufacturer's instructions with 75 nM of the template collection, 1.5 mM of $MgCl_2$ and 0.2 mM of dNTPs. The protocol consisted of a pre-denaturation of 2 min at 95 ºC, followed by 3 cycles of denaturation at 95 °C for 20 s, annealing at 63 °C for 3 min, and extension at 74 °C for 2 min, and a final extra extension of 5 min at 74 ºC. Subsequently, an alkaline washing step (Figure 9) removed the collection of templates and any other reactive or waste products.

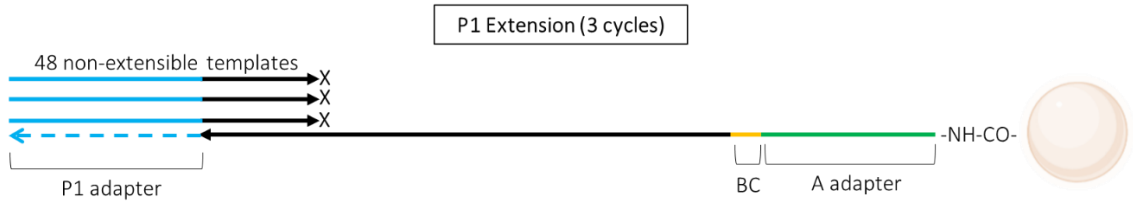

Figure 13. *Diagram of the P1 extension PCR. Arrowheads simulate the 3' ends. It must be noted this is just a scheme, beads are covered all over with DNA molecules. BC: barcode.*

4.1.5. Releasing the complete library from the TCRβ-beads

The last step involves a simple PCR. To release the final TCRβ libraries from the TCRβ-beads we employed a pair of primers, each of them targeting one of the adapters (Figure 14). This ensures that only complete libraries will be obtained at the end of the methodology. It must be noted that the forward primer has three additional nucleotides at its 5' end which will fully complete the P1 adapter. The final PCR was carried out following the manufacturer's instructions with 200 nM of each primer, 1.5 mM of $MgCl_2$ and 0.2 mM of dNTPs in a total volume of 30 µl. The protocol consisted of a pre-denaturation of 2 min at 95 ºC, followed by 40 cycles of denaturation at 95 °C for 20 s, annealing at 65 °C for 30 s, and extension at 74 °C for 2 min, and a final extra extension of 5 min at 74 ºC.

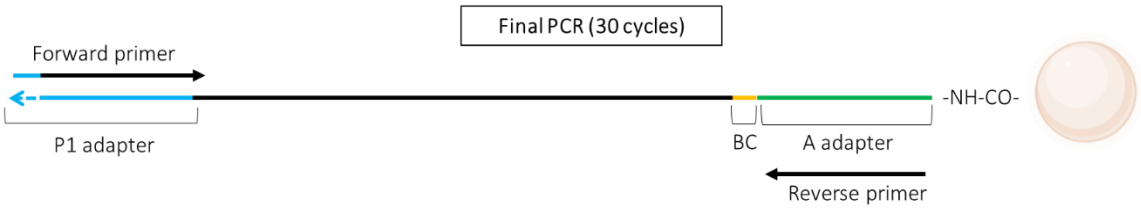

Figure 14. *Diagram of the final PCR. Arrowheads simulate the 3' ends. It must be noted this is just a scheme, beads are covered all over with DNA molecules. BC: barcode.*

Final amplicons, from 5' to 3', are composed of the P1 adapter; in some cases, a variable number of nucleotides from the 5' UTR (depending on where the enrichment primer hybridizes); the complete TRVB segment, which comprises exons 1 and 2 (including CDR1 and CDR2); the hypervariable region, CDR3; the complete TRJB segment; the 5' beginning of TRCB segment; a barcode; and the A adapter (Figure 15).

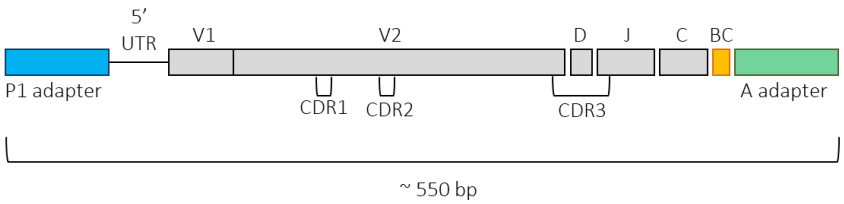

Figure 15. *Diagram of the final amplicons obtained by the TCRβ library preparation. UTR: untranslated region, V1: exon 1 from TRVB (variable) segment, V2: exon 2 from TRVB segment, CDR: complementary determining region, D: TRDB (diversity) segment, J: TRJB (joining) segment, C: TRCB (constant) segment, BC: barcode. bp: base pair.*

Chapter II:

Human adipose tissue as a major reservoir of cytomegalovirus-reactive memory T cells

4.2. Chapter II: Human adipose tissue as a major reservoir of cytomegalovirus-reactive memory T cells

Throughout this chapter, we present the results published in Redruello-Romero et al., 2023. In this study, we collected blood, omental white adipose tissue (oWAT), subcutaneous white adipose tissue (sWAT), and liver samples from 11 obese patients to characterize the adaptive immune landscape of human white adipose tissue, considering both phenotypic and immune receptor specificity aspects.

High-throughput sequencing analyses demonstrated that the TCR repertoires in oWAT, sWAT, and liver samples exhibit lower diversity and are characterized by the dominance of hyperexpanded clones, in contrast to blood samples. Furthermore, we predicted the presence of TCRβs specific to viral epitopes, particularly those from CMV, a prediction confirmed by ELISpot assays. Notably, we observed a higher proportion of CMV-reactive T cells in oWAT compared to blood or sWAT. Finally, flow cytometry analyses revealed that the majority of lymphocytes infiltrating WAT were tissue-resident effector memory CD8 T cells.

4.2.1. Baseline characteristics of patients

Table 7 provides information on basic parameters of patients from subcohort 1, including age, gender, type 2 diabetes (T2D), hypertension, body mass index (BMI), glucose level, glycated haemoglobin (HbA1c), C-reactive protein (CRP), cholesterol, triglycerides, high-density lipoprotein (HDL), low-density lipoprotein (LDL), and seropositivity for CMV and EBV (Epstein-Barr virus). Around 64% of patients were women, a trend caused by social constraints in our geographic environment. Additionally, Table 8 provides information on the HLA genotype for each patient.

Table 7. *Clinical baseline characteristics of patients from subcohort 1.*

	Patient										
Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Age (years)	37	36	55	55	53	50	44	65	38	48	55
Female / Male	F	F	M	M	F	M	F	F	M	F	F
T2D	+	-	-	+	-	+	-	-	-	-	-
Hypertension	+	-	+	-	+	+	-	+	-	-	+
BMI (kg/m2)	38.2	39	33.8	50	36.9	40.6	45.5	37.1	48.1	41.8	42.8
Glucose (mg/dL)	151	88	75	98	92	318	116	88	90	105	101
Hb1Ac (%)	5.5	5	5.7	7	5.4	5.5	5.7	6.2	5.4	5.5	5.5
CRP (mg/L)	18.7	5	9.3	22	11.9	2.7	20.9	5.9	5.3	4.1	16.6
Cholesterol (mg/dL)	128	118	156	125	125	201	131	143	143	188	149
Triglycerides (mg/dL)	175	156	104	105	48	12	202	175	85	264	120
HDL (mg/dL)	39	29	39	41	61	34	27	52	35	34	38
LDL (mg/dL)	54	58	96	63	80	134	64	56	91	101	87
CMV (IgG/IgM)	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/-	+/-
EBV (IgG/IgM)	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/-	+/-

Table 8. *HLA genotype of patients.*

	HLA-I genotyping					
Patient	**HLA-A**		**HLA-B**		**HLA-C**	
1	A*66:01	A*68:01	B*15:01	B*44:03	C*03:04	C*16:01
2	A*01:01	A*31:01	B*07:02	B*67:01P	C*07:02	C*12:03
3	A*02:02	A*03:01	B*41:01	B*57:01	C*06:02	C*17:01
4	A*29:02	A*30:04	B*44:03	B*53:01	C*15:05	C*16:01
5	A*02:01	A*02:01	B*18:01	B*56:01	C*01:02	C*04:01
6	A*02:01	A*29:02	B*35:01	B*57:01	C*04:01	C*07:01
7	A*01:01	A*25:01	B*08:01	B*18:01	C*07:01	C*12:03
8	A*02:01	A*02:01	B*18:01	B*38:01	C*07:01	C*12:03
9	A*11:01	A*29:02	B*44:03	B*52:01	C*12:02	C*16:01
10	A*30:02	A*33:01	B*14:02	B*18:01	C*05:01	C*08:02
11	A*01:01	A*29:02	B*37:01	B*44:03	C*06:02	C*16:01
	HLA-II genotyping					
Patient	**HLA-DRB1**		**HLA-DPB1**		**HLA-DQB1**	
1	DRB1*04:01	DRB1*07:01	DPB1*03:01P	DPB1*04:01	DQB1*02:02	DQB1*03:02
2	DRB1*14:01P	DRB1*15:01	DPB1*02:01	DPB1*04:01	DQB1*05:03	DQB1*06:02
3	DRB1*07:01	DRB1*11:02	-	-	DQB1*03:03	DQB1*03:19
4	DRB1*13:02	DRB1*16:01	DPB1*02:01P	DPB1*17:01P	DQB1*05:02	DQB1*06:04
5	DRB1*01:01	DRB1*11:04	DPB1*04:01	DPB1*06:01	DQB1*03:01	DQB1*05:01
6	DRB1*04:01	DRB1*07:01	DPB1*04:01	DPB1*04:02	DQB1*02:02	DQB1*03:02
7	DRB1*03:01	DRB1*15:01	DPB1*03:01P	DPB1*23:01	DQB1*02:01	DQB1*06:02
8	DRB1*04:05	DRB1*13:01	DPB1*03:01P	DPB1*105:01	DQB1*03:02	DQB1*06:03
9	DRB1*07:01	DRB1*15:02	DPB1*02:01	DPB1*11:01	DQB1*02:02	DQB1*06:01
10	DRB1*01:02	DRB1*03:01	DPB1*04:01	DPB1*15:01	DQB1*02:01	DQB1*05:01
11	DRB1*01:03	DRB1*03:01	DPB1*01:01	DPB1*04:02	DQB1*02:01	DQB1*03:01

4.2.2. TCRβ repertoire analysis

These analyses were carried out in collaboration with María S. Benítez Cantos from the Department of Biochemistry and Molecular Biology III and Immunology. Here, we will exclusively present the results, omitting the details of the bioinformatic process, which can be checked in Redruello-Romero et al., 2023.

To study the immune receptor landscape across diverse tissue compartments, we examined the diversity, clonality and overlap of the TCRβ repertoires. For that purpose, we analyzed samples of blood, oWAT, sWAT, and liver from obese patients. A total of 5 million reads, covering full V-D-J regions (400-600 bp long), were obtained. We found 86,005 unique TCRs among our samples after identifying the TRVB and TRJB segments and the CDR3 sequence.

In terms of richness (number of different clonotypes in a sample), blood repertoires showed higher diversity than solid tissue repertoires (4201 ± 2556 clonotypes in blood, 343 ± 166 in oWAT, 441 ± 181 in sWAT and 296 ± 173 in liver, $p < 0.0001$) (Figure 16A). No significant differences were observed between solid tissues.

Regarding clonality, we observed the opposite. Solid tissue samples significantly presented more clonality than blood repertoires (0.11 ± 0.09 in blood, 0.23 ± 0.10 in oWAT, 0.24 ± 0.11 in sWAT and 0.26 ± 0.11 in liver, $p < 0.05$) (Figure 16B). Likewise, no significant differences were observed between solid tissues. Additionally, to assess clonality, we studied the proportion of the repertoire occupied by clonotypes of a given frequency. We catalogued clonotypes into four groups: small (clonotype frequency ≤ 0.0001), medium (0.0001-0.001), large (0.001-0.01) and hyperexpanded (>0.01). Blood repertoires were predominantly composed of medium clones, whereas hyperexpanded clones dominated solid tissues (Figure 16C).

Concerning overlap between blood and tissue repertoires, we observed a low overlap with blood for the three solid tissues. Overlap values range from 0 to 1, indicating that samples can either have no shared clonotypes (0) or share every TCRβ in the same proportion (1) (Figure 16D). The highest value observed was for patient 1 between blood and liver repertoires (0.13). Moreover, no significant differences were observed between solid tissues and their overlap with blood. ($p > 0.05$). Additionally, we studied the total number of TCRβs that overlap between samples of the same patient (Figure 16E). Likewise, we observed low or no overlap between TCRβ repertoires of blood and solid tissues. Thus, the subsets of tissue-TCRβ are very different from the others and some TCRβs are even undetectable in blood.

Altogether, these findings indicated that: blood TCRβ repertoires were more diverse, solid tissue TCRβ clonotypes were more expanded, and TCRβ repertoires of blood and solid tissues were different between them.

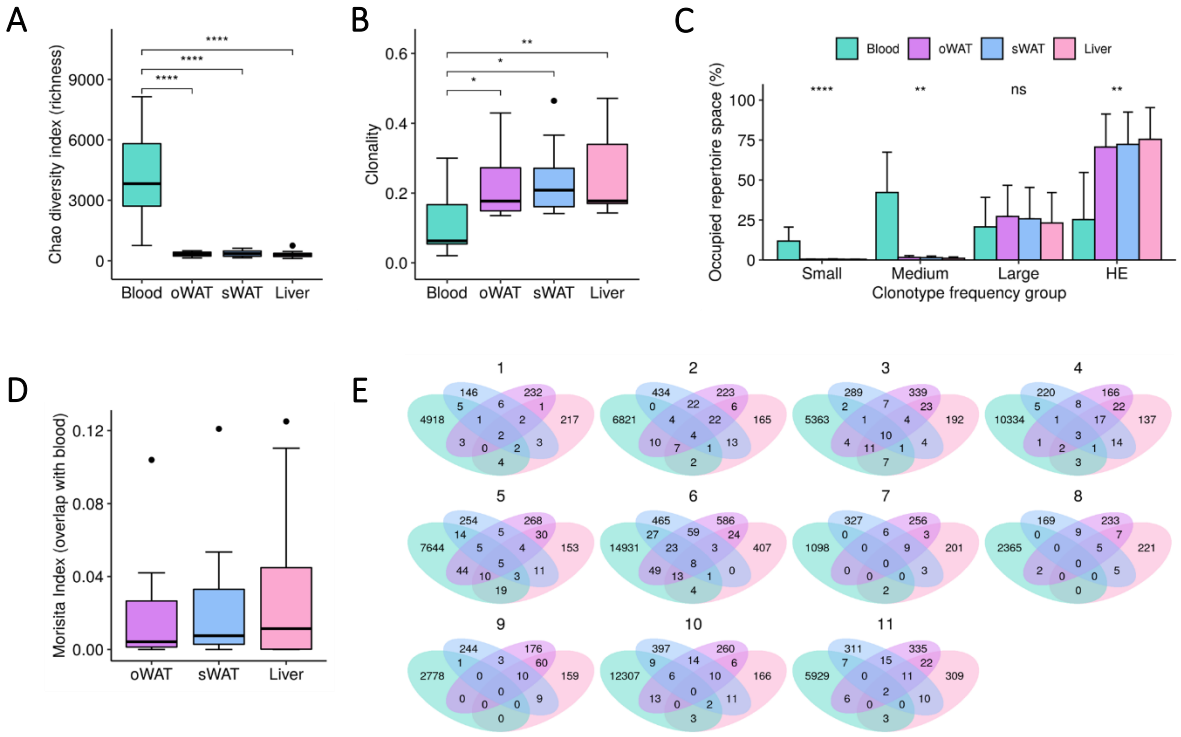

Figure 16. *TCRβ repertoire diversity measurements.* ***(A)*** *Richness analysis.* ***(B)*** *Clonality degree of the repertoires.* ***(C)*** *Percentage of occupied repertoire space by clones of a given frequency: small (≤ 0.0001), medium (0.0001 – 0.001), large (0.001 – 0.01) and hyperexpanded (HE, > 0.01).* ***(D)*** *Degree of overlap of the solid tissue repertoires with their respective blood repertoires.* ***(E)*** *Venn diagrams representing the number of shared clonotypes among the different TCRβ repertoires from the same patient. ns: non-significant, *: p ≤ 0.05, **: p ≤ 0.01, ***: p≤ 0.001, ****: p ≤ 0.0001.*

4.2.3. Antigen specificity prediction

These analyses were carried out in collaboration with María S. Benítez Cantos from the Department of Biochemistry and Molecular Biology III and Immunology. Here, we will exclusively present the results, omitting the details of the bioinformatic process, which can be checked in Redruello-Romero et al., 2023.

To predict the antigen specificity of TCRβ repertoires, the 86,005 TCRβs discovered were clustered by sequence similarity with the sequences on the VDJdb database. Only 363 TCRβs clustered with a CDR3 sequence of VDJdb, which is only 0.42% of our data. This highlights the fact that these databases are still under development and their necessity to be enlarged with more *in vitro* assays.

Our repertoires exhibited specificity to 30 epitopes derived from 19 antigens, present in blood and solid tissues. They included various viral epitopes from CMV, EBV, and Influenza virus A (Figure 17), which are the most studied viruses and therefore there is more information about them in the database. For each TCR-epitope pair, the database provides information about the HLA allele used in the antigen specificity assay. Remarkably, 24 out of the 30 epitopes are presented by HLA class I. The remaining 6 are presented by HLA class II but were undetected in oWAT. However, acknowledging the known promiscuity of HLA alleles in antigen presentation (Meyer et al., 2018; Rao et al.,

2011), we included all discovered antigen specificities, even when the reported HLA allele in the database does not match those carried by our patients.

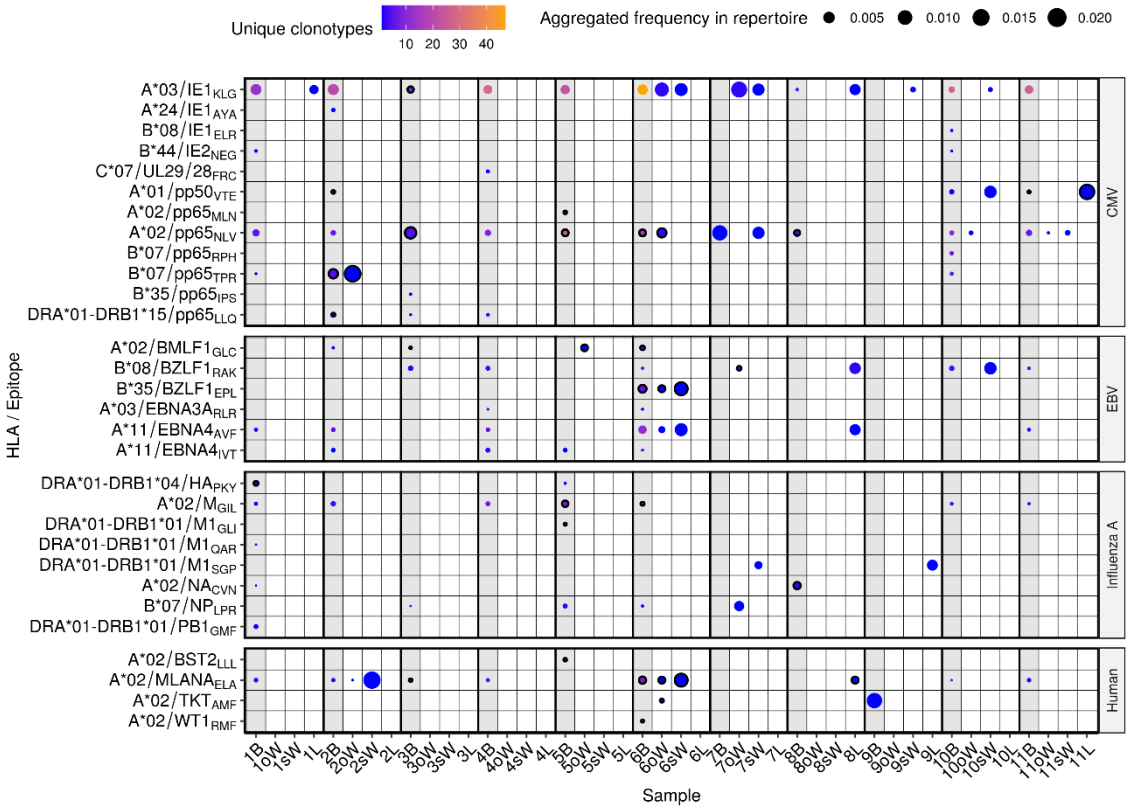

Figure 17. *Predicted antigen specificity of the 44 TCRβ repertoires analyzed. Epitopes (y axis) are identified by the HLA allele presenting them, the protein of origin and their first three amino acids. Each circle represents the specificity of TCRβs to each epitope in a given repertoire. Color scale stands for the number of different clonotypes recognizing that epitope, while size represents the relative frequency of those clonotypes in the repertoire. Blood samples from each donor are highlighted in light grey and outlined circles indicate that the patient carries the HLA allele presenting the epitope according to the VDJdb database. B, blood; oW, omental white adipose tissue; sW, subcutaneous white adipose tissue; L, liver tissue.*

TCRβs exhibiting specificity to 8 CMV epitopes were exclusively detected in blood samples. Meanwhile, specificity to 4 epitopes was observed in both blood and WAT in 5 out of 11 patients (45.5%). On the other hand, specificity to EBV epitopes was found to a lesser extent in blood and solid tissues. On the contrary, Influenza A epitopes were primarily predicted to be recognized by TCRβs from blood. We would like to highlight the following epitopes:

- A*03/IE1$_{KLG}$ (CMV) was predicted to be recognized with specific TCRs found in oWAT from patient 6 and sWAT from patients 6, 7, 9, and 10. It was identified as the most widely recognized CMV epitope in our cohort. It is worth mentioning that this epitope has the highest number of known TCRs in the VDJdb database, potentially influencing the results.
- Specificity for A*02/pp65$_{NLV}$ (CMV) was predicted for all the patients except patient 9 (who tested negative in CMV serology). Likewise, it was predicted in

oWAT from patients 6 (carrier of the HLA-A*02 allele, Table 8), 10 and 11, and sWAT from patients 7 and 11.

- TCRβs against B*07/pp65$_{TPR}$ (CMV) were found in oWAT from patient 2 (carrier of the HLA-B*07 allele, Table 8). It occupied approximately 16.8% of the repertoire.
- Cognate TCRβs for B*35/BZLF1$_{EPL}$ (EBV) and A*11/EBNA4$_{AVF}$ (EBV) epitopes were found in WAT from patient 6, showing the broadest specificity against EBV epitopes.
- Patient 7 exhibited specificity against epitopes B*07/NP$_{LPR}$ (Influenza A) in oWAT and DRA*01-DRB1*01/M1$_{SGP}$ (Influenza A) in sWAT.

Specificity to self-antigens, particularly to A*02/MLANA$_{ELA}$, was also predicted in blood, oWAT and sWAT from patients 2 and 6. This epitope is derived from the melanoma antigen recognized by T cells 1 (MART1 or MLANA). Reactive T cells against MLANA are associated with certain pathological conditions such as vitiligo and melanoma (Mandelcorn-Monson et al., 2003). Remarkably, TCRβs against this epitope occupied 21.1% of the sWAT repertoire of patient 2 and 10.7% of patient 6. Interestingly, patient #6 was diagnosed with basal cell carcinoma afterwards.

4.2.4. ELISpot assays

One remarkable outcome from the *in silico* antigen-specificity prediction was the notable presence of CMV-specific TCRs in oWAT (always considering that the VDJdb database is biased to the most studied diseases). To prove these findings, we performed a functionality test through ELISpot assays isolating T cells from blood, oWAT, and sWAT of all patients (Figure 18). We must highlight that liver samples were exclusively used for TCRβ repertoire analyses due to the limited amount of hepatic tissue.

The average number of spots of CMV-reactive T cells in oWAT (1174.45 ± 207.86) was significantly higher than in PBMCs (255.09 ± 194.37, $p = 0.013$) and sWAT (45.36 ± 15.94, $p<0.001$) (Figure 19A). It is important to clarify that we tried to seed the same number of mononuclear cells in each well. The mononuclear fraction in the SVF from both types of adipose tissues studied, oWAT and sWAT, comprises 40-50% T lymphocytes (García-Rubio et al., 2018). On the other hand, T lymphocytes from the mononuclear fraction of blood are typically 40-70%. In this study, flow cytometry revealed that these percentages are narrowed down to 45-55% across all samples, meaning that the number of cells seeded was similar.

Another noteworthy finding lies in the spot size. It serves as an indicator of the amount of interferon-γ (IFN-γ) secretion post-TCR stimulation. In oWAT, a significant portion of cells (approximately 15.5%) exhibited what we have addressed as "high-secretion level" of IFN-γ (spot area >2000 μm2). In contrast, in blood, only 3.5% of spots exhibited a high-secretion level, whereas in sWAT is absent (Figure 19B). To better illustrate this, we showed in Figure 19C that the first decile of oWAT had spot sizes significantly larger than those in sWAT ($p<0.001$) and blood ($p<0.001$).

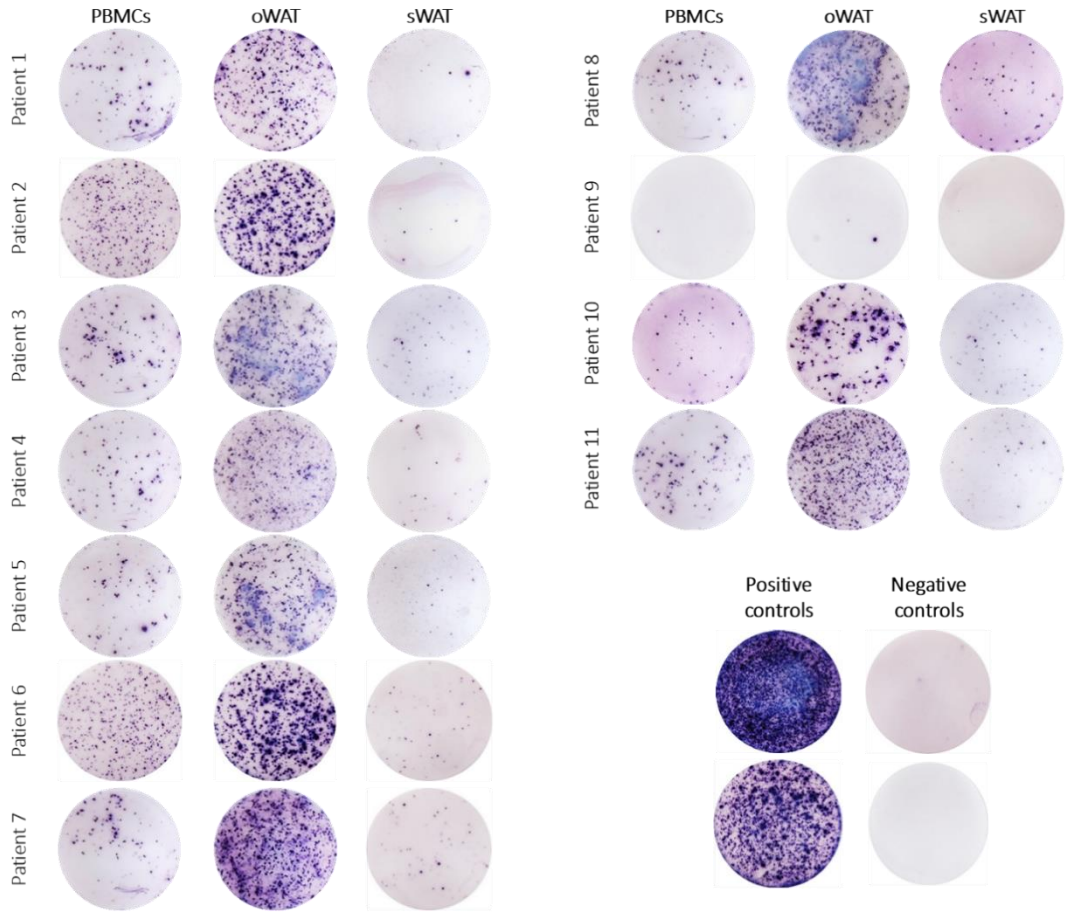

Figure 18. *ELISpot assays of all 11 patients, and their positive and negative controls. It is worth noting that there is only one seronegative patient, number 9, who also tested negative in the ELISpot assay.*

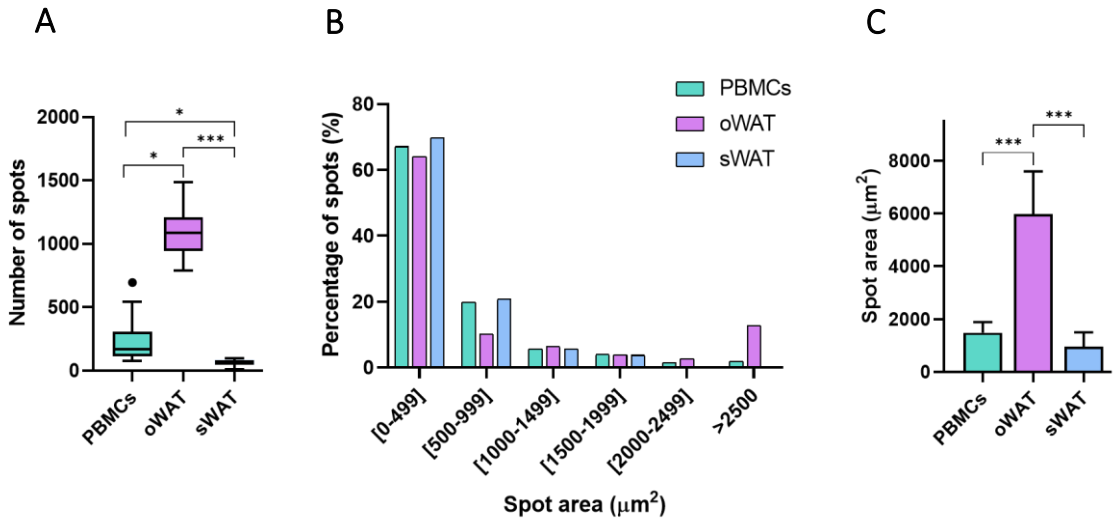

Figure 19. *CMV ELISpot assays measurements.* ***(A)*** *Total number of spots in PBMCs, oWAT and sWAT for all seropositive patients.* ***(B)*** *Ranges of spots' areas in µm2 and their percentages for PBMCs and adipose tissues.* ***(C)*** *First 10% of the biggest spots in the three different tissues for all seropositive patients. *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$.*

4.2.5. Phenotypic analyses of T cells by flow cytometry

We used two different panels to assess the phenotype of T lymphocytes in the different samples: blood, oWAT and sWAT. The first panel identified the naïve phenotype (CCR7+/CD45RA+) and three memory populations: T central memory (Tcm: CCR7+/CD45RA-), T effector memory (Tem: CCR7-/CD45RA-) and T effector memory re-expressing CD45RA (Temra: CCR7-/CD45RA+). The second panel included tissue-residence markers like CD69, CD49a, and CD103. Once again, it must be noted that liver samples were completely used for the TCRβ analyses.

Figure 20A shows the SSC *versus* FSC plot and the gating strategy to identify the two main T cell populations (CD4+ and CD8+) by Panel 2 in the SVF of oWAT. Notably, blood and tissue samples exhibited different flow cytometry profiles (Figures 21 and 22), suggesting negligible contamination of tissue samples with peripheral blood. Previous studies reported a higher abundance of FALCs in oWAT (Bénézech et al., 2015; Moro et al., 2010), which is consistent with our results showing significantly more T cells (both CD4+ and CD8+) in oWAT than in sWAT (Figure 20B). Likewise, the CD4+:CD8+ ratio differed significantly between both types of adipose tissue (Figure 20C), with CD8+ cells dominating the T cell pool in oWAT and CD4+ cells in sWAT. Consequently, the increased number of CD8+ T cells in oWAT was the main driving force of the T cell differences between both types of adipose tissue.

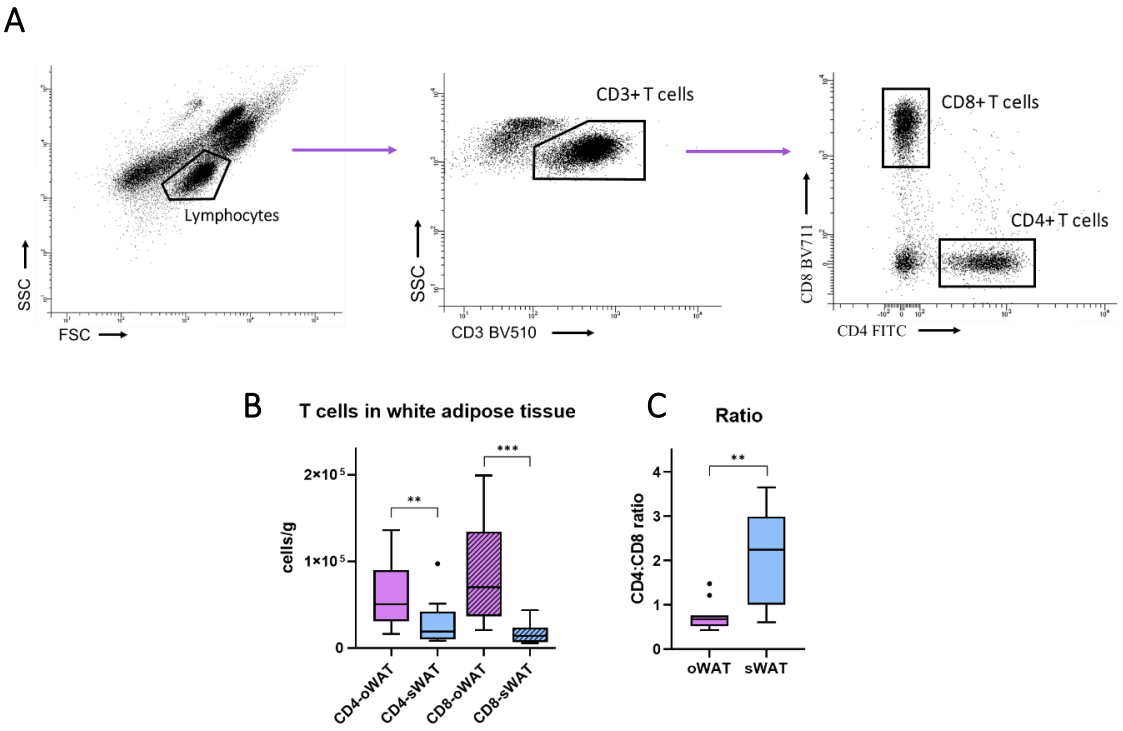

Figure 20. (A) *SSC versus FSC plot of oWAT SVF and the gating strategy used to identify CD4+ and CD8+ subpopulations by Panel 2.* ***(B)*** *Total number of CD4+ and CD8+ T lymphocytes per gram of oWAT and sWAT.* ***(C)*** *Ratio CD4:CD8 in omental and subcutaneous adipose tissue. *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$.*

Panel 1 showed one of the main differences between blood and adipose tissue. The naïve population was abundant in blood but almost absent in tissue (22.2% blood, 1.9% oWAT, 1.7% sWAT, $p<0.001$, Figure 21), which agrees with previous studies (Farber et al., 2014). The most abundant CD4+ subset in both types of adipose tissue was the T effector memory (Tem), followed by the T central memory (Tcm) and the T effector memory re-expressing CD45RA (Temra) (Figures 21B-C and 21E-F). In the same way, Tem was also the most abundant population in the CD8+ subset of adipose tissue, but now followed by Temra and Tcm (Figures 21H-I and 21K-L). On the other hand, Temra was the most abundant CD8+ subset in blood (Figure 21G and 21J), contrasting with its almost absent presence in the CD4+ cells (Figure 21A and 21D). These findings are consistent with previous reports that described circulant Temra CD8+ as a major subset increasing with age (Koch et al., 2008).

Panel 2 showed that the proportion of T cells positive for tissue-resident markers varied between blood and adipose tissue. In tissue-resident memory T cells (Trm), CD69 is usually co-expressed with CD49a and/or CD103. Remarkably, nearly all T cells in tissue displayed a memory phenotype. The majority of adipose tissue T lymphocytes were positive for CD69 (CD4+ in oWAT: 86.8 %, Figures 22B and 22E; CD8+ in oWAT: 93.3 %, Figures 22H and 22K; CD4+ in sWAT: 72.5% Figures 22C and 22F; CD8+ in sWAT: 81.7%; Figures 22I and 22L). On the contrary, blood only exhibited a small fraction of CD69+ T cells (CD4+: 2.7 %, Figures 22A and 22D; CD8+: 4.4 %, Figures 22G and 22J), while CD49a and CD103 were almost absent (Figures 22A, 22D, 22G and 22J, and Figures 23A and 23D, respectively). Likewise, CD103+ T cells were minimal in WAT (Figure 23B, 23C and 23D), but CD49a+ showed high levels, particularly in oWAT (Figure 22B, 22D, 22H and 22K).

Two conclusions can be drawn from the CD69 and CD49a expression patterns. Firstly, nearly all CD49a+ T cells co-expressed CD69. Secondly, the double positive population (CD69+/CD49a+) was the most abundant in oWAT (especially in CD8+) (Figure 22H and 22K), whereas no predominant population was observed in sWAT.

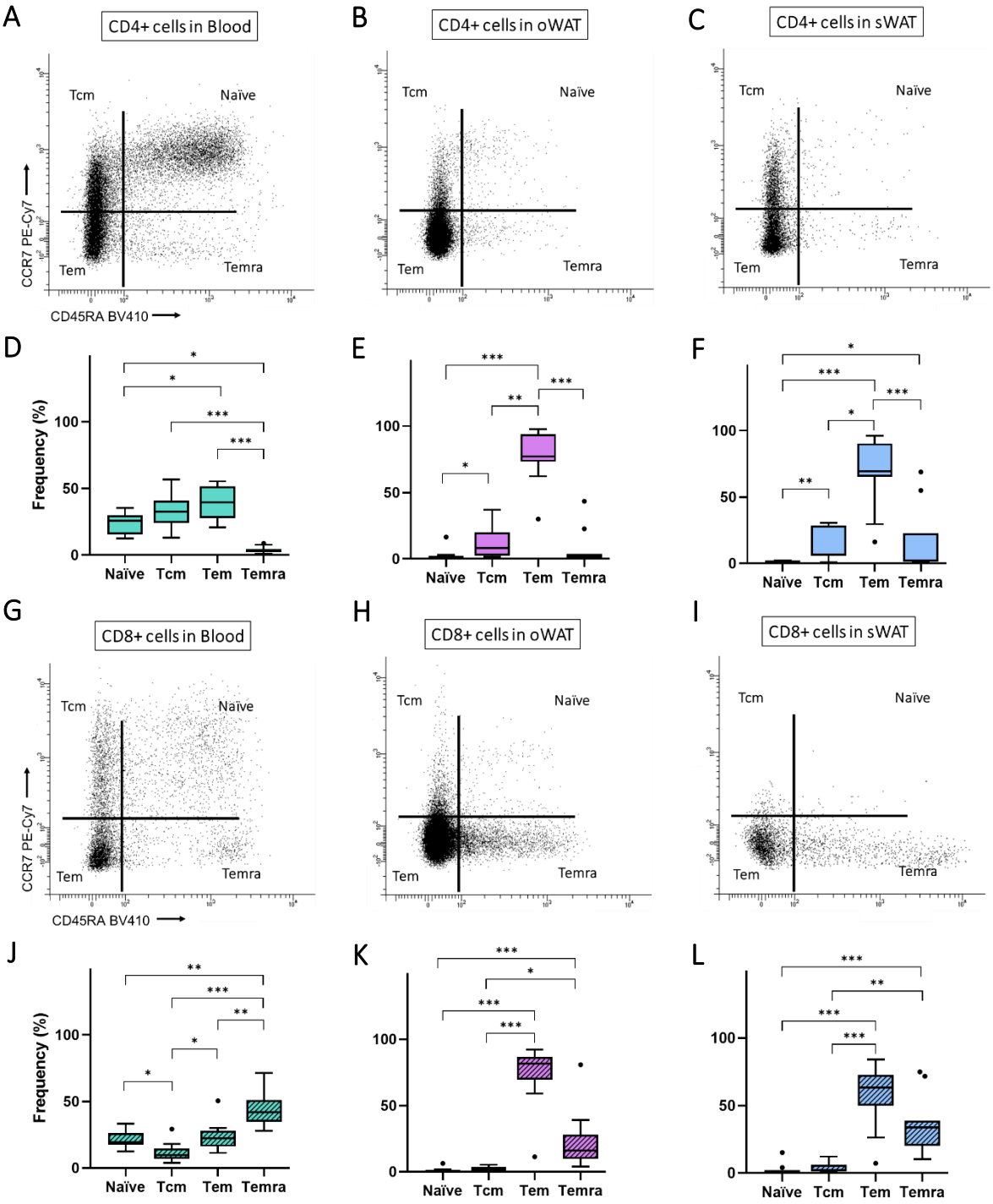

Figure 21. *CD4+ T cell subpopulations (naïve, T central memory (Tcm), T effector memory (Tem), T effector memory re-expressing CD45RA (Temra)) identified by Panel 1 in:* ***(A)*** *blood,* ***(B)*** *oWAT, and* ***(C)*** *sWAT.* ***(D)*** *CD4+ subsets in blood.* ***(E)*** *CD4+ subsets in oWAT.* ***(F)*** *CD4+ subsets in subcutaneous sWAT. CD8+ T cell subpopulations identified by Panel 1 in:* ***(G)*** *blood,* ***(H)*** *oWAT, and* ***(I)*** *sWAT.* ***(J)*** *CD8+ subsets in blood.* ***(K)*** *CD8+ subsets in oWAT.* ***(L)*** *CD8+ subsets in sWAT. *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$.*

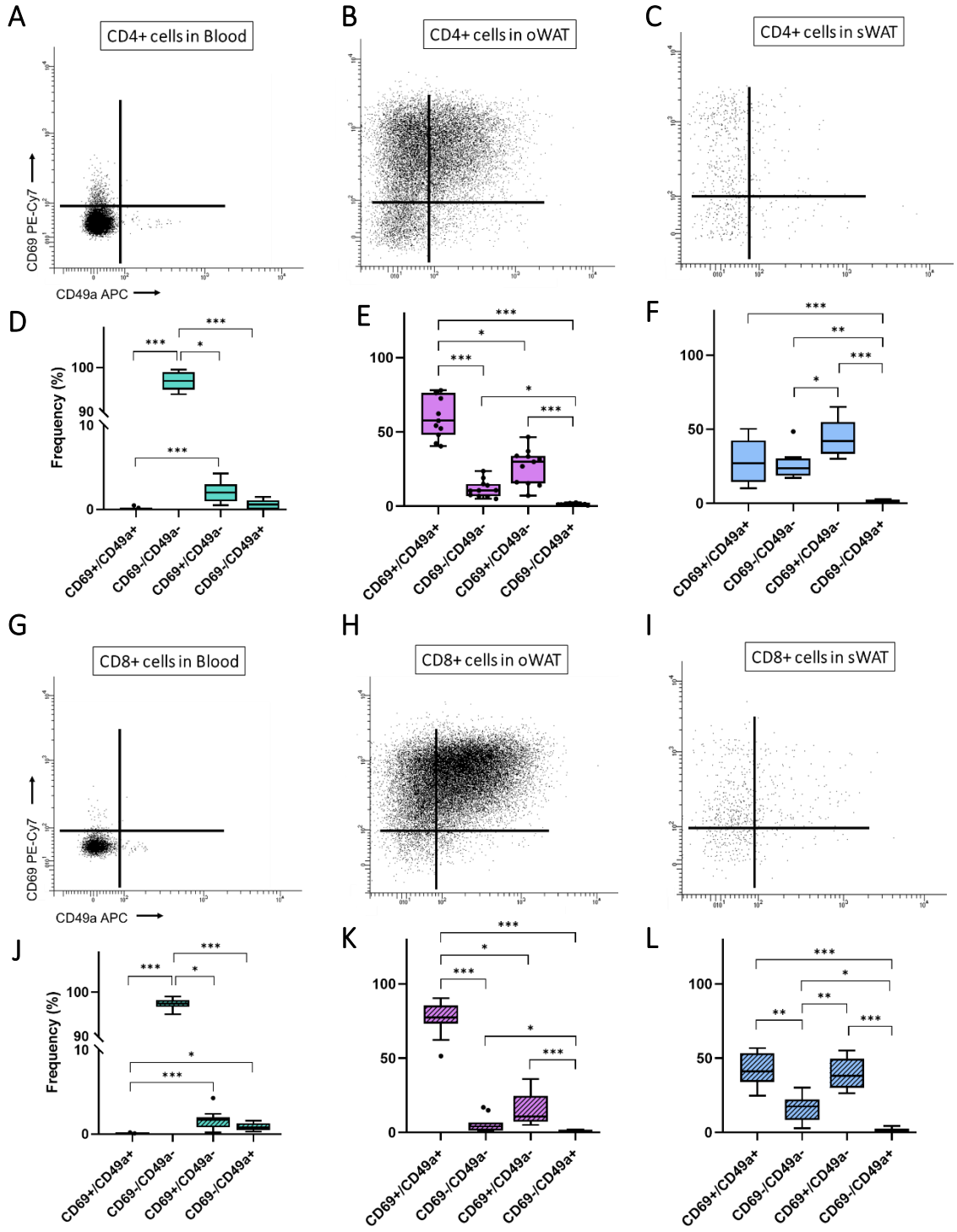

Figure 22. *CD4+ T cell subpopulations regarding memory phenotype (CD69+/CD49a+, CD69-/CD49a-, CD69+/CD49a-, CD69-/CD49a+) identified by Panel 2 in:* ***(A)*** *blood,* ***(B)*** *oWAT, and* ***(C)*** *sWAT.* ***(D)*** *CD4+ subsets in blood.* ***(E)*** *CD4+ subsets in oWAT.* ***(F)*** *CD4+ subsets in sWAT. CD8+ T cell subpopulations identified by Panel 2 in:* ***(G)*** *blood,* ***(H)*** *oWAT, and* ***(I)*** *sWAT.* ***(J)*** *CD8+ subsets in blood.* ***(K)*** *CD8+ subsets in oWAT.* ***(L)*** *CD8+ subsets in sWAT. *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$.*

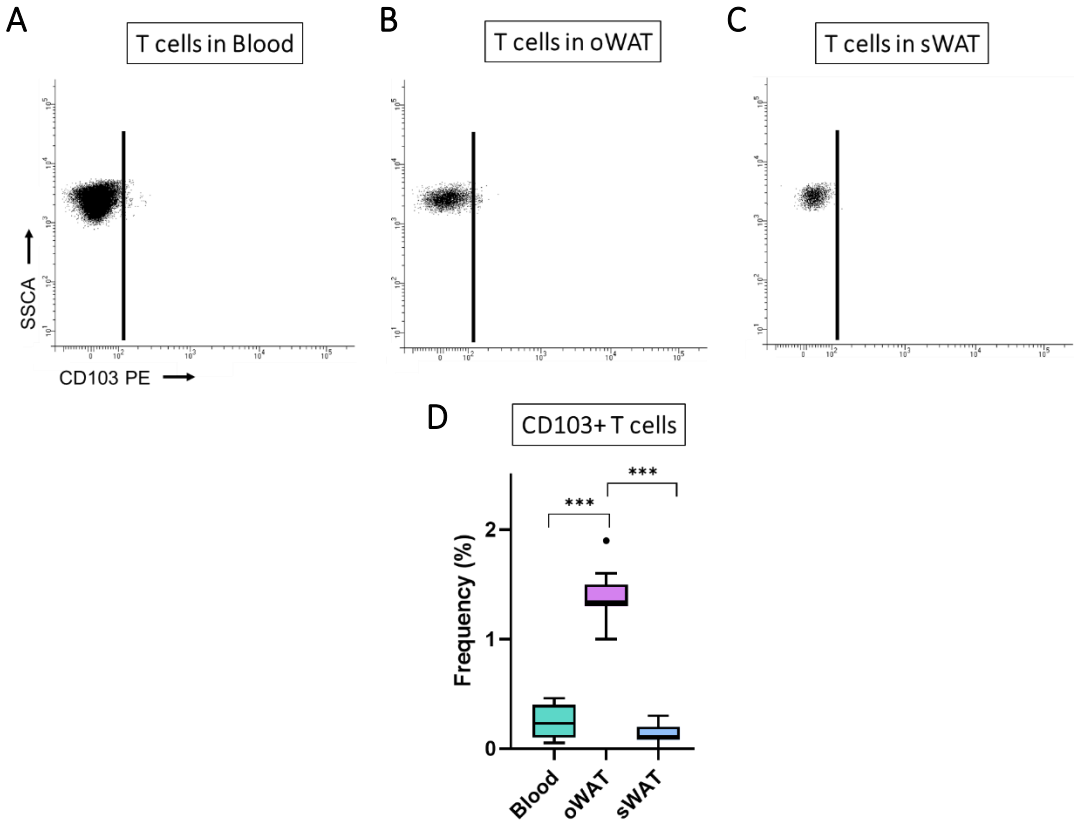

Figure 23. *T cells CD103 staining in: **(A)** blood, **(B)** oWAT, and **(C)** sWAT. **(D)** Frequency of CD103+ T cells in blood, oWAT and sWAT. ***: $p \leq 0.001$.*

Chapter III:

Non-canonical TCRβ rearrangements

4.3. Chapter III: Non-canonical TCRβ rearrangements

For this chapter, we obtained blood from 24 obese patients, and omental white adipose tissue (oWAT), subcutaneous white adipose tissue (sWAT), and liver samples from 11 to study the possibility of functional non-canonical TCRβ rearrangements. High-throughput sequencing demonstrated the presence of TCRβs comprised of the recombination of two TRVB segments. Not only did they appear in PBMCs, but in solid tissue, where T cells have already undergone a selection process after an HLA-antigen encounter. Furthermore, we verified that these were not artefacts arising from the library preparation or sequencing processes and provided evidence that Jurkat-KOβ cells can express a complete CD3-TCR complex when a non-canonical rearrangement is introduced.

4.3.1. Baseline characteristics of patients

Table 9 provides information on basic parameters of patients from subcohort 2, including age, gender, type 2 diabetes (T2D), hypertension, body mass index (BMI), glucose level, glycated haemoglobin (HbA1c), C-reactive protein (CRP), cholesterol, triglycerides, high-density lipoprotein (HDL) and low-density lipoprotein (LDL).

Subcohort 2 is divided into two groups: A and B. Group A includes a total of 11 patients, where analyses were performed on samples of PBMCs, oWAT, sWAT, and liver. Group B includes the same 11 patients from group A, with an additional 13 patients whose PBMCs were exclusively studied.

Table 9. *Clinical baseline characteristics of patients from subcohort 2 (groups A and B). Continuous variables are expressed in mean ± SD (standard deviation).*

Variables	Group A	Group B
Age (years)	47.63 ± 9.64	50.12 ± 8.56
Female / Male	5 / 6	17 / 7
T2D (Yes / No)	3 / 8	5 / 19
Hypertension (Yes / No)	7 / 4	12 / 12
BMI (kg/m2)	42.65 ± 6.59	42.33 ± 6.56
Glucose (mg/dL)	115.36 ± 70.56	108.04 ± 47.81
Hb1Ac (%)	5.6 ± 0.63	5.69 ± 0.55
CRP (mg/L)	8.38 ± 5.65	7.2 ± 5.84
Cholesterol (mg/dL)	130.55 ± 26.43	141.5 ± 27.62
Triglycerides (mg/dL)	115.8 ± 42.21	131.26 ± 50.70
HDL (mg/dL)	36.27 ± 9.51	37.37 ± 8.14
LDL (mg/dL)	72.45 ± 24.41	78.67 ± 28.42

4.3.2. Bioinformatic analyses of non-canonical TCRβs rearrangements

These analyses were carried out in collaboration with A. Jesús Muñoz Pajares from the Department of Genetics. Here, we will explain the bioinformatic pipeline employed for the analyses and present the results obtained, while omitting the details on the code created for the analyses.

The pipeline consisted of the following steps:

1. Data preparation: The first step involved preparing the data for analysis. For this purpose, reference sequences for each TRVB and barcodes for each sample were read. Reference sequences were obtained from the IMGT database.

2. Size filtering: Secondly, sequences were filtered based on size, focusing on sequences of ≥400 bp (Figure 24) to ensure the maximum coverage of the TCRβ length. Likewise, it will allow better identification of the complete TRVB.

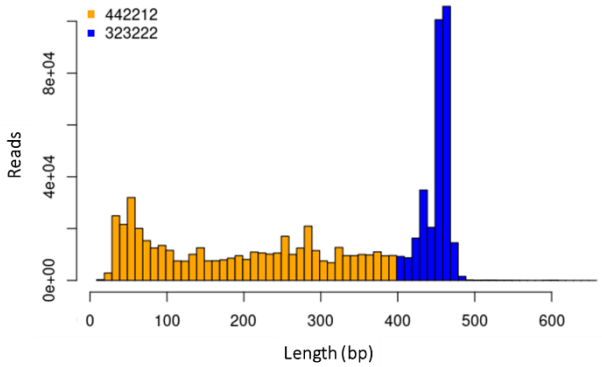

Figure 24. *Example of a histogram of the selected (blue) and discarded sequences (orange) after size-filtering.*

3. BLAST analysis: Sequences of ≥400 bp were further divided into fragments of 100 bp for analysis using BLAST (Basic Local Alignment Search Tool). This bioinformatic tool compares a query sequence against a database of reference sequences for each TRVB, identifying regions of similarity. BLAST assigns scores to aligned regions based on the similarity between the query and the database sequences. Thus, multiple BLAST analyses were performed per sequence and the consistency of the TRVB identified across the query sequence was used to identify non-canonical rearrangements. According to that, individual sequences were categorized as “No BLAST" for those showing no significant similarities with the database; "No Recomb" for sequences aligning with TRBVs belonging to the same family across the entire sequence; and "Recomb" for sequences aligning with reference TRVBs from different families (Figure 25, Table 10).

Figure 25. *Illustration of the BLAST analysis workflow. TRVBs from the same family are represented with the same colour (i.e., purple) but different shading. Different TRVBs are represented in different colours (i.e., purple and green). We must emphasize that it is a mere representation. There can be more fragments, and the rearrangement can be observed from any point along the whole sequence.*

Table 10. *Results from the BLAST alignment of PBMCs (Group B, N=24) and solid tissues (Group A, N=11). "No BLAST": number of sequences with no significant similarities with the database, "No Recomb": number of sequences aligning with TRBVs belonging to the same family across the entire sequence, and "Recomb": number of aligning with reference TRVBs from different families.*

Sample	No BLAST	No Recomb	Recomb	Sample	No BLAST	No Recomb	Recomb
1-PBMCs	0	14736	55	5-oWAT	0	2316	1
2-PBMCs	6	25572	94	6-oWAT	0	3057	6
3-PBMCs	0	14793	55	7-oWAT	1	2715	2
4-PBMCs	1	12767	44	8-oWAT	10	2284	1
5-PBMCs	1	16294	39	9-oWAT	0	1583	1
6-PBMCs	0	15817	40	10-oWAT	0	3014	1
7-PBMCs	0	10968	32	11-oWAT	1	2879	8
8-PBMCs	0	12501	49	1-sWAT	36	1560	1
9-PBMCs	4	21614	45	2-sWAT	219	1383	1
10-PBMCs	0	24550	63	3-sWAT	0	3391	1
11-PBMCs	0	17754	7	4-sWAT	7	4552	1
12-PBMCs	1	20693	21	5-sWAT	18	154	0
13-PBMCs	0	17229	40	6-sWAT	1	2420	2
14-PBMCs	39	18841	28	7-sWAT	0	1801	0
15-PBMCs	1	20890	42	8-sWAT	11	2260	0
16-PBMCs	1	18100	44	9-sWAT	0	2864	2
17-PBMCs	7	62001	77	10-sWAT	0	2289	1
18-PBMCs	1	48144	96	11-sWAT	3	3738	4
19-PBMCs	1	59976	74	1-Liver	22	1169	3
20-PBMCs	2	23493	110	2-Liver	114	2577	2
21PBMCs	0	11147	16	3-Liver	0	3467	2
22-PBMCs	3	3266	7	4-Liver	4	2814	1
23-PBMCs	3	13194	39	6-Liver	5	2470	7
24-PBMCs	0	39963	54	7-Liver	0	2879	3
1-oWAT	8	1995	1	8-Liver	12	864	2
2-oWAT	152	3289	2	9-Liver	0	1929	1
3-oWAT	1	15190	13	10-Liver	0	3398	5
4-oWAT	2	1638	0	11-Liver	0	2401	1

4. Frequency calculation: Finally, the ratio of reads categorized as "Recomb" to "No Recomb" was calculated to determine the frequency of non-canonical TCRβs observed in a sample.

Figure 26 shows that PBMCs exhibited a significant higher frequency of non-canonical rearrangements compared to solid tissues, but no significant differences were observed among oWAT, sWAT, and liver. We must highlight that every PBMCs sample was included in this study as differences between group A and B were not significant (Group A: 0.28 ± 0.1, Group B: 0.24 ± 0.11, p= 0.253).

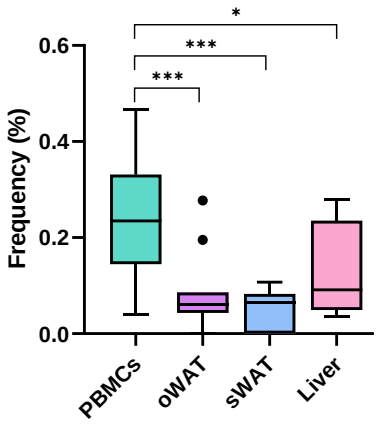

Figure 26. *Frequency of non-canonical TCRβs in PBMCs (Group B, N=24) and solid tissues (Group A, N=11).*

As expected, the presence of a more diverse TCR repertoire in blood might be associated with a higher number of non-canonical TCRβ rearrangements in blood. These lymphocytes are still in their naïve phase, having not undergone selection through HLA-antigenic encounters. Interestingly, these rearrangements are also identified, albeit at a lower frequency, within solid tissue. This phenomenon indicates that these TCRs have undergone a selection process and are now part of the immunological memory. These findings suggest the functionality of these atypical TCRβ rearrangements, revealing the possibility of an additional layer in the generation of TCR diversity. One might speculate that such low-frequency rearrangements could arise from artefacts in the experimental process. Therefore, a meticulous analysis in the subsequent sections aims to elucidate their authenticity.

4.3.3. Confirmation of non-canonical TCRβs rearrangements by PCR

Considering the results obtained from the bioinformatic analyses, we wanted to verify that the non-canonical TCRβ rearrangements were not an artefact from the library preparation or sequencing process. For this purpose, we designed an experiment involving two nested PCRs, PCR-A and PCR-B. PCR-A was intended to enrich the sample with the specific molecule of interest (i.e., a non-canonical rearrangement), while PCR-B aimed to amplify a smaller molecule that contains the hotspot where the rearrangement is allegedly taking place.

We designed different sets of primers pointing at possible rearrangements, but very few resulted in positive outcomes. In the end, we decided to focus our experiment on the rearrangements between TRVB6-1 and TRBV27. These segments shared only 66.8% sequence similarity, which is lower compared to other segments (usually from the same family) that can share more than 98% sequence similarity. There are two longer homologous sequences between TRBV6-1 and TRBV27. The first one is located between CDR1 and CDR2 and the second one is at the 3' end. The homology study between these segments is detailed below:

- Shaded in purple: nucleotides form the TRBV27 segment shared with the TRBV6-1 segment.
- Bold: start codon.
- Bold and underlined: CDR1 and CDR2.
- Shaded in blue: longest homologous sequences with 21 and 20 nucleotides, respectively.

acacctgcctgatgctgcc**atg**ggccccagctccttggctatgtggtcctttgccttctagga
gcaggcccctggaagcccaagtgacccagaacccaagatacctcatcacagtgactggaaaga
agttaacagtgacttgttctcagaat**atgaaccatgagtat**atgtcctggtatcgacaagaccc
agggctgggcttaaggcagatctactat**tcaatgaatgttgaggtg**actgataagggagatgtt
cctgaagggtacaaagtctctcgaaaagagaagaggaatttcccctgatcctggagtcgccca
gccccaaccagacctctctgtacttctgtgccagcagttatc

To minimize the chance of introducing artefacts during the PCR process, we decided to amplify small amplicons. It is important to highlight that the amplified zone contained the first longer homologous sequence. Amplicons from PCR-A were 141 bp long, while amplicons from PCR-B were 92 bp long. Primers were composed of 30 nucleotides (Table 11), leaving a 32-bp non-hybridizing region. This design prevented any potential PCR artefacts. Noteworthy, primers were carefully designed so that the nucleotides at the 3' end specifically hybridized with a unique sequence from either TRBV6-1 or TRBV27. PCR-A consisted of 20 cycles, while PCR-B consisted of 40 cycles, with an annealing temperature of 68 ºC for both.

Table 11. *Sequences of primers employed for the verification of TCRβ rearrangement between TRBV6-1 and TRBV27.*

PCR	Forward primers		Reverse primers	
	Name	Sequence	Name	Sequence
A	V27-A(F)	acagtgactggaaagaagttaacagtgact	V6-1-A(R)	gcttctgagggtaccactgacaaaggagaa
	V6-1-A(F)	ctgaagacaggacagagcatgacactgcag	V27-A(R)	atgaatgttgaggtgactgataagggagat
B	V27-B(F)	acttgttctcagaatatgaaccatgagtat	V6-1-B(R)	actgaggctgatttattactcagcttctga
	V6-1-B(F)	cagtgtgcccaggatatgaaccataactcc	V27-B(R)	cttaaggcagatctactattcaatgaatgt

Likewise, we opted to perform a random experiment using 10 mixtures of 10 mixed cDNA samples of PBMCs from random patients (with 150 ng of each sample). Following verification of amplicon size through agarose electrophoresis, we validated their sequence by Sanger sequencing. Among the total mixtures, 2 of them (mixtures 3 and 10) presented the TRVB27/TRVB6-1 rearrangement, while another 2 (mixtures 6 and 7) exhibited the TRVB6-1/TRVB27 rearrangement (Figure 27). We must highlight that we considered a non-canonical rearrangement when more than 5 out of 10 colonies in the same mixture presented the same rearrangement.

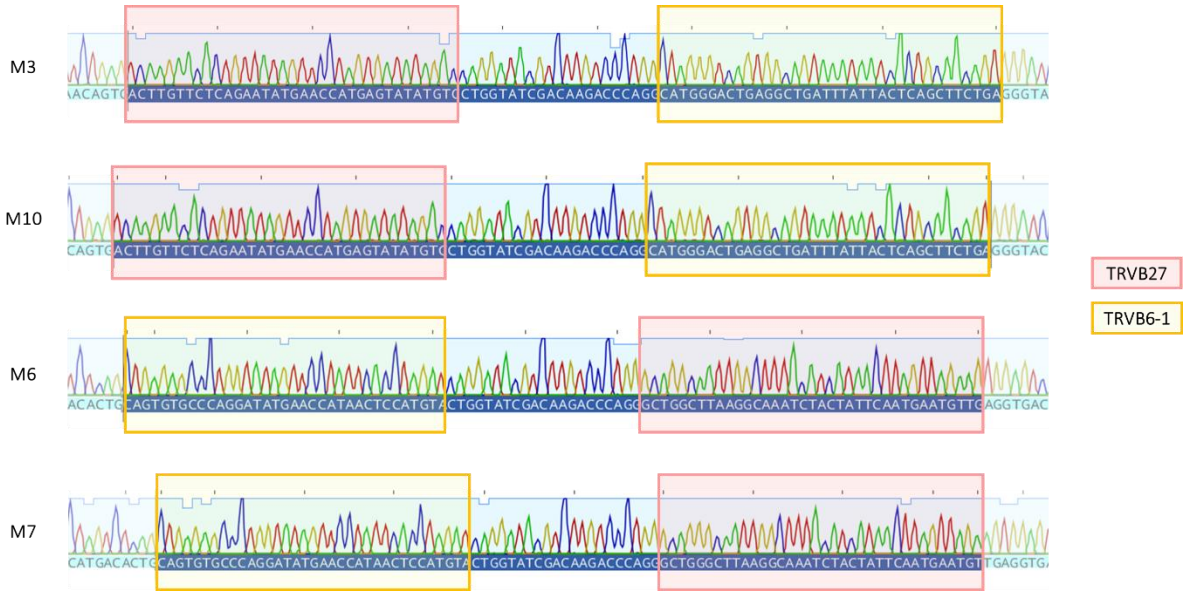

Figure 27. *Electropherograms from mixtures M3, M10, M6, and M7 show their different non-canonical rearrangements. The middle part, not marked in any square, is the longest common sequence between TRVB27 and TRVB6-1.*

Interestingly, negative results in this scenario further supported our hypothesis. The lack of amplification with alternative sets of primers could be attributed to either a low or nonexistent frequency of the rearrangement. It is crucial to highlight that under conditions of low frequency of the target molecule in a PCR, the likelihood of obtaining PCR artefacts rises, as primers may indiscriminately hybridize due to the limited availability of specific binding sites. Therefore, in such cases, it would be expected to observe many unspecific amplifications, potentially indicative of artefacts.

Altogether these findings suggest that non-canonical rearrangements are natural and likely originated at some point in the V(D)J recombination process. Despite their limited frequency, these rearrangements may contribute to an additional layer of diversity in the TCRβ biology.

4.3.4. *In vitro* assays

4.3.4.1. Cells characterization

Both cell cultures, Jurkat-WT and Jurkat-KOβ, were characterized through flow cytometry. One unique panel was used to differentiate living cells expressing a CD3-TCR complex on their surface (ZombieRed-/CD3+) from living cells not expressing CD3-TCR complexes (ZombieRed-/CD3-) or dead cells (ZombieRed+). As depicted in Figures 28A and 28C, little mortality was observed. On the other hand, there was a clear significance regarding the CD3-TCR complex between Jurkat-WT (Figure 28B) and Jurkat-KOβ (Figure 28D), indicating the successful TCRβ knock-out.

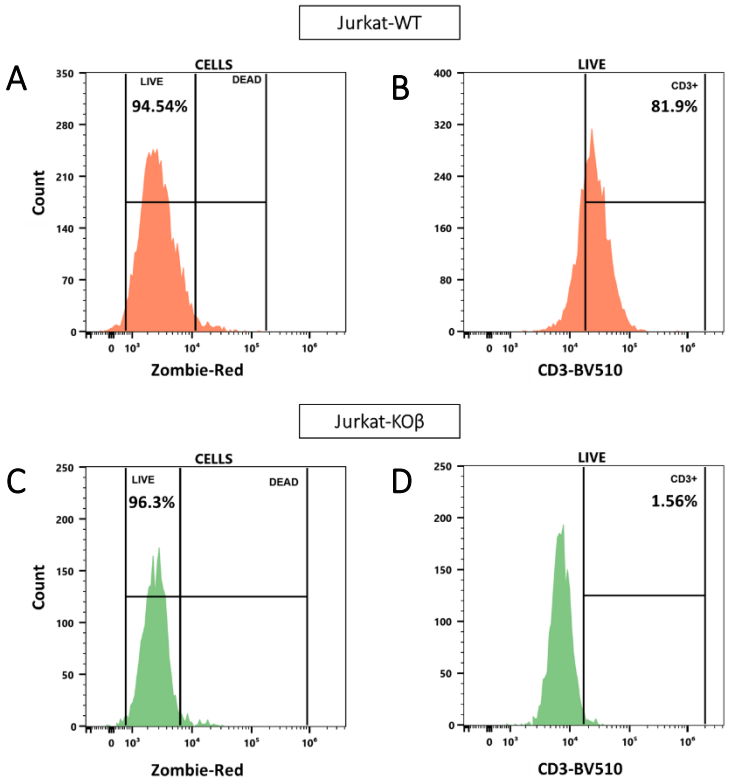

Figure 28. ***(A)*** *Jurkat-WT viability assay.* ***(B)*** *CD3 expression of the living Jurkat-WT cell line.* ***(C)*** *Jurkat-KOβ viability assay.* ***(D)*** *CD3 expression of the living Jurkat-KOβ cell line.*

Likewise, we decided to characterize both cell cultures through confocal microscopy. On this occasion, we used Hoescht, CD45-TxRed as a cell membrane marker and CD3-FITC to determine CD3-TCR complexes. Figures 29A and 29B differed drastically on the CD3-FITC signal, where it can be observed that Jurkat-KOβ cells did not express any CD3-TCR complex. Once again, this indicated that the knock-out for TCRβ was 100% efficient.

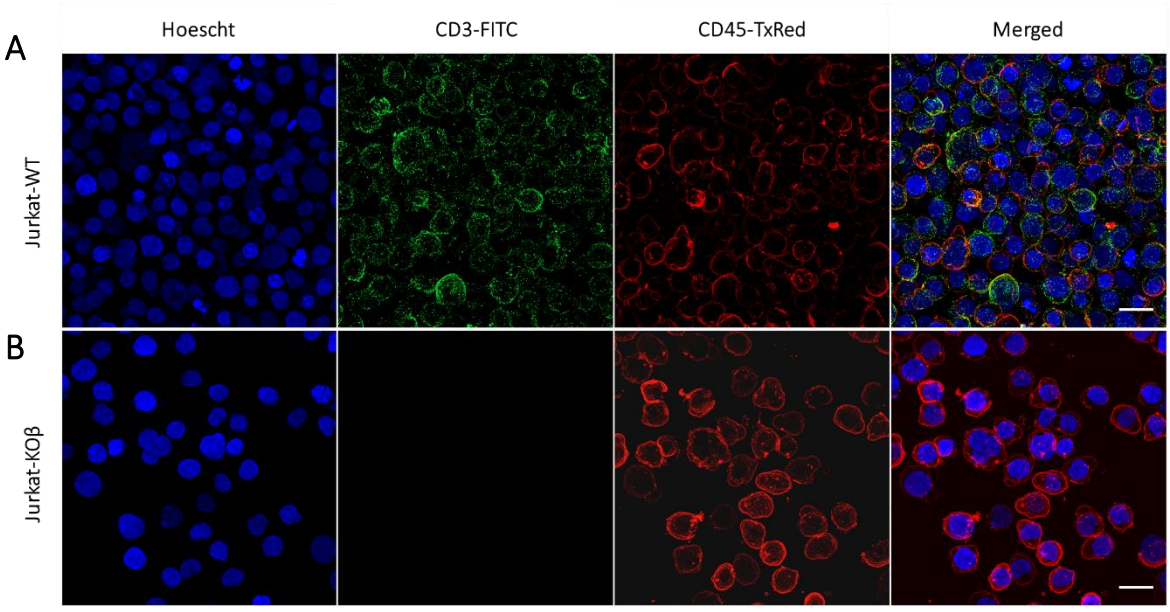

Figure 29. ***(A)*** *Characterization of Jurkat-WT cell line.* ***(B)*** *Characterization of Jurkat-KOβ cell line. Maximum intensity projection. Scale bar: 20 µm.*

4.3.4.2. Cells transfection with different TCRβ constructions

Three different synthetic constructions were individually introduced into the Jurkat-KOβ cells. These constructs are detailed in section 3.13. Synthetic TCRβs. To recap, the different constructions consisted of a conventional TCRβ ("Canonical" – C), a non-canonical rearrangement where the TRVB segment was a fusion between TRVB6-1 and TRVB27 ("Non-Canonical" – NC), and a shorter TCRβ where TRCB was not present ("Non-Productive" – NP).

Figures 30A-D and 30F-I show flow cytometry plots illustrating cell viability and CD3 expression, respectively, after the electroporation process. Unstained controls of Jurkat-WT and Jurkat-KOβ cell cultures were also performed to verify autofluorescence after the process (Figure A 4).

Regarding viability, no significant differences were observed between Jurkat-WT cells and any Jurkat-KOβ transfection (Figure 30E). Similarly, CD3 expression, indicating the presence of CD3-TCR complexes on the cell surface, showed no significant differences between WT cells and those with the C construct. On the other hand, cells with the NC (66.57% ± 6.84) or NP (14.27% ± 2.12) construction exhibited significant differences compared to the WT cells (93.45% ± 0.91) (Figure 30J). Remarkably, despite a slight decrease in CD3-TCR expression in NC cells compared to WT or C cells, a significant percentage of Jurkat-KOβ cells demonstrated the capability to express this non-canonical rearrangement, evidenced by the notable increase regarding NP cells ($p=0.0033$).

Once again, these findings shed light on the potential functionality of non-canonical TCRβs in the immune repertoire. The detection of these variants, albeit at low frequency, in PBMCs from patients and, importantly, in solid tissue, indicates their presence in physiological contexts. Furthermore, the observation that these TCRβs are not mere artefacts, coupled with the demonstration of their expression in Jurkat-KOβ cells, adds depth to our understanding of the complexity and diversity of TCR repertoires.

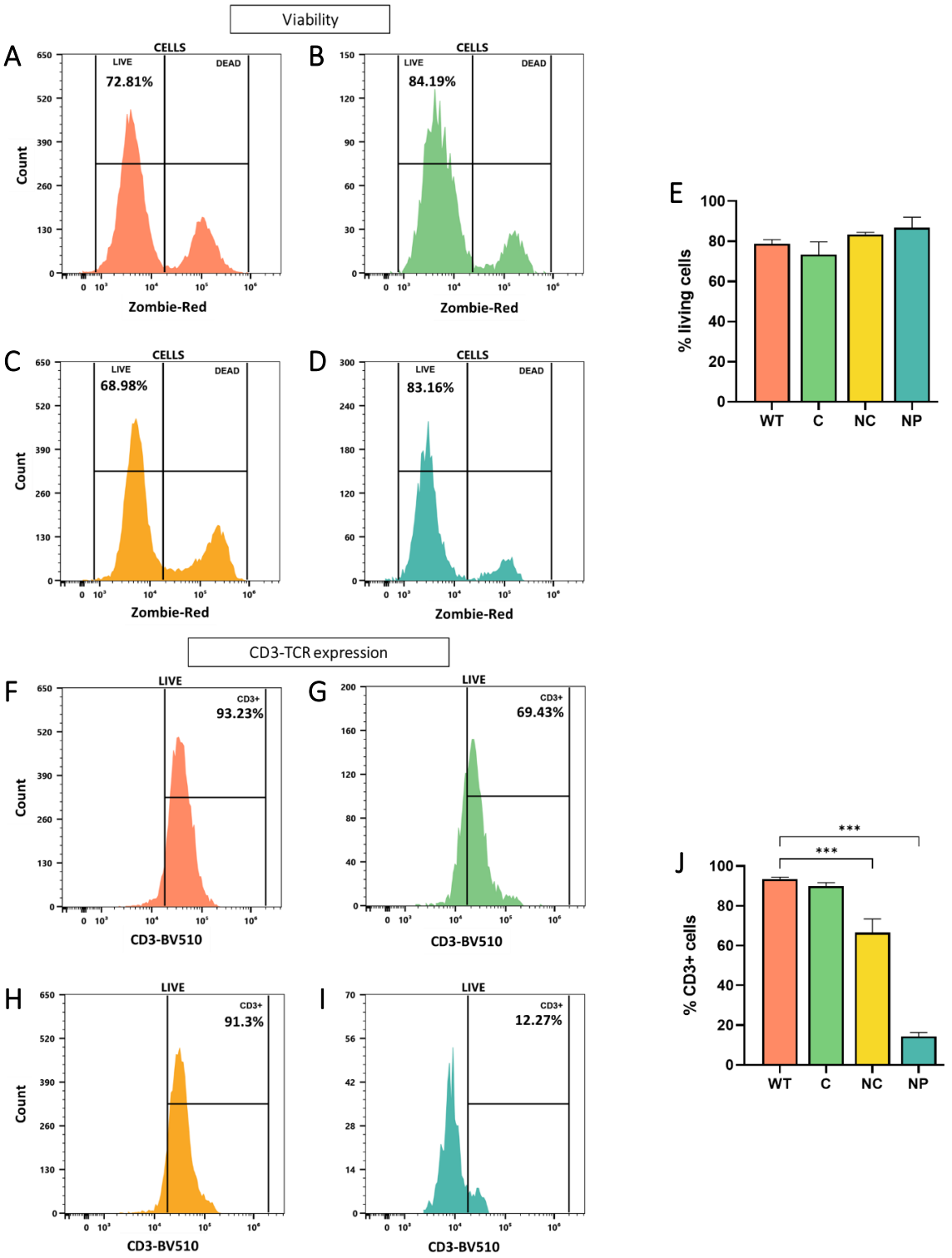

Figure 30. *Transfection through electroporation assays. Cell viability after electroporation of* ***(A)*** *Jurkat-WT cells,* ***(B)*** *Jurkat-KOβ cells transfected with the canonical TCRβ construction,* ***(C)*** *Jurkat-KOβ cells transfected with the non-canonical TCRβ construction, and* ***(D)*** *Jurkat-KOβ cells transfected with the non-productive TCRβ construction.* ***(E)*** *Viability statistical analysis (N=3). CD3-TCR expression after electroporation in living cells of* ***(F)*** *Jurkat-WT,* ***(G)*** *Jurkat-KOβ transfected with the canonical TCRβ construction,* ***(H)*** *Jurkat-KOβ transfected with the non-canonical TCRβ construction, and* ***(I)*** *Jurkat-KOβ transfected with the non-productive TCRβ construction.* ***(J)*** *CD3-TCR expression statistical analysis (N=3). WT: wildtype cells, C: cells with the canonical TCRβ, NC: cells with the non-canonical TCRβ, NP: cells with the non-productive TCRβ. ***: $p \leq 0.001$.*

Chapter IV:

Small RNA NGS-library preparation

4.4. Small RNA NGS-library preparation

This chapter details the complete process of preparing small RNA libraries for NGS (Figure A 5). It is important to emphasize that this methodology was designed to prepare miRNA libraries specifically. However, it captures all kinds of small RNAs. Contrary to the TCRβ library preparation, this technology is designed for the Illumina sequencing platform, which uses other sequences for its adapters. In this case, the methodology created was designated with the name Double-Tailing Trap (DTT) (patent number: P202031119).

This technology arises from the one explained in Chapter 1. In this case, it emerges from the need to create a method capable of sequencing all miRNA variants in a biological sample without knowing their sequence and avoiding the ligation step. This reaction is the main source of biases. Moreover, the bias is greater when the molecule is smaller, which under or overestimates particular variants in the case of miRNAs.

Oligonucleotides and templates used for this procedure were also purchased from IDT (Coralville, IA, USA) and they are shown in Table A 6. To test the methodology, synthetic miRNAs were also purchased from IDT, as well as a Simulator (synthetic RNA with a poly-A already present at its 3' end) to perform positive controls (Table A 7).

4.4.1. Functionalization of miRNA-beads

Diverse beads were also tested for this procedure, both commercial and non-industrial crafted by collaborators. The best results were obtained by the Sera-Mag™ SpeedBead Carboxylated-Modified [E3] Magnetic Particles (Cytiva, Marlborough, MA, USA). Hereafter, these microparticles, once functionalized, will be referred to as "miRNA-beads". Other beads tested exhibited either incompatibility with the procedure or lower efficiency than the chosen ones. The predominant issue was the instability during the alkaline washes, leading to an aggregation of particles, and inhibition of common molecular biology reactions.

Sera-Mag™ SpeedBead Carboxylated-Modified [E3] Magnetic Particles are spheres of ~1 µm diameter with a double layer of magnetite, improving their magnetic response. The presence of carboxylic groups on their surface, negatively charged, ensures their stability, and facilitates the covalent coupling of amino groups. These beads are well-suited for our technology due to their elevated magnetism, enhanced mobility, low sedimentation rate, non-inhibition of reactions, and compatibility with high temperatures and extreme pH conditions.

The first step of the preparation of NGS libraries of small RNAs is the functionalization of microparticles. We coupled the oligonucleotide SP1-F (hybridised with its complementary so that nitrogen bases do not interact with carboxylic groups on the beads' surface) (Table A 6) that included the sequence of adapter SP1 and a poly-T tail of 20 thymines at its 3' end (Figure 31). This T-tail will be the part "trapping" the molecules of interest (4.4.3. "Trapping" miRNAs), hence the name of the technology.

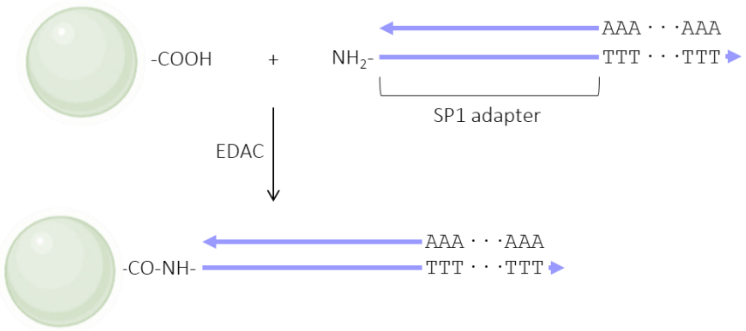

Figure 31. *Diagram of the process of the miRNA-bead functionalization. Arrowhead simulates the 3' end of the oligonucleotide. It must be noted this is just a scheme, beads are covered all over with the same oligonucleotide.*

Each hybridised-oligonucleotide (0.2 nmol per milligram of particles) is coupled via a 5' amino (NH2) modification of the forward primer to the beads (10 mg/ml) in a 1M NaCl + 100 mM MES at pH 5 medium supplemented with 1.25 M of EDAC. The latest establishes the amide bond between the 5' amino end of the oligonucleotides by activating the carboxylic groups on the surface of the microparticles. The reaction needs to be vortex for 10 min and then incubated for 4 hours with continuous rotation.

4.4.2. Polyadenylation of miRNAs

The main purpose of this methodology is to discover new variants of miRNAs. For this reason, after the miRNA extraction of a sample, miRNAs need to be adenylated so that we can capture all variants without knowing any specific sequence.

Adenines are included with a poly(A) polymerase (PAP) enzyme derived from yeast (Jena Bioscience, Jena, Germany) following the manufacturer's instructions with an incubation time of 30 min at 37 ºC. We must highlight that this process has been set up so that the adenines added are more than 20. In this way, the miRNA can be efficiently captured by the poly-T tail, which is 20 bases long. The reaction is stopped by adding 100 µl of "hybridization buffer" (TRIS-HCl 50 mM + NaCl 250 mM at pH 7.7) supplemented with EDTA 12 mM.

4.4.3. "Trapping" miRNAs and retrotranscription on the surface of the beads

Once adenylated, miRNAs are ready to be trapped by the poly-T tail in the "miRNA-beads" (Figure 32A). To do so, 10 µl of functionalized particles are incubated with the product of the adenylated miRNAs for 1 h at 37 ºC, 1 h at 30 ºC, 1 h at 20 ºC and overnight at room temperature with continuous rotation. The following day, beads are washed with the "hybridization buffer" twice to eliminate anything not bound to them.

Next step is to retrotranscribe miRNAs into cDNA on the surface of the particle (Figure 32B). To do so, we use the M-MLV reverse transcriptase from Promega (Madison, WI, USA) and follow the manufacturer's instructions with 1 hour at 37 ºC incubation. Finally, beads are washed with NaOH 250 mM (Figure 9), and pH is restored with TRIS-

HCl 200 mM and 50 mM. Thus, everything not covalently bound to the particles will be eliminated and we will be left with single-stranded DNA to proceed with the methodology.

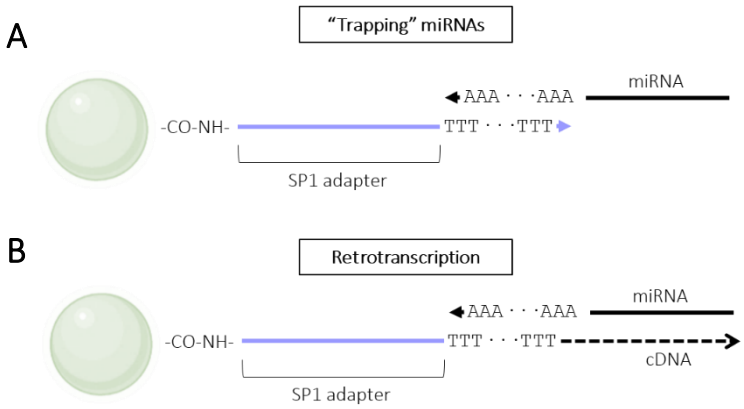

Figure 32. ***(A)*** *Diagram of the "trapping" miRNAs procedure.* ***(B)*** *Diagram of the retrotranscription process. Arrowheads simulate the 3' ends.*

4.4.4. Second tailing of the trap

To continue with the procedure, we will include the second tail of nucleotides at the 3' end of the molecules coupled to the beads (Figure 33). In this case, the tail will be composed of guanines. For this purpose, the terminal deoxynucleotidyl transferase (TdT) enzyme is used (Promega, Madison, WI, USA). This enzyme prefers guanines although it can include any other nucleotide. In this case, we only add dGTPs to the reaction and follow the manufacturer's instructions. During the development of the technology, incubation time was set up at 20 min at 37 ºC. Beads must be immediately washed (Figure 9) to stop the reaction with NaOH and so that reagents are eliminated.

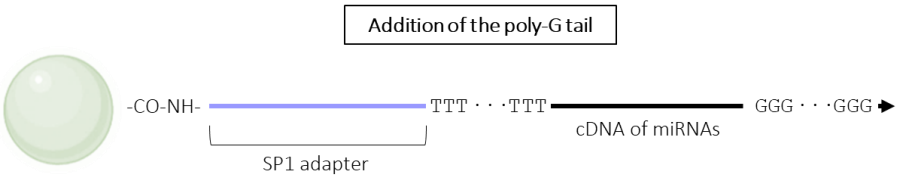

Figure 33. *Diagram of the addition of the second tail of nucleotides, a poly-G tail. Arrowhead simulates the 3' end.*

4.4.5. Including the SP2 adapter sequence

Thanks to the poly-G tail added in the previous step, we can include the sequence of the SP2 adapter. We use a template consisting of a poly-C tail at its 3' end and the sequence of the SP2 adapter at its 5' end. The poly-C tail hybridizes with the poly-G tail and the DNA strands on the beads elongate the SP2 adapter sequence (Figure 34). In this case, the tail of cytosines of the template must be larger than the guanines one. It is important to emphasize that this template has its 3' end blocked with a chemical modification. Thus, the template does not acquire the rest of the sequence and will not

disturb the reaction. As in previous steps, beads are washed to eliminate any reagents or DNA not covalently bound to them (Figure 9).

By an "extension PCR" of 10 cycles with GoTaq® G2 Flexi DNA Polymerase (Promega, Madison, WI, USA), the single-stranded DNA included the SP2 adapter sequence. miRNA-beads were resuspended in the "extension PCR" reaction volume (100 µl), which followed the manufacturer's instructions with 200 nM of the template oligonucleotide, 1.5 mM of MgCl2 and 0.2 mM of dNTPs. The protocol consisted of a pre-denaturation of 2 min at 95 ºC, followed by 10 cycles of denaturation at 95 °C for 30 s, annealing at 62 °C for 5 min, and extension at 72 °C for 30 s, and a final extra extension of 5 min at 72 ºC.

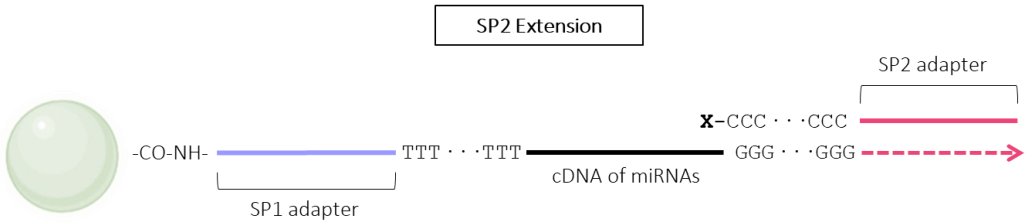

Figure 34. *Diagram of the elongation of the SP2 adapter. X represents the blocked 3' end of the template. Arrowhead simulates the 3' end.*

4.4.6. Releasing the complete library from the miRNA-beads and completing adapters

The last step of the DTT methodology is to release the molecules from the beads and complete the sequence of the Illumina adapters. At the same time, an index will be included to identify the samples. This stage involves two subsequent PCRs with GoTaq® G2 Flexi DNA Polymerase (Promega, Madison, WI, USA).

The first one releases the libraries from the miRNA-beads. It employs a pair of primers against the SP1 and SP2 adapters. The forward primer will add a part of the P5 adapter, and the reverse primer will add the index and a part of the P7 adapter (Figure 35A). The second one will complete the P5 and P7 adapters (Figure 35B). To continue with the second PCR, it is not necessary to carry out a purification of the first PCR product. Instead, 5 µl of the first PCR product will be included in the second PCR reaction.

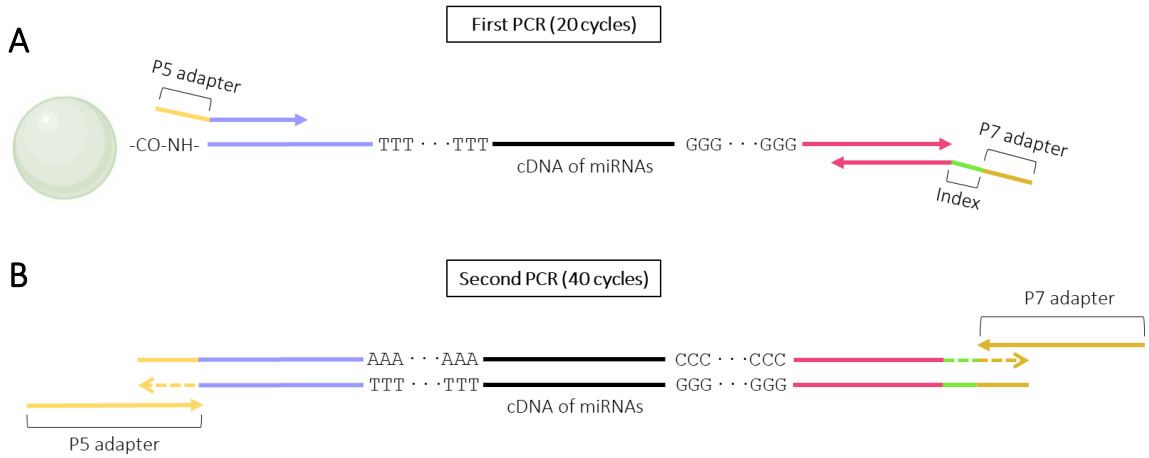

Figure 35. (A) *Diagram of the first PCR, which releases the molecules from the beads and adds an index a part of P5 and P7 adapters.* ***(B)*** *Diagram of the second PCR, which completes P5 and P7 adapters. Arrowheads simulate the 3' ends.*

The first PCR will be carried out in a volume of 20 µl with 75 nM of both primers, 1.5 mM of MgCl2 and 0.2 mM of dNTPs. The protocol consisted of a pre-denaturation of 2 min at 95 ºC, followed by 20 cycles of denaturation at 95 °C for 30 s, annealing at 64 °C for 30 s, and extension at 72 °C for 30 s, and a final extra extension of 5 min at 72 ºC.

The second PCR will be carried out in a volume of 50 µl with 5 µl of the first PCR product, 200 nM of both primers, 1.5 mM of MgCl2 and 0.2 mM of dNTPs. The protocol consisted of a pre-denaturation of 2 min at 95 ºC, followed by 40 cycles of denaturation at 95 °C for 30 s, annealing at 64 °C for 30 s, and extension at 72 °C for 30 s, and a final extra extension of 5 min at 72 ºC. Final amplicons, from 5' to 3', are composed of the P5 + SP1 adapter, a poly-A tail, the population of miRNAs, a poly-C tail, the SP2 adapter, an index and the P5 adapter as seen in Figure 36.

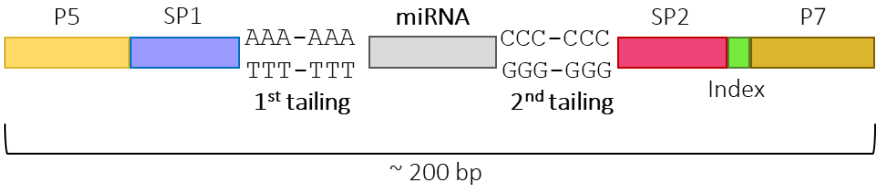

Figure 36. *Diagram of the final amplicons obtained by the Double-Tailing Trap methodology. bp: base pair.*

4.4.7. Blocking miRNA-free oligonucleotides

When trapping miRNAs some of the oligonucleotides attached to beads will not capture any. This excess of miRNA-free oligonucleotides will continue with the process, leading to final "empty libraries" (without any miRNA sequence between the two tails) and worsening the methodology efficiency. For this reason, those oligonucleotides need to be blocked before adding the poly-G tail. We tested different types of blocking methods. Here, we present two different blocking techniques, which are the most efficient ones for this methodology.

The first one, referred to as "hairpin-blocking", consists of an incubation of the miRNA-beads with a hairpin-like primer to block the miRNA-free oligonucleotides (Figure 37A). Incubation is carried out in the "hybridization buffer" and the "decreasing temperature protocol". Later, beads are washed only with the "hybridization buffer" to eliminate anything not bound to them.

The second one, referred to as "dideoxy-blocking", consists of adding a dideoxynucleotide that will prevent the elongation through 3' of the miRNA-free oligonucleotides. To do so, we used a blocking-template, that hybridizes with the miRNA-free oligonucleotides attached to the miRNA-beads, with a longer tail of adenines to efficiently incorporate a dideoxythymidine (ddT) (Figure 37B). For this procedure, we selected the Therminator^{TM} DNA Polymerase (New England Biolabs, Ipswich, MA, USA). The total volume of the reaction was 100 µl with 300 nM of the blocking-template, 1.5 mM of $MgCl_2$ and 0.2 mM of ddTTPs. The protocol consisted of a pre-denaturation of 2 min at 95 ºC, followed by 40 cycles of denaturation at 95 °C for 20 s, annealing at 60 °C for 20 s, and extension at 72 °C for 20 s. Finally, beads are washed to eliminate everything not covalently bound to them (Figure 9).

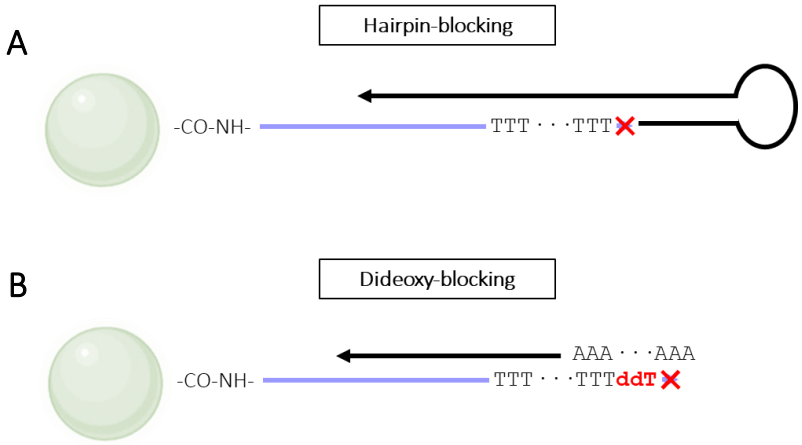

Figure 37.(A) *Diagram of the hairpin-blocking procedure.* ***(B)*** *Diagram of the dideoxy-blocking procedure. Arrowheads simulate the 3' ends. ddT: dideoxythymidine.*

To compare both methods, we prepared libraries without "trapping" any miRNAs (as blanks of blocking). Of the two blocking methods examined, the enzymatic addition of ddTTP (dideoxy-blocking) demonstrated greater efficiency compared to the hairpin-blocking, as illustrated in Figure 38. The use of the hairpin-like oligonucleotide resulted in a higher quantity of DNA (lane 2) and similar size to the positive control (+). The little difference in size indicated libraries without any miRNA but with the rest of the process completed. On the other hand, the dideoxy-blocking result (lane 1) exhibited a reduction in the amount of DNA and size of the molecule, indicating a more effective miRNA-free oligonucleotide blocking.

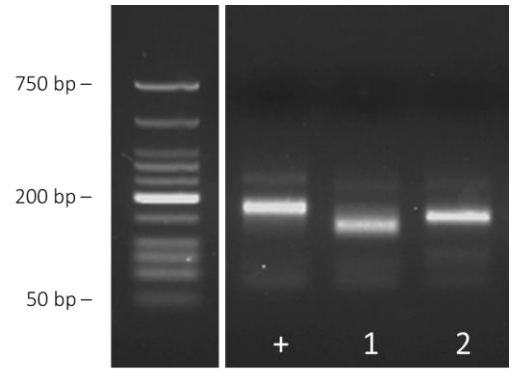

Figure 38. *Comparison between dideoxy-blocking (lane 1) and hairpin-blocking (lane 2). +: positive control with 1 nM of Simulator. Agarose 2.2%. Modified image.*

As depicted in Figure 38, it is evident that the dideoxy-blocking was not flawless. Thus, we aimed to improve this step to prevent "empty libraries" from interfering with the following reactions and masking the completed ones (especially when handling samples with lower concentrations of miRNAs). To improve the efficiency of blocking, modifications were made to other steps of the DTT methodology. These modifications resulted in an evolved version of the DTT technology, which is elaborated in the following sections of the thesis.

4.5. Improving the Double-Tailing Trap methodology

Double-Tailing Trap evolved to DTT-Click Chemistry (CC) (patent number: EP23383196) reducing work time, increasing efficiency, and reducing its detection threshold. With the incorporation of click chemistry into the technology, we can perform the initial steps in solution rather than on the beads' surface, which improves the methodology sensitivity. Likewise, the process of blocking miRNA-free oligonucleotides is replaced with an elimination of those oligonucleotides instead.

A detailed schematic overview of the evolved DTT technology is depicted in Figure A 6. Oligonucleotides and templates employed for this procedure were identical except for the 5'-modification on the oligonucleotide "trapping" miRNAs and the incorporation of the one designed to eliminate the miRNA-free oligonucleotides (Table A 6).

4.5.1. Functionalization of miRNA-beads for CC

The process mirrored that outlined in section 4.4.1, but instead of using an oligonucleotide with a poly-T tail, we coupled another molecule called propargylamine to the beads (Figure 39). Propargylamine is an organic compound with the formula $HC{\equiv}CCH_2NH_2$. On one end it has an alkyne group and an amine group on the other. The latter was bound to the particles using EDAC, which activated the carboxylic groups on the particles' surface. A total of 8 µmol of propargylamine per milligram of beads was used in a medium consisting of 1M NaCl + 100 mM MES at pH 5 supplemented with

1.25 M of EDAC. Hereafter, these microparticles, once functionalized, will be referred to as "miRNA-beads for CC".

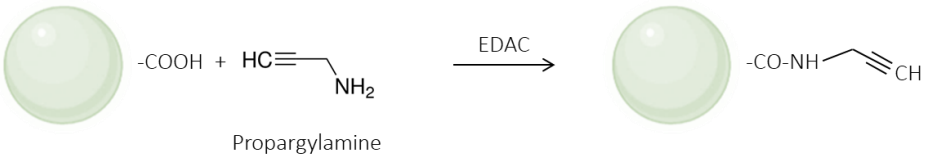

Figure 39. *Diagram of the process of the miRNA-bead for CC functionalization. It must be noted this is just a scheme, beads are covered all over with propargylamine molecules.*

4.5.2. Steps in solution

It is important to highlight that the chain reactions described in this section were carried out with the M-MLV buffer (Promega, Madison, WI, USA). No significant differences were observed between the M-MLV buffer and the buffers specific to each enzyme.

- **Polyadenylation of miRNAs** was similar to the one explained in section 4.4.2. On this occasion, to stop the reaction, EDTA was reduced to 3.7 mM to prevent it from impeding the next reactions.

- **"Trapping" miRNAs and retrotranscription** were performed in solution, not over the particles' surface. miRNAs, once adenylated, were trapped by the same oligonucleotide with a poly-T tail at its 3' end but with a 5'-azide modification (Figure 40A). This 5'-modification is essential for the future click chemistry reaction. A total volume of 20 µl was added to the product of the adenylated miRNAs containing 1x M-MLV buffer, 50 nM of the 5'-azide-modified oligonucleotide, 7,5 mM of $MgCl_2$ and 0.2 nM of dNTPs. This reaction was incubated following the "decreasing temperature protocol". After the incubation, the reaction was supplemented with 200 units of the M-MLV reverse transcriptase enzyme to proceed with the retrotranscription of miRNAs into cDNA. Then, the reaction was incubated for 30 min at 37 ºC (Figure 40B).

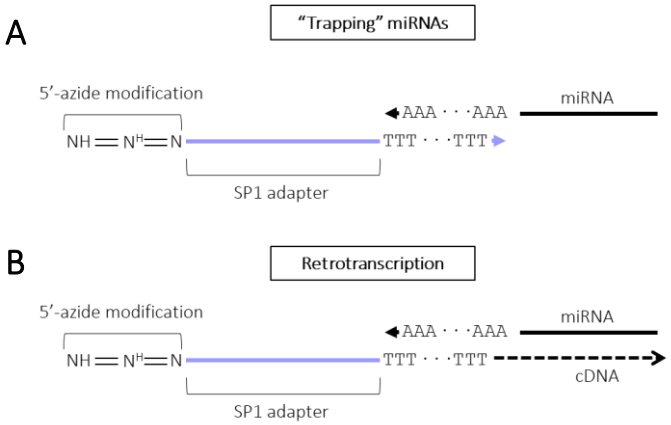

Figure 40. (A) *Diagram of the "trapping" miRNAs procedure.* ***(B)*** *Diagram of the retrotranscription process. Arrowheads simulate the 3' ends.*

4.5.3. Elimination of the miRNA-free oligonucleotides

Thanks to the previous reactions being in solution, we can get rid of the miRNA-free oligonucleotides using magnetic beads instead of trying to block them. For this aim, we used the Sera-Mag™ Streptavidin-Coated Magnetic Beads (Cytiva, Marlborough, MA, USA) and a 5'-biotin-modified oligonucleotide that hybridizes with part of the miRNA-free oligonucleotides (Figure 41).

For this purpose, we continued with another chain reaction where we added 40 µl to the volume of the preceding reactions. These 40 µl contained 1x M-MLV buffer, 200 nM of the 5'-biotin-modified oligonucleotide and 10 µl of the streptavidin-coated beads (10 mg/ml). This reaction was incubated according to the "decreasing temperature protocol", followed by a subsequent 1 h incubation at room temperature with continuous rotation. Afterwards, beads were placed on a magnet to discard them and keep the supernatant.

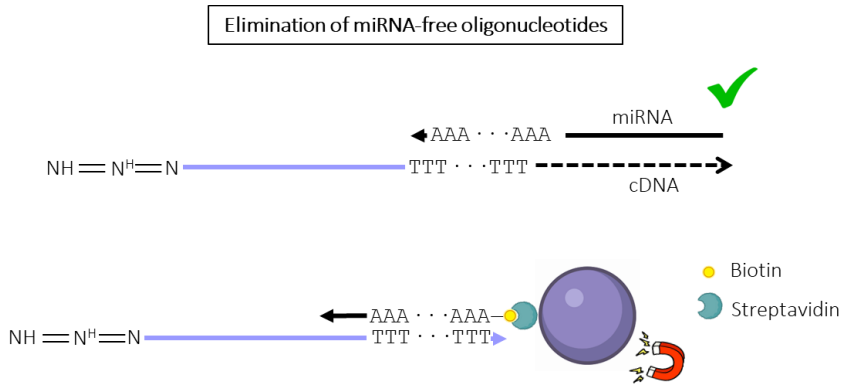

Figure 41. *Diagram of the elimination of miRNA-free oligonucleotides through streptavidin-coated magnetic beads. Arrowheads simulate the 3' ends.*

4.5.4. Click chemistry

Click chemistry is the term used to describe the union of alkyne and azide groups into a triazole using Cu(I) as a catalyst. The reaction shows very high yields at room temperature and requires no solvents. While this reaction (the Huisgen azide-alkyne cycloaddition) has been known since 2001, in the last decade, it has found numerous applications in biotechnology due to its ability to assemble macromolecular or colloidal components quickly and simply.

To accomplish the click chemistry reaction, we placed 10 µl of the miRNA-beads for CC on a magnet to remove the TRIS-HCl medium and resuspend them with the resulting volume from the previous reactions supplemented with 1 nM of $CuSO_4$ and 10 mM of sodium ascorbate (Figure 42). This reaction was carried out in a nitrogen atmosphere (to avoid oxidation of any compound) and incubated for 1h at room temperature with continuous rotation. Finally, particles were washed to eliminate any reagents or DNA not covalently bound to them (Figure 9).

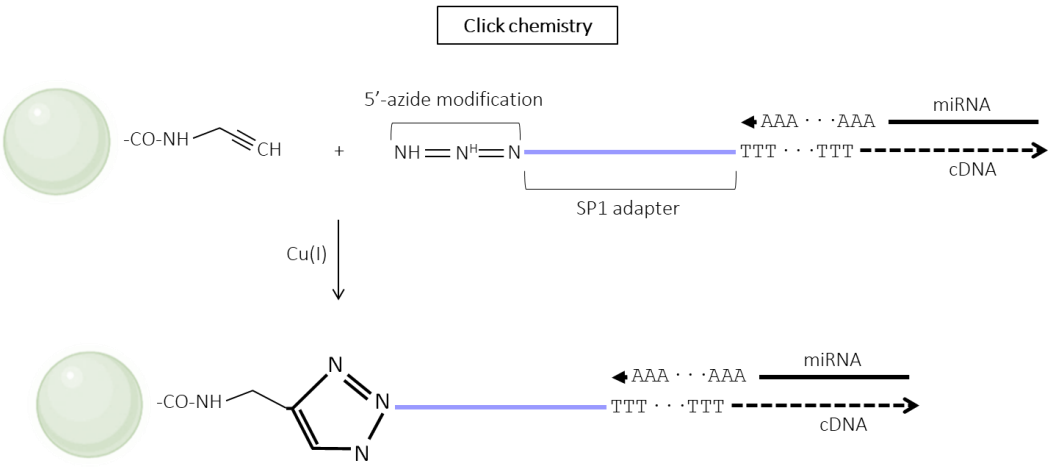

Figure 42. *Diagram of the click chemistry reaction. Arrowheads simulate the 3' ends.*

4.5.5. Steps on the surface of the beads

From this point forward, the rest of the process is the same as in the initial DTT methodology.

- Second tailing of the "trap".
- Including the SP2 adapter sequence.
- Releasing the complete library from the miRNA-beads for CC and completing adapters.

4.5.6. Improvements achieved with the evolution of the methodology

The DTT-CC version improved the original methodology in diverse aspects outlined in Table 12. These aspects consisted of a lowered detection threshold, decreased in working hours and days, removal of the overnight incubation, reduction of washing steps, and modification of an inefficient blocking step.

Table 12. *Comparison between DTT and DTT-CC methodologies.*

	DTT	DTT-CC
Detection threshold	700 pg	0.7 pg
Working days	3	2
Laboratory work*	8 h 40 min	7 h 30 min
Overnight incubations	1	None
Washing steps	5	3
Capacity to eliminate miRNA-free oligonucleotides	No	Yes

* Overnight incubations and washing steps not included

The most important one is the sensitivity reduction. To do so, we employed synthetic miRNAs to identify our minimum threshold. As depicted in Figure 43A, using the DTT original methodology, we could "trap" miRNAs when present in concentrations up to 1 nM (= 700 pg) (lane 2). When concentrations were lower like 0.1 nM (= 70 pg) (lane 3) or 0.01 nM (= 7 pg) (lane 4), a slight decrease in size is observed, suggesting that

libraries are "empty". On the other hand, using DTT-CC methodology, we observed that even the lower concentration (0.01 nM = 7 pg, lane 5) maintains the correct library size (Figure 43B).

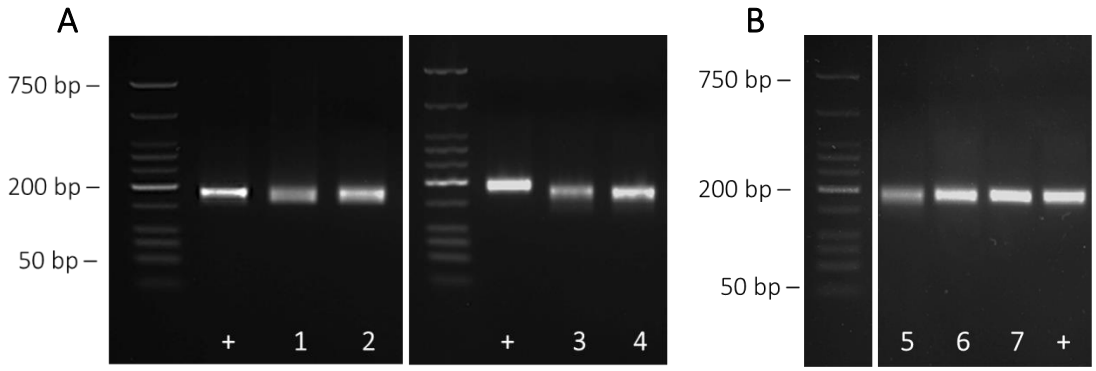

Figure 43. *Detection threshold experiment with different concentrations of synthetic miRNAs.* ***(A)*** *Using DTT methodology. Lane 1: 10 nM. Lane 2: 1 nM. Lane 3: 0.1 nM. Lane 4: 0.01 nM.* ***(B)*** *Using DTT-CC methodology. Lane 5: 0.01 nM. Lane 6: 0.1 nM. Lane 7: 1 nM. +: positive control with 1 nM of Simulator. Agarose 2.2%. Modified image.*

Another important improvement is the capability to eliminate miRNA-free oligonucleotides. Figure 44 shows the differences in final libraries when blocking these oligonucleotides with either hairpin-blocking or dideoxy-blocking and when eliminating them. As explained before, neither of the blocking systems was complete as both final libraries show waste products of a lower size compared to the positive control, indicating "empty" libraries once again (Figure 44A). Nevertheless, as shown in Figure 44B, eliminating miRNA-free oligonucleotides from the reaction via streptavidin-coated magnetic beads resulted in no "empty" libraries at all, just a few primer dimers at the end of the process can be observed at the bottom (lane 3).

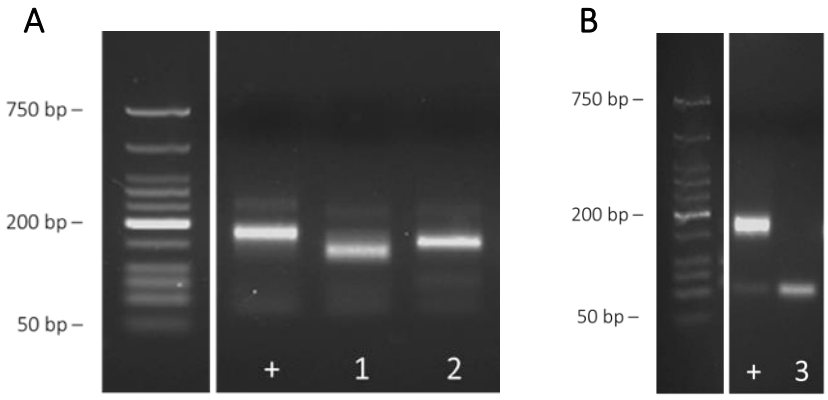

Figure 44. *Comparison between* ***(A)*** *dideoxy-blocking (lane 1) and hairpin-blocking (lane 2) with DTT methodology,* ***(B)*** *and the elimination of miRNA-free oligonucleotides in the DTT-CC methodology (lane 3). +: positive control with 1 nM of Simulator. Agarose 2.2%. Modified image.*

5. Discussion

5.1. The massive sequencing era

After the discovery of the structure of DNA in 1953 (J. D. Watson & Crick, 1974), 24 years later Sanger sequencing was developed to decipher genomes (Sanger et al., 1977). However, using Sanger sequencing to sequence large genomes was extremely expensive and laborious. Next-generation sequencing (NGS) appeared in the early 2000s, capable of generating millions of short reads in a few days. Thus, increasing the number of genomes sequenced in the following years. Moreover, NGS techniques did not require bacterial cloning and electrophoresis separation as Sanger sequencing did.

Nowadays, diverse NGS platforms exist and NGS technologies have become a standard tool in diverse fields such as the study of biological processes, genetic diagnosis, evolution, and agronomical research, etc. Nevertheless, despite the transformative impact of NGS, it has limitations. As addressed in the introduction, the principal sources of error and limitation lie in the library preparation, read length, error rates, or the GC content. Companies are in constant development of new technologies or solutions to overcome these challenges and improve the reproducibility and reliability of data.

One of the main problems is the read length. Genomes need to be assembled by piecing together thousands of short reads. Sometimes, when studying single-nucleotide variants (SNV) or indels, short reads are sufficient, but they cannot phase alleles to their parental homolog. On the other hand, larger structural variants (SVs) are, on most occasions, undetectable. Similarly happens with repeated sequences, which may be longer than the read. Furthermore, NGS processes rely on pre-amplification by PCR, adding new potential sources of bias due to inefficient amplification of certain regions, like GC-rich sequences. Therefore, there is a need to improve methodologies to surpass these challenges (van Dijk et al., 2018).

In 2011 and 2014, long-read sequencing (LRS) emerged from PacBio (single molecule real-time, SMRT) and ONT (nanopore sequencing), respectively. Besides the capability of generating long-read data, both technologies are characterized by the real-time sequencing process and the absence of pre-amplification. In the beginning, these technologies presented low accuracy rates compared to NGS (60-70% in ONT and 80% in PacBio) (van Dijk et al., 2023). In the last decade, LRS has been in continuous improvement and both technologies have evolved to address their limits.

The technology behind Pacific Biosciences, SMRT, employs a circular ssDNA (single-stranded DNA), referred to as SMRTbell, formed by attaching hairpin adapters to both ends of the target dsDNA (double-stranded DNA). One primer and polymerase are added to the adapter, and the library is loaded onto a specialized flow cell known as zero mode waveguides (ZMWs). A single polymerase is immobilized at the bottom of each ZMW, which is transparent. During the replication of the target molecule, it emits fluorescence signals during each fluorescently-label nucleotide incorporation in real time. A camera records the colour and the duration of emitted light. Furthermore, the interpulse duration (IPD), the time between nucleotide incorporations, enables the detection of base

modifications such as methylations. The closed-circle SMRTbell structure allows continuous sequencing of both DNA strands until polymerase exhaustion, resulting in a continuous long read (CLR). This CLR can be divided into subreads removing adapter sequences, culminating in a circular consensus sequence (CCS) with enhanced accuracy (Figure A 7) ((Eid et al., 2009; van Dijk et al., 2018)).

This approach solves part of the limitations of NGS as it does not need a long library preparation procedure with potential error points, beyond the adapters' attachment. Nevertheless, this still can introduce biases. Another major achievement is the absence of a pre-amplification process, being able to sequence single molecules and doing it in real time. In the past, researchers needed to choose between noisy long reads or accurate shorter reads of around 2 kb. Notably, sequencing chemistries have improved read-length and accuracy, reaching 200 kb or HiFi 10-25 kb with 99% accuracy thanks to CCS (Logsdon et al., 2020; Wenger et al., 2019). Longer HiFi read lengths could potentially be achieved through advancements in polymerase chemistry or the reduction of raw error rates, as well as improvements in software to generate high-quality consensus sequences like DeepConsensus (Baid et al., 2023).

On the other hand, nanopore sequencing by ONT takes place within a flow cell, featuring two compartments (*cis* and *trans*) filled with ionic solutions and separated by a membrane that has individual nanopores. An applied constant voltage induces an ionic current through the nanopore. When a DNA or RNA molecule translocates, a discernible change in the ionic current is observed. One of the steps that remains in this sequencing process as well is the library preparation. For nanopore sequencing, it involves end-repairing DNA fragments, followed by adapter ligation. Adapters, forming DNA–protein complexes, incorporate a polymerase or helicase enzyme, facilitating stepwise DNA movement through the pore via a ratcheting mechanism. dsDNA unwinds at the pore, and one strand passes through it (Figure A 8). For enhanced accuracy, the mechanism "$1D^2$" was developed, ensuring the sequencing of the second strand after the completion of the first one. To do so, specialized adapter sequences are used to increase the likelihood of the second strand following the first one through the nanopore (van Dijk et al., 2018).

Similarly to SMRT sequencing, nanopore sequencing avoids long-library preparation processes as well as pre-amplification of the target molecules, eliminating important sources of bias. However, its performance was not ideal at the beginning, with 30-40% error rates due to the unique sequencing of a single strand ("1D"). ONT reduced error rates to 7-9% when the sequencing of dsDNA was achieved ("2D") with a hairpin adapter. Finally, "$1D^2$" replaced the latest system, reaching up to 99.6% accuracy when setting lower temperatures (28 ºC) and the slowest sequencing speed (van Dijk et al., 2023). Apart from improvements in its chemistry, ONT has been developing different pores and base-calling software, which have also reduced error rates. One of the major achievements of nanopore sequencing is that read lengths are supposedly limited by the target molecule. Thus, with extremely intact DNA, reads can reach up to several megabases (Payne et al., 2019).

Alternatively, Illumina launched a library preparation kit (Complete Long-Read) to generate “linked-reads”. Link-read technologies are not the same as LRS. It is based on short reads corresponding to a longer DNA that includes tags all along the fragment for a better assembly afterwards. This new technology is an affordable option in laboratories with Illumina platforms, but sequencing data are not real long-reads, and it still requires a library preparation process (van Dijk et al., 2023).

Regarding short reads, Illumina is still the gold standard due to low error rates and high throughput. Nonetheless, PacBio has introduced a short-read sequencing system called “Onso” based on sequencing by binding technology, claiming higher accuracy. Similarly, ONT now provides a solution for short-read sequencing using the same technology designed for long reads. While nanopore sequencing was not initially designed for small templates, the recent introduction of the 'short fragment mode' allows the inclusion of reads as short as 20 nucleotides (van Dijk et al., 2023).

5.2. Theoretical comparison between TCRβ library preparation methodologies

During this thesis, there has not been enough time to compare the methodology created for the TCRβ library preparation with other commercially available. For that reason, we will carry out a theoretical comparison with the existing commercial methods. However, we are aware that we cannot prove that our technology works better or worse than any of the commercial kits.

There are 5 main commercial technologies for TCR sequencing that obtain whole-TCRβ sequences. Each of those uses different library preparation methodologies, which can be divided into those that perform multiplex PCR (iRepertoire), multiplex PCR with adapters’ ligation (Illumina and Thermo Fisher), 5’RACE (Rapid Amplification of cDNA ends) (Takara Bio) and end-repair with one adapter ligation (QIAGEN) (Figure A 1). Of all of them, our technology is more similar to the first one (iRepertoire), but with the addition of magnetic particles that allow us to capture the molecules of interest (i.e., TCRβ) and concatenate reactions without noise and side products in between.

iRepertoire performs their patented arm-PCR (amplicon rescued multiplex-PCR). It consists of an RT-PCR1 with nested primers (inside and outside primers) that selectively amplify all TRVB and TRCB genes and incorporate communal adaptors to perform the PCR2. They claim that in the first cycles of RT-PCR1 outside primers improve the sensitivity of the reaction (enriching the sample with the molecules of interest) by first-strand synthesis. Then, inside primers amplify the templates and incorporate communal adapters at both ends. PCR1 is “rescued” by a clean-up step or by transferring part of the product directly into PCR2. During PCR2, only molecules with the communal adapters will be exponentially amplified and will incorporate sequencing adapters and indexes.

One of the things that catches our attention is that iRepertoire can perform a nested PCR in the TRVB region of the TCRβ. In our technology, we carry out a nested PCR only

through the TRCB segment. For that reason, we referred to it as semi-nested PCR. We tried to develop primers for a complete multiplex-nested PCR but primers in either exon 1 or 2 of the segment performed significantly worse than those in the 5' UTR. However, we cannot forget that in many cases, chemical modifications are added to improve primers' efficiency and reduce primer dimers.

Regarding the "rescue" part of the process, they suggest two choices for rescuing PCR1. The first alternative involves a purification step with beads and the second alternative is the transfer of the PCR1 product into the PCR2 mix. Depending on the final goal, researchers must choose from one of those. They claimed the first option enhances both the discovery of unique CDR3 and the overall quality of the library. The second one is quicker and simpler to execute, yet it still ensures good unique CDR3 discovery. In our opinion, the second option will introduce a lot of noise in the PCR2 amplification due to the high number of primers used in the RT-PCR1, despite being diluted. On the other hand, a purification step after 10-15 cycles of PCR1 will reduce the process' efficiency or eliminate certain TCR variants from the system. Moreover, this loss is dependent on the purification technique and kit used. On the contrary, our methodology does not accumulate waste products or previous reagents thanks to the use of magnetic microparticles. Likewise, washing steps simulate a clean-up step but with negligible loss of sequences.

Lastly, PCR2 is in charge of adding the complete sequencing adapters and indexes, which for Illumina are around 60 bp long. This "primer-tail" minimizes the amplification efficiency too, as it may be longer than the part of the primer hybridizing with the template. With our methodology, one of the adapters is already in the system thanks to the oligonucleotide covalently bound to the beads, that binds to the TRCB segment of the TCRβs. The 5' adapter is elongated afterwards with the completion of the last 3 nucleotides in the final PCR. Nevertheless, we must highlight that we developed the methodology for Ion Torrent, which adapters are around 30 bp long. If adapted for Illumina, we must check the viability of longer primers to complete sequencing adapters in the last PCR amplification of the process.

The rest of the technologies include one or two of the following steps: multiplex PCR, 5' RACE and ligation reaction/s. Multiplex PCR methods use a set of forward primers complementary to most known TRVB gene segments and a set of reverse primers for either J or C regions, depending on whether the starting material is genomic DNA (gDNA) or complementary DNA (cDNA), respectively. Generally, this kind of amplification also incorporates the adapters' sequences for subsequent sequencing on high-throughput platforms. Despite the potential introduction of biases and errors through multiple PCR rounds, several strategies exist to mitigate such issues. Using a smaller set of reverse primers to target the TRCB region (only two segments that can be incorporated in one primer) when starting with mRNA reduces PCR error compared to using 13 primers for TRJB segments in gDNA. Conversely, starting from gDNA eliminates the need for reverse transcription, minimizing errors introduced during cDNA synthesis (Pai & Satpathy, 2021).

In this case, we also use multiplex PCR in our technology. As explained in the results section, we carefully designed the collection of primers and tested individually their efficiency. Likewise, we decided to work with mRNA to avoid the excessive number of primers when working with gDNA. Additionally, we reduced the number of forward primers by selecting the 5' UTR conserved region of TRVB, amplifying various segments with the same primer. Thus, reducing primer dimers and noise in the process.

An alternative method instead of multiplex PCR involves 5′ RACE. In this technique, RNA undergoes RT using a reverse transcriptase enzyme with terminal transferase activity, adding untemplated C nucleotides to the 3′ end of the cDNA. A template switch oligonucleotide (TSO) with a small tail of complementary Gs nucleotides hybridizes to this untemplated region. This process allows the reverse transcriptase to switch templates and extend the cDNA to the end of the TSO, which includes part of one of the adapter sequences (Chen & Patton, 2001). Consequently, a single pair of primers targeting the 5′ adaptor and the TRCB segment amplifies all TCR rearrangements. It must be noted that reverse transcriptase includes 2-5 non-templated nucleotides with a preference for cytosine so for the TSO to hybridize accurately, the chemistry of the process might be permissive. Moreover, to perform 5' RACE the starting material must be high-quality RNA, which in some cases is not feasible (Woodsworth et al., 2013). Despite requiring fewer PCR rounds, 5′ RACE strategies remain susceptible to errors from PCR, template switching, or sequencing.

Illumina, Thermo Fisher and QIAGEN carry out the addition of adapters by a ligation reaction. Both Illumina and Thermo Fisher first enrich the sample with a multiplex PCR and partially digest amplicons for an adapters' ligation afterwards. In contrast, QIAGEN performs a selective RT of the TCR, then it ligates one adapter. Afterwards, it enriches the sample with a semi-nested PCR that selectively captures the TCRβ chain by its TRCB segment. It seems more optimal to ligate one of the adapters (which can be ligated to both ends of the amplicons) and then selectively enrich the sample. In this way, you avoid artefacts associated with a multiplex PCR. However, adapter ligation can be biased towards one end because of a higher affinity for certain sequences (Alon et al., 2011; Hafner et al., 2011; Jayaprakash et al., 2011). In that case, with the subsequent semi-nested enrichment PCR, those specific sequences would be lost because the adapter is not ligated at the appropriate end.

In like manner, the ligation of both adapters at the same time (Illumina and Thermo Fisher) may introduce potential artefacts. We must highlight that Illumina adapters are twice as long as the Thermo Fisher ones. Firstly, as previously discussed, certain sequences may be under or overrepresented due to a differential ligation efficiency of adapters (Alon et al., 2011; Hafner et al., 2011; Jayaprakash et al., 2011). Secondly, adapters may adhere together, creating "empty libraries". Thirdly, certain TCRβ sequences may incorporate the same adapters at both ends, leading to an under or overrepresentation once again. Lastly, adapters may bind in the opposite direction.

One of the key points in our methodology is the total absence of ligation steps. The first adapter is already included in the system through its covalent coupling to the

magnetic beads. After "fishing" TCRβ molecules with the beads, those oligonucleotides will elongate the complete TCRβ sequences. Likewise, thanks to a collection of templates, DNA strands covalently bound to the beads will elongate towards their 3' ends to incorporate the second adapter's sequence. In this way, using magnetic beads and a simple but high-fidelity DNA polymerase, we can include both adapters' sequences without the use of ligases.

Regarding clean-up steps, all the commercial kits perform at least one purification (either by beads or column) in addition to the final one. Every clean-up process involves the loss of a small percentage of the initial material. This loss may be due to an incomplete elution in cases of purification columns, where the DNA is retained and not fully recovered, or inefficient washes where DNA may be lost along with impurities. In our methodology, we take advantage of the particles' magnetism to wash them between the different steps. Thus, we carry out a purification-substitute process that allows us to change reagents easily to accumulate reactions without the waste products of the previous ones.

One of our weak points is the absence of unique molecular identifiers (UMIs). UMIs are short sequences of random nucleotides incorporated into individual DNA or RNA molecules during the initial steps of library preparation. The main goal of UMIs is to distinguish true biological molecules from those multiplied during the library preparation or sequencing processes. In a recent study, researchers benchmarked nine different TCRβ protocols using either 5' RACE or multiplex PCR, and with or without the use of UMIs to test their reproducibility, replicability, and sensitivity (Barennes et al., 2021). The results revealed method-specific repertoire profiles that exhibited consistency among replicates, indicating unique biases imposed by each protocol on capturing the TCR repertoire. The study highlighted the crucial role of starting material quantity. Finally, they suggested that, for studies aiming to capture maximal TCR diversity or identify rare clonotypes in highly diverse samples, non-UMI are more sensitive than UMI-based methods. Conversely, when the objective is to construct a representative clonal structure or identify TCRs of interest based on expansion levels, UMI correction was considered more suitable (Barennes et al., 2021). In this study, our interests rely on unique clonotypes, so UMIs were not strictly necessary. Nevertheless, such sequences could be added to our methodology in the future.

The conclusion we can draw from all the presented information is that all methodologies will have steps prone to introducing errors or biases in the TCRβ library preparation. Therefore, it is essential to methodically design each step and strive to minimize errors. Likewise, each researcher must analyse the specific case they are facing and consider which approach is most suitable for their study. In addition, researchers should remain vigilant and stay informed about the latest advancements in TCRβ sequencing methodologies. Continuous updates and improvements in techniques may offer new solutions to address potential challenges and enhance the accuracy and reliability of TCR library preparation.

5.3. Insights from human omental white adipose tissue (oWAT) and viral antigen responses

First of all, we must address the main differences between human white adipose tissue and murine WAT since the majority of data from this tissue is derived from studies on mice. A growing body of research emphasizes the importance of WAT in the immune system (reviewed in Trim & Lynch, 2022). However, there are significant differences in anatomy and physiopathology between species (Luong et al., 2019). In humans, intraperitoneal fat depots include omental and mesenteric compartments, while mice have mesenteric and well-developed gonadal depots, but lack the omental one (except for obese animals) (Bagchi & MacDougald, 2019). Unlike humans, who mainly expand omental fat depots during visceral expansion, rodents expand their gonadal depot, and both species can expand their subcutaneous fat depot. Notably, murine gonadal fat depot and human omental fat depot quite differ. Murine fat drains to the systemic circulation, while human omental adipose tissue drains to the portal circulation. Human oWAT also shows a higher abundance of immune cells and FALCs (Bénézech et al., 2015), providing oWAT with the strategic capability to filter a more diverse antigen repertoire. Both species exhibit obesity-related pathologies within the excessive expansion of intraperitoneal fat depots (Virtue & Vidal-Puig, 2008). Unfortunately, mesenteric depots remain less explored due to limited murine tissue and challenges in human biopsies collection.

This research pioneers the investigation into the immunological aspects of oWAT in humans. However, when interpreting the data, it is crucial to consider the context of morbid obesity, where the extensive expansion of body fat depots alters the composition and phenotype of adipose tissue leukocytes. Obesity leads to increased neutrophil infiltration, reduces resident mast cell number (García-Rubio et al., 2018; Lopez-Perez et al., 2021), induces a proinflammatory phenotype shift in macrophages, and disrupts mast cell function (Lopez-Perez et al., 2022; Lumeng et al., 2008). This chronic low-level inflammation, combined with factors like adipocyte hypertrophy, fibrosis, or tissue hypoxia, disrupts the normal functioning of adipose tissue, ultimately contributing to obesity-associated pathologies.

In a prior study of our group, we observed that substantial weight loss in individuals with morbid obesity did not result in changes to the proportion of T lymphocytes in adipose tissue, considered within the non-adipose cell ensemble, or the CD4+:CD8+ ratio (García-Rubio et al., 2018). Further investigations are required to understand how obesity influences the normal function of the adipose tissue T resident memory cells (Trm) reservoir. Additionally, the Trm population may play a role in the development of obesity-related pathologies. A murine model study suggested that the acute-phase infection of murine cytomegalovirus (mCMV) might impair glucose metabolism (Contreras et al., 2019). Several unresolved issues persist, including the immunological differences between human omental and mesenteric depots, as well as their relationship with mucosa-associated immunity. On the other hand, despite the status of Trm in the adipose tissue of lean individuals remaining unknown, it is reasonable to assume that their visceral fat depots also serve as significant reservoirs of Trm.

The noteworthy finding in this study underscores the significance of the omental fat depot. It is also in line with the fact that oWAT has a higher number of lymphocytes and FALCs than sWAT. The balance between CD4+ and CD8+ T lymphocytes, controlled by various factors, is biologically relevant. Although the thymus produces more CD4+ than CD8+ lymphocytes (Sinclair et al., 2013), the ratio between CD4+ and CD8+ T lymphocytes varies significantly between peripheral tissues (Sathaliyawala et al., 2013). In general, CD4+ T cells dominate the T cell pool throughout the body, except in the intestine, where CD4+ and CD8+ T lymphocytes are present in similar amounts (Sathaliyawala et al., 2013). Remarkably, in oWAT the predominance is led by CD8+ T cells.

Both T cells' subpopulations exhibit differences in expansion under immunological stimulation, such as viral infections. CD4+ T cell expansion is coupled with IFN-g-mediated apoptosis, resulting in minimal changes in their numbers after stimulation. Conversely, CD8+ T cells, lacking an apoptosis mechanism coupled with expansion, significantly increase after stimulation (Sckisel et al., 2017). The expansion of CD8+ T cells is predominantly driven by TCR recognition of the cognate epitope in an HLA class I molecule during low immunological stimulation. In high immunological stimulation, CD8+ T cells can proliferate in a bystander manner (TCR independent), characteristic of effector memory T cells (Tem). That is why the dominance of CD8+ T cells in the T cell pool in peripheral tissues serves as a marker of immunological stimulation.

CD69, a receptor belonging to the type II C-lectin family, functions as a classical early indicator of lymphocyte activation and tissue retention (Cibrián & Sánchez-Madrid, 2017). While it is commonly found in Trm lymphocytes (Kumar et al., 2017; Walsh et al., 2019), its presence in circulating T cells, typically undetectable during steady-state conditions, may increase in various pathological scenarios (Cibrián & Sánchez-Madrid, 2017). CD69 often coexists with other Trm markers such as integrins CD49a and CD103 in skin and mucosal tissues. The arrangement of Trm markers dictates distinct phenotypes. For example, CD49a plays a crucial role in the prolonged persistence of Trm cells, contributing to survival signals through interactions with collagen IV (Richter & Topham, 2007; Walsh et al., 2019). Likewise, CD49a expression correlates with increased T-cell mobility (Reilly et al., 2020). Viral infections' recurrence amplifies the proportion of CD49a+ cells in Trm lymphocytes. This aligns with the prominence of the CD49a+ subset in WAT T cells (especially in CD8+ cells from oWAT) (Chapman & Topham, 2010). Conversely, the expression of CD103 is negligible in WAT and does not seem relevant as it does in epithelial tissue (Mackay et al., 2012; Walsh et al., 2019). Our study identified three primary populations based on the expression patterns of CD69 and CD49a, each likely assuming distinct roles. These findings present an exciting avenue for research, prompting future studies to integrate new Trm markers and explore the functions of diverse T cell populations in adipose tissue depots.

The cytometric results provide valuable information, offering insights into the composition and characteristics of T cells within adipose tissue. Having explored the

cytometric landscape of T cells within adipose tissue, we now shift our focus to the intricate domain of TCR-antigen prediction, aiming to unravel antigenic specificities and interactions within this complex immunological landscape.

The optimal approach for TCR annotation in the study of TCR-antigen specificity, considering the current challenges in prediction algorithms (Hudson et al., 2023), involves cross-referencing annotations with established databases like VDJdb. At the time of our analysis, 34,685 TCR-antigen-HLA trios were featured in the VDJdb. We annotated a mere 0.42% of our TCRs. A major drawback of existing databases is their limited coverage of antigen-specificity information, addressing only a little fraction of the potential TCRs generated through somatic recombination. Additionally, there is a current overrepresentation of viral antigens in these databases, leaving numerous non-viral epitopes lacking experimentally validated cognate TCRs (Hudson et al., 2023). Additionally, we deliberately refrained from implementing strict HLA matching filters to exclude antigen specificity. The complexity of the HLA genetic system, characterized by hyper-polymorphism, multigenicity, and codominance (Meyer et al., 2018; Rao et al., 2011), adds intricacy to this decision. Furthermore, evidence suggests that a proinflammatory environment can alter the presentation of peptides by HLA class I and II molecules (Jurewicz & Stern, 2019; Prasad et al., 2016). In like manner, the interaction between TCR and the antigen-HLA complex exhibits a broad spectrum, ranging from high-affinity and high-specific interactions to low-affinity and promiscuous engagements.

In spite of the acknowledged limitations, we could successfully predict the presence of CMV-specific T lymphocytes infiltrating adipose tissue in our *in silico* analysis. This finding was subsequently validated through functional assays. However, future updates to the VDJdb would improve this analysis, offering a more comprehensive TCR annotation. By doing so, we would unravel the more complete antigen specificity landscape within adipose tissue-infiltrating T lymphocytes.

Functional analyses by ELISpot assays confirmed the presence of CMV-reactive T cells within WAT and unveiled distinctive features within oWAT. Firstly, the number of spots observed in oWAT significantly surpassed those in blood and sWAT, indicating a notable enrichment in CMV-reactive T cells. Secondly, although specific experiments confirming that T cells reactive in the ELISpot assay are Trm-like were not performed, a noteworthy observation emerges. Over 95% of oWAT T lymphocytes exhibit characteristics of memory T cells based on flow cytometry. Additionally, minimal contamination with peripheral blood in tissues results in very few naïve T cells appearing in the tissue. Thirdly, the spot size observed in oWAT was notably larger, a feature potentially influenced by cell phenotype and/or TCR binding affinity. Interestingly, a prior study in humans found substantial frequencies of T cells targeting CMV antigens in diverse samples (blood, bone marrow, and lymphoid nodes). In contrast, the presence was negligible in intestinal mucosa (Gordon et al., 2017). However, this study did not extend its analysis to adipose tissue samples.

Some discrepancies between *in silico* antigen-specificity predictions and functional ELISpot assays were observed in certain patients of the cohort. While *in silico*

analyses predicted the presence of TCRs reactive to CMV-epitopes only in blood but not in WAT (considering the 0.0001 frequency threshold filtering), ELISpot assays showed that all these cases tested positive in both blood and WAT. Several reasons, both methodological and biological, could contribute to this disparity. In the first place, sampling bias or quality controls and filters during data processing might have led to the loss of some TCR clonotypes. On a different note, the significant divergence in clonotype composition between blood and oWAT, coupled with the majority of VDJdb database annotations originating from experiments with PBMCs, may result in the omission of pertinent information.

In uncovering the intricate immunological landscape of adipose tissue, our bioinformatics assays not only identified virus-specific T cells but also highlighted the potential diversity in the TCR repertoire within adipose tissue. Subsequently, functional assays confirmed the presence of T cells specifically reactive to CMV, shedding light on the dynamic interactions occurring within this unique microenvironment. As we delve into the adaptive immune response in the context of viral infection, understanding the nuances of T cell activation and their implications for viral clearance becomes paramount.

The adaptive immune response to pathogens initiates with a swift expansion of clones responsive to their antigens. After a successful control of the infection, there is a recession of this T cell population to prevent detrimental effects from excessive activation. Later, three distinct scenarios may arise: 1) complete control and clearance of the pathogen resulting in a stable memory population, safeguarding against reinfection; 2) incomplete control leads to chronic infections (as observed in cases like HIV or hepatitis C virus), where the persistent presence of pathogen antigens continually challenges T cells, resulting in the accumulation of exhausted memory cells – a term undergoing reevaluation in the context of chronic inflammation by some authors (Speiser et al., 2014); 3) complete control but incomplete clearance of the pathogen (being undetectable) establishes a latent infection, these pathogens remain unnoticed by the immune system during the lysogenic phase but can be periodically activated during the lytic phase and reactivate the infection (Nikolich-Žugich, 2008).

In contrast to persistent viruses, latent viruses do not exhibit a constant presence of antigens, preventing the accumulation of exhausted T cells. However, the recurrent reactivation of the infection induces the growth of CD8+ Tem cells through a phenomenon known as memory inflation. Notably, this subset of memory T cells displays distinct behaviour compared to typical memory T cell populations. Under normal circumstances, Tem cells have a restricted expansion capacity and depend on the proliferation and differentiation of Tcm cells (long-lived classical memory). In latent infections, Tem cells significantly enhance their proliferation rate and sustain their pool independently, without the support of Tcm cells (Muschaweckh et al., 2016). This gives rise to a novel long-lived non-classical memory population within the Tem pool (Klenerman & Oxenius, 2016).

Intriguingly, latent infections not only impact the phenotype and dynamics of the memory T cell population but also alter the TCR repertoire. TCRs lack a mechanism for

affinity maturation, unlike antibodies, creating a challenge for generating a robust T cell response and memory population under normal circumstances. Viruses could exploit this limitation with minimal mutations to escape T cell response. Luckily, coreceptor molecules CD4 and CD8 address this issue by binding conserved regions in the HLA. Thus, enhancing the stability of the TCR-peptide-HLA complex. This enables TCRs with moderate affinity to contribute to the memory compartment, promoting diversity and preventing immune escape through antigen mutation (Turner et al., 2006). The efficiency of this mechanism is notable during the acute phase of infection, where high-affinity TCRs are saturated, allowing low-affinity TCRs to engage with the remaining antigens. However, in the latent phase with a smaller antigen load, high-affinity TCRs outcompete low-affinity ones, resulting in an oligoclonal response directed toward the most immunodominant antigens (such as CMV phosphoprotein pp65 and immediate-early protein 1 (IE1)) (Klenerman & Oxenius, 2016). This skewed response may gain significance with age, potentially facilitating reinfection with genetically different variants. This process is termed superinfection (Hansen et al., 2010). Importantly, this phenomenon poses challenges in developing vaccines against viruses like CMV or EBV (Berry et al., 2020), as these viruses employ various immunoevasins to escape immune responses.

In conclusion, our findings highlight a comprehensive immune response against latent CMV infection in adipose tissue, particularly oWAT, indicating potential responses to other viruses (including EBV and influenza) or pathogens. Such suggests a more robust immunological function in oWAT compared to sWAT. The presence of CMV-reactive T lymphocytes, inflation of the CD8+ population, and a skewed TCR repertoire against immunodominant antigens presented by HLA class I alleles strongly imply oWAT's significance as an immunological organ for antiviral responses in humans. While our study specifically examined oWAT and sWAT, it suggests the possibility for other adipose tissue types, even beyond the intraperitoneal cavity, to play crucial roles in immunological responses.

5.4. The existence of non-canonical TCRs

One curious finding in this research lies in the identification of non-canonical TCRs that presented two different TRVB segments fused. Albeit at low frequency, this discovery gains significant relevance by suggesting that these peculiar lymphocytes, present in blood and solid tissue, might have an active role and potentially be incorporated into the immunological memory. The coexistence of these TCRs in both environments raises the intriguing possibility that these lymphocytes have been activated through the recognition of HLA-antigen complexes. Moreover, this phenomenon may have fundamental implications for understanding the dynamics in the TCR rearrangement.

Non-canonical TCRs were observed in significantly higher numbers in peripheral blood mononuclear cells (PBMCs) compared to solid tissues. Chapter 2 findings demonstrated that TCRβ repertoires in blood exhibited greater diversity, and clonotypes

were less expanded compared to those in solid tissues. This disparity is attributed to the heightened presence of naïve T cells in the bloodstream relative to tissues.

The abundance of naïve T cells circulating in the blood indicates their continuous journey until encountering a selective interaction with an HLA-peptide complex. After several days without such an encounter, most naïve T cells undergo apoptosis. Consequently, it is anticipated that a substantial portion of blood TCRs will not be selected. Given the higher presence of "waste" in blood, there is a higher probability of identifying non-canonical rearrangements.

One might expect them to be eliminated from the system due to the low frequency of non-canonical TCRs. However, these rare rearrangements were also detected in solid tissues (WAT and liver), even at a diminished frequency. This presence in solid tissues indicates a prior antigen encounter, as these TCR variants were clonally expanded and preserved in the immunological memory. Interestingly, after PBMCs, liver tissue was the next sample exhibiting a higher number of non-canonical TCRs, due to the close relationship between blood and liver (Kalra et al., 2024).

Upon discovering these aberrant rearrangements, the initial concern naturally revolved around the possibility of them being artefacts arising from the library preparation or sequencing procedures. To address this, a meticulous analysis was undertaken to verify the existence of these atypical TCR variants. Subsequent experiments were designed to investigate whether these unconventional rearrangements could indeed generate a functional protein and assemble into a TCR-CD3 complex through *in vitro* assays. Remarkably, all conducted experiments consistently affirmed the genuine presence and functionality of these non-canonical TCRs.

Several key observations further emphasize the authenticity of these findings. Firstly, if the rare TCRs were mere artefacts originating from the experimental processes, a uniform distribution across samples or sequencing rounds would be anticipated. However, the data showed variation, reinforcing the idea that they are not artefacts. Secondly, the comparison of TCRβ repertoires in tissues revealed lower diversity and hyperexpanded clonotypes. In the context of the experimental procedures, particularly during PCRs in both library preparation and sequencing, the TCR-homogeneity of tissue samples could facilitate the hybridization of the more similar sequences. The contrasting diversity in tissue and blood repertoires adds weight to the argument against their artificial origin. Lastly, the inability to amplify these non-canonical rearrangements using alternative sets of primers is a critical piece of evidence. This lack of amplification strongly suggests a low or even nonexistent frequency of certain rearrangements. It is important to bear in mind that the risk of PCR artefacts increases in conditions of low target molecule frequency, where primers may non-specifically bind due to limited specific binding sites, resulting in the generation of artefacts.

This study has not investigated where or how these rearrangements arise. Nevertheless, other examples of aberrant TCR rearrangements have been reported in the past. To recap, TCRs are created after the somatic recombination of their gene segments.

For the β chain, these segments are TRVB, TRDB and TRJB. Somatic recombination initiates with the lymphoid-specific recognition of conserved recombination signal sequences (RSSs). These RSSs comprise a conserved block of seven nucleotides ("the heptamer") and flank the gene segments, contiguous with the coding sequence. Following the heptamer, there is a non-conserved region called "the spacer", of 12 or 23 nucleotides long and a second conserved block of nine nucleotides, the nonamer" (Bassing et al., 2002). The recombination process takes place following the 12/23 rule, between gene segments flanked by RSSs with different spacer lengths. Given that the DNA helix completes a full turn approximately every 10 ± 2 nucleotides (Levitt, 1978), the 12/23 rule elucidates the recombination process of TRVB-TRDB and TRDB-TRJB genes by correlating one and two turns of the DNA helix during VDJ recombination.

Nevertheless, in 1989 it was first confirmed that, despite the 12/23 rule, the direct joining of two D gene segments (D-D fusion) can occur in B cells. Remarkably, all D segments are flanked by RSSs of 12 nucleotides (Meek et al., 1989). Certain immunologists perceived V(DD)J recombination as an important mechanism behind the creation of longer CDR3s (Yu & Guan, 2014). Others downplayed its importance or even dismissed it as an artefact (Corbett et al., 1997; L. C. Watson et al., 2006). Notably, it has been later demonstrated that V(DD)J recombination is rather a crucial aspect of the immune response, leading to the formation of extended CDR3s that contribute to the development of substantial clonal lineages within antigen-stimulated repertoires (Safonova & Pevzner, 2020). D-D fusions are identified in approximately 5% of genes encoding antibodies (Larimore et al., 2012; Meek et al., 1989).

A similar phenomenon has been described in 2% of productive TCRβs in humans. However, the identification of the TRDB segment poses challenges due to its small size and nucleotide insertion/deletion at both ends during V(D)J recombination, resulting in nearly 20% of sequences remaining unidentified (Liu et al., 2014). Our results showed a lower frequency of non-canonical TCRs. It must be highlighted that mindful of potential imprecise identifications by BLAST, we excluded rearrangements within the same TRVB family. Considering our restriction in our analyses, it is worth considering the possibility that some rearrangements within the same TRVB family could represent aberrant TCRs formed by the fusion of two TRVB segments. Thus, increasing the initial percentage calculated.

On the contrary, TCRβ with TRVB direct to TRJB rearrangements and the absence of TRDB have been reported in humans. This phenomenon, observed in 0.7% of the productive unique sequences analysed, maintains a typical CDR3 size. Interestingly, they observed that nucleotide deletion was reduced in cases of direct TRVB-TRJB rearrangement compared to a canonical V(D)J recombination, compensating for the lack of the TRDB segment for a usual CDR3 size. Moreover, they demonstrated a different preferential use of TRVB segments when direct TRVB-TRJB rearrangements occurred (Ma et al., 2016).

Moving on to the TCRα chain formation, one peculiarity of its locus is that it has the *tcrδ* locus embedded between the TRAV and TRAJ gene segments. Typically, the TRD

genes are expected to exclusively encode TCRδs. However, certain TRVD gene segments can form junctions with TRJA genes, allowing them to participate in the synthesis of TCRα. Specific CD8+ T cell clones characterized by TRVD1+ TCRα have been identified for their ability to recognize and eliminate HIV-infected target cells (Ueno et al., 2003). On the other hand, other TRVD1+ CD8+ T cell clones are implicated in contributing to chronic neutropenia (Bank et al., 2003).

The identification of aberrant TCRs that deviate from the conventional V(D)J recombination patterns of a canonical TCR is a noteworthy observation. Our study specifically revealed the presence of TCRs in which the TRVB segment is formed through the fusion of two different segments. While this phenomenon manifests in a relatively limited number of productive TCRs, its prevalence is not significantly distant from the frequencies observed in other atypical rearrangements (Larimore et al., 2012; Liu et al., 2014; Ma et al., 2016; Meek et al., 1989). Unfortunately, none of the studies clarify the mechanisms that lead to the existence of these non-canonical TCRs, apart from the hypothesis of the DNA helix turns and structure.

The significance of these findings lies in the potential impact on our comprehension of diversity generation within the TCR and BCR repertoires. Although the occurrence of these unconventional TCR rearrangements may be relatively rare, their exploration holds promise for shedding light on novel aspects of immune response variability. Investigating these phenomena in future studies may uncover additional layers of complexity in the mechanisms governing the adaptive immune system. This, in turn, could contribute to a more comprehensive understanding of how immune cells generate diverse repertoires to effectively recognize and respond to a wide range of antigens.

5.5. Theoretical comparison between small RNA library preparation methodologies

During this thesis, there has not been enough time to compare the methodology created for the small RNA library preparation with any commercial kit. For that reason, we will carry out a theoretical comparison. However, we cannot prove that our technology has better or worse performance than any of the alternatives.

There are 9 main commercial technologies for miRNA sequencing. Each of those uses different library preparation methodologies, which can be divided into two groups: those that perform ligations reactions to incorporate the adapters' sequence (QIAGEN, New England Biolabs, Illumina, TriLink BioTechnologies, Bioo Scientific, Lexogen, and SeqMatic) and those that perform 5'RACE (Rapid Amplification of cDNA ends) (Takara Bio and Diagenode). Our technology does not resemble any of those. However, there are shared steps with some of the commercial technologies like polyadenylation of miRNAs or completion of the sequence of adapters via PCR.

Numerous investigations have provided evidence suggesting that the bias introduced during adapter ligation steps can be substantial. This phenomenon arises from the influence of RNA sequence and structure, leading to a preferential ligation of specific

small RNAs with a particular adapter (Fuchs et al., 2015; Hafner et al., 2011; Jayaprakash et al., 2011). Several approaches have been developed to try to mitigate this problem and improve the fidelity of NGS results.

The introduction of randomized nucleotides in adapter sequences near the ligation junction has been proposed. This strategy is followed by the NEXTflex kit from Bioo Scientific where 4 random nucleotides are present in the ligation junction sequences. Interestingly, one study revealed that randomized regions do not need to be adjacent to the ligation junction; instead, they can be internal. Nevertheless, they observed that certain miRNAs, despite the use of randomized adapters, exhibit a preference for sequences that facilitate the formation of particular structures. Thus, indicating that RNA structure also plays an important role in the ligation bias (Fuchs et al., 2015).

Another problem associated with the ligation reaction is the generation of waste products like adapter dimers. Companies have approached this problem from different angles. The most common is the elimination of excess 3' adapter before the 5' adapter ligation. These strategies may include purification steps (Bioo Scientific, Lexogen, and SeqMatic), the use of complementary sequences to block the 3' adapter (New England Biolabs) or the chemical inactivation of the adapter (Illumina and Bioo Scientific). Other approaches like the incorporation of chemical modifications that impede dimer formation between adapters are used (TriLink BioTechnologies). These diverse strategies provide options for minimizing the appearance of side products and improving the efficiency of small RNA library preparation.

Instead of altering the adapters themselves, certain research has focused on mitigating bias by optimizing reaction conditions. Some laboratories explored the use of the thermostable DNA/RNA ligase, 'MthRnl' from New England Biolabs, which works at 65 °C, a temperature at which it is assumed that the RNA does not have complex structures. However, results in reduction bias have been inconsistent with one study reporting a reduction (Z. Zhang et al., 2013) and others not observing any discernible effect (Jackson et al., 2014; Song et al., 2014). Still, some kits have chemically modified ligases that allegedly improve the reaction, such as the QIAseq kit by QIAGEN.

The latest studies also explored ligation at different temperatures (18 ºC, 25 ºC, or 37 ºC), obtaining better results at 25 ºC (Song et al., 2014). Notably, they also found that the addition of PEG enhanced ligation efficiency, and significantly diminished bias (Song et al., 2014; Z. Zhang et al., 2013). However, the optimal PEG concentration varied for different miRNAs tested (Song et al., 2014), indicating once again that miRNA composition plays an important role. However, 2 companies (New England Biolabs and Bioo Scientific) incorporate PEG in their ligation reactions.

For all these reasons, in our methodology, we decided to suppress any form of ligation. Thanks to the double trap created by the nucleotide tails and the use of magnetic beads, we achieve the elongation and incorporation of the adapters' sequences into the small RNA population.

In like manner, some companies opt to avoid the adapter ligation completely. This is the case of Takara Bio and Diagenode. Both procedures rely on the polyadenylation for the incorporation of the 3' adapter and the addition of the 5' adapter through reverse transcriptase template-switching (5' RACE). The difference between them is that the second (D-Plex by Diagenode) includes UMI within its 5' adapter. However, poly(A) polymerases (PAPs) are also affected by RNA structures, particularly when derived from *E. coli* (Yehudai-Resheff & Schuster, 2000). Thus, pre-heating the sample for RNA denaturation is crucial. During this study, PAPs derived from *E. coli* and yeast were tested. The latter exhibited better results as the number of adenines added remained almost constant, with a dispersion of 1.82. Furthermore, RNA 3' modification like the 2'-O-methylation impacts the tailing efficiency (Munafó & Robb, 2010). This modification is present in plants' miRNAs and mammals' piRNAs, among others (Ghildiyal & Zamore, 2009). Therefore, the decrease in tailing efficiency and the resulting alteration in expression levels could be of particular significance when analysing these types of samples.

Other important aspects discussed in section 5.2 are shared with small RNA library preparation protocols. Intermediate purification steps are carried out in several procedures either between 3' and 5' adapters' ligations (Bioo Scientific, Lexogen, and SeqMatic) or after the reverse transcription (QIAGEN). In every purification step, there is a slight loss of the initial material. This problem can be exacerbated the lesser the material to be purified, as in the case of small RNA ligation, where the starting material is already scarce, and no enrichment has taken place. In our methodology, we leverage the magnetic properties of the microparticles to facilitate washings between various steps, mimicking a purification process. Likewise, 5'RACE is another approach that requires high-quality starting material (Woodsworth et al., 2013). Although outside the scope of this thesis, it is worth mentioning that the extraction of small RNA is also complex. The available kits developed for this purpose differ in performance. The dilemma arises in the choice between kits that yield very high-quality yet low-quantity material and those providing higher quantities of material but with higher contamination of other molecules.

In terms of total RNA or small RNA input, all commercial kits differed greatly in their recommendations. According to the data provided by the companies, the minimum amount of total RNA required is 50 pg in the Lexogen kit, while the minimum for small RNA is 10 pg in the Diagenode kit. Notably, 5 out of the 9 kits do not specify the small RNA amount required, just the total RNA. Diverse studies have tested specifically low-input protocols (Heinicke, Zhong, Zucknick, Breidenbach, Sundaram, T. Flåm, Leithaug, Dalland, Farmer, et al., 2020; Herbert et al., 2020). The lowest amount used was 10 ng in the TriLink Biotechnologies and QIAGEN procedures (Herbert et al., 2020). Remarkably, with our methodology, we are capable of initiating the protocol with 0.7 pg.

Additionally, the introduction of errors and biases can appear during the PCR process. RNA GC content impacts the amplification efficiency, being GC-rich molecules less readily amplified and remaining underrepresented or even unidentified in the end (Raabe et al., 2014). Some studies have evaluated PCR buffers composition to identify

which ones perform better and reduce final distortion (Aird et al., 2011; Dabney & Meyer, 2012). For that reason, using DNA polymerases with enhanced activity on GC-rich regions must be considered. On the other hand, including identifiers or barcodes during the amplification process eliminates the biases of introducing them by ligation (Alon et al., 2011). Notably, all commercial kits including ligation do not incorporate identifiers or barcodes in the reaction.

Lastly, it is worth highlighting that only two commercially available kits include UMIs in their technology (QIAGEN and Diagenode). In the same way, as in the methodology for the TCRβ, we do not incorporate UMIs, although it could be easily achieved in the future. In contrast to the TCR library preparation kits, the use of UMIs has not been studied and compared among the miRNA library preparation kits.

Once more, the conclusion derived from all the information gathered is that each methodology will have its own flaws depending on the approach carried out. Likewise, each procedure includes steps that are susceptible to introducing errors and/or biases. Therefore, to achieve results that can be effectively compared and to assess relative changes in the final quantification of small RNA variants, researchers must carry out experiments and analyses under precisely identical conditions. Additionally, it is important to highlight that variation not only applies to inter-sample comparison but also intra-sample comparison. Furthermore, researchers need to be aware of the most recent developments in small RNA sequencing methodologies as continuous advancements are occuring.

5.6. Comparison between DTT and DTT-CC methodologies

During this study, we developed the DTT methodology which was patented and is currently protected in Europe and the USA. However, the technology has continued to evolve to improve its efficiency, resulting in the DTT-CC methodology. Several aspects must be highlighted to compare both technologies.

The most notable improvement is the reduction of the detection threshold. The detection threshold is defined as the minimum amount of small RNA material required for capture using miRNA-beads in the original technology or the 5'-azide-modified oligonucleotides in the evolved version. This threshold was reduced 1000 times from the DTT methodology to the DTT-CC. We believe this enhancement in sensitivity comes from the performance of the first steps in solution. The environment for small RNA trap is most effective when all components of the system are in solution. However, the fact that the oligonucleotide that traps the miRNA population is bound to a particle from the beginning of the process may reduce the interaction with the molecules.

Regarding working days, they were reduced from 3 to 2 days. Likewise, the total time spent on laboratory work was also shortened by 70 min. It should be highlighted that laboratory work is calculated based on the reaction time, not just the handling time. While this might seem like a modest time difference, we must highlight that overnight incubations and washing steps are not included in this estimation. The original DTT

methodology includes one overnight incubation, extending working days to 3. It is worth noting that both methodologies can be longer than 3 or 2 days, respectively, as many "safe-stop" points can be chosen based on individual preferences. On the other hand, washing steps are reduced from 5 to 3. The impact of washing time becomes more pronounced with an increased number of samples, as each washing step for a single sample takes 2-3 minutes. Thus, differences in laboratory work time are greater than the initial 70-minute reduction. Additionally, washing steps may be detrimental to the beads, which further improves the methodology.

Another important enhancement is the ability to remove miRNA-free oligonucleotides from the reaction. Previously, we attempted to block those oligonucleotides to prevent them from continuing in the process and potentially masking completed libraries. Different blocking methods were tested, with dideoxy-blocking being more satisfactory than hairpin-blocking. Nevertheless, even this blocking method was not 100% efficient. In the evolved version of DTT, the elimination of miRNA-free oligonucleotides is achieved thanks to the first reactions taking place in solution. Oligonucleotides are free in solution since the beginning of the process. Therefore, capturing this oligo with another bead via streptavidin-biotin interaction to get rid of the excess is feasible.

Unfortunately, we have not been able to conclude this methodology by performing a massive sequencing with real samples to analyse its efficiency. Despite this limitation, we believe that our methodology holds the potential to enhance current methodologies, also considering that there might be potential associated biases or sources of error.

Recently, we have identified alternative applications for this methodology beyond the discovery of new miRNA variants. A notable challenge in miRNA analysis via microarray lies in its low sensitivity. Our methodology addresses this limitation by the enrichment of the entire small RNA population in a sample. This may be achieved through the addition of universal sequences, similar to NGS adapters but of a smaller size, to later amplify the small RNA pool using these sequences as binding sites for primers.

Moreover, the adaptability of our methodology extends to the enrichment or sequencing of unidentified sequences present in highly degraded samples, such as ancient DNA. This versatile capability broadens the scope of potential applications, demonstrating the methodology's resilience in diverse scenarios. As we continue to explore and refine this approach, it opens avenues for advancements not only in small RNA research but also in the analysis of degraded or challenging samples across various biological contexts.

6. *Limitations and future perspectives*

Regarding limitations and future perspectives related to Chapter I, we must highlight the fact that we did not compare our methodology to prepare TCRβ libraries with any of the commercially available alternatives (including long-read sequencing). This aspect represents a significant gap that warrants further exploration to comprehensively assess the efficiency and potential errors associated with our technology. A comparative analysis with existing commercial methods would contribute valuable insights into the strengths and weaknesses of our approach.

In like manner, it would be interesting to include UMIs at the beginning of the process. This enhancement becomes particularly relevant when the study's objective is the quantitative analysis of TCR repertoires. UMIs could provide a valuable tool for more accurate quantification and enable a deeper understanding of the intricacies within the TCR repertoire.

Moving on to Chapter II, as we have addressed several times, VDJdb is very scarce at the moment. As the repository accumulates more comprehensive information on antigen-TCR interactions, particularly regarding non-viral epitopes, future re-analyses will yield a more accurate representation of the TCR repertoire within WAT-infiltrating lymphocytes. The evolution of VDJdb remains a key factor in unlocking deeper insights into the immune landscape.

In an ideal scenario where liver biopsy is less invasive and dangerous, the collection of more liver samples could broaden the scope of analyses beyond the possibilities explored in this study. Techniques such as flow cytometry and ELISpot assays, currently constrained to blood and WAT samples, could be more extensively employed to enrich the TCR repertoire knowledge. Furthermore, in this ideal scenario, obtaining samples of other types of fat depots like the adipose tissue covering the heart will be fascinating. In that way, we could confirm if WAT plays a crucial role in protecting such important organs.

Additionally, the feasibility of ELISpot assays for other viruses like EBV or influenza would not only corroborate bioinformatic analyses but also contribute to a better understanding of the immune responses. Moreover, the potential for further experiments such as single-cell TCR-seq becomes interesting. The implementation of single-cell TCR-seq could provide a detailed examination of individual TCRs, unravelling intricate patterns within the immune system.

Concerning Chapter III, it would be interesting to carry out a more in-depth exploration of other types of non-canonical TCRβs. While our study has unveiled the presence of TCRs where two TRVB gene segments are joined, it is noteworthy that literature reported the presence of TCRs with two TRDB segments. This observation implies the likelihood of encountering novel forms of rearrangements that could further diversify our understanding of atypical TCRβs.

Furthermore, we had originally intended to incorporate confocal microscopy to verify CD3-TCR assembly. Unfortunately, this plan encountered a handicap during electroporation, where associated cell death resulted in substantial debris with

autofluorescence, making difficult the ability to distinguish positive signals. An alternative approach, involving the infection of cells with lentiviruses, could potentially overcome this issue. By doing so, we aim to determine if this method allows for the identification of cells that have successfully incorporated and are expressing the construct. This adaptation provides a potential solution to the challenges encountered in the initial experimental setup.

Moreover, our aspirations extend to designing new constructs that mimic aberrant TCRs. This addition to our experimental repertoire would enable a more comprehensive exploration of atypical TCR rearrangements.

Finally, turning our attention to Chapter IV, diverse aspects must be further studied beyond the fine-tuning of the DTT and DTT-CC methodologies presented in this thesis Firstly, a massive sequencing must be performed to address the methodology's efficiency. This will facilitate an in-depth analysis of its performance, shedding light on which small RNAs are captured and the potential presence of side products. Subsequently, with such insights we could analyse whether the technique could be improved in any further aspect.

Secondly, to determine the comparative advantages of our technology, it is imperative to conduct a thorough comparison with alternative commercial kits. This comparative analysis is crucial in determining whether our methodology effectively mitigates errors or biases inherent in other existing technologies.

Additionally, the incorporation of UMIs into the methodology must be considered, especially when conducting quantification studies. This enhancement has the potential to improve the precision of quantification, providing a valuable tool for studies that demand accurate and nuanced measurement. For example, to study miRNA expression variations in a particular pathological context.

As for the beads, as we are functionalising them ourselves to add the propargylamine so that they can be used in a click chemistry reaction, we must include quality controls for each batch. Continuous efforts in the laboratory are dedicated to refining and expanding these quality controls. Despite the initiation of this process, there is still the need to design experiments that enable a comprehensive characterization of the particles, ensuring consistency and reliability in their performance.

7. Conclusions

1. We successfully designed and developed a novel methodology to prepare TCRβ libraries for massive sequencing. This innovative approach allows the identification of the complete TCRβ chain, including the three complementary determining regions. The technology strategically employs magnetic beads to avoid the ligation of adapters' sequences.

2. Our findings shed light on the immune response of white adipose tissue (WAT) against latent infections, particularly in omental WAT (oWAT). Although this study focuses on CMV, it hints at potential responses to other viruses such as Epstein-Barr virus and influenza or pathogens.

3. Results suggest a pivotal role of oWAT as an immunological organ for antiviral response in humans. oWAT exhibited an increased presence of CMV-reactive T cells and CD8+ population, as well as a TCR repertoire biased against immunodominant antigens presented by HLA class I alleles.

4. The predominant T cell subpopulation in white adipose tissue is the effector memory subset, in both CD4+ and CD8+ T cells. When considering the expression of CD69 and CD49a, the double-negative subpopulation prevails in blood, whereas in oWAT is the double-positive. Notably, no predominant subset is observed in sWAT. In contrast, CD103 expression was absent in T cells from blood or WAT.

5. The discovery of non-canonical TCRβs, formed by the fusion of two TRVB segments present in both blood and tissue, suggests the potential emergence of a previously unrecognized level of TCR diversity generation.

6. We successfully designed and developed an innovative methodology for the preparation of small RNA libraries for next-generation sequencing. This nanotechnology-based approach eliminates the principal source of bias present in most of the commercial methods, which is the ligation bias.

Bibliography

Aalto, A. P., & Pasquinelli, A. E. (2012). Small Non-coding RNAs Mount a Silent Revolution in Gene Expression. *Current Opinion in Cell Biology*, *24*(3), 333–340. https://doi.org/10.1016/j.ceb.2012.03.006

Aird, D., Ross, M. G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D. B., Nusbaum, C., & Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, *12*(2), R18. https://doi.org/10.1186/gb-2011-12-2-r18

Alles, J., Fehlmann, T., Fischer, U., Backes, C., Galata, V., Minet, M., Hart, M., Abu-Halima, M., Grässer, F. A., Lenhof, H.-P., Keller, A., & Meese, E. (2019). An estimate of the total number of true human miRNAs. *Nucleic Acids Research*, *47*(7), 3353–3364. https://doi.org/10.1093/nar/gkz097

Alon, S., Vigneault, F., Eminaga, S., Christodoulou, D. C., Seidman, J. G., Church, G. M., & Eisenberg, E. (2011). Barcoding bias in high-throughput multiplex sequencing of miRNA. *Genome Research*, *21*(9), 1506–1511. https://doi.org/10.1101/gr.121715.111

Antoniou, A., Baptista, M., Carney, N., & Hanley, J. G. (2014). PICK1 links Argonaute 2 to endosomes in neuronal dendrites and regulates miRNA activity. *EMBO Reports*, *15*(5), 548–556. https://doi.org/10.1002/embr.201337631

ATDBio—Next generation sequencing. (n.d.). Retrieved 1 February 2024, from https://atdbio.com/nucleic-acids-book/Next-generation-sequencing#figure-emulsion-pcr

Attaf, M., Huseby, E., & Sewell, A. K. (2015). Aβ T cell receptors as predictors of health and disease. *Cellular & Molecular Immunology*, *12*(4), 391–399. https://doi.org/10.1038/cmi.2014.134

Attaf, M., Legut, M., Cole, D. K., & Sewell, A. K. (2015). The T cell antigen receptor: The Swiss army knife of the immune system. *Clinical & Experimental Immunology*, *181*(1), 1–18. https://doi.org/10.1111/cei.12622

Bagchi, D. P., & MacDougald, O. A. (2019). Identification and Dissection of Diverse Mouse Adipose Depots. *Journal of Visualized Experiments : JoVE*, *149*, 10.3791/59499. https://doi.org/10.3791/59499

Baid, G., Cook, D. E., Shafin, K., Yun, T., Llinares-López, F., Berthet, Q., Belyaeva, A., Töpfer, A., Wenger, A. M., Rowell, W. J., Yang, H., Kolesnikov, A., Ammar, W., Vert, J.-P., Vaswani, A., McLean, C. Y., Nattestad, M., Chang, P.-C., & Carroll, A. (2023). DeepConsensus improves the accuracy of sequences with a gap-aware sequence transformer. *Nature Biotechnology*, *41*(2), Article 2. https://doi.org/10.1038/s41587-022-01435-7

Bank, I., Cohen, L., Kneller, A., De Rosbo, N. K., Book, M., & Ben-Nun, A. (2003). Aberrant T-Cell Receptor Signalling of Interferon-γ- and Tumour Necrosis Factor-α-Producing Cytotoxic CD8+ Vδ1/Vβ16 T Cells in a Patient with Chronic Neutropenia. *Scandinavian Journal of Immunology*, *58*(1), 89–98. https://doi.org/10.1046/j.1365-3083.2003.01272.x

Baran-Gale, J., Kurtz, C. L., Erdos, M. R., Sison, C., Young, A., Fannin, E. E., Chines, P. S., & Sethupathy, P. (2015). Addressing Bias in Small RNA Library Preparation for Sequencing: A New Protocol Recovers MicroRNAs that Evade Capture by Current Methods. *Frontiers in Genetics*, *6*. https://www.frontiersin.org/articles/10.3389/fgene.2015.00352

Barennes, P., Quiniou, V., Shugay, M., Egorov, E. S., Davydov, A. N., Chudakov, D. M., Uddin, I., Ismail, M., Oakes, T., Chain, B., Eugster, A., Kashofer, K., Rainer, P. P., Darko, S., Ransier, A., Douek, D. C., Klatzmann, D., & Mariotti-Ferrandiz, E. (2021). Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. *Nature Biotechnology*, *39*(2), Article 2. https://doi.org/10.1038/s41587-020-0656-3

Barman, B., & Bhattacharyya, S. N. (2015). mRNA Targeting to Endoplasmic Reticulum Precedes Ago Protein Interaction and MicroRNA (miRNA)-mediated Translation Repression in Mammalian Cells. *The Journal of Biological Chemistry*, *290*(41), 24650–24656. https://doi.org/10.1074/jbc.C115.661868

Bassing, C. H., Swat, W., & Alt, F. W. (2002). The Mechanism and Regulation of Chromosomal V(D)J Recombination. *Cell*, *109*(2), S45–S55. https://doi.org/10.1016/S0092-8674(02)00675-X

Bénézech, C., Luu, N.-T., Walker, J. A., Kruglov, A. A., Loo, Y., Nakamura, K., Zhang, Y., Nayar, S., Jones, L. H., Flores-Langarica, A., McIntosh, A., Marshall, J., Barone, F., Besra, G., Miles, K., Allen, J. E., Gray, M., Kollias, G., Cunningham, A. F., … Caamaño, J. H. (2015). Inflammation-induced formation of fat-associated lymphoid clusters. *Nature Immunology*, *16*(8), 819–828. https://doi.org/10.1038/ni.3215

Berry, R., Watson, G. M., Jonjic, S., Degli-Esposti, M. A., & Rossjohn, J. (2020). Modulation of innate and adaptive immunity by cytomegaloviruses. *Nature Reviews Immunology*, *20*(2), Article 2. https://doi.org/10.1038/s41577-019-0225-5

Bilotti, K., Potapov, V., Pryor, J. M., Duckworth, A. T., Keck, J. L., & Lohman, G. J. S. (2022). Mismatch discrimination and sequence bias during end-joining by DNA ligases. *Nucleic Acids Research*, *50*(8), 4647–4658. https://doi.org/10.1093/nar/gkac241

Bracken, C. P., Scott, H. S., & Goodall, G. J. (2016). A network-biology perspective of microRNA function and dysfunction in cancer. *Nature Reviews Genetics*, *17*(12), Article 12. https://doi.org/10.1038/nrg.2016.134

Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D., & Craig, D. W. (2016). Translating RNA sequencing into clinical diagnostics: Opportunities and challenges. *Nature Reviews. Genetics*, *17*(5), 257–271. https://doi.org/10.1038/nrg.2016.10

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, *136*(4), 642–655. https://doi.org/10.1016/j.cell.2009.01.035

Chapman, T. J., & Topham, D. J. (2010). Identification of a Unique Population of Tissue-Memory CD4+ T cells in the Airways after Influenza Infection that is Dependent on the Integrin VLA-1. *Journal of Immunology (Baltimore, Md. : 1950)*, *184*(7), 3841–3849. https://doi.org/10.4049/jimmunol.0902281

Chen, D., & Patton, J. T. (2001). Reverse Transcriptase Adds Nontemplated Nucleotides to cDNAs During 5′-RACE and Primer Extension. *BioTechniques*, *30*(3), 574–582. https://doi.org/10.2144/01303rr02

Cheng, M., Yang, J., Zhao, X., Zhang, E., Zeng, Q., Yu, Y., Yang, L., Wu, B., Yi, G., Mao, X., Huang, K., Dong, N., Xie, M., Limdi, N. A., Prabhu, S. D., Zhang, J., & Qin, G. (2019). Circulating myocardial microRNAs from infarcted hearts are carried in exosomes and mobilise bone marrow progenitor cells. *Nature Communications*, *10*(1), 959. https://doi.org/10.1038/s41467-019-08895-7

Cibrián, D., & Sánchez-Madrid, F. (2017). CD69: From activation marker to metabolic gatekeeper. *European Journal of Immunology*, *47*(6), 946–953. https://doi.org/10.1002/eji.201646837

Coenen-Stass, A. M. L., Magen, I., Brooks, T., Ben-Dov, I. Z., Greensmith, L., Hornstein, E., & Fratta, P. (2018). Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biology*, *15*(8), 1133–1145. https://doi.org/10.1080/15476286.2018.1514236

Contreras, N. A., Sitnik, K. M., Jeftic, I., Coplen, C. P., Čičin-Šain, L., & Nikolich-Žugich, J. (2019). Life-long control of cytomegalovirus (CMV) by T resident memory cells in the adipose tissue

results in inflammation and hyperglycemia. *PLOS Pathogens*, *15*(6), e1007890. https://doi.org/10.1371/journal.ppat.1007890

Corbett, S. J., Tomlinson, I. M., Sonnhammer, E. L., Buck, D., & Winter, G. (1997). Sequence of the human immunoglobulin diversity (D) segment locus: A systematic analysis provides no evidence for the use of DIR segments, inverted D segments, 'minor' D segments or D-D recombination. *Journal of Molecular Biology*, *270*(4), 587–597. https://doi.org/10.1006/jmbi.1997.1141

Culshaw, A., Ladell, K., Gras, S., McLaren, J. E., Miners, K. L., Farenc, C., van den Heuvel, H., Gostick, E., Dejnirattisai, W., Wangteeraprasert, A., Duangchinda, T., Chotiyarnwong, P., Limpitikul, W., Vasanawathana, S., Malasit, P., Dong, T., Rossjohn, J., Mongkolsapaya, J., Price, D. A., & Screaton, G. R. (2017). Germline bias dictates cross-serotype reactivity in a common dengue-virus-specific CD8+ T cell response. *Nature Immunology*, *18*(11), Article 11. https://doi.org/10.1038/ni.3850

Dabney, J., & Meyer, M. (2012). Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques*, *52*(2), 87–94. https://doi.org/10.2144/000113809

Dard-Dascot, C., Naquin, D., d'Aubenton-Carafa, Y., Alix, K., Thermes, C., & Dijk, E. van. (2018). Systematic comparison of small RNA library preparation protocols for next-generation sequencing. *BMC Genomics*, *19*. https://doi.org/10.1186/s12864-018-4491-6

Delves, P. J., Martin, S. J., Burton, D. R., & Roitt, I. M. (2017). *Roitt's essential immunology* (Thirteenth edition). Wiley Blackwell.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., deWinter, A., Dixon, J., … Turner, S. (2009). Real-Time DNA Sequencing from Single Polymerase Molecules. *Science*, *323*(5910), 133–138. https://doi.org/10.1126/science.1162986

Elliott, J. I., & Altmann, D. M. (1995). Dual T cell receptor alpha chain T cells in autoimmunity. *The Journal of Experimental Medicine*, *182*(4), 953–959. https://doi.org/10.1084/jem.182.4.953

Farber, D. L., Yudanin, N. A., & Restifo, N. P. (2014). Human memory T cells: Generation, compartmentalization and homeostasis. *Nature Reviews. Immunology*, *14*(1), 24–35. https://doi.org/10.1038/nri3567

Fransquet, P. D., & Ryan, J. (2018). Micro RNA as a potential blood-based epigenetic biomarker for Alzheimer's disease. *Clinical Biochemistry*, *58*, 5–14. https://doi.org/10.1016/j.clinbiochem.2018.05.020

Fuchs, R. T., Sun, Z., Zhuang, F., & Robb, G. B. (2015). Bias in Ligation-Based Small RNA Sequencing Library Construction Is Determined by Adaptor and RNA Structure. *PLoS ONE*, *10*(5), e0126049. https://doi.org/10.1371/journal.pone.0126049

Garcia-Elias, A., Alloza, L., Puigdecanet, E., Nonell, L., Tajes, M., Curado, J., Enjuanes, C., Díaz, O., Bruguera, J., Martí-Almor, J., Comín-Colet, J., & Benito, B. (2017). Defining quantification methods and optimizing protocols for microarray hybridization of circulating microRNAs. *Scientific Reports*, *7*, 7725. https://doi.org/10.1038/s41598-017-08134-3

García-Rubio, J., León, J., Redruello-Romero, A., Pavón, E., Cozar, A., Tamayo, F., Caba-Molina, M., Salmerón, J., & Carazo, Á. (2018). Cytometric analysis of adipose tissue reveals increments of adipocyte progenitor cells after weight loss induced by bariatric surgery. *Scientific Reports*, *8*(1), 15203. https://doi.org/10.1038/s41598-018-33488-7

Gebert, L. F. R., & MacRae, I. J. (2019). Regulation of microRNA function in animals. *Nature Reviews. Molecular Cell Biology*, *20*(1), 21–37. https://doi.org/10.1038/s41580-018-0045-7

Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: An expanding universe. *Nature Reviews. Genetics*, *10*(2), 94–108. https://doi.org/10.1038/nrg2504

Giudicelli, V. (2006). IMGT/LIGM-DB, the IMGT(R) comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Research*, *34*(90001), D781–D784. https://doi.org/10.1093/nar/gkj088

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews. Genetics*, *17*(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Gordon, C. L., Miron, M., Thome, J. J. C., Matsuoka, N., Weiner, J., Rak, M. A., Igarashi, S., Granot, T., Lerner, H., Goodrum, F., & Farber, D. L. (2017). Tissue reservoirs of antiviral T cell immunity in persistent human CMV infection. *Journal of Experimental Medicine*, *214*(3), 651–667. https://doi.org/10.1084/jem.20160758

Guay, C., & Regazzi, R. (2013). Circulating microRNAs as novel biomarkers for diabetes mellitus. *Nature Reviews Endocrinology*, *9*(9), Article 9. https://doi.org/10.1038/nrendo.2013.86

Guerrini, V., & Gennaro, M. L. (2019). Foam cells: One size doesn't fit all. *Trends in Immunology*, *40*(12), 1163–1179. https://doi.org/10.1016/j.it.2019.10.002

Hafner, M., Renwick, N., Brown, M., Mihailović, A., Holoch, D., Lin, C., Pena, J. T. G., Nusbaum, J. D., Morozov, P., Ludwig, J., Ojo, T., Luo, S., Schroth, G., & Tuschl, T. (2011). RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA*, *17*(9), 1697–1712. https://doi.org/10.1261/rna.2799511

Han, S.-J., Glatman Zaretsky, A., Andrade-Oliveira, V., Collins, N., Dzutsev, A., Shaik, J., Morais da Fonseca, D., Harrison, O. J., Tamoutounour, S., Byrd, A. L., Smelkinson, M., Bouladoux, N., Bliska, J. B., Brenchley, J. M., Brodsky, I. E., & Belkaid, Y. (2017). White Adipose Tissue Is a Reservoir for Memory T Cells and Promotes Protective Memory Responses to Infection. *Immunity*, *47*(6), 1154-1168.e6. https://doi.org/10.1016/j.immuni.2017.11.009

Hansen, S. G., Powers, C. J., Richards, R., Ventura, A. B., Ford, J. C., Siess, D., Axthelm, M. K., Nelson, J. A., Jarvis, M. A., Picker, L. J., & Früh, K. (2010). Evasion of CD8+ T cells is critical for super-infection by cytomegalovirus. *Science (New York, N.Y.)*, *328*(5974), 102–106. https://doi.org/10.1126/science.1185350

He, B., Zhao, Z., Cai, Q., Zhang, Y., Zhang, P., Shi, S., Xie, H., Peng, X., Yin, W., Tao, Y., & Wang, X. (2020). miRNA-based biomarkers, therapies, and resistance in Cancer. *International Journal of Biological Sciences*, *16*(14), 2628–2647. https://doi.org/10.7150/ijbs.47203

Heinicke, F., Zhong, X., Zucknick, M., Breidenbach, J., Sundaram, A. Y. M., T. Flåm, S., Leithaug, M., Dalland, M., Farmer, A., Henderson, J. M., Hussong, M. A., Moll, P., Nguyen, L., McNulty, A., Shaffer, J. M., Shore, S., Yip, H. K., Vitkovska, J., Rayner, S., … Gilfillan, G. D. (2020). Systematic assessment of commercially available low-input miRNA library preparation kits. *RNA Biology*, *17*(1), 75–86. https://doi.org/10.1080/15476286.2019.1667741

Heinicke, F., Zhong, X., Zucknick, M., Breidenbach, J., Sundaram, A. Y. M., T. Flåm, S., Leithaug, M., Dalland, M., Rayner, S., Lie, B. A., & Gilfillan, G. D. (2020). An extension to: Systematic assessment of commercially available low-input miRNA library preparation kits. *RNA Biology*, *17*(9), 1284–1292. https://doi.org/10.1080/15476286.2020.1761081

Herbert, Z. T., Thimmapuram, J., Xie, S., Kershner, J. P., Kolling, F. W., Ringelberg, C. S., LeClerc, A., Alekseyev, Y. O., Fan, J., Podnar, J. W., Stevenson, H. S., Sommerville, G., Gupta, S., Berkeley, M., Koeman, J., Perera, A., Scott, A. R., Grenier, J. K., Malik, J., … Levine, S. S. (2020). Multisite Evaluation of Next-Generation Methods for Small RNA Quantification. *Journal of Biomolecular Techniques : JBT*, *31*(2), 47. https://doi.org/10.7171/jbt.20-3102-001

Hirschberger, S., Hinske, L. C., & Kreth, S. (2018). MiRNAs: Dynamic regulators of immune cell functions in inflammation and cancer. *Cancer Letters*, *431*, 11–21. https://doi.org/10.1016/j.canlet.2018.05.020

Ho, P. T. B., Clark, I. M., & Le, L. T. T. (2022). MicroRNA-Based Diagnosis and Therapy. *International Journal of Molecular Sciences*, *23*(13), Article 13. https://doi.org/10.3390/ijms23137167

Hudson, D., Fernandes, R. A., Basham, M., Ogg, G., & Koohy, H. (2023). Can we predict T cell specificity with digital biology and machine learning? *Nature Reviews Immunology*, 1–11. https://doi.org/10.1038/s41577-023-00835-3

Iraci, N., Leonardi, T., Gessler, F., Vega, B., & Pluchino, S. (2016). Focus on Extracellular Vesicles: Physiological Role and Signalling Properties of Extracellular Membrane Vesicles. *International Journal of Molecular Sciences*, *17*(2), 171. https://doi.org/10.3390/ijms17020171

Jackson, T. J., Spriggs, R. V., Burgoyne, N. J., Jones, C., & Willis, A. E. (2014). Evaluating bias-reducing protocols for RNA sequencing library preparation. *BMC Genomics*, *15*(1), 569. https://doi.org/10.1186/1471-2164-15-569

Jaitin, D. A., Adlung, L., Thaiss, C. A., Weiner, A., Li, B., Descamps, H., Lundgren, P., Bleriot, C., Liu, Z., Deczkowska, A., Keren-Shaul, H., David, E., Zmora, N., Shibolet, O., Hill, D. A., Colonna, M., Lazar, M. A., Ginhoux, F., Shapiro, H., … Amit, I. (2019). Lipid-associated macrophages control metabolic homeostasis in a Trem2-dependent manner. *Cell*, *178*(3), 686-698.e14. https://doi.org/10.1016/j.cell.2019.05.054

Jayaprakash, A. D., Jabado, O., Brown, B. D., & Sachidanandam, R. (2011). Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Research*, *39*(21), e141. https://doi.org/10.1093/nar/gkr693

Jurewicz, M. M., & Stern, L. J. (2019). Class II MHC Antigen Processing in Immune Tolerance and Inflammation. *Immunogenetics*, *71*(3), 171–187. https://doi.org/10.1007/s00251-018-1095-x

Kalla, R., Ventham, N. T., Kennedy, N. A., Quintana, J. F., Nimmo, E. R., Buck, A. H., & Satsangi, J. (2015). MicroRNAs: New players in IBD. *Gut*, *64*(3), 504–513. https://doi.org/10.1136/gutjnl-2014-307891

Kalra, A., Yetiskul, E., Wehrle, C. J., & Tuma, F. (2024). Physiology, Liver. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK535438/

Kamerkar, S., LeBleu, V. S., Sugimoto, H., Yang, S., Ruivo, C. F., Melo, S. A., Lee, J. J., & Kalluri, R. (2017). Exosomes Facilitate Therapeutic Targeting of Oncogenic Kras in Pancreatic Cancer. *Nature*, *546*(7659), 498–503. https://doi.org/10.1038/nature22341

Kim, V. N., Han, J., & Siomi, M. C. (2009). Biogenesis of small RNAs in animals. *Nature Reviews Molecular Cell Biology*, *10*(2), Article 2. https://doi.org/10.1038/nrm2632

Klenerman, P., & Oxenius, A. (2016). T cell responses to cytomegalovirus. *Nature Reviews Immunology*, *16*(6), Article 6. https://doi.org/10.1038/nri.2016.38

Koch, S., Larbi, A., Derhovanessian, E., Özcelik, D., Naumova, E., & Pawelec, G. (2008). Multiparameter flow cytometric analysis of CD4 and CD8 T cell subsets in young and old people. *Immunity & Ageing : I & A*, *5*, 6. https://doi.org/10.1186/1742-4933-5-6

Kumar, B. V., Ma, W., Miron, M., Granot, T., Guyer, R. S., Carpenter, D. J., Senda, T., Sun, X., Ho, S.-H., Lerner, H., Friedman, A. L., Shen, Y., & Farber, D. L. (2017). Human tissue-resident memory T cells are defined by core transcriptional and functional signatures in lymphoid and mucosal sites. *Cell Reports*, *20*(12), 2921–2934. https://doi.org/10.1016/j.celrep.2017.08.078

Larimore, K., McCormick, M. W., Robins, H. S., & Greenberg, P. D. (2012). Shaping of Human Germline IgH Repertoires Revealed by Deep Sequencing. *The Journal of Immunology*, *189*(6), 3221–3230. https://doi.org/10.4049/jimmunol.1201303

Lawrie, C. H., Soneji, S., Marafioti, T., Cooper, C. D. O., Palazzo, S., Paterson, J. C., Cattan, H., Enver, T., Mager, R., Boultwood, J., Wainscoat, J. S., & Hatton, C. S. R. (2007). Microrna expression distinguishes between germinal center B cell-like and activated B cell-like subtypes of diffuse large B cell lymphoma. *International Journal of Cancer*, *121*(5), 1156–1161. https://doi.org/10.1002/ijc.22800

Levitt, M. (1978). How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(2), 640–644.

Liu, P., Liu, D., Yang, X., Gao, J., Chen, Y., Xiao, X., Liu, F., Zou, J., Wu, J., Ma, J., Zhao, F., Zhou, X., Gao, G. F., & Zhu, B. (2014). Characterization of human αβTCR repertoire and discovery of D-D fusion in TCRβ chains. *Protein & Cell*, *5*(8), 603–615. https://doi.org/10.1007/s13238-014-0060-1

Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nature Reviews Genetics*, *21*(10), Article 10. https://doi.org/10.1038/s41576-020-0236-x

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, *30*(5), Article 5. https://doi.org/10.1038/nbt.2198

Lopez-Perez, D., Redruello-Romero, A., Garcia-Rubio, J., Arana, C., Garcia-Escudero, L. A., Tamayo, F., Puentes-Pardo, J. D., Moreno-SanJuan, S., Salmeron, J., Blanco, A., Galvez, J., Leon, J., & Carazo, Á. (2021). In Patients With Obesity, the Number of Adipose Tissue Mast Cells Is Significantly Lower in Subjects With Type 2 Diabetes. *Frontiers in Immunology*, *12*, 664576. https://doi.org/10.3389/fimmu.2021.664576

Lopez-Perez, D., Redruello-Romero, A., Garcia-Rubio, J., Arana, C., Garcia-Escudero, L. A., Tamayo, F., Salmeron, J., Galvez, J., Leon, J., & Carazo, Á. (2022). In Obese Patients With Type 2 Diabetes, Mast Cells in Omental Adipose Tissue Decrease the Surface Expression of CD45, CD117, CD203c, and FcεRI. *Frontiers in Endocrinology*, *13*, 818388. https://doi.org/10.3389/fendo.2022.818388

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A., Downing, J. R., Jacks, T., Horvitz, H. R., & Golub, T. R. (2005). MicroRNA expression profiles classify human cancers. *Nature*, *435*(7043), Article 7043. https://doi.org/10.1038/nature03702

Lumeng, C. N., DelProposto, J. B., Westcott, D. J., & Saltiel, A. R. (2008). Phenotypic Switching of Adipose Tissue Macrophages With Obesity Is Generated by Spatiotemporal Differences in Macrophage Subtypes. *Diabetes*, *57*(12), 3239–3246. https://doi.org/10.2337/db08-0872

Luong, Q., Huang, J., & Lee, K. Y. (2019). Deciphering White Adipose Tissue Heterogeneity. *Biology*, *8*(2), 23. https://doi.org/10.3390/biology8020023

Ma, L., Yang, L., Bin Shi, He, X., Peng, A., Li, Y., Zhang, T., Sun, S., Ma, R., & Yao, X. (2016). Analyzing the CDR3 Repertoire with respect to TCR—Beta Chain V-D-J and V-J Rearrangements in Peripheral T Cells using HTS. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep29544

Mackay, L. K., Stock, A. T., Ma, J. Z., Jones, C. M., Kent, S. J., Mueller, S. N., Heath, W. R., Carbone, F. R., & Gebhardt, T. (2012). Long-lived epithelial immunity by tissue-resident memory

T (TRM) cells in the absence of persisting local antigen presentation. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(18), 7037–7042. https://doi.org/10.1073/pnas.1202288109

Madi, A., Shifrut, E., Reich-Zeliger, S., Gal, H., Best, K., Ndifon, W., Chain, B., Cohen, I. R., & Friedman, N. (2014). T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. *Genome Research*, *24*(10), 1603–1612. https://doi.org/10.1101/gr.170753.113

Mandelcorn-Monson, R. L., Shear, N. H., Yau, E., Sambhara, S., Barber, B. H., Spaner, D., & DeBenedette, M. A. (2003). Cytotoxic T Lymphocyte Reactivity to gp100, MelanA/MART-1, and Tyrosinase, in HLA-A2-Positive Vitiligo Patients. *Journal of Investigative Dermatology*, *121*(3), 550–556. https://doi.org/10.1046/j.1523-1747.2003.12413.x

McCombie, W. R., McPherson, J. D., & Mardis, E. R. (2019). Next-Generation Sequencing Technologies. *Cold Spring Harbor Perspectives in Medicine*, *9*(11), a036798. https://doi.org/10.1101/cshperspect.a036798

McInerney, P., Adams, P., & Hadi, M. Z. (2014). Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Molecular Biology International*, *2014*, 287430. https://doi.org/10.1155/2014/287430

Meek, K. D., Hasemann, C. A., & Capra, J. D. (1989). Novel rearrangements at the immunoglobulin D locus. Inversions and fusions add to IgH somatic diversity. *The Journal of Experimental Medicine*, *170*(1), 39–57. https://doi.org/10.1084/jem.170.1.39

Meyer, D., C. Aguiar, V. R., Bitarello, B. D., C. Brandt, D. Y., & Nunes, K. (2018). A genomic perspective on HLA evolution. *Immunogenetics*, *70*(1), 5–27. https://doi.org/10.1007/s00251-017-1017-3

Meza-Perez, S., & Randall, T. D. (2017). Immunological functions of the omentum. *Trends in Immunology*, *38*(7), 526–536. https://doi.org/10.1016/j.it.2017.03.002

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., Peterson, A., Noteboom, J., O'Briant, K. C., Allen, A., Lin, D. W., Urban, N., Drescher, C. W., Knudsen, B. S., Stirewalt, D. L., Gentleman, R., Vessella, R. L., Nelson, P. S., Martin, D. B., & Tewari, M. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(30), 10513–10518. https://doi.org/10.1073/pnas.0804549105

Moro, K., Yamada, T., Tanabe, M., Takeuchi, T., Ikawa, T., Kawamoto, H., Furusawa, J., Ohtani, M., Fujii, H., & Koyasu, S. (2010). Innate production of TH2 cytokines by adipose tissue-associated c-Kit+Sca-1+ lymphoid cells. *Nature*, *463*(7280), Article 7280. https://doi.org/10.1038/nature08636

Munafó, D. B., & Robb, G. B. (2010). Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. *RNA*, *16*(12), 2537–2552. https://doi.org/10.1261/rna.2242610

Muschaweckh, A., Buchholz, V. R., Fellenzer, A., Hessel, C., König, P.-A., Tao, S., Tao, R., Heikenwälder, M., Busch, D. H., Korn, T., Kastenmüller, W., Drexler, I., & Gasteiger, G. (2016). Antigen-dependent competition shapes the local repertoire of tissue-resident memory CD8+ T cells. *The Journal of Experimental Medicine*, *213*(13), 3075–3086. https://doi.org/10.1084/jem.20160888

Muzzey, D., Evans, E. A., & Lieber, C. (2015). Understanding the Basics of NGS: From Mechanism to Variant Calling. *Current Genetic Medicine Reports*, *3*(4), 158–165. https://doi.org/10.1007/s40142-015-0076-8

Nanopore DNA Sequencing. (n.d.). Retrieved 26 February 2024, from https://www.genome.gov/genetics-glossary/Nanopore-DNA-Sequencing

Nikolich-Žugich, J. (2008). Ageing and life-long maintenance of T-cell subsets in the face of latent persistent infections. *Nature Reviews Immunology*, *8*(7), Article 7. https://doi.org/10.1038/nri2318

O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology*, *9*. https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2018.00402

Outters, P., Jaeger, S., Zaarour, N., & Ferrier, P. (2015). Long-Range Control of V(D)J Recombination & Allelic Exclusion: Modeling Views. *Advances in Immunology*, *128*, 363–413. https://doi.org/10.1016/bs.ai.2015.08.002

Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D., & Zamore, P. D. (2019). PIWI-interacting RNAs: Small RNAs with big functions. *Nature Reviews Genetics*, *20*(2), Article 2. https://doi.org/10.1038/s41576-018-0073-3

Pai, J. A., & Satpathy, A. T. (2021). High-throughput and single-cell T cell receptor sequencing technologies. *Nature Methods*, *18*(8), 881–892. https://doi.org/10.1038/s41592-021-01201-8

Paramasivam, A., & Vijayashree Priyadharsini, J. (2020). MitomiRs: New emerging microRNAs in mitochondrial dysfunction and cardiovascular disease. *Hypertension Research: Official Journal of the Japanese Society of Hypertension*, *43*(8), 851–853. https://doi.org/10.1038/s41440-020-0423-3

Pauley, K. M., Cha, S., & Chan, E. K. L. (2009). MicroRNA in autoimmunity and autoimmune diseases. *Journal of Autoimmunity*, *32*(3–4), 189–194. https://doi.org/10.1016/j.jaut.2009.02.012

Payne, A., Holmes, N., Rakyan, V., & Loose, M. (2019). BulkVis: A graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics*, *35*(13), 2193–2198. https://doi.org/10.1093/bioinformatics/bty841

Peng, Y., & Croce, C. M. (2016). The role of MicroRNAs in human cancer. *Signal Transduction and Targeted Therapy*, *1*, 15004. https://doi.org/10.1038/sigtrans.2015.4

Potapov, V., Ong, J. L., Langhorst, B. W., Bilotti, K., Cahoon, D., Canton, B., Knight, T. F., Thomas C Evans, J., & Lohman, G. J. S. (2018). A single-molecule sequencing assay for the comprehensive profiling of T4 DNA ligase fidelity and bias during DNA end-joining. *Nucleic Acids Research*, *46*(13), e79. https://doi.org/10.1093/nar/gky303

Prasad, S., Starck, S. R., & Shastri, N. (2016). Presentation of cryptic peptides by MHC I is enhanced by inflammatory stimuli. *Journal of Immunology (Baltimore, Md. : 1950)*, *197*(8), 2981–2991. https://doi.org/10.4049/jimmunol.1502045

Prieur, X., Mok, C. Y. L., Velagapudi, V. R., Núñez, V., Fuentes, L., Montaner, D., Ishikawa, K., Camacho, A., Barbarroja, N., O'Rahilly, S., Sethi, J. K., Dopazo, J., Orešič, M., Ricote, M., & Vidal-Puig, A. (2011). Differential Lipid Partitioning Between Adipocytes and Tissue Macrophages Modulates Macrophage Lipotoxicity and M2/M1 Polarization in Obese Mice. *Diabetes*, *60*(3), 797–809. https://doi.org/10.2337/db10-0705

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., & Gu, Y. (2012). A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, *13*(1), 341. https://doi.org/10.1186/1471-2164-13-341

Raabe, C. A., Tang, T.-H., Brosius, J., & Rozhdestvensky, T. S. (2014). Biases in small RNA deep sequencing data. *Nucleic Acids Research*, *42*(3), 1414–1426. https://doi.org/10.1093/nar/gkt1021

Rao, X., Hoof, I., Fontaine Costa, A. I. C. A., van Baarle, D., & Keşmir, C. (2011). HLA class I allele promiscuity revisited. *Immunogenetics*, *63*(11), 691–701. https://doi.org/10.1007/s00251-011-0552-6

Redruello-Romero, A., Benitez-Cantos, M. S., Lopez-Perez, D., García-Rubio, J., Tamayo, F., Pérez-Bartivas, D., Moreno-SanJuan, S., Ruiz-Palmero, I., Puentes-Pardo, J. D., Vilchez, J. R., López-Nevot, M. Á., García, F., Cano, C., León, J., & Carazo, Á. (2023). Human adipose tissue as a major reservoir of cytomegalovirus-reactive T cells. *Frontiers in Immunology*, *14*. https://www.frontiersin.org/articles/10.3389/fimmu.2023.1303724

Reilly, E. C., Lambert Emo, K., Buckley, P. M., Reilly, N. S., Smith, I., Chaves, F. A., Yang, H., Oakes, P. W., & Topham, D. J. (2020). TRM integrins CD103 and CD49a differentially support adherence and motility after resolution of influenza virus infection. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(22), 12306–12314. https://doi.org/10.1073/pnas.1915681117

Richter, M. V., & Topham, D. J. (2007). The α1β1 Integrin and TNF Receptor II Protect Airway CD8+ Effector T Cells from Apoptosis during Influenza Infection1. *The Journal of Immunology*, *179*(8), 5054–5063. https://doi.org/10.4049/jimmunol.179.8.5054

Robins, H. S., Srivastava, S. K., Campregher, P. V., Turtle, C. J., Andriesen, J., Riddell, S. R., Carlson, C. S., & Warren, E. H. (2010). Overlap and effective size of the human CD8+ T-cell receptor repertoire. *Science Translational Medicine*, *2*(47), 47ra64. https://doi.org/10.1126/scitranslmed.3001442

Ronaghi, M., Uhlén, M., & Nyrén, P. (1998). A Sequencing Method Based on Real-Time Pyrophosphate. *Science*, *281*(5375), 363–365. https://doi.org/10.1126/science.281.5375.363

Rosenfeld, N., Aharonov, R., Meiri, E., Rosenwald, S., Spector, Y., Zepeniuk, M., Benjamin, H., Shabes, N., Tabak, S., Levy, A., Lebanony, D., Goren, Y., Silberschein, E., Targan, N., Ben-Ari, A., Gilad, S., Sion-Vardy, N., Tobar, A., Feinmesser, M., … Barshack, I. (2008). MicroRNAs accurately identify cancer tissue origin. *Nature Biotechnology*, *26*(4), 462–469. https://doi.org/10.1038/nbt1392

Safonova, Y., & Pevzner, P. A. (2020). V(DD)J recombination is an important and evolutionarily conserved mechanism for generating antibodies with unusually long CDR3s. *Genome Research*, *30*(11), 1547–1558. https://doi.org/10.1101/gr.259598.119

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467.

Sathaliyawala, T., Kubota, M., Yudanin, N., Turner, D., Camp, P., Thome, J. J. C., Bickham, K. L., Lerner, H., Goldstein, M., Sykes, M., Kato, T., & Farber, D. L. (2013). Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets. *Immunity*, *38*(1), 187–197. https://doi.org/10.1016/j.immuni.2012.09.020

Scheja, L., & Heeren, J. (2019). The endocrine function of adipose tissues in health and cardiometabolic disease. *Nature Reviews Endocrinology*, *15*(9), Article 9. https://doi.org/10.1038/s41574-019-0230-6

Schuldt, N. J., & Binstadt, B. A. (2019). Dual TCR T Cells: Identity Crisis or Multitaskers? *Journal of Immunology (Baltimore, Md.: 1950)*, *202*(3), 637–644. https://doi.org/10.4049/jimmunol.1800904

Sckisel, G. D., Mirsoian, A., Minnar, C. M., Crittenden, M., Curti, B., Chen, J. Q., Blazar, B. R., Borowsky, A. D., Monjazeb, A. M., & Murphy, W. J. (2017). Differential phenotypes of memory

CD4 and CD8 T cells in the spleen and peripheral tissues following immunostimulatory therapy. *Journal for Immunotherapy of Cancer*, *5*, 33. https://doi.org/10.1186/s40425-017-0235-4

Sell, H., Habich, C., & Eckel, J. (2012). Adaptive immunity in obesity and insulin resistance. *Nature Reviews Endocrinology*, *8*(12), Article 12. https://doi.org/10.1038/nrendo.2012.114

Sharma, P., & Kranz, D. M. (2018). Subtle changes at the variable domain interface of the T-cell receptor can strongly increase affinity. *Journal of Biological Chemistry*, *293*(5), 1820–1834. https://doi.org/10.1074/jbc.M117.814152

Shi, H., Zhou, Y., Jia, E., Pan, M., Bai, Y., & Ge, Q. (2021). Bias in RNA-seq Library Preparation: Current Challenges and Solutions. *BioMed Research International*, *2021*, e6647597. https://doi.org/10.1155/2021/6647597

Sinclair, C., Bains, I., Yates, A. J., & Seddon, B. (2013). Asymmetric thymocyte death underlies the CD4:CD8 T-cell ratio in the adaptive immune system. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(31), E2905–E2914. https://doi.org/10.1073/pnas.1304859110

Song, Y., Liu, K. J., & Wang, T.-H. (2014). Elimination of Ligation Dependent Artifacts in T4 RNA Ligase to Achieve High Efficiency and Low Bias MicroRNA Capture. *PLoS ONE*, *9*(4), e94619. https://doi.org/10.1371/journal.pone.0094619

Speiser, D. E., Utzschneider, D. T., Oberle, S. G., Münz, C., Romero, P., & Zehn, D. (2014). T cell differentiation in chronic infection and cancer: Functional adaptation or exhaustion? *Nature Reviews Immunology*, *14*(11), Article 11. https://doi.org/10.1038/nri3740

Szeto, C., Lobos, C. A., Nguyen, A. T., & Gras, S. (2021). TCR Recognition of Peptide–MHC-I: Rule Makers and Breakers. *International Journal of Molecular Sciences*, *22*(1), Article 1. https://doi.org/10.3390/ijms22010068

Tian, G., Yin, X., Luo, H., Xu, X., Bolund, L., & Zhang, X. (2010). Sequencing bias: Comparison of different protocols of MicroRNA library construction. *BMC Biotechnology*, *10*(1), 64. https://doi.org/10.1186/1472-6750-10-64

Trim, W. V., & Lynch, L. (2022). Immune and non-immune functions of adipose tissue leukocytes. *Nature Reviews Immunology*, *22*(6), Article 6. https://doi.org/10.1038/s41577-021-00635-7

Turchinovich, A., Weiz, L., & Burwinkel, B. (2013). Isolation of circulating microRNA associated with RNA-binding protein. *Methods in Molecular Biology (Clifton, N.J.)*, *1024*, 97–107. https://doi.org/10.1007/978-1-62703-453-1_8

Turner, S. J., Doherty, P. C., McCluskey, J., & Rossjohn, J. (2006). Structural determinants of T-cell receptor bias in immunity. *Nature Reviews. Immunology*, *6*(12), 883–894. https://doi.org/10.1038/nri1977

Ueno, T., Tomiyama, H., Fujiwara, M., Oka, S., & Takiguchi, M. (2003). HLA class I-restricted recognition of an HIV-derived epitope peptide by a human T cell receptor α chain having a Vδ1 variable segment. *European Journal of Immunology*, *33*(10), 2910–2916. https://doi.org/10.1002/eji.200324148

Unamuno, X., Gómez-Ambrosi, J., Rodríguez, A., Becerril, S., Frühbeck, G., & Catalán, V. (2018). Adipokine dysregulation and adipose tissue inflammation in human obesity. *European Journal of Clinical Investigation*, *48*(9), e12997. https://doi.org/10.1111/eci.12997

Van Braeckel-Budimir, N., Gras, S., Ladell, K., Josephs, T. M., Pewe, L., Urban, S. L., Miners, K. L., Farenc, C., Price, D. A., Rossjohn, J., & Harty, J. T. (2017). A T Cell Receptor Locus Harbors

a Malaria-Specific Immune Response Gene. *Immunity*, *47*(5), 835-847.e4. https://doi.org/10.1016/j.immuni.2017.10.013

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends in Genetics*, *34*(9), 666–681. https://doi.org/10.1016/j.tig.2018.05.008

van Dijk, E. L., Jaszczyszyn, Y., & Thermes, C. (2014). Library preparation methods for next-generation sequencing: Tone down the bias. *Experimental Cell Research*, *322*(1), 12–20. https://doi.org/10.1016/j.yexcr.2014.01.008

van Dijk, E. L., Naquin, D., Gorrichon, K., Jaszczyszyn, Y., Ouazahrou, R., Thermes, C., & Hernandez, C. (2023). Genomics in the long-read sequencing era. *Trends in Genetics*, *39*(9), 649–671. https://doi.org/10.1016/j.tig.2023.04.006

Vickers, K. C., Palmisano, B. T., Shoucri, B. M., Shamburek, R. D., & Remaley, A. T. (2011). MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nature Cell Biology*, *13*(4), 423–433. https://doi.org/10.1038/ncb2210

Virtue, S., & Vidal-Puig, A. (2008). It's Not How Fat You Are, It's What You Do with It That Counts. *PLoS Biology*, *6*(9), e237. https://doi.org/10.1371/journal.pbio.0060237

Walsh, D. A., da Silva, H. B., Beura, L. K., Peng, C., Hamilton, S. E., Masopust, D., & Jameson, S. C. (2019). The functional requirement for CD69 in establishment of resident memory CD8+ T cells varies with tissue location. *Journal of Immunology (Baltimore, Md. : 1950)*, *203*(4), 946–955. https://doi.org/10.4049/jimmunol.1900052

Wanjalla, C. N., McDonnell, W. J., Ram, R., Chopra, A., Gangula, R., Leary, S., Mashayekhi, M., Simmons, J. D., Warren, C. M., Bailin, S., Gabriel, C. L., Guo, L., Furch, B. D., Lima, M. C., Woodward, B. O., Hannah, L., Pilkinton, M. A., Fuller, D. T., Kawai, K., … Koethe, J. R. (2021). Single-cell analysis shows that adipose tissue of persons with both HIV and diabetes is enriched for clonal, cytotoxic, and CMV-specific CD4+ T cells. *Cell Reports Medicine*, *2*(2), 100205. https://doi.org/10.1016/j.xcrm.2021.100205

Watson, J. D., & Crick, F. H. C. (1974). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, *248*(5451), Article 5451. https://doi.org/10.1038/248765a0

Watson, L. C., Moffatt-Blue, C. S., McDonald, R. Z., Kompfner, E., Ait-Azzouzene, D., Nemazee, D., Theofilopoulos, A. N., Kono, D. H., & Feeney, A. J. (2006). Paucity of V-D-D-J rearrangements and VH replacement events in lupus prone and nonautoimmune TdT-/- and TdT+/+ mice. *Journal of Immunology (Baltimore, Md.: 1950)*, *177*(2), 1120–1128. https://doi.org/10.4049/jimmunol.177.2.1120

Wei, Y., Li, L., Wang, D., Zhang, C.-Y., & Zen, K. (2014). Importin 8 Regulates the Transport of Mature MicroRNAs into the Cell Nucleus. *The Journal of Biological Chemistry*, *289*(15), 10270–10275. https://doi.org/10.1074/jbc.C113.541417

Weisberg, S. P., McCann, D., Desai, M., Rosenbaum, M., Leibel, R. L., & Ferrante, A. W. (2003). Obesity is associated with macrophage accumulation in adipose tissue. *The Journal of Clinical Investigation*, *112*(12), 1796–1808. https://doi.org/10.1172/JCI19246

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N. D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A. M., Schatz, M. C., Myers, G., DePristo, M. A., … Hunkapiller, M. W. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, *37*(10), Article 10. https://doi.org/10.1038/s41587-019-0217-9

Winter, J., Jung, S., Keller, S., Gregory, R. I., & Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature Cell Biology*, *11*(3), Article 3. https://doi.org/10.1038/ncb0309-228

Woodsworth, D. J., Castellarin, M., & Holt, R. A. (2013). Sequence analysis of T-cell repertoires in health and disease. *Genome Medicine*, *5*(10), 98. https://doi.org/10.1186/gm502

Wu, J., Cohen, P., & Spiegelman, B. M. (2013). Adaptive thermogenesis in adipocytes: Is beige the new brown? *Genes & Development*, *27*(3), 234–250. https://doi.org/10.1101/gad.211649.112

Wu, J., Yang, J., Cho, W. C., & Zheng, Y. (2020). Argonaute proteins: Structural features, functions and emerging roles. *Journal of Advanced Research*, *24*, 317–324. https://doi.org/10.1016/j.jare.2020.04.017

Xiao, M., Li, J., Li, W., Wang, Y., Wu, F., Xi, Y., Zhang, L., Ding, C., Luo, H., Li, Y., Peng, L., Zhao, L., Peng, S., Xiao, Y., Dong, S., Cao, J., & Yu, W. (2016). MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biology*, *14*(10), 1326–1334. https://doi.org/10.1080/15476286.2015.1112487

Yehudai-Resheff, S., & Schuster, G. (2000). Characterization of the E.coli poly(A) polymerase: Nucleotide specificity, RNA-binding affinities and RNA structure dependence. *Nucleic Acids Research*, *28*(5), 1139–1144.

Yohe, S., & Thyagarajan, B. (2017). Review of Clinical Next-Generation Sequencing. *Archives of Pathology & Laboratory Medicine*, *141*(11), 1544–1557. https://doi.org/10.5858/arpa.2016-0501-RA

Yu, L., & Guan, Y. (2014). Immunologic Basis for Long HCDR3s in Broadly Neutralizing Antibodies Against HIV-1. *Frontiers in Immunology*, *5*. https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2014.00250

Zhang, J., Ji, Z., & Smith, K. N. (2019). Chapter Twenty-Two—Analysis of TCR β CDR3 sequencing data for tracking anti-tumor immunity. In L. Galluzzi & N.-P. Rudqvist (Eds.), *Methods in Enzymology* (Vol. 629, pp. 443–464). Academic Press. https://doi.org/10.1016/bs.mie.2019.08.006

Zhang, L., Guerrero-Juarez, C. F., Hata, T., Bapat, S. P., Ramos, R., Plikus, M. V., & Gallo, R. L. (2015). Dermal adipocytes protect against invasive Staphylococcus aureus skin infection. *Science (New York, N.Y.)*, *347*(6217), 67–71. https://doi.org/10.1126/science.1260972

Zhang, L., Zhang, S., Yao, J., Lowery, F. J., Zhang, Q., Huang, W.-C., Li, P., Li, M., Wang, X., Zhang, C., Wang, H., Ellis, K., Cheerathodi, M., McCarty, J. H., Palmieri, D., Saunus, J., Lakhani, S., Huang, S., Sahin, A. A., … Yu, D. (2015). Microenvironment-induced PTEN loss by exosomal microRNA primes brain metastasis outgrowth. *Nature*, *527*(7576), 100–104. https://doi.org/10.1038/nature15376

Zhang, Z., Lee, J. E., Riemondy, K., Anderson, E. M., & Yi, R. (2013). High-efficiency RNA cloning enables accurate quantification of miRNA expression by deep sequencing. *Genome Biology*, *14*(10), R109. https://doi.org/10.1186/gb-2013-14-10-r109

Appendix

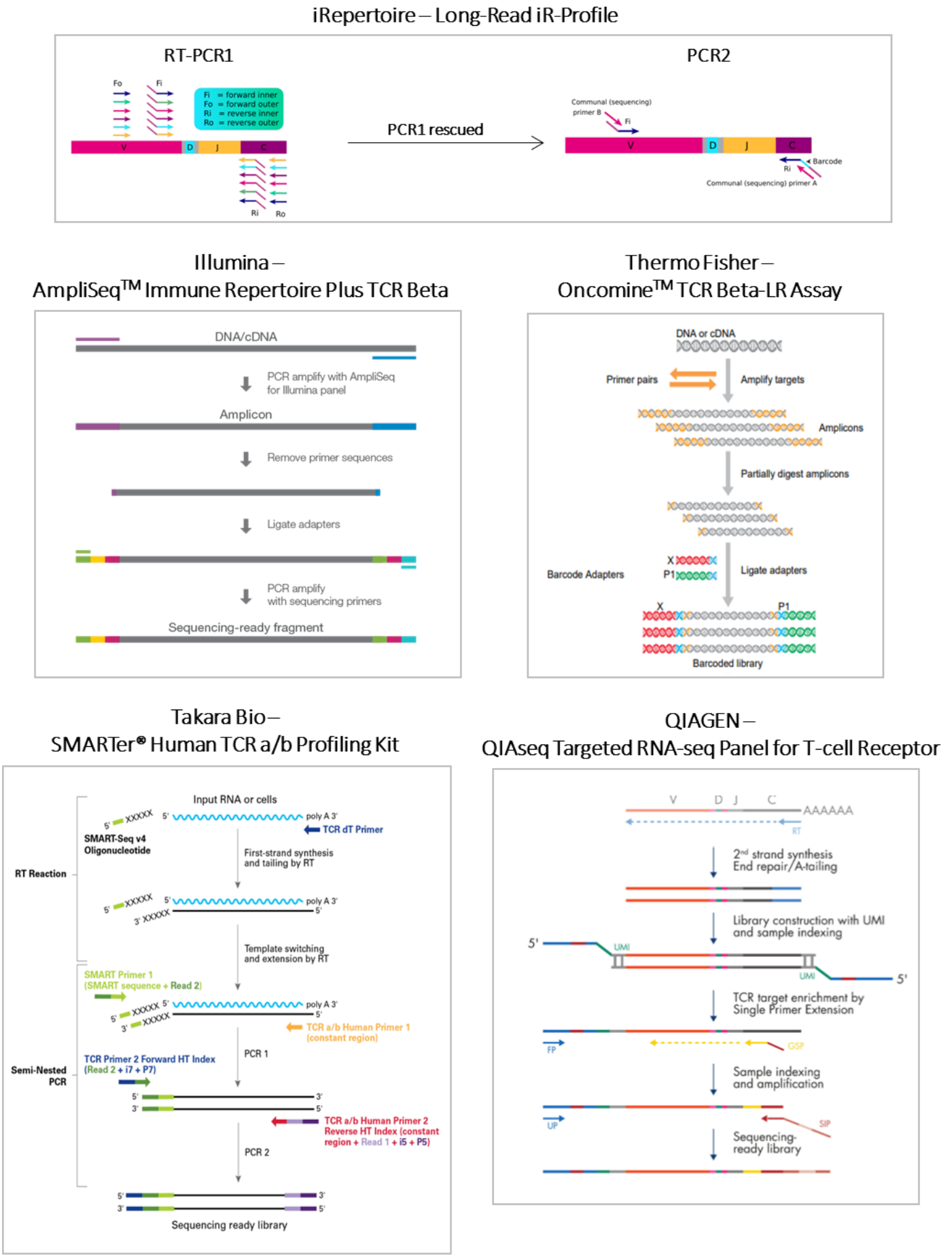

Figure A 1. *Graphical representation of commercial TCRβ library preparation kits from different companies. Modified from each company's user manual (iRepertoire, Illumina, Thermo Fisher, Takara Bio, and QIAGEN).*

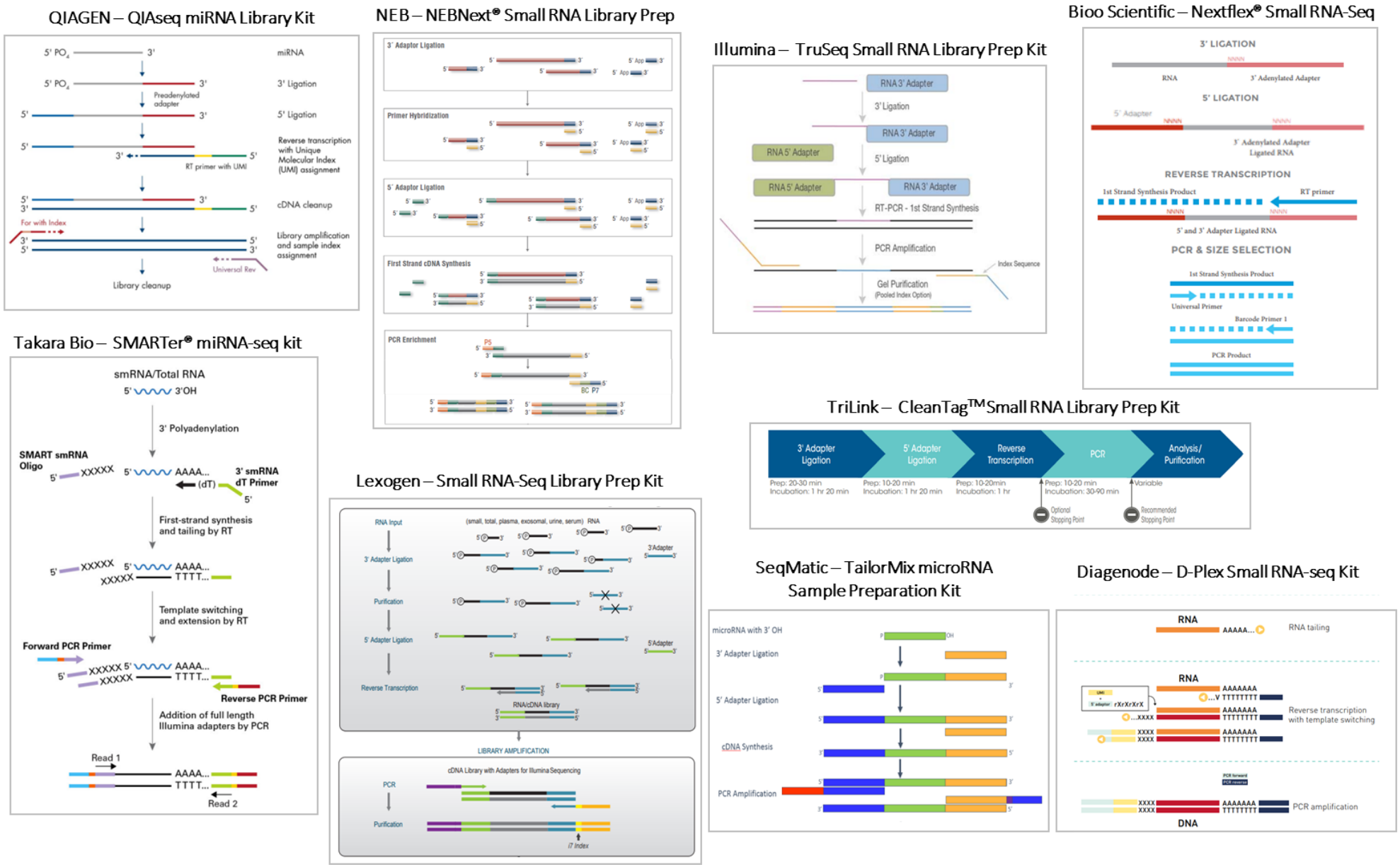

Figure A 2. Graphical representation of commercial TCRβ library preparation kits from different companies. Modified from each company's user manual (QIAGEN, New England Biolabs (NEB), Illumina, Bioo Scientific, Takara Bio, Lexogen, SeqMatic, TriLink BioTechnologies and Diagenode).

Table A 1. *Confocal microscopy characteristics.*

Fluorochrome	Laser (nm)
Hoescht 33342	405
FITC	488
Tx-Red	561

Table A 2. *BD FACS Aria IIIu characteristics.*

Fluorochrome	Laser (nm)	Filter
BV711	405	710/50
BV421	405	450/40
BV510	405	510/50
FITC	488	530/30
BB515	488	530/30
PE-Cy7	561	780/60
PE	561	582/15
APC	633	660/20

Table A 3. *Cytek Northern Lights characteristics.*

Fluorochrome	Laser (nm)	Filter
BV510	405	V7
Zombie Red (PE-TxRed)	488	B6

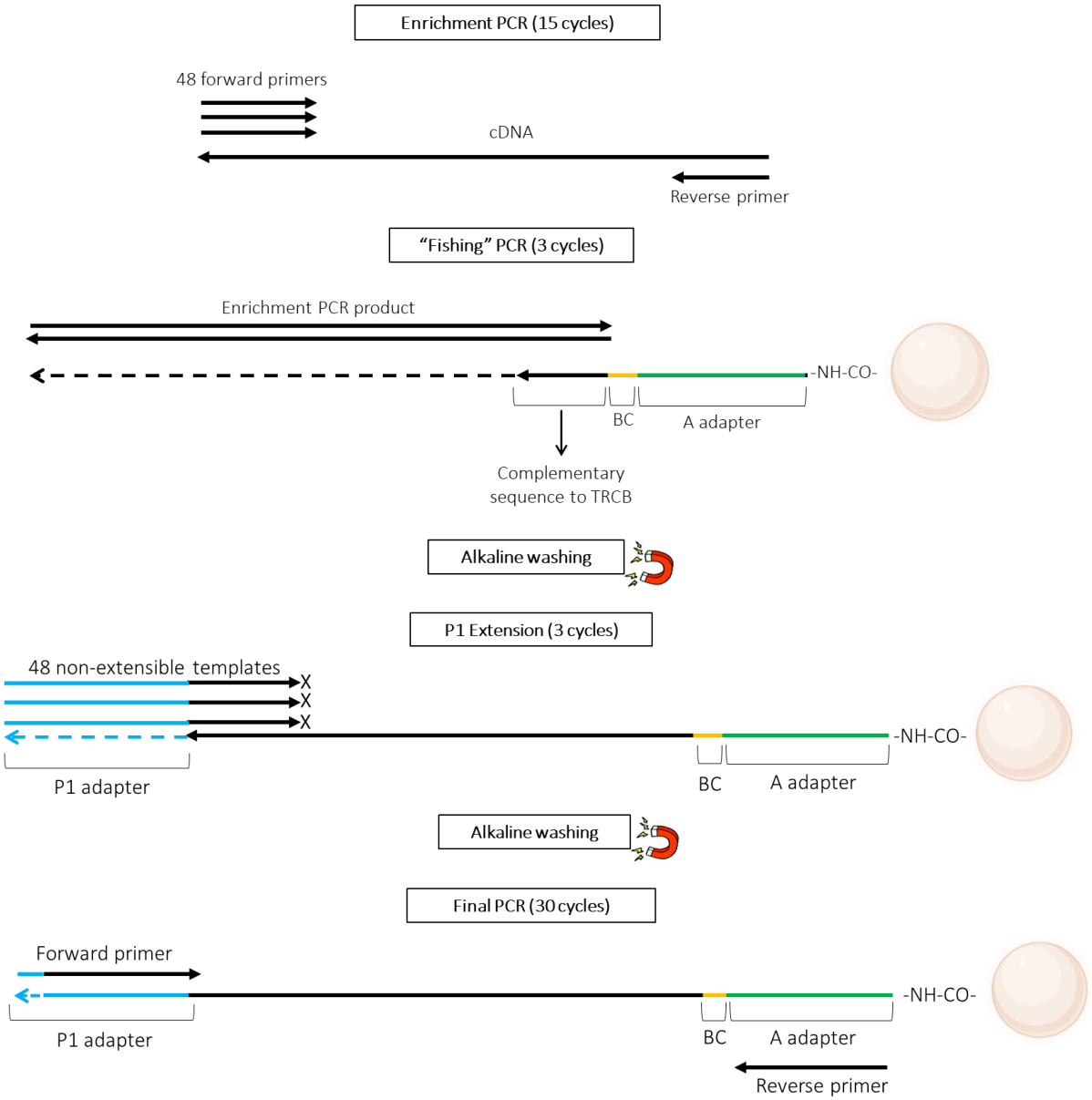

Figure A 3. *Overview of the TCRβ library preparation. Arrowheads simulate the 3' end of the DNA strands.*

Table A 4. List of oligonucleotides/templates for the TCRβ library preparation. B=G +T + C. Xs in yellow simulate the barcode sequence. In green, sequence of the A adapter. In blue, sequence of the P1 adapter.

Step	Forward oligonucleotide(s)/templates		Reverse oligonucleotide	
	Name	Sequence	Name	Sequence
Enrichment PCR	1.1-F	atcctgtagcacctgccat	Mx-R	tgtgggagatctctgcttctgttggctcaaa
	1.2-F	gtcctggagcacctgccat		
	1.3-F	atcctgcagttcctgccat		
	1.4-F	tcctgcagcccctgccat		
	1.5-F	cctg**B**cctgaccctgccat		
	1.6-F	ccttccctgaccctgccat		
	1.7-F	ctggcctgaccctgccat		
	1.8-F	agctaggagatcctgccat		
	1.9-F	cctgctgtgatcctgccat		
	1.10-F	ggggtggtattcctgccat		
	2.1-F	cctgccccactgtgccat		
	2.2-F	tcctgtccctctgtgccat		
	2.3-F	cctggcccactgtgccat		
	2.4-F	gcctgccccattgtgccat		
	3.1-F	gtctgcctcactctgccat		
	3.2-F	tctgctctcactctgccat		
	3.3-F	acctgccccaccctgccat		
	3.4-F	cccccagctccttggctat		
	4.1-F	tgtcttgtagcatctgccat		
	4.2-F	cgtcttgcagcatctgccat		
	5-F	ctagccccaactgtgccat		
	6-F	ctccatcctgcctcttcat		
	7-F	ttccttcttcaaagcagccat		
	8-F	ggctgctctgctgtgcgg		
	9.1-F	tcctgccctggagctgaaa		
	9.2-F	tcctggcctggacctgaaa		
	10.1-F	cctgccctgactctgtcat		

Enrichment PCR	10.2-F	cctggtctgacactgtcat		
	11.1-F	gaaggtggtgtgaggccat		
	11.2-F	ttctcaggggagaggccat		
	12-F	cactgcctgacctcaccat		
	13-F	caccaggctcctctgccat		
	14-F	gggcaccaggctcctcgg		
	15-F	tttacctgggtcctgccat		
	16-F	cctgtgggctccctccat		
	17-F	aaggcccccattgcactat		
	18-F	cccagactagctgaaggaaa		
	PG1.1-F	tccactgtggtgtggtccat		
	PG3.2-F	gctgcaggctcctctgctat		
	PG5.2-F	gcctgctacgctgtgccat		
	PG7.5-F	cctggcctgaccctgccat		
	PG8.2-F	ctcactctgaccctaccat		
	PG12.1-F	ttctgctctcactctgccat		
	PG12.2-F	atctgttctcactctgccat		
	PG16-F	tcttcccctaattctgccat		
	PG21.1-F	cttctctgctgtgtggccat		
	PG22.1-F	gagctgggtcctctgctat		
	PG26-F	cttctctgctgtgtgatcat		
“Fishing PCR”	-	-	A-BCX-R	5' Amino Modifier C12-ccatctcatccctgcgtgtctccgactcagXXXXXXXXXXcgatcagcgacctcgggtgggaaca
P1 extension	1.1-P1	ctacgcctccgctttcctctctatgggcagtcggtgatatcctgtagcacctgccat	-	-
	1.2-P1	ctacgcctccgctttcctctctatgggcagtcggtgatgtcctggagcacctgccat		
	1.3-P1	ctacgcctccgctttcctctctatgggcagtcggtgatatcctgcagttcctgccat		
	1.4-P1	ctacgcctccgctttcctctctatgggcagtcggtgattcctgcagcccctgccat		
	1.5-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcctg**B**cctgaccctgccat		
	1.6-P1	ctacgcctccgctttcctctctatgggcagtcggtgatccttccctgaccctgccat		
	1.7-P1	ctacgcctccgctttcctctctatgggcagtcggtgatctggcctgaccctgccat		
	1.8-P1	ctacgcctccgctttcctctctatgggcagtcggtgatagctaggagatcctgccat		
	1.9-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcctgctgtgatcctgccat		
	1.10-P1	ctacgcctccgctttcctctctatgggcagtcggtgatggggtggtattcctgccat		

P1 extension	2.1-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcctgccccactgtgccat	-	-
	2.2-P1	ctacgcctccgctttcctctctatgggcagtcggtgattcctgtccctctgtgccat		
	2.3-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcctggcccactgtgccat		
	2.4-P1	ctacgcctccgctttcctctctatgggcagtcggtgatgcctgccccattgtgccat		
	3.1-P1	ctacgcctccgctttcctctctatgggcagtcggtgatgtctgcctcactctgccat		
	3.2-P1	ctacgcctccgctttcctctctatgggcagtcggtgattctgctctcactctgccat		
	3.3-P1	ctacgcctccgctttcctctctatgggcagtcggtgatacctgccccaccctgccat		
	3.4-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcccccagctccttggctat		
	4.1-P1	ctacgcctccgctttcctctctatgggcagtcggtgattgtcttgtagcatctgccat		
	4.2-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcgtcttgcagcatctgccat		
	5-P1	ctacgcctccgctttcctctctatgggcagtcggtgatctagccccaactgtgccat		
	6-P1	ctacgcctccgctttcctctctatgggcagtcggtgatctccatcctgcctcttcat		
	7-P1	ctacgcctccgctttcctctctatgggcagtcggtgatttccttcttcaaagcagccat		
	8-P1	ctacgcctccgctttcctctctatgggcagtcggtgatggctgctctgctgtgcgg		
	9.1-P1	ctacgcctccgctttcctctctatgggcagtcggtgattcctgccctggagctgaaa		
	9.2-P1	ctacgcctccgctttcctctctatgggcagtcggtgattcctggcctggacctgaaa		
	10.1-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcctgccctgactctgtcat		
	10.2-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcctggtctgacactgtcat		
	11.1-P1	ctacgcctccgctttcctctctatgggcagtcggtgatgaaggtggtgtgaggccat		
	11.2-P1	ctacgcctccgctttcctctctatgggcagtcggtgatttctcaggggagaggccat		
	12-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcactgcctgacctcaccat		
	13-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcaccaggctcctctgccat		
	14-P1	ctacgcctccgctttcctctctatgggcagtcggtgatgggcaccaggctcctcgg		
	15-P1	ctacgcctccgctttcctctctatgggcagtcggtgatttacctgggtcctgccat		
	16-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcctgtgggctccctccat		
	17-P1	ctacgcctccgctttcctctctatgggcagtcggtgataaggcccccattgcactat		
	18-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcccagactagctgaaggaaa		
	PG1.1-P1	ctacgcctccgctttcctctctatgggcagtcggtgattccactgtggtgtggtccat		
	PG3.2-P1	ctacgcctccgctttcctctctatgggcagtcggtgatgctgcaggctcctctgctat		
	PG5.2-P1	ctacgcctccgctttcctctctatgggcagtcggtgatgcctgctacgctgtgccat		
	PG7.5-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcctggcctgaccctgccat		
	PG8.2-P1	ctacgcctccgctttcctctctatgggcagtcggtgatctcactctgaccctaccat		

	PG12.1-P1	ctacgcctccgctttcctctctatgggcagtcggtgatttctgctctcactctgccat	-	-
	PG12.2-P1	ctacgcctccgctttcctctctatgggcagtcggtgatatctgttctcactctgccat		
	PG16-P1	ctacgcctccgctttcctctctatgggcagtcggtgattcttcccctaattctgccat		
	PG21.1-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcttctctgctgtgtggccat		
	PG22.1-P1	ctacgcctccgctttcctctctatgggcagtcggtgatgagctgggtcctctgctat		
	PG26-P1	ctacgcctccgctttcctctctatgggcagtcggtgatcttctctgctgtgtgatcat		
Final PCR	Final-F	ccactacgcctccgctttcctctctatgg	Final-R	ccatctcatccctgcgtgtctccgactca

Table A 5. *Conditions for the different PCRs of the library preparation for TCRβ repertoire.*

	Enrichment PCR	**"Fishing" PCR**	**P1 extension**	**Final PCR**
Forward primer/s	75 nM	-	75 nM	200 nM
Reverse primer	150 nM	Microparticles	-	200 nM
$MgCl_2$	1.5 nM	-	1.5 nM	1.5 nM
dNTPs	0.2 mM	-	0.2 mM	0.2 mM
Microparticles	-	12µl	Microparticles after a washing step	Microparticles after a washing step
Total volume	20 µl	32 µl	100 µl	30 µl
Cycles	15	3	3	40
Annealing temperature (time)	62˚C (30'')	68˚C (1')	63˚C (3')	65˚C (30")
Elongation temperature (time)	74˚C (2')	74˚C (2')	74˚C (5')	74˚C (2')
Amplicon length	~550 bp	~500 bp	~550 bp	~550 pb

The amplicon length changes depending on the V segment and CDR3 composition

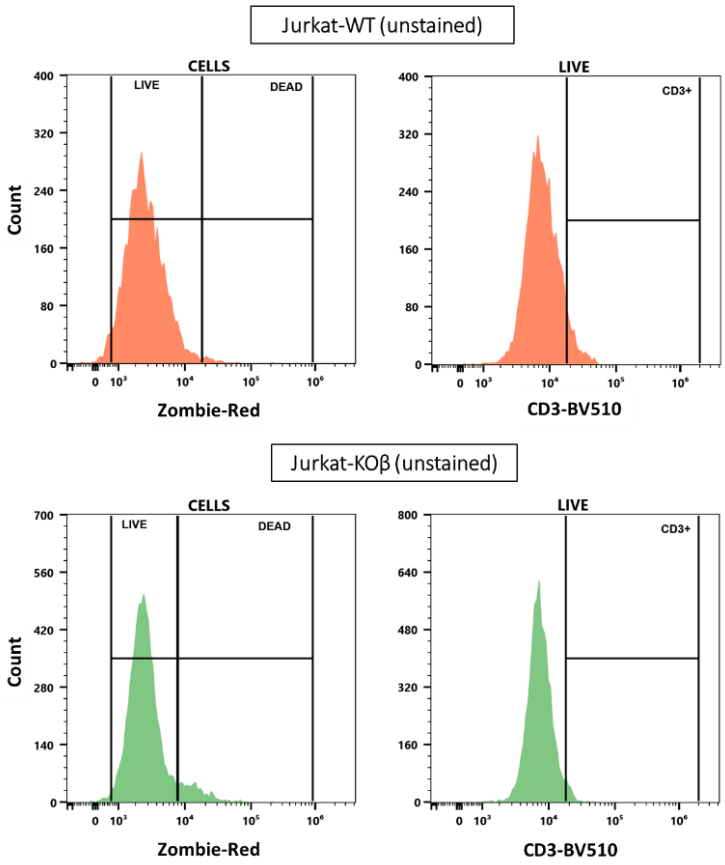

Figure A 4. *Unstained controls for cell viability and CD3-TCR expression of Jurkat-WT and Jurkat KOβ after the electroporation process.*

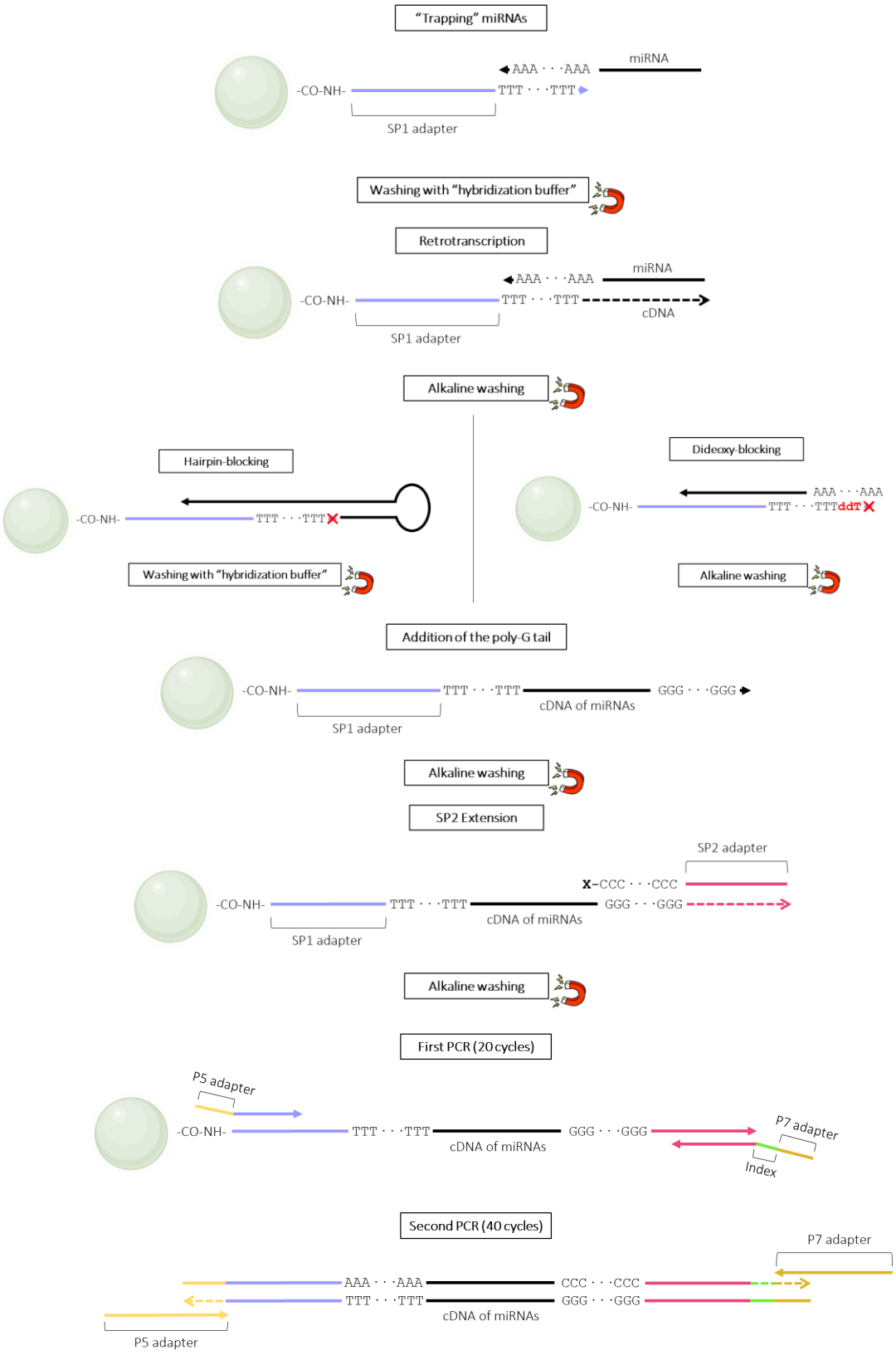

Figure A 5. *Overview of the Double-Tailing Trap methodology for small-RNA NGS library preparation. Arrowheads simulate the 3' end of the DNA strands.*

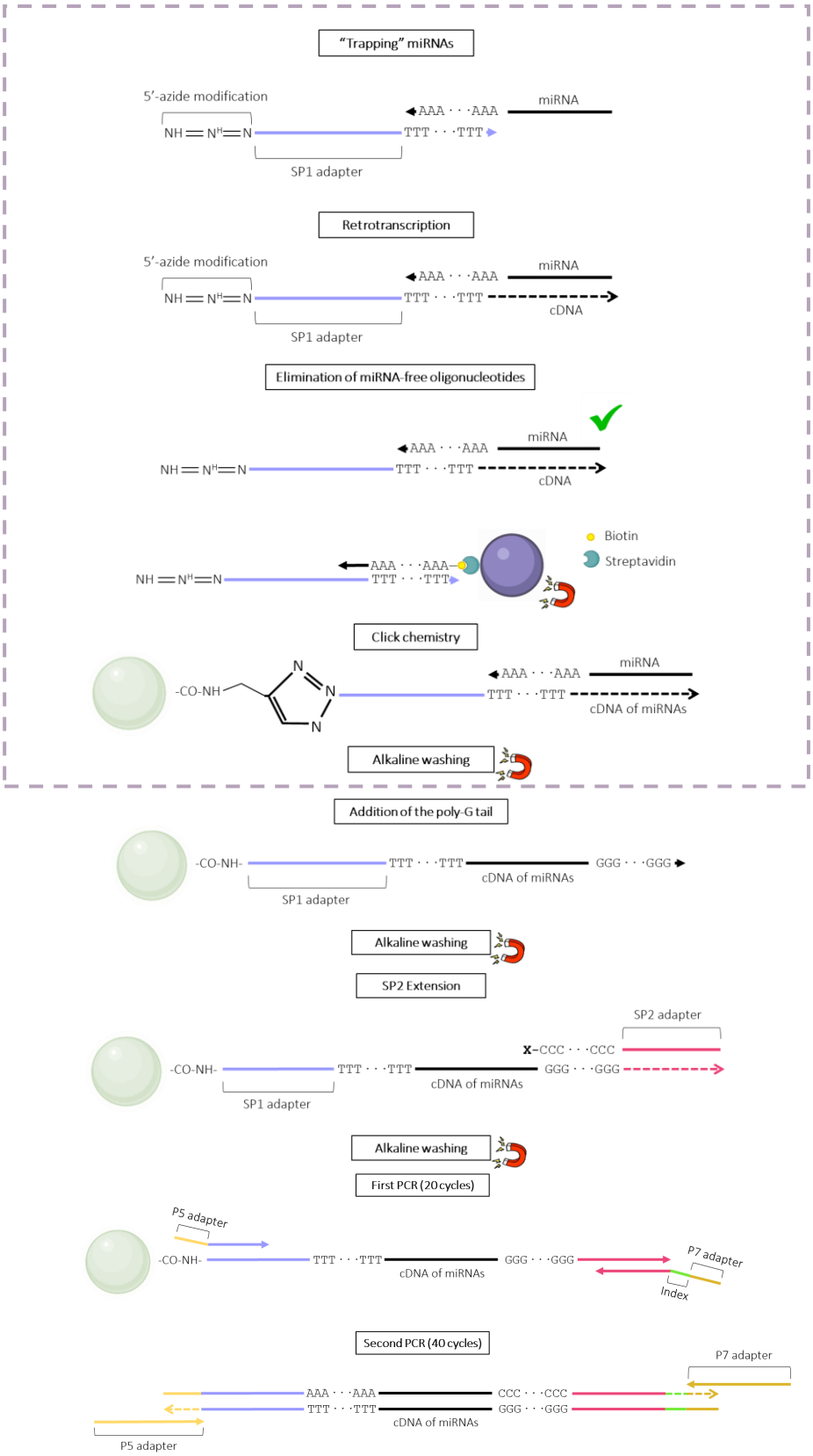

Figure A 6. *Overview of the DTT-Click Chemistry (CC) methodology for small-RNA NGS library preparation with the incorporation of click chemistry. Highlighted in the discontinuous rectangle are the steps that differ from the previous methodology. Arrowheads simulate the 3' end of the DNA strands.*

Table A 6. List of oligonucleotides/templates for the small-RNA library preparation by Double-Tailin Trap (DTT) and DTT-Click Chemistry (CC). In violet, sequence of the SP1 adapter. Underlined, free nucleotides of the hairpin's loop. In pink, sequence of the SP2 adapter. In yellow, sequence of the P5 adapter. Xs in green simulate the index sequence. In brown, sequence of the P7 adapter.

Method	Step	Forward oligonucleotides		Reverse oligonucleotides/templates	
		Name	Sequence	Name	Sequence
DTT	"Trapping" miRNAs	P1	5' Amino Modifier C12-acactctttccctacacgacgctcttccgatctttttttttttttttttt	-	-
DTT-CC		P1-az	5' Azide (NHS Ester)-acactctttccctacacgacgctcttccgatctttttttttttttttttt	-	-
DTT	Hairpin-blocking	-	-	Hairpin	gcgccacgattcaacgtggcgcaaaaaaaaaaaaaaaaaaagatcggaagagcgtcgtgta
	Dideoxy-blocking	-	-	Dideoxy	aaaaaaaaaaaaaaaaaaaaagatcggaagagcgtcgtg
DTT-CC	miRNA-free oligos elimination	-	-	Eliminator-biotin	5' Biotin-aaaaaaaagatcggaagagcgtcgtgtag
DTT	SP2 extension	-	-	SP2-ext	gtgactggagttcagacgtgtgctcttccgatcccccccccccccccccccc-3' Inverted dT
DTT-CC					
DTT	Final PCR A	A-F	caccgagatctacactctttccctacacga	A-R	cggcatacgagatXXXXXXgtgactggagttcagacg
DTT-CC					
DTT	Final PCR B	B-F	aatgatacggcgaccaccgagatctaca	B-R	caagcagaagacggcatacgag
DTT-CC					

Table A 7. Sequences of Simulator and synthetic miRNAs.

Name	Sequence
Simulator	agaagcagagaucucacacaaaaaaaaaaaaaaaaaaaaa
P1-miR-135a	uauggcuuuuuauuccuauguga
P2-miR-135b	uauggcuuuucauuccuauguga
P3-miR-26b	uucaaguaauucaggauaggu
P4-miR-34a	uggcagugucuuagcugguugu
P5-miR-34b	uaggcagugucauuagcugauug
P6-miR-34c	aggcaguguaguuagcugauugc
P7-miR-145	guccaguuuucccaggaaucccu
P8-miR-125b	ucccugagacccuaacuugugaaacttgtga
P9-miR-320a	aaaagcuggguugagagggcga
P10-miR-23b	aucacauugccagggauuaccac
P11-miR-17	caaagugcuuacagugcagguag
P12-miR-18a	uaaggugcaucuagugcagauag
P13-miR-20a	uaaagugcuuauagugcagguag
P14-miR-21	uagcuuaucagacugauguuga
P15-miR-23a	aucacauugccagggauuucc
P16-miR-24	uggcucaguucagcaggaacag
P17-miR-29c	uagcaccauuugaaaucgguua
P18-miR-125a	ucccugagacccuuuaaccuguga
P19-miR-130a	cagugcaauguuaaaagggcau
P20-miR-150	ucucccaacccuuguaccagug
P21-miR-200c	uaauacugccggguaaugaugga
P22-miR-210	cugugcgugugacagcggcuga
P23-miR-221	agcuacauugucugcuggguuuc
P24-miR-301a	cagugcaauaguauugucaaagc
P25-miR-365a	uaaugccccuaaaaauccuuau
P26-miR-454	uagugcaauauugcuuauagggu
P27-miR-663b	gguggcccggccgugccugagg
P28-miR-synthetic1	gaaguccucuuccaacuccuguug
P29-miR-synthetic2	guagguucacuccuguguuccuac
P30-miR-synthetic3	uggaucaccauggacuacca

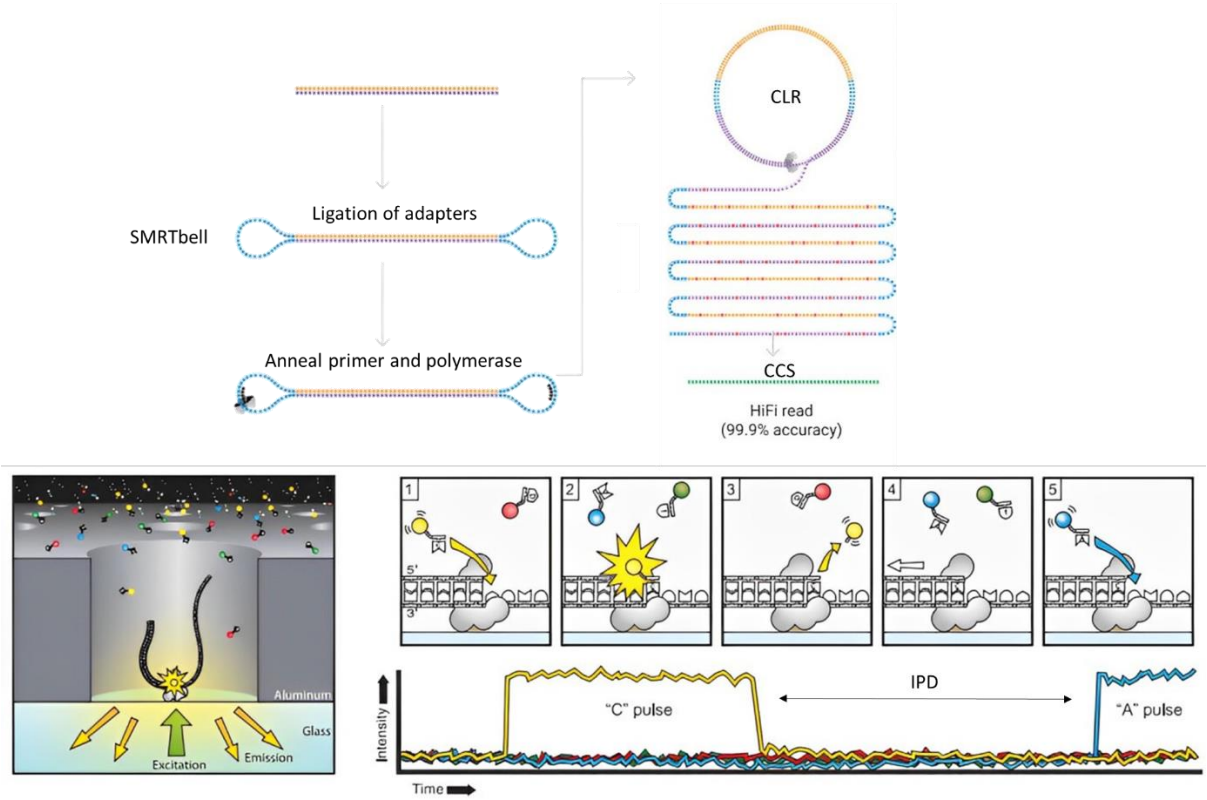

Figure A 7. *Schematic representation of the Pacific Biosciences (PacBio) technology. CLR: continuous long read, CCD: circular consensus sequence, IPD: interpulse duration. Modified from PacBio and Eid et al., 2009.*

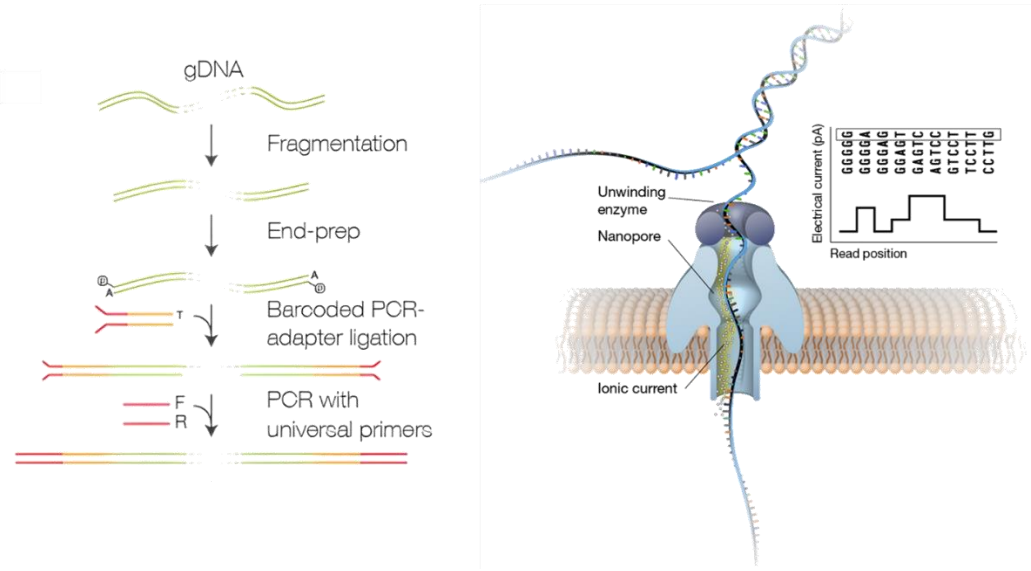

Figure A 8. *Schematic representation of the Oxford Nanopore Technologies (ONT). Modified from ONT and (*Nanopore DNA Sequencing*, n.d.)*