Development of a High-Throughput Single-Cell Sequencing Platform for the Discovery of Shared-Antigen and Neoepitope-Specific T-Cell Receptors

Dissertation zur Erlangung des Grades
"Doktor der Naturwissenschaften"

Am Fachbereich Biologie
Der Johannes Gutenberg-Universität Mainz

Shaheer El Bardisy

Geb. am 12.05.1993 in Kairo

Mainz, 2020

Datum der mündlichen Prüfung: 12 März 2020

Statutory declaration

I hereby declare that the submitted doctoral thesis "Development of a High-Throughput Single-Cell Sequencing Platform for the Discovery of Shared-Antigen and Neoepitope-Specific T-Cell Receptors" is, to the best of my knowledge and belief, in all parts my original work and that I have not received assistance from outside other than acknowledged. I have clearly indicated and referenced all used material and sources. This work has not been submitted, either substantially or in whole, for examination purposes at this or any other University before.

………………………………   ………………………………

Mainz   Shaheer El Bardisy

Dedication

This thesis is dedicated to my late Grandfather, my Mother, and my Father. Your unyielding love, ceaseless encouragement, and fountain of knowledge have laid the foundation of this, and many more endeavors to come.

Summary

Engineering tumor antigen-specific αβ T-cell receptor (TCR) genes into autologous T cells for adoptive T cell therapy (ACT) has demonstrated prolonged tumor regression in numerous cancers. Clinical advancement of ACT is hindered by the difficulty in retrieving potent and tumor-specific TCRs. Further, the paucity in murine TCR candidates impedes pre-clinical studies needed to advance the clinical success of TCR-based therapies. Despite advances in single cell TCR sequencing (scTCRseq) methods, drawbacks still exist, including low-throughput, high costs and complex protocols.

This thesis aimed to develop a high-throughput and easily applicable murine scTCRseq platform to facilitate the discovery of shared tumor antigen and neoepitope-specific murine TCRs.

A plate-based scTCRseq platform compatible with next-generation sequencing (NGS) was developed. Sample barcoding permits early pooling of single cell samples and 96-well plate multiplexing allows little hands-on work at a high-throughput. The use of standard laboratory equipment renders the platform easily applicable. Only 25 primers are used and both α and β TCR chains are amplified in the same polymerase chain reaction. An automated data analysis pipeline retrieves αβTCR information from the 2x300 bp paired-end MiSeq data. As an advantage over synthetic gene orders, TCRs of interest can directly be cloned for functional characterization using remaining cDNA. The platform was optimized improving αβTCR detection rates and minimizing platform noise, costs and complexity. Performance evaluation revealed an average paired αβTCR detection rate of 49 %. The scTCRseq platform revealed its superior performance compared to a previously published scTCR cloning platform as well as its ability to provide insight into the TCR repertoire.

The scTCRseq platform was applied to establish a library of human and murine HLA-restricted TCRs specific for oncoviral and cancer/testis epitopes, as well as neoantigens. The latter were strictly mutation specific and the majority of the TCRs could recognize the respective tumor cell line. We show that frequent retrieval of the same αβTCR, high read counts and CDR3 homology were indicative of antigen specificity.

Moreover, a gene-specific primer panel of commonly used T cell markers was integrated into the platform to allow the additional NGS-based phenotyping of single T cells. This scTCRPhenoSeq platform was applied to determine the phenotype and clonality of chimeric antigen receptor T cells after ACT and subsequent RNA-based *in vivo* expansion.

Together, a powerful tool was developed for the rapid discovery of TA-specific murine TCRs allowing the full exploitation of such TCRs for pre-clinical and clinical advancements of ACT for the benefit of cancer patients.

Zusammenfassung

In ersten Studien hat der adoptive Transfer αβ T-Zell-Rezeptor (TCR) transgener T-Zellen bei zahlreichen Krebsarten zu anhaltenden Tumorregressionen geführt. Der klinische Fortschritt von adoptiven T-Zell-Therapien ist jedoch wesentlich durch die Schwierigkeiten bei der Identifizierung potenter tumorspezifischer TCRs beeinträchtigt. Weiterhin besteht ein Mangel an murinen TCR Kandidaten für die Durchführung notwendiger präklinischer Studien. Trotz der Weiterentwicklung etablierter Einzelzell-TCR-Sequenzierungsverfahren (engl.: scTCRseq) existieren noch immer Schwächen wie ein zu niedriger Durchsatz, zu hohe Kosten und eine sehr hohe Komplexität der Protokolle.

Das Ziel dieser Arbeit war deshalb die Entwicklung einer einfach durchzuführenden scTCRseq Plattform zur Identifizierung von murinen TCRs im Hochdurchsatz. Im Anschluss sollte die Plattform eingesetzt werden, um TCRs zu isolieren, die tumorassoziierte (engl.: shared) und tumorspezifische (Neo) Antigene erkennen.

Im Rahmen dieser Arbeit wurde dementsprechend eine *Next-Generation Sequencing* (NGS) kompatible scTCRseq Plattform im 96-Well-Platten Format entwickelt. Dabei ermöglicht das Einführen von Barcodes sowohl ein frühes Zusammenführen von Einzelzellproben, als auch die Multiplexhandhabung einer Vielzahl von 96-Well-Platten, wodurch ein hoher Durchsatz gewährleistet ist. Darüber hinaus ist die Plattform durch die Verwendung von Standardlaborgeräten einfach zu etablieren und zu handhaben. Innerhalb des Prozesses werden insgesamt lediglich 25 Primer verwendet. Sowohl α als auch β TCR-Ketten werden in der gleichen Polymerase-Kettenreaktion amplifiziert. Eine automatisierte Datenanalyse-Pipeline ruft αβTCR Informationen aus den gepaarten 2x300 Basenpaaren langen MiSeq Reads ab. Relevante TCR Kandidaten die funktionell charakterisiert werden sollen, können aus der verbliebenen cDNA kloniert werden und müssen nicht kostenintensiv synthetisiert werden.

Die Plattform wurde hinsichtlich der αβTCR Erkennungsraten, Reduzierung von Hintergrundsignalen und Komplexität der Plattform, sowie Minimierung der Kosten optimiert. Die Leistungsevaluation ergab eine durchschnittliche Identifikationsrate von 49 % gepaarter αβTCR. Im Vergleich zu einer zuvor veröffentlichten scTCR Klonierungsplattform zeigte die neue scTCRseq Plattform eine verbesserte Lesitungspotential. Sie gewährt zudem Einblicke in das TCR Repertoire.

Im Rahmen dieser Arbeit wurde die scTCRseq Plattform eingesetzt, um eine Bibliothek von humanen und murinen MHC-restringierter TCRs zu etablieren, welche spezifisch onkoretrovirale, *Cancer/Testis* und Neoantigene erkennen. Letztere waren strikt mutationsspezifisch und für die Mehrheit derselben konnte Tumorzelllinienerkennung gezeigt werden. Es konnte außerdem gezeigt werden, dass häufiges Auffinden desselben αβTCRs, sowie hohe Readzahlen für einen TCR und CDR3-Homologien mehrerer TCR Kandidaten auf Antigenspezifität hindeuten.

Abschließend wurde ein Primer-Panel für häufig verwendete T-Zell-Marker in die Plattform integriert, um zusätzlich die NGS-basierte Phänotypisierung der T-Zellen zu ermöglichen. Diese scTCRPhenoSeq-Plattform wurde nachfolgend eingesetzt, um den Phänotyp und die Klonalität von chimären Antigenrezeptor tragenden (CAR) T-Zellen nach adoptivem Transfer und anschließender RNA-basierter *in vivo* Expansion zu bestimmen.

Im Rahmen dieser Arbeit wurde somit ein leistungsstarkes Werkzeug für die einfache und schnelle Identifizierung von murinen, tumorspezifischen TCRs entwickelt. Dies legt den Grundstein für die präklinische und klinische Weiterentwicklung, damit letztlich vermehrt hochpotente TCRs für die Immuntherapie von Krebspatienten zur Verfügung stehen.

Contents

List of Figures

List of Tables

Abbreviations

5′-RACE	5′ rapid amplification of cDNA ends
aa	Amino acid
ACT	Adoptive cell therapy
ADCC	Antibody-dependent cell-mediated cytotoxicity
Adpgk	ADP dependent glucokinase
APC	Antigen presenting cell
BCR	B cell receptor
BMDC	Bone marrow-derived dendritic cells
C/T	Cancer/testis
CAR	Chimeric antigen receptors
CBC	Column-specific barcode
CDC	Complement-dependent cytotoxicity
CDR	Complementary determining region
CLDN	Claudin
cTECs	Cortical thymic epithelial cells
CTL	Cytotoxic T cell
CTLA-4	Cytotoxic T cells antigen-4
DC	Dendritic cell
DN	Double negative
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
eGFP	Enhanced green fluorescent protein
ES	Exon-spanning
gDNA	Genomic DNA
GSP	Gene specific primer
HBV	Hepatitis B virus
HLA	Human leukocyte antigen
HPV	Human papillomavirus
ICI	Immune checkpoint inhibitor
IFNγ	Interferon-γ
Ig	Immunoglobulins
IL	Interleukin
ITAMs	Immunoreceptor tyrosine-based activation motifs
IVT	*In vitro* transcription
KKLC1	Kita-kyushu lung cancer antigen 1
LPX	Lipopolyplex
MAGE	Melanoma-associated antigen
MART-1	Melanoma differentiation antigen-1
MDSC	Myeloid-derived suppressor cell
MHC	Major histocompatibility complex complexes
MMLV	Murine leukemia viruses
mRNA	Messenger RNA

mTECs	Medullary thymic epithelial cells
NGS	Next-generation sequencing
NK	Natural killer cell
NY-ESO	New York Esophageal Squamous cell carcinoma
PBC	Plate-specific barcode
PBL	Peripheral blood lymphocyte
PCR	Polymerase chain reaction
PD-L1	Programmed cell death 1 ligand 1
PE	Paired-end
PMA	Phorbol 12-myristate 13-acetate.
PSA	Prostate-specific antigen
RBC	Row-specific barcodes
Reps1	RALBP1 Associated Eps Domain Containing 1
RNA	Ribonucleic acid
RSS	Recombination signal sequence
RT	Reverse transcription
RTase	Reverse transcriptase
Sc	Single cell
SEB	Staphylococcal enterotoxin B
SMAC	Supramolecular activation cluster
SMART	Switching Mechanism At 5' end of RNA Transcript
SNP	Single nucleotide polymorphism
SP	Single positive
TA	Tumor antigen
TAA	Tumor-associated antigen
TCR	T cell receptor
TdT	Terminal deoxynucleotidyl transferase
Th	T helper cell
TIL	Tumor-infiltrating T lymphocyte
TNBC	Triple-negative breast cancer
TNFα	Tumor necrosis factor-α
TRAC	TCR α chain constant gene
TRBC	TCR β chain constant gene
Treg	T regulatury cell
TS	Template switching
TSA	Tumor specific antigen
UTR	Untranslated region

1. Introduction

1.1. T cell immunity

1.1.1. An overview of the immune system

The immune system protects the host by self-nonself discrimination and subsequent pathogen destruction. It achieves this through an interactive network of immune cells, lymphoid organs, receptors and extracellular signalling proteins. Immunity is divided into two distinct yet interacting systems: innate and the adaptive immunity[1].

The innate immune response is rapid and composed of low antigen-specificity defence mechanisms such as pattern recognition receptors (e.g. Toll-like receptors) expressed on phagocytes and antigen presenting cells (APCs). In parallel, the complement pathway leads to pathogen lysis and pathogen opsonisation facilitating phagocytosis by dendritic cells (DCs). Antigen uptake results in DC maturation and migration to secondary lymphoid tissues to present antigen on major histocompatibility complex complexes (MHC) to T cells resulting in the activation of the adaptive immune response[2].

The adaptive immune response takes days to develop and manifests exquisite specificity for antigens via antigen-specific receptors expressed on lymphocytes: T and B cells. T and B cell receptors (TCR and BCR) induce lymphocyte proliferation upon antigen encounter. Naïve lymphocytes populate lymphoid tissues (e.g. lymph nodes) where they are primed and differentiate into effector cells following antigen encounter. Effector lymphocytes home to the site of infection where the second antigen encounter initiates an effector response. CD8[+] cytotoxic T cells (CTLs) induce apoptosis of infected cells. CD4[+] T helper (Th) cells orchestrate either a cellular response against intracellular pathogens or a humoral B-cell antibody-mediated response against extracellular antigens. T cells therefore play a cardinal role in the regulation and effector functions of an immune response[3].

1.1.2. T cell development and central tolerance

Contrary to B cells, T cells do not complete their development in the bone marrow but in the thymus (**Figure 1.1**)[3]. Common lymphoid progenitor cells migrated into the thymus at the corticomedullary junction, engraft as thymocytes and encounter networks of cortical thymic epithelial cells (cTECs)[4]. The thymocytes lack expression of CD4 and CD8 co-receptors and are termed double negative (DN). The DN population is divided into differentiation stages according to CD44 (adhesion molecule) and CD25 (IL-2 receptor α chain) expression. Most DN thymocytes give rise to αβ TCR chain heterodimer expressing T cells; however, 5% bear the γδ TCR chains for intraepithelial mucosal immunity. DN thymocyte differentiation is accompanied by movement towards the outer cortex of the

sub-capsular zone. Successful β TCR chain selection (together with a pre-α TCR chain) leads to DN cell survival, proliferation and expression of both CD4 and CD8 generating double positive (DP) cells. Failure to select a β TCR chain leads to apoptosis. DP thymocytes rearrange their α TCR chain loci, to produce an αβ TCR. αβ TCR-expressing DP cells interact with self-antigens expressed on MHC class I or II on cTECs during positive selection. Those cells that engage antigen-MHC complexes with an appropriate affinity survive, whereas those with a weaker affinity undergo apoptosis. T cells with adequate MHC I self-peptide complex affinity, retain expression of CD8 whilst those that recognize an MHC II self-peptide complex retain CD4 expression. CD4 or CD8 single-positive (SP) cells move to the medulla to undergo negative selection[5]. Also known as central tolerance, this process eliminates T cells strongly reactive to self-peptides. There, thymocytes are presented with self-antigens on medullary TECs (mTECs). The majority of thymocytes interact too strongly and undergo apoptosis. Those with a moderate interaction survive and exit the thymus circulating in the periphery. Some T cells also progress to become natural Tregs (nTregs)[6]. With age, the thymus atrophies. Maintenance of the naïve T cell repertoire relies heavily on the long life spans (6-9 years) of these cells and the homeostatic peripheral proliferation promoted by Interleukin (IL)-7 and self-peptide MHC complexes[7–9].
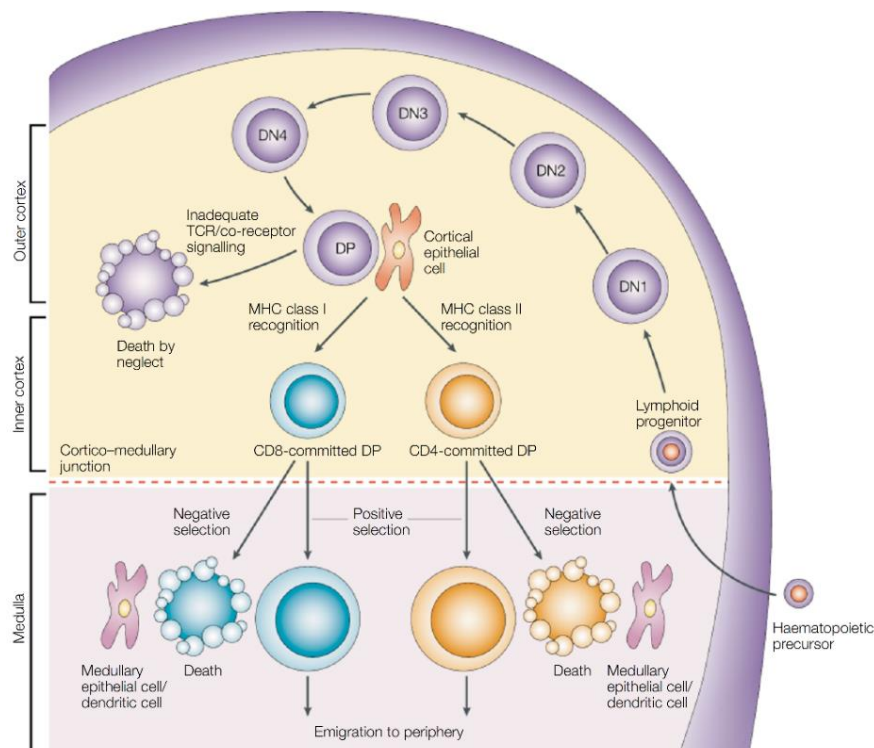
Figure 1.1| Thymic T cell development. Lymphoid progenitors migrate to the thymus. Early DN thymocytes lack TCR, CD4 and CD8 expression. At stages DN2 to DN4, they express a β TCR chain and a pre-α chain. Successful pre-TCR expression causes proliferation during the DN4/DP transition and rearrangement of the α TCR chain. αβ TCR DP thymocytes interact with cTECs expressing self-peptides. Intermediate TCR signalling induces MHC class commitment. Negative selection takes place in the medulla. Thymocytes are presented with self-antigens and the majority interact too strongly and undergo apoptosis. Those with a moderate interaction exit the thymus[10].

1.1.3. T cells: priming, activation and effector functions

After leaving the thymus, naïve T cells circulate between secondary lymphoid organs and the blood[2]. In secondary lymphoid organs, T cell priming occurs, where naïve T cells encounter, for the first time, their cognate antigen presented by DCs, resulting in IL-2 production, proliferation, and differentiation into different effector T cell subtypes, which then migrate to different tissues for local antigen patrol. In addition to TCR stimulation, co-stimulatory signals such as B7-CD28 interaction are crucial for T cell priming and activation. Other co-stimulatory molecules, which are upregulated during T cell activation, include CD40, OX40 and 41BB molecules. Engagement of the TCR and epitope-MHC complex in absence of costimulation leads to T cell anergy. The TCR signal is enhanced by the binding of the CD8 and CD4 T cell coreceptors to the MHC class I and II molecules, respectively[11].

Following TCR activation and depending on the cytokine milieu secreted by DCs, naïve CD4$^+$ Th cells differentiate into distinct Th subsets: Th1, Th2, Tregs and Th17 cells **(Figure 1.2)**[12]. CD4 Th cells secrete cytokines that act on other immune cells. Th1 cells secrete IFNγ and activate a cell-mediated immunity such as macrophages and CTLs against intracellular pathogens. IFNγ also enhances antigen presentation by MHC molecules. Alternatively, Th2 cells interact with antigen-primed B cells via MHC class II complexes in the germinal centres within secondary lymphoid organs and secrete IL-4 and IL-13 inducing immunoglobulin (Ig) class switching and B cell differentiation into antibody-producing plasma B cells. Th2 cells also enhance eosinophil, basophil and mast cell activation. Th1 cells can also lead to B cell responses. nTregs, generated in the presence of transforming growth factor-β (TGF-β) and IL-2, suppress immune cell functions by secreting immunosuppressive cytokines (e.g. IL-10) and immune cell checkpoint receptors (e.g. cytotoxic T cells antigen-4 (CTLA-4))[13].

On the other hand, naïve CD8$^+$ T cell priming in secondary lymphoid organs generates CTLs capable of directly killing pathogen-infected cells **(Figure 1.2)**[14]. CTLs recognize peptides bound to MHC-I molecules. In addition to producing cytokines such as IFNγ, mature CD8$^+$ T cells induce cell death by releasing cytotoxins such as perforin, granzymes, and granulysin into the target cell. CTLs can also induce apoptosis via Fas-Fas ligand cell-surface interaction. Some T and B cells remain in the nonlymphoid tissues as resident memory cells that guard against pathogen reinfection. Memory precursor effector CD8$^+$ T cells, which are thought to have received less stimulation, contribute to the memory population[15,16].
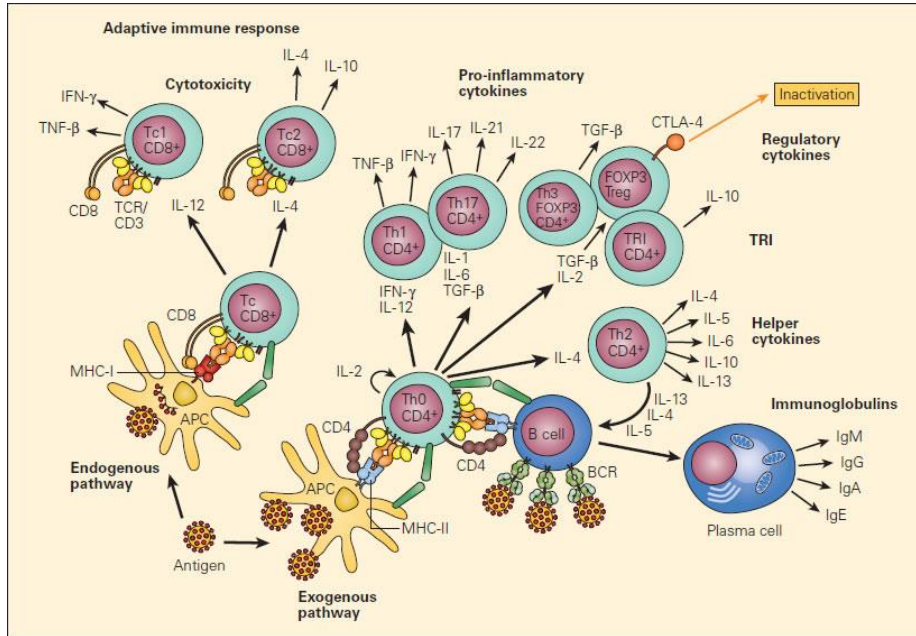
Figure 1.2| T cell priming and differentiation. T cell priming occurs in secondary lymphoid organs. Naïve T cells encounter antigens presented by DCs, resulting in their differentiation into effector cells. Depending on the cytokine milieu, naïve CD4⁺ cells differentiate into distinct Th subsets

that secrete cytokines acting on other immune cells. Naïve CD8⁺ T cell priming generates Tc1 and Tc2 CTLs capable of killing pathogen-infected or tumor cells as well as extracellular pathogens and partaking in allergy, respectively.

1.2. The T-cell receptor

1.2.1. TCR Loci and somatic TCR gene rearrangement

TCRs are clone-specific, transmembrane heterodimeric polypeptide chains consisting of either an alpha (α) and beta (β) chain or gamma (γ) and delta (δ) chains linked by a disulphide bond[17]. TCR genes are present in the germline genomic DNA. In mice, the TCRα and TCRδ genes are located on chromosome 14 **(Figure 1.3)** and the TCRβ and TCRγ genes on chromosome 6 [18]. In humans, the TCRα and TCRδ genes are also located on chromosome 14 but the TCRβ and TCRγ genes are located on chromosome 7.

The TCRα locus is composed of variable (Vα) and joining (Jα) genes and a single constant (Cα) gene[19]. The TCRβ locus contains diversity (Dβ) genes in addition to Vβ and Jβ genes as well as two C genes (Cβ1 and 2). In mice, the TCRα genes consist of 73 functional Vα genes, and 38 Jα genes upstream of a Cα gene. A duplication of 40 Vα genes spanning more than 400 kb characterizes the locus. Furthermore, in the TCRα locus of the commonly used Mus musculus C57BL/6J mouse strain, an insertion of 300 kb corresponds to a triplication of 34 out of the duplicated 40 Vα genes. The TCRα locus is interrupted between the Jα and Vα genes by the TCRδ locus. Compared to the TCRα locus, the TCRβ locus contains 22 functional Vβ genes. The Vβ genes are located upstream of two clusters each cluster containing a Dβ gene (D1 or D2) upstream of 7 Jβ

genes and a single Cβ gene (C1 or C2) each. Each α and β V gene is preceded by two exons encoding the leader sequences: L-PART1 and L-PART2 required for translation. The Cα and Cβ genes are encoded by three and four exons, respectively. In contrast, in humans, there are 47 Vα and 50 Jα, 48 Vβ, 2 Dβ, 13 Jβ, functional genes[20].
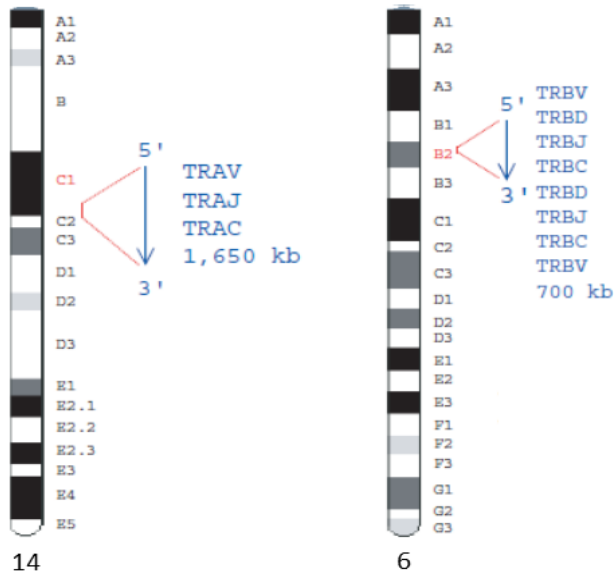
Figure 1.3| TCR gene loci. The mouse TCRα locus is located on chromosome 14 and spans 1,650 kb. It comprises 98 Vα genes (73 functional and 25 non-functional genes) upstream of 60 TRAJ genes (38 functional and 22 non-functional genes) and of a single Cα gene. The mouse TCRβ locus is located on chromosome 6 B2 and spans 700 kb. It comprises 35 Vβ genes (22 functional and 13 non-functional genes). Except Gene *TRBV31*, which is downstream of the Cβ2 gene in an inverted orientation, all other Vβ genes are located upstream of a duplicated cluster of one Dβ, seven Jβ and a single Cβ gene.

The TCR genes are somatically rearranged through V-(D)-J-C recombination during thymic ontogeny[2]. Recombination of TCRδ, TCRγ and TCRβ occurs in the DN2 and DN3 stages of thymocyte development upon recombinase expression. Successful recombination of TCRδ and TCRγ promotes a γδTCR[21]. Otherwise, the cell proceeds to rearrange the TCRβ locus. Successful TCRβ chain recombination promotes assembly of the TCRβ chain with a pre-TCRα chain to form a pre-TCR. Pre-TCR signals downregulate recombinase expression, induce several rounds of proliferation and the differentiation to DP cells where the recombinase genes are re-expressed allowing TCRα gene rearrangement. Different TCRα rearrangement in the proliferated DP thymocytes originating from the same DN thymocyte (same TCRβ chain) can generate T cells possessing the same TCRβ but different TCRα chains[2,22].

TCR gene recombination is mediated by Recombinases (RAG1 and RAG2) that insert double-stranded breaks at recombination signal sequences (RSSs) that flank TCR gens[21,23]. These breaks are resolved by non-homologous end joining. The RSS comprises two conserved sequences, a heptamer separated by either 12 (5' of D and J gens) or 23 (3′ of V gene) random nucleotides from a consensus nonamer. Recombination takes place between genes with different spacers. This "12–23" rule means that V genes can only recombine with D or J genes. During rearrangement, random nucleotides are added in the

junctions between the V-(D), V-J and J-C J genes by a terminal deoxynucleotidyl transferase (TdT)[24].

For the β chain, the Dβ gene is joined to a Jβ gene. The Vβ gene is then joined to the DJ gene at the DNA level. After transcription, the VDJ exon is spliced to the respective Cβ gene **(Figure 1.4)**. For the TCRα chain, the Vα gene randomly rearranges to a Jα gene at the DNA level[2]. After transcription, the primary transcript is spliced to the Cα gene generating the mature V-J-C mRNA. The mature mRNA of both chains are translated and the leader peptides are cleaved off following entry into the endoplasmic reticulum (ER) producing a mature TCR chain.
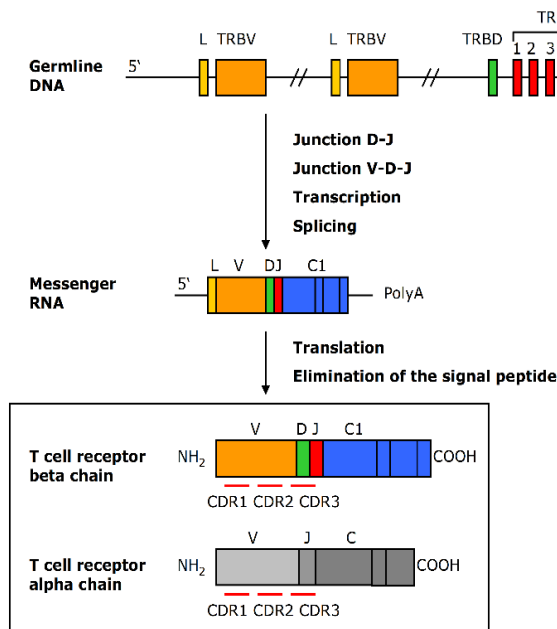
Figure 1.4| TCR ß chain biosynthesis. A Dß gene rearranges with a Jß gene, then with a Vß gene. Transcription and splicing of the V-D-J exon to a Cß generates the mRNA, which is translated into a TCR ß chain. The Vα and J α genes are rearranged and the functional VJ-region exon is transcribed and spliced to the Cα gene. The mRNA is translated into an α TCR chain. Both TCR chains pair into a heterodimer. L = Leader; Ex = Exon.

TCR allelic exclusion is the expression of TCR chains from a single allelic copy[25]. Allelic exclusion is regulated differently for the α and β TCR chains[26]. TCRβ gene arrangement occurs on one chromosome and continues on the other if the first attempt generated a non-productive gene rearrangement[27]. Monoallelic initiation of TCRβ gene rearrangement is regulated by nuclear localization and histone modifications. The expression of a functional TCRβ chain (together with the pre-TCRα chain) enforces allelic exclusion by inhibiting further recombination on the other TCRβ allele via a feedback inhibition signal reducing RAG activity or silencing germline V genes. In contrast, TCRα rearrangement occurs simultaneously on both chromosomes and is not ceased by productive TCRα gene formation. Therefore, T cells can express two TCRα chains[28]. If one allele generates an in-frame gene, then that is expressed on the surface. If both allele products are out-of-frame, the thymocyte dies via apoptosis. If both allele rearrangements generate in-frame genes, only one TCRα chain is expressed in the TCR heterodimer. This is regulated by

post-transcriptional mechanisms (phenotypic allelic exclusion) including competition between the TCRα chains for the TCRβ chain[27,29].

1.2.2. TCR repertoire diversity and dynamics

TCR diversity is not only a consequence of combinatorial diversity resulting from V-(D)-J rearrangement, but also of junctional diversity due to nucleotide insertion and deletion events at the junctional sites through TdT activity and error-prone non-homologous end joining[2]. Diversity is further augmented through α-β TCR chain pairing. Unlike immunoglobulins (Ig), TCRs do not undergo somatic hypermutation. Evidence however suggests that TCR chains can be re-edited in the periphery by RAG re-activation in a process known as TCR revision[30]. The total theoretical αβTCR diversity is thought to exceed 10^{18} in humans and 10^{15} in mice[31]. Due to thymic selection, the sum of unique T cell clonotypes in humans and mice is predicted to be greater than 10^7 and 10^6, respectively[32,33].

Despite such diversity, public T cell clones (TCR chains shared amongst individuals) have been observed in humans, mice and other species[34]. The reasons are poorly understood but proposed mechanisms include: 1) Convergent evolution where thymic selection favours the same TCRs subsets. 2) Convergent recombination where differently rearranged genes and nucleotide insertion/deletion converge producing the same TCR amino acid (aa) sequence. 3) Recombinatorial bias including biases in genes usage and nucleotide addition/deletion.

In addition to its diversity and specificity against countless antigens, the TCR repertoire is highly dynamic and changes rapidly due to the clonal expansion of antigen-specific T cells in response to infection, autoimmunity and malignancy followed by regression upon antigen clearance[35].

1.2.3. TCR surface structure and TCR-peptide: MHC interaction

Unlike BCRs, TCRs only recognize foreign antigens in the form of short peptides (epitopes) presented in the peptide-binding groove of MHC molecules. Optimal MHC class I peptides are usually nine aa's long with defined residues at particular sites for MHC docking[36]. MHC class II peptides vary from 11 to 30 aa's since the peptide-binding groove is open at both ends[37].

The α and β TCR chains consist of approximately 240 and 280 aa's, respectively. Each human T cell bears around $3x10^4$ TCR molecules of one specificity on its surface[2]. The number of surface TCR molecules for murine T cells is not known. In 1984, T. Mak and M. Davis first discovered and cloned the human and mouse TCRs, respectively[38,39]. α and

β TCR chains each consist of a variable (V) amino-terminal region and a constant (C) region forming antiparallel β-sheets **(Figure 1.5A)**[40]. Both extracellular portions fold into two Ig-like domains. The juxtaposition of both V domains forms the antigen-binding site[2]. The constant region is followed by a hydrophobic transmembrane region and a short intracellular cytoplasmic tail. The C domain of both TCR chains consists of cysteine residues that form the inter-chain disulphide bond in the hinge domain. Unlike the BCR, which binds Fc receptors, the TCR is docked in membrane.

The TCR chains have no intrinsic intracellular signalling capability and signal by noncovalently coupling with a CD3 complex after translation **(Figure 1.5B)**. The complex is composed of the CD3δε, CD3γε, and CD3ζζ modules in a 1:1:1:1 stoichiometry. These dimers contain the immunoreceptor tyrosine-based activation motifs (ITAMs) required for TCR downstream signalling. The CD3γ, δ and ε subunits each consist of a single extracellular Ig domain and an ITAM, whereas the CD3ζ has a short extracellular domain and three ITAMs[17].
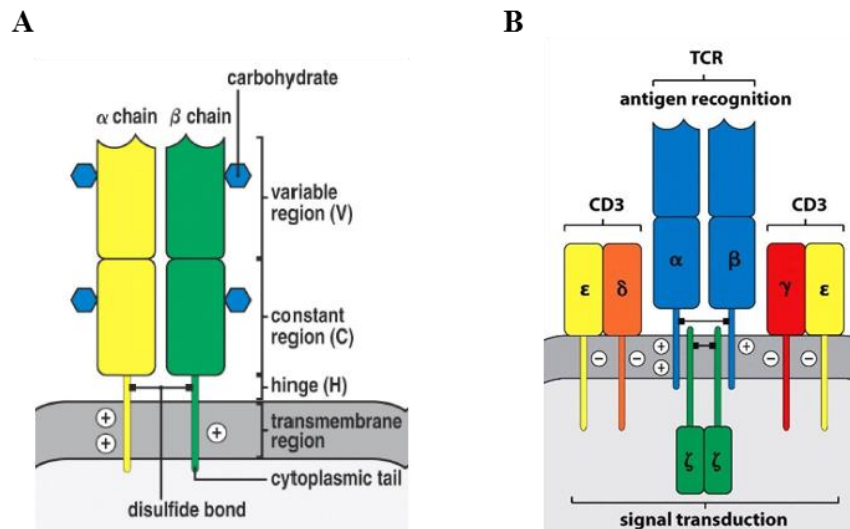
Figure 1.5| Structure of the TCR-CD3 complex. **(A)** The TCR heterodimer is composed of a trans-membrane α and a β glycoprotein chain. The extracellular region of each chain consists of a variable (V) and constant (C) domain. The juxtaposition of both V domains forms the site for antigen recognition. A short hinge domain connects the chains via an inter-chain disulphide bond. The trans-membrane helices of both chains contain positively charged residues. **(B)** The CD3 T cell co-receptor complex provides T cells with intracellular signalling abilities. The CD3 complex contains four distinct chains: CD3γ, CD3δ chain, two CD3ε chains. The TCR, ζ-chains, and CD3 molecules together constitute the TCR complex. The CD3 chains contain a single extracellular immunoglobulin domain and the transmembrane region of the CD3 chains is negatively charged allowing them to interact with the positively charged TCR chains[2].

The V domains of the α and β TCR chains have three complementarity determining regions (CDRs) each, generating six loops in total that determine the TCR specificity **(Figure 1.6A)**. The CDR1 and 2 loops are germline derived and constant for each gene. The CDR3 loop is the most variable resulting from V-D-J rearrangement and the main one responsible for interacting with the antigen peptide **(Figure 1.6B)**. The α-chain CDR1

8

interacts with the antigenic peptide N-terminal, while the β-chain CDR1 interacts with the peptide C-terminal. The CDR2 recognizes the MHC molecule[2,41].
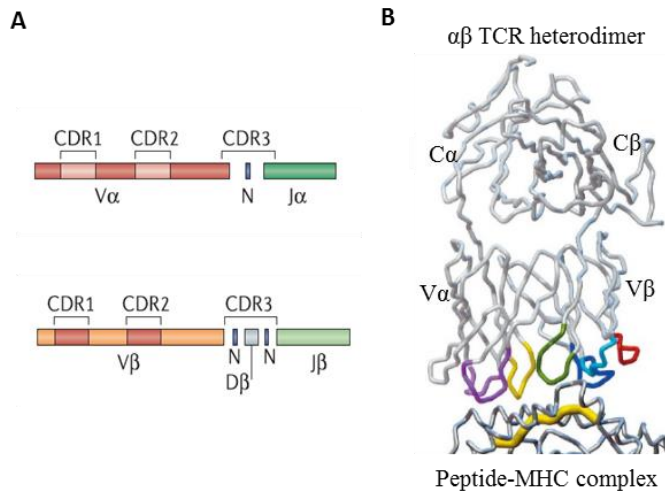
Figure 1.6| TCR CDR loops. (A) The CDR regions of the α (top) and β (bottom) TCR chains are shown. The CDR3 region of each chain shows the highest variability due to combinatorial and junctional diversity. **(B)** The αβ TCR heterodimer interacts with the top of an MHC molecule. The presented eight-mer peptide is depicted in yellow. The CDR regions of the TCR are indicated in colour: CDR1 and 2 of the α chain in light and dark purple, respectively, the CDR1 and 2 of the ß chain in light and dark blue, respectively. The CDR3 loop of the α chain is marked in yellow, the CDR3 loop of the ß chain in green. The additional ß chain hypervariable loop 4 (HV4) marked in red, is involved in the recognition of viral superantigens.

High on/off-rate is characteristic for TCR-peptide-MHC interaction. TCRs have a high degree of antigen specificity, despite weak binding to the peptide/MHC ligand (affinity dissociation constant 1-100 μM)[42]. Low affinity allows degeneracy (many TCRs recognize the same antigen and multiple antigens are recognized by the same TCR). Antigen-experienced T cells have higher sensitivity because they depend less on costimulatory signals[43]. TCR signalling is thought to be regulated by TCR microcluster formation increasing T cell sensitivity via an avidity-based mechanism[44]. Immunological synapses (Supramolecular activation cluster-SMAC) are formed between the T cell and APC composed of central, peripheral and distal regions in which TCR/MHC, adhesion and costimulatory molecules exist, respectively. TCR and CD4/8 interaction with the peptide: MHC complex initiates a cascade of protein recruitment and phosphorylation steps leading to the activation of multiple signalling pathways such as the NF-κB pathway (Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1) that induce transcriptional changes, cytokine production and effector functions[45].

1.3. Cancer and T-cell receptor therapy

1.3.1. Cancer pathophysiology and existing therapies

Cancer is a group of diseases where acquired or inherited genetic mutations and epigenetic alterations lead to the uncontrolled proliferation of cells some of which can then

metastasise to distant sites. According to the World Health Organization, cancer is the leading cause of death before age 70 years in most countries. The global burden was estimated at 18 million new cases and 9.6 million deaths in 2018[4746,47]. Lung cancer is the most common cancer (11.6%) and primary cause of cancer-related death (18.4%). This is followed by colorectal (9.2%), liver (8.2%) and stomach cancer (8.2%) for mortality.

90% of cancer cases occur due to genetic mutations acquired from environmental factors and 10% owing to genetic predispositions[48]. Environmental factors include physical carcinogens such as ultra-violet light that induces DNA thymine dimers. Chemical carcinogens such as Ethyl Methanesulfonate in tobacco can form DNA-adducts. Biological carcinogens such as Human papillomavirus (HPV) and Helicobacter pylori cause neoplastic transformation through oncoviral proteins and by inducing inflammatory environments.

For neoplastic transformation, multiple non-synonymous mutations must occur in cell division, differentiation and apoptosis related genes[49]. Gain-of-function mutations in oncogenes (e.g. Ras GTPase) or over-expression can lead to increased cell division. Loss-of-function mutations in tumor-suppressor or 'gatekeeper' genes (e.g. p53) permit excessive replication or inhibit apoptosis. Furthermore, caretaker gene mutations (e.g. DNA mismatch repair protein Msh2) lead to genomic instability and mutation accumulation[50].

The large armamentarium of conventional anti-cancer therapeutics suffers from low efficacy and life-threatening adverse effects. With the aim of removing or debulking tumors, surgical resection remains the mainstay of treatment in solid tumors and is typically combined with chemo or radio-(neo)adjuvant therapy[51]. Surgery is ineffective in metastatic cancers. Chemotherapy and radiotherapy show efficacy in a small subset of patients and can adversely affect healthy proliferating cells. Moreover, multidrug resistance is a common cause of low chemotherapy efficacy[52]. Efforts to circumvent off-target cytotoxicity laid the foundation for targeted therapy using small molecules such as Tyrosine Kinase inhibitors[53]. Despite reduced toxicity, intra-tumoral spatiotemporal genomic heterogeneity caused by mutation accumulation remains a major cause of patient relapse[54]. Furthermore, inter-patient tumor heterogeneity plays a big role in the ineffectiveness of these therapies. These limitations, together with the discovery of cancer immune surveillance paved the way for cancer immunotherapy.

1.3.2. T cells in anti-tumor immunity

In 1909, P. Ehrlich hypothesized the immune system's ability to prevent neoplastic cells from developing into tumors[55]. In 1943, L. Gross demonstrated the immunogenicity of sarcomas through intradermal immunization of mice[56]. After, M. Burnet and L. Thomas

proposed the cancer immunosurveillance theory postulating that lymphocytes recognize transformed cells through the recognition of tumor antigens (TAs)[57]. DNA cloning and CTL reactivity determination approaches, MHC-binding prediction algorithms and peptide elution from tumor derived MHC molecules have now generated large libraries of well-studied TAs[58–60].

TAs are classified based on their expression pattern. Tumor specific antigens (TSA) include *oncoviral antigens*. Viral proteins (e.g. HPV E6, E7 or hepatitis B virus (HBV) protein s produced inside the infected tumor cells are endogenously processed and presented on MHC I molecules eliciting a CTL response[61]. Other TSA's include *neoantigens* arising from genetic mutations. Insertions or deletion (Indels) and non-synonymous mutations can generate immunogenic neoepitopes perceived as foreign by the immune system[62]. Advances in next-generation sequencing (NGS) and *in silico* epitope prediction algorithms have facilitated the identification of patient-specific neoantigens[63]. Other TSAs include *cancer/testis (C/T)antigens* expressed in gametes and trophoblasts (no antigen presentation) in addition to many cancers but not in normal somatic tissue (e.g. Melanoma-antigen encoding (MAGE) genes and Kita-kyushu lung cancer antigen 1 (KKLC1))[64,65]. Other antigens include tumor-associated antigens (TAAs) such *Tissue differentiation antigens* including Prostate-specific antigen (PSA) in prostate carcinomas and CD20 in B cell lymphomas. These are expressed on cells of the tumor origin and not in other tissues. Other less specific TAs include overexpressed proteins (e.g. Epidermal growth factor receptor 2 (HER2/neu) in breast cancer)[66].

Multiple immune mechanisms, mainly mediated by tumor-infiltrating T lymphocytes (TILs), control malignant tumor development. Tumor-associated DCs endocytose dead neoplastic cells and present TAs in the draining lymph nodes for T cell priming. CD8[+] CTLs recognize and directly kill neoplastic cells expressing TAs on HLA molecules[67]. Correlations between tumor control and CD8[+] T cell infiltration rates have been well established[68]. CD4[+] Th cells can mediate anti-tumor humoral immune responses and help prime CD8[+] T cells[69]. CD4[+] T cells with cytotoxic functions have also been described in several cancer types[70]. IFNγ secretion by CD8[+] T cells and CD4[+] T cells can also induce M1 macrophage-mediated tumor destruction. Furthermore, natural killer (NK) cells recognize cancer cells that lost their MHC expression[71].

Due to selective pressure, the high mitotic rate of tumor cells leads to the outgrowth of cells with reduced immunogenicity. This immunoediting can occur through TA downregulation as well as reduced MHC surface expression[72]. Tumor cells also establish immunosuppressive environments by releasing TGF-β and recruiting Tregs and myeloid-derived suppressor cells (MDSC). Tumors can also express immune checkpoint molecules such as (CD80/86) and programmed cell death 1 ligand 1 (PD-L1). T cell exhaustion is also a major cause of the immune system's failure to eliminate tumor cells.

Despite the ability of tumors to escape immunosurveillance, the potential of the immune system to eliminate tumors with remarkable specificity and long-lasting memory has motivated researchers to modify the patient's immune system to target tumor cells[73].

1.3.3. Cancer immunotherapy and T cell-based therapies: opportunities and challenges

In 1891, W. Coley harnessed the immune system to treat cancer patients by injecting bacterial strains into patient tumors achieving surprising responses in sarcoma, lymphoma, and testicular carcinoma[74]. Since then, exploitation of tumor-specific antibodies and cellular immune effector mechanisms has culminated in the clinical approval of various immunotherapeutic modalities[73].

IL-2 administration to expand TA-specific T cells was the first Food and Drug administration (FDA) approved immunotherapy capable of mediating complete, durable responses in renal cancer and melanoma patients[75]. Currently many more unspecific immunomodulatory modalities have been approved (e.g. IFN-α2b)[76].

Following the first monoclonal antibody (mAB) approval for Non-Hodgkin's lymphoma (Rituximab; anti-CD20) in 1997, many more have been approved. In addition to inducing antibody dependent cell mediated cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC), other mABs inhibit signalling pathways (e.g. Cetuximab for head and neck cancer) or act as antibody-drug conjugates[77,78]. Cancer cells have the ability to activate immunosuppressive immune checkpoint pathways. The development of mABs that target such immune checkpoints was a cardinal milestone in immunooncology[79]. Following the approval of Ipilimumab (anti-CTLA-4 receptor), a plethora of new immune checkpoint inhibitors (ICIs) entered clinical pipelines. Pembrolizumab and Nivolumab (anti-PD-1), showed an objective response rate (ORR) of 40–45% in melanoma and non-small cell lung carcinoma (NSCLC) patients. In Hodgkin's lymphoma, Nivolumab showed an ORR of 87% and 17% complete response. Despite such successes, ICI shows efficacy in only the fraction of patients with a pre-existing anti-tumor immune response. As a result, increasing efforts are given to combining ICIs with other immunization approaches that can generate *de novo* anti-tumor immune responses[80,81].

Active immunization approaches using TA-encoding therapeutic vaccines have been developed to generate *de novo* T cell responses. Such vaccines include peptide, DNA or RNA-based vaccines and pulsed DCs or allogenic whole cell vaccines[82]. First-generation cancer vaccines were based on non-mutant patient-shared antigens (e.g. MART-1, gp100). Although immunogenic, many induced responses in a minority of late-stage cancer patients due to central tolerance, T cell exhaustion, and immunosuppressive tumor microenvironments[83]. Advances in mutation identification and vaccine production

enabled the development of patient-individualized vaccines using mutation-encoding RNA-based vaccines or long peptides[63,84,85]. Several clinical trials showed clinical benefits associated with elevated neoantigen-specific T cell responses[86]. Other vaccination approaches use oncolytic viruses e.g. T-VEC (Herpes simplex-1 virus (HSV1) for advanced melanoma[87].

Despite TA-specific T cell enrichment using vaccination approaches, tumors can progress owing to immunosuppressive tumor microenvironments and T cell exhaustion; encouraging the development of adoptive T cell therapy (ACT)[88]. These involve the removal of autologous T cells from immunosuppressive environments, expanding and selecting them *ex vivo* and re-infusing them back in large numbers **(Figure 1.7)**. The first effective ACT immunotherapy revealed that the IL-2-driven *ex vivo* expansion of TILs from metastatic melanoma patients and their re-infusion with high-dose IL-2 mediated an overall objective response rate of 34%[89]. The effectiveness of this approach was dramatically improved with preparative lymphodepletion using chemotherapy (e.g. cyclophosphamide) alone or together with irradiation[75]. This enhanced T cell persistence by eliminating Tregs and other lymphocytes that competed for homeostatic cytokines. Improvements in culturing, expansion and selection protocols further improved outcomes. Follow-up studies showed that autologous TIL transfer could mediate curative responses in refractory metastatic melanoma patients. The advantages of this approach are that only a small number of anti-tumour cells must be identified and expanded and the exact populations and effector functions can be selected *ex vivo*.
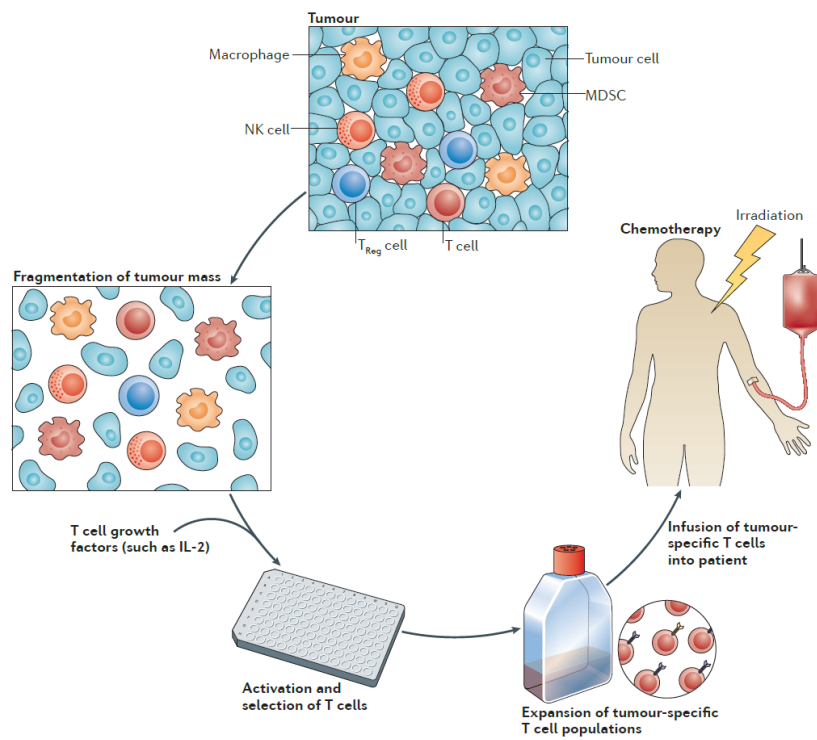
Figure 1.7| Adoptive TIL therapy. Tumour masses can be surgically resected and fragmented. The cells can be cultured and the TILs expanded using IL-2. T cell populations with the desired antigen-specificity can be selected and autologously transferred back into the patients together with IL-2 infusion. Prior to adoptive transfer, patients are immunodepleted using chemotherapy alone or together with total-body irradiation. MDSC, myeloid-derived suppressor cell; NK, natural killer cell; T_{Reg}, Regulatory T cell.

Despite some success, TIL-based expansion therapies rely on the patient's pre-existing immunity and the availability of resectable material with infiltrated T cells that can be expanded *ex vivo*. The difficulty in identifying T cells with high anti-tumour avidity in many patients led to the development of strategies that genetically engineer autologous PBLs with unique TA specificities **(Figure 1.8, bottom)**. One approach involves adding "chimeric antigen receptors" (CAR; a single chain antibody fused to intracellular immunostimulatory domains)[90]. CAR-engineered T cells can recognize cell surface TAs in an MHC-independent manner. To date, two anti-CD19 CAR-based therapies have been FDA approved (Kymriah and Yescarta) showing up to 92% full recovery in in end-stage B-cell lymphoma patients. Despite such impressive results in haematological malignancies, the use of CAR T cells for solid tumors has been disappointing so far due to T cell persistence.
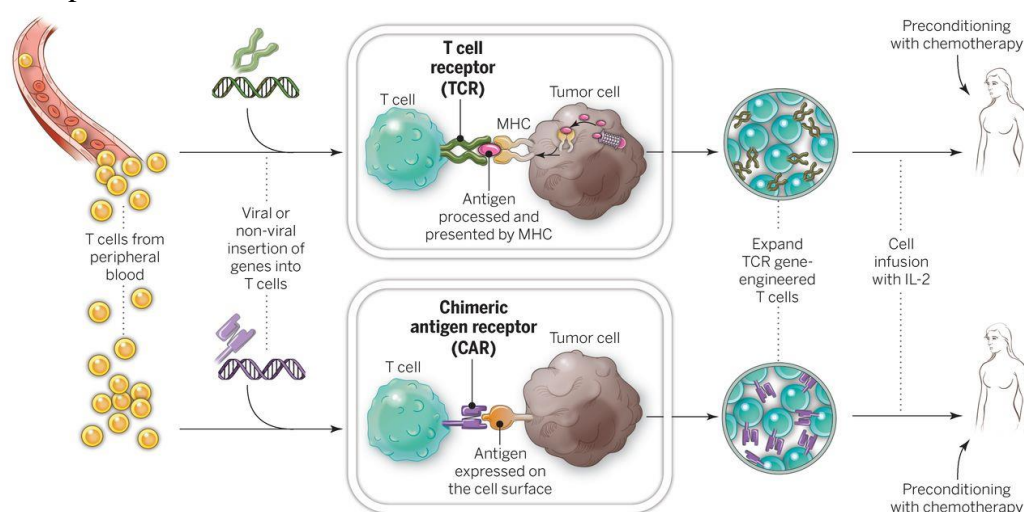
Figure 1.8| Gene-modification of peripheral blood lymphocytes. The top panel depicts the insertion of natural αβ TCRs into patient T cells, followed by expansion and re-infusion. The bottom panel shows CAR insertion into patient T cell, followed by expansion and re-infusion. Re-infusion is preceded by immunodepletion in both approaches.

Another promising approach involves administrating patients with autologous PBLs after retroviral transduction with genes encoding conventional TA-specific αβTCRs **(Figure 1.8, top)** [91,92]. High-affinity TCRs can be isolated through antigen-specific *in vitro* priming of peripheral blood lymphocytes (PBLs) from healthy donors, discovery of autologous neoantigen-reactive T cells that escaped negative selection in the thymus or by immunizing human leukocyte antigen (HLA) transgenic mice against TAs thereby avoiding central tolerance[64]. TCRs can be further mutagenized to increase their TA-affinity[93]. Mispairing of the engineered chains with endogenous TCR chains can be avoided using murine constant genes or by inserting cysteine residues favouring engineered chain pairing[94]. The first successful application used a Melanoma

differentiation antigen (MART-1)-specific TCR mediating objective regressions in 13% of treated patients with metastatic melanoma including liver and lung metastases regression[95]. Subsequent trials using improved ACT protocols and human or murine TCRs targeting other TAs including melanoma-associated antigen 3 (MAGE-A3), glycoprotein 100 (gp100), and New York esophageal squamous cell carcinoma-1 (NYESO-1) demonstrated prolonged regression in melanoma and sarcoma patients[96–99]. Some patients did however experience on-target off-site toxicities due to the expression of these shared TAs in other tissues and due to affinity maturation of the used TCRs that later showed unpredicted reactivity to self-antigens.

Unlike targeting shared TAs, targeting tumor-specific neoantigens can overcome such toxicities **(Figure 1.9)**. Increasing evidence discloses neoepitopes as the primary targets of TILs[100–106]. Targeting mutations directly using *ex vivo*-selected neoantigen-specific TILs mediates complete durable regression in patients with metastatic breast cancer, metastatic colorectal cancer and metastatic cholangiocarcinoma[107]. As a result, engineering mutation-specific TCR genes into T cells for ACT holds the promise of a truly tumor-specific and effective therapy. For such an approach, appropriate neoantigens as well as potent TCRs must be identified in a patient-specific manner. Furthermore, the dynamic nature and intra-tumoral heterogeneity of the mutational landscape emphasize the importance of targeting multiple neoantigens, primarily driver mutations, to avoid tumor resistance[108].
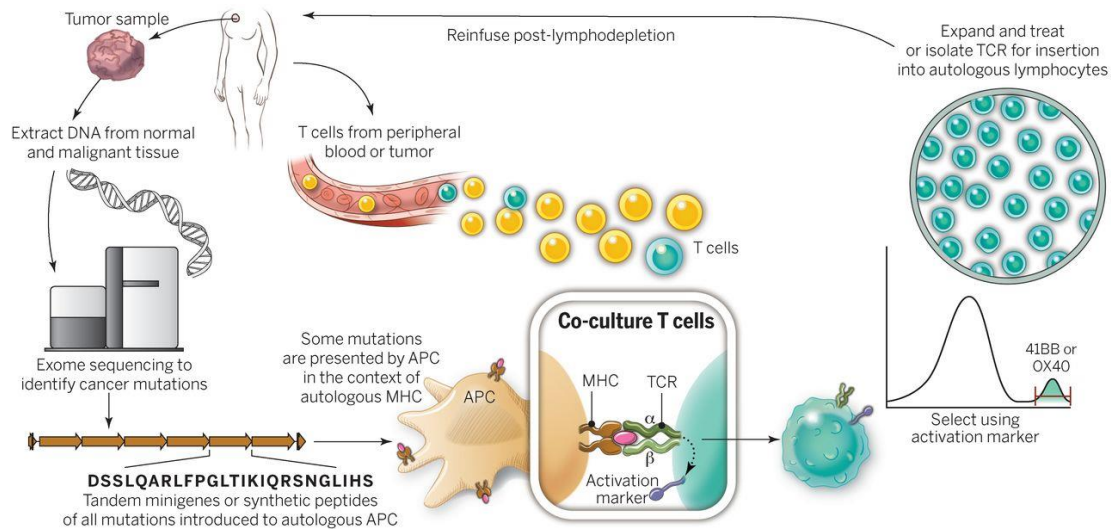
Figure 1.9| Treatment of patients with neoantigen-specific TCRs. Exome DNA from tumor and normal cells from a patient are compared to identify tumor mutations. Minigenes or polypeptides encoding the chosen mutations can be expressed on autologous APCs co-cultured with autologous T cells. Neoantigen-specific T cells be isolated by selecting activated T cells (e.g. 41BB[high] and OX40[high]) using flow cytometry. Neoantigen-specific T cells are expanded and given back to the patient or TCRs can be isolated and cloned to be used for genetic engineering of autologous T cells before reinfusion into the tumor-bearing patient.

1.3.4.   Pre-clinical studies to improve ACT

Further improvements to ACT therapy depend greatly on a better understanding of the complex immunological processes behind tumor (neo)-antigen-specific T cell activity. Our understanding so far has been greatly influenced by ACT studies in mice. Recent findings include the essential contribution of tumor-specific CD4[+] T cells to effective antitumor responses or that central memory cells are more effective than highly differentiated cells for ACT[85,109]. However, more mode of action studies are required to further improve *in vivo* T cell persistence, efficient trafficking into tumors and reversing immunosuppressive tumor environments. The expansion of transferred T cells *in vivo* through combination with antigen-specific vaccination approaches is also a field of interest to augment as well as prolong anti-tumor activity of the infused T cells. Identifying the minimum number of mutations that must be targeted for an effective response to avoid immune escape is also essential for effective neoantigen-specific ACT. A more detailed understanding of such questions through pre-clinical *studies* will likely be translated into the clinic for improved ACT protocols. However, the lack of available shared tumor antigen and neoantigen-specific TCRs caused by the difficulty to efficiently identify their sequence remains a considerable obstacle to pre-clinical investigations and subsequent clinical application of ACT therapy[110].

1.4. T-cell receptor sequencing and cancer

Identifying TCR sequences from TILs, healthy tissue-infiltrating T cells and PBLs in humans and mice is highly beneficial for monitoring the immune system since the TCR is unique to each T cell clone[111]. In addition to identifying TA-specific TCRs for pre-clinical and clinical use, TCR sequencing also provides valuable information for studying the immunostatus and immunodynamics[112].

The first efforts to dissect the TCR repertoires of a population of T cells used V-gene specific antibody staining and flow cytometry[113]. In addition to the limited availability of Vβ-gene–specific antibodies, such an approach lacked information regarding the CDR3 regions. The first CDR3 sequence-based approach, known as CDR3 Spectratyping, analysed the CDR3 sequence length distribution in a sample[114]. TCR transcripts were amplified using Vβ-gene and Cβ-gene specific primers (GSPs). The relative frequencies of the differently sized products within an individual TRBV subfamily were measured using electrophoresis. A Gaussian distribution implied a polyclonal population while a dominant fragment suggested clonal enrichment. Advances in DNA sequencing methods have led to the development of two contrasting methods to sequence TCRs: bulk TCR repertoire profiling and single cell TCR sequencing.

1.4.1. Bulk TCR repertoire profiling

NGS technologies parallelize DNA sequencing, generating millions of sequences. Many approaches have been developed, with the Illumina platform being the most commonly used. This method uses reversible deoxynucleoside triphosphate (dNTP) terminators. DNA amplicons are attached to a flow cell and amplified with dideoxy-NTPs (ddNTPs) labelled with a fluorescent dye and blocked at the 3′-OH. After one ddNTP addition, remaining nucleotides are washed and the added ddNTP is recorded. The blocking label is removed and the next sequencing cycle starts.

TCR repertoire profiling involves lysing bulk T cell samples, extracting total RNA and reverse transcribing or genomic DNA (gDNA)[115–117]. For reverse transcription (RT), 5′ rapid amplification of cDNA ends (5′-RACE) is performed using dT-priming. **(Figure 1.10)**. The template switching (TS) protocol (Switching Mechanism At 5' end of RNA Transcript-SMART) catalysed by the recombinant Moloney Murine Leukemia Virus (M-MuLV)-derived reverse transcriptase (RTase) is typically used. SMART cDNA synthesis allows for amplification of full-length TCR (and other) mRNAs irrespective of their variable 5' ends. During first-strand cDNA synthesis, the RTase's terminal transferase activity adds a few nucleotides mostly ribo-cytosines to the 3' end of the cDNA to which a second primer can bind via complementary ribo-guanines. This leads to the RTase switching templates and incorporating the sequence of the second primer into the transcript for subsequent global preamplification via PCR. This process leads to the incorporation of PCR handles on each end. The cDNA encoding all chains is collectively amplified typically using TCR constant gene-specific primers in semi-nested PCRs or using primers complementary to the PCR handles. Other approaches use gDNA as starting material and typically use multiplex V and J-GSPs for TCR amplification. This approach, however, can insert some repertoire bias due to different primer-binding abilities. Amplicons are deep sequenced using NGS. The reads are then aligned to TCR references using available software tools (e.g. MiXCR[118]) to identify the α and β TCR chains. Such approaches benefit highly from the processing of millions of cells at once. The overview of all the unique α and β TCR chains with their relative clone abundance provides a highly accurate snapshot of the TCR repertoire at a given time point. The β TCR chain has always been the focus in repertoire studies due to its greater diversity and its monoallelic expression per cell.
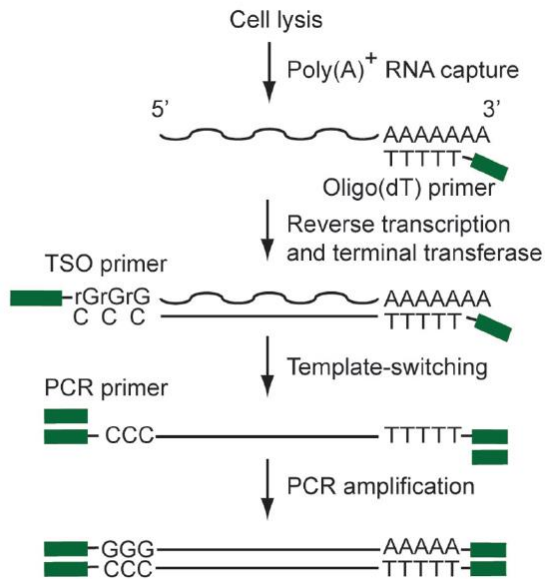
Figure 1.10| Template switching technology. Reverse transcription is primed with a primer containing a PCR handle (green box) and oligo(dT) sequence. Template switching oligonucleotide (TSO) adds a PCR handle to the other end of cDNA via complementarity with terminal cytosines (Saliba et al., 2014).

TCR repertoire profiling has been extensively applied in humans and mice to monitor T cell diversity, selection and expansion in response to cancer immunotherapy[119,120]. It is also used for prognostic purposes as well as diagnostic biomarker discovery. For example, studies have tracked the presence and abundance of TA-specific T cells in tumors and peripheral blood. TCR profiling has also been used to monitor minimal residual disease after lymphoid malignancy treatment[121]. Other studies have focused on comparing the repertoire diversity and clonal overlap between healthy and malignant tissue as well as identifying public TCR clones between patients[119,122–124]. Further work has also focused on determining intra-tumoral spatial heterogeneity of TILs[125].

The parallel processing of bulk T cell samples however leads to loss of pairing information of corresponding α and β TCR chains that constitute a TCR. Frequency-based matching (FBM) to predict the chain pairing involves matching α and β TCR chains with similar relative frequencies (e.g. matching most frequent α with most frequent β TCR chain etc.). This however is not accurate, only works for highly expanded clones and requires subsequent confirmation through other methods. Another approach, known as pairSEQ involves splitting up bulk T cells into samples and sequencing each sample using repertoire profiling[126]. Combinatorial statistics is used to predict α-β pairing. This however can only lead to the discovery of pairs from clonally expanded clones that exist in many of the sub samples and if a lot of patient material is available. The lack of αβTCR chain pairing information, which is required for further downstream functional analysis of the detected TCRs can therefore only be achieved through single cell approaches.

1.4.2. Single cell TCR sequencing technologies

Single cell (sc) TCR isolation strategies can detect the paired αβTCR chains of a single T cell. Only a few single cell approaches were described at the beginning of this PhD project.

One flow cytometry based-single cell sorting approach was first developed by Omokoko *et al.* (2014, 2015)[64,127]. Single human or murine T cells were sorted using selected T cell markers into 96-well plates (one cell per well). RNA was isolated for each cell using a silica-based RNA isolation method. First strand cDNA synthesis was performed using the TS protocol and used for subsequent global preamplification via PCR. This was followed by a Vβ-gene multiplex nested PCR. For the single cells with a successfully amplified β TCR chain, the α TCR chain was amplified using a Vα-gene multiplex nested PCR. Each of these PCRs was composed of a primer pool of V-gene leader specific primers. Amplified chains were cloned into a plasmid vectors allowing direct *in vitro* transcription (IVT) of full-length TCR chains after confirmative Sanger sequencing. They used this approach to discover and functionally validate TA-specific TCRs. Despite being robust, this approach suffers from low throughput and high costs due to the individual cloning of each TCR chain from the single cells, the use of Sanger sequencing and the need for costly Polymerases with highest processivity and fidelity.

Han *et al.* (2014) went a step further by also developing a flow cytometry sorting-based scTCR sequencing (scTCRseq) but integrating it with NGS technology[128]. Following human single T cell sorting into 96-well plates, multiplex RT-PCR was performed using 76 TCR V-GSPs and 34 phenotyping GSPs. The products of each well were used in three separate multiplex PCR reactions for α TCR chain, β TCR chain and phenotyping gene amplification. Another PCR reaction was then performed to add row and column-specific barcodes into each well allowing pooling of the products. A barcode is a short nucleotide sequence that uniquely tags cell transcripts to their well of origin allowing the tracing back of the amplicons to the cell of origin after sample pooling. The samples were then sequenced using the Illumina MiSeq platform. TCR sequences were obtained from the paired-end (PE) sequences by aligning them to a reference genome using an alignment software tool: VDJFasta. They applied the technology to characterize T cell heterogeneity in human colorectal carcinoma TILs. Despite having a higher throughput than cloning-based strategies, this approach relied on a complex PCR protocol using many primers and separate PCRs for α and β TCR chains. The late barcoding and pooling of the samples makes the technology costly and require a lot of hands-on work. Furthermore, the technology was only developed for human T cells and not murine T cells.

Despite advances in scTCRseq methods, many drawbacks still exist limiting the advancement of TCR discovery. Such disadvantages include low-throughput, high costs and complex protocols. Therefore, there is a strong demand for novel scTCRseq technologies with a simpler workflow, higher throughput and greater applicability.

1.5. Thesis outline

<u>Background to TCRs in cancer immunotherapy:</u> The large armamentarium of conventional anti-cancer therapeutic modalities suffers from low efficacy and life-threatening adverse effects. The general tumor elimination ability, remarkable specificity and long-lasting memory of TA-specific T cells has motivated researches to modify the patient's immune system to target tumor cells. Engineering shared antigen, or mutation-specific TCR genes into T cells for ACT in combination with active immunization approaches and ICI holds the promise of a truly tumor-specific and effective therapy.

Advancements are hindered by the difficulty in retrieving TA-specific TCRs. Furthermore, the paucity in suitable TCR candidates hinders pre-clinical mode of action studies and proof of concept investigations required to further understand and improve ACT protocols.

<u>Literature gap:</u> Despite advances in scTCRseq methods, many drawbacks still exist limiting the advancement of TCR discovery. Such disadvantages include low-throughput, high costs, complex protocols in addition to being limited to human T cells. There is a great need for a high-throughput, cost-efficient and easily applicable scTCRseq technology.

<u>Thesis aims:</u>

1. Development of a high-throughput and easily applicable scTCRseq platform for murine T cells.
2. Application of the developed scTCRseq platform to generate a library of shared TA and neoepitope-specific murine TCRs.
3. Incorporate T cell phenotyping into the developed scTCRseq platform to integrate TCR sequencing with the T cell phenotype.

<u>Thesis approach:</u>

ScTCRseq platform design
↓
Framework development
↓
Process optimization
↓
Process validation
↓
Performance evaluation
↓
Benchmarking
↓
Shared tumor antigen and neoepitope-specific TCR library generation
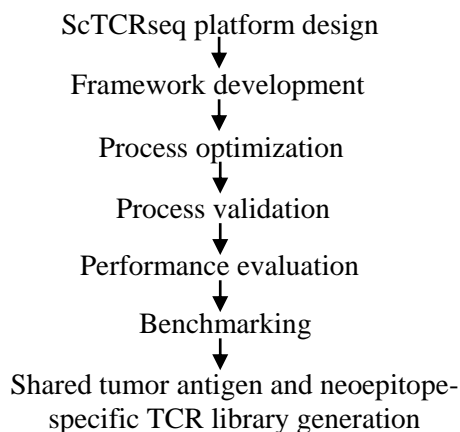
Figure 1.11| Thesis approach.

2. Materials and methods

2.1. Materials

2.1.1. Laboratory instruments

Name	Manufacturer
Agilent 2200 TapeStation	Agilent, Waldbronn, Germany
BioStep Argus X1 Transilluminator	Biostep, Burkhardtsdorf, Germany
AutoMACS Pro	Miltenyi Biotec, Bergisch Gladbach, Germany
BD FACSAria II	BD Biosciences, New Jersey, USA
BD FACSCanto II	BD Biosciences, New Jersey, USA
BioAnalyzer 2100	Agilent, Waldbronn, Germany
Biometra TAdvanced	Biometra, Göttingen, Germany
BVC Control Fluid Aspiration System	Vacuubrand, Wertheim, Germany
Centrifuge 5810/ 5810 R	Eppendorf, Hamburg, Germany
CKX31 Inverted Microscope System	Olympus, Düsseldorf, Germany
Dynamag-2 Magnet	Thermofisher, Dreieich, Germany
ECM 830 Square Wave Electroporation System	Bio-Rad, Hercules, USA
Electrophoresis Chamber	BioRad, Hercules, USA
Eppendorf Research Plus 10, 100, 1000	Eppendorf, Hamburg, Germany
Heraeus B6 Incubator	Thermofisher, Dreieich, Germany
Immunospot Series S Versa Elispot Analyzer	Immunospot, Bonn, Germany
Incu-Line	VWR, Darmstadt, Germany
Macmix Rotator	Miltenyi Biotec, Bergisch Gladbach, Germany
Miseq System	Illumina, San Diego, USA
Nanodrop	Thermofisher, Dreieich, Germany
Peqstar	Peqlab Biotechnologie, Erlangen, Germany
Powerpac High-Current Power Supply	BioRad, Hercules, USA
Qiaxcel Advanced System	Qiagen, Hilden, Germany
Qubit V3.0	Thermofisher, Dreieich, Germany
Rotational-Vacuum-Concentrator RVC	Christ, Osterode/Harz, Germany
Tecan Freedom EVO Workstation	Tecan, Wiesbaden, Germany
TW2 Water Bath	VWR, Darmstadt, Germany
Vi-CELL XR	Beckman Coulter, California , USA

2.1.2. Reagents and Antibiotics

Name	Manufacturer
3-(N-Morpholino) propanesulfonic acid (MOPS)	Sigma, St. Louis, USA
6x DNA Loading Dye	Fermentas, St. Leon-Rot, Germany
Agar	Difco, Detroit, USA
Agarose	Sigma, St. Louis, USA
Agencourt Ampure XP beads	Beckman Coulter, California , USA
Ampicillin	Sigma, St. Louis, USA
Bovine serum albumin (BSA)	Invitrogen, Karlsruhe, Germany
CD3/CD28 T-Activator Dynabeads, mouse	Thermofisher, Dreieich, Germany
CD4/ CD8 Microbeads, mouse	Miltenyi Biotec, Bergisch Gladbach, Germany
Diethylenpyrocarbonat (DEPC)	Sigma, St. Louis, USA
Dimethyl sulfoxide (DMSO)	Applichem, Darmstadt, Germany
Deoxynucleotide triphosphates mixed (dNTPs)	Biozym Scientific, Oldendorf, Germany
Ethanol 96-100%	Merck, Darmstadt, Germany
Ethidium bromide	Sigma, St. Louis, USA
Ethylendiamintetraacetat (EDTA)	Merck, Darmstadt, Germany
FACS-Clean	Becton Dickinson, Heidelberg, Germany
FACS-Flow	Becton Dickinson, Heidelberg, Germany
FACS-Rinse	Becton Dickinson, Heidelberg, Germany
Fetal calf serum (FCS)	Invitrogen, Karlsruhe, Germany
Ficoll-Hypaque	Amersham Bioscience, Little Chalfont, UK
Formaldehyde	Merck, Darmstadt, Germany
Generuler 50bp, 1 Kb DNA Ladder	Fermentas, St. Leon-Rot, Germany
Glycerin, water free	Merck, Darmstadt, Germany
HEPES Buffer (1M)	Invitrogen, Karlsruhe, Germany
IFNγ Microbeads, mouse	Miltenyi Biotec, Bergisch Gladbach, Germany
IFNγ Microbeads, human	Miltenyi Biotec, Bergisch Gladbach, Germany
Isopropanol	Merck, Darmstadt, Germany
Kanamycin	Sigma, St. Louis, USA
MEM non-essential amino acids (NEAA)	Invitrogen, Karlsruhe, Germany
Molecular biology grade water	Lonza, Basel, Switzerland
Penicillin-streptomycin	Invitrogen, Karlsruhe, Germany
Phix Sequencing Control	Illumina, San Diego, USA

Phosphate buffered saline (PBS)	Bio Whittaker Europe, Verviers, Belgium
QX 50 Bp-800 Bp Qiaxcel Ladder	Qiagen, Hilden, Germany
Recombinant IL-15, mouse	Miltenyi Biotec, Bergisch Gladbach, Germany
Recombinant IL-7, mouse	Miltenyi Biotec, Bergisch Gladbach, Germany
Recombinant IL-15, human	Miltenyi Biotec, Bergisch Gladbach, Germany
Recombinant IL-7, human	Miltenyi Biotec, Bergisch Gladbach, Germany
Riboruler High Range RNA Ladder	Thermo Fisher, Dreieich, Germany
Rpmi-1640 Glutamax-I	Invitrogen, Karlsruhe, Germany
S.O.C. Medium	Thermo Fisher, Dreieich, Germany.
Sodium acetate	Merck, Darmstadt, Germany
Sodium pyruvate	Invitrogen, Karlsruhe, Germany
SPRIselect beads	Beckman Coulter, California, USA
ß-Mercaptoethanol	Applichem, Darmstadt, Germany
Triton X-100	Sigma, St. Louis, USA
Trypan-Blue	Applichem, Darmstadt, Germany
Tween 20	Sigma, St. Louis, USA
X-Vivo 15 Medium	Lonza, Basel, Switzerland

2.1.3. Enzymes

Name	Manufacturer
BCIP/NBT	Sigma, St. Louis, USA
Calf intestine alkaline phosphatase (CIAP)	Fermentas, St. Leon-Rot, Germany
Concanavalin A	St. Louis, USA
DpnI	NEB, Frankfurt, Germany
EarI,	NEB, Frankfurt, Germany
EciI	NEB, Frankfurt, Germany
EcoRV	Fermentas, St. Leon-Rot, Germany
Exonuclease I	NEB, Frankfurt, Germany
Extravidin-Alkaline Phosphatase	Sigma, St. Louis, USA
HotStar Taq DNA Polymerase	Qiagen, Hilden, Germany
Maxima H Minus Reverse Transcriptase	Thermo Fisher, Dreieich, Germany
NotI	Fermentas, St. Leon-Rot, Germany
PfuUltra Hotstart High-Fidelity DNA Polymerase	Agilent, Waldbronn, Germany
PfuUltra II Fusion HS DNA Polymerase	Agilent, Waldbronn, Germany

RiboLock RNase Inhibitor	Thermo Fisher, Dreieich, Germany
RevertAid H Minus Reverse Transcriptase	Thermo Fisher, Dreieich, Germany
SacI	Fermentas, St. Leon-Rot, Germany
SapI	NEB, Frankfurt, Germany
T4 DNA Ligase	Fermentas, St. Leon-Rot, Germany

2.1.4. Antibodies

The following monoclonal antibodies were used for murine T cell staining: CD4-APC-Cy7 (BD), CD4-BV421 (BioLegend), CD8-eFluor506 (eBioscience), CD8-PE (BD), CD8-PE-Cy7 (ebioscience), TCRβ murine-APC (BD), TCRβ murine-PE (BD), IFNγ-eF450 (eBioscience), $Adpgk_{R304M}$-APC tetramer (MBL), $Reps1_{P45A}$-APC tetramer (MBL), CD40L-PE (BioLegend), OX40-PE-Cy7 (BioLegend) and TNFα-APC (TNFα-APC). Viability Dye-7-AAD (Beckman Coulter) was used.

2.1.5. Peptides

Peptides were synthesized by JPT Peptide Technologies (Berlin, Germany) with 90 % purity and were dissolved in water with 10 % DMSO to a final concentration of 2 mM. Overlapping peptide pools (PepMix^{TM}) encoding the whole length of HPV 16 Protein E6, HPV 16 Protein E7, as well as control antigen HIV-gag were used. Single peptides used for *ex vivo* stimulation of splenocytes from vaccinated mice or for APC-loading include: HPV $E7_{11-20}$ (YMLDLQPETT), HPV $E7_{86-93}$ (TLGIVCPI), HPV $E6_{127-142}$ (DKKQRFHNIRGRWTG), HPV $E6_{29-38}$ (TIHDIILECV), $HBVs_{126-138}$ (RGLYFPAGGSSSG), $HBVc_{128-140}$ (PPAYRPPNAPIL), $KKLC1_{66-74}$ (ILNNFPHSI), $Adpgk_{R304M}$ (ASMTNMELM), $Adpgk_{WT}$ (ASMTNRELM), $Reps1_{P45A}$ (AQLANDVVL), $Reps1_{WT}$ (AQLPNDVVL), $Ova_{257-264}$ (SIINFEKL) and $OVA_{323-339}$ (ISQAVHAAHAEINEAGR).

2.1.6. RNA

Antigen-encoding RNAs were synthesized by BioNTech RNA Pharmaceuticals GmbH. Adpgk and Reps1 RNAs encode 27 amino acids of the respective antigen with the mutated amino acid in the centre (position 14). Full-length antigen sequences were used for HPV E7, HPV E6 and KKLC1 vaccinations. The following RNAs were also used: Mouse CD1 spleen mRNA (Zyagen, San Diego, California), Brassica Napus Raps carrier RNA (synthesized in house), and Poly(A) carrier RNA (Qiagen, Hilden, Germany).

2.1.7. DNA Vectors

TCR chains and HLA antigens in pST1 expression vectors were used for *in vitro* transcription (IVT) as previously described[127]. The vector backbone was derived from the pST1-sec-2βgUTR-A(120)-Sap1 plasmid featuring a T7 transcription promotor site. PJET1.2/blunt vectors were used for blunt-end cloning (Thermo Fisher, Dreieich, Germany).

2.1.8. Buffers

Name	Components
Cell-capture buffer	Amounts/well: 1 μL dNTP mix (10 mM), 5 ng Poly(A) carrier RNA, 8 U RiboLock RNase Inhibitor (40 U/μL), 0.2 % of volume, Triton X-100, 3' gene-specific primer (GSP) or dT30Vn primer, filled up to 4 μL.
FACS buffer	2 % FCS, 5 mM EDTA, 0.1 % sodium acid, in PBS.
MACS buffer	PBS, 0.5% BSA and 2 mM EDTA.
RevertAid and Maxima H Minus Reverse Transcriptase buffer (customized)	250 mM Tris-HCl (pH 8.3 at 25°C), 250 mM KCl, 30 mM MgCl$_2$, 50 mM DTT.

2.1.9. Media

Name	Components
DC medium (human)	500 mL RPMI Medium Complete, 1000 U/mL IL-4, 5000 U/mL GM-CSF.
DC medium (mouse)	500 mL RPMI-1640 Glutamax, 10 % FCS, 1 % penicillin/streptomycin (Pen/Strep) solution (100 U/mL and 100 μg/mL), 0.1 % 50 μM ß-ME, 10 % GM-CSF supernatant, 100 U/mL GM-CSF, 25-50 U/mL IL-4.
ELIspot Blocking medium	2 % human albumin in X-Vivo 15 Medium.
LB medium	5 g Yeast extract, 10 g Tryptone, 10 g sodium chloride, 1 L distilled H$_2$O.
Mouse T cell medium	RPMI Glutamax 1%, sodium pyruvate 1%, NEM NEAA 1% , HEPES (1M) 1%, FCS heat inactivated (h.i.) 10%, Pen/Strep 0.5%, β-ME 0.1%.
RPMI murine T cell medium	500 mL RPMI-1640 Glutamax-I, 5 % h.i. human AB serum, 1 % sodium pyruvate (100 mM), 1 % NEAA, 0.5 % Pen/Strep.

2.1.10. Solutions

Name	Components
DEPC water	1 mL DEPC, 1 L distilled H_2O, shake, autoclave.
Avidin-Peroxidase complex solution	10 mL PBS/Tween 0.1 %, one drop each of reagent A and B (Vectastain kit), incubated 30 min before use.
LB agar	1 L LB medium, 15 g agar.
MOPS gel (1 %)	1.5 g Agarose, 120 mL DEPC water, dissolved by boiling, cooled down to 55 °C, 15 mL 10X MOPS stock solution and 15 mL 37 % formaldehyde.
TAE 50x stock solution	242 g TrisBase, 57.1 mL Glacial Acetic acid, 100 mL 0.5 M EDTA, pH 8.0, 1 L distilled H_2O.
TAE gel (1.5 %)	2.25 g agarose, 150 mL 1X TAE, dissolved agarose by boiling, cooled down, 75 µL ethidium bromide (EtBr) added.

2.1.11. Cells, cell lines and animals

The human chronic myeloid leukemia cell line K562 was obtained from ATCC. K562 cells transiently or stably transfected with HLA allelotypes were used for A2DR1-mice derived TCR specificity determination. One Shot Top 10 (Thermo Fisher) or XL1-Blue (Agilent) Competent E. coli cells were used for heat-shock transformations. The MC38 tumor cell line was kindly provided by Genentech. For immunization experiments, female C57BL/6 mice were purchased from Envigo and Janvier. A2DR1 (HLA-A*0201[+]/DRB1*0101[+]) double transgenic mice embryos were purchased from TAAM-CNRs[129]. Thy1.1[+]2D2 TCR MOG35-55 transgenic C57BL/6 mice, expressing a TCR recognizing MOG$_{35\text{-}55}$ in the context of MHC class II (MHC haplotype I-Ab) were provided by Prof. Dr. Ari Waisman (University Medical Centre of the Johannes Gutenberg University Mainz, Germany). All mice were bred in the animal facility of the BioNTech SE, Germany. All experiments were performed according to German animal experimentation regulations.

2.1.12. Commercially available kits

Name	Manufacturer
Agilent D1000 Screentape	Agilent, Waldbronn, Germany
Agilent High Sensitivity DNA Kit	Agilent, Waldbronn, Germany
Clonejet PCR Cloning Kit	Thermo Fisher, Dreieich, Germany
High Sensitivity D1000 Screentape	Agilent, Waldbronn, Germany
Magnetic Cell Sorting (MACS) Kit	Miltenyi Biotec, Bergisch Gladbach, Germany
Miseq Reagent Kit V2 (500 Cycles)	Illumina, San Diego, USA

Miseq Reagent Kit V3 (600 Cycles)	Illumina, San Diego, USA
Nucleospin Plasmid Mini Kit	Macherey-Nagel, Düren, Germany
Qiafilter Plasmid Midi Kit	Qiagen, Hilden, Germany
Qiaquick Gel Extraction Kit	Qiagen, Hilden, Germany
Qiaquick Minelute Gel Extraction Kit	Qiagen, Hilden, Germany
Qubit Dsdna HS Assay Kit	Thermo Fisher, Dreieich, Germany
Quikchange XL Site-Directed Mutagenesis Kit	Agilent, Waldbronn, Germany
Rneasy Micro Kit	Qiagen, Hilden, Germany
Truseq DNA Library Preparation Kit V2 A/B	Illumina, San Diego, USA

2.1.13. Consumables

Name	Manufacturer
6/24-well plates	Becton Dickinson, Heidelberg, Germany
8-Strip tubes, 0.2 mL	Molecular BioProducts, San Diego, USA
96-well ELIspot plates	Millipore, Bedford, USA
96-well microtiter plates	Nunc, Wiesbaden, Germany
96-Well PCR hard-shell plates	Bio-Rad, Rudigheim, Germany
Adhesive PCR foil	VWR, Darmstadt, Germany
Cell culture flasks	Becton Dickinson, Heidelberg, Germany
Cell strainer, 40 µm, Nylon	Becton Dickinson, Heidelberg, Germany
Electroporation cuvettes, 4 mm gap, sterile	BioRad, Hercules, USA
Eppendorf tubes, 1.5 and 2 mL	Eppendorf, Hamburg, Germany
FACS tubes	Becton Dickinson, Heidelberg, Germany
Falcon tubes	Greiner BioChemica, Flacht, Germany
Filter tips	Eppendorf, Hamburg, Germany
PCR tubes, 0.5 mL	Molecular BioProducts, San Diego, USA
RNase-free tubes (1,5 and 2 mL)	Eppendorf, Hamburg, Germany

2.1.14. Computational biology tools

Tool	Tool
Bioedit Sequence Alignment Editor	MiXCR
Eurofins Genomics PCR Primer Design Tool	MobaXterm Xserver
Eurofins Primer Properties Tool	NetBeans
Flowjo	Nucleoblast
Gatcviewer 1.0	PyCharm
Graphpad Prism	RSEM
HT-Seq-Count	UCSC Genome Browser
Imgt/V-Quest	WinSCP

2.2. Methods

2.2.1. Molecular biology

2.2.1.1. Total RNA isolation

Total RNA was isolated from single T cells and from mouse-derived single cell spleen suspensions. Spin column-based nucleic acid purification, was performed using the RNeasy Micro Kit (Qiagen). Nucleic acids bind to silica under certain salt conditions. Four main steps are involved, cell lysis, nucleic acid binding to the Silica gel membrane, nucleic acid washing and nucleic acid elution. Each frozen sample was harvested by adding 75 µL RLT-β-ME buffer: a chaotropic lysing agent guanidine thiocyanate. The cells were transferred to a 1.5 mL reaction tube containing 275 µL RLT-β-ME buffer. Samples were homogenized by vortexing for 45 s. 350 µL 70 % ethanol was added to the lysates and mixed to create a silica-binding environment. Reaction contents were transferred to the RNeasy columns and centrifuged for (8000 ×g 15 s) before being placed in new collection tubes. The nucleic acid binds the membrane as the solution passes through. RNA molecules longer than 200 nucleotides bind the membrane. 5.8S rRNA, 5S rRNA and tRNA are excluded. 700 µL ethanol-containing RW1 buffer were added to each sample and the columns centrifuged. This step was repeated using 500 µL ethanol-containing RPE buffer and 500 µL 80 % ethanol to remove remaining salts on the membrane. For elution, 14 µL RNase-free water were used.

Total RNA of single T cells was isolated in 14 µL water. Since the total volume of the cDNA strand synthesis reaction is 10 µL, the eluted RNA was concentrated to 4 µL using the Rotational Vacuum Concentrator. RNA samples were placed open lidded into the centrifuge chamber connected to a vacuum pump. Upon lowering of the pressure, the boiling point drops and the solvents slowly evaporate. The centrifugal force keeps the RNA at the bottom of the sample.

2.2.1.1. Reverse transcription

Reverse transcription (RT) was performed to synthesize first strand cDNA using mRNA from single sorted or bulk T cell splenocytes. Reactions were performed using the template switching (TS) protocol described by Picelli *et al.* (2013) catalyzed by a recombinant Moloney Murine Leukemia Virus (M-MuLV)-derived Reverse Transcriptase (RTase) (Thermo Fisher)[130]. Mastermix 1 (RT-MMX1) **(Table 2.1)** was added to the RNA samples. Different 3' primers were used in different experiments (scTCRseq vs scTCRphenoSeq).

Table 2.1| RT-MMX1.

Component (stock concentration)	Volume (Final concentration)
dNTP mix (10 mM each)	1 μL (1 mM each)
TS primer (100 μM)	0.12 μL (1.2μM)
3' Primers d(T)30 VN (100 μM) TRAC-GSP (100 μM) TRBC-GSP (100 μM)	 0.12 μL (1.2μM) 0.265 μL (2.64μM) 0.13 μL (1.32μM)
RNase-free Water	Fill up to 7 μL

Samples were denatured at 70 °C for 5 min before being placed on ice for 2 min. Mastermix 2 (RT-MMX2) **(Table 2.2)** was added to reach a 10 μL reaction volume. Samples were incubated at 42 °C for 90 min; 70 °C for 15 min when using the RevertAid H Minus RTase. For RT reactions using the Maxima H Minus RTase, samples were incubated as follows 50 °C for 90 min; 70 °C for 15 min.

Table 2.2| RT-MMX2.

Component	Volume/reaction
Customized reaction buffer, 30 mg Mg^{2+} (5X)	2 μL
M-MLV RevertAid H Minus Reverse Transcriptase or Maxima H Minus Reverse Transcriptase	1 μL

2.2.1.2. Polymerase chain reactions

Three Polymerase were used:

- **PfuUltra Hotstart High-Fidelity DNA Polymerase** (Agilent) exhibits proof reading activity (3'-5' exonuclease activity allows excision of incorrect inserted base pairs).
- **PfuUltra II Fusion HS DNA Polymerase** (Agilent) possesses even higher fidelity (accuracy) at faster extension times.
- **HotStar Taq DNA Polymerase** (Qiagen) is a less costly polymerase and without proof reading activity. It was used for some platform optimization experiments.

Standard PCRs were performed as shown in **table 2.3** with cycles ranging from 12-35. Unchanged parameters included: 50 uL total reaction volume, 1 μL Polymerase, 1 μL dNTP mix (10 mM) and 5 μL 10X Polymerase buffer. Template and primer amounts varied from one experiment to the other.

Table 2.3| PCR standard programs used for each Polymerase.

Property/application	Temp (°C)	PfuUltra	PfuFusion	Taq
Heat-activation	95	2 min	1 min	15 min
Denaturation	94	30 s	20 s	30 s
Annealing	T_m - ~5	30 s	20 s	30 s
Elongation	72	60 s	30 s	60 s
Final elongation	72	6 min	3 min	6 min

2.2.1.3. Site-directed mutagenesis PCR

Purified mini bacterial colonies that had acquired false nucleotides were mutagenized back to the correct sequence using the QuikChange XL Site-Directed Mutagenesis Kit (Agilent). For mutagenesis, two complimentary primers containing the desired mutation, flanked by unmodified nucleotide sequences were designed. The two site-specific primers, each complementary to the opposite vector strands, are extended using the HS Pfu DNA Polymerase. The PCR reaction used 10 ng dsDNA template, 125 ng of each primer, in a final volume of 50 µL. The PCR program used is shown in **Table 2.4**. After the PCR, 1 µL DpnI restriction enzyme (10 U/µL) (target sequence: 5´-Gm6ATC-3´) was added and incubated at 37 °C for 1 h to digest the parental *E.coli*-derived methylated dsDNA.

Table 2.4| QuikChange XL Method PCR cycling parameters

Temperature	Time	Cycles
95 °C	1 min	1
95 °C	50 s	
60 °C	50 s	18
68 °C	1 min/kb of vector length	
68 °C	7 min	1

2.2.1.4. Magnetic bead-based DNA size selection

Solid Phase Reversible Immobilization (SPRI)-based bead purification was used to purify PCR amplicons. The SPRIselect (Beckman Coulter) technology was performed according to the manufacturer's instructions. SPRIselect beads were later replaced with the less costly Agencourt AMPure XP beads (Beckman Coulter). The paramagnetic beads are composed of polystyrene surrounded by magnetite coated with Carboxyl molecules. Polyethylene glycol (PEG) in the bead solution causes the negatively charged DNA to bind the carboxyl groups. Increasing the bead:sample volume ratio increases the efficiency of smaller fragment binding. Left-size selection was performed to eliminate smaller fragments. Elution volumes ranged from 30–50 µL according to the application.

2.2.1.5. Exonuclease I treatment

Exonucleases catalyze single stranded DNA (primers) phosphodiester bond hydrolysis in a 3'>5' direction. They require a free 3' or 5' end and are not sequence specific. Exonuclease I was used for removal of single-stranded primers for nested PCR reactions. Reactions were performed using 5 U or 10 U and 10X Exonuclease I buffer and incubated at 37 °C 40 min; 80 °C 20 min.

2.2.1.6. Measuring nucleic acid concentration and size

The NanoDrop (Thermo Fisher) is an ultraviolet–visible light spectrophotometer used to quantify total DNA, RNA and protein. Samples measured included primers, vectors, spleen-derived total RNA and purified PCR products (2-15,000 ng/uL). The optical system delivers light at the absorbance peak for RNA and DNA (260 nm). The nucleic acid-absorbed light is proportional to the nucleic acid amount. Light absorbance is also measured at the protein absorbance peak (280) to determine sample purity.

The Qubit Fluorimeter (Invitrogen) was used to measure DNA sample concentration before and after the NGS library preparation (10 pg/µL-100 ng/µL).

The Qiaxcel (Qiagen) uses capillary electrophoresis to image PCR products. A current is applied into a capillary filled with gel. Negatively charged DNA molecules migrate towards the positively charged terminus. The movement signal is converted into a gel image.

The Agilent 2200 TapeStation System is an automated capillary electrophoresis platform that uses consumable ScreenTapes. The high sensitivity D1000 ScreenTape Assay has a quantitative range of 10-1000 pg/µL with a size range of 35-1500 bp for NGS library preparation QC checks. The Agilent 2100 Bioanalyzer quantificaties 50-7000 bp long DNA molecules down to 5 pg/µL.

Agarose gel electrophoresis (AGE) was used to separate DNA molecules. Agarose is a 3D matrix held by hydrogen bonds. A voltage is applied at the cathode forcing DNA molecules to migrate towards the anode. Smaller molecules move faster and further than larger ones. EtBr, a DNA-intercalating agent was used to fluorescently label DNA. AGE was used for visualizing and purifying TCR chains. Samples were mixed with 5 % bromophenol blue marker (Applichem). Gels were placed in TAE buffer. Samples were run at 80V for ~1 h. Gel images were created using the BioStep Argus X1 Transilluminator (Biostep).

Denaturing formaldehyde agarose gel electrophoresis was used to assess RNA length and integrity after IVT RNA synthesis. RNA tends to form secondary structures. Therefore, formaldehyde (Merck), a denaturing agent is incorporated into the gel to ensure that the

heat-denatured RNA remains single stranded in the gel. 1 μL RNA was mixed with 1 μL EtBr-containing RNA Loading Dye (Fermentas). All samples were heat-denatured at 70 °C for 10 min before being placed on ice for 3 min. Samples were run at ~80 V for ~50 min.

2.2.1.7.   Next-generation library preparation and sequencing

Prepared DNA libraries are captured on a flow cell lawn of surface-bound primers complementary to the library adapters. Each fragment is amplified into clonal clusters through bridge amplification, after which templates are ready for sequencing. The reversible terminator method used by the Illumina MiSeq platform detects single bases as they are incorporated into the DNA template strands. Paired-end (PE) sequencing (2x300bp) involves sequencing both ends of the DNA fragments. Libraries where given a unique adaptor index. The TruSeq DNA Library Preparation Kit v2 (Illumina) was used. 40 ng of each sample was given one of 24 unique indices. The preparation was performed according to the manufacture's protocol and consisted of the following steps: DNA end repair and 5' phosphorylation, 3' end adenylation, adaptor ligation, PCR enrichment and sequencing primer incorporation. Library samples were diluted to 2 nM using elution buffer. The dilutions were pooled to reach 14 μL of the pooled volume. 10 μL of this mixture were mixed with 10 μL of 0.1 M diluted NaOH. 14 μL of that mixture were mixed with 2 μL of 2 nM PhiX (Illumina) and incubated at RT for 5 min. Pre-chilled HT1 Buffer was added to the samples to reach a concentration of 12 pM cluster density. The sample was added to the inlet of the thawed MiSeq cartridge. Two types of flow cells were used depending on the number of samples sequenced **(Table 2.5)**.

Table 2.5| MiSeq Reagent kit flow cell properties.

Property	MiSeq v3	MiSeq Nano v2
Read depth	25×10^6	1×10^6
Cycles (sequencing length)	600	500

2.2.1.8.   Sanger sequencing

Sanger sequencing involves the incorporation of chain-terminating dideoxynucleotides during DNA replication initiated by a sequencing primer and was used to sequence cloned TCR fragments. Sequencing was performed by GATC Biotech, Konstanz, Germany.

2.2.1.9.   Blunt-end cloning

The Clonejet PCR Cloning Kit (Thermo Fisher) was used during platform development to sequence PCR amplicons. The blunt-end cloning kit uses an Ampicillin-resistant pJET1.2/blunt vector. Recircularized pJET1.2/blunt vector molecules lacking an insert

express a lethal restriction enzyme. The ligation reaction consisted of 1 µL T4 DNA Ligase, 10 µL 2X Reaction Buffer, 30 ng AGE-purified PCR product and 1 µL pJET1.2/blunt Cloning Vector (50 ng/µL) in a total volume of 20 µL. The ligation reaction was incubated at RT for 5 min. 2.5 µL of the reaction were directly used for transformation in Top10 *E.coli* cells.

2.2.1.10. TCR chain cloning using remaining first strand cDNA

After identifying the αβ TCR chain pairs of interest, first strand cDNA from the respective well was cloned for downstream functional characterization. The cloning protocol was adapted and modified from Omokoko *et al.* (2014) (**Figure 2.1**)[127].
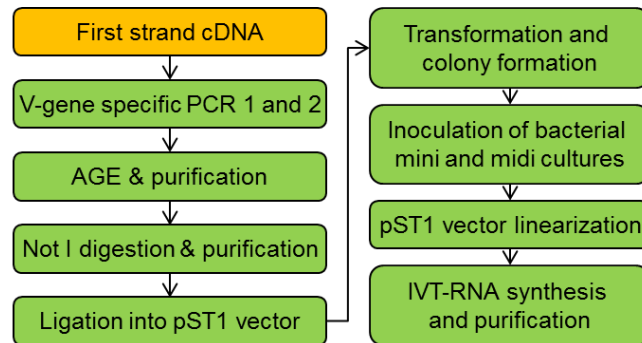
Figure 2.1| Cloning αβTCR chains of interest after NGS using first strand cDNA.

10 µL of the first strand cDNA were amplified in the first PCR composed of: 2 µL Vgene leader-specific primer (10 µM), 2 µL TRAC or TRBC GSP, 2 µL PfuUltra High-Fidelity DNA Polymerase, 10 µL 10X reaction buffer and 2 µL dNTP mix (10 mM) in a 100 µL final reaction volume. The PCR program was as follows: 95°C 2 min; 94°C 30 s, T_m-5 °C 30 s, 72°C 60 s x 21 cycles; 72°C 6 min. 6 µL of the product were used for a subsequent PCR composed of: 1 µL PfuUltra II Fusion HotStart DNA Polymerase, 5 µL 10X Reaction Buffer, 1 µL dNTP mix (10 mM) and 3 µL of each of the same primers used in PCR 1. The reaction was incubated as follows: 95°C 1 min; 94°C 30 s, T_m-5 °C 30 s, 72°C 30 s x 40 cycles; 72°C 3 min. PCR products were purified using AGE and the Qiaquick Gel Extraction Kit according to the manufacturer's protocol. DNA was eluted in 50 µL nuclease-free water.

AGE-purified TCR chains were NotI digested at the 5' end for ligation into the NotI-linearized pST1 vector. The amplicon's 3' end remained blunt end and was ligated to the vector through the PCR reverse primer overhanging phosphate. The 1 h reaction at 37 °C consisted of 50 µL purified DNA sample, 0.75 µL NotI enzyme (10 U/µL) (Fermentas), 5 µL Cut Smart Buffer and 0.25 µL nuclease-free water. NotI digested TCR chains were purified using the Qiaquick Minelute Kit (Qiagen) according to the manufacturer's instructions. DNA was eluted in 15 µL nuclease-free water.

NotI-digested TCR chains were ligated into the pST1 vector entailing the respective TRC gene. The following reagents were mixed to reach a volume of 10 µL before being incubated for 2 h at RT **(Table 2.6).** After ligation, the pST1-TCR vector was transformed into *E.coli* cells.

Table 2.6. Ligation reaction master mix for different DNA concentrations.

pST1 vector 50ng/µL	T4 Ligase buffer	T4 Ligase	Water Low Middle High	DNA sample Low Middle High
1 µL	1 µL	1 µL	/	7 µL
			2 µL	5 µL
			4 µL	3 µL

;where 'high' corresponds to concentrations above 20 ng/µL, 'middle' between 15-20 ng/µL and 'low' below 15 ng/µL.

2.2.1.11. Heat-shock transformation

Transformation is the process by which bacterial cells, in a state of competence, take up exogenous genetic material from the surrounding medium. One Shot Top 10 (Thermo Fisher) or XL1-Blue (Agilent) competent *E. coli* cells were used. Cells were thawed on ice. For XL1-Blue cells, 3.6 µL β-mercaptoethanol were added and incubated on ice for 10 min. 5 µL or 10 ng of the ligation reactions and mutagenesis reactions were added to the cells, respectively. Cells were incubated on ice for 30 min. Cell suspensions were heat-shocked at 42 °C in a water bath (XL1-Blue cells 45 s, Top 10 cells 30 s) and put back on ice for 2 min. 200 µL preheated S.O.C. medium (Thermo Fisher) were added and samples were incubated for 1 h at 37 °C in a thermal shaker shaking at 700 rpm. 100 µL transformed cell suspension was pipetted onto Kanamycin (Ampicillin for the CloneJet vector) agar plates. Plates were stored for 16 h in an incubator at 37 °C.

2.2.1.12. Preparation of plasmid DNA from *E. coli* cells

Inoculation is the introduction of transformed *E. coli* cells into a culture medium allowing growth of a chosen bacterial colony. Mini bacterial cultures were inoculated using the transformed *E.coli* cells to generate enough plasmid for a Sanger sequencing quality control. Midi bacterial cultures were then inoculated using the mini cultures of the correctly cloned TCRs for further amplification required for subsequent plasmid linearization needed for IVT.

For mini cultures, the appropriate amount of the chosen medium was distributed across the tubes/wells **(Table 2.7)**. Picked colonies were transferred into a vessel and incubated

for 16 h at 37 °C and 150 rpm. For midi cultures, 40 mL of LB medium (0.025 % kanamycin or 0.1 % ampicillin) were added to 250 mL Erlenmeyer flasks. 100 μL retain sample from the respective mini culture was added to the respective flask and incubated for 16 h at 37 °C and 150 rpm.

Table 2.7| Inoculation vessels for bacterial Mini or Midi cultures

Sample number and type	Inoculation vessel	Inoculation medium	Volume
≤ 24 Minis	15 mL tubes	LB	3 mL
25-64 Minis	48-deep-well plate	LB	3 mL
≥ 65 Minis	96-deep-well plate	TB	1.5 mL
Midis	Erlenmeyer flask	LB	40 mL

Following the growth of the DNA plasmid-containing *E.coli* cells in a bacterial culture, cells were lysed and the DNA plasmid purified. Different commercial kits were used based on the number of samples and expected plasmid DNA yield **(Table 2.8)**. Purification was performed according to the manufacturer's protocols. Cells were lysed under alkaline conditions in a buffer containing RNase A and sodium dodecyl sulfate (SDS). SDS solubilizes phospholipids leading to lysis while the alkaline conditions denature DNA and proteins. The lysate is neutralized and adjusted to high-salt binding conditions allowing denatured proteins, chromosomal DNA, cellular debris, and SDS to precipitate, while the smaller plasmid DNA renature correctly and stay in solution after centrifugation. Supernatants are applied to the Silica membranes and washed. For midi preparation, 300 μL culture were used for glycerol stock generation for later re-inoculations by mixing with 3 volumes 75% glycerol.

Table 2.8| Vector-DNA purification kits used.

Samples	Purification kit	Elution volume
≤ 24 Minis	NucleoSpin Plasmid Mini	40 μL
25-64 Minis	NucleoSpin 8 Clean-up	2 elution steps: 75 and 50 μL
≥ 65 Minis	NucleoSpin 96 PCR Clean-up	2 elution steps: 75 and 50 μL
Midis	Qiafilter Plasmid Midi	200 μL

After, 10 μg purified plasmid were linearized using one of three possible restriction enzymes **(Table 2.9).** The linearized products were purified using the Qiagen Gel Extraction kit to be used for IVT.

Table 2.9| pST1-553 vector linearization.

Restriction enzyme	Cut Smart Buffer	Restriction enzyme	Final reaction volume	Incubation time
Ear I	3 μL	2 μL	30 μL	37 °C for 3 h
Eci I	5.5 μL	5.5 μL	55 μL	37 °C for 4 h
Sap I	10 μL	10 μL	100 μL	37 °C for 15 h > 65 °C for 20 min

2.2.1.13. Generation of *in vitro* transcribed RNA

IVT was performed to synthesize RNA using a T7 RNA Polymerase promoter present in the pST1-TCR DNA. All steps were performed under RNase-free conditions. 10 μg of linearized and purified vector were used for the IVT RNA generation. The reaction consisted of 3 mM CAP, 7.5 mM ATP, 7.5 mM CTP, 7.5 mM UTP, 1.5 mM GTP, 10 μL 10X T7-Transcription buffer, 10 μL T7 Enzyme Mix (200 U/μL) in a 100 μL final volume and was incubated for 37 °C for 1 h. 1.5 μL GTP (100 mM) were added to each sample and incubated at 37 °C for another hour. GTP is added at a lower amount during the first incubation to prevent its competition with the CAP molecules. For the degradation of the remaining DNA, 5 μL Turbo DNase (2 U/μL) were added to each sample and incubated at 37 °C for 15 min. After synthesis, the RNA is purified using the RNeasy Micro Kit. QC was performed using MOPS gel and RNA was stored at -80 °C.

2.2.2. Cell culture and immunological assays

2.2.2.1. Spleen isolation and single cell suspension generation

Single T cell suspensions were generated from isolated mouse spleens. The left lateral abdomen of the sacrificed mouse was disinfected with 70 % ethanol and the spleen excised using surgical scissors and disinfected forceps. The spleen was kept in 2 mL PBS on ice. Freshly isolated spleen was grinded through a cell strainer into a 50 mL falcon tube using a sterile syringe plunger. Remaining cells were rinsed with 10 mL RPMI. The suspension was centrifuged (300 xg 8 min, 4 °C) before being incubated for 5 min in 1 mL erythrocyte-lysis buffer. The reaction was terminated by adding 15 mL cold RPMI. Cells were resuspended in mouse medium and counted via the Vi-Cell Counter to adjust the cell concentration accordingly.

2.2.2.2. Cultivation of single cell suspensions and cell lines

Single cell suspensions were cultivated at 37 °C, 5 % CO_2 and 95 % relative humidity. The MC38 tumor cell line was cultured in T75 flasks at a density of 2 x10^4 cells/cm^2 in RPMI (2mM L-glutamine and 10% FCS) and subcultured every 2 days.

2.2.2.3. Liposomes, RNA-LPX preparation and immunization

Liposomes with positive (cationic) net charge were used to complex RNA for the formation of RNA-LPX and comprised of the cationic lipid DOTMA (Merck & Cie) or DOTAP (Merck Eprova), and the helper lipid DOPE (Avanti Polar Lipids or Corden Pharma) or cholesterol (Sigma-Aldrich). Liposomes were produced either by protocols based on the thin film hydration method (HPV E7 and HPV E6 vaccinations) or by an adopted proprietary protocol based on the ethanol injection technique (KKLC1, neoantigen vaccinations). Liposome manufacture and LPX formation were performed as previously described[131]. For HPV E7, E6 and KKLC1 vaccinations, 30 µg RNA-LPX were injected intravenous on day 1, 8, 15 and 22. For HBVs and HBVc vaccinations, 100 µg respective peptide were injected subcutaneous with 50 µg polyinosinic:polycytidylic acid on day 1, 8 and 15. Vaccination with neoantigen-encoding RNA-LPX was performed intravenously on day 1, 8 and 15 using 20 µg RNA or 0.5 µg and 39.5 µg irrelevant RNA (vaccine backbone without encoding antigen). Splenocytes were harvested one week after the last vaccination.

2.2.2.4. *In vitro* expansion, activation or restimulation of murine T cells

For unspecific activation of murine T cells in the phenotyping experiments, CD3/CD28 Dynabeads were used (Thermofisher) at a 1:1 bead:cell ratio for 4, 8, 12, or 24 h following the manufacturer's instructions in the presence of 5ng /mL IL-7 and IL-15, each. T cells were also activated using 100 ng/mL PMA and 750 ng/mL ionomycin. Beads were removed using a Dynamag-2 Magnet (Life Technologies). Activation of 2D2 T cells was performed by incubating the cells for 12 h in mouse medium with IL-7, IL-15 and Concanavalin A (5 ng/mL, 5 ng/mL and 2 µg/mL, respectively) at 1×10^6 cells /mL.

For antigen-specific TCR discovery, 1×10^6 cells /mL single splenocytes of vaccinated mice were restimulated *in vitro* using 2-4 µg/mL of the respective antigen peptide for 12 h. Activated cells were labelled using IFNγ-secretion assay-APCkit as per manufacturer's instructions (Miltenyi) prior to sorting.

2.2.2.5. Flow cytometry and single T cell sorting

In flow cytometry, single cells pass an excitation source. Light is scattered or absorbed and re-emitted at a higher wavelength (fluorescence), which is collected by different detectors (one for each fluorescent emission peak). The forward scatter (FSC) and side scatter (SSC) describe the cell volume and the granularity, respectively. After addition of an optimal concentration of each fluorochrome labeled anti-antibody, samples were incubated in the dark at 4 °C for 30 min. For the detection of dead cells, propidium iodide

(PI), a fluorescent intercalating dye, was added in a concentration of 2 µg per mL. PI is taken up by dead cells with damaged cell membranes. After staining, cells were first washed using 5 mL and resuspended in 300 µL FACS Buffer. Flow cytometry studies were performed on a BD FACSAria II and BD FACSCanto II (BD Biosciences).

$CD8^+CD3^+IFN\gamma^+$ or $CD4^+CD3^+IFN\gamma^+$ single T cells were sorted into 4 µL mild hypotonic lysis cell-capture buffer at the flow cytometry facility of the University Medical Center of the Johannes Gutenberg University Mainz. V-bottom, high-profile, semi-skirted 96-well PCR hard-shell plates (BioRad) were used. After sorting, plates were covered with adhesive foil, centrifuged and placed on dry ice and stored at -80 °C.

2.2.2.6. Microscopic manual single cell picking

Single T cells were picked by hand when only a few T cells were needed. 100 µl of the single cell suspension was diluted in 2 mL PBS in a 6-well cell-culture plate. Looking into the microscope, the tip of a 2 µl-pipette (set to 1 µl and pressed to the first pressure point) was slowly guided to the vicinity of a single cell. The pipette-plunger was slowly released to apply a suction force allowing the cell to travel into the pipette tip. After the cell passed into the opening of the tip, the plunger-release process was halted and the cell was transferred into the cell-capture buffer well and placed on dry ice.

2.2.2.7. Magnetic-activated cell sorting

Using Magnetic-Activated Cell Sorting (MACS) technology (Miltenyi Biotec), surface antigen-specific paramagnetic-coupled antibodies are applied to isolate cell populations expressing the respective antigens: CD8, CD4 or IFNγ. After washing, the number of cells was adjusted to 1×10^7 cells per 40 µL cold MACS Buffer. Cells were magnetically labeled with 10 µL respective Biotin-Antibody Cocktail and incubated for 5 min at 4 °C. 40 µL MACS Buffer were added as well as 20 µL Anti-Biotin Microbeads. A chilled MACS Column LS was rinsed with cold MACS Buffer. The cell suspension was added to the column placed in the magnetic field. For negative isolation, the flow-through was collected. 3 mL of MACS Buffer were added and flow-through was collected. For positive isolation, after removal of the column from the magnetic field, labelled cells were eluted by addition of 5 mL of MACS Buffer by pushing a plunger into the column. For $IFN\gamma^+$ T cell isolation, 1×10^7 cells were resuspended in 80 µL cold MACS Buffer. 20 µL IFNγ Catch Reagent (per 1×10^7 cells) were added and stored for 5 min on ice. 10 mL warm (37°C) mouse medium was added to dilute the cells and incubated for 45 min at 37°C under slow continuous rotation by using a Macmix Rotator. Cells were washed twice and resuspended in 60 µL cold MACS Buffer and 20 µL mouse IFNγ Detection Antibody (APC) were added and incubated for 10 min on ice. Cells were washed and resuspended

in 60 µL cold MACS Buffer and 20 µL Anti-APC MicroBeads were added and stored for 15 min in the refrigerator at 8 °C. Cells were washed and resuspended in 500 µL MACS Buffer before being added to the MACS Column LS.

2.2.2.8. Cell counting

Cell numbers were counted using a Vi-Cell XR cell Counter (Beckman Coulter). Cells were diluted 1:10 in a fixed volume of 500 µL. The Vi-Cell determines the concentration and viability of the cells based on their size, contrast, circularity and Trypan-Blue staining of dead cells.

2.2.2.9. RNA Electroporation

Human PBMCs were isolated by Ficoll-Hypaque (Amersham Biosciences) density-gradient centrifugation from buffy coats from healthy donors obtained from the Transfusion Center of the University Medical Center of the Johannes Gutenberg University Mainz. For C57BL/6-derived TCRs, murine T cells were activated and expanded prior to electroporation. CD8$^+$ or CD4$^+$ T cells were activated with 2 µg/mL Concanavalin A (ConA) in the presence of 5 ng/mL rh IL-7 and 5 ng/mL rh IL-15 or using CD3/28 Dynabeads in the presence of 50 U rh IL-2, respectively. After 3 days, CD8$^+$ or CD4$^+$ T cells were washed or debeaded and expanded with 5 ng/mL rh IL-7 and 5 ng/mL rh IL-15 or using 50 U/mL rh IL-2, respectively. Viable cells were separated from debris prior electroporation using 1.084 Ficoll-Paque PREMIUM (GE Healthcare) density-gradient centrifugation. Depending on the TCR transfected, CD8$^+$ or CD4$^+$-depleted T cells were used. 10 µg IVT RNA was added to T cells (10 µg for K562 cells) suspended in X-VIVO 15 in a precooled 4-mm gap sterile electroporation cuvette (Bio-Rad). Electroporation was performed with an ECM 830 Square Wave Electroporation System (human and murine T cells: 500 V/ 3 ms/ 1 pulse; K562: 200 V / eight ms / 3 pulses). Transfection efficiency was assessed via flow cytometry 24 h after electroporation. For murine T cells, eGFP RNA electroporated T cells served as mock control.

2.2.2.10. Isolation of human peripheral blood mononuclear cells

Human PBMCs were isolated from buffy coats generated from the centrifugation of the donated blood. The cell suspension was diluted 1:1 with PBS. 35 mL of the diluted cell suspension was slowly layered above 15 mL Ficoll-Hypaque (Amersham Biosciences) in a 50 mL Falcon-tube. Tubes were centrifuged (300 xg 45 min). After the density gradient centrifugation, the cloudy layer composed of T cells, B cells, monocytes and NK cells was

transferred into a new tube and cells were washed using cold PBS/EDTA at 300 xg for 8 min at 4 °C.

2.2.2.11. Enzyme Linked Immuno Spot Assay

The IFNγ ELISpot was performed to test the IFNγ secretion from murine or human T cells transfected with TCRs discovered from BALB/c or A2DR1 vaccinated mice, respectively. Multiscreen Filter Plates (Merck Millipore), pre-coated with antibodies specific for IFNγ (Mabtech) were washed with PBS and blocked with X-VIVO 15 (Lonza) containing 2% human Serum Albumin (CSL-Behring) for 1-5 hours. For human A2DR1-derived TCRs, 2.5×10^5 TCR-transfected CD4$^+$ or CD8$^+$ T cells from healthy donors were co-cultured with 5×10^4 peptide-pulsed (4 µg/mL) or antigen-encoding RNA-electroporated K562 cells transfected with HLA-class I or II. For BALB/c-derived TCRs, 1×10^5 effector cells/well were stimulated for 16-20 h with 1×10^4 antigen peptide-loaded (2 µg/mL) BMDCs. For tumor cell line recognition experiments, 2×10^5 transfected murine T cells were co-cultured with 1×10^4 MC38 tumor cells. All tests were performed in duplicate and included positive controls: Staphyloccocus Enterotoxin B (Sigma Aldrich), Con A (Sigma Aldrich) as well as TCRs with known reactivity to OVAI, OVAII and KKLC1[64]. Spots were visualized with a biotin-conjugated anti-IFNγ antibody (Mabtech) followed by incubation with ExtrAvidin-Alkaline Phosphatase (Sigma-Aldrich) and BCIP/NBT substrate (Sigma-Aldrich). Plates were scanned using CTL's ImmunoSpot Series S five Versa ELISpot Analyzer and analyzed by ImmunoCapture software. Spots were summarized as median values for each duplicate.

2.2.3. Computational biology and data analysis

2.2.3.1. ScTCRseq bioinformatics pipeline development

Demultiplexing the reads to single TCR cells was implemented in a Python script using the integrated development environment (IDE) PyCharm. To aid automatic identification of TCR chain pairs, the TCR Chain Filter has been developed in Java using the IDE NetBeans. Both in-house tools and the external tools were connected via Bash into a pipeline to automate most steps of the analysis (written in Notepad++/PyCharm). The tools and scripts were deployed on a Linux server with 8 CPUs and 64GB of RAM. On this server all analyses were carried out accessing it via WinSCP and a terminal application (MobaXterm).

2.2.3.2. Bcl2fastq

The MiSeq generates per-cycle base call (BCL) files at the end of the sequencing run. Downstream analysis applications use per-read FASTQ files as input. Bcl2fastq2 Conversion Software was used to convert the former into the latter as well as to abstract the spiked in scTCRseq sequences from other unrelated samples sequenced on the same flow cell. For the plate demultiplexing, the software reorganizes the FASTQ files based on the index sequencing information for each plate given in the SampleSheet.csv and RunInfo.xml. The following command was used in the directory: bcl2fastq -i /MiSeq_run/Data/Intensities/BaseCalls/ -o /datafolder/ --minimum-trimmed-read-length 30

2.2.3.3. MiXCR

MiXCR is a universal complex Java-based tool that processes big immunome (murine and human TCR and BCR repertoire) data from raw sequences to quantitated clonotypes[118]. MiXCR uses paired and single-end reads, considers sequence quality and corrects PCR errors. MiXCR uses all available sequence information including sequences upstream of V and downstream of J gene segments. MiXCR is composed of three main steps: (1) Align: aligning sequencing reads to V, D, J and C genes references (2) Assemble: assembling clonotypes using alignments obtained to extract specific gene regions e.g. CDR3 (3) Export: exporting alignment or clones to a human-readable text file.

2.2.3.4. Phenotyping gene sequence alignment tools

To determine the read count levels of the chosen T cell phenotyping genes, sequences that are not aligned to TCR genes I MiXCR are further processed by being aligned to the reference sequences of the phenotyping marker genes using STAR aligner. Successfully aligned reads are quantified for each phenotyping gene with htseq-count using the all-union model including non-unique hits (the latter option enables alignment to multiple splice variants of the same gene). Finally, phenotyping gene expression per cell is summarized into a tab-separated table, enabling quick insights into the data.

2.2.3.5. IMGT

IMGT/V-Quest (http://www.imgt.org/IMGT_vquest/vquest) is an online alignment tool for human and murine TCR and immunoglobulin (IG) nucleotide sequences. IMGT/V-QUEST compares the input sequence with the IMGT/V-Quest reference directory sets. TCR V-J and V-D-J rearrangement analysis was performed using the tool. The tool output displays the aligned genes and alleles, the CDR3 region as well as the alignment scores

and frameshifts. The IMGT/Gene-DB (database) was used to obtain V, D and J-gene sequences and relevant information such as the leader sequences and allele differences.

2.2.3.6. BLASTn

The Basic Local Alignment Search Tool (BLASTn) (https://blast.ncbi.nlm.nih.gov/Blast) was applied to identify DNA sequences that could not be aligned to TCR genes (e.g. PCR byproducts). The online tool compares nucleotide sequences to sequence databases and calculates the similarity.

2.2.3.7. Primer design and properties

The list of primers used can be found in the **supplementary section 6.1**. The Eurofins Genomics primer analysis tool was used to calculate DNA primer properties: GC content, T_m, Extinction coefficient, Maximum annealing score and the self-dimer formation score.

PCR primers were designed with the following properties: GC: ~40 %, T_m: 62-68 °C (without tag) (where the Annealing temperature= T_m-5°C), length: 21-25 bp, Eurofins Genomic primer maximum annealing score: <14. It was aimed to have more than one 1 C or G residue at the 3' end of the primer to ensure strong template binding. RT primers had a T_m of. For exon-spanning primers, ~25% of the primer was complimentary to the first exon and 75% to the second. Mutagenesis primers were designed individually according to the desired mutation. For each mutagenesis, two 25-35mer complementary primers were designed. The desired mutation (deletion or insertion) was placed in the middle of the primer with ~10–15 bp of the correct sequence on both sides.

2.2.3.8. BioEdit

BioEdit is a sequence alignment editor allowing alignment and manipulation of sequences. The automated ClustalW alignment feature was used to align Sanger sequenced TCR chains during the cloning process to ensure the absence of mutations. It was also used to determine the success of mutagenesis reactions.

3. Results

3.1. Development of an NGS-based platform for high-throughput detection of αβTCRs from single murine T cells

3.1.1. Design of the scTCRseq platform

There are two contrasting approaches for TCR analysis. Although single cell (sc)TCR cloning can detect paired αβTCR chains, cloning and Sanger sequencing limits throughput and cost-efficiency. In contrast, NGS-based TCR repertoire profiling of bulk T cell samples is high-throughput, but αβTCR chain pairing information is lost.

We sought to integrate the high-throughput nature of TCR repertoire profiling technologies with the αβTCR chain pairing ability of single cell approaches into one affordable scTCR sequencing (scTCRseq) platform. To ensure high-throughput, affordability and applicability, the following platform attributes were considered during platform design:

- Compatibility with NGS technology
- Early sample barcoding and plate multiplexing to minimize hands-on work and reagent costs
- Implementation of template-switching (TS) Polymerase chain reaction technology and shared primer tags to minimize use of different primers in later PCR steps
- Minimize use of specialized equipment to increase platform applicability
- Automatable wet-lab steps
- Automated TCR sequence retrieval and data analysis pipeline
- Ability to clone full-length TCR sequences of interest after data analysis for TCR downstream applications.

A workflow was devised (**Figure 3.1A**) in which single murine T cells are sorted by flow-cytometry into 96-well plates based on cytokine secretion or surface marker expression (**1**). One cell is sorted per well into a volume of 50 μL containing 10^3 human CCD-1079Sk carrier fibroblasts. The content of each well is harvested into a 1.5 mL tube for lysis. Total RNA extraction is performed using a Silica-membrane based spin-column method (**2**). The RNA eluate is concentrated using a speed-vacuum concentrator since the subsequent TS cDNA synthesis step is performed in a 10 μL total reaction volume (**3**). First strand cDNA synthesis is performed using *TRAC/TRBC* gene-targeting primers (**4, 5**).

A well-of-origin-based sample barcoding and pooling approach is adopted which uses a unique triple index combination to maintain the corresponding αβTCR chain information (**Figure 3.1A and 1B**). Row-specific barcodes (RBC) are incorporated at the cDNA 5' end in the first PCR (**6**). The eight samples of each column are pooled (**7**) generating twelve pools per plate (**8**). These pools are purified using magnetic beads (**9**). In a second PCR, the TCR chains of each column pool receive their respective column-specific barcode (CBC) at the 3' cDNA end (**10**). The remaining row is then pooled into one final sample (**11, 12**). This sample contains the amplicons from all cells on the plate and is purified once more (**13**). During NGS library preparation, each plate-derived sample is given a plate-specific barcode (PBC) using unique Illumina indices, allowing further multiplexing of up to 96 plates into one library (**14, 15**). 2x300 bp paired-end sequencing is performed using an Illumina MiSeq device (**16**). If two identified α and β TCR chains originated from the same cell (well), then they would possess the same PBC, RBC and CBC combination (**Figure 3.1C**).
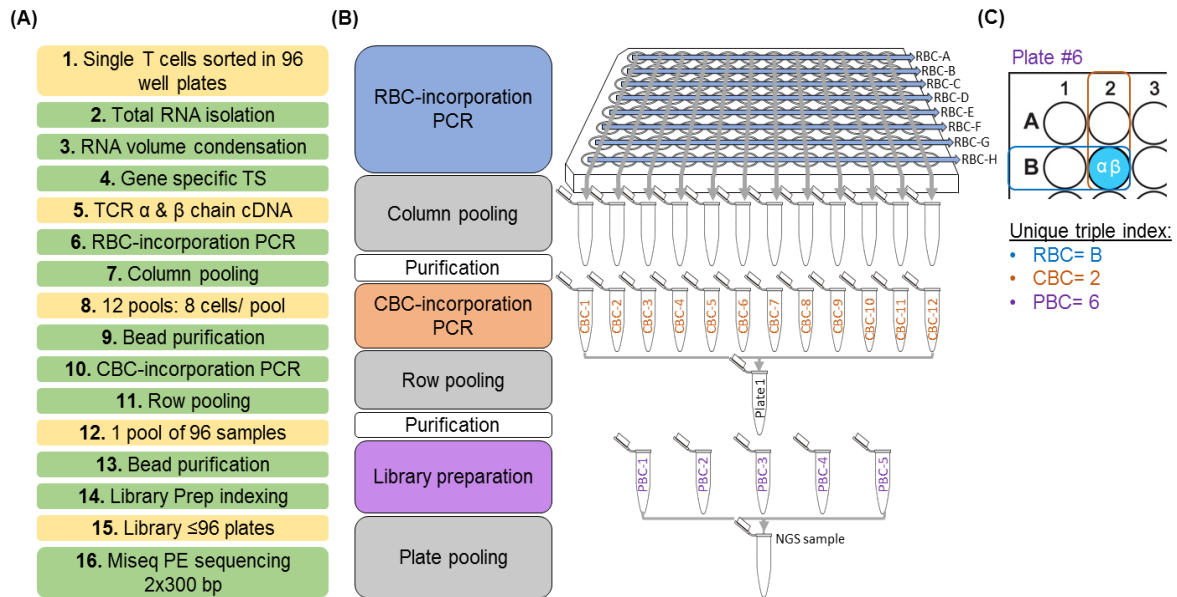
Figure 3.1| Schematic representation of the devised scTCRseq platform. (A) Workflow of the platform. Green and yellow text boxes indicate workflow action or input/output material, respectively. **(B)** Barcoding and pooling strategy. **(C)** Unique triple indexing of a representative αβTCR chain pair in plate 6, well 2A. TS: template-switching, bp: Base pairs, PE: Paired-End sequencing.

The workflow at the molecular level is illustrated in **Figure 3.2**. Every TCR V-(D)-J-C gene rearrangement can be reverse transcribed using only two gene-specific primers (GSP) targeting the *TRAC* and *TRBC* genes (see section 3.1.2). Due to TS sequence incorporation at the cDNA 5' end, no *TRAV/TRBV*-specific forward primers are required in the subsequent PCRs. Additionally, the use of GSPs during RT minimizes background that may be amplified during the following PCRs in comparison to a dT-priming approach.

In the RBC-incorporation PCR, the forward primer is composed of the TS sequence, the respective five base pair (bp) long RBC and a shared primer tag (Tag-A). The reverse primers are nested *TRAC* and *TRBC* GSPs with another shared primer tag (Tag-B) attached. The nested primers should enhance the specificity of this PCR. The incorporation of Tag-A and Tag-B allows the use of the same primers for all TCR chains in the subsequent PCR simplifying the workflow. In the CBC-incorporation PCR, a reverse primer composed of Tag-B and the five bp long CBC is used. The forward primers are composed of Tag-A and 4 or 5 random nucleotides at the 5' end essential for accurate fluorophoric camera detection during NGS.
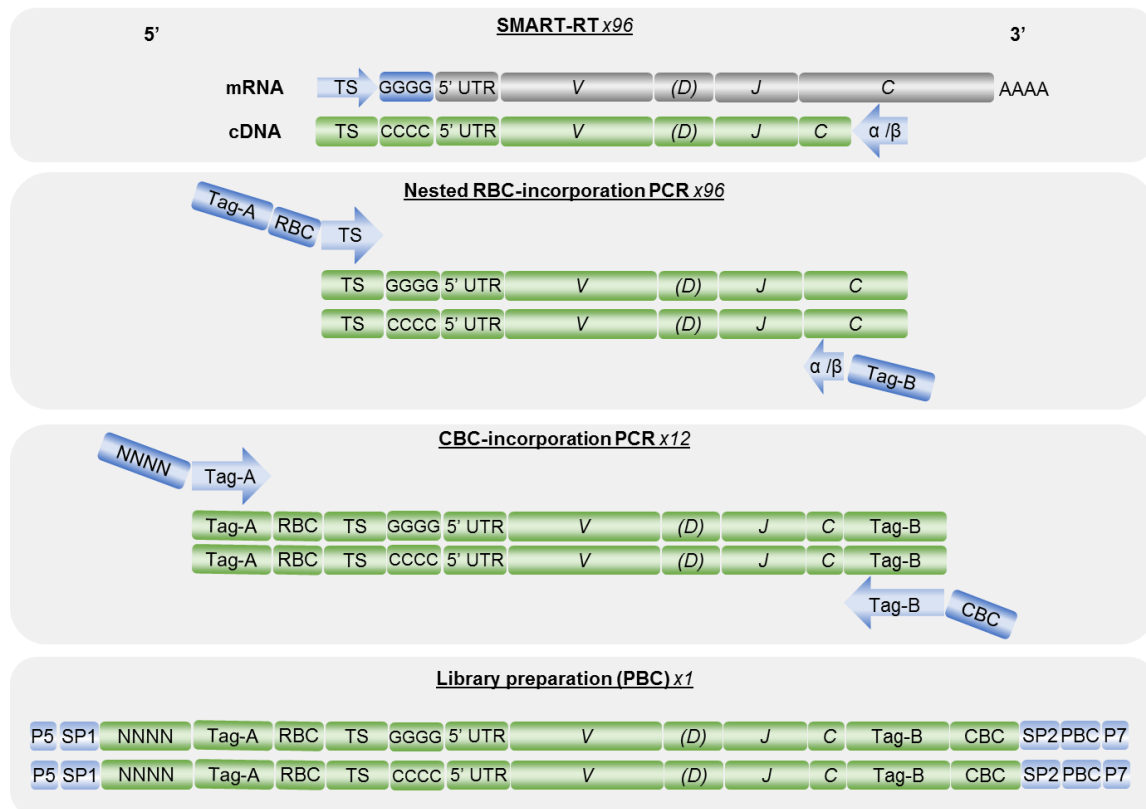
Figure 3.2| Schematic representation of devised workflow at the molecular level. Following TCR gene-specific priming and reverse transcription, RBCs, CBCs and PBCs are incorporated during the first PCR, second PCR and the library preparation, respectively. V: Variable gene, D: Diversity gene, J: Joining gene, C: Constant gene, TS: Template-switch sequence, CCCC: Cytosines added by the terminal transferase activity of the MMLV RTase, GGGG: Riboguanines in TS primer and Guanines in cDNA, UTR: Untranslated region, RBC: Row-specific barcode, CBC: Column-specific barcode, PBC: Plate-specific barcode (illumina indices), NNNN: 5' variation, SP1/2: sequencing primer 1/2. Primers and adaptors shown in blue, mRNA in grey and cDNA in green.

3.1.2. Establishment of the reverse transcription and PCR steps using mouse spleen RNA

For reverse transcription (RT), one primer was designed for each TCR constant gene (RT-TRAC and RT-TRBC). They bind within the first exon of their respective genes. For the RBC-incorporation PCR, two nested reverse primer variants (v) were designed for each TCR constant gene for testing (TRACv1, TRACv2, TRBCv1 and TRBCv2). The RT and PCR *TRBC*-specific primers bind both *TRBC1* and *TRBC2* genes. To design the Tag-A and Tag-B sequences, fifty different 25 bp long sequences with a 50-60 % GC-content and a melting temperature above 65 °C with no predicted binding to the murine transcriptome were designed. Sequence-130 and sequence-146 were chosen based on their predicted secondary structures (low hairpin and low self and hetero-dimer formation scores) and were tested as Tag-A and Tag-B, respectively. The RBCs and CBCs were adopted from Han *et al.,* 2015[128]. These barcodes differ by two bp from each other. Thereby at least two PCR or sequencing errors in a barcode are required for it to be mistaken for another one; reducing the chances of false barcode detection.

The first experiment focused on: 1) testing the functionality of the TCR GSPs for RT, 2) testing and selecting the best nested GSPs (without Tag-B attached) for the subsequent PCR and 3) testing that the selected nested GSPs are not affected by attachment of the Tag-B. Total RNA was isolated from mouse splenocytes. TS cDNA synthesis was performed using 250 ng RNA and the designed GSPs. The first strand cDNA was diluted ten-fold and aliquots of 10 % were used for the subsequent PCR. The TS primer was used without the barcode and Tag-A attached as to test only one variable at a time. PCRs were performed using 30 cycles, PfuUltra HS High-Fidelity DNA Polymerase and 66 °C annealing temperature. Both nested PCR primer variants for each TCR constant gene could amplify the TCR cDNA without and with the Tag-B after RT with the designed GSPs **(Figure 3.3A)**. Amplicon sizes ranged from 550-650 bp in length (with Tag-B) due to the differently rearranged TCR genes in the spleen RNA. TRACv1-Tag-B and TRBCv2-Tag-B showed stronger amplification of α and β TCR chains, respectively, compared to their counterparts and were therefore chosen for all further experiments.

The next experiment focused on testing that the functionality of the TS primer is not affected by the attachment of 1) the Tag-A and 2) the different row barcodes. Twelve barcode variants were tested. Eight functional RBC primers that showed strong and robust amplification were selected as RBCA to H **(Figure 3.3B).**

(A)

(B)
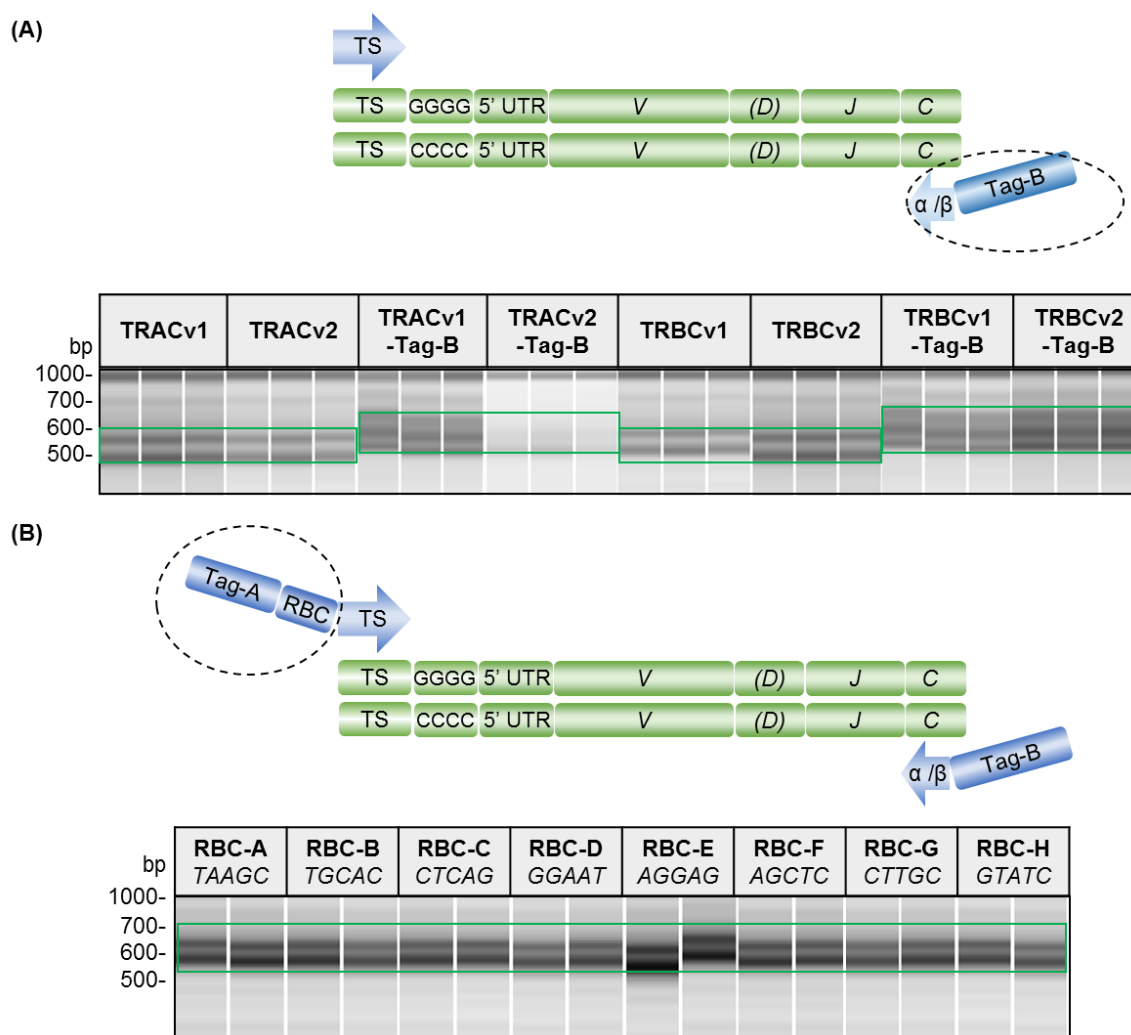
Figure 3.3| Amplification of mouse spleen RNA using the designed nested TCR GSPs for RT and PCR. Qiaxcel images of PCR amplicons (10uL/lane). TS RT was performed in a volume of 10 μl using the designed GSPs. First strand cDNA was pooled and diluted 10-fold. PCRs were performed using 30 cycles, PfuUltra HS High-Fidelity DNA Polymerase, 66 °C annealing temperature and 10 μL diluted first strand cDNA as template. **(A)** The forward TS primer was used together with two reverse primer versions specific for the TRAC and TRBC genes, respectively. The latter were tested with and without Tag-A attached. **(B)** The forward TS primer variants with respective 5 bp RBC (A-H) and Tag-A attached were tested together with the previously chosen TRACv1-Tag-B and TRBCv2-Tag-B primers. The final functional Tag-A-RBC-TS forward primer panel set is shown. The five bp sequences indicate the chosen RBCs. V: Variable gene, D: Diversity gene, J: Joining gene, C: Constant gene, TS: Template-switch sequence, CCCC: Cytosines added by the terminal transferase activity of the MMLV RTase, GGGG: Guanines, RBC: Row-specific barcode.

After column pooling, each pool is purified using magnetic bead-based size selection. Based on the obtained band ranges in **Figure 3.3B**, a bead: template ratio of 0.75 was used and elution was performed in 30 μL water. Following purification, each column pool receives a unique CBC in the subsequent CBC-incorporation PCR. Therefore, the

functionality of the reverse Tag-B-CBC primer variants was tested. 10 µL of the respective purified column pool were used as template for the CBC-incorporation PCR. The CBC-incorporation PCRs were performed using 30 cycles, the PfuUltra II Fusion HS DNA Polymerase and 64 °C annealing temperature. **Figure 3.4** shows the successful amplification using all twelve CBCs.
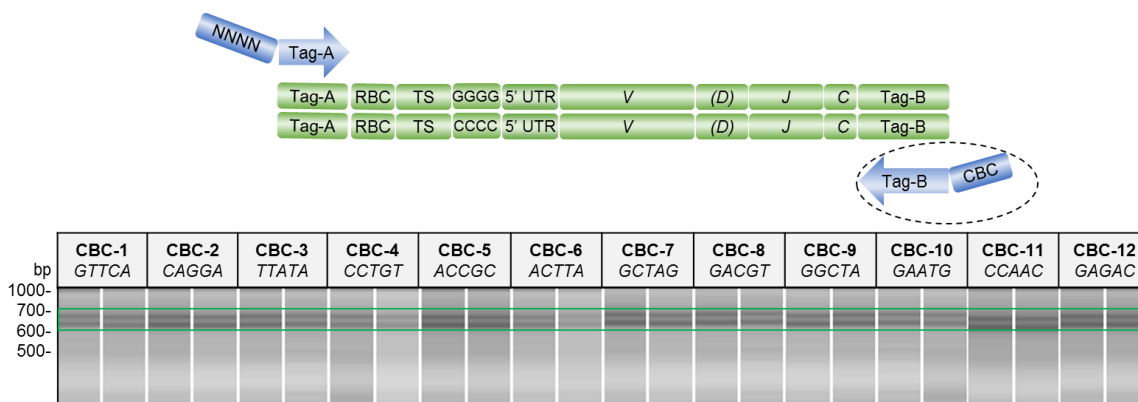
Figure 3.4| Amplification of purified column pools using a reverse Tag-B-CBC primer panel. Qiaxcel images of PCR amplicons (10uL/lane). PCRs were performed using 30 cycles and the PfuUltra II Fusion HS DNA Polymerase and 64 °C annealing temperature. The forward NNNN-Tag-A primer was used together with reverse Tag-B-CBC1-12 primers. 10 µL purified column pool was used as template. The five bp sequences indicate the chosen CBCs. V: Variable gene, D: Diversity gene, J: Joining gene, C: Constant gene, TS: Template-switch sequence, CCCC: Cytosines added by the terminal transferase activity of the MMLV RTase, GGGG: Guanines, RBC: Row-specific barcode, CBC: Column-specific barcode.

3.1.3. TCRseq test run using a TCR chain-encoding IVT RNA library as input

The next experiment focused on down titrating the input RNA amounts and testing the barcoding approach using IVT RNA encoding different TCR sequences. The used IVT RNA library consisted of 12 unique α and β TCR chains each **(supplementary section 6.2)**. In the starting plate, each well received 0.25 pg IVT RNA of a unique α and β TCR chain combination **(Figure 3.5A).** The experiment workflow is depicted in **Figure 3.5B**. The chosen parameters led to the successful amplification of the TCR-encoding IVT-RNA **(Figure 3.5C)**. Amplicons were 500-700 bp in length after the CBC-incorporation PCR. To sequence the amplicons, 25 ng of each CBC-incorporation PCR product was cloned into the pJET1.2 blunt end vector. 96 colonies (eight colonies/column pool) were Sanger sequenced to get a random selection of amplicons from the plate. Barcode demultiplexing was performed manually and TCR sequence characterization was performed using IMGT/V-Quest.

93/96 (97%) sequenced colonies contained a TCR chain insert. 85/93 (91%) insert-positive sequences had a RBC, 90 (97%) had a CBC and 81 (88%) had both. Altogether

42 unique sequences were identified and allocated to 25 different wells **(Figure 3.5A)**. All TCR chains were allocated to the correct well of origin. Paired αβTCR chains were detected in 16/25 wells (64%). In 6/25 (24%) and 2/25 wells (8%) only the α TCR or β TCR chain, was detected, respectively. One well contained a TCR chain that was not familiar to the generated IVT RNA library. For all demultiplexed sequences, no TCR sequence errors were detected when aligned to the IMGT database.

In conclusion, Sanger sequencing a random sample of the platform output confirmed the feasibility of amplifying 0.25 pg of spiked-in αβTCR-encoding IVT RNA as well as the functionality of the barcoding system.
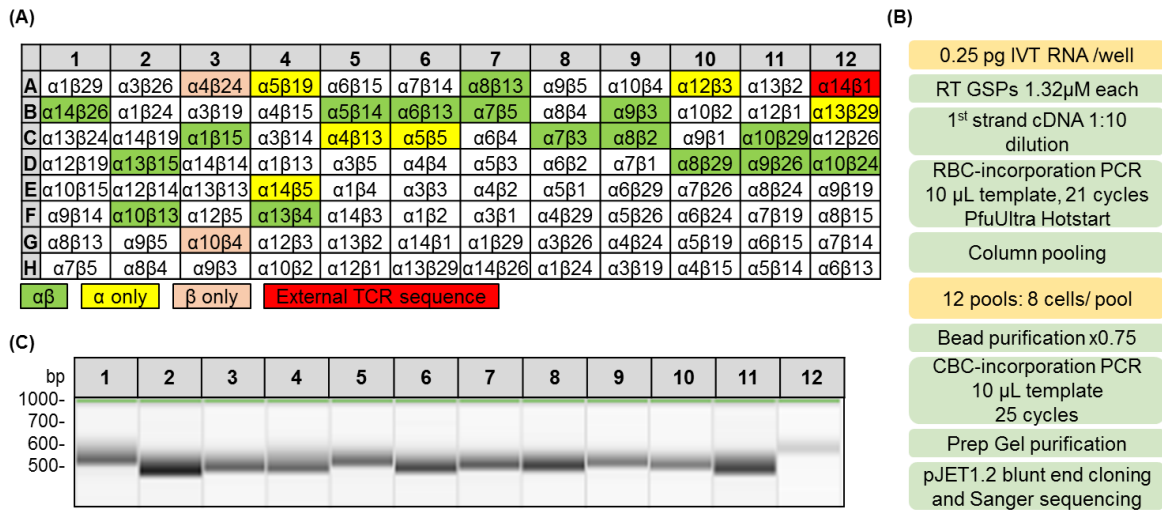
Figure 3.5| Successful amplification of a αβTCR-encoding in vitro transcribed RNA library. (A) Platform input and output sequences. Spiked-in α and β TCR chain encoding IVT RNA combinations indicated. 0.25 pg of IVT RNA of each chain were used per well. **(B)** Performed platform workflow and chosen parameters. 25 ng DNA of each CBC-incorporation PCR product (column pool) was cloned into the pJET1.2 blunt end vector for subsequent Sanger sequencing analysis. Barcode detection and demultiplexing was performed manually and TCR sequence characterization was performed using IMGT/V-Quest. **(C)** Amplified cDNA library after the CBC-incorporation PCR. Each lane represents the product of the respective CBC-incorporation PCR. Qiaxcel images of PCR amplicons (10uL/lane). IVT RNA: In vitro transcribed RNA, RBC: Row-specific barcode, CBC: Column-specific barcode, Prep Gel: Preparative agarose gel electrophoresis.

3.1.4. First application of the platform using single T cells

To test the platform for the intended single cell level, single T cells were sorted into a 96-well plate. The workflow was performed using the same PCR parameters as in section 3.1.3. However, additional cycles were added to each PCR (24 and 27 final cycle number, respectively) due to the lower TCR mRNA RNA amounts in single T cells. **Figure 3.6A** shows the successful amplification of the single cell-derived TCR mRNA after the CBC-incorporation PCR. Different band sizes could be seen for each column pool (600 +/- 100

bp), possibly indicating the amplification of the different lengthed TCR chains from the different cells within each column pool. 25 ng DNA of each CBC-incorporation PCR product were cloned, sequenced and analyzed as described in section **3.1.3**.

84/96 (87 %) sequences colonies contained a RBC, CBC and a TCR sequence. 57 unique sequences were identified and allocated to 33 different wells **(Figure 3.6B)**. Paired αβTCR chains were found in 6/33 wells (18 %). 11/33 (33 %), 7/33 (21 %) and 5/33 (15 %) wells possessed only a productive α TCR chain, a productive β TCR chain or an unproductive α TCR chain, respectively. Due to biallelic expression, up to two α TCR and one β TCR chains are generally expected. 4/33 wells (12 %) possessed more than two α TCR chains indicating potential noise caused by well-to-well contamination or barcoding malfunction.

Taken together, Sanger sequencing confirmed the ability of the technology to amplify TCR genes from single cells, however, with a low detection rate and the presence of noise. Furthermore, the manual demultiplexing and TCR sequence retrieval were time consuming. Therefore, subsequent work focused on:

- Developing an automated data analysis pipeline for the demultiplexing and retrieval of TCR sequences from NGS data **(Section 3.2)**.
- Platform optimization to enhance αβTCR detection rates and minimize platform noise, costs and hands-on-work time **(Section 3.3)**.

(A)

(B)

	1	2	3	4	5	6	7	8	9	10	11	12
A	α1	β1	α2 α3 α4 α5	α6 α7	α8 α9 α10 α11 α12		α13 β2		β3	α6 α14		
B	α15 β2 β4		α16				α17					
C							α18 β5	α19				
D		β4			α20				α21			
E				α22				α23 α24 α25 α26				β6
F		α27 β7										
G				α28	α29	α30				α31 β4	α32 α33 α34	
H	β5	β7	α35	α36	α37 α38 α39				β8		α39	α40 α41 β13
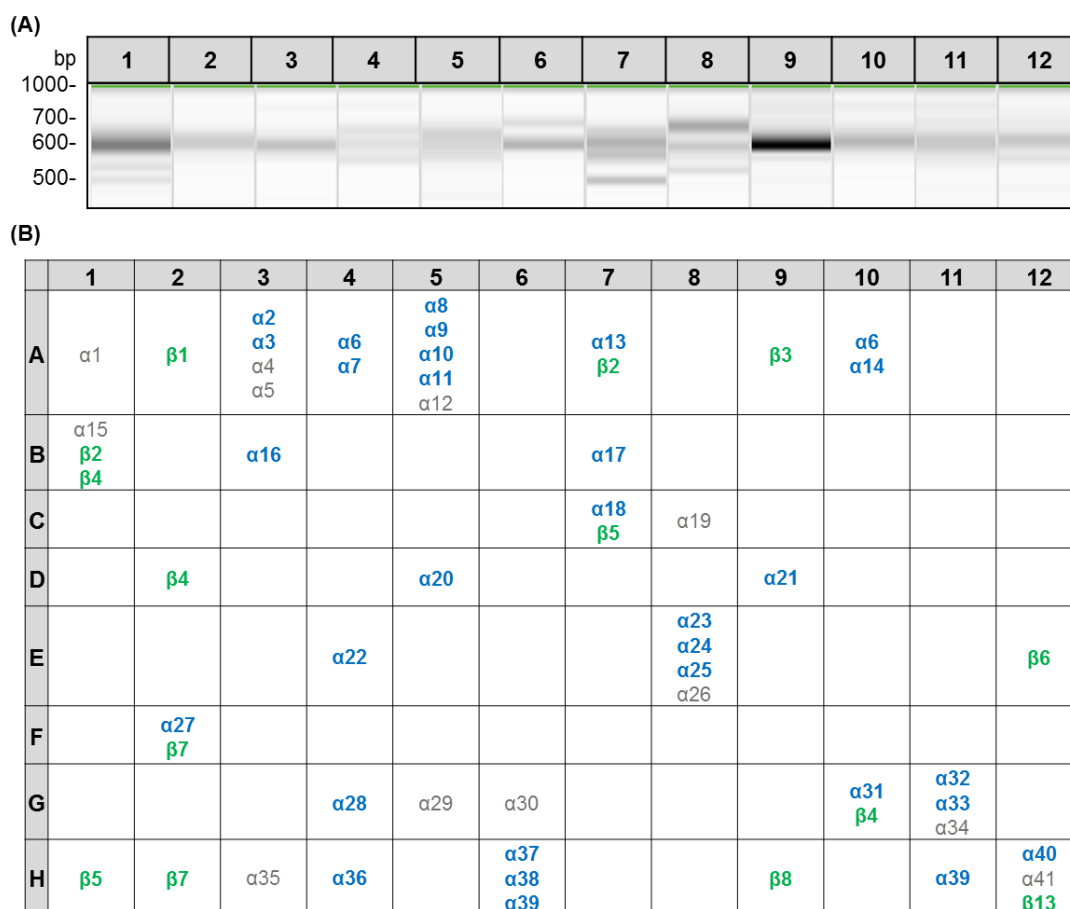
Figure 3.6| Successful amplification of single cell derived-TCR mRNA. (A) Amplification of single cell derived-TCR cDNA after the CBC-incorporation PCRs of the respective columns. Qiaxcel images of PCR amplicons (10uL/lane). CD8+CD137+ T cells were sorted. RBC-incorporation PCR: 24 cycles. CBC-incorporation PCR: 28 cycles. **(B)** Identification of single cell derived TCR chains using Sanger sequencing. 25 ng DNA of each CBC-incorporation PCR product was cloned blunt end into the pJET1.2 vector for subsequent Sanger sequencing. Barcode detection and demultiplexing was performed manually and TCR sequence characterization was performed using IMGT/V-Quest. Blue, green and grey font represent productive α, productive β and unproductive chains, respectively. α: α TCR chain; β: β TCR chain; 1-41: TCR chain identity.

3.2. Development of an automated NGS data analysis pipeline for retrieval of paired αβTCR sequences

Manual demultiplexing and TCR sequence analysis was highly time consuming. Therefore, there was a need for an automated bioinformatic pipeline for demultiplexing and TCR sequence retrieval (**Figure 3.7A**). 2x300 bp NGS reads are demultiplexed based on their unique Illumina primers. Sequences of each plate are allocated to the well of origin by identifying the respective RBCs and CBCs. V(D)J genes of the TCR chains in

each well and their read counts are determined using MiXCR, a tool for analysis of TCR sequencing data. The read counts are normalized based on plate, row and column total read count to eliminate technically induced read count variation. A developed TCR chain filter script reduces technical noise and determines the paired αβTCR chains. Figures and excel files summarizing the data are generated allowing final manual data reviewing and analysis. **Figure 3.7B** shows the respective scripts developed and the directories in which they were performed. To help set up the pipeline, the CBC-incorporation PCR products from the first single cell experiment (section 3.1.4) were sequenced 2x300 bp using the MiSeq with a read depth of $1x10^6$ reads and used as a sample data set.
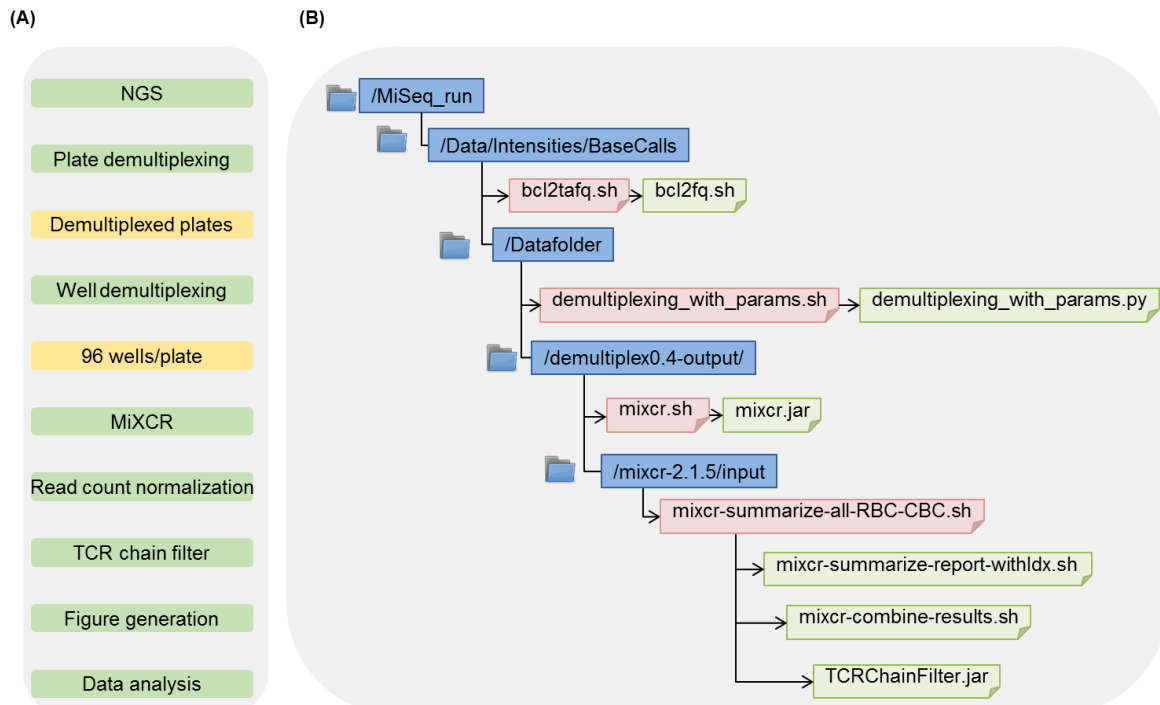
Figure 3.7| The developed automated NGS data analysis pipeline for the retrieval of paired αβTCR sequences. (A) Workflow and **(B)** bioinformatic pipeline scripts and directories. Well demultiplexing was implemented in a Python script using the integrated development environment PyCharm. The TCR Chain Filter was developed in Java. In-house tools and the external tools were connected via Bash into a pipeline to automate most steps of the analysis. On this server, all analyses were carried out accessing it via WinSCP and a terminal application (MobaXterm).

3.2.1. Establishment of a barcode demultiplexing and adaptor trimming script

Manual demultiplexing was highly time consuming. In this regard, a script was developed to allocate NGS sequences to the well of origin by identifying the respective RBCs and CBCs (**Figure 3.8**). Following the (1) detection of the TS primer sequence, (2) five bp upstream are marked as the RBC. (3) The Tag-B is then located in the adjacent read. The

TS primer and Tag-B sequence are searched for in both directions. Only half of the TS primer and Tag-B sequence are searched for allowing assignment of reads with potential PCR sequencing errors in those regions. (4) The CBC is defined as the five bp downstream of the Tag-B sequence. For the barcodes, one mismatch is allowed. Following demultiplexing, the sequences of each well are trimmed to leave behind the TCR-encoding sequence.
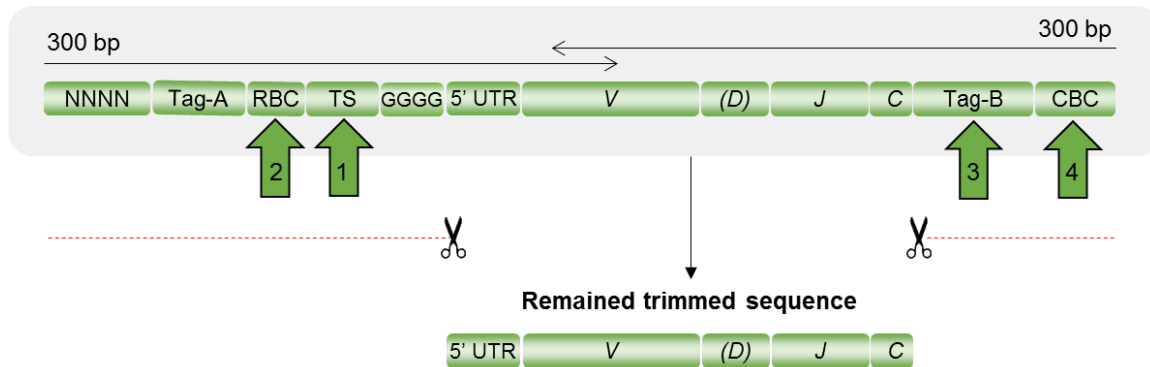
Figure 3.8| Establishment of a barcode demultiplexing and adaptor trimming script. Sequenced amplicon is shown. Sequences 1, 2, 3 and 4 are searched for during the demultiplexing process. Sequences are trimmed after being demultiplexed. V: Variable gene, D: Diversity gene, J: Joining gene, C: Constant gene, TS: Template-switch sequence, GGGG: Guanines, RBC: Row-specific barcode, CBC: Column-specific barcode.

To confirm the presence and determine the locations of the shared primer tags and barcodes in the NGS reads, the different motifs were mapped on all amplicons from the first NGS run (from section 3.1.4). The locations of the primer and tag sequences were as expected **(Figure 3.9)**. Only 11 % of the paired-end sequences could be overlapped in accordance with the PCR amplicons which where 600 +/- 100 bp. The regions required to characterize the barcodes and primers as well as the TCR chain CDR3 regions existed in the beginning and middle of the paired-end reads, respectively (high quality-sequencing region).
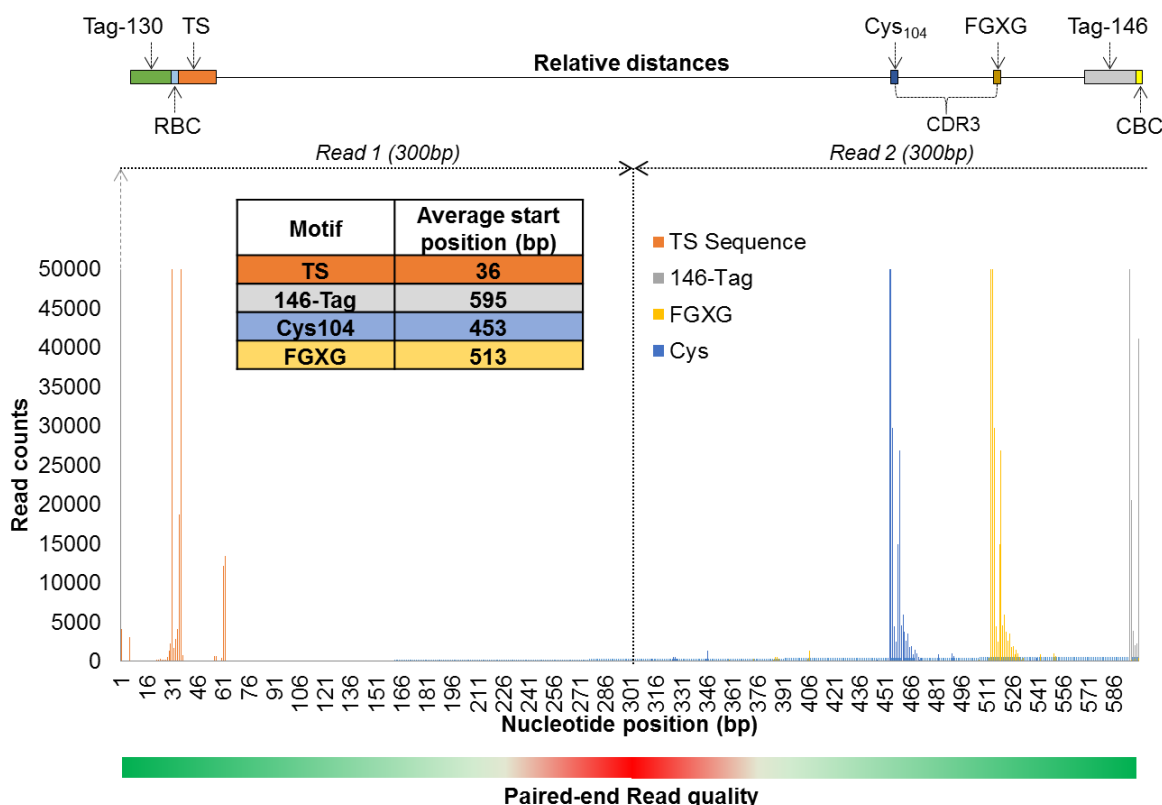
Figure 3.9| The CDR3, barcodes and primers exist in the high quality-sequencing region. TCR motifs, shared primer tags and barcodes were mapped on the NGS-sequences. Platform products from the first single cell trial were used. The transition of high quality (green) to low quality (red) base positions is shown in the bottom bar. TS: Template-switch sequence, RBC: Row-specific barcode, CBC: Column-specific barcode; bp: base pairs. CDR3 is determined as the region between the Cysteine$_{104}$ and the Phenylalanine of the FGXG motif; 130-Tag: Tag-A; 146-Tag: Tag-B.

3.2.2. Modification of the MiXCR TCR repertoire analysis tool for single cell analysis and confirmation of noise

Following well demultiplexing, α and β TCR chains of each T cell and their read counts are determined. MiXCR is an available computing tool that processes big immunome repertoire data from raw sequences to clonotypes with read counts[118]. It is composed of three main steps: (1) aligning sequencing reads to TRAV/TRBV, TRBD and TRAJ/TRBJ genes, (2) assembling clonotypes to extract gene regions e.g. CDR3, (3) exporting clones to a readable text file. Since, MiXCR is a program for bulk T cell repertoire profiling, multiple parameters were modified to tailor it for single cell application (**Table 3.1**).

Parameter	Value	Description
OParameters.geneFeatureToAlign	VTranscript	In order to utilize information from both reads and let MiXCR use the 5' sequence regions to align the V gene 5'UTRs for more accurate TCR detection.
OjParameters.parameters.floating RightBound	False	Specifies whether right bound of alignment is fixed or float: if set to false, the right bound of either target or query will be aligned. Since the target amplicons have no primer sequences in the J gene (i.e. library was amplified using TCR constant-specific primers) the value was set to false to increase J gene identification accuracy and overall specificity of alignments. Default: True
OcParameters.parameters.minAli gnmentLength	10	Minimal length of aligned region. Default= 15. Was decreased to increase TRJ gene detection which was low.
OcParameters.parameters.absolut eMinScore	30	Minimal score of alignments: alignments with smaller score will be dropped. Default=40. Was decreased to increase alignment rates.
OallowChimeras	False	Chimeric alignment is defined as as having V, J or C genes from the incompatible chains, e.g. TRBV / TRAJ or / TRBC, etc…).

Table 3.1| MiXCR parameter optimization for use with single T cells.

Running the demultiplexing and MiXCR scripts using the first sample data set (from section 3.1.4) indicated the successful amplification and possible automated retrieval of single cell derived TCR sequences **(Figure 3.10A)**. The MiXCR output of all wells was directly combined into one Excel file using a data merging script.

NGS data confirmed the presence of more α or β TCR chains than expected in most wells as seen with in the Sanger sequencing data (Section 3.1.4). In addition to the presence of one to two chains with high read counts in a specific well, the results showed the presence of other TCR chains in the same well with lower read counts, which we considered as noise **(Figure 3.10A)**. Furthermore, in depth analysis revealed that the same TCR chains existed in more than one well in the same row or column hinting at systematic technical noise (**Figure 3.10B**). Despite the successful amplification of single cell-derived α and β TCR chains, the difficulty in detecting the real αβTCR chain pair emphasized the need for addressing the issue of noise by platform wet-lab optimization as well as the development of a TCR chain noise elimination filter.

(A)

CBC	RBC	Read count	V	D	J	CDR3
1	A	406	TRBV26*01	TRBD2*01	TRBJ2-5*01	CASSLTGGAQDTQYF
1	A	288	TRAV14N-2*01		TRAJ4*01	CAASLWAGSFNKLTF
1	A	194	TRBV14*01	TRBD2*01	TRBJ2-1*01	CASSLTLGGNYAEQFF
1	A	15	TRBV13-2*01	TRBD2*01	TRBJ2-5*01	CASGFQGDTQYF
1	A	11	TRBV19*01	TRBD2*01	TRBJ2-7*01	CASSDRVYEQYF
1	B	13957	TRBV26*01	TRBD2*01	TRBJ2-5*01	CASSLTGGAQDTQYF
1	B	10783	TRAV14N-2*01		TRAJ4*01	CAASLWAGSFNKLTF
1	B	6157	TRBV14*01	TRBD2*01	TRBJ2-1*01	CASSLTLGGNYAEQFF
1	B	45	TRAV7D-4*01		TRAJ26*01	CAANNYAQGLTF
1	C	3205	TRBV4*02	TRBD1*01	TRBJ1-1*01	CASSSWDSTEVFF
1	C	3112	TRAV7-3*01		TRAJ31*02	CASSSWDSTEVFF
1	C	2372	TRAV5D-4*01		TRAJ49*01	CAAKGTGYQNFYF
1	C	120	TRBV26*01	TRBD2*01	TRBJ2-5*01	CASSLTGGAQDTQYF
1	C	95	TRAV14N-2*01		TRAJ4*01	CAASLWAGSFNKLTF
1	D	2186	TRAV12D-2*02		TRAJ22*01	CALIPGSWQLIF
1	D	922	TRBV4*02	TRBD1*01	TRBJ1-1*01	CASSSWDSTEVFF
1	D	895	TRAV7-3*01		TRAJ31*02	CASSSWDSTEVFF
1	D	89	TRAV14N-2*01		TRAJ4*01	CAASLWAGSFNKLTF
1	D	17	TRBV26*01	TRBD2*01	TRBJ2-5*01	CASSLTGGAQDTQYF

(B)

CBC	RBC	Read count	CDR3
1	B	13957	CASSLTGGAQDTQYF
1	F	2070	CASSLTGGAQDTQYF
11	B	1922	CASSLTGGAQDTQYF
12	B	1650	CASSLTGGAQDTQYF
1	G	1283	CASSLTGGAQDTQYF
9	B	811	CASSLTGGAQDTQYF
3	B	695	CASSLTGGAQDTQYF
1	E	580	CASSLTGGAQDTQYF

Figure 3.10| Successful single cell-derived TCR chain retrieval from the NGS sequences. (A) Example of single cell derived TCR chains identified using MiXCR. Platform products from the first single cell trial were used. **(B)** Detection of row and column-based systematic technical noise. An example is shown for CDR3 CASSLTGGAQDTQYF. RBC: Row-specific barcode, CBC: Column-specific barcode, V: Variable gene, D: Diversity gene, J: Joining gene, CDR3 Complementary determining region 3.

3.2.3.   Construction of a TCR chain filter script

Due to biallelic expression, up to two rearranged α TCR and one rearranged β TCR chains (in a few cases two) are expected. Additional TCR chains can be considered as noise. Therefore, a TCR chain filter was developed to separate technically induced noise from real TCR chains **(Figure 3.11)**. α and β TCR chains of each well were processed separately in this script. Non-productive chains were kept in the data set, but were excluded from subsequent cut-off filters. For each well, the chain with the highest read count was defined (for TCR α and β). Sequences then went through two cut-offs:

1. Since the presence of clones with high read counts generated noise in neighboring wells of the same row and column, a 'clone-specific base clipping filter' was developed. Any read of a unique clonotype below 2% of the highest read count of that clonotype in the same row or column was filtered-out.
2. Following that, a well-specific cut-off was adopted in which any read count smaller than 10 % of the highest chain in the well was removed.

A FilteredClones.xlsx file was generated containing the remaining TCR chains. All filtered-out and flagged clones were also kept to allow subsequent manual reviewing.
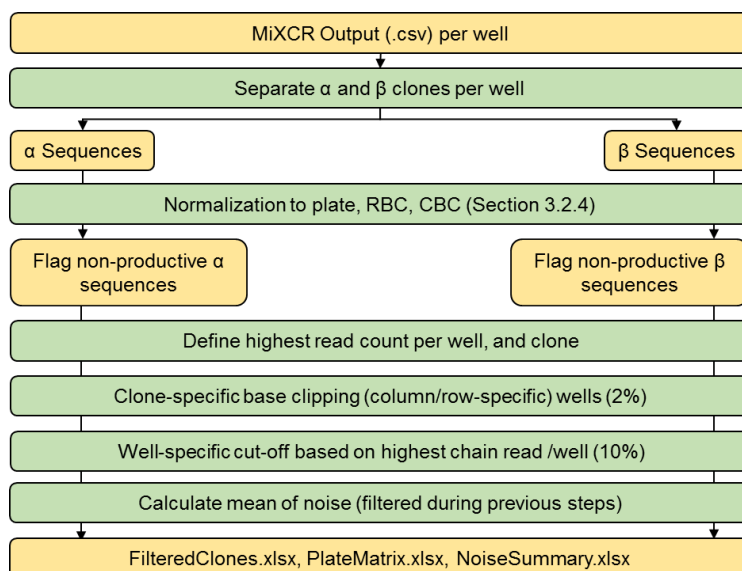
Figure 3.11| Development of a platform noise-elimination and TCR chain filter pipeline. α and β TCR chains are processed separately. Displayed cut-off values are the default and can be changed by the user. Green and yellow text boxes indicate pipeline action or input/output material, respectively. α: α TCR chains, β: β TCR chains, CDR3: Complementary determining region 3.

3.2.4. Development of a read count normalization script

To generate a homogenous read count level across all wells, all read counts were normalized to eliminate primer-induced read count variation between different plates, rows and columns **(Figure 3.12)**. The α and β TCR chains were normalized separately. Three subsequent normalization steps were performed: normalization by total plate read count ($R_{raw}>R_{plate}$), then by total row read count ($R_{plate} >R_{row}$) then by total column read count normalization ($R_{row}>R_{column}$). For each normalization step, the read count of each chain was divided by the total read counts per plate, RBC or CBC, respectively.
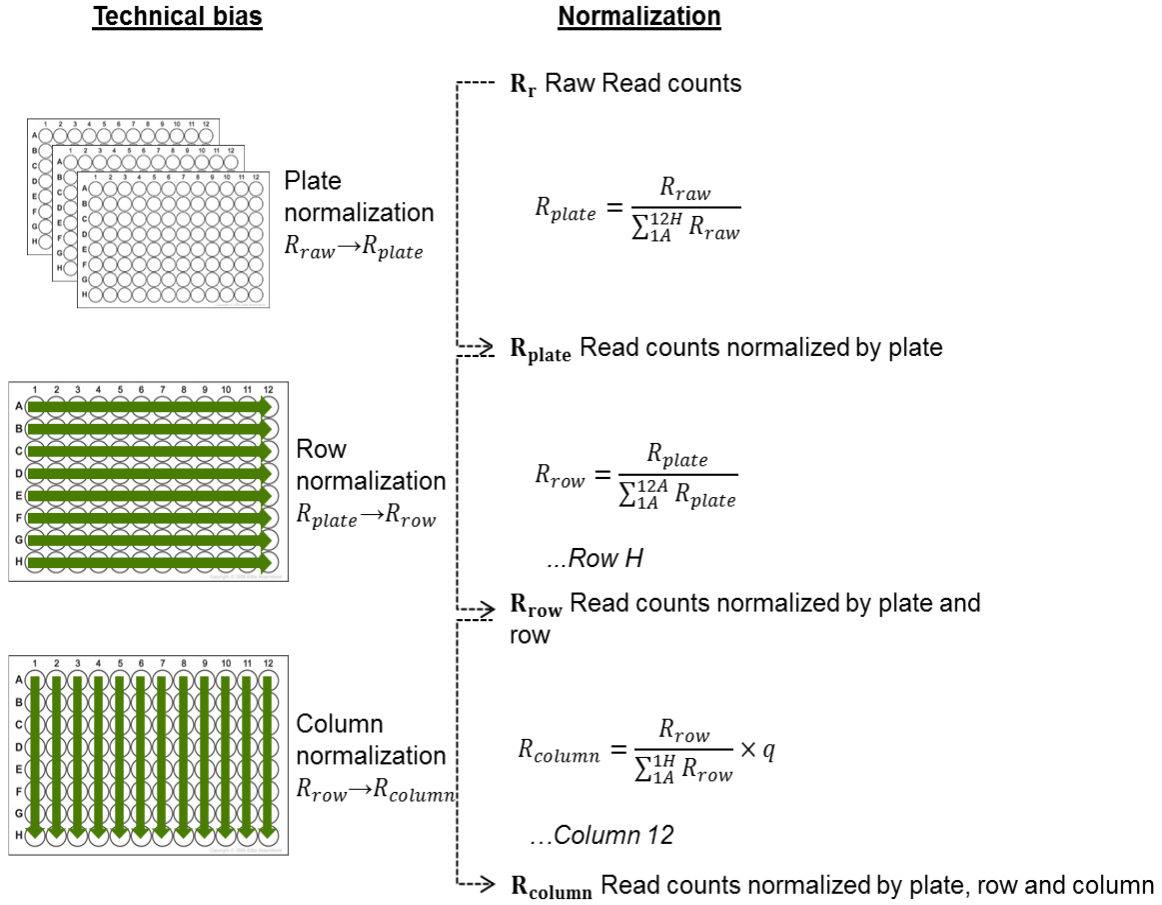
Figure 3.12| Development of a plate, row and column-based normalization script to eliminate technically induced read count variation. α and β TCR chains were normalized separately. Three normalization steps were performed consecutively for plate, column and row. For each normalization step, the read count of each chain was divided by the total read counts per plate then RBC then CBC, respectively. R_{raw}: raw read counts, R_{plate}: read counts normalized by plate, R_{row}: read counts normalized by plate and row, R_{column}: read counts normalized by plate, row and column.

3.3. Optimization of the scTCRseq platform

The initial experiments established a functional scTCRseq platform framework but with low αβTCR detection rates and the presence of noise. Accordingly, the subsequent work focused on:

- Increasing detection rates of paired αβTCR chains
- Minimizing platform read count noise
- Reducing platform hands-on-work
- Decreasing platform costs.

For efficient process optimization (and later platform performance evaluation), quantitative performance indicators were used **(Table 3.2)**. Only one platform parameter was modified at a time. Following data analysis, the modification was kept if it helped reach the aims mentioned above.

Experiment performance indicator	Quantitative reference parameter	Platform optimization	Performance evaluation
PCR yield *(and efficiency)*	PCR product concentrations *(compared to cost)*	✓	
αβTCR pair detection rate	Percentage of wells containing both TCR α and β chains compared to total number of wells	✓	✓
Platform accuracy	Percentage of sequences with PCR error rate (provided by MiXCR alignment report)	✓	
Noise levels	Read count of 3rd highest TCR α and β chain compared to the highest chain in the same well	✓	✓
Selectivity	False positives in negative control wells		✓
Sensitivity	True positives (spiked-in single T cell clones)		✓
Reproducibility	Technical duplicates	✓	
Barcoding performance	Read counts per barcode	✓	
TCR αβ chain bias	Read count of α vs β chains per well	✓	
Barcode bias	Comparison of each barcodes read counts throughout plate	✓	
Demultiplexing rates	Number of sequences with Illumina index, RBC and CBC compared to sequences with an Illumina index		✓
MiXCR alignment rate	The rate of demultiplexed sequences which are TCRs	✓	✓
External contamination	Presence of high read count chains that were previously observed in irrelevant experiments (*The possibility of such chains being public TCRs was taken into consideration*).	✓	✓

Table 3.2| Quantitative reference parameters used for platform optimization and measuring platform performance.

3.3.1. Direct cell-capturing into lysis buffer increases αβTCR detection rates and simplifies workflow

The RNA-isolation procedure was first addressed. Single T cells were sorted into a volume of 50 µL before being harvested, lysed and the RNA isolated using a spin-column based method. In addition to being time-consuming and costly, column-based RNA isolation leads to RNA loss[132]. A direct-cell capture protocol was devised in which cells are sorted into a small volume of cell lysis buffer making the mRNA accessible for RT while maintaining its integrity. The buffer should also not affect the downstream RT reaction. This optimization would eliminate the harvesting, column-based RNA isolation and speed-vacuum concentrator steps **(Figure 3.13A)**. We compared (1) the originally adopted column-based RNA isolation protocol to the direct cell-capturing approach using (2) round-bottom (CellStar; M9436) or (3) V-bottom (Bio-Rad; HSP9621) semi-skirted 96-

well plates. 20 single sorted T cells were used per condition. For the direct cell-capturing, 4 μL hypotonic cell-capturing buffer was used per well consisting of 0.2 % Triton X-100, RNase inhibitor (2 U/ μL) and 5 ng Poly (A) carrier RNA. Compared to the column-based RNA isolation where 2/20 paired αβTCR chains were identified (10 % αβ TCR detection rate), the cell-capturing approach using the round-bottom plate and the V-bottom plate led to the detection of 6/20 (30 %) and 12/20 (60 %) paired αβTCR chains, respectively **(Figure 3.13B)**.

An additional 56 T cells were processed to confirm the increased detection rates using direct cell-capture. αβTCR pairs were found in 34/56 T cells (61 %) confirming the detection rates of the first experiment **(Figure 3.13C)**. In 8/56 (14 %), 6/56 (11 %) and 5/56 wells (9 %) only a α TCR chain, a β TCR chain or no chains were detected, respectively. 42/56 (75 %) of all α TCR chains and 40/56 (71 %) of all β TCR chains were detected.

In conclusion, the direct cell-capture approach using the V-bottom Bio-Rad plates significantly increased the αβTCR detection rate six-fold and strongly reduced the hands-on-time. This approach was adopted for all further experiments.
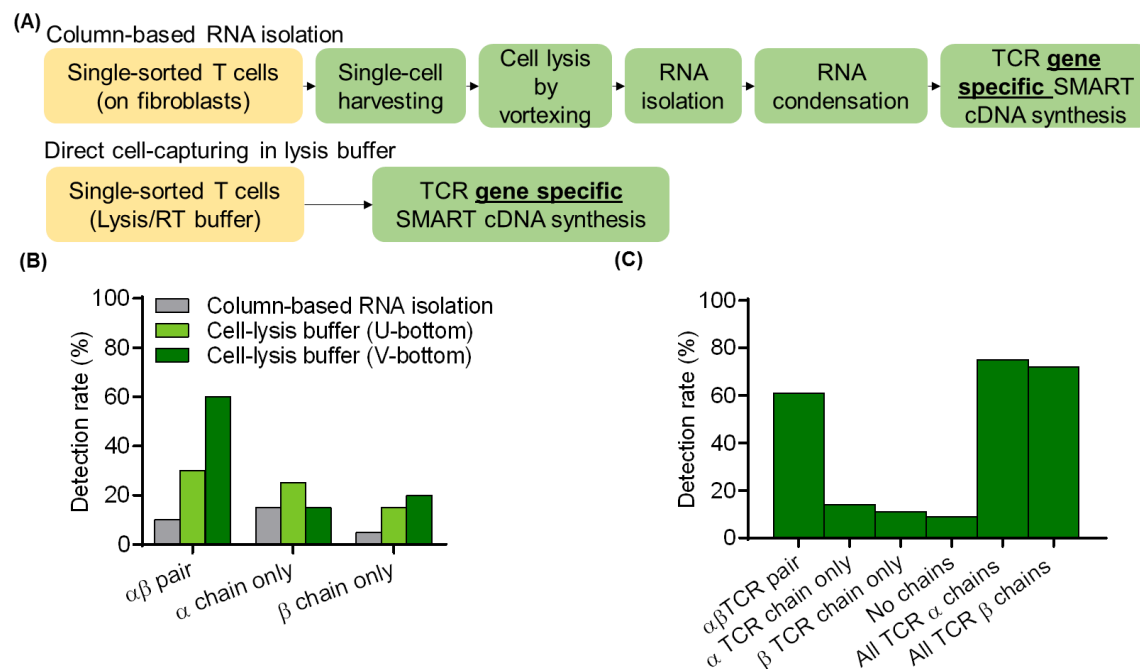
Figure 3.13| Direct cell-capturing in lysis buffer increases αβTCR detection rates and reduces platform duration and costs. (A) Comparison of the column-based RNA isolation and the direct cell-capturing approaches. **(B)** Direct cell-capturing increases αβTCR detection rates. 60 CD8[+] single T cells were sorted. 20 T cells were processed per condition: the originally adopted column-based RNA isolation protocol, direct cell-capturing approach using round-bottom or V-bottom 96-well plates. **(C)** Confirmation of the increased detection rates using a direct cell-capturing approach. 56 CD8[+] T cells obtained from another mouse were sorted and processed. For direct cell-capture the first strand cDNA was diluted 1:10

and 10 µL were used for the RBC-incorporation PCR. RBC and CBC-incorporation PCRs were performed using 25 and 28 cycles, respectively. All cells were sequenced with a read depth of at 14,000 reads.

3.3.2. Optimization of the reverse transcription reaction for robust detection of paired αβTCR chains

To further increase detection rates and minimize noise, efforts were made to optimize the RT reaction. In comparison to the β TCR chains, the α TCR chains had lower read counts **(Figure 3.14A)**. To adjust the α: β TCR chain read count ratio, the amount of RT-TRAC primer was doubled to a 2.4 µM final RT reaction amount. This increase led to the successful equalization of the α: β TCR chain read count ratio but did not have a significant effect on αβTCR detection rates (41.6 and 43.8 %, respectively) **(Figure 3.14A)**. Nevertheless, the 2:1 ratio was adopted to homogenize the α TCR and β TCR read counts.

To rule out genomic DNA amplification using the designed single-exon binding RT-primers as a causative factor for the observed PCR noise, exon-spanning (ES) RT primer variants were designed for the *TRAC, TRBC1 and TRBC2* genes. No common primer could be designed for both *TRB1* and *TRBC2* genes due to their different sequences at the exon-exon junctions. The new exon-spanning primer variants were used in all combinations to reverse transcribe 250 ng total spleen RNA followed by a 30 cycle RBC-incorporation PCR. TRAC ESv1, TRBC1 ESv1 and TRBC2 ESv1 in combination showed the best amplification ability **(Figure 3.14B)**. Sorted single T cells were processed using either the previously used single-exon binding primers or the newly designed chosen exon-spanning primer combination. The single-exon binding primers showed higher detection rates compared to the exon-spanning primers **(Figure 3.14C)**. Further, the use of exon-spanning primers had no effect on the presence of noise in the NGS data ruling out genomic DNA as a causative factor for the observed noise (Data not shown). Therefore, the single-exon binding primers were further used as before.

Another approach to reduce platform noise involved reducing the amounts of reagents for RT; GSP, dNTPs and TS. 15 single T cells were processed per condition. Standard reagent amounts led to the detection of 47 % of the paired αβTCR chains **(Figure 3.14D)**. Any amount reduction of these RT reagents led to the complete elimination of paired αβTCR chain detection. This reduction was accompanied with a decreased ability to align NGS reads to TCR genes **(Figure 3.14E)**. As a result, the RT reagent amounts were left unchanged.

(A)

(B)

(C)

(D)

(E)
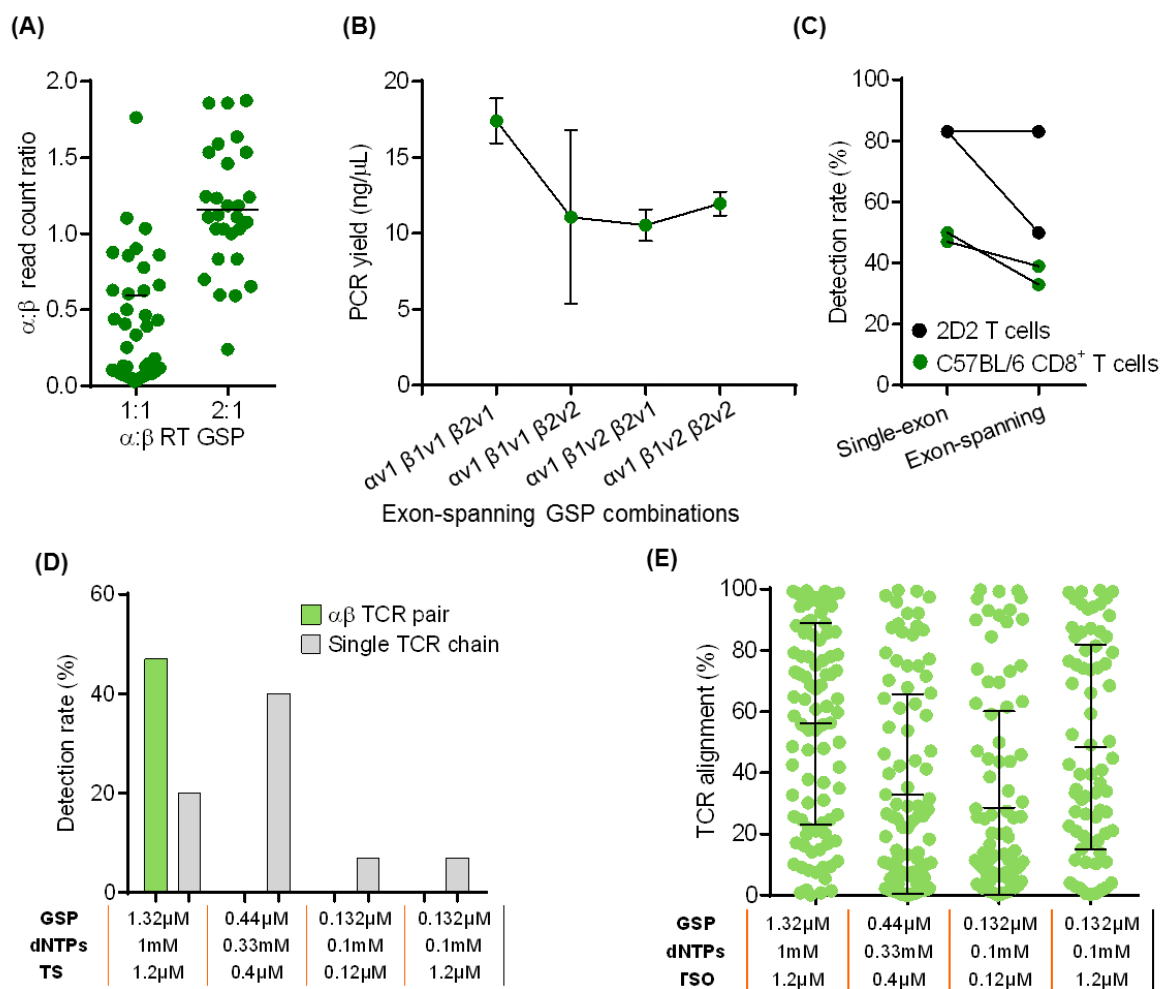
Figure 3.14| Optimizing the scTCRseq RT reaction. (A) A 2:1 RT-TRAC: RT-TRBC gene-specific primer (GSP) ratio equalizes the α:β TCR chain read count ratio. Single sorted CD4⁺ T cells were processed using either a 1:1 or a 2:1 RT-TRAC: RT-TRAC primer ratio. **(B)** TRAC v1, TRBC1 v1 and TRBC2 v1 exon-spanning primers (ES) show together the best spleen RNA-derived TCR mRNA amplification ability compared to their ES primer counterparts. RT was performed using 250 ng total spleen RNA. PCRs were performed using 30 cycles and the PfuUltra Hotstart High-Fidelity DNA Polymerase and 66 °C annealing temperature. 10 µL first strand cDNA was used as template for the RBC-incorporation PCR. Following bead purification, the concentrations were measured using the Qubit. **(C)** Single-exon binding primers show superior detection rates compared to ES RT primers. Two 96-well plates with single sorted CD8⁺ T cells were processed separately. One-half of each plate was processed using either the single-exon binding primers or the exon-spanning primers. Twelve CD4⁺ 2D2 T cells with known TCR sequences were used per plate as positive controls (six cells per condition). Joint dots indicate detection rates from the same plate. **(D)** Reducing the amounts of RT reagents leads to the complete elimination of αβTCR chain detection and **(E)** a reduction in the ability to align TCR sequences. 60 single sorted CD4⁺ T cells were processed. 15 T cells were sorted per condition. TCR alignment rates were obtained from the MiXCR generated summary report.

3.3.3. RBC-incorporation PCR optimization

Due to the strong amplification of single cell derived first strand cDNA, it is crucial that the first PCR is as accurate as possible. Therefore, the RBC-incorporation PCR was optimized to increase yield and reduce costs and noise whilst ensuring maximum accuracy.

In the first few cycles of the RBC-incorporation PCR, only the TS and GSP parts of the forward and reverse primers, respectively, bind to their respective regions in the first strand cDNA allowing the use of a lower annealing temperature **(Figure 3.15A)**. After the first cycles, the whole primer sequence binds to the template permitting an increase in the annealing temperature. In contrast to the standard PCR where a 66 °C annealing temperature is used throughout all cycles, a step-up PCR involves a lower starting annealing temperature allowing increased yield followed by a raise in temperature to the previously used temperature to ensure amplification specificity and minimize PCR noise in the later cycles. In contrast, in a touch-up PCR, a lower starting annealing temperature is used followed by (in this case) two increases in the annealing temperature, with the second increase reaching higher temperatures than before. The yield of the RBC-incorporation PCR was thus compared using the standard PCR at 66 °C, a step-up PCR and a touch-up PCR protocol finally reaching 72 °C annealing temperature. 0.25 pg of TCR-encoding IVT RNA were used per well. Compared to the standard and step-up PCR, the touch-up program generated higher yields and was therefore adopted **(Figure 3.15A)**. In a parallel experiment processing twelve single cell sorted CD8$^+$ T cells per condition, the touch-up PCR led to the detection of six unique T cell clones compared to three clones using the standard PCR program (Data not shown).

Taking advantage of this increased yield, we aimed to reduce the PCR cycle number to minimize potential PCR errors. Instead of the previously used 25 cycles, the RBC-incorporation touch-up PCR was tested using 14, 16, 18, 20 and 25 cycles. Single CD8$^+$ T cells were processed and 15 % of the diluted first strand cDNA was used for each PCR. Whilst the paired αβTCR chain detection rate stayed relatively constant (45, 47, 48, 49 and 46 %, respectively) (data not shown), processing the single T cells using 18 cycles showed the highest TCR sequence alignment rate **(Figure 3.15B)** and higher than that of the previously used protocol (41 %; data not shown)**.** Thereafter, the RBC-incorporation touch-up PCR was performed using 18 cycles.

Moreover, we predicted that leftover TS primer from the RT reaction competes with the Tag-A-RBC-TS forward PCR primer for the first strand cDNA in the RBC-incorporation PCR, thereby reducing PCR efficiency. The elimination of unused TS primer after RT using Exonuclease I was tested. 250 ng total spleen RNA was used as template. Compared to control, treatment using 2.5 U increased the RBC-incorporation PCR yield by

discarding unused TS primer after RT and preventing its competition with the Tag-A-RBC-TS forward PCR primer for the first strand cDNA in the RBC-incorporation PCR. **(Figure 3.15C)**. Next, single CD4[+] T cells were processed comparing the use of either 0 U, 5 U or 10 U Exonuclease I after RT. Although there was no significant increase in the paired αβTCR detection rates (44, 47 and 46 %, respectively), treatment using 5 U showed the highest TCR sequence alignment rate **(Figure 3.15D).** The use of 5 U was also associated with a 5.2% and 3.2 % increase in the ability to detect the RBC and CBCs in the sequences (data not shown). The treatment of the first strand cDNA using 5 U Exonuclease I was therefore adopted thereafter.

Moreover, all bead purification steps were performed using SPRI beads (Beckman Coulter) which were very costly. The Agencourt AMPure XP bead kit (Beckman Coulter), which is 25 % cheaper than the SPRI beads, is not sold as a DNA size selection bead kit. With the aim of minimizing platform costs, we tested its ability to perform DNA size selection and compared it to the SPRI beads using a DNA ladder. Size selection using different bead: sample ratios showed the exact same results when using both bead products **(Figure 3.15E)**. Thereafter, the less costly Agencourt AMPure XP beads were used for all size-selection bead purifications.
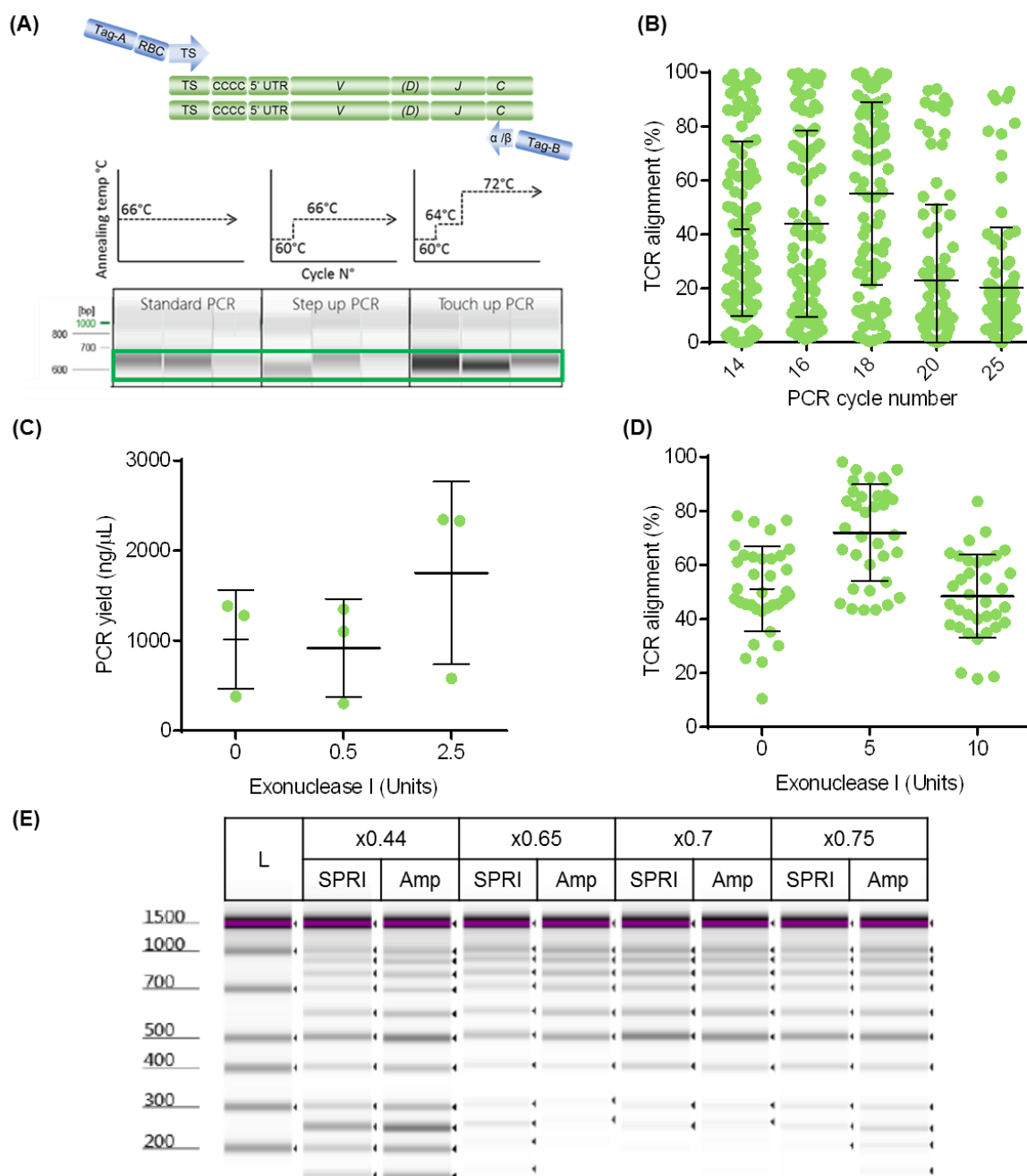
Figure 3.15| Optimizing the scTCRseq RBC-incorporation PCR. (A) The touch-up PCR generates a higher yield compared to the standard and step-up PCR. 0.25 pg of TCR-encoding IVT RNA (αV9 and βV15) were used per well. Three PCRs were performed per condition. PCRs were performed using a total of 35 cycles. Graphs represent the different annealing elevations performed for each PCR program. Read outs were performed using a Qiaxcel. **(B)** RBC-incorporation touch-up PCR using 18 cycles shows the highest TCR alignment rate. CD8[+] T cells were sorted into a 96-well plate. After RT and 1:10 cDNA dilution, 10 µL of the diluted first strand cDNA were used for each RBC-incorporation PCR. The CBC-incorporation PCR was performed as before using a 25 cycles standard PCR and 10 µL template. NGS libraries were prepared for each pooled plate and sequenced at a depth of $1x10^6$ reads per plate. (95°C 2

min) x1, (94°C 30 sec, 60°C 30 sec, 72°C 60 sec) x5, (94°C 30 sec, 64°C 30 sec, 72°C 60 sec) x5, (94°C 30 sec, 72°C 120 sec) 4, 6, 8, 10 or 15 cycles, (72°C 6 min) x1. **(C)** Treatment of first strand cDNA (prior to dilution) derived from RT of 250 ng bulk splenocyte-derived total RNA using different amounts of Exonuclease I. Yield of the subsequent RBC-incorporation PCR was measured using the Qubit. **(D)** Treatment of first strand cDNA using 10 U Exonuclease I enhances TCR alignment rate. CD4+ T cells were sorted into a 96-well plate. After RT, first strand cDNA was treated using (0, 5 or 10 U Exonuclease I) for 30 minutes at 37°C, then 20 minutes at 80°C. The platform was performed till the end and the pooled plate sample was sequenced at a depth of $1x10^6$ reads. **(E)** Agencourt AMPure XP beads can perform size-selection based DNA purification as a cheaper alternative. A 100 bp DNA ladder was purified using SPRI or AMPure XP beads using a bead: sample ratio of x0.44, x0.65, x0.70 and x0.75. Post-purification samples were run on the Qiaxcel.

3.3.4.   CBC-incorporation PCR optimization

For the CBC-incorporation PCR, a standard PCR at 64 °C annealing temperature using 27 cycles was used so far. With the aim of optimizing this PCR, the PCR was tested using a standard program at 64 or 66 °C annealing temperature as well as a 64-66 °C step-up program. No touch-up program was tested since the complete primer sequences including Tag-A and Tag-B bind to the template during the first few cycles **(Figure 3.16A)**. Each PCR program was performed for 27 or 30 PCR cycles using the PfuUltra II Fusion HS DNA Polymerase **(Figure 3.16A)**. The standard PCR at 64 °C showed the highest yield compared to the standard PCR at 66 °C and the 64>66 °C step-up program **(Figure 3.16B)**. There was no significant increase in yield when using 30 cycles compared to the 27 PCR cycles used before.

The best PCR condition using the PfuUltra II Fusion HS DNA Polymerase, which had been previously used as well, (standard PCR at 64 °C, 27 cycles) was then compared to the less costly Taq Polymerase at 27 cycles. The use of PfuUltra II Fusion HS DNA Polymerase showed a significantly higher yield compared to the Taq Polymerase **(Figure 3.16C)**.

Never the less, we tested the ability of both Polymerases to function at smaller reaction volumes and Polymerase amounts to reduce PCR costs. Reduction in the total reaction volume or Polymerase amount when using the PfuUltra II Fusion HS DNA Polymerase significantly reduced PCR yield **(Figure 3.16D)**. In contrast, the Taq Polymerase showed a similar yield to standard volumes when the total reaction volume and Polymerase amount were halved. Despite its cost-saving advantage, the Taq Polymerase, in accordance with its lower accuracy, showed a higher PCR error rate compared to the Pfu Fusion **(Figure 3.16E)**. As a result, the higher accuracy of the Pfu Fusion was given priority over its higher costs. The Pfu Fusion was therefore further used for further experiments as before. Afterwards, different PCR cycle numbers and template amounts were tested. The Pfu Fusion showed the highest αβTCR detection rates when using 24

PCR cycles and 10 μL template **(Figure 3.16F)**. In conclusion, the PfuUltra II Fusion HS DNA Polymerase, a standard PCR at 64 ºC annealing temperature and 24 cycles as well as 10 μL template were the conditions adopted for the CBC-incorporation PCR.
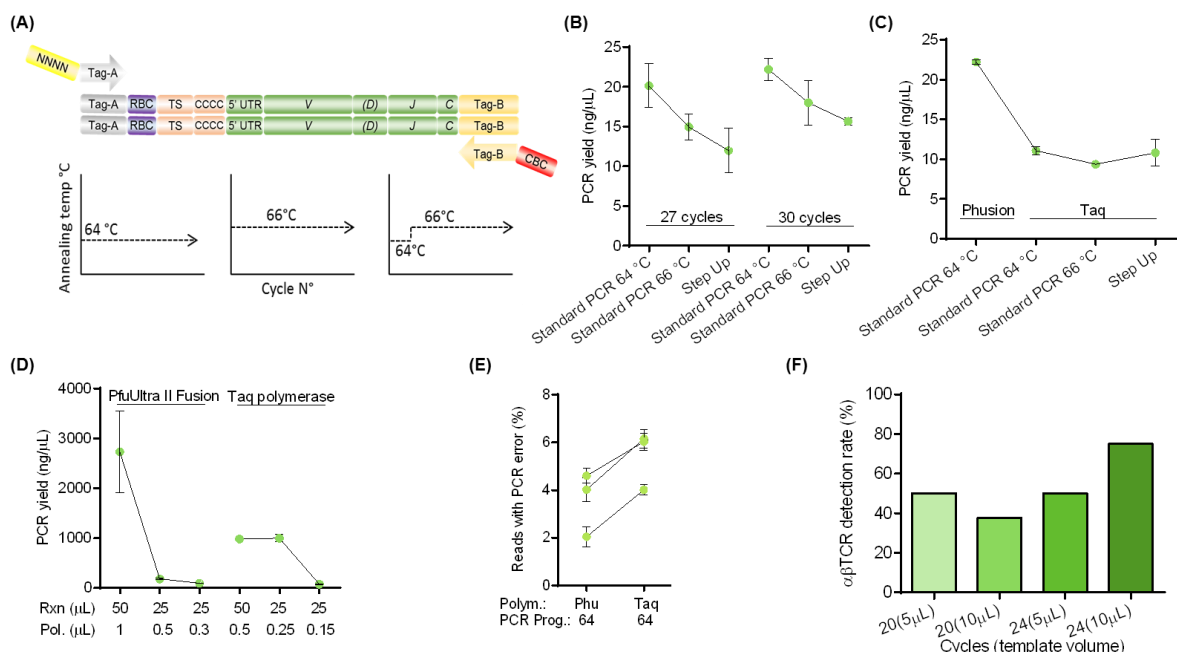
Figure 3.16| Optimizing the scTCRseq CBC-incorporation PCR. Purified RBC-incorporation PCR products were generated from 60 single sorted CD4+ T cells. For all experiments below, the same column pool (after the RBC-incorporation PCR and bead purification) was split as template for all CBC-incorporation PCRs of the respective experiment. **(A)** Graphical depiction of the different PCR programs tested. **(B)** Comparison of different PCR programs at 27 and 30 cycles using the PfuUltra II Fusion HS DNA Polymerase. **(C)** Comparing PCR yield when using the PfuUltra II Fusion HS DNA or the Taq Polymerase using 27 PCR cycles. **(D)** Testing ability of both Polymerases to function with reduced total reaction and Polymerase volume. **(E)** Comparing the PfuUltra II Fusion HS DNA and Polymerase Taq Polymerase PCR error generation rate. CBC-incorporation PCR performed at 64 ºC for 27 cycles. Reads with error values obtained from the MiXCR summary report. **(F)** αβTCR detection rates using different PCR cycle numbers and template volumes in the CBC-incorporation PCR.

3.3.5. Noise eradication through post-PCR Exonuclease I treatment and using more template cDNA

So far, platform detection rate was increased and platform duration and costs reduced, however, noise remained. The presence of noise was hypothesized to originate from leftover RBC-primers being falsely incorporated into cDNA from wells of other rows in the second PCR after column pooling **(Figure 3.17A)**. As a result, two modifications were tested simultaneously to eliminate such noise: 1) use of larger amounts of first strand

cDNA (10 uL of 1:3 diluted first strand cDNA instead of 10 uL of 1:10 diluted first strand cDNA). This could potentially increase the 'real TCR chain' to 'noise TCR chain' read count ratio. 2) Eliminating leftover RBC-primers after column pooling and bead purification using 10 U Exonuclease I treatment thereby preventing unused primer barcodes from being incorporated into other amplicons from other rows. These modifications, together with the implementation of the TCR chain filter script, led to the complete elimination of platform noise as indicated by the ranking of the 2D2-derived TCR chains in respective well **(Figure 3.17B)**. Prior to these modifications, the α and β TCR chains of the 2D2 T cells did not have the highest reads in their wells due to the presence of noise with higher read counts. After optimization, the 2D2 TCR chains had the highest read counts in their respective well. The exonuclease I treatment eliminated leftover RBC-primers after the first PCR and thereby prevented them from being falsely incorporated into cDNA from other wells in the second PCR after column pooling.
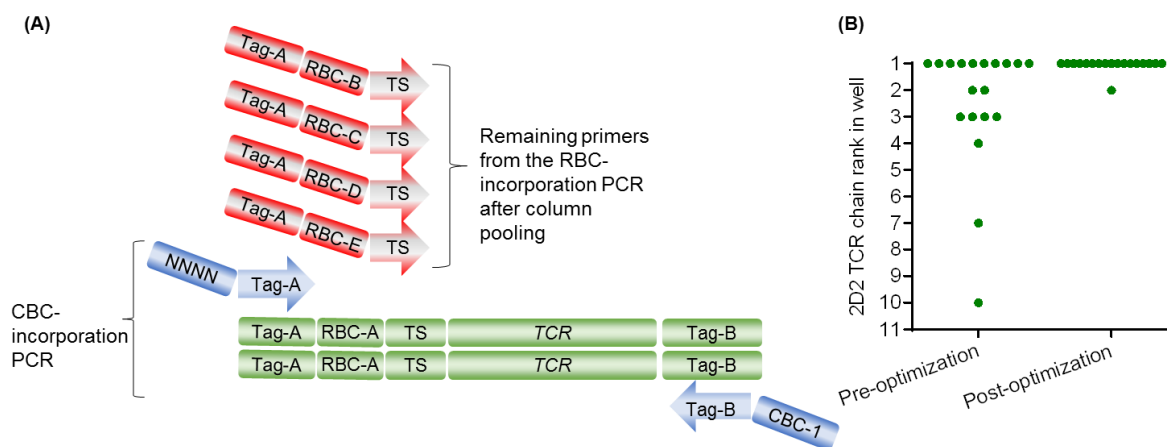
Figure 3.17| Elimination of platform noise. **(A)** Schematic of the hypothesized reason behind the observed noise. Leftover RBC-primers are falsely incorporated into cDNA from wells of other rows in the second PCR after column pooling. **(B)** Elimination of platform noise. The lowered chain ranks of the spiked-in 2D2 TCR chains in their respective wells prior to optimization was used to indicate noise levels.

Together, the optimization of various platform parameters led to a significant rise in the platform efficiency by increasing its paired αβTCR detection rates as well as reducing its noise, costs and hands-on time. **Supplementary section 6.4** contains the optimized scTCRseq platform protocol.

3.4. Performance evaluation and validation of the optimized scTCRseq platform

3.4.1. Quantitative performance analysis of the optimized platform.

Following optimization, the platform performance was evaluated in seven separate platform runs, each composed of one processed 96-well plate using the quantitative reference parameters introduced in section 3 (see Table 3.2). For each plate, $CD3^+CD4^+$ or $CD3^+CD8^+$ T cells were sorted into 72 test wells (light green) **(Figure 3.18A)**. $CD3^+CD4^+TRBV16^+$ splenocytes from 2D2 MOG_{35-55}-specific TCR transgenic mice were added to positive control wells (dark green). Negative control wells contained no cells (grey). The location of the positive and negative controls were distributed over all rows and columns and kept the same through all runs, however, different amounts of control wells were used in the different experiments.

An average of only $1,371,532 \pm 415,024$ read counts were used per plate (i.e. $14,286 \pm 4,423$ reads/ well) **(Figure 3.18B)**. Efficient and consistent incorporation of primer-tags and barcodes could be shown across all experiments **(Figure 3.18C)**. Although there is a decreasing 'per base' sequence quality with increasing base position during NGS, the amplified CDR3 regions, primer-tags and the barcodes existed in the good quality read regions [133]. 81 % of all NGS reads were successfully demultiplexed indicating efficient barcoding and demultiplexing. 90 % of the NGS reads could be mapped to TCR genes signifying accurate TCR gene amplification and sequencing **(Figure 3.18D)**. The paired αβTCR chains of 77 % of the spiked-in 2D2 T cells were detected showing platform sensitivity **(Figure 3.18E).** The paired αβTCR chains of 49% of the test T cells were detected and detection rates were consistent across different experiments. Consistent with literature, in 15 % of the T cells with a detected α TCR chain, a second α TCR chain could be identified **(Figure 3.18F)**. There was no significant correlation between read count depth and paired αβTCR chain detection rate **(Figure 3.18G)**. 3.6 % of the negative control wells contained some TCR reads indicating a low false positive rate **(Figure 3.18H).** Together, the developed scTCRseq platform can efficiently identify paired αβTCR chains from single T cells in a robust manner.
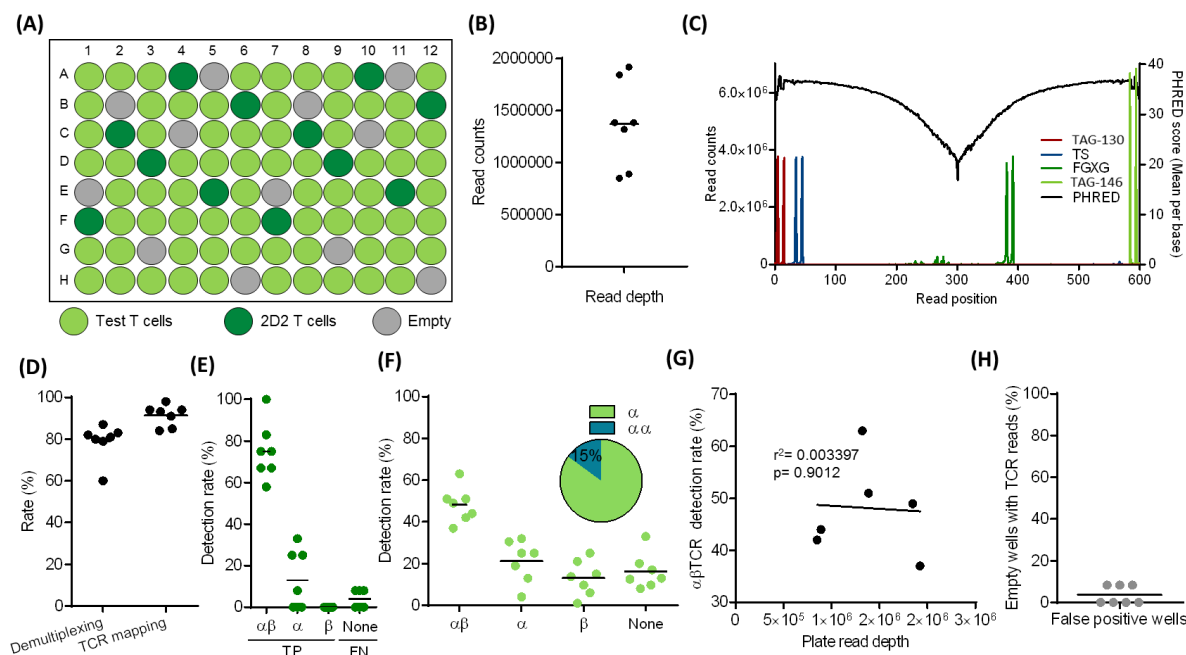
Figure 3.18| ScTCRseq platform performance evaluation reveals efficient and consistent paired αβTCR chain retrieval. **(A)** Plate layout used. **(B)** Average read depth per plate. **(C)** Percentage of NGS reads containing Tag-130 (Tag-A), RBC, TS, Cys_{104}, FGXG, Tag-146 (Tag-B) and CBC elements. Graph bars indicate average positions covered by the respective elements. PHRED score indicates read quality. **(D)** Demultiplexing rate displays plate reads successfully allocated to a well. TCR mapping rate shows demultiplexed reads mapped to a TCR sequence. Detection rates of paired αβTCR chains per well from **(E)** 2D2 T cells and **(F)** test T cells. TP, true positive; FN, false negative. Detection rate of single α TCR or single β TCR chain or no chain per well is also shown. Pie chart specify fraction of test T cells detected with more than one α TCR. **(G)** Correlation between plate read depth and paired αβTCR detection rates of test T cells. **(H)** False positive rates calculated as the percentage of negative control wells with TCR read counts (noise level) after running the TCR chain filter. The whole performance evaluation was performed using seven separate experiments (one 96-well plate each).

3.4.2. Throughput and costs of the platform

Following platform optimization, the throughput and costs of the platform were calculated (**Figure 3.19**). Due to the early sample barcoding and pooling approach adopted, up to three 96-well plates can be processed within three hand-on work hours. This is accompanied with another three hours of incubation/waiting time. After the processing of the desired plate numbers, the library preparation can be prepared in 2.5 hand-on work hours. The RBC-incorporation PCR was automated using a TECAN robot to reduce the hands-on time spent. Looking at the cost dissection of the developed platform, 66% of the costs are allocated to the two Polymerases used. Furthermore, there is no use of specific equipment increasing platform applicability. The average cost per plate for obtaining TCR

sequences is 600€ if 15 plates are to be sequenced per flow cell. This is much lower compared to the pre-existing scTCR cloning platform (6000€ per plate)[127]. Overall, an easily applicable and cost-efficient platform was developed.
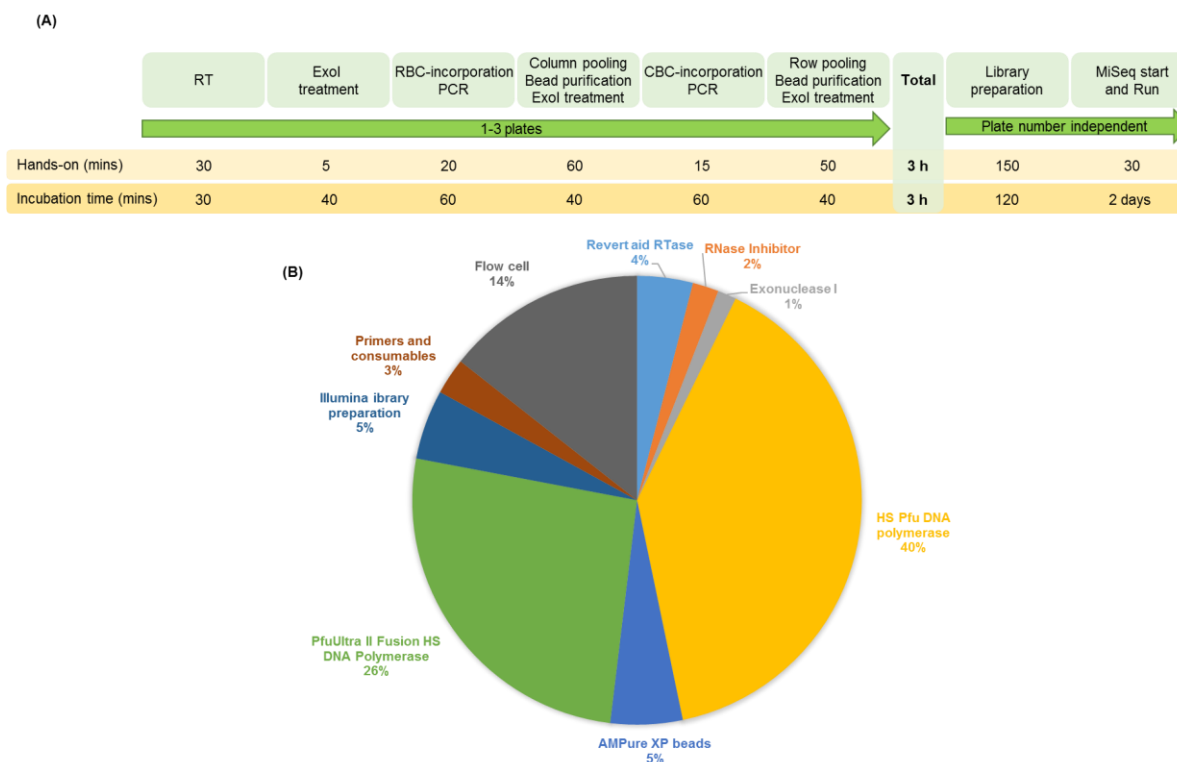
Figure 3.19| Throughput and cost dissection of the developed scTCRseq platform.

3.5. Identification of functional tumor-antigen-specific αβTCRs

TA-encoding RNA or peptide-based vaccinations can induce strong immune responses in mice[85,134]. The optimized scTCRseq platform was therefore applied together with RNA-lipoplex (LPX) or peptide-based vaccination to generate a library of human and murine HLA-restricted murine TCRs specific for TA of different classes: oncoviral antigens, cancer/testis (C/T) antigens and neoantigens. The discovery of TCRs against such antigens would encourage pre-clinical ACT studies for different cancers.

Human papillomavirus (HPV) and Hepatitis B virus (HBV) antigens were chosen due to their tumor-specific expression and clinical significance. In HPV16-induced cancers, oncoproteins E6 and E7 are critical for the oncogenesis and maintenance of the transformed state[61]. Other oncoviral antigens such as HBV-derived surface-antigen (HBVs) and core-antigen (HBVc) represent attractive targets for HBV-positive hepatocellular carcinoma (HCC)[135].

A2DR1 and C57BL/6 mice were vaccinated using RNA-LPXs (full length HPV 16 E7 or E6) or peptides (HBVs$_{126-138}$ or HBVc$_{128-140}$), respectively. A2DR1 mice express human HLA genes (HLA-A*0201$^+$/DRB1*0101$^+$) and can therefore express human epitopes. Three consecutive vaccinations were performed with one-week intervals. The spleens were isolated seven days after the last vaccination. 12 h after ex vivo antigen peptide-specific restimulation of splenocytes, activated CD8$^+$ (HPV E7) or CD4$^+$ (HPV E6, HBVc and HBVs) T cells expressing IFNγ were single cell sorted **(Figure 3.20A)**. Samples from the HPV E7 and HBVs vaccinated mice were also processed using the single T cell cloning platform (Omokoko et al. 2014) as a benchmark.

scTCRseq of CD8$^+$ and CD4$^+$ activated T cells from an HPV E7 and an HPV E6 vaccinated A2DR1 mouse, respectively, led to the identification of four unique αβTCRs per antigen **(Figure 3.20B)**. HPVE7_TCR1 was also detected using the scTCR cloning platform validating the results achieved via the novel platform (data not shown).The scTCR cloning platform did not detect any additional TCRs. HPVE7_TCR1 was detected 27 times using the scTCRseq platform compared to ten times using the scTCR cloning platform indicating a higher sensitivity of the former. The three other TCRs detected using the scTCRseq platform were detected either once or twice.

All αβTCRs were cloned using the remaining first strand cDNA from the respective well of origin. IVT RNA encoding the paired α and β chains of each TCR was generated. TCR encoding IVT RNA was transfected into primary T cells and ELISPOT assays performed. Six out of the eight cloned TCRs (HPV E7 and E6) showed significant expression on the surface of TCR-electroporated human T cells. **(Figure 3.20C)**. The two TCRs not detected on the surface might have been caused by the detection of the second α TCR chain allele and not the first. The highly expanded HPVE7_TCR1 as well as HPVE6_TCR1 and HPVE6_TCR4 mediated strong reactivity against their respective antigen using peptide-pulsed as well as antigen-encoding RNA electroporated K562 antigen presenting cells (APCs) **(Figure 3.20D)**. HPVE7_TCR1 was specific for the HPVE7$_{11-20}$ minimal peptide, whereas both HPVE6_TCR1 and HPVE6_TCR4 were specific for the HPVE6$_{127-142}$ minimal peptide **(Figure 3.20E)**.

Sequencing CD4$^+$ T cells from HBV-derived surface-antigen (HBVs) and core-antigen (HBVc) peptide-vaccinated C57BL/6 mice, led to the identification of seven and six unique αβTCRs, respectively. The clonotype frequency varied greatly amongst the different clones **(Figure 3.20F)**. For the HBVs vaccination, HBVs_TCR1 was also detected using the scTCR cloning platform in a clonally expanded manner acting as a benchmark validation. The scTCRseq platform could identify six additional αβTCRs. However, only the highly expanded HBVs_TCR1 mediated reactivity against HBVs$_{126-138}$ peptide loaded bone marrow derived dendritic cells (BMDCs). HBVc_TCR5, which was only detected once, strongly recognized HBc$_{128-140}$ peptide loaded BMDCs **(Figure**

3.20G). The other HBVc TCRs, including the clonally expanded ones showed no antigen recognition (data not shown).

Taken together, by processing T cells from RNA and peptide vaccinated C57BL/6 and HLA-transgenic mice, the developed scTCRseq platform enabled straightforward detection of five oncoviral-specific TCRs as well as TCR clonality analysis.
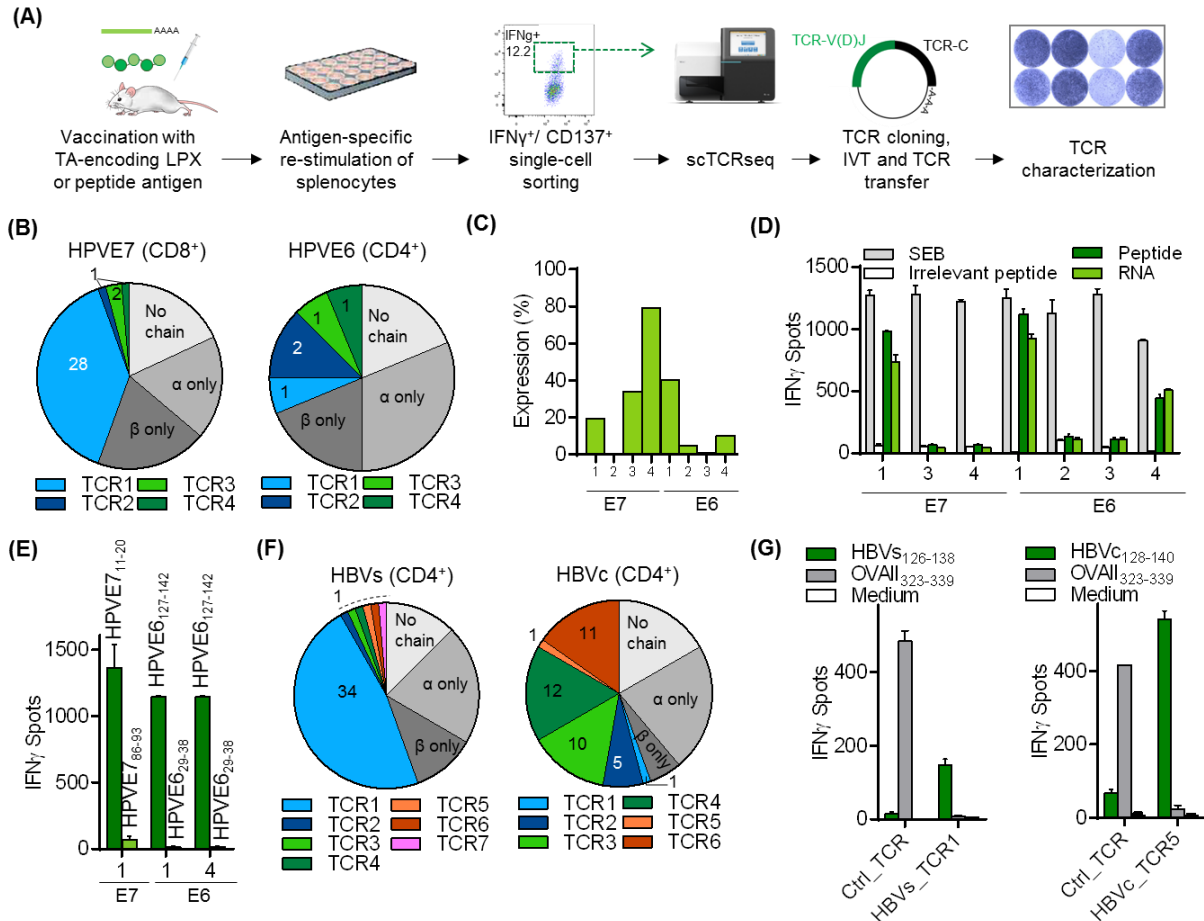
Figure 3.20| Discovery of human and murine HLA-restricted TCRs specific for oncoviral epitopes. **(A)** Antigen-specific TCR discovery process. **(B)** Identified absolute TCR clonotype frequencies after scTCRseq of 72 and 16 single cell sorted T cells from an HPV E7 or E6 vaccinated A2DR1 mouse, respectively. **(C)** Surface expression determined by β TCR chain V-gene specific flow cytometry analysis after co-electroporation of α and β TCR chain-encoding RNA into human PBMC-derived CD8$^+$ or CD4$^+$ T cells. **(D)** TCR-electroporated human CD8$^+$ or CD4$^+$ T cells were co-cultured with K562_A*0201 or K562_DRB1*0101 cells in an IFNγ ELISPOT assay, respectively. K562 cells were peptide loaded or electroporated with antigen-encoding RNA. Controls; HIV GAG peptide loaded K562 cells and Staphylococcal Enterotoxin B (SEB)-treated T cells. Only TCRs expressed on surface are shown. **(E)** TCR-electroporated human CD8$^+$ and CD4$^+$ T cells were co-cultured with HPV E7-minimal peptide loaded K562_A*0201 or HPV E6-minimal peptide loaded K562_DRB1*0101 cells, respectively. HPV E7$_{11-20}$ (YMLDLQPETT), HPV E7$_{86-93}$ (TLGIVCPI), HPV E6$_{127-142}$ (DKKQRFHNIRGRWTG), HPV E6$_{29-38}$ (TIHDIILECV). **(F)** Absolute TCR clonotype frequencies after scTCRSeq of HBVs or HBVc-vaccinated

C57BL/6 mice. **(G)** TCR-electroporated murine T cells were co-cultured with $HBVs_{126-138}$, $HBc_{128-140}$ or $OVA_{323-339}$ (ISQAVHAAHAEINEAGR) peptide-loaded bone marrow derived dendritic cells (BMDCs). Positive control; $OVA_{323-339}$-specific TCR electroporated T cells. Mean + SD from technical duplicates is depicted.

The C/T antigen KKLC1 is expressed in 79% of early stage gastric cancers[136]. Colleagues previously showed by transcriptional profiling that KKLC1 expression is strictly restricted to testis, is immunogenic and expressed at high level and frequency in triple-negative breast cancer (TNBC) tumors[64]. T cell-mediated immunotherapy targeting KKLC1 may therefore represent an attractive approach for gastric and TNBC cancers. We therefore sought to apply the developed scTCRseq platform to identify KKLC1 -specific TCRs as well as study the antigen-specific T cell response in combination with repertoire profiling.

Three naïve A2DR1 mice were vaccinated with full-length KKLC1-encoding RNA-LPX. ScTCRseq of 72 T cells from mouse one (M1) allowed the discovery of eleven unique TCRs **(Figure 3.21A)**. KKLC1_TCR1, #2 and #3 were clonally expanded. All TCRs were expressed on the surface of human $CD8^+$ T cells except KKLC1_TCR2 (0.38%) and #8 (0.26%). KKLC1_TCR1, #3, #4 and #6 could strongly mediate reactivity against peptide-pulsed and antigen-electroporated K562 APCs **(Figure 3.21B left and right)**. CDR3 similarity within a repertoire has been previously described as a potential predictor of antigen-specific TCRs[137]. The β TCR chains of KKLC1_TCR1 and #6 displayed the same functional CDR3 amino acid sequence CASSPGLGGGYEQYF generated by convergent recombination of two different *trbv* genes **(Figure 3.21C)**. Shared CDR3 aa motifs were identified amongst all β TCR chains four antigen-specific TCRs representing potential paratopes **(Figure 3.21D)**. However, the α TCR chains of the respective TCRs did not show any similarity.

The scTCR cloning platform (Omokoko *et al.* 2014) as well as TCR repertoire profiling analysis were performed as benchmarks for the same sample (mouse M1). The newly developed ScTCRseq platform was superior to the cloning platform and could identify all αβTCRs found by the latter in addition to six additional unique TCRs, one of which was functional (KKLC_TCR1) **(Figure 3.21E)**.

ScTCRseq data correlated nicely with the repertoire profiling data from the same mouse M1 **(Figure 3.21F)**. The α and β TCR chains of the most frequently detected TCR in the scTCRseq data were also the most abundant α and β TCR chains in the TCR repertoire data. Furthermore, the α and β TCR chains of the second most frequently detected TCR in the scTCRseq data were the second and third most abundant α and β TCR chains in the TCR repertoire data. The scTCRseq data also revealed that the much greater frequency of the top α TCR chain compared to the top β TCR chain in the repertoire is due to the pairing of the same α TCR chain with another β TCR chain. The top three α TCR chains in the

repertoire and the scTCRseq data were the same and represented the α TCR chains of three out of the four KKLC1-specific TCRs. In addition, six out of the eleven identified αβ TCR pairs obtained using the scTCRseq platform were in the top nine clones of the repertoire.

TCR repertoire profiling of the spleens from all three vaccinated mice (M1, 2 and 3) revealed the presence of three out of the four antigen-specific TCRs in the repertoire of more than one vaccinated mouse **(Figure 3.21G)**. Such TCRs are termed "public" TCRs.

Therefore, in addition to the identification of KKLC1-specififc TCRs, the data reveals superior performance of the scTCRseq platform over the scTCR cloning platform as well as the ability of the former to provide basic insight into the repertoire.
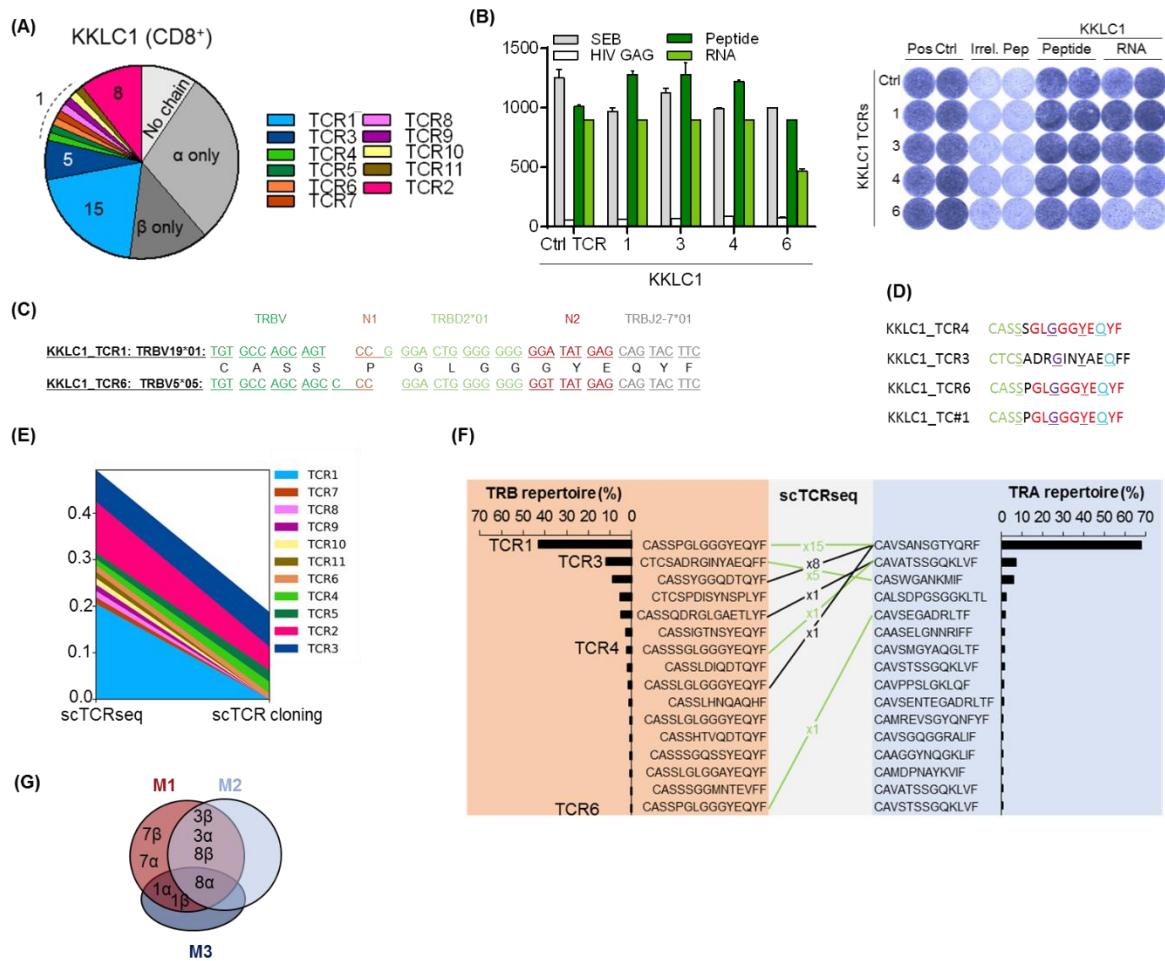
Figure 3.21| Identification and unbiased analysis of cancer/testis-specific T cell responses at the single cell level. (A) Absolute TCR clonotype frequencies identified using scTCRseq of 72 sorted CD8[+]IFNγ[+] T cells from an A2DR1 mouse (M1) vaccinated with full-length KKLC1 RNA-LPX. Three mice were vaccinated in total. 12 h *in vitro* antigen-specific restimulation was performed. **(B)** TCR-electroporated human CD8[+] T cells were co-cultured with peptide loaded or antigen-encoding RNA electroporated K562_A*0201 cells and an IFNγ ELISPOT assay performed. HIV GAG peptide loaded K562 cells and a previously discovered KKLC1-specific TCR (Paret *et al.* 2015) were used as negative and positive controls,

respectively[64]. Only functional TCRs are shown. Right, representative Elispot. **(C)** Convergent recombination of two different *trbv* genes to TRBD2 and TRBJ2-7 generating the same CDR3 through codon degeneracy and random N base insertion (N1 and N2). **(D)** β TCR chain CDR3 regions of the four antigen specific TCRs show shared amino acid usage. **(E)** Sorted $CD8^+IFN\gamma^+CD137^+$ T cells from M1 were processed using the scTCR cloning platform as a benchmark. Venn diagram showing the TCRs identified from KKLC1 vaccinated M1 using scTCRseq and the scTCR cloning platform. **(F)** Matching TRB and TRA repertoire data from the splenocytes of KKLC1-vaccinated mouse M1 using scTCRseq data (black lines) and positioning KKLC1-specific TCRs in the splenocyte repertoire (green lines). **(G)** Venn diagram showing identical KKLC1 antigen-specific TCRs in the post-vaccination repertoires of all three vaccinated mice (M1, M2, M3).

Increasing evidence discloses neoepitopes as the primary targets of tumor-infiltrating lymphocytes (TILs)[100–102]. Targeting such antigens using T cells mediates complete regression in cancer patients. Tumor mouse models are valuable for testing the efficacy of such approaches. A recent study combined exome and transcriptome sequencing with mass spectrometry to identify neoepitopes in the widely used MC38 colorectal carcinoma mouse model[138]. Amongst others, two immunogenic MHCI-restricted neoepitopes Adpgk$_{R304M}$ or Reps1$_{P45A}$ have been described. The former involved in glycolysis and the latter in endocytosis and cytoskeletal changes. Progress of neoantigen-specific ACT is hindered by the limited availability of TCRs specific for neoantigens. Thus, we aimed to apply the scTCRseq platform to identify neoantigen-specific αβTCRs.

Splenic T cells from naïve C57BL/6 mice vaccinated with Adpgk$_{R304M}$ and Reps1$_{P45A}$ -encoding RNA-LPXs were processed using the scTCRseq platform. Sequencing TCRs from Adpgk$_{R304M}$ or Reps1$_{P45A}$ RNA-LPX vaccinated C57BL/6 mice led to the identification of nine and seven unique TCRs, respectively **(Figure 3.22A)**. Adpgk_TCR1 (most clonally expanded T cell) as well as #7 (fourth most expanded clone) mediated a strong response against the Adpgk$_{R304M}$ peptide loaded BMDCs **(Figure 3.22B)**. Reps1_TCR1 (most clonally expanded T cell), #4 (detected 8 times) and #7 (detected once) strongly recognized Reps1$_{P45A}$ peptide loaded BMDCs **(Figure 3.22C)**. Adpgk_TCR1 and #7 were highly mutation specific and did not recognize the WT peptide. Reps1_TCR1 was also highly mutation-specific whereas Reps1_TCR4 and #7 showed some recognition of the WT peptide at this peptide amount **(Figure 3.22D)**. Adpgk_TCR1 and #7 as well as Reps1_TCR1 and #4 could mediate reactivity against the neoantigen-expressing MC38 tumor cell line. Adpgk_TCR1 showed the strongest tumor cell line recognition. IFNγ stimulation of the tumor cell line to upregulate MHC class I expression could enhance the TCR-mediated response. Reps1_TCR7 could only show a weak cell line recognition **(Figure 3.22E)**.

Together, the data shows the successful application of the scTCRseq platform to identify TCRs recognizing endogenously expressed neoantigens in tumor cell lines. Adpgk_TCR1

and #7 are presented as top candidates for pre-clinical neoantigen-specific ACT studies using the MC38 mouse model.
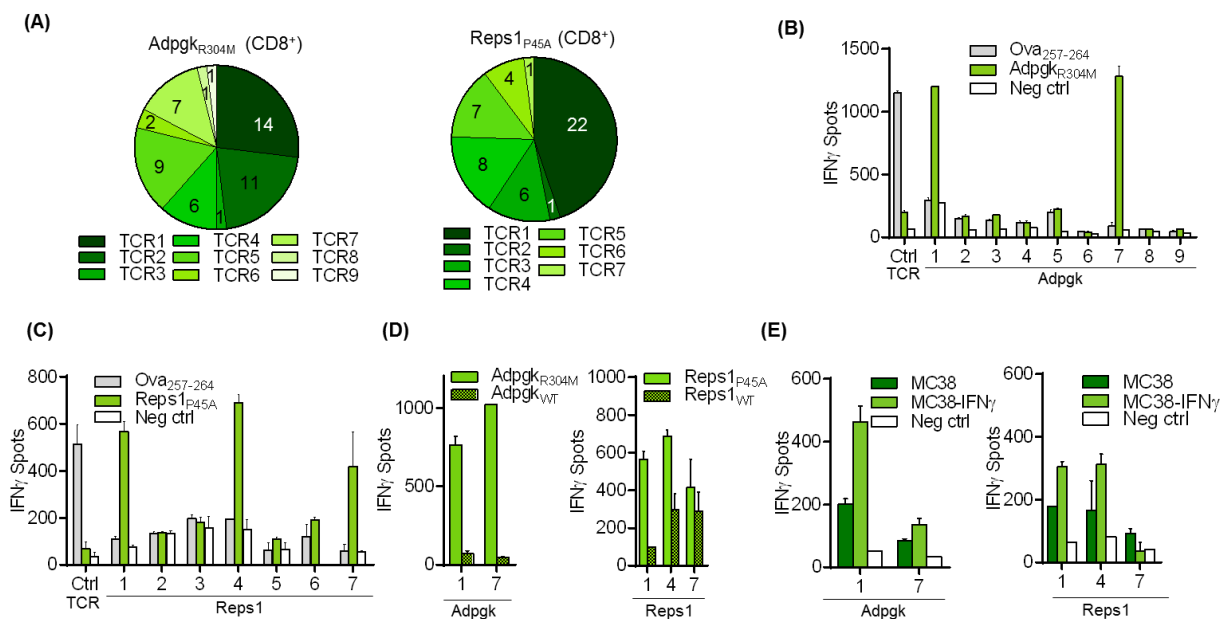
Figure 3.22| Neoantigen-specific TCRs discovered using scTCRseq mediate MC38 tumor reactivity. **(A)** Absolute TCR clonotype frequencies identified using scTCRseq of sorted CD8$^+$IFNγ^+ T cells from C57BL/6 WT mice vaccinated with 20 µg Adpgk$_{R304M}$ (ASMTNMEL) or Reps1$_{P45A}$ (AQLANDVVL) encoding RNA-LPXs. 12 h *in vitro* antigen-specific restimulation was performed. Characterization of **(B)** Adpgk$_{R304M}$ and **(C)** Reps1$_{P45A}$-specific TCRs using IFNγ ELISPOT assay. TCR-electroporated C57BL/6 T cells were co-cultured with BMDCs loaded with the respective peptide. Irrelevant peptide; Ova$_{257-264}$ (SIINFEKL), Positive control; Ova$_{257-264}$-specific TCR. **(D)** Adpgk_TCR1, Adpgk-TCR#7 and Reps1_TCR1 show neoepitope specificity. TCR-electroporated C57BL/6 T cells were co-cultured with BMDCs loaded with the respective WT or mutated peptide. **(E)** Adpgk_TCR1, Adpgk_TCR7, Reps1_TCR1 and Reps1_TCR4 recognize the MC38 tumor cell line (5x10^4 cells/well) with and without 24 h IFNγ stimulation to upregulate MHC class I expression. Mean + SD from technical duplicates is depicted.

Since a lower RNA-LPX vaccination dose could potentially induce the clonal expansion of higher-affinity TCR-bearing T cells only, a low RNA-LPX vaccination dose was combined together with tetramer sorting and scTCRseq in a next experiment, aiming at discovering better tumor cell line-recognizing TCR candidates. Vaccination of C57BL/6 mice with a 40-fold less RNA-LPX dose could induce an immune response for Adpgk$_{R304M}$ but not for Reps1$_{P45A}$ **(Figure 3.23A)**. This was in accordance with Adpgk$_{R304M}$'s higher immunogenicity as previously described in Yadav *et al.* (2014)[139]. ScTCRseq of tetramer positive splenocytes isolated from mouse M1, which showed the strongest response, led to the identification of seven additional unique TCRs (Adpgk_TCR10-16) **(Figure 3.23B)**. Surprisingly, Adpgk_TCR12 and #13 did share the same β TCR chain as the previously identified tumor cell line-recognizing Adpgk_TCR1

and #7, respectively. Furthermore, their respective α TCR chains were nearly identical. These TCRs were predicted to be antigen-specific. Out of the seven newly detected TCRs, five (Adpgk_TCR11-15) could mediate reactivity against the mutated Adpgk peptide **(Figure 3.23C)**. These TCRs were mutation specific. Adpgk_TCR12 and #13 did recognize the peptide as predicted. All five TCRs could recognize the tumor cell line with and without IFNγ stimulation compared to control **(Figure 3.23D)**. The newly identified Adpgk_TCR11 and #13 showed the strongest response out of all TCRs including those previously identified.

Co-culturing the TCR-electroporated C57BL/6 T cells with BMDCs loaded with decreasing Adpgk$_{R304M}$ peptide amounts showed a higher affinity of the low dose vaccination-derived TCRs **(Figure 3.23D)**. In particular, Adpgk_TCR12 showed higher affinity compared to the highly similar Adpgk_TCR1 derived from the higher dose vaccination. Further, Adpgk_TCR13 showed greater affinity compared to the highly similar Adpgk_TCR7 from the higher dose vaccination. Adpgk_TCR11 from the low RNA dose vaccination showed the highest affinity at lower peptide concentrations.

Therefore, combining low dose RNA vaccination with tetramer sorting led to the identification of higher affinity Adpgk-specific TCRs, two of which could mediate a stronger response against the MC38 tumor cell line compared to the high RNA dose-derived TCRs.
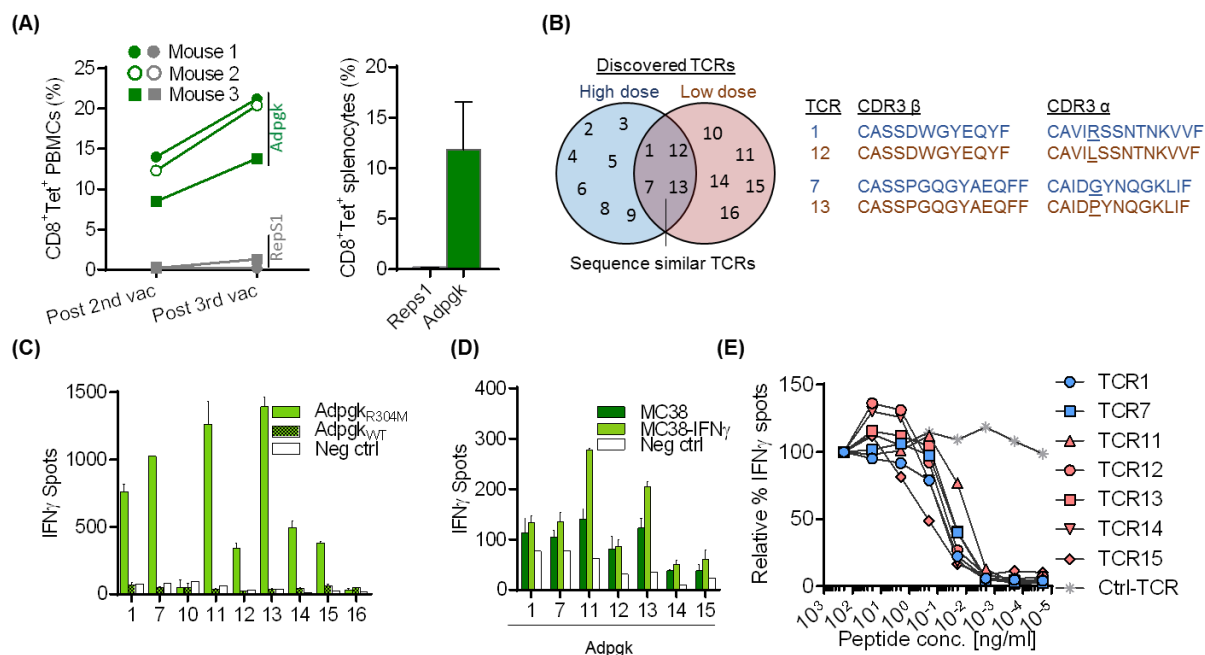
Figure 3.23| Detection of additional neoantigen-specific TCRs despite the use of a lower RNA vaccination dose. Vaccination of three C57BL/6 mice was performed using 0.5 µg Adpgk$_{R304M}$ (ASMTNMEL), 0.5 µg Reps1$_{P45A}$ (AQLANDVVL) and 39 µg irrelevant RNA. **(A)** Mice could generate an

immune response against Adpgk$_{R304M}$ (Green) but not Reps1$_{P45A}$ *(Grey)* as determined by tetramer staining n=3. After third vaccination, CD8$^+$ Adpgk$_{R304M}$ tetramer$^+$ splenocytes were sorted and processed using scTCRseq from the best responding mouse (M1). **(B)** α and β TCR chain CDR3 regions of two newly discovered TCRs (Adpgk_TCR12 and #13 following 0.5 μg RNA vaccination) depicted compared to Adpgk_TCR1 and #7 which were discovered following 20 μg RNA vaccination. **(C)** TCR-electroporated C57BL/6 T cells were tested for reactivity against mutated or wild type Adpgk peptide-loaded BMDC. **(D)** TCR-electroporated C57BL/6 T cells were tested for reactivity against the MC38 tumor cell line with and without 24 h IFNγ stimulation to upregulate MHC class I expression. Mean + SD from technical duplicates is depicted. **(E)** TCR-electroporated C57BL/6 T cells were tested for reactivity against BMDCs loaded with decreasing Adpgk$_{R304M}$ peptide amounts. Positive control; OvaI$_{257-264}$-specific TCR.

3.6. ScTCRPhenoSeq: Integrating T cell phenotyping into the scTCRseq platform

3.6.1. Platform concept design

Combining TCR sequencing with T cell gene expression analysis can integrate information on T cell clonality as well as T cell function. Although flow cytometry-based staining can determine the phenotype of a T cell population prior to single cell sorting, intracellular staining of secreted proteins and cellular fixation is not compatible with NGS needed for TCR identification. Furthermore, antibody staining can be quite costly.

We therefore aimed to develop and optimize an improved version of the scTCRseq platform that detects the gene-expression levels of a selected panel of commonly used T cell markers (scTCRphenoSeq platform). Due to the use of the scTCRseq platform for TA-specific TCR detection, we chose a gene panel composed of three general T cell markers, CD3, CD8α and CD4, as well as five T cell activation markers: IFNγ, 41BB, CD40, TNFα and OX40. Such an approach might allow the replacement of the flow sorter with a simpler single cell dispenser.

The devised gene-targeted workflow at the molecular level is illustrated in **Figure 3.24**. Compared to the developed scTCRseq platform, the scTCRphenoSeq platform entailed four main differences:

- The use of a dT30VN priming during cDNA synthesis in addition to the TCR GSPs.
- Additional nested GSPs for the selected phenotyping markers during the RBC-incorporation PCR.
- Removal of the Poly(A) carrier RNA in the cell-capture buffer during cell sorting due to the use of dT-priming.
- The use of the Maxima H Minus RTase due to its higher processivity compared to the RevertAid H Minus RTase.
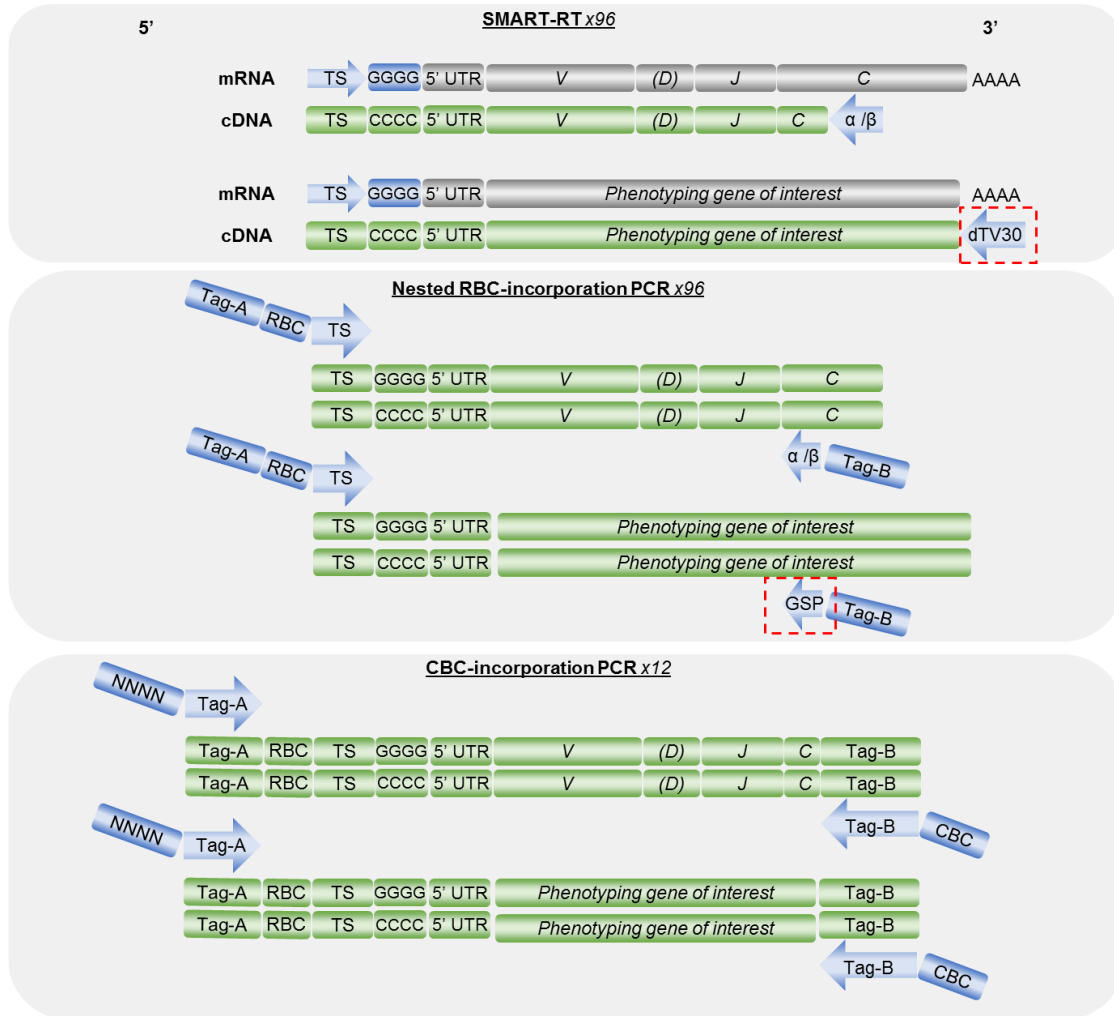
Figure 3.24| Schematic representation of devised workflow of the scTCRphenoSeq platform at the molecular level. Changes to the scTCRseq platform shown using a red box outline. During the cDNA synthesis reaction, TCR GSPs and dT-priming is performed. As before, RBCs and CBCs are incorporated during the first PCR and second PCR, respectively. During the RBC-incorporation PCR, additional phenotyping marker GSPs are used. V: Variable gene, D: Diversity gene, J: Joining gene, C: Constant gene, TS: Template-switch sequence, CCCC: Cytosines added by the terminal transferase activity of the MMLV RTase, GGGG: Riboguanines in TS primer and Guanines in cDNA, UTR: Untranslated region, RBC: Row-specific barcode, CBC: Column-specific barcode, NNNN: 5' variation. Primers and adaptors shown in blue, mRNA in grey and cDNA in green.

3.6.2. Selection of functional PCR primers for phenotyping gene amplification

For the chosen marker panel, GSP primers were designed for the RBC-incorporation PCR for all nine genes. Primers were designed to generate amplicons similar in length to the

TCR amplicons to allow purification of both phenotyping and TCR genes using the same bead-based size selection ratio (0.75x). All primer properties were kept similar to the TCR GSPs, to allow amplification of both phenotyping and TCR genes in the same PCR reaction.

Out of all designed primers for the RBC-incorporation PCR, the primer variants showing the strongest amplification of the phenotyping genes using spleen-derived bulk cDNA as a template (dT-primed), were chosen **(Figure 3.25)**. The band sizes of the amplicons of the different genes ranged from 550 to 800 bp. The Ampure bead-based size selection ratio was therefore slightly increased from x0.75 to x0.8.
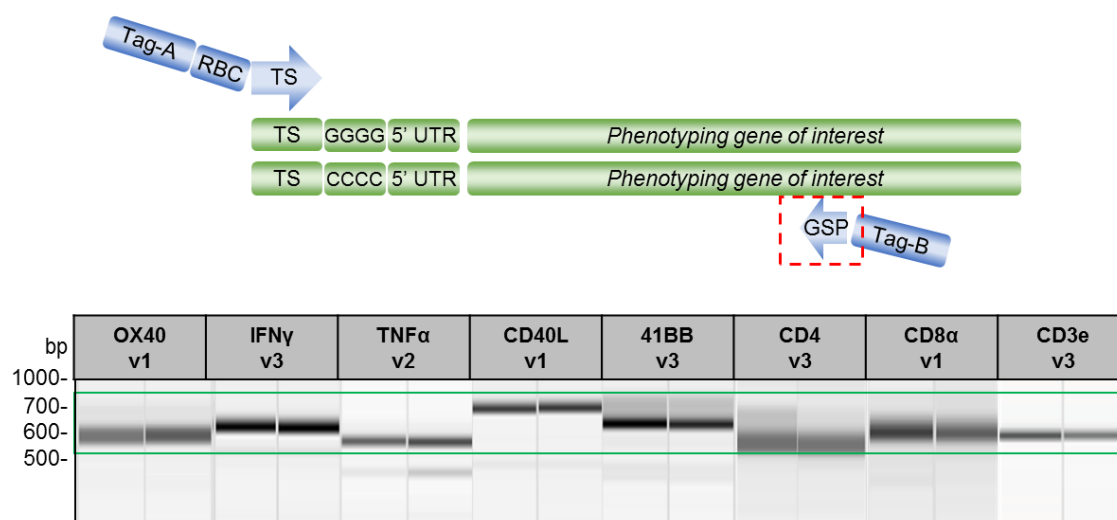
Figure 3.25| Amplification of mouse spleen RNA using the designed nested phenotyping marker GSPs in the RBC-incorporation PCR. Qiaxcel images of PCR amplicons (10uL/lane). Murine splenocytes were activated *in vitro* using CD3/28 beads for 24 hours prior to total-RNA isolation. TS RT was performed using dT-priming. First strand cDNA was pooled and diluted three-fold. PCRs were performed using scTCRseq conditions: PfuUltra HS DNA Polymerase, 0.2 μM GSP-Tag-B primer, 0.2 μM forward TS-RBC-Tag-A primer in 50 μL total reaction volume. 95°C 2 min; 94°C 30 s, 60°C 30 s, 72°C 1 min x 5 cycles; 94°C 30 s, 64°C 30 s, 72°C 1 min x 5 cycles; 94°C 30 s, 72°C 2 min x 8 cycles; 72°C 6 min. 10 μL first strand cDNA were used as template. TS: Template-switch sequence, CCCC: Cytosines added by the terminal transferase activity of the MMLV RTase, GGGG: Guanines, RBC: Row-specific barcode, UTR: untranslated region.

3.6.3. Phenotyping bioinformatic pipeline development

In order to determine the read counts of the amplified phenotyping genes of the processed T cells, the scTCRseq bioinformatic pipeline was extended to process non-TCR reads **(Figure 3.26)**. As before, 2x300 bp PE reads are demultiplexed, and the sequences successfully allocated to a well are processed via MiXCR. Sequences that are not aligned to TCR genes are further processed by being aligned to the reference sequences of the selected phenotyping marker genes. Successfully aligned reads were quantified to provide a read count level for each gene in the phenotyping marker panel for each cell. This data was then summarized into a .CSV file to facilitate manual reviewing.
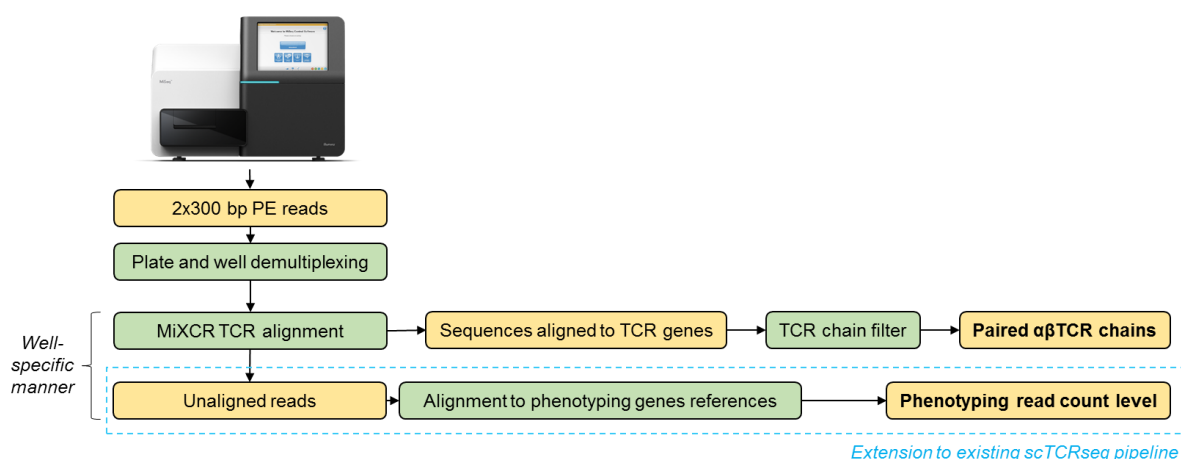
Figure 3.26| Extending the bioinformatic pipeline to allow alignment of NGS reads to the phenotyping gene reference sequences. NGS reads which are not aligned to TCR genes via MiXCR are further processed via the pipeline extension. MiXCR unaligned reads are aligned to the phenotyping gene reference sequences. Expression data is then summarized in a .CSV file to summarize the read counts of each of the chosen phenotyping markers for each cell. Green and yellow text boxes indicate pipeline action or input/output material, respectively.

3.6.4. ScTCRPhenoSeq of in vitro activated single T cells

To test the feasibility of the scTCRphenoSeq platform, single sorted CD8$^+$ and CD4$^+$ splenocytes, which were either unstimulated or stimulated for 12 or 24 hours *in vitro* using CD3/CD28 beads or PMA/Ionomycin, were processed **(Figure 3.27A)**. The latter activation reagent was used due to its stronger ability to upregulate CD40L expression on T cells. The BD FACS Melody was used to perform indexed cell sorting to enable the correlation of surface protein levels determined by antibody staining with RNA gene expression data (NGS). 96-well plates were sequenced at (3 x10^6 reads /plate or 3 x10^5 reads /cell).

Unstimulated T cells were sorted as CD8$^+$41BB$^-$ and CD4$^+$OX40$^-$, whereas stimulated T cells were sorted as CD8$^+$41BB$^+$ and CD4$^+$OX40$^+$ **(Figure 3.27B)**. An average read depth of 1.9 x 10^4 reads /cell (two thirds of the planned read depth) was obtained. The average demultiplexing rate was 83 % **(Figure 3.27C)**. 64 % of the reads aligned to TCR genes and 31 % to phenotyping genes. 80 % of the spiked in 2D2 αβTCR pairs were detected and 67 % of the αβTCR chains of the sorted single splenocytes were detected. Minimal noise was observed. Aligning the remaining 31 % of NGS reads to the phenotyping marker reference sequences allowed a clear differentiation between CD8$^+$ and CD4$^+$ T cells. Compared to the flow-cytometry antibody staining, the NGS data showed a true positive rate of 63 % and 60 % for CD8$^+$ and CD4$^+$ T cells, respectively **(Figure 3.27D).** All sorted T cells showed a CD3$^+$ true positive rate of 40 %.

The gene expression data revealed the presence of cells with increased read count levels of the activation markers following *in vitro* stimulation using CD3/28 beads and /or PMA/Ionomycin compared to no stimulation **(Figure 3.27E)**. OX40, 41BB, IFNγ and TNFα were expressed in both CD8$^+$ and CD4$^+$ T cells, whereas CD40L expression was limited to CD4$^+$ T cells[140–142]. 85 % of all stimulated T cells expressed at least one activation marker (data not shown). The different activation marker genes showed different read count level kinetics at the different time points as well as when using the two different *in vitro* T cell stimulation methods. CD3/28-induced increases in OX40, 41BB, CD40L and TNFα expression on CD4$^+$ and CD8$^+$ T cells as well as IFNγ on CD8$^+$ T cells persisted at 24 hr stimulation. Whereas PMA/Ionomycin-induced increases in activation marker read count levels at 12 hr stimulation were followed by a downregulation at the 24 hr stimulation time point. Representative examples of the α and β TCR chain CDR3 regions are annotated for random cells **(Figure 3.27E).**

There was no direct correlation between NGS-based gene read counts and flow cytometry-based surface protein expression. A representative example for OX40 expression in CD4$^+$ T cells is shown in **Figure 3.27F**.

Together, the data reveals that the scTCRphenoSeq platform is able to perform basic phenotyping of the single sorted T cells as well as identify their αβTCR pairs. The scTCRphenoSeq platform protocol is shown in **Supplementary section 6.4**.

3.6.5. ScTCRPhenoSeq of in vivo expanded CAR+ T cells

The application of chimeric antigen receptor (CAR)-redirected T cells for therapy of solid cancers is hampered by the poor expansion and persistence of transferred T cells. Colleagues have recently introduced the tight junction molecule Claudin-6 (CLDN6) as a new CAR target in solid cancers (Reinhard *et al.*, unpublished). They demonstrated that a nanoparticulate RNA vaccine designed to deliver the Claudin-6 antigen into lymphoid

compartments stimulates adoptively transferred CAR T cells, promoting their expansion and leading to an improved engraftment of CAR T cells after transferring lower CAR$^+$ T cell numbers **(Fig 3.28A)**.

To better understand the mechanism of action of such a vaccine on CAR T cells, we used the scTCRphenoSeq platform to compare the phenotype of CLDN6 CAR-transduced CD8$^+$ T cells adoptively transferred into irradiated C57BL/6 mice vaccinated with either relevant or irrelevant antigen-encoding RNA-LPX at day 7 post adoptive transfer. Their clonality was determined based on the endogenous TCRs **(Figure 3.28A)**. Single CD8$^+$CAR$^+$ T cells were sorted from pooled splenocytes of five mice each at 10 (TP1) and 14 days (TP2) post adoptive transfer. 264 CAR$^+$ T cells were processed with an average αβTCR detection rate of 73.25 % **(Figure 3.28B)**. We show the ability to detect 41BB, IFNγ and OX40 expression in the *in vivo* expanded CD8$^+$ T cells **(Figure 3.28B)**. At TP1, there is an enhanced expression of all three activation markers (41BB, IFNγ and OX40) in CD8$^+$CAR$^+$ T cells from the CLDN6-vaccinated mice compared to the CD8$^+$CAR$^+$ T cells from the control group mouse vaccinated with irrelevant RNA-LPX **(Figure 3.28C)**. At TP2, only 41BB and OX40 showed elevated levels in T cells from the CLDN6 vaccinated mice compared to T cells from the control mouse whilst IFNγ showed no difference. As in the previous experiment, no CD40L or TNFα expression was seen (data not shown). Analysis of the endogenous αβTCRs of the CD8$^+$CAR$^+$ T cells revealed no clonal expansion at TP1, whereas some clonal expansion was observed at TP2 **(Figure 3.28D)**. In particular, two clones (blue and yellow) showed an expansion in the vaccinated group compared to the control group. This is potentially indicative of the expansion effect of the RNA-LPX vaccination on the CAR$^+$ T cells. TCR chains found at TP1 and TP2 were unique.
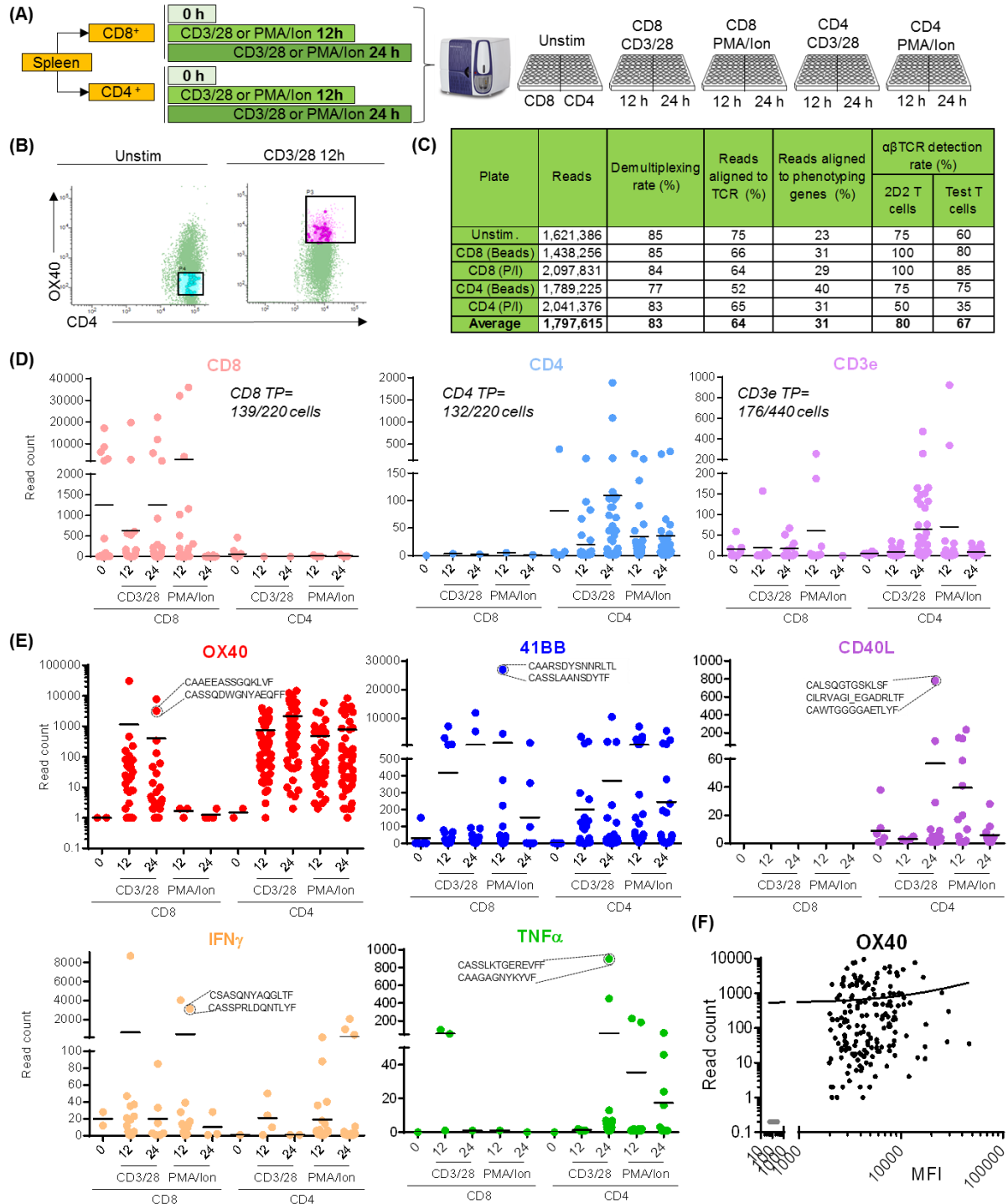
Figure 3.27| Performance evaluation and benchmarking of the scTCRPhenoSeq platform using single murine T cells. (A) Experimental workflow. Single sorted CD8+ and CD4+ splenocytes, which were either unstimulated or stimulated for 12 or 24 hours *in vitro* using CD3/CD28 beads or PMA/Ionomycin, were processed using the scTCRphenoSeq platform. **(B)** Representative gating strategy of unstimulated and stimulated T cells. **(C)** Quantitative platform performance indicators for all sequenced plates. Mean read count level of phenotyping genes determined using the scTCRphenoSeq platform for **(D)** CD8, CD4 and

85

CD3 as well as for **(E)** OX40, 41BB, CD40L, IFNγ and TNFα. Representative examples of the α and β TCR chain CDR3 regions are annotated for random cells. TP, true positive value based on flow cytometry staining. **(F)** Correlation between gene read counts and protein surface expression (MFI; mean fluorescence intensity). OX40 expression in CD4+ T cells shown as an example.
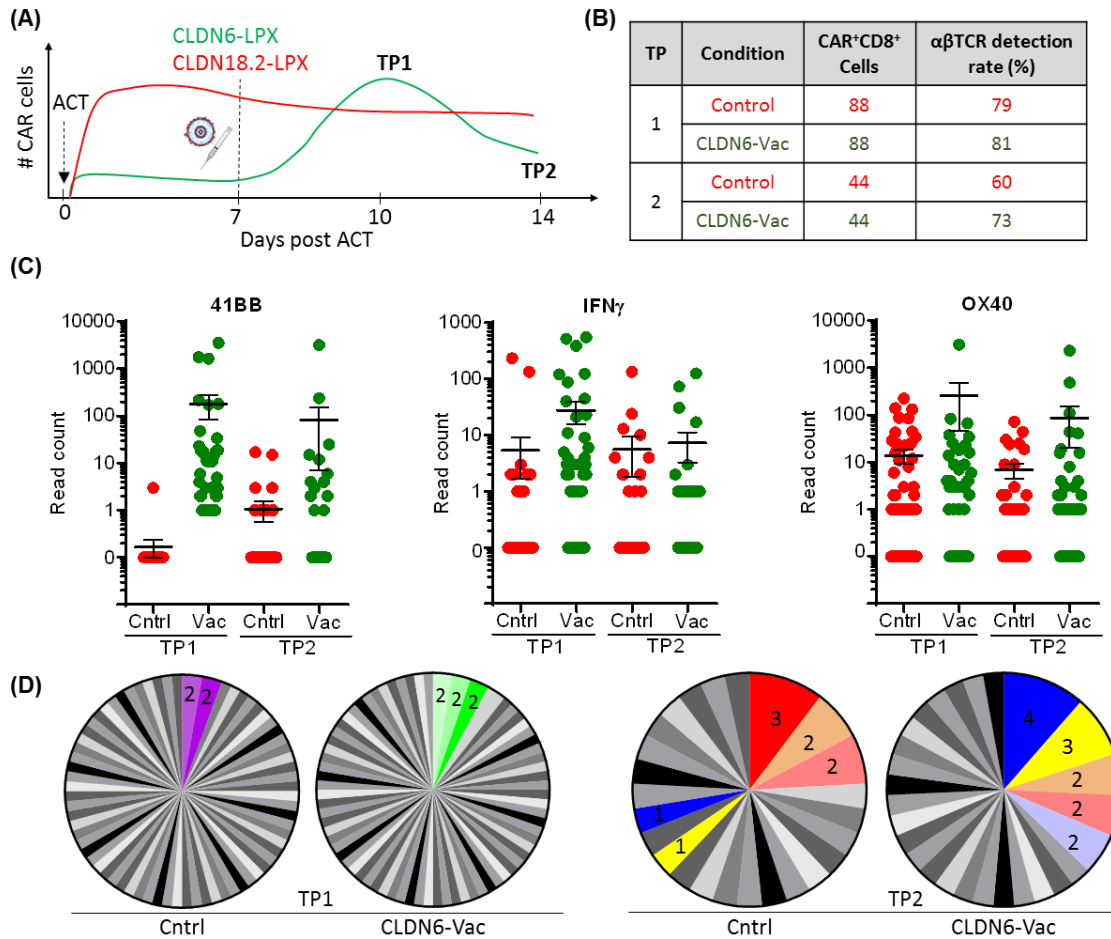
Figure 3.28| Phenotyping CLDN6 CAR+ T cells expanded *in vivo* using RNA-LPX vaccination reveals an enhanced activation profile. (A) Experimental workflow and expected CAR+ T cell frequencies *in vivo* after RNA-LPX treatment. Pre-conditioned naïve C57BL/6 Thy1.1+ mice (n=20) were engrafted with Thy1.1+ eGFP+ CLDN6 CAR T cells at day 0 (1.5 x10⁶ and 7.5 x10⁶ CAR+ T cells for CLDN6-vaccinated (n=10) and control group (n=10), respectively). The test mice (green line) subsequently received a CLDN6-LPX vaccination at day 7. The control mice received a vaccination at day 7 with an irrelevant antigen CLDN18.2 (red line). CLDN6 CAR+CD8+ T cells from the pooled splenocytes of mice (n=5 mice per TP per group) were sorted at time point (TP) 1 and TP 2 (day 10 and 14) after the vaccination and were processed using the scTCRphenoSeq platform. **(B)** Number of CLDN6 CAR+ T cells processed per condition and respective αβTCR detection rates. **(C)** Read count levels of phenotyping genes of CAR+CD8+ T cells from control (cntrl) and CLDN6-vaccinated (vac) mice at TP 1 and TP 2. Mean ±SD shown. **(D)** Clonality analysis of endogenous αβTCR of CD8+CAR+ T cells at TP1 and TP2. Grey shades indicate a clonal abundance equal to one. Colors indicate a clonal abundance greater than one and the same non-grey shade colors indicate the same clone.

4. Discussion

4.1. Development of an NGS-based platform for high-throughput detection of tumor antigen-specific αβTCRs from single murine T cells

Genetically engineering autologous T cells to bear tumor antigen (TA)-specific T cell receptors (TCRs) can generate *de novo* anti-tumor immune responses after reinfusion into cancer patients [96–98]. The vision of establishing TCR libraries for "off-the-shelf" therapy and the importance of targeting multiple TAs to circumvent tumor escape necessitates efficient TCR discovery. Further, murine TCRs targeting murine TA are essential for pre-clinical mode of action studies and proof of concept investigations required to further understand and improve ACT protocols. Existing single cell TCR sequencing methods suffer from cumbersome workflows, low throughput and high costs resulting in a paucity of suitable murine TCR candidates[127,128,143].

Encouraged by this, this thesis aimed to develop a high-throughput and easily applicable scTCRseq platform that would facilitate the discovery of shared and neoepitope-specific murine TCRs.

In this thesis, a scTCRseq approach was developed that adopts a plate-based workflow compatible with NGS and requires only standard laboratory equipment. Sample barcoding permits early pooling of single cell samples and multiplexing of 96-well plates allowing little hands-on work at a high-throughput. We integrate an end-to-end data analysis pipeline to retrieve the paired αβTCR information from the 2x300 bp paired end MiSeq data. The platform makes use of template switching (TS) technology to minimize amplification bias and introduces universal primer sequences to reduce PCR complexity. The platform was optimized to improve αβTCR detection rates six-fold and eradicate platform noise, reduce costs significantly and reduce the protocol duration to three hours hands-on work until the library preparation. The technology was validated using spiked-in TCR-encoding *in vitro*-transcribed (IVT) RNA libraries and 2D2 mouse-derived CD4$^+$ T cells with known transgenic TCR genes. Furthermore, the developed platform showed superior performance compared to the *in house* used single cell TCR cloning platform developed by Omokoko *et al.*(2014). Comparing the scTCRseq data to TCR repertoire profiling showed the ability of the former to perform moderate repertoire analysis. We demonstrate the utility of this technology by combining it with TA-encoding RNA or peptide vaccination in mice to rapidly compile a library of shared and neo-epitope-specific TCRs. Parallel work focused on extending the platform to allow the integration of TCR data with T cell phenotyping information using a targeted gene panel of commonly used T cell markers.

4.2. Positioning the developed murine scTCRseq platform within the field of TCR sequencing

At the start of this thesis, the two principal plate-based cell TCR sequencing technologies were the human and mouse Sanger-sequencing-based TCR cloning platforms described by Omokoko *et al.* and the Next-generation sequencing (NGS)-based human scTCR sequencing platform by Han *et al.* [64,127,128].

Due to the strong demand for TCR analysis and discovery, several NGS-compatible single cell TCR sequencing technologies have emerged throughout this study's duration (**Figure 4.1**)[144–146]. The emerging spectrum of single cell TCR sequencing technologies can be classified into plate or fluidic-based technologies. Both approaches show compatibility with human and/ or murine T cells. Despite using plates, advancements in DNA barcoding and multiplexing strategies have enabled plate-based approaches to reach a throughput of up to thousands of single T cells. A non-plate based and commercially available approach described involved TCR sequence rescue from single cell RNA-seq data using the valve-based microfluidic Fluidigm C1 system[143]. Cells are loaded onto an integrated fluidics circuit and captured into small chambers. After cell lysis, SMART dT-primed cDNA synthesis is performed. The chemistry does not allow targeted TCRseq and therefore TCR sequence reconstruction is performed using complex bioinformatic pipelines. An example of this is TraCeR, which aligns the fragmented TCR sequences to reconstruct full-length TCR contigs sequences from single cell RNAseq data[143]. In parallel, recent literature describes droplet-based microfluidic devices that in contrast to valve-based systems encapsulate hundreds of thousands of T cells separately into oil emulsions where they are lysed, the mRNA reverse transcribed and barcoded; thus, benefiting from less hands-on work compared to plate-based approaches.
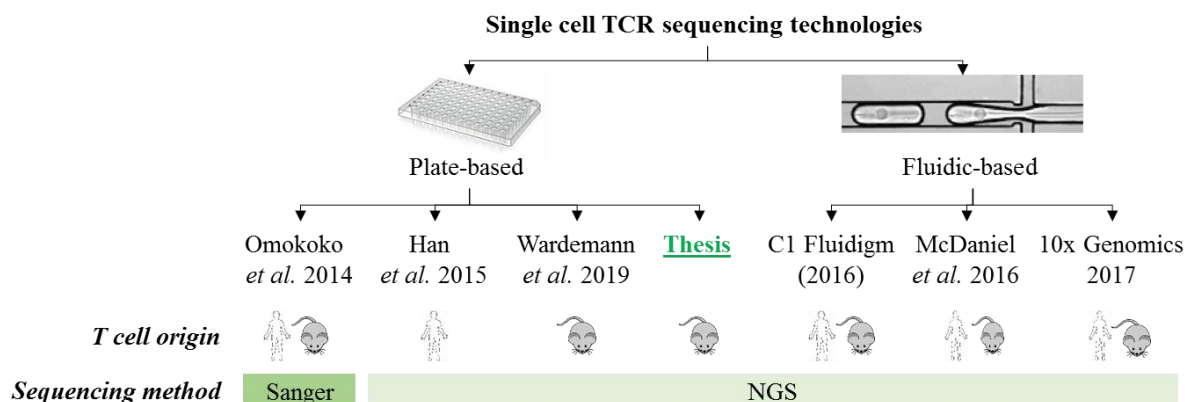
Figure 4.1| Current landscape of single cell TCR sequencing technologies. Technologies are classified as plate or fluidic-based. The murine scTCRseq technology during this thesis is shown in green font.

In spite of the higher throughput of droplet-based microfluidic devices, the applicability of plate-based approaches is higher. Whilst commercially available fluidic instruments such as the 10x system are easily accessible but very costly, fluidic approaches described in the literature for self-assembly such as that described by McDaniel *et al.* (2016) are cheaper but not easily implemented. Further, full-length gene products for direct cloning to enable functional TCR analyses are typically not readily available when using fluidic approaches. We present the suitability of the developed scTCRseq for TA-specific TCR discovery by benefiting from remaining first strand cDNA, which can be used after data analysis for cloning of full-length TCR sequences of interest.

In this discussion, the developed scTCRseq platform is predominantly compared to other plate based approaches, and where possible, to fluidic-based approaches.

αβTCR detection efficiency

Quantitative performance evaluation of the developed scTCRseq platform revealed that 81 % of NGS reads were demultiplexed indicating efficient barcoding and demultiplexing. Along with the integrated demultiplexing bioinformatic pipeline, this might be attributed to the primer-tags and barcodes existing in the high quality locations within the NGS reads. Han *et al.* similarly showed 80% demultiplexing rates. Additionally, we show successful alignment of 90 % of the sequenced reads suggesting accuracy of the developed platform. This is in accordance with only 3.4 % of sequences entailing PCR errors and might be attributed to the use of two high-fidelity Polymerases and the use of only two PCRs each with low cycle numbers (total 45 cycles). Omokoko *et al.*, Han *et al*. and Wardemann *et al.* perform 61, 86 and 100 PCR cycles, respectively.

The established scTCRseq platform showed detection of αβTCR pairs in 49 % of processed T cells compared to 40, 82 and 52 % obtained using the Omokoko *et al.*, Han *et al.* and Wardemann *et al.* technologies, respectively. Despite limiting its throughput due to the very deep sequencing per cell (minimum of 1×10^6 reads/cell), the TCR rescue from RNA-Seq data using the C1 Fluidigm system reported 70-93 % αβTCR detection rates. The 10x Genomics technology reports typical detection rates between 40 and 70 % depending on the read depth. McDaniel *et al.* do not mention detection rates. The exact noise-filtering strategies used by most other technologies, if at all, is not known and our use of a strict noise-filtering system to remove 'unreal' TCR chains could be a contributing factor to not having even higher detection rates. This precaution is at large driven by the high costs of subsequent TCR cloning and downstream TCR antigen-specificity determination experiments.

15 % of cells possessed a second α TCR chain in line with previous observations reporting the biallelic expression of α TCR chains. The platform's sensitivity was tested by spiking

in 2D2 T cells. The paired αβTCR chains of 77 % of 2D2 T cell was detected indicating high platform sensitivity. Having higher αβTCR detection rates from the 2D2 T cells compared to the T cells from vaccinated mice could in part be due to higher TCR mRNA expressed in the former cells due to constant MOG$_{35-55}$ epitope encounter in the host 2D2 transgenic mouse model.

Throughput

The use of library preparation dual-indices and subsequent plate multiplexing in the developed platform permits a maximum throughput of 9,216 cells; higher than the C1 and Wadermann *et al.* (2019) approaches (96 and 3456 cells, respectively) and similar to Han *et al.* (5,000-10,000 cells). Despite impressive throughput of recent fluidic approaches such as that described by McDaniel *et al.* (greater than $1x10^6$) or the 10x Genomics system (8×10^5), such technologies suffer from 35-50 % cell loss during cell encapsulation together with 40-70 % paired αβ TCR retrieval rates.

Throughput is based at large on sample barcoding and pooling. Fluidic-based technologies benefit the most from this due to barcode attachment during RT. The developed scTCRseq platform incorporates its well-based barcodes during the first PCR allowing column pooling and collapsing of samples from 96 to 12. This is followed by row pooling after the second PCR and the collapsing of samples down to one. This early sample pooling thus facilitates parallel plate processing, reduces protocol complexity and costs. In contrast to that, the TCR cloning platform from Omokoko *et al.* does not use sample barcoding and Han *et al.* as well as Wardemann *et al.* only perform barcoding during the last PCR resulting in handling of either 96 or 384 samples till and including the last PCR.

Throughput gains from multiplexing come with an added layer of complexity since 'barcode hopping' can be a concern[147]. This is the incorrect assignment of sequences from an expected barcode to a different barcode. An issue faced during the development of the scTCRseq platform was the presence of clones with high read counts generated noise in neighbouring wells of the same row and column. This noise was hypothesized to originate from leftover RBC primers being falsely incorporated into cDNA from other wells in the second PCR after column pooling. We significantly reduced this noise by eliminating leftover RBC primers after column pooling using Exonuclease I treatment. A recent paper by Guo *et al.* (2018) also address such noise during RNA-Seq using the same approach [148].

Moreover, all previously published plate-based technologies perform separate PCRs for the α and β TCR chains doubling platform pipetting steps and reagent use. By taking advantage of the incorporated TS sequence in the first strand cDNA and the additional primer Tag-B in the first PCR, we can use the same primers for the α and β TCR chains,

allowing the amplification of both chains in one reaction and further minimizing the required PCRs. The established scTCRseq platform, similar to the TCR cloning platform Omokoko *et al.* and Wardemann *et al.*, also benefits from only having two PCR steps compared to Han *et al.* who has a third PCR for barcode incorporation.

Flexibility in throughput represents an advantage for the processing of tumor-derived T cells. Fluidic approaches possess a high lower-limit of input cell numbers. Together with the described cell loss during encapsulation, both mentioned issues would pose a concern when sequencing tumor-derived T cells, where the availability of resectable material with infiltrated T cells is often a problem.

Read depth per cell

Throughput per run is also grounded upon the read depth necessary per cell since NGS flow cells have limited read counts. The gene-targeted approach used in the established scTCRseq platform, which was also adopted by Han *et al.* (2015), imposes an advantage over RNA-seq based approaches. Although the latter, as a standard, provides gene expression data, the required read-depth significantly restricts throughput and surges sequencing costs. RNA-seq technologies require at least $1x10^6$ reads per cell and dedicated bioinformatic tools to fish out and reconstruct TCR sequences from RNA-seq data. With an average read depth of 14,287 reads per cell, the established scTCRseq platform requires 70-fold less sequencing depth for the same cell number. Han *et al.* use 7,382 per cell. We show that in our platform, there is no direct correlation between read depth and αβTCR detection rates suggesting the ability to further reduce read depth as shown in some platform runs with 9000 reads/cell. The adopted targeted approach therefore further increases throughput and cuts platform costs when using the Illumina MiSeq platform.

Primers

In addition to early sample pooling, a platform's robustness and costs are influenced by the number of primer used. In the established platform, only 27 primers are used throughout the process. The use of a forward TS-containing primer during RT avoids the need for TCR Variable (V) gene-specific primers (GSPs). Omokoko *et al.* , Han *et al.* and Wardemann *et al.*, alone in their multiplex PCRs, use 60, 76 and 41 primers, respectively. Furthermore, Wardemann *et al.* use random-hexamer-primers for cDNA synthesis as well as 47 primers for the multiplex PCRs. McDaniel *et al.* also use multiplex PCRs. Therefore, indicating that the primer-PCR strategy adopted poses an advantages over other platforms. TS inefficiency has been previously described in TS-based systems[130]. Although efficient enough in our hands, the cDNA library yield using TS can be further improved through the modification of the RT reaction contents as described by Picelli *et*

al. (2013) or the use of RTases with higher processivity such as the Maxima H Minus Reverse Transcriptase[130].

TCR repertoire profiling

We highlight the ability of the developed scTCRseq platform to provide basic insight into the repertoire by comparing its data to the TCR repertoire profiling data of splenocytes from the same KKLC1-vaccinated mouse. Relative frequencies obtained using the scTCRseq platform were in accordance to those obtained using repertoire profiling suggesting that information from a representative sample of the T cell population can be gathered using the developed platform. The α and β TCR chains of the two most frequently detected TCRs in the scTCRseq data were also the two most abundant α and β TCR chains in the TCR repertoire data. The top three α TCR chains in the repertoire profiling and the scTCRseq data were the same and represented the α TCR chains of three out of the four KKLC1-specific TCRs. Frequency-based matching of α and β TCR chains has been attempted to determine αβTCR pairs α and β from TCR repertoire data. By using the pairing data from the scTCRseq platform, we show that this is only possible for the top clones and clonal repertoires but the chances of identifying the corresponding chains of the less frequent clones or in less clonal repertoires is highly unlikely.

T cell phenotyping

Omokoko *et al.* and Wardemann *et al.* perform phenotyping prior to sorting using flow cytometry antibody staining. Although index sorting can correlate protein level expression with the sorted cell, as performed by the latter platform, intracellular staining of secreted proteins and fixation is not compatible with NGS. Furthermore, antibody-staining can be quite costly. RNA-seq approaches, as with the C1 Fluidigm and 10x Genomics systems can provide detailed RNA expression data but require very deep sequencing. A plate-based phenotyping approach was performed by Han *et al.* for 34 genes. However, in that approach, TCR and phenotyping genes were amplified in different PCRs making the approach more tedious and costly. Here, we apply a chosen phenotyping panel of eight commonly used T cell markers to phenotype T cells. This targeted approach allows basic phenotyping without requiring too much read depth (2×10^4 reads/cell). Unlike Han *et al.*, we show the ability to amplify the α and β TCR chains as well as all the phenotyping genes in the same PCR; probably due to the lower numbers of used primers (TCR V-specific multiplex and phenotyping primers).

Section conclusion

Therefore, the developed scTCRseq platform displays high throughput and αβTCR detection rates **(Table 4.1)**. Compared to other plate-based approaches, the established

platform benefits from earlier barcoding and sample pooling, the use of less primers and PCR reactions. We show the ability to provide basic insight into the repertoire and phenotype T cells. Compared to fluidic approaches, we show both easy accessibility and application as well as the benefit of cloning TCRs of interest for downstream application.

Attribute	scTCRseq	Omokoko et al.(2014)	Han et al.(2015)	Wardemann et al. (2019)	C1 Fluidigm	Fluidics
αβ TCR deteciton rate	++	++	+++	+++	+++	++
Cell number throughput	+++	+	+++	++	+	+++
Early barcoding and pooling	+++	+	+	++	+++	+++
Read depth necessary per cell	+	N/A	+	+	+++	+
Number of primers used	+	+++	+++	+++	+	+
Ease of laboratory equipment and practices	+++	++	+++	+++	+	+
Ability to clone TCR of interest	+++	+++	-	+++	+	+
Ability to process low cell numbers	+++	+++	+++	+++	+++	+
Ability to perform repertoire analysis	++	+	++	++	+	+++
Ability to phenotype T cells	+++	-	+++	-	+++	+++

Table 4.1| Comparing technical attributes of existing scTCRseq technologies. +: Low, ++: Medium, +++: High, - : none, N/A: not applicable.

4.3. Rapid discovery of shared and neoepitope-specific T cell receptors

Tumor immunology studies have generated libraries of well-studied TAs including cervical cancer-inducing HPV antigens E6 and E7 as well as HBV antigens S and C proteins in Hepatocellular carcinoma[61,149,58]. Others include cancer-germline antigens such as KKLC1 expressed in breast and gastric cancers[64]. Preceding work has revealed the efficiency of eliciting TA-specific T cell responses by immunizing mice by means of TA-encoding RNA or peptides[85,134]. Therefore, we combined our newly developed scTCRseq platform with TA-encoding RNA vaccination to identify nine TA-specific TCRs specific for these TAs **(Table 4.2).**

Antigen		TCR	Epitope	MHC restriction	Reference	TCR β chain			TCR α chain		
						CDR3	V	J	CDR3	V	J
Oncoviral	E7	E7_TCR1	YMLDLQPETT	I	A2DR1	Ressing *et al.* 1996	CASSQDPGAEQFF	2*01	2-1*01	CAVTALSNYNVLYF	9D-4*04	21*01
	E6	E6_TCR1	DKKQRFHNIRGRWTG	II	A2DR1	Man *et al.* 2005; Piersma *et al.* 2008	CTCSADWGDEQYF	1*01	2-7*01	CAASMPNYAQGLTF	7-2*02	J6*01
		E6_TCR4					CTCSADRDLNTLYF	1*01	2-4*01	CAVSSASSGSWQLIF	7N-5*01	22*01
	HBs	HBs_TCR1	RGLYFPAGGSSSG	II	C57BL/6	Pajot *et al.* 2004, Pajot *et al.* 2006	CASSYWSNYNSPLYF	4*01	1-6*01	CAAGNQGGRALIF	7-4*02	15*01
	HBc	HBc_TCR5	PPAYRPPNAPIL	II	C57BL/6	Milich *et al.* 1995	CAWSPGGVYAEQFF	31*01	2-1*01	CAVESNYNVLYF	3D-3*02	21*01
C/T	KKLC1	KKLC1_TCR1	ILNNFPHSI	I	A2DR1	Sahin *et al.* 2015	CASSPGLGGGYEQYF	19*03	2-7*01	CAVSANSGTYQRF	3D-3*02	13*01
		KKLC1_TCR3					CTCSADRGINYAEQFF	1*01	2-1*01	CASWGANKMIF	12D-2*01	47*01
		KKLC1_TCR4					CASSSGLGGGYEQYF	4*02	2-7*01	CAVATSSGQKLVF	3D-3*2	16*01
		KKLC1_TCR6					CASSPGLGGGYEQYF	5*05	2-7*01	CAVSEGADRLTF	3-3*01	45*01
Neoantigens	Adpgk	Adpgk_TCR1	ASMTNMEL	I	C57BL/6	Yadav *et al.* 2005	CASSDWGYEQYF	13-1*02	2-7*01	CAVIRSSNTNKVVF	7-1*01	34*02
		Adpgk_TCR7					CASSPGQGYAEQFF	5*01	2-1*01	CAIDGYNQGKLIF	13D-2*01	23*01
		Adpgk_TCR11					CASSLVGGARGAETLYF	3*01	2-3*01	CAVSRATGGNNKLTF	9N-3*01	56*01
		Adpgk_TCR12					CASSDWGYEQYF	13-1*02	2-7*01	CAVILSSNTNKVVF	7-1*01	34*02
		Adpgk_TCR13					CASSPGQGYAEQFF	5*01	2-1*01	CAIDPYNQGKLIF	13D-2*01	23*01
		Adpgk_TCR14					CASSETGFYEQYF	17*01	2-7*01	CALGIYQGGRALIF	6D-4*01	15*01
		Adpgk_TCR15					CASSRTGGYEQYF	17*01	2-7*01	CALAIYQGGRALIF	6D-4*01	15*01
	Reps1	Reps1_TCR1	AQLANDVVL	I	C57BL/6		CASSPGQSQNTLYF	12-1*01	2-4*01	CALSSSSGQKLVF	6D-7*04	16*01
		Reps1_TCR4					CASSFYSNSGNTLYF	14*01	1-3*01	CALRRNSNNRIFF	12D-2*01	31*01
		Reps1_TCR7					CASSDARNSYNSPLYF	13-1*02	1-6*01	CALGDHYNQGKLIF	6D-6*02	23*01

Table 4.2| Shared tumor antigen and neoantigen-specific TCR library assembled using scTCRseq. The TCR V(D)J genes are indicated using IMGT nomenclature. C/T: Cancer/testis antigen; V: variable; D: diversity; J: joining; C: constant; MHC: major histocompatibility complex

From a clinical perspective, murine TCRs have been used to genetically engineer T cells for human ACT[150,151,97,152]. Particularly due the ability to avoid human central tolerance and the ease of obtaining samples when trying to identify TA-specific TCRs. Since we vaccinated human MHC-expressing A2DR1 mice, we provide new TCR candidates that target human TA epitopes for the potential treatment of E6, E7 and KKLC1 expressing tumors. For the former, such tumors include HPV positive head and neck or cervical cancer, and, for the latter, cancers such as breast and gastric cancers. As a next step, these TCR candidates would have to be validated for their safety and efficacy.

In parallel, it was shown in clinical trials that by targeting mutations using neoantigen-specific T cells, complete regression in metastatic cancer patients can be achieved[105,107,153]. Improvement of ACT-based approaches has relied in large part on the use of pre-clinical mouse models. To facilitate the investigation of the therapeutic effects of targeting neoantigens using ACT, we applied the scTCRseq platform to identify TCRs specific for mutated peptides described in the MC38 colorectal carcinoma mouse model. Seven and three TCRs specific for two immunogenic MHC class I epitopes Adpgk and Resps1 were discovered, respectively **(Table 4.2).** Furthermore, we combine low RNA dose vaccination with tetramer-based sorting to isolate higher affinity TCRs. We present nine MC38 neoantigen-specific CD8 TCRs that can recognize the endogenously expressed mutation in the MC38 tumor cell line. From a pre-clinical perspective, the discovered cell line-recognizing TCRs could be used in the context of the MC38 colon adenocarcinoma mouse models to optimize neoantigen-specific ACT protocols. Such

optimizations could include testing the ability to avoid tumor immunoediting and escape by using multiple TCRs or targeting both neoepitopes. Further, recent findings display the essential contribution of tumor-specific CD4$^+$ T cells to effective antitumor responses[85]. Using the developed scTCRseq platform to identify MHC II epitope-specific TCRs to study the anti-tumoral mechanisms of CD4$^+$ T cells is of great importance. A more detailed understanding of such questions through pre-clinical studies will likely be translated into the clinic for improved ACT protocols.

Together, we show that we could identify at least one functional TCR for each TA indicating platform robustness. Furthermore, when benchmarked against the platform developed by Omokoko *et al.*, the scTCRseq platform could identify additional E6 and KKLC1-specific TCRs indicating higher TA-specific TCR detection efficiency.

Interestingly, all TA-specific TCRs could be retrospectively characterized using three properties: read count level, clonal expansion and CDR3 similarity **(Figure 4.2).** All nineteen functional TA-specific TCRs had α and β TCR chain read counts in the upper quartile of read count levels across the plate. This might be due to increased TCR mRNA expression following antigen encounter[106]. Of those, ten TA-specific TCRs were found in more than one well. This is in concert with the proliferative capacity of these TA-specific TCR-bearing T cells following antigen encounter[154]. A third category of TA-specific TCRs entailed six TCRs that recognized the same peptide and shared similar CDR3 aa sequences. These conserved aa positions might represent paratopes essential for peptide interaction. Some TCRs, such as KKLC1_TCR1 and KKLC1_TCR6 shared the same β but not α TCR chain, which is in accordance with the bigger role of β TCR chains when interacting with the peptide-MHC complex[2]. We could also show the sharing of the same β TCR chain in Adpgk_TCR1 and Adpgk_TCR12 or Adpgk_7_TCR1 and Adpgk_TCR13, which were obtained from two different mice; in agreement with the narrower TCR repertoire of mice compared to humans[155]. Despite having such similar sequences, we show how a single aa change in the α TCR chain can enhance the affinity of the respective TCRs for the Adpgk$_{R304M}$ mutation. For TCRs KKLC1_TCR1 and KKLC1_TCR6, different TCR genes rearranged to generate the same respective CDR3 sequence. This phenomenon is known as convergent recombination and has been described previously in the literature[156–158,158]. Some TCRs existed in multiple categories. Although in some cases, non-TA-specific TCRs also fell into these categories (e.g. TCR HBVc_TCR6 was clonally expanded but did not recognize the antigen; probably due to recognizing another unknown peptide in the culture medium), these three attributes can be used to classify TCRs as potentially 'higher confidence' antigen-specific TCRs. This delivers a possibly new approach to select TCRs for cloning and validation instead of testing all TCRs. This is of great importance when sequencing larger T cell numbers. Indeed, several studies have shown CDR3 paratope sharing when retrospectively

analysing large antigen-specific TCR data sets[137]. This has also encouraged efforts to predict, *in silico,* the antigen-specificity of TCRs when the antigen is unknown. We therefore, in part, provide evidence for the feasibility of such an approach.
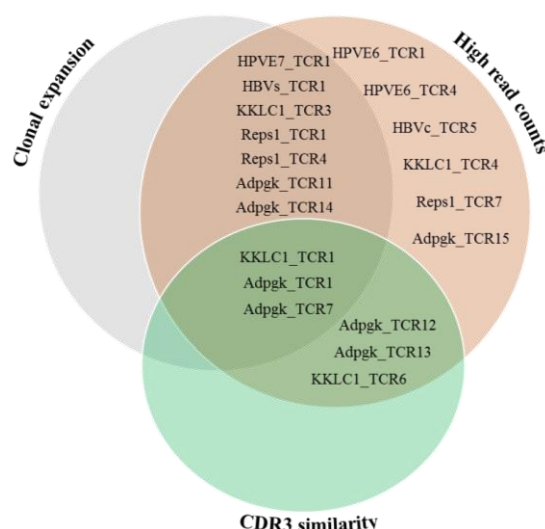
Figure 4.2| Classification of discovered TA-specific TCRs according to clonal expansion, read count level and CDR3 similarity.

4.4. Combing TCRseq with NGS-based phenotyping

There is great synergy in combining TCR sequencing with T cell gene expression data[128]. The former can deliver information on T cell ancestry and antigen specificity, whereas the latter can provide insight into T cell function; providing comprehensive T cell analysis. In this thesis, we developed and optimized an extended version of the scTCRseq platform that enables the read count level detection of a selected panel of commonly used T cell markers. The use of a targeted gene panel allows, together with paired αβTCR sequencing, to phenotype T cells at shallow sequencing depth. Due to the application of the scTCRseq platform for TA-specific TCR detection, we developed a panel composed of three general T cell markers, CD3, CD8a and CD4, and five T cell activation markers: IFNγ, 41BB, CD40, TNFα and OX40. Depending on the use of the platform, such a panel may be subject to addition, removal or replacement of the genes of interest by changing the used GSPs in the RBC-incorporation PCR. Further, due to its phenotyping ability, the scTCRphenoSeq platform might allow the replacement of the flow cytometer with simple single cell dispensing technologies such as the Nanocellect or the cellenOne platforms; rending the platform even more applicable for TA-TCR discovery based on activation marker levels.

The workflows of the scTCRseq and scTCRphenoSeq platform differ mainly in the GSPs used in the first PCR and the use of a RTase with higher processivity, abandonment of

carrier RNA and an increased read depth of $2x10^4$ reads/cell for the latter. Primers were designed in a way that all amplicons, including the TCR genes, are of equal length so that gene length does not affect read count levels.

We show the ability to determine RNA expression levels without the use of housekeeping genes for read count normalization. This is in concert with the adopted GSP targeted approach and simplifies data analysis as well as reduces the number of primers needed.

We demonstrate the ability of the scTCRphenoSeq platform to determine whether the processed cell is a T cell based on CD3, CD4 or CD8 expression in 89 % of cells. For 62 % of the cells, the platform can determine, based on RNA expression, whether it is a CD4[+] or CD8[+] T cell. Furthermore, the platform can be used to study T cell activation kinetics at the RNA expression level. Increasing RNA expression could be detected of all chosen T cell activation marker genes with increasing CD3/28 and PMA/Ionomycin–induced activation followed by reduction after 24 hour stimulation in some cases. As in concert with literature, we show 41BB expression on CD8[+] and CD4[+] T cells and favourable expression of OX40 and CD40L on CD4[+] T cells[140–142]. Furthermore, Han *et al.* could only detect yes/no expression signals for the different genes in their developed T cell phenotyping platform. In this thesis, we show the ability of the developed platform to detect gradual increases and decreases in activation marker levels following increasing T cell activation allowing the study of T cell activation kinetics.

Integrating phenotype gene amplification with TCR gene amplification showed no effect on TCR detection rates. We furthermore, show that there is no correlation between RNA and protein expression levels, which is in concert with literature describing rapid and dynamic protein surface expression and internalization[159–162].

Furthermore, we apply the scTCRphenoSeq platform to integrate phenotyping and TCR data from CLDN6 CAR[+] T cells with or without RNA-based *in vivo* expansion. The activation marker data showed a more activated state in the CAR[+] T cells of the CLDN6-vaccinated mouse. This indicates that the CLDN6-encoding RNA vaccine enhances the activation state of the adoptively transferred CAR[+] T cells. In parallel, we used the endogenous TCR sequences to study CAR[+] T cell clonality. The TCR information revealed the absence of strongly expanded T cell clones indicating the absence of clonal expansion caused by retroviral insertional mutagenesis; a frequent concern when using retroviral transduction for gene transfer. However, the TCR data revealed the ability of the RNA vaccine to induce CAR[+] T cell expansion. Together, this data reveals the ability of the scTCRphenoSeq platform to study not only unmodified T cells, but also genetically engineered CAR T cells as well as its ability to detect the activation of *in vivo* stimulated T cells.

4.5. Further improvement of the developed scTCRseq platform

Despite the adopted optimizations, there is room for further improvement. One approach involves adopting 384-well plates. This would require the establishment of additional eight compatible RBCs and twelve compatible CBCs. Since that would significantly increase manual pipetting, such an approach would work best with platform automation, which would also minimize the occurrence of pipetting errors. One further step could be the use of nano-well plates and the attachment of sample barcodes during RT. This too, would require the miniaturization of PCR reaction volumes and reagent amounts and implementation of nanoliter liquid handling robots. We show that unlike the PfuUltra HS DNA Polymerase, the Taq HS Polymerase permits volume reduction, however, is accompanied with reduced PCR accuracy. This suggests the importance of testing other proofreading Polymerases such as the Q5 High-Fidelity DNA Polymerase for compatibility with reduced volume. Another approach to enhance platform throughput, would be to eliminate the library preparation step, which calls for 2.5 hours hands-on work and 4.5 hours in total. This could be performed by modifying the CBC-incorporation PCR to entail the Illumina indices as well as the flow cell binding sequences and sequencing primer sites. This would allow direct loading of the pooled and purified CBC-incorporation PCR products onto the flow cell for NGS.

Another modification to the scTCRseq platform is its translation to human T cells. *This is currently being successfully performed in a side project.* It predominantly requires the identification of α and β TCR constant-region GSPs in the RT reaction as well as the nested primers in the RBC-incorporation PCR that are compatible with all barcode and Tag sequences. This translation would allow the discovery of TCRs from patient peripheral blood lymphocyte (PBL) or TIL samples.

4.6. Conclusions and future prospects

Existing scTCRseq methods suffer from cumbersome workflows, low throughput and high costs. The absence of efficient scTCR sequencing technologies has limited ACT progress. The scope of this thesis was the development of a high-throughput and easily applicable murine scTCRseq platform. This was addressed by developing a plate-based and NGS-compatible scTCRseq workflow.

After establishment, the platform was optimized to improve αβTCR detection rates and minimize platform noise, costs and workflow complexity. Platform performance validation and evaluation showed robust TCR gene amplification, demultiplexing, alignment and αβTCR detection rates with minimal noise. The scTCRseq platform was

supplemented with a panel of primers for selected T cell phenotyping genes to allow the detection of TCRs from T cells that encountered their cognate antigen.

Early sample pooling, minimal primer use and low PCR reaction numbers as well as low read depth facilitated TCR sequencing. TCR discovery was further eased by the use of common laboratory practices, the ability to clone TCRs of interest from existing cDNA and the integrated end-to-end data analysis pipeline.

By sequencing TCRs of single T cells obtained from vaccinated A2DR1 and C57BL/6 mice, we compiled a library of shared TA and neoepitope-specific TCR candidates. Whilst the A2DR1 mice-derived TCRs can be used for clinical testing, the C57BL/6 mice-derived TCRs provide a valuable opportunity for pre-clinical ACT studies.

We show that the discovery of TA-specific TCRs can be facilitated using sequencing information indicative of clonal expansion, high read count levels and CDR3 sequence similarity. Furthermore, the developed platform provides the ability to study described TCR phenomena such as convergent recombination and the occurrence of public TCRs.

In summary, a high-throughput, easily applicable and robust scTCRseq platform was developed and exploited as a powerful tool for the rapid discovery of TA-specific murine TCRs.

5. References

1. Chaplin, D. D. Overview of the Immune Response. *The Journal of allergy and clinical immunology* 125, S3-23; 10.1016/j.jaci.2009.12.980 (2010).
2. Charles A Janeway, JR, Travers, P., Walport, M. & Shlomchik, M. J. *Immunobiology* (Garland Science, 2001).
3. Bonilla, F. A. & Oettgen, H. C. Adaptive immunity. *The Journal of allergy and clinical immunology* 125, S33-40; 10.1016/j.jaci.2009.09.017 (2010).
4. Takaba, H. & Takayanagi, H. The Mechanisms of T Cell Selection in the Thymus. *Trends in Immunology* 38, 805–816; 10.1016/j.it.2017.07.010 (2017).
5. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *nri* 14, 377–391; 10.1038/nri3667 (2014).
6. Owen, D. L. *et al.* Thymic regulatory T cells arise via two distinct developmental programs. *Nat Immunol* 20, 195–205; 10.1038/s41590-018-0289-6 (2019).
7. Nguyen, V., Mendelsohn, A. & Larrick, J. W. Interleukin-7 and Immunosenescence. *Journal of Immunology Research* 2017; 10.1155/2017/4807853 (2017).
8. Holder, A., Mella, S., Palmer, D. B., Aspinall, R. & Catchpole, B. An Age-Associated Decline in Thymic Output Differs in Dog Breeds According to Their Longevity. *PLOS ONE* 11, e0165968; 10.1371/journal.pone.0165968 (2016).
9. Surh. Homeostasis of Naive and Memory T Cells. *Immunity* 29, 848–862; 10.1016/j.immuni.2008.11.002 (2008).
10. Germain, R. N. T-cell development and the CD4–CD8 lineage decision. *Nat Rev Immunol* 2, 309–322; 10.1038/nri798 (2002).
11. Schwartz, R. H. T cell anergy. *Annual review of immunology* 21, 305–334; 10.1146/annurev.immunol.21.120601.141110 (2003).
12. Pennock, N. D. *et al.* T cell responses: naïve to memory and everything in between. *Advances in Physiology Education* 37, 273–283; 10.1152/advan.00066.2013 (2013).
13. Wan, Y. Y. Regulatory T cells: immune suppression and beyond. *Cellular and Molecular Immunology* 7, 204–210; 10.1038/cmi.2010.20 (2010).
14. Martínez-Lostao, L., Anel, A. & Pardo, J. How Do Cytotoxic Lymphocytes Kill Cancer Cells? *Clin Cancer Res* 21, 5047–5056; 10.1158/1078-0432.CCR-15-0685 (2015).
15. Mescher, M. F. *et al.* Signals required for programming effector and memory development by CD8+ T cells. *Immunological reviews* 211, 81–92; 10.1111/j.0105-2896.2006.00382.x (2006).
16. Parish, I. A. & Kaech, S. M. Diversity in CD8(+) T cell differentiation. *Current opinion in immunology* 21, 291–297; 10.1016/j.coi.2009.05.008 (2009).
17. Dong, D. *et al.* Structural basis of assembly of the human T cell receptor–CD3 complex. *Nature* 573, 546–552; 10.1038/s41586-019-1537-0 (2019).
18. Glusman. Comparative Genomics of the Human and Mouse T Cell Receptor Loci. *Immunity* 15, 337–349; 10.1016/S1074-7613(01)00200-X (2001).
19. Toor, A. A., Toor, A. A. & Manjili, M. H. *On The Organization Of Human T Cell Receptor Loci* (2015).
20. Lefranc, M. P. *et al.* IMGT, the international ImMunoGeneTics database. *Nucleic acids research* 27, 209–212; 10.1093/nar/27.1.209 (1999).
21. Krangel, M. S. Mechanics of T cell receptor gene rearrangement. *Current opinion in immunology* 21, 133–139; 10.1016/j.coi.2009.03.009 (2009).
22. Joachims, M. L., Chain, J. L., Hooker, S. W., Knott-Craig, C. J. & Thompson, L. F. Human αβ and γδ Thymocyte Development: TCR Gene Rearrangements, Intracellular TCRβ Expression and γδ Developmental Potential – Differences between Men and Mice1,2. *Journal of immunology (Baltimore, Md. : 1950)* 176, 1543–1552 (2006).
23. Sadofsky, M. J. The RAG proteins in V(D)J recombination: more than just a nuclease. *Nucleic acids research* 29, 1399–1409 (2001).
24. Parkinson, N. J. *et al.* Violation of the 12/23 rule of genomic V(D)J recombination is common in lymphocytes. *Genome Research* 25, 226–234; 10.1101/gr.179770.114 (2015).
25. Brady, B. L., Steinel, N. C. & Bassing, C. H. Antigen Receptor Allelic Exclusion: An Update and Reappraisal. *Journal of immunology (Baltimore, Md. : 1950)* 185, 3801–3808; 10.4049/jimmunol.1001158 (2010).

26. Gascoigne, N. R. & Alam, S. M. Allelic exclusion of the T cell receptor alpha-chain: developmental regulation of a post-translational event. *Seminars in immunology* 11, 337–347; 10.1006/smim.1999.0190 (1999).

27. Rybakin, V. *et al.* Allelic Exclusion of TCR α-Chains upon Severe Restriction of Vα Repertoire. *PLOS ONE* 9, e114320; 10.1371/journal.pone.0114320 (2014).

28. Honey, K. Understanding why T-cell receptors remain single. *Nat Rev Immunol* 5, 95; 10.1038/nri1560 (2005).

29. Levin-Klein, R. & Bergman, Y. Epigenetic Regulation of Monoallelic Rearrangement (Allelic Exclusion) of Antigen Receptor Genes. *Frontiers in Immunology* 5; 10.3389/fimmu.2014.00625 (2014).

30. Hale, J. S. & Fink, P. J. T-cell receptor revision: friend or foe? *Immunology* 129, 467–473; 10.1111/j.1365-2567.2010.03250.x (2010).

31. Venturi, V., Price, D. A., Douek, D. C. & Davenport, M. P. The molecular basis for public T-cell responses? *Nature reviews. Immunology* 8, 231–238; 10.1038/nri2260 (2008).

32. Casrouge, A. *et al.* Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. *Journal of immunology (Baltimore, Md. : 1950)* 164, 5782–5787; 10.4049/jimmunol.164.11.5782 (2000).

33. Arstila, T. P. *et al.* A direct estimate of the human alphabeta T cell receptor diversity. *Science (New York, N.Y.)* 286, 958–961; 10.1126/science.286.5441.958 (1999).

34. Li, H., Ye, C., Ji, G. & Han, J. Determinants of public T cell responses. *Cell Res* 22, 33–42; 10.1038/cr.2012.1 (2012).

35. Simone, M. de, Rossetti, G. & Pagani, M. Single Cell T Cell Receptor Sequencing: Techniques and Future Challenges. *Frontiers in Immunology* 9; 10.3389/fimmu.2018.01638 (2018).

36. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. & Rammensee, H. G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 351, 290–296; 10.1038/351290a0 (1991).

37. Rammensee, H. G., Friede, T. & Stevanoviic, S. MHC ligands and peptide motifs: first listing. *Immunogenetics* 41, 178–228; 10.1007/bf00172063 (1995).

38. Hedrick, S. M., Cohen, D. I., Nielsen, E. A. & Davis, M. M. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* 308, 149–153; 10.1038/308149a0 (1984).

39. Yanagi, Y. *et al.* A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* 308, 145–149; 10.1038/308145a0 (1984).

40. Clambey, E. T., Davenport, B., Kappler, J. W., Marrack, P. & Homann, D. Molecules in medicine mini review: the αβ T cell receptor. *Journal of molecular medicine (Berlin, Germany)* 92, 735–741; 10.1007/s00109-014-1145-2 (2014).

41. Wong, W. K., Leem, J. & Deane, C. M. Comparative analysis of the CDR loops of antigen receptors. *bioRxiv,* 709840; 10.1101/709840 (2019).

42. Birnbaum. Deconstructing the Peptide-MHC Specificity of T Cell Recognition. *Cell* 157, 1073–1087; 10.1016/j.cell.2014.03.047 (2014).

43. Kumar. Increased Sensitivity of Antigen-Experienced T Cells through the Enrichment of Oligomeric T Cell Receptor Complexes. *Immunity* 35, 375–387; 10.1016/j.immuni.2011.08.010 (2011).

44. Dustin, M. L. The immunological synapse. *Cancer immunology research* 2, 1023–1033; 10.1158/2326-6066.CIR-14-0161 (2014).

45. Paul, S. & Schaefer, B. C. A new look at TCR signaling to NF-κB. *Trends in immunology* 34, 269–281; 10.1016/j.it.2013.02.002 (2013).

46. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 68, 394–424; 10.3322/caac.21492 (2018).

47. Ferlay. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *European Journal of Cancer* 103, 356–387; 10.1016/j.ejca.2018.07.005 (2018).

48. Golemis, E. A. *et al.* Molecular mechanisms of the preventable causes of cancer in the United States. *Genes & Development* 32, 868–902; 10.1101/gad.314849.118 (2018).

49. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* 144, 646–674; 10.1016/j.cell.2011.02.013 (2011).

50. Deininger, P. Genetic Instability in Cancer: Caretaker and Gatekeeper Genes. *The Ochsner Journal* 1, 206–209 (1999).

51. Tohme, S., Simmons, R. L. & Tsung, A. Surgery for Cancer: A Trigger for Metastases. *Cancer research* 77, 1548–1552; 10.1158/0008-5472.CAN-16-1536 (2017).

52. Housman, G. *et al.* Drug Resistance in Cancer: An Overview. *Cancers* 6, 1769–1792; 10.3390/cancers6031769 (2014).

53. Arora, A. & Scholar, E. M. Role of tyrosine kinase inhibitors in cancer therapy. *The Journal of pharmacology and experimental therapeutics* 315, 971–979; 10.1124/jpet.105.084145 (2005).

54. Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: implications for targeted therapeutics. *British Journal of Cancer* 108, 479–485; 10.1038/bjc.2012.581 (2013).

55. Ehrlich Paul. Ueber den jetzigen Stand der Karzinomforschung (1909).

56. Gross, L. Intradermal Immunization of C3H Mice against a Sarcoma That Originated in an Animal of the Same Line. *Cancer Res* 3, 326–333 (1943).

57. BURNET, M. Cancer: a biological approach. III. Viruses associated with neoplastic conditions. IV. Practical applications. *British medical journal* 1, 841–847; 10.1136/bmj.1.5023.841 (1957).

58. Vigneron, N., Stroobant, V., Van den Eynde, Benoît J. & van der Bruggen, P. Database of T cell-defined human tumor antigens: the 2013 update. *Cancer Immunity* 13 (2013).

59. Cox, A. L. *et al.* Identification of a peptide recognized by five melanoma-specific human cytotoxic T cell lines. *Science (New York, N.Y.)* 264, 716–719; 10.1126/science.7513441 (1994).

60. Boon, T., Gajewski, T. F. & Coulie, P. G. From defined human tumor antigens to effective immunization? *Immunology today* 16, 334–336 (1995).

61. Gameiro, S. F. *et al.* Analysis of Class I Major Histocompatibility Complex Gene Transcription in Human Tumors Caused by Human Papillomavirus Infection. *Viruses* 9; 10.3390/v9090252 (2017).

62. Smith, C. C. *et al.* Alternative tumour-specific antigens. *Nat Rev Cancer* 19, 465–478; 10.1038/s41568-019-0162-4 (2019).

63. Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Available at https://www.nature.com/articles/nature23003 (2017).

64. Paret, C. *et al.* CXorf61 is a target for T cell based immunotherapy of triple-negative breast cancer. *Oncotarget* 6, 25356–25367 (2015).

65. Scanlan, M. J., Gure, A. O., Jungbluth, A. A., Old, L. J. & Chen, Y.-T. Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. *Immunological reviews* 188, 22–32; 10.1034/j.1600-065x.2002.18803.x (2002).

66. Even-Desrumeaux, K., Baty, D. & Chames, P. State of the Art in Tumor Antigen and Biomarker Discovery. *Cancers* 3, 2554–2596; 10.3390/cancers3022554 (2011).

67. van der Bruggen, P. *et al.* A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science (New York, N.Y.)* 254, 1643–1647; 10.1126/science.1840703 (1991).

68. Durgeau, A., Virk, Y., Corgnac, S. & Mami-Chouaib, F. Recent Advances in Targeting CD8 T-Cell Immunity for More Effective Cancer Immunotherapy. *Frontiers in Immunology* 9, 14; 10.3389/fimmu.2018.00014 (2018).

69. Hung, K. *et al.* The Central Role of CD4+ T Cells in the Antitumor Immune Response. *The Journal of Experimental Medicine* 188, 2357–2368 (1998).

70. Takeuchi, A. & Saito, T. CD4 CTL, a Cytotoxic Subset of CD4+ T Cells, Their Differentiation and Function. *Frontiers in Immunology* 8; 10.3389/fimmu.2017.00194 (2017).

71. Calmeiro, J. *et al.* *Highlighting the Role of DC-NK Cell Interplay in Immunobiology and Immunotherapy* (IntechOpen, 2018).

72. Mittal, D., Gubin, M. M., Schreiber, R. D. & Smyth, M. J. New insights into cancer immunoediting and its three component phases — elimination, equilibrium and escape. *Current opinion in immunology* 27, 16–25; 10.1016/j.coi.2014.01.004 (2014).

73. Emens, L. A. *et al.* Cancer immunotherapy: Opportunities and challenges in the rapidly evolving clinical landscape. *European journal of cancer (Oxford, England : 1990)* 81, 116–129; 10.1016/j.ejca.2017.01.035 (2017).

74. Coley, W. B. The treatment of malignant tumors by repeated inoculations of erysipelas. With a report of ten original cases. 1893. *Clinical orthopaedics and related research,* 3–11 (1991).

75. Rosenberg, S. A. IL-2: The First Effective Immunotherapy for Human Cancer. *Journal of immunology (Baltimore, Md. : 1950)* 192, 5451–5458; 10.4049/jimmunol.1490019 (2014).

76. Cancer immunotherapy: The interferon-α experience. *Seminars in Oncology* 29, 18–26; 10.1053/sonc.2002.33078 (2002).

77. Maloney, D. G. *et al.* IDEC-C2B8 (Rituximab) Anti-CD20 Monoclonal Antibody Therapy in Patients With Relapsed Low-Grade Non-Hodgkin's Lymphoma. *Blood* 90, 2188–2195; 10.1182/blood.V90.6.2188.2188_2188_2195 (1997).

78. Reichert, J. & Pavlou, A. Monoclonal antibodies market. Available at https://www.nature.com/articles/nrd1386.

79. Bengsch, B. & Thimme, R. Success of immune checkpoint blockade therapies - mechanisms and implications for hepatology. *Zeitschrift fur Gastroenterologie* 57, 74–86; 10.1055/a-0805-6936 (2019).

80. Massarelli, E. *et al.* Combining Immune Checkpoint Blockade and Tumor-Specific Vaccine for Patients With Incurable Human Papillomavirus 16-Related Cancer: A Phase 2 Clinical Trial. *JAMA oncology* 5, 67–73; 10.1001/jamaoncol.2018.4051 (2019).

81. Lopes, A. *et al.* Combination of immune checkpoint blockade with DNA cancer vaccine induces potent antitumor immunity against P815 mastocytoma. *Sci Rep* 8, 1–11; 10.1038/s41598-018-33933-7.

82. Lopes, A., Vandermeulen, G. & Préat, V. Cancer DNA vaccines: current preclinical and clinical developments and future perspectives. *J Exp Clin Cancer Res* 38, 1–24; 10.1186/s13046-019-1154-7 (2019).

83. Maeng, H. M. & Berzofsky, J. A. Strategies for developing and optimizing cancer vaccines. *F1000Research* 8; 10.12688/f1000research.18693.1 (2019).

84. Ott, P. A. *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 547, 217–221; 10.1038/nature22991 (2017).

85. Kreiter, S. *et al.* Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* 520, 692–696; 10.1038/nature14426 (2015).

86. Peng, M. *et al.* Neoantigen vaccine: an emerging tumor immunotherapy. *Molecular Cancer* 18; 10.1186/s12943-019-1055-6 (2019).

87. Conry, R. M., Westbrook, B., McKee, S. & Norwood, T. G. Talimogene laherparepvec: First in class oncolytic virotherapy. *Human Vaccines & Immunotherapeutics* 14, 839–846; 10.1080/21645515.2017.1412896 (2018).

88. Perica, K., Varela, J. C., Oelke, M. & Schneck, J. Adoptive T Cell Immunotherapy for Cancer. *Rambam Maimonides Medical Journal* 6; 10.5041/RMMJ.10179 (2015).

89. Rosenberg, S. A. *et al.* Treatment of Patients With Metastatic Melanoma With Autologous Tumor-Infiltrating Lymphocytes and Interleukin 2. *J Natl Cancer Inst* 86, 1159–1166; 10.1093/jnci/86.15.1159 (1994).

90. Salmikangas, P., Kinsella, N. & Chamberlain, P. Chimeric Antigen Receptor T-Cells (CAR T-Cells) for Cancer Immunotherapy – Moving Target for Industry? *Pharmaceutical Research* 35; 10.1007/s11095-018-2436-z (2018).

91. Hammerl, D., Rieder, D., Martens, J. W., Trajanoski, Z. & Debets, R. Adoptive T Cell Therapy: New Avenues Leading to Safe Targets and Powerful Allies. *Trends in Immunology* 39, 921–936; 10.1016/j.it.2018.09.004 (2018).

92. Gattinoni, L. Adoptive T cell transfer: imagining the next generation of cancer immunotherapies. *Seminars in immunology* 28, 1–2; 10.1016/j.smim.2016.03.019 (2016).

93. Timothy. Understanding TCR affinity, antigen specificity, and cross-reactivity to improve TCR gene-modified T cells for cancer immunotherapy. Available at https://www.springermedizin.de/understanding-tcr-affinity-antigen-specificity-and-cross-reactiv/17250666.

94. Schmitt, T. M., Ragnarsson, G. B. & Greenberg, P. D. T Cell Receptor Gene Therapy for Cancer. *Human Gene Therapy* 20, 1240–1248; 10.1089/hum.2009.146 (2009).

95. Morgan, R. A. *et al.* Cancer regression in patients after transfer of genetically engineered lymphocytes. *Science (New York, N.Y.)* 314, 126–129; 10.1126/science.1129003 (2006).

96. Abate-Daga, D. *et al.* Development of a T cell receptor targeting an HLA-A*0201 restricted epitope from the cancer-testis antigen SSX2 for adoptive immunotherapy of cancer. *PloS one* 9, e93321; 10.1371/journal.pone.0093321 (2014).

97. Johnson, L. A. *et al.* Gene therapy with human and mouse T-cell receptors mediates cancer regression and targets normal tissues expressing cognate antigen. *Blood* 114, 535–546; 10.1182/blood-2009-03-211714 (2009).

98. Robbins, P. F. *et al.* Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. *Journal of clinical oncology :*

official journal of the American Society of Clinical Oncology 29, 917–924; 10.1200/JCO.2010.32.2537 (2011).

99. Sharpe, M. & Mount, N. Genetically modified T cells in cancer therapy: opportunities and challenges. *Disease Models & Mechanisms* 8, 337–350; 10.1242/dmm.018036 (2015).

100. Tran, E. *et al.* Cancer Immunotherapy Based on Mutation-Specific CD4+ T Cells in a Patient with Epithelial Cancer. *Science (New York, N.Y.)* 344, 641–645; 10.1126/science.1251102 (2014).

101. Dudley, M. E. *et al.* Randomized Selection Design Trial Evaluating CD8+-Enriched Versus Unselected Tumor-Infiltrating Lymphocytes for Adoptive Cell Therapy for Patients With Melanoma. *Journal of Clinical Oncology* 31, 2152–2159; 10.1200/JCO.2012.46.6441 (2013).

102. Gros, A. *et al.* Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nature medicine* 22, 433–438; 10.1038/nm.4051 (2016).

103. Hinrichs, C. S. & Rosenberg, S. A. Exploiting the curative potential of adoptive T-cell therapy for cancer. *Immunological reviews* 257, 56–71; 10.1111/imr.12132 (2014).

104. Lu, Y.-C. *et al.* Efficient Identification of Mutated Cancer Antigens Recognized by T Cells Associated with Durable Tumor Regressions. *Clin Cancer Res* 20, 3401–3410; 10.1158/1078-0432.CCR-14-0433 (2014).

105. Tran, E. *et al.* T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *The New England journal of medicine* 375, 2255–2262; 10.1056/NEJMoa1609279 (2016).

106. Paillard, F., Sterkers, G. & Vaquero, C. Transcriptional and post-transcriptional regulation of TcR, CD4 and CD8 gene expression during activation of normal human T lymphocytes. *The EMBO Journal* 9, 1867–1872 (1990).

107. Zacharakis, N. *et al.* Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nature medicine* 24, 724–730; 10.1038/s41591-018-0040-8 (2018).

108. Yarchoan, M., Johnson, B. A., Lutz, E. R., Laheru, D. A. & Jaffee, E. M. Targeting neoantigens to augment antitumour immunity. *Nature reviews. Cancer* 17, 209–222; 10.1038/nrc.2016.154 (2017).

109. Klebanoff, C. A. *et al.* Memory T cell–driven differentiation of naive cells impairs adoptive immunotherapy. *The Journal of Clinical Investigation* 126, 318–334; 10.1172/JCI81217.

110. Omokoko, T., Simon, P., Türeci, Ö. & Sahin, U. Retrieval of functional TCRs from single antigen-specific T cells: Toward individualized TCR-engineered therapies. *Oncoimmunology* 4; 10.1080/2162402X.2015.1005523 (2015).

111. Li, B. *et al.* Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nature genetics* 48, 725–732; 10.1038/ng.3581 (2016).

112. Rosati, E. *et al.* Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnology* 17; 10.1186/s12896-017-0379-9 (2017).

113. van Beemd, R. d. *et al.* Flow cytometric analysis of the Vβ repertoire in healthy controls. *Cytometry* 40, 336–345; 10.1002/1097-0320(20000801)40:4<336::AID-CYTO9>3.0.CO;2-0 (2000).

114. Koga, M., Yuki, N., Tsukada, Y., Hirata, K. & Matsumoto, Y. CDR3 spectratyping analysis of the T cell receptor repertoire in Guillain-Barre and Fisher syndromes. *Journal of neuroimmunology* 141, 112–117; 10.1016/s0165-5728(03)00212-1 (2003).

115. Brown, S. D., Raeburn, L. A. & Holt, R. A. Profiling tissue-resident T cell repertoires by RNA sequencing. *Genome medicine* 7, 125; 10.1186/s13073-015-0248-x (2015).

116. Li, B. *et al.* Ultrasensitive detection of TCR hypervariable-region sequences in solid-tissue RNA-seq data. *Nature genetics* 49, 482–483; 10.1038/ng.3820 (2017).

117. Rosati, E. *et al.* Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnology* 17; 10.1186/s12896-017-0379-9 (2017).

118. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *nmeth* 12, 380–381; 10.1038/nmeth.3364 (2015).

119. Jiang. Ushering in Integrated T Cell Repertoire Profiling in Cancer. *Trends in Cancer* 5, 85–94; 10.1016/j.trecan.2018.11.005 (2019).

120. Cui, J.-H. *et al.* TCR Repertoire as a Novel Indicator for Immune Monitoring and Prognosis Assessment of Patients With Cervical Cancer. *Frontiers in Immunology* 9; 10.3389/fimmu.2018.02729 (2018).

121.   Weng, W.-K. *et al.* Minimal Residual Disease Monitoring with High-Throughput Sequencing of T Cell Receptors in Cutaneous T Cell Lymphoma. *Science Translational Medicine* 5, 214ra171-214ra171; 10.1126/scitranslmed.3007420 (2013).

122.   Jia, Q. *et al.* Diversity index of mucosal resident T lymphocyte repertoire predicts clinical prognosis in gastric cancer. Available at https://www.tandfonline.com/doi/full/10.1080/2162402X.2014.1001230 (2015).

123.   Reuben, A. *et al.* TCR Repertoire Intratumor Heterogeneity in Localized Lung Adenocarcinomas: An Association with Predicted Neoantigen Heterogeneity and Postsurgical Recurrence. *Cancer Discov* 7, 1088–1097; 10.1158/2159-8290.CD-17-0256 (2017).

124.   Brown, P. D. & Patel, P. R. Nanomedicine: a pharma perspective. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* 7, 125–130; 10.1002/wnan.1288 (2015).

125.   Feng, L. *et al.* Heterogeneity of tumor-infiltrating lymphocytes ascribed to local immune status rather than neoantigens by multi-omics analysis of glioblastoma multiforme. *Sci Rep* 7, 1–10; 10.1038/s41598-017-05538-z.

126.   Howie, B. *et al.* High-throughput pairing of T cell receptor alpha and beta sequences. *Science Translational Medicine* 7, 301ra131; 10.1126/scitranslmed.aac5624 (2015).

127.   Simon, P. *et al.* Functional TCR Retrieval from Single Antigen-Specific Human T Cells Reveals Multiple Novel Epitopes. *Cancer Immunol Res* 2, 1230–1244; 10.1158/2326-6066.CIR-14-0108 (2014).

128.   Han, A., Glanville, J., Hansmann, L. & Davis, M. M. Linking T-cell receptor sequence to functional phenotype at the single cell level. *Nature biotechnology* 32, 684–692; 10.1038/nbt.2938 (2014).

129.   Pajot, A. *et al.* A mouse model of human adaptive immune functions: HLA-A2.1-/HLA-DR1-transgenic H-2 class I-/class II-knockout mice. *European Journal of Immunology* 34, 3060–3069; 10.1002/eji.200425463 (2004).

130.   Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *nmeth* 10, 1096–1098; 10.1038/nmeth.2639 (2013).

131.   Kranz, L. M. *et al.* Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature* 534, 396–401; 10.1038/nature18300 (2016).

132.   Svec, D. *et al.* Direct Cell Lysis for Single cell Gene Expression Profiling. *Frontiers in Oncology* 3; 10.3389/fonc.2013.00274 (2013).

133.   Wu, L. *et al.* Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiol* 15, 1–12; 10.1186/s12866-015-0450-4 (2015).

134.   van Hall, T. & van der Burg, Sjoerd H. Mechanisms of peptide vaccination in mouse models: tolerance, immunity, and hyperreactivity. *Advances in immunology* 114, 51–76; 10.1016/B978-0-12-396548-6.00003-2 (2012).

135.   Koh. A Practical Approach to Immunotherapy of Hepatocellular Carcinoma Using T Cells Redirected Against Hepatitis B Virus. *Molecular Therapy - Nucleic Acids* 2, e114; 10.1038/mtna.2013.43 (2013).

136.   Futawatari, N. *et al.* Early gastric cancer frequently has high expression of KK-LC-1, a cancer-testis antigen. *World journal of gastroenterology* 23, 8200–8206; 10.3748/wjg.v23.i46.8200 (2017).

137.   Glanville, J. *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature* 547, 94–98; 10.1038/nature22976 (2017).

138.   Yadav, M. *et al.* Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* 515, 572–576; 10.1038/nature14001 (2014).

139.   Yadav, M. *et al.* Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* 515, 572–576; 10.1038/nature14001 (2014).

140.   Dawicki, W. & Watts, T. H. Expression and function of 4-1BB during CD4 versus CD8 T cell responses in vivo. *European Journal of Immunology* 34, 743–751; 10.1002/eji.200324278 (2004).

141.   Quezada, S. A., Jarvinen, L. Z., Lind, E. F. & Noelle, R. J. CD40/CD154 interactions at the interface of tolerance and immunity. *Annual review of immunology* 22, 307–328; 10.1146/annurev.immunol.22.012703.104533 (2004).

142.   Watts, T. H. TNF/TNFR family members in costimulation of T cell responses. *Annual review of immunology* 23, 23–68; 10.1146/annurev.immunol.23.021704.115839 (2005).

143.   Stubbington, M. J. T. *et al.* T cell fate and clonality inference from single cell transcriptomes. *nmeth* 13, 329–332; 10.1038/nmeth.3800 (2016).

144.   McDaniel, J. R., DeKosky, B. J., Tanno, H., Ellington, A. D. & Georgiou, G. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nature protocols* 11, 429–442; 10.1038/nprot.2016.024 (2016).

145.   Zemmour, D. *et al.* Single cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat Immunol* 19, 291–301; 10.1038/s41590-018-0051-0 (2018).

146.   Ludwig, J., Huber, A., Bartsch, I., Busse, C. E. & Wardemann, H. High-throughput single-cell sequencing of paired TCRα and TCRβ genes for the direct expression-cloning and functional analysis of murine T-cell receptors. *European Journal of Immunology* 49, 1269–1277; 10.1002/eji.201848030 (2019).

147.   Costello, M. *et al.* Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics* 19, 1–10; 10.1186/s12864-018-4703-0 (2018).

148.   Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* 172, 1091-1107.e17; 10.1016/j.cell.2018.02.001 (2018).

149.   Koh. A Practical Approach to Immunotherapy of Hepatocellular Carcinoma Using T Cells Redirected Against Hepatitis B Virus. *Molecular Therapy - Nucleic Acids* 2, e114; 10.1038/mtna.2013.43 (2013).

150.   Chinnasamy, N. *et al.* A TCR targeting the HLA-A*0201-restricted epitope of MAGE-A3 recognizes multiple epitopes of the MAGE-A antigen superfamily in several types of cancer. *Journal of immunology (Baltimore, Md. : 1950)* 186, 685–696; 10.4049/jimmunol.1001775 (2011).

151.   Cohen, C. J. *et al.* Recognition of fresh human tumor by human peripheral blood lymphocytes transduced with a bicistronic retroviral vector encoding a murine anti-p53 TCR. *Journal of immunology (Baltimore, Md. : 1950)* 175, 5799–5808; 10.4049/jimmunol.175.9.5799 (2005).

152.   Parkhurst, M. R. *et al.* Characterization of genetically modified T-cell receptors that recognize the CEA:691-699 peptide in the context of HLA-A2.1 on human colorectal cancer cells. *Clinical cancer research : an official journal of the American Association for Cancer Research* 15, 169–180; 10.1158/1078-0432.CCR-08-1638 (2009).

153.   Stevanovic, S. *et al.* Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. *Science (New York, N.Y.)* 356, 200–205; 10.1126/science.aak9510 (2017).

154.   Goldrath, A. W. & Bevan, M. J. Selecting and maintaining a diverse T-cell repertoire. *Nature* 402, 255–262; 10.1038/46218 (1999).

155.   Izraelson, M. *et al.* Comparative analysis of murine T-cell receptor repertoires. *Immunology* 153, 133–144; 10.1111/imm.12857 (2018).

156.   Li, H. *et al.* Recombinatorial Biases and Convergent Recombination Determine Interindividual TCRβ Sharing in Murine Thymocytes. *The Journal of Immunology* 189, 2404–2413; 10.4049/jimmunol.1102087 (2012).

157.   Venturi, V. *et al.* Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *PNAS* 103, 18691–18696; 10.1073/pnas.0608907103 (2006).

158.   Quigley, M. F. *et al.* Convergent recombination shapes the clonotypic landscape of the naïve T-cell repertoire. *PNAS* 107, 19414–19419; 10.1073/pnas.1010586107 (2010).

159.   Croft, M., So, T., Duan, W. & Soroosh, P. The Significance of OX40 and OX40L to T cell Biology and Immune Disease. *Immunological reviews* 229, 173–191; 10.1111/j.1600-065X.2009.00766.x (2009).

160.   Gramaglia, I., Weinberg, A. D., Lemon, M. & Croft, M. Ox-40 ligand: a potent costimulatory molecule for sustaining primary CD4 T cell responses. *Journal of immunology (Baltimore, Md. : 1950)* 161, 6510–6517 (1998).

161.   Taraban, V. Y. *et al.* Expression and costimulatory effects of the TNF receptor superfamily members CD134 (OX40) and CD137 (4-1BB), and their role in the generation of anti-tumor immune responses. *European Journal of Immunology* 32, 3617–3627; 10.1002/1521-4141(200212)32:12<3617::AID-IMMU3617>3.0.CO;2-M (2002).

162.   Eertwegh. In vivo CD40-gp39 interactions are essential for thymus-dependent humoral immunity. I. In vivo expression of CD40 ligand, cytokines, and antibody production delineates sites of cognate T-B cell interactions. *The Journal of Experimental Medicine* 178, 1555–1565 (1993).

6. Supplementary information

6.1. Primer sequences

Supplementary table 1| Primer sequences

Primer	Sequence
RT	
RT-TRAC	CCAGCACAGTTTTGTCAGTGATG
RT-TRBC	CACGTGGTCAGGGAAGAAGC
dT30Vn	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
TS	AAGCAGTGGTATCAACGCAGAGTACGCrGrGrG (r, ribonucleotide)
RBC-incorporation PCR	
130-RBCA-TS	CGATCCAGACTAGACGCTCAGGAAGTAAGCAAGCAGTGGTATCAACGCAGAGT
130-RBCB-TS	CGATCCAGACTAGACGCTCAGGAAGTGCACAAGCAGTGGTATCAACGCAGAGT
130-RBCC-TS	CGATCCAGACTAGACGCTCAGGAAGCTCAGAAGCAGTGGTATCAACGCAGAGT
130-RBCD-TS	CGATCCAGACTAGACGCTCAGGAAGGGAATAAGCAGTGGTATCAACGCAGAGT
130-RBCE-TS	CGATCCAGACTAGACGCTCAGGAAGCGAGGAAGCAGTGGTATCAACGCAGAGT
130-RBCF-TS	CGATCCAGACTAGACGCTCAGGAAGAGGAGAAGCAGTGGTATCAACGCAGAGT
130-RBCG-TS	CGATCCAGACTAGACGCTCAGGAAGTGTTGAAGCAGTGGTATCAACGCAGAGT
130-RBCH-TS	CGATCCAGACTAGACGCTCAGGAAGCAACTAAGCAGTGGTATCAACGCAGAGT
146-TRACv1	CAATATGTGACCGCCGAGTCCCAGGGTACACAGCAGGTTCTGGGTTCTGG
146-TRBCv2	CAATATGTGACCGCCGAGTCCCAGGAGACCTTGGGTGGAGTCACATTTCTCAGATCC
146-CD8	CAATATGTGACCGCCGAGTCCCAGGCACAGGTGAGGGAGTTCGCAGC
146-CD4	CAATATGTGACCGCCGAGTCCCAGGGCTGGTACCCGGACTGAAGGTCACTTTG
146-CD3	CAATATGTGACCGCCGAGTCCCAGGCTGCTACCAGCACCGGTTCCTCG
146-41BB	CAATATGTGACCGCCGAGTCCCAGGCCAAGCTACAGGTTTTGCAACCCTG
146-OX40	CAATATGTGACCGCCGAGTCCCAGGCTGTCCTCACAGACTGCGTCCAAGC
146-IFNG	CAATATGTGACCGCCGAGTCCCAGGCCTGTGGGTTGTTGACCTCAAAC
146-CD40L	CAATATGTGACCGCCGAGTCCCAGGCCTTTCTTGGCCCACTGTAGAACGG
146-TNFa	CAATATGTGACCGCCGAGTCCCAGGGCTCCTCCACTTGGTGGTTTGTGAG
CBC-incorporation PCR	
N4-130	NNNNCGATCCAGACTAGACGCTCAGGAAG
N5-130	NNNNNCGATCCAGACTAGACGCTCAGGAAG
CBC1-146	GTTCACAATATGTGACCGCCGAGTCCCAGG
CBC2-146	CAGGACAATATGTGACCGCCGAGTCCCAGG
CBC3-146	TTATACAATATGTGACCGCCGAGTCCCAGG
CBC4-146	CCTGTCAATATGTGACCGCCGAGTCCCAGG
CBC5-146	ACCGCCAATATGTGACCGCCGAGTCCCAGG
CBC6-146	ACTTACAATATGTGACCGCCGAGTCCCAGG
CBC7-146	GCTAGCAATATGTGACCGCCGAGTCCCAGG
CBC8-146	GACGTCAATATGTGACCGCCGAGTCCCAGG
CBC9-146	GGCTACAATATGTGACCGCCGAGTCCCAGG
CBC10-146	GAATGCAATATGTGACCGCCGAGTCCCAGG
CBC11-146	CCAACCAATATGTGACCGCCGAGTCCCAGG
CBC12-146	GAGACCAATATGTGACCGCCGAGTCCCAGG
Library preparation	
All primers were used from the TruSeq Nano DNA library kit (Illumina)	

6.2. IVT RNA TCR library

Supplementary table 2| IVT RNA TCR library

TCR chain	V J C	TCR chain	V D J C
α1	V1 J27 C	β1	V1 D1 J1.6 C1
α3	V3.3 J50 C	β2	V2 D1 J1.5 C1
α4	V4D.3 J2 C	β3	V3 D2 J2.5 C2
α5	V5.1 J27 C	β4	V4_2 D1 J1.1_1 C1
α6	V6D.7 J45 C	β5	V5_5 D1 J1.3 C1
α7	V7.2_02 J34_02 C	β13	V13.1_2 D2 J2.5 C2
α8	V8D.2_2 J13 C	β14	V14 D2 J2.1 C2
α9	V9D.3_2 J33 C	β15	V15 J2.2 C2
α10	V10 J7 C	β19	V19 D2 J2.7 C2
α12	V12D.2 J22 C	β24	V24_3 D1 J1.4 C1
α13	V13N.2 J9_2C	β26	V26 D2 J2.5 C2
α14	V14.2 J4 C	β29	V29 D1 J2.7 C2

6.3. ScTCR (pheno) Seq protocol

Cell-capture plate preparation

1. Generate cell-capture master mix. *No carrier RNA used for scTCRphenoSeq.

Component	Amount
dNTP	1mM each
RNase inhibitor	2 U/μL
Carrier RNA*	5ng
Triton X-100	0.2%

2. Add 4 μL of the mastermix to the cell-capture plates.

Single Cell Sorting

3. After plate sort, cover using adhesive seal.
4. Centrifuge plate and store at -80 °C.

GSP SMART cDNA synthesis

5. Generate the RT-MMX1. *Only for scTCRPhenoSeq.

Reagent	Volume/well μL
TS (100μM)	0.12 (1.2 μM)
TRAC-RT (100μM)	0.265 (2.64 μM)
TRBC-RT (100μM)	0.1325 (1.32 μM)
+/- dT30Vn (100μM)*	0.12(1.2 μM)

6. Add 3 μL of mastermix to each well.
7. Place into the cycler for 5 minutes at 70 °C. Place the plate on ice for 2 min.

8. Add 3 µL of master mix 2 to each well. * Use Maxima RTase for scTCRphenoSeq.

Reagent	Volume/well (µL)
Revert-aid Buffer	2
Revert aid RTase*	1

9. Place the plate into the cycler for 90 min at 42 °C (Revert aid) or 50°C (Maxima), then 15 min at 70°C for both.

Exonuclease I treatment and product dilution

10. Add 0.5 µL Exo I to each well (5 U). Place the plate into a cycler for 30 min at 37 °C, then 20 min at 80 °C.
11. Add 20 µL of water to each well using a stepper.

RBC-incorporation PCR

12. Generate 8 master mixes in the pre-PCR lab template-free hood using the table below:

Reagent	Volume (µL)	Volume for 13 wells (µL)
RNase-free water	31	403
Pfu HS buffer	5	65
dNTPs 10mM	1	13
TRAC/TRBC mix 5µM each	1	13
(Pheno GSP mix)	(5)	(65)
RBC oligos A-H 10µM	1	13
Pfu HS	1	13

13. Add 40 µL of the correct row-barcoded master mix to the respective wells in the new PCR plates.
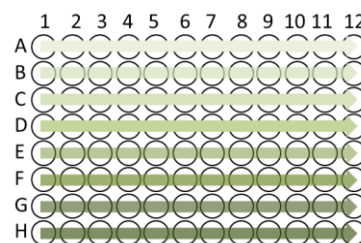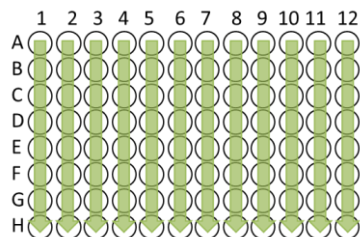

14. Transfer 10µL of the diluted 1st cDNA strand to the respective wells in the new PCR plate.

15. PCR program:

95°C	2 min	
94°C	30 sec	
60°C	30 sec	5 cycles
72°C	60 sec	
94°C	30 sec	
64°C	30 sec	5 cycles
72°C	60 sec	
94°C	30 sec	
72°C	120 sec	8 cycles
72°C	6 min	

109

16. Transfer the wells of each column into one 1.5 mL tube.

17. Carry out x0.75 (0.8) Ampure XP size-selection. Elute in 44uL.

18. Add 6μL of the Exo I master mix to each pool. Place the plate into a cycler for 30 min at 37 °C, then 20 min at 80°C.

	Exo I Buffer (µL)	Exo I (µL)
For one Sample	5	1

19. Carry out x0.75 (0.8) Ampure XP size-selection. Elute in 30μL.

PCR2, row pooling and bead purification

20. Generate the PCR 2 master mix:

Reagent	Volume (µL)	Volume for 1 plate (µL) (13 reactions)
RNase-free water	31	403
HS Pfu Phusion buffer	5	65
dNTPs 10mM	1	13
(5µM each) 5/4 N 5'oligo aliquots	1	13
HS Pfu Phusion	1	13

21. Add 39 μL of the PCR2 master mix to 12 0.5 mL PCR tubes. Add 1μL of the respective 10μM column-specific barcode oligo. Add 10μL of template to respective PCR tube.

22. PCR program:

95°C   1 min
94°C   20 sec
64°C   20 sec   24 cycles
72°C   30 sec
72°C   3 min

23. Pool the contents of all 12 PCR tubes together into a1.5 mL tube.

24. Perform bead purification and Exo I treatment as in steps 17-20. Elute in 44uL.

25. Quality control: Qubit and Tape station.

26. Library prep: using 40ng DNA.

27. Quality control: Qubit and Tape station.

Shaheer El Bardisy CV

Personal Information

Address: Am Winterhafen 16, 55131, Mainz, Germany
Mobile Number: +201224450450
Email: sbardissi@hotmail.com
Date of Birth: 12/05/1993
Nationality: German

Profile

I am an ambitious and target-oriented person inspired by the long-term aim of translating cutting-edge therapeutic approaches to developing countries, especially in the unmet medical fields of gene therapy and immuno-oncology. I persistently seek new challenges in fast-paced environments with high potential for professional and personal growth.

Education

"PhD in immuno-oncology" – BioNTech Cell and Gene therapies
(Jan 2015- Ongoing)

The main aim of my PhD project was to develop a cost-efficient and easy-to-perform single T cell receptor (TCR) sequencing platform to allow the rapid detection of shared antigen and neoepitope- specific TCRs for pre-clinical and proof-of-concept Adoptive T cell therapy studies.

"MSc Molecular Medicine" – University College London - <u>*Distinction award (76%)*</u>
(Sep 2013- Aug 2014)
- Introduction to Molecular Medicine (*Grade* 73%)
- Molecular Basis of Disease (70%)
- Data Interpretation (74%)
- Laboratory Practicals (73%)
- Immunological Basis of disease (67%)
- Cell and Gene therapy: Molecular and clinical aspects (72%)
- Data mining and analysis thesis *"Comparing senescence in T cells and fibroblasts"* (74%)
- Laboratory Research project "*Investigating the contribution of lipid metabolism in APCs to iNKT cell phenotypes in health and SLE*"; Centre of Rheumatology, UCL (*83%*)

"BSc Biomedical Sciences"- King's College London- <u>*First Class Honors degree*</u>
(September 2010-June 2013)

<u>Third year (2012-2013)</u>:
- Molecular Immunology (74%)
- Biology of Cancer (80%)
- Cell & Molecular Pharmacology (72%)
- Cardiovascular Pharmacology (68%)
  - Pharmacology laboratory research project: "*Platelet and smooth muscle cell interaction in asthma*" (78%)

Second year (2011-2012):
The immune system in Health & disease, Drugs & disease, Molecular Medicine, Drug Discovery & Development, Gene Cloning & Expression, Cell Biology.

First year (2010-2011):
From cells to systems, Biochemistry & Molecular Biology, Biochemistry & Molecular Biology, The Practice Of Biomedical Sciences, Fundamentals Of Pharmacology, Physiological Systems.

'International Baccalaureate'- **Modern English School Cairo** *(2008-2010) - <u>Total 34 points</u>*
(Marks out of 7, HL= Higher level, SL= Standard level): Biology- HL (6); Chemistry-HL (6); Business & management- HL (6); English A1-SL (5); French B- SL (6); Math & Statistics- HL (4); Bonus Points- (1)
'IGCSE's'- **Modern English School Cairo** *(2006-2008)*:
Physics (A*); Mathematics (A*); Biology (A); Chemistry (A); Business (A); Accounting (C); English (C); French (B).

Honors and Awards

2014- "John & Elizabeth Sherris Graduate prize for most outstanding student on the course and for achieving the highest overall grade"-MSc, Molecular medicine, UCL

2012 -"Oxford University Press Achievement in Biosciences Prize" –second Year, BSc, KCL

Publications

Rapid discovery of shared tumor antigen and neoepitope-specific T-cell receptors by high-throughput single-cell seq
Shaheer Bardissi, Tana A. Omokoko, Özlem Türeci, Ugur Sahin, 2019
Manuscript in preparation

Rapid discovery of tumor (neo) antigen-specific TCRs via high-throughput single-cell TCR sequencing.
CIMT 2019
Shaheer Bardissi, Rolf Hilker, Juliane Quinkhardt, Bettina Steege, Anke Oelbermann, Oliver Klein, Katharina Reinhard, Lisa Hebich, Claudia Tolliver, Mathias Vormehr, Ugur Sahin, Tana Omokoko

Using single cell TCRseq to study tumor antigen-specific T cell responses in mice.
CIMT 2017
Shaheer Bardissi, Rolf Hilker, Juliane Schwarz, Claudia Tolliver, Anke Oelbermann, Bettina Steege, Oliver Klein, Nele Brüne, Christian Stofft, Ugur Sahin, Tana Omokoko

Developing a High-throughput, Cost-Efficient and Easy-to-Perform Single Cell TCR NGS Platform: A Big Step Towards Individualized Adoptive T Cell Therapy.
CIMT 2016
Shaheer Bardissi, Tana Omokoko, Özlem Türeci, Ugur Sahin

Luciferase mRNA Transfection of Antigen Presenting Cells Permits Sensitive Nonradioactive Measurement of Cellular and Humoral Cytotoxicity
Tana A. Omokoko,Uli Luxemburger, Shaheer Bardissi, Petra Simon, Magdalena Utsch, Andrea Breitkreuz, Özlem Türeci and Ugur Sahin

Altered lipid metabolism contributes to defective iNKT-cell function in patients with systemic lupus erythematosus (SLE)
Smith EC, El Bardisy S, Torra IP, Isenberg DA, Jury EC

Skills and Attributes

Languages:

According to the European reference framework for languages:
English = proficient
German = independent C (Proficient)
French = basic A2 (elementary)
Arabic= independent B2 (intermediate)

Information Technology:

Microsoft Office-*(Excellent Knowledge)*
Bioinformatics- Ability to identify amino acid and nucleotide sequences, molecular properties and 3D structures of known proteins; acquired through tutorials and Courseworks. *(Good Knowledge)*
Linux based language *(Good knowledge)*
Flowjo-Flow cytometery data analysis *(Excellent Knowledge)*
GraphPad Prism- graphing, statistical analysis, and data organization- *(Excellent knowledge)*
VisualBasic- Programing Skills *(Good knowledge)*
Photoshop- Photo-retouching Skills *(Excellent Knowledge)*
TrueSpace 3D-Animation Skills *(Good knowledge)*

Research Methodologies and Lab Techniques:

Polymerase chain reactions
ELISpots
Next-generation sequencing
Extracellular and intracellular staining
Operating Flow cytometry and detailed results analysis
ELISA technique
Working under tissue culture and sterile area standards
PBMC and platelet isolation
Monocyte, B cell and T cell negative isolation
Dendritic cell culture setup and iNKT cell line development
Culturing monocytes, B cells, T cell subsets and Airway smooth muscle cells
Northern, Western and southern blotting
Measuring cellular Chemotaxis
DNA Cloning into Vectors/ Ligation reactions
PCR amplification reactions
DNA analysis- Agarose gel electrophoresis
Protein analysis- SDS-PAGE
Transformation of *E. coli* competent cells with plasmid DNA
Microbiology techniques – aseptic spreading of bacteria on agar plates
Restriction enzyme analysis of plasmid DNA
Induction of the expression of histidine tagged protein in *E. coli* using tetracycline
Radioligand Binding Assays and detailed results analysis
Histological slide preparation
Measuring Erythrocyte sedimentation rate
Using a coagulometer to measure Prothrombin time and Partial Thromboplastin time
Using a spectrophotometer to measure blood glucose levels
Using Chemiluminescence to carry out a complete blood picture
Using a microscope and counter to measure relative leukocyte percentage in blood sample

Transferable Skills:

Experiment planning throughout the 4 years of my doctoral studies

Scientific discussions and contributions through many journal clubs, data clubs, scientific seminars and conferences
Discipline and good time management: Ability to coordinately manage projects and meet deadlines on time.
Results oriented, self-motivated and determined: Constant improvements in grades throughout educational career.
Desire to learn, improve and hard worker: BSc, MSc and applying for PhD at a very young age.
Ability to work with group members: acquired through many group and lab projects throughout the past 4 years.
Written skills and communication: many essays and written projects of different lengths.
Oral communication: Achieved high grades in all presentations.
Literature research skills: Critically analyze and classify data as shown in
Critical thinking and convey research findings clearly and meaningfully: Article and literature analysis seminars.
Creativity, innovation and strategic thinking: through course works of different natures and experimental planning.
Co-ordination of scientific projects: Pharmacology experimental research project (BSc Thesis) and Laboratory Research project (MSc thesis).

Lab research experience

PhD doctoral thesis project "Rapid discovery of tumor (neo) antigen-specific TCRs via high-throughput single-cell TCR sequencing"

For patients lacking pre-existing anti-tumoral immunity, vaccination and adoptive T-Cell therapy (ACT) targeting somatic mutations represent highly promising treatment options. Despite advances in the former, progress of neoantigen-specific ACT is limited by the unavailability of neoantigen-specific T-Cell Receptor (TCR) sequences required for mode of action and proof of concept studies. To facilitate the discovery of high-affinity (neo)-antigen-specific TCRs, we developed a high-throughput single-cell TCR sequencing (scTCRseq) platform, integrating an easily applicable wet-lab process and data analysis pipeline. The platform was highly optimized to increase efficiency and throughput and minimize costs. We sequenced $\alpha\beta$TCR chains of IFNγ+ single splenocytes from mice immunized with Adpgk or Reps1 MC38 neoantigen encoding RNA vaccines. Two out of nine Adpgk and three out of eight Reps1 unique TCRs recognized the respective mutated, but not corresponding wild type epitope. Moreover, four TCRs specifically recognized the mutation-expressing MC38 cell line. (A forty-fold reduction in the RNA dose used for immunization led to the discovery of five additional unique TCRs with greater cell line-affinity). Clonal expansion, high read counts and CDR3 homology were indicative of TCR-antigen specificity. Additionally, ten high-affinity TCRs specific for commonly investigated tumor-restricted antigens (HPV E6, HPV E7, HBVc, HBVs and KKLC1) were identified using our technology.

MSc Laboratory Research project "*Investigating the contribution of lipid metabolism in APCs to iNKT cell phenotypes in health and SLE*": For my Master's laboratory project, I have worked with Dr. Liz Jury, a prominent figure in the field of Systemic Lupus Erythematosus, and her research group at the Centre of Rheumatology, UCL. In only a short period, I was able to reveal that monocytes and B-cells possess very different membrane lipid profiles, which explains their differential abilities to activate iNKT cells. I was also able to show that APC membrane lipid-profiles are defective in SLE patients and this results in the altered iNKT phenotypes seen in these patients. Moreover, the results revealed that the LXRβ nuclear receptor is the cause of this defect in SLE B-cells. This study further supported the role of lipid metabolism in SLE pathogenesis and exposed the role of APC lipid metabolism in regulating iNKT cell activity. Thus, this study contributed to understanding the interaction between APCs and iNKT cells in health and disease and shed light on targeting defective lipid metabolism as a new therapeutic modality in SLE.

BSc Pharmacology laboratory research project (2013): "*Platelet and smooth muscle cell interaction in asthma*" I worked in close collaboration with Dr. Varsha Kanabar in the renowned

Sackler Institute of Pulmonary Pharmacology (KCL) for four months. Platelets have been shown to participate in chronic airway remodeling in severe asthma. I was able to show that mast cell- derived heparin increases platelet chemotaxis *ex vivo* and that this is dependent on specific structural heparin moieties**.** Furthermore, I was able to demonstrate that unlike healthy airway smooth muscle cells (SMC), asthmatic SMCs have the ability to induce platelet chemotaxis. In turn, these platelets participate in enhancing the inflammatory milieu by inducing IL-8 secretion from healthy airway SMCs in a contact -dependent and -independent manner. Since Il-8 is a potent inflammatory mediator and platelet chemokine, I was able to establish an IL8-SMC-platelet positive feedback model, which contributes to the chronic phase in asthma.

I have also been involved in many independent practical courses and tutorials, which helped build my experimental knowledge basis.

Work experience

"PhD doctoral studies at BioNTech Cell and Gene therapies"– Whilst pursuing my PhD studies for five years at BioNTech, the corporate environment has taught me the importance of discipline, team work and strategic thinking. Oral communication and accurate documentation were also strong deliverables from this experience. The opportunity to work within a rapidly growing business provided me with a rare experience.

 "Cairo center for clinicopathological laboratories" – I worked in a Clinical pathology lab for 3 weeks in cooperation with other technicians and under the supervision of a professor of clinical pathology in the Faculty of Medicine Suez Canal University in Egypt. My responsibilities included carrying out hematological analysis such as: measuring Erythrocyte sedimentation rate, measuring Prothrombin time and Partial Thromboplastin time using a coagulometer and measuring blood glucose levels using a spectrophotometer. I also prepared complete blood pictures using Chemiluminescence. I was then able to participate in the interpretation of the results by comparing the values to the normal human ranges.

"Amplifon Middle East"- I worked in a branch of Amplifon, a multinational and world leader hearing service provider for one month. My responsibilities included the use of audiometers to assess hearing in patients with hearing impairment referred from ENT's and other clinicians. I gathered accurate recordings for up to 30 patients whilst demonstrating customer orientation following the quality system as implemented in this organization.