Immune cell tracking following hematopoietic cell and gene therapy

Natalia Izotova

Thesis submitted for the degree of

Doctor of Philosophy

University College London

Molecular and Cellular Immunology Section,

Infection, Immunity, and Inflammation Programme

Great Ormond Street Institute of Child Health

2022

Declaration

I, Natalia Izotova, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

In this project we combined standard cellular and molecular assays with a custom PCR-based technology based on high throughput sequencing to track genetically engineered cells in treated patients by means of viral integration sites (IS) analysis. We leveraged this analytical pipeline to 1) assess whether naïve T cells can still be produced for many years even in absence of any supply by multipotent progenitors in the bone marrow and 2) to investigate the origin of CAR-T cells that mediate anti-leukaemic responses or long-term immune surveillance. In a clinical trial X-linked Severe Combined Immunodeficiency patients received an infusion of autologous hematopoietic stem/progenitor cells corrected via retroviral vector encoding the interleukin-2 common cytokine receptor gamma chain. In these patients, many years after gene therapy only vector-positive T and NK cells persist while no other genetically engineered blood cell populations are detectable. By a comprehensive long-term immunophenotypic, molecular and functional characterization we demonstrated that the thymus is actively producing a new and diverse repertoire of vector-positive naïve T cells (TN). This suggests that, even if gene corrected HSC are absent, a de novo production of genetically engineered T cell is maintained by a population of gene-corrected long term lymphoid progenitors (Lt-LP). Moreover, tracking IS clonal markers overtime, we inferred that Lt-LP can support both T and NK cells production. In a separate clinical trial of CD19-CAR-T cells for the treatment of haematological malignancies, we used IS analysis to investigate the origin of short- and long-term circulating CAR-T cells that mediate early anti-leukaemic responses or long-term immune surveillance. We compared IS between the product and CAR-T cells at

early/late timepoints in vivo. This analysis suggested that T memory stem cells (TSCMs) contained in the infused cell product, contributed the most to the generation of CAR-T cell clones during the peak response phase as well as to the generation of long-term persisting CAR-T cells.

Impact Statement

Findings of this project are important for a better understanding of CAR T cell kinetics in patients and human immune cell development. In this project, we established integration site analysis technology at UCL and used it to unravel novel insights about human haematopoiesis in severe combined immunodeficiency gene therapy patients and CAR T cell clonal dynamics and persistence in CARPALL trial patients.

We here showed that protocols we used for integration site analysis in gene therapy patients are efficient, robust, and consistent at generating high quality data for not only assessing safety profiles post gene therapy but also for answering important biological questions about human haematopoiesis and immune cell clonal kinetics. These established and optimised integration site analysis protocols can now be used by our group and others to assess safety profiles post gene therapy and to better understand clonal kinetics of different immune cells directly in vivo in humans.

Here we showed that vector positive T and NK cells can be produced in humans for decades in the absence of gene corrected stem cells and that these cells might be originating from common transduced long-term lymphoid progenitors. Identification and exploitation of such human progenitor population may be of significant benefit in the development of next generation gene therapy and cancer immunotherapy approaches.

Our analysis represents one of the longest follow up studies of gene therapy patients to date (up to 18 years). We showed that in 5 patients analysed, all T and NK cells

remain vector positive and functional for many years post gene therapy and we also demonstrated that to date there is no clonal expansion of potentially oncogenic clones. This is an important finding, considering that few other patients on the same gene therapy trial (SCID-X1) developed leukaemia.

We here also showed that T memory stem cells are responsible for the long-lived CAR T cells in CARPALL patients. Based on our data, we hypothesise that assessing the frequency and polyclonality of T memory stem cells in patient products can help to predict early CAR T cell loss in some acute lymphoblastic leukaemia patients, allowing more time for finding bone marrow transplant for these patients. Next step would be to look at larger cohort of patients to verify these findings.

In the future, we are planning to use our integration site analysis technology to identify different progenitor cell types that comprise human thymus to obtain a better understanding of thymic microenvironment and what shapes T cell development.

Acknowledgements

I would like to thank my supervisors and everyone at GOS ICH and Camelia Botnar Labs who helped me with this project.

I am very grateful to my primary supervisor Adrian Thrasher for giving me this opportunity and all his support with this project and my thesis. Thank you for giving me this space to grow and develop.

I would also like to say a big thank you to Luca Biasco for all his mentoring and support. For helping me with my presentations, conference abstracts, data analysis, writing papers and this thesis. I really appreciate you always taking your time to answer my questions and to help and support me whenever I needed it, not only as a supervisor but also as a friend.

I would like to express my gratitude to Christine Rivat, my first supervisor at ICH. You were a great mentor to me. You inspired and encouraged me to start this PhD project. You stayed long hours with me to help me with experiments and even when you left ICH, you were still there to help and support me whenever I needed it.

I would also like to thank Kimberly Gilmour. You were always so kind and supportive. Thank you for always finding the time to help me with advice. Because of your kind support I was able to do both patient monitoring and my PhD at the same time, and actually enjoyed doing both.

Table of contents

List of tables

List of figures

List of Abbreviations

ACT   Adoptive T cell therapy

ADA   Adenosine deaminase

ALL   Acute Lymphoblastic Leukaemia

APC   Antigen Presenting Cell

ApoB   Apolipoprotein B

BM   Bone Marrow

CAR   Chimeric Antigen Receptors

CCR7   C-C chemokine receptor 7

cDNA   Complimentary DNA

CDR   Complementarity Determining Regions

         Cellular Indexing of Transcriptomes and Epitopes by Sequencing

CITESeq   analyses (

CLP   Common Lymphoid Progenitors

CMP   Common Myeloid Progenitors

CMV   Cytomegaloviruses

D   Diversity segment

ddPCR   digital PCR

DNA   Deoxyribonucleic Acid

ETP   Early T Progenitors

FACS   Fluoresence-activated cell sorting

FcγRIIIa   Fragment crystallisable (Fc) receptor

FcεRIγ-   Fc fragment of the high affinity receptor for IgE antibody

GMP   Granulocyte-Monocyte Progenitors

GOSH   Great Ormond Street Hospital

GR   GammaRetroviral

GT   Gene Therapy

HIV-1   Human Immunodeficiency type 1

HSC   Haematopoietic Stem Cell

HSCT   Hematopoietic Stem Cell Transplantation

IFNg	Interferon Gamma
IL	Interleukin
IL2RG	IL-2 receptor common gamma chain gene
ILC	Innate Lymphoid Cells
IS	Integration Site
ITAM	Intracellular Immunoreceptor Tyrosine-based Activation Motif
J	Joinin segment
LAM-PCR	Linear Amplification Mediated Polymerase Chain Reaction
LC	Linker cassette
Lin-	Lineage negative
LM-PCR	Linker mediated Polymerase Chain Reaction
LMO2	LIM domain only 2
LPMP	Lymphoid Primed Multipotent Progenitors
Lt-LP	Long - term Lymphoid Progenitors
ltHSC	Long-term Haematopoietic Stem Cell
LTR	Long Terminal Repeat
MEP	Megakaryocyte-Erythrocyte Progenitors
MFI	Mean Fluorescence Intensity
mFTOC	mouse Foetal Thymic Organ Cultures
MHC	Major Histocompatibility Complex
MLV	Murine Leukaemia Virus
MLP	Multi-lymphoid progenitors
MPP	Multipotent Progenitors
NGS	Next Generation Sequencing
NK	Natural Killer cell
NKT	Natural Killer T cell
nrLM-PCR	Non-restrictive Linker Mediated Polymerase Chain Reaction
NSG mice	Next Generation Sequencing
PBMC	Peripheral Blood Mononuclear cells
PCR	Polymerase Chain Reaction
PGK	human Phosphoglycerate Kinase

PID	Primary Immune Deficiencies
PMA	Phorbol 12-myristate 13-acetate
preB/NK	B cell and NK cell progenitors
qPCR	quantitative PCR
RB	Rainbow Beads
ScFv	Single Chain Variable Fragment
SCID	Severe Combined Immune Deficiency syndrome
SCID-X1	X-linked Severe Combined Immune Defeciency syndrome
SFFV	Spleen Focus Forming Virus
SIN	Self-Inactivating viral vectors
T-ALL	T Cell Acute Lymphoblastic Leukaemia
Tc	cytotoxic T cell
TCM	Central Memory T cell
TCR	T Cell Receptor
TEM	Effector Memory T cell
TEMRA	T cell Effector memory CD45RA+
Th	T helper cell
TIL	Tumour Infiltrating Lymphocytes
TN	Naïve T cell
TRECs	T cell Receptor Excision Circles
Treg	Regulatory T cell
TRM	Tissues resident memory T cells
TSCM	T Memory Stem Cells
UMI	Unique Molecular Identifier
V	Variable segmenet
VCN	Vector Copy Number
VISA	Vector Integration Site Analysis
WAS	Wiskott-Aldrich Syndrome Gene Therapy
WGA	Whole Genome Amplification
αβ	Alpha Beta
γδ	Gamma Delta

1 Introduction

1.1 Overview of the human immune system

The human immune system is designed to protect the host from damaging agents of both exogenous and endogenous nature, ensuring that all invading pathogens, toxic and/or allergenic substances are detected and removed from a body in a timely and efficient manner. Immune cells can recognize structural molecular features that are unique for these pathogens or toxins and are not present in the host.

The human immune system is conventionally subdivided into two distinct arms: adaptive and innate immune systems. The innate immune system consists of cells that are hard-wired to respond quickly to evading pathogens or toxins by recognizing molecular patterns that are shared between different foreign antigens but are absent in humans. These pattern recognition receptors are encoded in the germ-line genes of the host. The adaptive immune system consists of cells that can produce much slower but highly specific response to their antigens. These responses are encoded by gene elements that somatically rearrange to produce antigen-binding receptors specifically tailored to recognize these foreign antigens.

The human immune system comprises a vast array of various immune cell types that differ in their function, location preferences and self-renewal ability. Some of these cells belong to the innate part of the immune system (e.g., macrophages and dendritic cells), others belong to the adaptive immune system (e.g. T and B cells), whereas some represent a bridge between the two systems (e.g. NKT cells and γδ T cells).

One of the key features of cells that belong to the adaptive immune system is that after the initial response and clearance of the foreign antigen, they are able to produce long-lived memory cells that persist in the organism for decades and are able to initiate effector responses very quickly upon stimulation with the same antigen.

1.2 Haematopoiesis – the generation of immune system and blood.

Haematopoiesis is a lifelong process of production of new blood and immune cells in higher organisms like humans. An average human adult produces around one trillion blood cells a day in the bone marrow (BM) during steady state haematopoiesis and even larger cell numbers can be generated in response to bleeding or infection (Gordon, Lewis and Marley, 2002). Haematopoiesis is conventionally described as a hierarchal process in which hematopoietic stem cells (HSC) reside at the apex of the hierarchical differentiation tree and are responsible for the replenishment of the entire hematopoietic system (Till and McCulloch, 1980) (Spangrude *et al.*, 1991). There are various types of cells that constitute the immune system and blood and are classically divided into two main branches – lymphoid branch and myeloid branch. However, only HSC have a long-term self-renewal potential and have an ability to generate all the immune and blood lineages – lymphoid and myeloid/erythroid cells. Human hematopoietic stem cells that possess the highest degree of multilineage potential are phenotypically defined as CD34+CD38-Thy(CD90)+CD45RA- and represent a very rare population in the human BM with approximately only one in a million cells being a true HSC (Civin *et al.*, 1984) (Baum *et al.*, 1992) (Murray *et al.*, 1995) (Bhatia *et al.*, 1997) (Conneally *et al.*, 1997) (Lansdorp, Sutherland and Eaves, 1990) (Wang, Doedens and Dick, 1997). HSC have an ability to undergo three types of cell division. Symmetric self-renewal results in two identical HSC, symmetric differentiation

generates two haematopoietic progenitor cells, and asymmetric self-renewal results in one HSC and one haematopoietic progenitor cells (Yamamoto, Wilkinson and Nakauchi, 2020). Haematopoietic progenitor cells then will further differentiate into various progenitor intermediates that will eventually produce a mature immune or blood cell. The precise trajectory of the differentiation pathway from HSC to mature immune/blood cell is not yet clear and is a subject of intense research efforts.

The classical model of haematopoiesis was initially proposed based on murine studies and suggested that long-term HSC gradually differentiate into intermediate HSC, short-term HSC and subsequently multipotent progenitors (MPP) and eventually give rise to oligopotent and unipotent progenitors with much shorter lifespans and more lineage restriction (Yang *et al.*, 2005) (Doulatov *et al.*, 2012) (Benveniste *et al.*, 2010) (Notta *et al.*, 2011). It was thought that MPPs then either give rise to common lymphoid progenitors (CLP) that are able to generate all lymphoid cells but unable to differentiate into myeloid/erythroid cells, or to common myeloid progenitors (CMP) that can replenish all the myeloid/erythroid cells but unable to generate any lymphocytes (Akashi *et al.*, 2000) (Kondo, Weissman and Akashi, 1997). Oligopotent CMPs then further differentiate into more restricted granulocyte/monocyte progenitors that generate granulocytes and monocytes, and megakaryocyte/erythroid progenitors that generate platelets and red blood cells. Oligoclonal CLPs further differentiate into T, B and NK cell progenitors.

The classical model has been a useful tool for understanding the basics of haematopoiesis. However, some recent studies have challenged this model of hierarchical gradual hematopoietic stem cell differentiation in both human and mice. Indeed, in mice several studies now suggest that there might be an early specification

of some HSC that directly differentiate into megakaryocyte/erythroid progenitors without any intermediates, this being the first potential branching of the hematopoietic tree, with the rest of the cells later diverging into a distinct population of lymphoid primed multipotent progenitors (LPMP) that are capable of producing T cells, B cells, granulocytes and monocytes (Adolfsson *et al.*, 2005) (Månsson *et al.*, 2007). Interestingly, it has been demonstrated that the majority of LPMP cells are biased towards generating certain lineages with almost half of the compartment being specified towards producing dendritic cells and only a small fraction (around 3%) having a multi-lineage unbiased potential (Naik *et al.*, 2013). In addition to LPMP, two other MPP subsets were identified – MPP2 and MPP3 that in contrast, demonstrated a myeloid-biased potential (Wilson *et al.*, 2009) (Pietras *et al.*, 2015).

One of the disadvantages of early studies of haematopoiesis was the isolation of cells in bulk using cell surface markers assuming that all cells that have an identical phenotype (based on known markers) are also functionally identical. The emergence of single cell analytics, that allowed us to study individual cells and their differentiation into distinct subsets, further challenged the classical model of haematopoiesis. Based on a single cell analysis, one study has identified subsets of phenotypically true HSC, that still had a long-term repopulating ability, to be committed to the megakaryocyte, megakaryocyte/erythroid or common myeloid lineages, suggesting that loss of self-renewal ability and gradual differentiation through multiple progenitor stages is not essential for lineage commitment of HSC (Yamamoto *et al.*, 2013). This and other studies have also demonstrated that molecular programs that are associated with myeloid/erythroid/megakaryocyte lineages, can directly arise in MPP, suggesting again that at least myeloid differentiation might not be a gradual process and that it

could happen without the progression through the CMP stage (Månsson *et al.*, 2007) (Yamamoto *et al.*, 2013) (Pronk *et al.*, 2007).

In humans, several studies suggested that the first major lineage split might, like in mice, also occur when HSC directly generate megakaryocyte/erythroid progenitors and myelomonocytic lineage co-segregates together with lymphoid lineage (Doulatov *et al.*, 2010) (Notta *et al.*, 2016). It has also been shown that human multi-lymphoid progenitors that represent the earliest lymphoid differentiation precursors, phenotypically defined as CD34+CD38-Thy-1(CD90) neg-lowCD45RA+, can generate all lymphoid cell types as well as monocytes, macrophages and dendritic cells suggesting that myeloid potential is retained in multi-lymphoid progenitors (Doulatov *et al.*, 2010). Human CMP could be isolated from CD34+CD38+ progenitor cell population, with cells expressing CD123 (interleukin (IL) - 3 receptor alpha chain) and CD135 (FLT3) markers having myeloid but not erythroid potential (Manz *et al.*, 2002; Doulatov *et al.*, 2010). The differentiation of CMP into granulocyte/monocyte progenitors is accompanied by acquisition of progenitor cells of the CD45RA marker. CMP cells that were CD135+CD45RA- were able to reconstitute all myeloid lineages but had no lymphoid differentiation potential in vitro and in vivo after the transplant. This suggests that human CMP, in contrast to human MLP, might follow a predicted classical model of lineage differentiation.

Overall, haematopoiesis in both human and mice is a highly complex process with relatively high degree of plasticity. Some HSC and MPP seem to be biased towards production of a specific blood cell type despite being capable of producing multiple lineages, and some oligopotent/unipotent progenitors retain their multilineage potential despite their commitment to the specific branch/lineage. This plasticity could be an

adaptation mechanism so that in case of injury, disease or any other immunological challenge, various progenitor and HSC types can efficiently generate larger numbers of certain blood and immune cells to protect a host from damage.

1.3   T cells

As mentioned before, the immune system in higher organisms like humans consists of many different blood and immune cells with vast array of different functions and responsibilities. One of the major players of the adaptive arm of the immune system are T cells. T cells comprise a heterogeneous population of cells with distinct functional properties, various localisation preferences and different survival ability. T cells have a unique molecular feature called T cell receptor (TCR)- on their surface that they use to recognise their target antigens and generate target specific responses. As antigens get into the body they are processed and presented by various antigen-presenting cells (e.g., dendritic cells macrophages and B cells) via major histocompatibility complex (MHC) molecules and are recognised by TCR on the surface of T cells. Upon stimulation of their TCR with a cognate antigen, in cooperation with stimulation with various cytokines, chemokines, co-stimulatory molecules, integrins and different metabolites, T cells become activated and differentiate into effector cells that will eliminate this antigen from the system (Daniels and Teixeiro, 2015). The TCR consists of two different polypeptide chains, with the vast majority of human T cells carrying alpha beta (αβ) TCRs and a small proportion of human T cells carrying gamma delta (γδ) TCRs (Haas, Pereira and Tonegawa, 1993). The TCR plays an essential role in antigen recognition but does not have any signalling domains, hence all TCR signalling is done in cooperation with CD3 coreceptors, that contain intracellular signalling motifs (Birnbaum, Berry, *et al.*, 2014). The TCR is non-covalently associated with CD3

complex that is comprised of CD3γ, CD3δ, CD3ε and CD3ζ subunits (Meuer *et al.*, 1983; Wucherpfennig *et al.*, 2010). Both TCRα and TCRβ chains have extracellular domains that consist of variable and constant regions, a membrane-proximal connecting peptide (transmits signals from TCRαβ heterodimer to CD3ζ complex, responsible for antigen responsiveness), a transmembrane segment, and a cytoplasmic tail (Bäckström *et al.*, 1996; Garboczi *et al.*, 1996; Dong *et al.*, 2019). The CD3γ, CD3δ and CD3ε segments are organised in the similar way to TCR with extracellular immunoglobulin domain, a short connecting peptide (with conserved motif in CD3ε subunit being essential for TCR signalling and T cell development), a transmembrane segment and an intracellular immunoreceptor tyrosine-based activation motif (ITAM) (Dong *et al.*, 2019) (Wang *et al.*, 2009). The CD3ζ subunit, however, is organised in a slightly different way, with only short extracellular sequence, three ITAMs and a transmembrane segment (Love and Hayes, 2010). Interaction of extracellular and transmembrane domains of CD3 and TCR complexes is thought to be responsible for the transmission and propagation of a signal, that is generated in response to binding of TCRs to MHC-peptide complexes, to ITAM motifs in the cytoplasmic regions of CD3 subunits (Kane, Lin and Weiss, 2000; Gaud, Lesourne and Love, 2018). It has recently been demonstrated that apart from MHC-peptide recognition, TCRs on some T cells (NKT cells, γδ T cells) are capable of recognising lipid antigens presented by various isoforms of CD1 molecules (MHC class-I like molecules) and metabolic intermediates presented by MR1 molecules to mucosal associated invariant T cells in the gut (Mori and De Libero, 2012) (Reantragoon *et al.*, 2012).

The TCR diversity and, as a consequence, the ability of T cells to recognise a large number of different antigens is achieved by somatic recombination of genes in developing thymocytes that encode TCRα and TCRβ chains (or TCRγ and TCRδ chains), and results in the generation of a large number of unique TCRs (Kurosawa *et al.*, 1981; Davis and Bjorkman, 1988; Robins *et al.*, 2010; Murugan *et al.*, 2012). Some estimates based on deep sequencing analysis suggested that there could be between $1 - 4 \times 10^6$ unique TCRβ chains in the human peripheral blood paired on average with at least 25 different TCRα chains (Robins *et al.*, 2009, 2010), with the lower bound of overall TCR diversity in humans predicted to be around 2.5×10^7 unique TCR specificities (Arstila *et al.*, 1999). Genes that encode TCRα chain rearrange their V-J segments, whereas TCRβ genes rearrange their V-D-J segments (V - variable, J - joining, D – diversity segments) and similar process forms γδ T cell TCRs, with TCRγ rearranging V-J segments, and TCRδ rearranging V-D-J segments (Kurosawa *et al.*, 1981; Davis and Bjorkman, 1988; Legut, Cole and Sewell, 2015). Recombination of these multiple VDJ genes yields numerous T cells with distinct TCRs. The specificity of each TCR to its ligand is determined by the diversity of their six complementarity determining regions (CDR). In particular, CDR3 regions that represent amino acid sequences primarily responsible for the direct binding to antigenic peptides bound to MHC molecules on the surface of antigen presenting cells (Garboczi *et al.*, 1996; Al-Lazikani, Lesk and Chothia, 2000; Wong, Leem and Leem, 2019; Fernández-Quintero *et al.*, 2020)

An important feature of TCRs is that they, unlike most of other cell surface receptors, are cross-reactive in their ligand recognition and are able to recognise different ligands (e.g., self vs non-self) with various affinities (Birnbaum, Mendoza, *et al.*, 2014). The

bonding affinity between the TCR and MHC molecules with bound peptides determines the type of the biological response that this engagement is going to initiate, with different outcomes generated in response to self or non-self (foreign antigens) detection (Gaud, Lesourne and Love, 2018). For example, in the thymus where T cells develop and mature, the intensity of binding between the TCR and self-antigens presented by other cells, determines whether a given T cell will survive or will be eliminated from further selection. Those T cells that express TCRs that bind with high affinity to self-antigens will die by apoptosis (negative selection) and TCRs that will not be able to bind to self-antigens with sufficient affinity will also die by apoptosis. Only those T cells whose TCRs engage with self-antigens with appropriate affinity will receive survival signals and will proceed to the next stage of T cell development. This mechanism of positive and negative selection is important to ensure that only T cells with fully functional TCRs will be expelled from the thymus into the periphery. The low affinity binding of TCRs to self-MHC-peptide complexes in the periphery promotes homeostatic survival of naïve T cells, whereas high-affinity engagement with foreign antigens will promote naïve T cell activation (Hochweller *et al.*, 2010).

1.3.1   Thymic seeding progenitors

As mentioned earlier, while most of the blood cell lineages complete their maturation in BM, T cells mature in the thymus. Thymus seeding progenitors migrate out of the BM, and via circulation enter the thymus, where they undergo cell proliferation, T cell specification and commitment, as well as positive and negative selections (Zúñiga-Pflücker, 2004) (Bhandoola *et al.*, 2007; Zlotoff and Bhandoola, 2011) (Rothenberg, Moore and Yui, 2008) (Shah and Zúñiga-Pflücker, 2014). The phenotype of these thymic seeding progenitors is not currently well defined. It has been proposed that the

most uncommitted immature population of cells in the human thymus expresses the highest levels of CD34 cell surface glycoprotein (widely accepted as the marker of hematopoietic and progenitors cells in the bone marrow and thymus) and are negative for the CD1 expression (CD34$^+$CD1$^-$ cells), followed by CD34$^+$CD1$^+$ thymocytes (Galy *et al.*, 1993). One of the most commonly used markers to define T cells and T cell progenitors in the thymus, and one of the earliest markers in T cells ontogeny, in general, is CD7 (Sutherland, Rudd and Greaves, 1984; Haynes *et al.*, 1988, 1989). It has been shown that CD34$^+$CD7^{high} thymocytes are fully committed to T cell lineage and lack myeloid differentiation potential(Schmitt *et al.*, 1993). In contrast, CD34$^+$CD7$^-$ cells represent a very rare subset of immature CD1$^-$ cells and have been shown to be enriched with thymocytes with lympho-myeloid-erythroid differentiation potential (Weerkamp *et al.*, 2006; Hao *et al.*, 2008). It has been demonstrated that based on the differential expression of CD7 molecule, three distinct thymic populations can be identified with regards to their T cell lineage commitment: CD7$^-$ thymocytes with lympho-myeloid-erythroid potential, CD7^{intermediate}CD10$^+$ multi-lymphoid progenitors (T, B and NK cells) and CD7^{high} T/NK progenitors (the most abundant CD34$^+$ population in the thymus) (Hao *et al.*, 2008). In a recent study, investigators combined single cell RNA sequencing and Cellular Indexing of Transcriptomes and Epitopes by Sequencing analyses (CITEseq), a technique that allows performing RNA sequencing along with gaining insights of the surface phenotype of the cell by using antibodies. The authors applied CITEseq to unravel the phenotype of the most immature postnatal thymic seeding cells in humans and identified two cell populations with distinct cell surface characteristics (Lavaert *et al.*, 2020). Both populations were able to seed the thymus and to differentiate into early T progenitors (ETP) being referred to as TSP1 and TSP2 cells. Immunophenotypically, they defined TSP1 cells as CD34+CD44hiCD7-CD10+

and TSP2 cells as CD34+CD44hiCD7intCD10-. Some other investigators suggested that CD34hiCD45RAhiCD7low cells (Haddad *et al.*, 2006), CD34+CD38−CD10−CD62L+ cells (Kohn *et al.*, 2012) and/or CD34+CD10+CD24− cells (Six *et al.*, 2007) could be the thymic seeding cells in humans.

Once thymocytes upregulate their CD1a expression they become irreversibly committed to T cell fate and start rearranging their TCR delta (TCRδ), TCR gamma (TCRγ), TCR beta (TCRβ) and TCR alpha (TCRα) loci, in this respective order (Ratajczak, 2008) (Dik *et al.*, 2005). It has been shown that only a small number of HSC derived progenitor clones that migrate into thymus after transplantation is required to generate a highly diverse TCR repertoire (Brugman *et al.*, 2015). Once TCR rearrangements are finalised and thymocytes have gone through positive and negative selection processes, they mature into naïve CD8/CD4 and regulatory αβ T cells (Germain, 2002) (Spits, 2002) (Spits, 2002). Some thymocytes will also differentiate into NKT cells or γδ T cells. After maturation process is complete, newly formed naïve T cells migrate into the periphery and start recirculating between secondary lymphoid organs and blood.

1.3.2   Naïve T cells

T cells remain functionally and phenotypically naïve before they encounter their cognate antigen. Phenotypically naïve T cells differ from memory and effector T cells in their expression of cell surface receptors responsible for lymphocyte migration to different organs. Naïve T cells express high levels of CD62L, CCR7 and CD45RA molecules on their surface, which allows the homing of these cells to secondary lymphoid organs. Conversely, true naïve T cells do not express the CD95 molecule on

their surface unlike memory T cells (Gattinoni *et al.*, 2011). Upon antigen stimulation, naïve T cells secrete significantly lower levels of cytokines including interferon gamma (IFNγ) and IL-2 and have a slower proliferation rate compared to memory T cells (Veiga-Fernandes *et al.*, 2000). On the other hand, naïve T cells have the highest differentiation potential and are able to generate all effector and memory subsets.

Naïve T cells have been shown to have the most diverse TCR repertoires, as thymus is the only place where T cells with new specificities can be produced (Kohler *et al.*, 2005). Because of their precursor nature, naïve T cells also have significantly higher T cell receptor excision circles (TRECs) content compared to memory and effector T cells (Kimmig *et al.*, 2002). TRECs are small circular DNA fragments that are formed in T cells as they progress through the thymus as part of the T cell maturation process. TRECs are not copied during cell division but instead are diluted with cell proliferation, leading to decline in the average number of TRECs per T cell. Therefore, measuring their relative content in the pool of circulating T cells can give an indirect measurement of thymic output (Lynch and Sempowski, 2013).

The primary mechanism of naïve T cell pool maintenance and their lifespan in humans remain still open questions mostly due to discrepancies in the phenotypic characterisation of true naïve T cells in the early studies. Some initial studies that focused on understanding proliferation rates in naïve T cells using isotype labelling showed that naïve T cells (both CD8 and CD4) proliferate at much slower rates than memory T cells (Macallan *et al.*, 2003) (Vrisekoop *et al.*, 2008). However, both studies used a now outdated way of phenotypically defining naïve T cells (CD45RA+CD27+ or CD45RO-CD27+), which, as a result, included TEMRA (effector memory CD45RA+) and TSCM populations (T memory stem cells, CD45RA+CD27+CD95+) in the

35

analysis. This is due to the fact that these subsets had not yet been identified at that time. In later studies, investigators included the CD95 marker to distinguish between true naïve T cells (CD27+CD45RA+CD95-) and TSCMs (CD27+CD45RA+CD95+) and confirmed previous findings that the naïve T cells have much lower proliferation rates than memory T cells (del Amo *et al.*, 2018). Another study that included the analysis of TRECs to distinguish between newly expelled naïve T cells from the thymus and naïve T cells produced by peripheral proliferation, showed that in adult humans the vast majority of new naïve CD4 T cells were produced by peripheral division and could be maintained for up to 6-10 years in contrast to adult mice whose naïve T cell pool is maintained by thymic output and individual naïve T cell clones persist only for 6-10 weeks (Vrisekoop *et al.*, 2008)(den Braber *et al.*, 2012). However, this study also used an old definition of naïve T cells (CD45RO-CD27+), meaning that the analysis also included TSCMs together with naïve T cells. Because it was showed that naïve T cells have much lower proliferation rates and higher TREC content compared to TSCMs this suggests that a significant proportion of T cell proliferation attributed to naïve T cells in this study could be due to TSCMs (del Amo *et al.*, 2018) (Gattinoni *et al.*, 2011) (Ahmed *et al.*, 2016). It was additionally demonstrated, that individual TSCM clones can persist in circulation from 9 (del Amo *et al.*, 2018) to 12 years (Biasco *et al.*, 2015), which could partially account for the long lifespan of naïve T cells in humans predicted by Vrisekoop, et al. Moreover, TSCM cells are much more abundant in humans than mice, suggesting that failure to exclude them from analysis could explain why more replacement by proliferation rather than thymic production was seen in humans more than in mice (Gattinoni *et al.,* 2011) (Zhang *et al.*, 2005). Overall, a more comprehensive analysis with inclusion of phenotypic markers that differentiate between the true naïve T cells and TSCMs is now required to better understand the

kinetics of naïve T cells in humans and how long individual naïve T cell clones are maintained in circulation.

1.3.3   Effector subsets

There are different types of T cells that can be generated in the thymus. Most of the T cells in the human body express surface TCRs that are MHC molecule (I/II) restricted for the antigen recognition and are comprised of glycoprotein alpha and beta chains. Effector αβ T cells can be subdivided into two broad categories – CD4 "helper" T cells (Th) and CD8 "cytotoxic" T cells (Tc). Both CD4 and CD8 effector cells comprise a variety of functionally different subsets that each have a distinct cytokine profile, transcriptome, phenotype, and that allow mounting an appropriate immune response depending on the type of antigen encountered. The most commonly described CD4 effector T cell subsets include Th1, Th2, Th17 and regulatory T cells (Treg) whereas the typical CD8 effector subsets include Tc1, Tc2, Tc9, Tc17 and Tc22 (Caza and Landas, 2015) (Raphael *et al.*, 2015) (Geginat *et al.*, 2014) (Mittrücker, Visekruna and Huber, 2014) (St. Paul and Ohashi, 2020) (Zhu, Yamane and Paul, 2010). There is a minor subset of cells that is called Natural Killer T (NKT) cells, which also expresses alpha beta TCR chains (these cells always express the same variable segment – V24, and the same joining segment - Ja18 in their alpha chain) as well as NK cell receptors such as CD56 and CD161 and recognise their antigens via interaction with CD1d MHC-class-I-like molecules (Eger *et al.*, 2006) (Borg, 2007) (Kumar and Delovitch, 2014). There are two main subsets that are described for NKT cells – NKT type I and NKT type II (Bandyopadhyay, Marrero and Kumar, 2016) (Berzins, Smyth and Baxter, 2011). Both NKT types can be further subdivided into Th-1 like T cells and Th-2 and Treg like T cells and have been shown to both be able to elicit potent antitumour

responses (Th-1 like cells) and promote tumour progression (Th-2 like and Treg like cells) depending on the surrounding environmental stimuli (Krijgsman, Hokland and Kuppen, 2018). Another type of T cells expresses TCRs that consist of γ and δ chains (Passoni *et al.*, 1997) (Cowan, Jenkinson and Anderson, 2015). Conventionally γδ T cells are subdivided into different subsets based on the TCR delta chain expression and are referred to as δ1, δ2, δ3, and δ5 that pair with various gamma chains (Kabelitz *et al.*, 2005) (Willcox *et al.*, 2012). The most prevalent subset in the human peripheral blood is Vγ9 Vδ2 (Dimova *et al.*, 2015). The Vγ9 Vδ2 T cell subset has potent anti-tumour activities and can inhibit cancer cells proliferation, angiogenesis and increase cancer cell apoptosis (Di Carlo *et al.*, 2013).

After the antigen has been cleared from the system, most of the effector cells will die by apoptosis (90-95% of expanded effector T cells) (Gasper, Tejera and Suresh, 2014) (Valbon, Condotta and Richer, 2016). However, a small proportion of activated effector cells specific for this antigen will be maintained as various subsets of memory T cells. Memory T cells have an ability to persist in circulation for decades and upon stimulation with their antigen rapidly differentiate into effector cells (Bartlett, 2004) (Hammarlund *et al.*, 2003) (Farber *et al.*, 2016). Reactivation of memory T cells, and their subsequent proliferation and expansion, requires T cell receptor stimulation with its cognate antigen although certain cytokines like IL-18 and IL-15 secreted by inflammatory monocytes during microbial invasion can trigger early differentiation of memory T cells into effectors before they encounter their antigen (Soudja *et al.*, 2012). Conventionally, T cells are subdivided into distinct subsets using differential expression of cell surface receptors. The classical memory T cell subsets include central memory T cells (TCM), effector memory T cells (TEM) and effector memory CD45RA+ (TEMRA) cells

(Sallusto *et al.*, 1999) (Jameson and Masopust, 2009). Other memory subsets include tissues resident memory T cells (TRM) (Thome *et al.*, 2014) and T memory stem cells (TSCM) (Gattinoni *et al.*, 2011).

1.3.4 Memory generation in CD8 and CD4 T cells

Immunological memory is defined as an ability of immune cells activated and expanded during acute response phase to be retained long-term for more rapid and effective response, in case of the future encounters with the same antigen (Ahmed and Gray, 1996) (Swain *et al.*, 2006) (Kaech, Wherry and Ahmed, 2002). There is strong evidence that memory differentiation of CD8 T cells follows a linear model in which less differentiated CD8 T cell subsets gradually lose or acquire various phenotypic and functional markers and change their gene expression as they progress through distinct stages (Gattinoni *et al.*, 2011) (Lugli *et al.*, 2013). The generation of memory in CD4 T cell compartment seems to be less straightforward and evidence suggests that several types of memory differentiation pathways could be employed depending on the type of CD4 effectors (e.g., Th1, Th2, Tregs or Th17) and environmental cues involved (Gasper, Tejera and Suresh, 2014) (Lanzavecchia and Sallusto, 2002) (Kaech, Wherry and Ahmed, 2002) (Chang, 2012). There is evidence that at least some CD4 T cells follow linear model of differentiation, in which Th1 and Th2 effectors give rise to memory T cells directly during primary response to their pathogens (Topham and Doherty, 1998) (Marshall *et al.*, 2011) (Pepper and Jenkins, 2011) (Harrington *et al.*, 2008) (Gasper, Tejera and Suresh, 2014). Several studies showed that within the linear differentiation framework, the strength and the length of antigen exposure could influence an ability of a given T cell to generate memory cells (Lanzavecchia and Sallusto, 2002) (Moulton *et al.*, 2006). Some results suggested that stronger and longer

exposure to antigenic signals tends to drive CD4 T cells into terminal differentiation making them less able and fit to respond to memory-supporting survival signals and subsequently results in their elimination (Lanzavecchia and Sallusto, 2002). Another study showed that shorter length of antigen exposure results in less differentiated CD4 effectors that mostly produced CD62L TCM cells, and an increased duration of antigen exposure resulted in more differentiated CD4 effectors producing TEM cells (Moulton *et al.*, 2006). Overall, memory cell generation in CD4 T cell compartment is currently less clear compared to CD8 T clear compartment but based on available data could also follow linear differentiation model.

1.3.5   Memory subsets

TCM and TEM are the two most described circulating memory subsets. They are typically differentiated from one another based on their phenotype, the level, and types of secreted cytokines and self-renewal properties. Phenotypically TCM and TEM cells can be differentiated based on their expression of lymphoid homing receptors such as CD62L and C-C chemokine receptor 7 (CCR7) (Sallusto *et al.*, 1999) (Sallusto, Geginat and Lanzavecchia, 2004). TCM cells, like naïve T cells, express high levels of CD62L and CCR7 receptors that enable them to travel to lymphoid tissues, whereas TEM cells lack CCR7 and express very low levels of CD62L on their surface. TEM cells are preferentially found in liver, lungs, and gut, and have an ability to recirculate between the blood and tissues (Sallusto, Geginat and Lanzavecchia, 2004) (Ahlers and Belyakov, 2010). The proportion of TCM and TEM cells in CD4 and CD8 compartments differs in blood, with CD4 cells having higher proportion of TCM cells, whereas CD8 cells predominantly being TEM (Sallusto, Geginat and Lanzavecchia, 2004). Upon antigenic stimulation CD4 TCM cells secrete high levels of IL-2 but very

low levels of IL-4 and IFN-γ, whereas TEM cells produce high levels of IL-4 and IFN-γ but very low levels of IL-2 (Sallusto *et al.*, 1999). It has also been shown that in addition to IL-4 and IFN-γ secretion, some CD8 TEM cells also contain intracellular perforin, with a subset of TEM cells that re-expresses CD45RA (TEMRA) containing the highest levels of perforin (Sallusto, Geginat and Lanzavecchia, 2004). It has been suggested that TEM cells serve as the first line of defence against re-infection readily secreting effector cytokines upon stimulation, whereas TCM cells are responsible for stimulation of B cells and dendritic cells in secondary lymphoid organs, as well as rapid expansion and proliferation (Sallusto *et al.*, 1999). Both subsets have low antigen activation thresholds, however, the ability of memory cells to expand and proliferate decreases from TCM to TEM to TEMRA due to reduction in telomere length and higher sensitivity to pro-apoptotic signals (Geginat, Lanzavecchia and Sallusto, 2003) (Sallusto *et al.*, 1999).

The TSCM population is one of the most recently described subsets of memory T cells that was shown to have stem cell like self-renewal properties, clonal long-term survival ability, and capability to differentiate into all other memory and effector subsets upon stimulation (multipotency) (Gattinoni *et al.*, 2011; Biasco *et al.*, 2015). In one study, investigators analysed the difference in expression of 900 genes that they identified were differentially expressed in four CD8 T cell subsets (TN, TSCM, TCM and TEM) (Gattinoni *et al.*, 2011). They showed that based on the gene expression pattern, TSCM population represents the least differentiated memory subset. The group used a multidimensional scaling analysis (technique that allows the visualisation of the level of similarity of individual cases in a data set) to assess the level of similarity in gene expression between true naïve T cells and 3 memory subsets. They showed that true

naïve T cells have more similarity in gene expression with TSCM subset (75 differentially expressed genes) than with TCM and TEM subsets (157 and 226 differentially expressed genes, respectively). The highest level of similarity was observed between TSCM and TCM cells – only 20 differentially expressed genes out of 900. It has been proposed that TSCM is the primary subset responsible for the long-term maintenance of immunological memory due to their ability to self-renew and their long lifespans (Gattinoni *et al.*, 2011; Biasco *et al.*, 2015) (Fuertes Marraco *et al.*, 2015). Recently, however, it was demonstrated that TSCM cells classically defined as CD45RO$^-$CD27^{bright}CCR7$^+$CD95$^+$ (Mahnke *et al.*, 2013) consist of at least two kinetically distinct populations (del Amo *et al.*, 2018). The majority of TSCM cells exhibit a very short half-life of typically 5 months – first population, whereas the other much smaller in size TSCM population has a half-life of approximately 9 years. In addition to longevity, the latter TSCM population also exhibited a high degree of self-renewal, suggesting that this small long-lived TSCM population could be the one responsible for the long-term immunological memory.

1.4   NK cells

Natural Killer (NK) cells were discovered in the 1970s and were initially described as lymphocytes with non-MHC-restricted cytolytic activity that are able to recognise their target antigens with germline encoded receptors without any prior antigen priming unlike other lymphocytes (Pross and Baines, 1976) (Kiessling, Klein and Wigzell, 1975). Now it is well established that apart from their cytotoxic activity, NK cells are also important players in the immune regulation via their interaction with other immune cells like monocyte/macrophages, dendritic cells, and T cells (Agaugué *et al.*, 2008) (Moretta, 2005). Since their first recognition, NK cells are considered to belong to the lymphoid lineage together with T and B lymphocytes based on their development from common lymphoid progenitor in the BM, their morphology and expression of lymphoid cell surface receptors (Vivier *et al.*, 2011). However, in contrast to T and B lymphocytes, they are classified as part of the innate immunity due to the lack of antigen specific receptors. Since the recognition of innate lymphoid cells (ILC), which represent a lineage of professional cytokine producing cells, NK cells were placed in this group and are often referred to as cytotoxic ILC subset (Vivier *et al.*, 2018). NK cells have two main immune recognition strategies of their target antigens - recognition of self and non-self. Detection of foreign antigens or "non-self" is aimed at recognising various microbes and viruses whereas detection of "missing self" is aimed at detecting cells that do not express sufficient levels of MHC class-I receptors on their surface - a common feature of virally infected and tumour cells (Kärre, 2008) (Croft *et al.*, 2019). Induced self-recognition involves an ability of NK cells to detect ligands of their target cells that are normally absent or present in very low amounts on healthy cells and are overly expressed on stressed, virally infected or tumour cells (Medzhitov and Janeway,

2002). NK cells recognise their targets using germline-encoded, non-rearranged activating or inhibitory receptors such as MHC class-I binding receptors that fall into two major categories: C-type lectin like receptors that are mostly formed by the combination of CD94 with either NKG2A (inhibitory receptors) or NKG2C (activating receptors); and killer immunoglobulin like receptors (KIRs) (Colonna, Navarro and López-Botet, 1999).

1.5   NK cell development in human and mice

Initial experiments demonstrated that intact bone marrow environment was essential for the generation of fully functional, cytotoxic NK cells in mice (Hackett, Bennett and Kumar, 1985). Murine NK cell progenitors were identified in the bone marrow based on their expression of IL2/IL15 (CD122) receptor beta chain (Rosmaraki *et al.*, 2001). Further studies allowed identification of the most primitive murine pre-NK cell progenitors in the bone marrow that were phenotypically defined as lineage negative (Lin−) CD27+ CD244+ CD122− IL-7Rα+ FLT3− cells and lacked any other typical NK cell associated receptors (such as NKp46 and NKG2D) on their surface (Fathman et al., 2011). It has been shown that in the NK cell development promoting culture conditions, these pre-NK cell progenitors upregulate CD122 receptor on their surface and acquire full NK cell potential (Lodolce *et al.*, 1998; Kennedy *et al.*, 2000; Fathman *et al.*, 2011).

There is still no definitive and well-established description of NK cell development in humans. Unlike murine NK cells that primarily mature in the BM, human NK cells were shown to differentiate in secondary lymphoid organs (Freud, Yu and Caligiuri, 2014; Freud *et al.*, 2017). Human NK cell progenitors are phenotypically defined as Lin−

CD34+ CD38+ CD123− CD45RA+ CD7+ CD10+ CD127− cells and were shown to be unable to produce any T cells, B cells, myeloid cells or innate-lymphoid like cells (Renoux *et al.*, 2015). Conventionally, mature committed NK cells are phenotypically described as Lin-CD56+ cells (lineage markers include CD3 to exclude T cells, CD19 or CD20 to exclude B cells, and CD14 to exclude monocytes) (Freud, Yu and Caligiuri, 2014). Differential levels of expression of CD56 marker can be used to split NK cells into two distinct subsets – CD56bright and CD56dim NK cells (Cooper, Fehniger and Caligiuri, 2001; Poli *et al.*, 2009). CD56dim NK cells are overrepresented in human peripheral blood, co-express high levels of CD16 markers and are thought to be the most mature NK cell subset with highest levels of cytotoxicity and target induced cytokine production ability (Nagler *et al.*, 1989; Fauriat *et al.*, 2010) (Zhang and Yu, 2010). CD56bright NK cells represent a small population in the peripheral blood. These cells have very low levels or no expression of CD16 receptor, exhibit low levels of natural cytotoxicity but have a better proliferation ability and increased cytokine secretion levels in response to IL-15 and IL-18 cytokines compared to CD56dim NK cells (Cooper, Fehniger and Caligiuri, 2001) (Fehniger *et al.*, 1999). CD56dim NK cells can be further subdivided into two distinct subsets based on the CD57 expression, with CD56dimCD57+ NK cells described as a terminally mature subset with the greatest killing capacity (Lopez-Vergès *et al.*, 2010). CD56bright cells represent only about 10% of all blood circulating NK cells in healthy people but are significantly enriched in secondary lymphoid tissues where they play an immunomodulatory role (Fehniger *et al.*, 2003). It has been shown that in addition of being a functionally distinct subpopulation, CD56bright NK cells also represent direct precursors of more mature CD56dim NK cell subset (Caligiuri, 2008) (Chan *et al.*, 2007) (Ouyang *et al.*, 2007) (Romagnani, 2007).

1.6 Memory NK cells

In the past decade, several studies demonstrated that NK cells are capable of eliciting memory-like antigen specific responses against various pathogens including herpesviruses, papillomaviruses, and cytomegaloviruses (CMV) (Orange, 2002). One of the first studies that showed the existence of NK memory cells, identified a subset of murine NK cells that expressed activating receptors - Ly49H on their surface that recognised a viral glycoprotein - m157, which is expressed on the surface of murine CMV infected cells (Sun, Beilke and Lanier, 2009). This study showed that infection with murine CMV results in activation of Ly49H+ NK cell subset and leads to their clonal expansion, with most effector cells being eliminated after antigen clearance. However, a small proportion of Ly49H+ NK cells is maintained as a memory pool that is able to elicit enhanced responses upon secondary challenge. The expansion of effector NK cells and a generation of memory occurs in receptor (Ly49H) and its ligand (m157) dependent manner, suggesting that this is a true antigen specific memory generation (Sun, Beilke and Lanier, 2009).

The existence of human CMV specific memory NK cells has also been demonstrated. It has been shown that individuals previously infected with CMV have an increased frequency of the NK cell subset that expresses CD94 and NKG2C receptor complex (Gumá *et al.*, 2006). Transfer of NKG2C+ NK cell subset into patients during haematopoietic stem cell transplantation resulted in NKG2C+ NK cell activation, expansion and potent IFN-γ production in response to CMV re-activation in these patients (Foley *et al.*, 2012). This NKG2C+ NK cell subset expresses high levels of maturation marker CD57, inhibitory immunoglobulin like transcript2 (ILT2) and KIRs, but have lower levels of natural cytotoxicity receptors, NKp30 and NKp36 receptors

(Gumá *et al.*, 2006; Lopez-Vergès, 2011). One recent study showed that only individuals with history of CMV infection harbour CD56dim FcεRIγ- (Fc fragment of the high affinity receptor for IgE antibody) NKG2C+ NK cells and FcεRIγ-NKG2C- and FcεRIγ+NKG2C+ NK cells, with each of these populations having distinct functional and phenotypical characteristics (Kim *et al.*, 2019). The CD56dim FcεRIγ-NKG2C+ NK cells exhibited the most restricted KIR repertoire suggestive of clonal expansion. This data is supported by another study that showed that NKG2C+ NK cells exhibit clonal-like expansion in response to human CMV UL40-encoded peptides presented by HLA-E (Redondo-Pachón *et al.*, 2017). It has been previously shown that long-lived memory CD8 T cells express low levels of Ki-67 protein (lower concentration is associated with slower divided cells) suggesting that these cells are quiescent but high levels of Bcl-2 protein, which allows CD8 memory T cells to be resistant to pro-apoptotic signals and as a result promotes their long-term persistence (Miller *et al.*, 2008) (Akondy *et al.*, 2017). Similarly to CD8 memory T cells, FcεRIγ-NKG2C+ NK cells expressed lower levels of Ki-67 compared to FcεRIγ-NKG2C- and FcεRIγ+NKG2C+ subsets, and significantly higher levels of CD57 marker (associated with high maturation and terminal differentiation of NK cells) – both features are associated with long-lived true memory cells (Kim *et al.*, 2019).

1.7 T/NK cell common origin hypothesis

One of the first studies which hypothesised that T and NK cells might share a common origin has been conducted by Lanier et al. in 1992 (Phillips *et al.*, 1992). In this study, they isolated NK cells from foetal tissue and showed that in contrast to adult NK cells from peripheral blood, foetal NK cells as well as a subset of NK cells in the cord blood expressed intracellular (but not surface) CD3δ and CD3ε. Because CD3 complex is a

T cell co-receptor that plays an essential role in TCR signalling, this finding led to the first hypothesis that T and NK cells might share a common progenitor that expresses components of the CD3 complex. The same year, another study found foetal thymocyte population in mice that lacked CD4 and CD8 receptors on their surface but expressed Fc gamma II/III receptors (these receptors belong to the immunoglobulin family of receptors commonly expressed on various immune cells including NK cells) prior to TCR upregulation (Rodewald *et al.*, 1992). They showed that if this CD4-CD8-FcgammaII/III+ cell precursor population remains in the thymic environment it generates functional T cells. However, if these precursors were removed from the thymus, they differentiated into canonical NK cells with full cytotoxic potential. The presence of T and NK cell biased progenitor in humans was suggested in two other early studies (Sánchez *et al.*, 1993, 1994). Based on the differential expression of CD56, CD5 and CD34 cell surface receptors, it has been shown that there are at least three populations of CD3-CD4-CD8- cells in the human thymus that could be identified using these markers. The CD56+CD5- cell population exhibited the phenotype like NK cells in the foetal liver, were able to generate NK cell clones, and were unable to produce T cells when cultured in mouse foetal thymic organ cultures (mFTOC), suggesting that these cells represent a more mature linage-restricted NK cell population in the thymus. The CD56-CD5- thymocytes had a similar phenotype to thymic NK cells (CD56+CD5-) and were able to generate NK cell clones that then upregulate their CD56 expression and acquire canonical NK cell cytolytic potential. Hence, the CD56-CD5- thymocyte population might represent an NK cell committed progenitor subset. The CD56-CD5+CD34+ thymocytes were only able to generate CD3+ CD8+/CD4+ T cells when cultured on mFTOC. In this study, investigators highlighted an interesting observation, that based on their data once a thymic

progenitor commits to NK cell lineage it completely loses its ability to differentiate along the T cell pathway.

Adaptive immunity has evolved in organisms resembling primitive jawed fish around 430 million years ago (Bayne, 2003). Before the possible lateral transfer of prokaryotic transposase and consequently the emergence of the process of somatic rearrangement of immunoglobulin receptors in T and B cells has evolved, genes encoding immunoglobulin superfamily receptors has already been expressed in early organisms for hundreds of millions of years (Agrawal, 2000). Therefore, innate myeloid and NK cells, that use their germline encoded receptors to recognise their targets might represent a more ancestral developmental programme of immune cell differentiation. As discussed previously in paragraph 1.2, haematopoiesis exhibits high levels of plasticity, with progenitors with some degree of commitment to B and T lineages still retaining their myeloid and NK cell potentials. There is some evidence that the commitment of progenitors to T cell lineage requires T cell determining factors to be enforced in precursor cells and in the absence of such, progenitors become committed to the NK cell lineage. One of these factors is Bc11lb zinc finger transcription factor, that has been shown in to play an essential role in the early T cell development (Wakabayashi *et al.*, 2003). This study demonstrated that Bcl11b-defficient mice had an impaired thymocyte development at CD4-CD8- double negative stage with inability to rearrange their TCR beta segments, without it having any effect on B cells or γδ T cells development. The presence of active Bcl11b gene was shown to repress NK cell associated genes and deletion of Bcl11b in conditional knockout mice resulted in the arrest of T cell development and redirection of progenitor cells towards the development of cells morphologically and transcriptionally resembling NK cells. These

NK cells were able to kill tumour cells in vitro and prevent tumour metastasis in vivo (Li *et al.*, 2010) (Ikawa *et al.*, 2010). This suggests that there might be a common T/NK cell progenitor and that early progenitors with T cell potential but with insufficient or absent expression of Bcl11b differentiate into NK cells. One interesting case report showed that an infant with mutant BCL11B variant, resulting in inability of wild-type BCL11B protein to bind to its targets, exhibited a leaky severe combined immune deficiency syndrome (SCID; leaky SCID means that a minimal level of immune activity is preserved) with very low T cell counts but normal NK cell numbers (Punwani *et al.*, 2016). This suggests that expression of Bcl11b might be essential for overcoming a more ancient ancestral NK-like cell developmental programme to allow progenitors to differentiate into T cells.

More support towards the developmental relationship between T and NK cells came from the study where investigators compared similarities in gene expression between two immune cell subsets and demonstrated using transcriptome-wide analysis that these cells shared expression of hundreds of genes (Bezman *et al.*, 2012). In addition, they demonstrated by using principal component analysis that NK and T cells clustered together within the complex that was distinct from other immune cell populations including B cells, dendritic cells, and macrophages.

Overall, there is compelling amount of evidence in both human and mice that suggests a close developmental relationship between T and NK cell lineages and the existence of T/NK bipotent progenitors.

1.8 T and NK cell deficiency in SCID-X1 patients

Mutations in genes that are essential for the differentiation and/or function of immune cells (e.g., cell surface receptors) can have devastating effects on the function of the immune system. Primary immune deficiencies (PIDs) are a group of inherited diseases that are caused by mutations in genes that are essential for the function of human immune system. Depending on what component of the immune system is affected and to which extent – these diseases have mild or severe clinical manifestations. One example is the X-linked severe combined immune deficiency disorder (SCID-X), which is a rare (1:50,000 – 1:100,000 incidents per births), inherited disorder of the immune system (Accetta Pedersen, Verbsky and Routes, 2011). SCID-X1 is caused by mutations in the gene that encodes interleukin 2 receptor common gamma chain (IL2RG) subunit, an essential component of multiple interleukin receptor complexes. These receptors induce the secretion of chemical signals called cytokines that are crucial for the development of different immune cell types. The glycoprotein that is encoded by IL2RG gene plays an essential role in the function of several different interleukin receptors like IL-2, IL-4, IL-7, IL-15, and IL-21. IL-7 and IL-15 are essential for the development of functional T and NK cells, respectively. Therefore, SCID-X1 is characterized by a significant reduction (leaky SCID) or almost complete absence of T and NK lymphocytes in affected individuals, which also subsequently affects B cell function (however, patients preserve normal B cell numbers) (Gaspar *et al.*, 2004) (Baum *et al.*, 2007). Absence of T and NK lymphocytes, and impaired B cell immunity is responsible for severe recurrent bacterial and viral infections in affected individuals. Patients normally present around 3-6 months of age, and SCID-X1 is almost always fatal unless treated. The only conventional treatment for this disorder is hematopoietic

stem cell transplantation (HSCT), which in the case of genetically fully matched donor results in 90% long-term survival rate (Antoine, et al., 2003). In the absence of a matched donor, patients could be offered ex vivo retroviral/lentiviral gene therapy (GT).

1.9 Gene therapy for SCID-X1

Somatic gene therapy for monogenic disorders, like SCID-X1, involves the use of viral vectors to carry a functional copy of the mutated gene. Some viruses like gamma-retroviruses and lentiviruses have a natural ability to enter cells and integrate their genetic material into the host cell genome. This feature is used in gene therapy, where viral gene sequences are replaced with therapeutic gene of interest. Resulting viral vectors carrying functional copy of a defective gene are then used to transduce CD34+ hematopoietic stem cells (HSCs) to restore protein expression and correct cell function.

CD34+ hematopoietic stem cells (HSCs) are a heterogeneous population of stem cells and cell progenitors that can reconstitute and maintain the entire immune system (Doulatov *et al.*, 2012). Therefore, all cells that will originate from gene corrected CD34 HSCs will permanently carry the corrected copy of the gene of interest, thereby restoring functional immunity in patients.

Viral vectors tend to integrate semi-randomly selecting preferred sites in the genome of the host cell (Wu *et al.*, 2003). Gamma-retroviral vectors prefer integrating at transcription start sites and CpG islands, whereas lentiviral vectors tend to integrate in the introns of active transcription units (Desfarges and Ciuffi, 2010). Vector integration sites create a molecular signature in the genome of CD34 HSCs, thereby permanently

marking individual CD34+ HSCs with a molecular identifier unique for each transduced cell. All progeny that will originate from this molecularly marked CD34 HSC will have identical insertion site in their genome.

In some instances, viral vectors cause insertional mutagenesis and result in malignant cell transformation in patients. This happened in a gamma retroviral GT trial for SCID-X1, conducted 20 years ago at Great Ormond Street Hospital (GOSH, London) and Necker Hospital (France). In this trial, the gamma retroviral MFG-gc vector was used to deliver a functional copy of IL-2 receptor common gamma chain (IL2RG) gene into CD34+ HSCs in patients (Hacein-Bey-Abina *et al.*, 2010) (Gaspar *et al.*, 2004). Without functional IL2RG gene, T cells are unable to develop. Therefore, in these treated patients, only gene corrected T cells can mature and proliferate, and eventually give rise to long-term persisting memory T cells. The MFG-gc vector had a strong enhancer element in its long terminal repeat (LTR) sequence driving the expression of the IL2RG gene. In 5 out of 20 patients (one patient in London and four patients in France), the vector integrated in the proximity to the proto-oncogene called LIM domain only 2 (LMO2) and caused overexpression of LMO2 (Howe *et al.*, 2008). In some patients, blast cells also contained integrations next to BMI1 and CCND2 proto-oncogenes (Hacein-Bey-Abina *et al.*, 2008). This overexpression led to the development of T cell acute lymphoblastic leukaemia (T-ALL) in 5 patients. A new generation of self-inactivating (SIN) viral vectors that lack these enhancer promoter elements to ensure safety and minimize the chances of transactivation of oncogenes have now been developed (Kraunus *et al.*, 2004).

Various conditioning regimens can be used in patients prior to HSCT to improve engraftment of transduced HSCs including myeloablative conditioning, reduced-

intensity conditioning and non-myeloablative conditioning regimens. In animal models, total body irradiation with high and medium radiation doses (1.5-3.0 Gy) significantly improved engraftment of transplanted HSCs compared to animals that had only very low level or no body irradiation prior to HSCT (Tomita, Sachs and Sykes, 1994). The choice of conditioning regimen depends on the disease type (malignant versus benign) and status, and overall health of a patient. In case of PID GT patients, the main justification for using conditioning prior to HSCT is to make space in BM HSC niches, thereby increasing levels of engraftment of transplanted gene corrected autologous HSCs.

However, because there is a selective advantage of transduced T cells in SCID-X1 patients and because even reduced intensity conditioning is still toxic for patients and might result in infertility or even cancer (Ferry and Socié, 2003), no cytoreductive conditioning was used in SCID-X1 trial.

Conventionally, it is thought that only HSCs are capable of long-term reconstitution of the entire immune system. Upon transplantation by peripheral blood infusion, hematopoietic stem and progenitor cells home to the BM, occupy specific niches and start to give rise to all blood cell lineages, most of which are short-lived and hence require constant replenishment by stem and progenitor cells from the BM (Calvi *et al.*, 2003) (Hackney *et al.*, 2002). Because no conditioning was done for patients in SCID-X1 trial, low levels of stem cell engraftment were expected in SCID-X1 patients. However, if genetically engineered haematopoiesis in these patients is still supported by transduced HSCs, we would expect to observe the transgene in all types of immune cell lineages. Conversely, if HSCs are lost overtime all mature transduced cells would progressively disappear from circulation with the exception of memory lymphocytes.

1.10 VISA for safety assessment and immune cell tracking during hematopoietic reconstitution

Vector integration site analysis is a molecular technique that allows the identification of vector integration sites. The amplification of vector-genome junctions by linear amplification mediated polymerase chain reaction (LAM-PCR) combined with - high throughput sequencing allows the retrieval of integrated viral sequences and their insertion sites in transduced HSCs and their progeny.

As part of patient monitoring routine after GT, this technique is used as a safety measure to ensure that clones with potentially oncogenic integrations are present at relatively low frequencies and that there is no clonal expansion in patients. Vector integration site analysis can also be used to study the dynamics and contribution of individual hematopoietic stem and progenitor cells (HSPC) to haematopoiesis in humans (Cavazzana-Calvo *et al.*, 2011) (Biasco *et al.*, 2016), as well as to better understand relationships between different cell lineages and specific subsets (Biasco *et al.*, 2015).

Using vector integration site analysis, the group led by Dr. Biasco in Milan has studied the differential contribution of various hematopoietic and progenitor cells (HSCPs) to haematopoiesis in six Wiskott-Aldrich syndrome gene therapy (WAS-GT) patients (Scala *et al.*, 2018). In this trial, patients received autologous CD34+ HSCs, transduced with lentiviral vector carrying functional copy of the wasp gene. Investigators collected bone marrow (BM) and peripheral blood (PB) samples from these patients at different time points. Based on the differential expression of cell surface markers, they isolated different populations of HSPCs including HSCs, early T cell progenitors (ETP),

common myeloid progenitors, granulocyte-monocyte progenitors (GMP), multipotent progenitors, multi-lymphoid progenitors (MLP), B and NK cell precursors (pre-B/NK) and megakaryocyte-erythrocyte progenitors (MEP). Comparing integration sites from these different progenitor types with cells isolated from PB at different time points post GT, they have shown that at the earliest stages (<12 months post GT) of hematopoietic reconstitution – MPPs tend to contribute to haematopoiesis more than HSCs. However, during stable phase (>2 years post GT) of hematopoietic reconstitution, HSCs took over haematopoiesis in these individuals. Interestingly, when they compared integration site sharing between PB blood lymphocytes, MPPs, MLPs, pre-B/NK cells and HSCs at late time points, they observed that clonal composition and output of early lymphoid precursors (MLPs and preB/NK cells) in the BM of GT patients appears to be statistically independent from the upstream contribution of MPPs or HSCs. This data suggested that there might be long-term persisting lymphoid progenitors (Lt-LP) that are able to contribute to haematopoiesis independently from HSCs many years post GT.

1.11 Preliminary observations at the basis of this PhD project

In our SCID-X1 GT clinical trial, as part of the patient follow up, vector copy number (VCN) analysis (detection of integrated vector sequences inside patients' DNA by means of quantitative PCR (qPCR) or digital PCR, see Chapter 6, sections 6.8 and 6.9 for protocols) of various immune cell populations, including T cells, NK cells, B cells, neutrophils, and monocytes, was performed at regular intervals. At around 27 months post GT only T and NK cells were still vector positive, without any detectable levels of transgene in all other immune cell types (see Chapter 3, section 3.1). Because only T and NK cells were still vector positive, we assumed that these patients lost their

genetically modified HSCs and now only have mature memory lymphocytes present in their circulation. We then decided to look closer at circulating T cells to better understand their subset composition. To our surprise, immunophenotypic analysis of T cells in these patients revealed that these individuals still have naive T cells present in their peripheral blood after up to 18 years post gene therapy, suggesting that there might be a de novo T cell production still occurring in these patients. If the vector-positive T and NK cells that we observe in our SCID-X1 patients have originated from conventionally described common lymphoid progenitors, we should expect to see also vector-positive circulating B cells. Because no transduced B cells are observed, such persistence of genetically modified T and NK cells raises an intriguing hypothesis that a population of genetically engineered T and NK specific lymphoid progenitors exist in these SCID-X1 patients. Observations from our SCID-X1 study taken together with data from Scala, et al. align in support for the hypothesis of the existence of long-term persisting lymphoid progenitor.

1.12 Using VISA to unravel the existence of a bipotent T/NK Lt-LP in SCID-X1 patients

For this project we exploited our unique SCID-X1 GT trial setting (no conditioning, low engraftment) to characterise vector positive naïve T cells phenotypically and functionally and demonstrate that T cell reconstitution is supported by Lt-LP capable of preserving the T naive thymic output.

To this purpose, we established five main aims in this project:

1) Confirm that circulating *bona fide* (genuine) naïve T cells are all vector positive.

2) Show that these circulating transduced naïve T cells are phenotypically naïve using combinations of cell surface markers to differentiate between T cell subsets.

3) Demonstrate that vector positive phenotypically naïve T cells are also functionally naïve (e.g., in their response to antigen stimulation), are still capable of differentiation into distinct T cells subsets and exhibit molecular signatures of true naïve cells (e.g., high levels of different TCR specificities).

4) Analyse vector integration sites, show that identical clones of naïve vector positive T cells can be detected over prolonged period, to demonstrate continuous output from Lt-LPs containing identical integration sites.

5) Compare integration profiles collected from circulating T and NK cells in SCID-X1 patients to prove that they originate from the same progenitor clone.

For this project, we analysed blood samples from SCID-X1 gene therapy UK patients, including the patient that developed T-ALL after treatment (PT8).

1.13 Cancer Immunotherapy using CAR T cells

The ability of T cells to recognize their cognate antigen is due to the function of their antigen specific TCRs. Under normal conditions, T cells are highly efficient at recognizing and destroying cancer cells. However, some cancers develop mechanisms that allow them to escape immune surveillance. The discovery of tumour infiltrating lymphocytes (TILs) that have a specificity for cancer cells and are in general associated with better survival in cancer patients, led to the development of adoptive T - cell therapies (ACT) for patients with advanced cancers (Clemente *et al.*, 1996)

(Zhang *et al.*, 2003). During ACT, TILs are isolated from the patient's tumour, activated, expanded in vitro, and infused back into the patient.

The effectiveness of T cell-based cancer immunotherapy depends on the capacity of T cells to recognize and subsequently eradicate cancer cells. T cell antigen specificity can be enhanced by genetic modification and redirection of their TCRs towards tumour antigens. In this type of therapy, T cells are isolated from patients and are genetically modified to express chimeric antigen receptors (CAR) that have an enhanced specificity towards specific antigens.

CARs are synthetic hybrid receptors (part antibody, part TCR) that consist of four main structures – an extracellular antigen binding domain for target recognition, a hinge region, a transmembrane domain and one or several intracellular signalling domains (Sterner and Sterner, 2021). The antigen binding single chain variable fragment (scFv) of CAR constructs conventionally consists of variable heavy and variable light chains of an antibody joined by a peptide spacer and linked to the intracellular signalling molecule that consists of ITAM-containing protein such as TCR CD3ζ chain and can also include optional co-stimulatory domains such as 4-1BB or CD28 (Davila, Sauter and Brentjens, 2015; Feins *et al.*, 2019; Sterner and Sterner, 2021). Introduction of CAR constructs into T cells adds on to their antigen specificity and allows them to target specific tumour antigens. Because scFv portion of CARs is derived conventionally from monoclonal antibodies, typically non-human, such as murine anti-human CD19 monoclonal antibody (for instance FMC63 or SJ25C1 antibodies), CAR T cells can recognise their target antigens directly via scFv portion of the CAR receptor, making tumour recognition process independent from MHC/peptide recognition/presentation root (Guedan *et al.*, 2019; Sterner and Sterner, 2021). This

enables them to recognise even those tumours that downregulated expression of their surface MHC molecules to escape immune surveillance (Feins *et al.*, 2019).

Different antibody clones that are used to produce scFv portions of CARs confer different specificities to CAR T cells and could result in different clinical outcomes. For example, in phase I clinical study a patient with metastatic seminal vesicle cancer has been infused with two different CAR-T cell lines to target MUC1 protein on the surface of tumour cells (You *et al.*, 2016). The first CAR T cell line had an scFv domain derived from SM3 antibody clone specific for the deglycosylated form of this protein and co-expressed IL-12 on their surface. The other CAR T cell line contained scFv that was based on the sequence recognised by HMFG2 clone that can recognise both glycosylated and deglycosylated forms of MUC1 protein. HMFG2 based CAR T cells showed better ability to destroy tumour cells and resulted in tumour necrosis, whereas SM3 based CAR T cells did not show any efficacy at decreasing tumour burden. The affinity of scFv region towards its target has also been shown to result in different clinical outcomes by changing the strength of T cell signalling and allowing CAR T cells to distinguish between overexpressed proteins on cancer cells and normal levels of the protein on healthy cells (Liu *et al.*, 2015) (Chmielewski *et al.*, 2004). High-affinity CAR T cells recognise tumour antigens that are expressed at any level, meaning that together with cancer cells they will also destroy healthy cells that express the same antigens but at low/normal levels. This happened to one patient who was infused with CAR T cells specific to ERBB2 protein overexpressed on her colon cancer cells (Morgan *et al.*, 2010). However, ERBB2 protein is also expressed at low levels on the lung epithelial cells, which resulted in the rapid infiltration of CAR T cells into lungs of the patient after infusion causing severe respiratory distress and eventually death.

CAR T cells with low affinity for an antigen were shown to maintain the same level of potent anti-tumour responses but only destroyed cancer cells that overexpressed target antigen with no damage to healthy cells (Liu *et al.*, 2015).

Overall, CAR T cell-based therapies showed high success rates in treatment of patients with relapsed/refractory acute lymphoblastic leukaemia, chronic lymphoblastic leukaemia and non-Hodgkin's lymphoma (Turtle *et al.*, 2016) (Davila *et al.*, 2014) (Porter *et al.*, 2011). However, CAR T cells tend to disappear from circulation relatively soon after infusion, which makes these patients more prone to a relapse (Maude *et al.*, 2014; Turtle *et al.*, 2016) (Lee *et al.*, 2015). There are multiple parameters that affect the ability of CAR T cells to expand and persist in patients including individual properties of infused T cells, the type of co-stimulatory domains used in CAR T cell design, ability of tumours to prevent CAR T cell proliferation and persistence, and immune anti-CAR T cell responses that destroy CAR T cells upon infusion (Kawalekar, et al., 2016). The non-human origin of antibodies that are used to create scFv domains in some cases elicited anti-CAR T cell responses with elimination of CAR T cells from circulation of these patients and as a result poor or absent CAR T cell expansion and persistence (Jensen *et al.*, 2010; Maude *et al.*, 2014; Turtle *et al.*, 2016). In one instance, a patient developed anti-CAR T cell responses and suffered from fatal anaphylactic reactions after multiple infusions of CAR T cell products (Maus *et al.*, 2013). The immunogenicity of CARs can be reduced by generating human or humanised scFv domains. There is a group working on validating the efficacy of CAR T cells with human derived scFvs targeting mesothelin protein (Adusumilli *et al.*, 2014). Another group generated CAR T cells carrying human scFvs specific for CD19 antigen, that recognised the same epitope with the same affinity as murine derived FMC63

scFvs used in approved CAR T cell therapies – Kymriah (Maude *et al.*, 2018) and Yescarta (Viardot *et al.*, 2019) (AlDallal, 2020) (Sommermeyer *et al.*, 2017). They showed that CAR T cells expressing human-CD19 scFvs had superior ability to recognise CD19 targets expressed on tumour cell lines and primary human leukaemia compared to FMC63 CAR T cells and were better at killing lymphoma xenografts in NSG mice.

An addition of co-stimulatory domains such as CD28 and 4-1BB was shown to greatly enhance CAR T cell activation, expansion, and persistence (Finney *et al.*, 1998; Krause *et al.*, 1998; Maher *et al.*, 2002; Finney, Akbar and Lawson, 2004; Imai *et al.*, 2004; Friedmann-Morvinski *et al.*, 2005). T cells use TCRs to recognise short epitopes (8-12 amino acids long) on the surface of their foreign (non-self) targets (Rossjohn *et al.*, 2015). The short length of TCR epitopes can result in TCRs recognising sequences that are similar to their target one, including similar epitopes expressed by host cell proteins (cross-reactivity). Binding of TCRs to host antigens could potentially lead to T cell activation and therefore host tissue damage, autoimmunity and even death. For this reason, T cells require at least two signals to get fully activated (Sharpe and Freeman, 2002). The initial recognition and binding of TCR to its target - is the first required signal, and the second signal is delivered through the binding of co-stimulatory molecules like CD28 on the T cell surface to CD80 or CD86 receptors that are typically expressed on the surface of antigen presenting cells (APCs) (Sharpe and Freeman, 2002) (Rossjohn *et al.*, 2015). Hence, if a T cell encounters an epitope with similar specificity on the surface of a normal (non-APC) cell, this cell will not be able to fully activate a T cell as it will not express any co-stimulatory receptors. On the other hand, during inflammation or infections, APCs are activated and start expressing CD80

and CD86 that enables them to fully activate T cells, which in turn results in T cell's target killing and long-term persistence (Rossjohn *et al.*, 2015) (Sharpe and Freeman, 2002). There are three different types of CAR T cells based on the number of co-stimulatory domains they contain – first generation CARs contain only CD3ζ chain and no co-stimulatory domains, second generation CARs contain one co-stimulatory domain and CD3ζ chain, and a third generation CARs contain multiple co-stimulatory domains and CD3ζ chain (Guedan *et al.*, 2019). The first study that clearly demonstrated in vivo in patients that second generation CARs containing CD28 co-stimulatory domains are superior in their survival and clonal expansion to the first-generation CARs containing only CD3ζ chain as a signalling domain was conducted by Savoldo and colleagues in 2011 (Savoldo *et al.*, 2011). Since then, multiple studies demonstrated efficacy of the second-generation CAR T cells both with CD28 molecule as a co-stimulatory domain (Davila *et al.*, 2014) (Lee *et al.*, 2015) as well as CAR T cells containing 4-1BB co-stimulatory receptor (Porter *et al.*, 2011) (Kalos *et al.*, 2011; Maude *et al.*, 2014, 2018; Schuster *et al.*, 2017) (Garfall *et al.*, 2015). The CD28 receptor co-stimulation is responsible for T cell activation and IL-2 secretion, whereas 4-1BB co-stimulation is associated with long-term survival of T cells (Long, 2015). It has been shown that tonic signalling that occurs due to the antigen independent clustering of CAR's scFvs can result in early exhaustion of CAR T cells. The co-stimulation of CD28 receptors amplified CAR T cells exhaustion, whereas the co-stimulation of 4-1BB molecules reduced CAR T cell exhaustion (Long, 2015). Reduction in tonic signalling in 4-1BB CAR T cells could be one of the reasons why they tend to have a longer persistence in patients compared to CD28 CAR T cells (Lee *et al.*, 2015) (Maude *et al.*, 2014). The efficacy of the third generation CAR T cells that contain both one of the immunoglobulin superfamily members (CD28 or ICOS)

together with tumour necrosis factor receptor superfamily member (4-1BB or OX40) is also being actively studied by several groups both pre-clinically and in patients (Carpenito *et al.*, 2009) (Zhao *et al.*, 2015) (Lai *et al.*, 2018) (Dai *et al.*, 2020) (Schubert *et al.*, 2019).

1.14 Preliminary observations at the basis of this PhD project

CARPALL is a phase 1 clinical trial designed to treat paediatric patients and young adults (25 years old and younger) with high-risk relapsed B cell acute lymphoblastic leukaemia (B-ALL) (NCT02443831) (Ghorashian *et al.*, 2019). In this trial, investigators used CAR T cells expressing new low-affinity CD19 CAR (CAT) scFv and 4-1BB to treat these patients. The affinity to CD19 antigen of the new CAT construct is more than 40 times lower than the affinity of the FMC63 construct used in Kymriah therapy and many other trials (Ghorashian *et al.*, 2019) (Lee *et al.*, 2015) (Turtle *et al.*, 2016) (Gardner *et al.*, 2017) (Maude *et al.*, 2018). In vitro, CAT CAR T cells showed higher level of cytotoxicity than FMC63 CAR T cells towards cancer cell lines that expressed high level of CD19 antigen and a similar level of cytotoxicity towards cell lines expressing low density of CD19. The CAT CAR T cells also showed higher level of proliferation and TNF-alpha secretion compared to FMC63 CAR T cells when were co-cultured with CD19 expressing cells. In pre-clinical experiments in NOD SCID gamma (NSG) mice, CAT CAR T cells led to tumour regression, whereas FMC63 cells decreased tumour growth but did not lead to tumour regression.

In the clinical trial, 17 patients were enrolled in total. Fourteen patients were successfully treated, and one patient passed away after infusion while in remission due to sepsis (for 3 patients the product could not be generated). The safety profile was

overall high with 13 out of 14 patients developing mild cytokine release syndrome and mild neurotoxicity. Cytopenias were common in patients, which can be attributed to lymphodepletion and prior treatments (with one patient developing multiple infections that led to sepsis and death). At day 90 post infusion 12 out of 14 patients achieved clinical remission, with 6 patients relapsing at some point after remission – 5 patients relapsed with CD19- disease and 1 patient relapsed with CD19+ disease. In the case of CD19+ relapse, anti-CAR T cells responses were detected in the patient. Overall survival was 84% at 6 months and 63% at 12 months. Event free survival was 63% at 6 months and 46% at 12 months.

Apart from safety and efficacy of new CAT CAR T cells, investigators also studied their kinetics after infusion and overtime. Most of the CAR T cells in the infused product exhibited a central memory and TSCM phenotype. Twelve out of fourteen patients had a robust expansion of CAT CAR T cells at around 14 days post infusion, with a median of 41% of circulating CAR T cells at that time. Overtime, the proportion of CAR T cells decreased substantially in patients but 11 out 14 patients still had detectable CAR T cells at the last follow up (up to 24 months for some patients), with median duration of persistence of 215 days (range, 14 - 728 days) at data cut off and the median half-life of CAR T cells was 34 days (range, 3 - 102 days). The peak expansion time, persistence and half-life were all higher than published data for other approved CAR T cell products.

1.15 Using VISA to study persistence of CAR T cells in CARPALL patients

In this project, we decided to use an extensive immunophenotyping combined with our VISA pipeline to study the kinetics of phenotypic changes in CAR T cells and to investigate the origin of the long-term circulating CAR T cells in two CARPALL patients with long term persistence compared to two patients whose CAR T cells disappeared shortly after infusion. An additional objective was to isolate CAR T cells from PB at different time points and track integration sites found in CAR T cells from PB back to the infused product, to identify which T cell subset that shares the highest number of identical integration sites with long-term circulating CAR T cells, and therefore being the subpopulation that contributed the most to the generation of long-term persisting CAR T cells.

2 Results Part I – Establishment of the VISA technology at UCL

The first part of my PhD project was to establish VISA technology in our lab. For this PhD project all samples were analysed using LAM-PCR combined with high throughput NGS technology. For future work and clinical GT patient monitoring, we also established non-restrictive (nr)LM-PCR technology.

2.1 Introduction to VISA technologies

VISA is a combination of techniques that, firstly, allow the retrieval of viral vector integrations followed by the identification of precise genomic locations of retrieved integrations in the genome of the host. There are various approaches that can be used for VISA. The earliest technologies of this kind included linker ligation mediated (LM) PCRs and inverse PCRs (Mueller and Wold, 1989) (Silver and Keerikatte, 1989). Both typically involved enzymatic digestion of DNA to create sticky ends, which then were followed by either linker ligation mediated PCR (linker cassettes, complementary to sticky ends of enzymatically digested fragments, are allowed to ligate so primers specific for vector sequences and linker cassettes can be used to start amplification by PCR) or inverse PCR (enzymatically digested fragments are allowed to self-ligate to form circular DNA template, which then is amplified by PCR), followed by shotgun cloning of the vector-genome amplicons into bacteria and Sanger sequencing of the positive colonies (Aiuti *et al.*, 2007) (Cattoglio *et al.*, 2007). Both methods were useful to detect major clonal imbalances, however, were time consuming, typically allowing processing of only hundreds of integrations per sample, and expensive. Later, a more

efficient (up to several thousand integrations per patient) and less expensive technology has been developed that became the method of choice for this PhD project – linear amplification mediated (LAM) PCR (Schmidt *et al.*, 2007). Initially, some groups combined LAM-PCR with 454 Roche pyrosequencing (Brugman *et al.*, 2013), however later the preferred choice of sequencing platform, including for this project, became Illumina next-generation sequencing (NGS) (Blanco *et al.*, 1989). In addition to being significantly more sensitive and cheaper compared to Sanger sequencing, NGS allows identification of the relative abundance of each integration site in the sample by means of sequence reads, which can be used to estimate a relative clonal size (Biasco *et al.*, 2018). A detailed protocol containing all steps and primer sequences that were used in this project together with statistical analysis can be found in Chapter 6, sections 6.4 and 6.6. Other VISA methods include non-restrictive (nr)LAM-PCR, nrLM-PCR, and transposase-based methods, all discussed later in this chapter (section 2.5.1) (Gabriel *et al.*, 2009) (Paruzynski *et al.*, 2010) (Sherman *et al.*, 2017) (Brady *et al.*, 2011) (Wang *et al.*, 2016).

2.2   Overview of VISA protocols for different types of GT vectors

The main goal of clinical GT is to treat diseases caused by defective gene variants by delivering a new functional copy of the defective gene into target cells. To have a long-term effect, therapeutic genes have to be maintained in the nucleus of target cells, replicated and passed on through cell divisions. The delivery of therapeutic genes is performed by carriers called vectors. These vectors can be used as vehicles to deliver a functional copy of the gene of interest into the patient cells for therapeutic purposes. There are two types of GT vectors that can be used to deliver therapeutic genes into target cells – viral and non-viral vectors (Nayerossadat, Ali and Maedeh, 2012).

Examples of non-viral vectors include polyethylenimine/DNA complexes and transferrin-polyethylene glycol-polyethylenimine/DNA complexes (Kursa *et al.*, 2003) (Kircheis *et al.*, 2001). Viral vectors in turn can be further subdivided into integrating and non-integrating viral vectors. Examples of non-integrating vectors include vectors based on adenoviruses and adeno-associated viruses (Nayerossadat, Ali and Maedeh, 2012) (Wold and Toth, 2014) (Naso *et al.*, 2017). Integrating viral vectors are modified viruses that are depleted from genes that allow them to cause infection, including genes that code for structural, enzymatic and envelope proteins, but that are still able to enter cells and integrate inside their genomes resulting in stable, permanent expression of therapeutic genes. VISA allows us to take advantage of this permanent host cell modification and study haematopoiesis in these patients using individual vector integration sites as markers. Because all progeny of each individual transduced CD34+ cell will have identical IS we can use VISA to better understand relationships between cells in the immune system, to track clonal dynamics overtime and to monitor the emergence of clonal dominance in patients for safety reason (Figure 2.2.1). Examples of integrating vectors include retroviral vectors used to treat patients in first generation SCID-X1 (gammaretroviral vector) and CARPALL (lentiviral vector) GT trials (Cavazzana-Calvo *et al.*, 2000; Ghorashian *et al.*, 2019) (Milone and O'Doherty, 2018) (Maetzig *et al.*, 2011)

In our first GT trial, gammaretroviral (GR) vectors were used to treat the X-linked severe combined immunodeficiency disorder (SCID-X1). The vector that was used to transduce CD34+ cells, isolated from SCID-X1 GT patients, was derived from Murine Leukaemia Virus (MLV) and carried a common chain γ-subunit of interleukin-2 receptor-γ gene under the MLV long terminal repeat (LTR) promoter/enhancer

(Cavazzana-Calvo *et al.*, 2000) (Gaspar *et al.*, 2004) (Hacein-Bey-Abina *et al.*, 2010). Therefore, the first protocol that we needed to set up was LAM-PCR specific for the retrieval of IS from MLV-based GR vectors. The second protocol was set up to perform VISA for CARPALL samples. PBMCs from CARPALL GT patients, were transduced with third generation self-inactivating lentiviral vectors (derived from human immunodeficiency virus (HIV) type 1) encoding αCD19CAT-41BBz cassette under the control of the human phosphoglycerate kinase (PGK) promoter (Ghorashian *et al.*, 2019) (Dull *et al.*, 1998).

Outside of this PhD project, for GT safety monitoring purposes, we also put in place a protocol for IS detection in patients from an adenosine deaminase (ADA) deficiency GT trial, in which a gammaretroviral vector encoding ADA complimentary DNA (cDNA) was used under the control of spleen focus forming virus (SFFV) LTR (Gaspar, Cooray, Gilmour, Parsley, Zhang, *et al.*, 2011) (Gaspar *et al.*, 2006). Lastly, we also established a new more sensitive and faster protocol for VISA using non-restrictive linker mediated PCR (nrLM-PCR) technology.
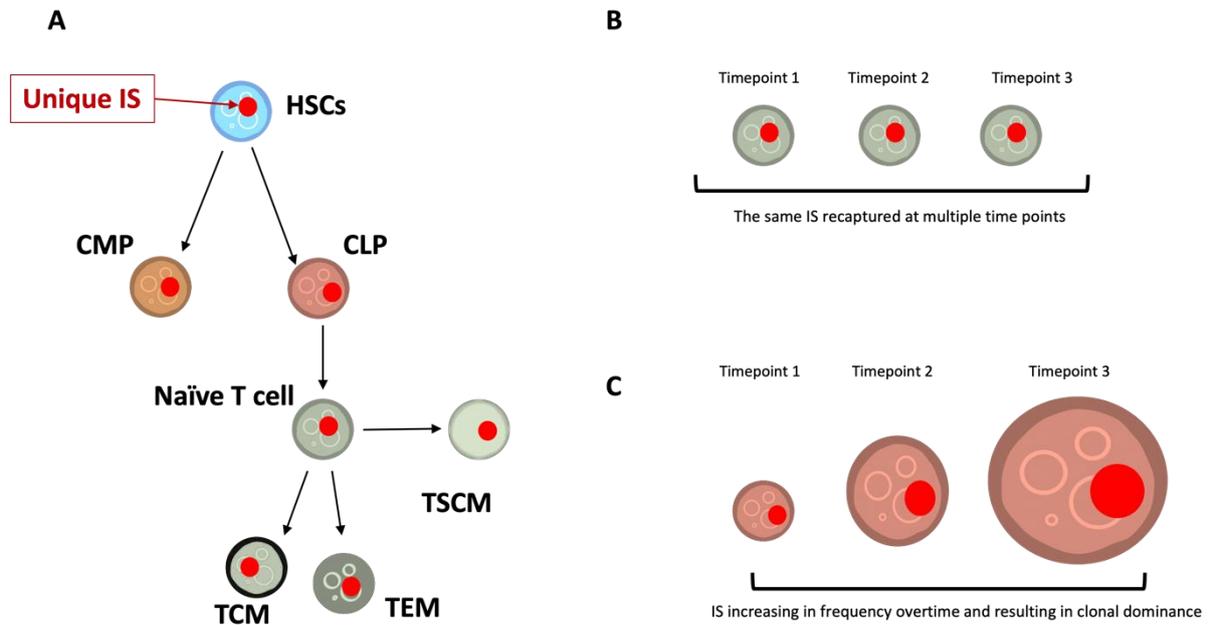
Figure 2.2.1 VISA can be used to study relationships between cells and to track clonal dynamics in humans.

a) As GT vectors integrate into genomes of stem cells, they mark them permanently with unique IS. This IS then gets passed on to all progeny of the cell, allowing us to track the differentiation pathway of stem and progenitor cells as well as the relationship between different immune cells. **b)** By tracking unique IS at different timepoints we can study persistence of different clones in patients over time. **c)** By tracking unique or potentially oncogenic IS overtime, we can spot the emergence of clonal dominance in GT patients, and in cases when it is suspected that it can lead to cancer, take action (performed as part of standard clinical monitoring of GT patients).

2.3 Overcoming low DNA concentration issue in patient samples

The LAM-PCR procedure starts with the extraction of DNA from samples. One of the frequent problems when working with patient samples is low DNA yields after extraction. Some of the samples that we used for the analysis in this project were cryopreserved 10 years ago, which severely affected their quality and the amount of

DNA we were able to extract. The starting DNA concentration required for high-quality VISA depends on the VCN of the sample, with fully transduced (100%) polyclonal samples needing 1-10 ng of DNA, whereas samples with 1-0.1% of transduced cells would need a 100 ng of starting DNA material for sufficient quality results (Schmidt *et al.*, 2007). This is because the larger the number of cells that carries integrated vectors the more likely it is for us to retrieve these integrations. If only very few cells are vector positive, they will get diluted in the untraduced population of cells and we would require significantly more DNA to be able to retrieve these integrations. Considering the age and subsequently the quality of some of our SCID-X1 samples as well as the very low frequency of transduced cells in our CARPALL samples at late time points (0.1%), we estimated that we would need at least 100 ng of DNA per sample per restriction enzyme (as discussed later in section 2.4.2, in this project we used 3 restriction enzymes to digest DNA, meaning 300 ng of DNA was required to start with) for high-quality results. In some cases, we had less than 10 ng of DNA extracted in total, which made it impossible to analyse samples. For this reason, we decided to use a whole genome amplification (WGA) technique that allows significant amplification from minimal starting concentrations of genomic DNA. We used a Qiagen REPLI-G kit, that utilises multiple displacement amplification method to amplify patient samples (See Chapter 6, section 6.3.1 for details). This kit uses Phi29 DNA polymerase (an enzyme isolated from the bacteriophage Phi29) that is able to replicate up to 70 kb of DNA without dissociating from it. The Phi29 polymerase has a 3'->5' exonuclease proofreading activity and with its high fidelity reduces the error rate to 1 in $10^6 - 10^7$ bases compared to classical *Taq* polymerase (1 in 9,000 bases) (Tindall and Kunkel, 1988) (Paez *et al.*, 2004). Overall, we were able to achieve significant amplification of genomic DNA isolated from patient samples using this method. To verify that

amplification is genuine and consistent we compared vector copy numbers in gene therapy patient samples using digital PCR in non-amplified genomic DNA and WGA DNA (Table 2.3.1) (See Methods Chapter 6, section 6.9 for detailed digital PCR protocol). As table 2.3.1 shows, in some WGA samples VCN was lower than in non-amplified samples. This suggested that we should expect to lose some integration sites in the process. It has previously been reported that during multiple displacement amplification process, some genomic regions are amplified better than others resulting in uneven coverage (Lage *et al.*, 2003). However, knowing that without the WGA we would not be able to start with required DNA concentrations and as a result would affect the quality of VISA, we decided to proceed by applying WGA to all samples, regardless of starting DNA concentrations, for normalisation and consistency of data.

Sample name	VCN (Genomic DNA) copies/cell	VCN (After WGA) copies/cell
PT5 TN	1.7	1.5
PT5 TCM	2.1	1.6
PT5 TEM	2.3	1.7
PT1 TN	2.6	1.8
PT1 TCM	3.6	1.0

Sample name	VCN (Genomic DNA) copies/cell	VCN (After WGA) copies/cell
PT1 TEM	3.6	2.2

Table 2.3.1 VCN comparison between genomic DNA and WGA DNA.

VCN in genomic DNA samples (2nd column) or VCN in samples after WGA (3rd column) from 2 patients (PT1 and PT5) in 3 cell subsets TN, TCM and TEM measured by digital PCR. Our digital PCR VCN assay allows us to calculate an average copy number of the vector sequences per cell relative to the reference gene (in this case ApoB gene) in the population of cells analysed (relative quantification).

2.4 LAM-PCR protocol

Overall, all steps for LAM-PCR protocol are the same for all type of vectors and protocols only differ in primer sequences, that are specific for each vector (see Chapter 6, section 6.4 for primer sequences for CARPALL and SCID-X1, see Methods in (Gaspar *et al.*, 2006) for primer sequences for GRV-ADA) (Figure 2.4.1).
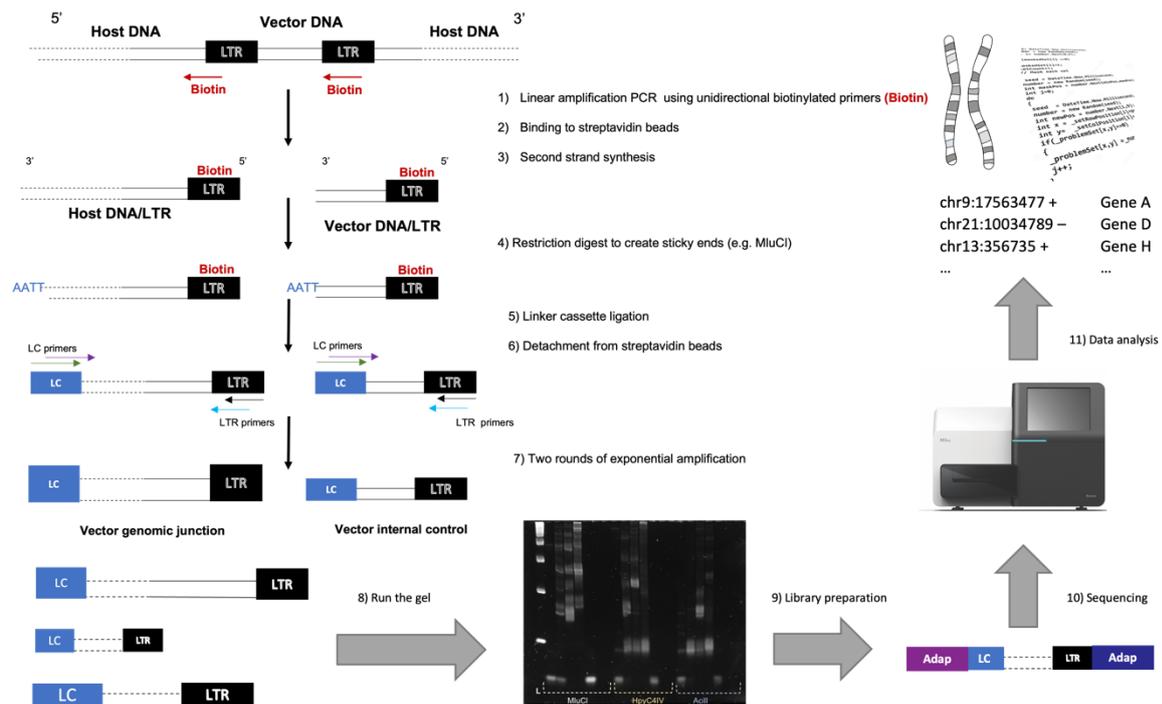
Figure 2.4.1 Schematic representation of LAM-PCR process.

LAM-PCR procedure starts with linear amplification of vector-genome junctions (Host DNA/LTR) using biotinylated primers (1), followed by binding to streptavidin beads (2) and second strand synthesis (3). Next, double stranded DNA is digested with 3 different enzymes (4) to generate sticky ends. This is followed by linker cassette (LC) ligation (5) and detachment from streptavidin beads (6). Two rounds of exponential PCRs are then performed with primers complimentary to LC and vector LTRs (7). Two nested PCR generate amplicons of various sizes that then can be visualised on the gel (8). This is then followed by library preparation, which involves Illumina specific adapter (Adap) ligation to LC and LTRs (9) and subsequent Illumina NGS (10). Finally, raw data is analysed to determine precise genomic locations of retrieved IS (11).

2.4.1 Linear PCR and binding to beads

After suitable quantities of DNA are obtained, the next step in the procedure, is the enrichment of target sequences by linear amplification PCR. When low levels of target sequences are present in the sample or when starting DNA concentrations are low,

forward and reverse primers that are used to amplify target DNA might not evenly hybridise to the entire DNA region of interest, resulting in some alleles being unamplified and as a result undetected (Grisedale and van Daal, 2014). Linear amplification PCR allows amplification of a single strand of DNA using one type of primer (either forward or reverse) resulting in the enrichment of target sequences with a reduced level of stochastic PCR effects (random variation) including the loss of alleles (Grisedale and van Daal, 2014). Biotinylated primers are used to isolate and concentrate target sequences for downstream processing.

We performed 100 cycles of linear amplification PCR in total using two different biotinylated primers (biotin modification was added at the 5'-terminus of primers to avoid interference with the extension of the growing strand) specific for MLV LTRs (SCID-X1) and lentiviral LTRs (CARPALL) to improve efficiency of target sequence amplification (see Chapter 6, section 6.4). For gammaretroviral ADA samples, we used only one type of SFFV LTR specific biotinylated primer and added double the amount instead. In all cases primers were annealing to LTRs in the orientation out towards genomic DNA junctions.

After linear PCR, samples were incubated overnight using streptavidin coupled magnetic beads (Invitrogen Dynabeads Kilobase Binder Kit) to ensure appropriate binding of target sequences to beads (streptavidin is a protein isolated from the bacterium *Streptomyces avidinii* and has a very high affinity for biotin (Weber *et al.*, 1989)). Next steps involved synthesis of second complimentary DNA strands by random hexanucleotide priming using Klenow enzyme to generate double stranded DNA molecules.

2.4.2 Restriction enzyme digest

At this point we had double stranded DNA fragments with one end having a known vector LTR sequence attached to the unknown genomic DNA sequence (Figure 2.4.1). To be able to amplify and subsequently sequence these fragments, we needed to attach another known sequence to the opposite end of the fragment. One approach is to create sticky ends with known sequences by digesting DNA with restriction enzymes, and to attach a linker cassette with known sequence to these ends to generate DNA fragments that will have a vector LTR on one end and a linker cassette on another end.

The type and the number of restriction enzymes used is highly important because it directly influences the extent of genome coverage and subsequently the sensitivity of LAM-PCR (Gabriel *et al.*, 2009). Restriction enzymes are able to digest the DNA at certain regions that they specifically recognise (restriction sites) (Roberts and Murray, 1976) (Kessler and Manta, 1990). These enzymes are naturally found in bacteria and archaea where they play a protective role against foreign DNA (e.g., phage DNA) (Kruger and Bickle, 1983). There are five different groups of restriction enzymes that differ structurally and in the way they recognise their target DNA sequences (Loenen *et al.*, 2014). The length of target sequences they recognise is typically between four to eight base pairs. The shorter the DNA fragments that are generated during restriction digest the better they are amplified during exponential PCRs (Gabriel *et al.*, 2009) (Bystrykh *et al.*, 2012). Therefore, efficient viral integration site retrieval requires restriction enzymes whose recognition sites are widespread in the genome and located within "amplifiable distance". Integration site collection is technically limited because precise IS locations can only be identified in the "analysable genome" – completely

aligned DNA sequences that represent 93% of the human genome (Lander *et al.*, 2001) (Bovee *et al.*, 2008) (Abdellah *et al.*, 2004). The analysable genome is then further restricted by an "amplifiable genome" – a 1000 bp region around all restriction sites that can be efficiently amplified by exponential PCR (Kent, 2002). Lastly, the amplifiable genome is limited by "mappable genome" – which is the collection of sequences that can be clearly aligned to the human genome. The overlap between these three genome access limitations is termed "accessible genome" and represents the portion of the genome where IS can be collected with enough precision.

It has been previously reported that six to eight base pair sequences that specific restriction enzymes can recognise are present in the human genome at low frequencies and that their use for VISA should be avoided (Gabriel *et al.*, 2009) (Harkey *et al.*, 2007) (Bystrykh *et al.*, 2012). In comparison, the frequency of four base pair cutter restriction enzymes is much higher, with some of them predicted to cover up to 85% of the accessible genome (Hayakawa *et al.*, 2009) (Bauer *et al.*, 2006) (Schmidt *et al.*, 2001) (Schwarzwaelder *et al.*, 2007). Another study reported that to cover 50% of the genome for VISA, at least two different restriction enzymes should be used with 4 base pair, no-CG (C -cytosine, G – guanine) recognition motifs, and that combining five the most potent no-CG motif cutting enzymes results in about 90% of genome coverage (Gabriel *et al.*, 2009). This is because integration sites containing CG motifs are less frequent in the human genome compared to no-CG motifs. However, the final combination of restriction enzymes should be based not only on the efficiency and frequency of digestion of the particular enzyme but also it should take into account specific vector integration site preferences. For example, no-CG motif digesting enzymes are more efficient in terms of genome coverage, however CG-rich DNA

regions which are typically located in promoter regions and CpG islands - that are often preferential loci for retroviral integrations (Felice *et al.*, 2009) (Lewinski *et al.*, 2006) (Cattoglio *et al.*, 2007), would not be covered by no-CG cutting enzymes, which in case of GR transduced SCID-X1 samples would represent a big loss of data. In addition, it is essential to choose restriction enzymes that do not have any restriction sites within the vector LTR sequence, to make sure it stays intact.

Taking all of this into account, we decided to use three restriction enzymes in this project to ensure high genome coverage but maintain reasonable processing time because the more enzymes are used in the process the more time consuming, laborious, and expensive LAM-PCR procedure becomes. We combined MluCI - the no-CG motif frequent cutter enzyme (restriction site: /AATT), AciI - a CG motif rare cutter (restriction site: CCGC (-3/-1), and HpyCH4IV - another rare cutter enzyme (restriction site: A/CGT).

2.4.3 Linker cassette ligation and nested PCR

After restriction enzyme digest, and subsequent generation of sticky ends, barcoded linker cassettes were ligated to target DNA sequences to generate DNA fragments with known vector LTR sequence at one end and known linker cassette sequence at the opposite end. The barcode or unique molecular identifier (UMI) in the linker cassette is an important quality control addition to target DNA prior to exponential amplification steps (Leonardelli *et al.*, 2016) (Firouzi *et al.*, 2014) (McCloskey *et al.*, 2007). These barcodes, together with Illumina barcodes that are added later in the process (discussed in section 2.4.5), serve as quality controls by allowing to minimise the frequency of collisions (identical IS in unrelated samples due to possible

contamination) as well as to reduce false positive reads during data analysis post sequencing (Wang *et al.*, 2016) (Greene *et al.*, 2013). Barcodes also allow to control for random variation in amplification of some fragments due to general heterogeneity in PCR amplification efficiency (Best *et al.*, 2015).

Next, two rounds of amplification by nested PCR were performed to amplify target DNA fragments so they can reach appropriate concentrations for sequencing. Nested PCR is a technique that allows to increase the sensitivity of PCR reaction by performing two rounds of amplification with the second PCR round employing primers that would bind inside the sequence that the first pair of primers was specific for (Figure 2.4.1) (Green and Sambrook, 2019). We used primers specific for MLV vector LTR and linker cassettes specific for sticky ends generated by MluCI, AciI or HpyCH4IV restriction enzymes digest for SCID-X1 patients (see Chapter 6, section 6.4). For CARPALL patients, we used primers specific for HIV-1 LTR and linker cassettes specific for sticky ends generated by MluCI, AciI or HpyCH4IV restriction enzymes digest for SCID-X1 patients (see Chapter 6, section 6.4).

2.4.4 Gel electrophoresis

The next step was the visualisation of amplicons using gel electrophoresis. For our studies, we performed gel electrophoresis purely for quality purposes, to ensure that we can see amplicons on the gel, meaning that the reaction has worked, and samples are ready for library preparation and sequencing. Gel electrophoresis allows separation of DNA fragments based on their molecular size (Kozulić and Heimgartner, 1991; Kozulić, 1994) (Lee *et al.*, 2012). During gel electrophoresis DNA fragments are pushed by an electrical field through the gel containing small pores, with smaller

fragments migrating more easily through the gel and as a result travelling faster and further away compared to larger DNA fragments. One end the gel is positively charged whereas the opposite end has a negative charge. Because the phosphate backbones of DNA molecules are negatively charged, DNA will be pulled towards the positively charged end of the gel.

Initially, we used precast high-resolution Spreadex gel (ElchromScientific). The Spreadex gels are characterised by exclusion limit meaning that DNA fragments above a certain size would not be able to travel through the gel. Due to restriction enzyme digest we expected many different DNA fragments of various sizes to be generated and amplified during exponential PCRs. Therefore, we chose Spreadex gels with 1200 bp exclusion limit that allow the optimum separation of DNA fragments within 250-800 base pairs range and allow resolution of separate DNA fragments with difference in length of only 2 bp. High resolution of Spreadex gels is achieved by polymers of the gel forming a unique 3-D structure that meets migrating DNA fragments with increasing resistance (Hourihan, O'Sullivan and Morgan, 2001). Selective delay in migration of longer DNA fragments results in increased spacing between fragments and visible spread of different fragments on the gel. Figure 2.4.2 shows an example of a Spreadex gel after running SCID-X1 samples (top left) and a gel visualising DNA fragment from CARPALL samples (top right). In both cases samples looked highly polyclonal (presence of many different IS) on the gel, which later was confirmed by sequencing (see Chapter 3, section 3.4 and Chapter 4, section 4.2 for IS analysis). To be able to run samples on Spreadex gels, the lab requires special electrophoresis apparatus such as ElchromScientific SEA 2000. The run of one gel typically takes about 1,5 - 2 hours and overall, the process is expensive and long. Therefore, after we were sure that the

protocol is working, we switched to running agarose gel for this step instead of Spreadex gels (Figure 2.4.2 bottom). In comparison, agarose gels are less sensitive when it comes to resolution of different bands but are cheap and can be run for 40 minutes for sufficient visualisation of amplified bands for quality control purposes only (as in our case).
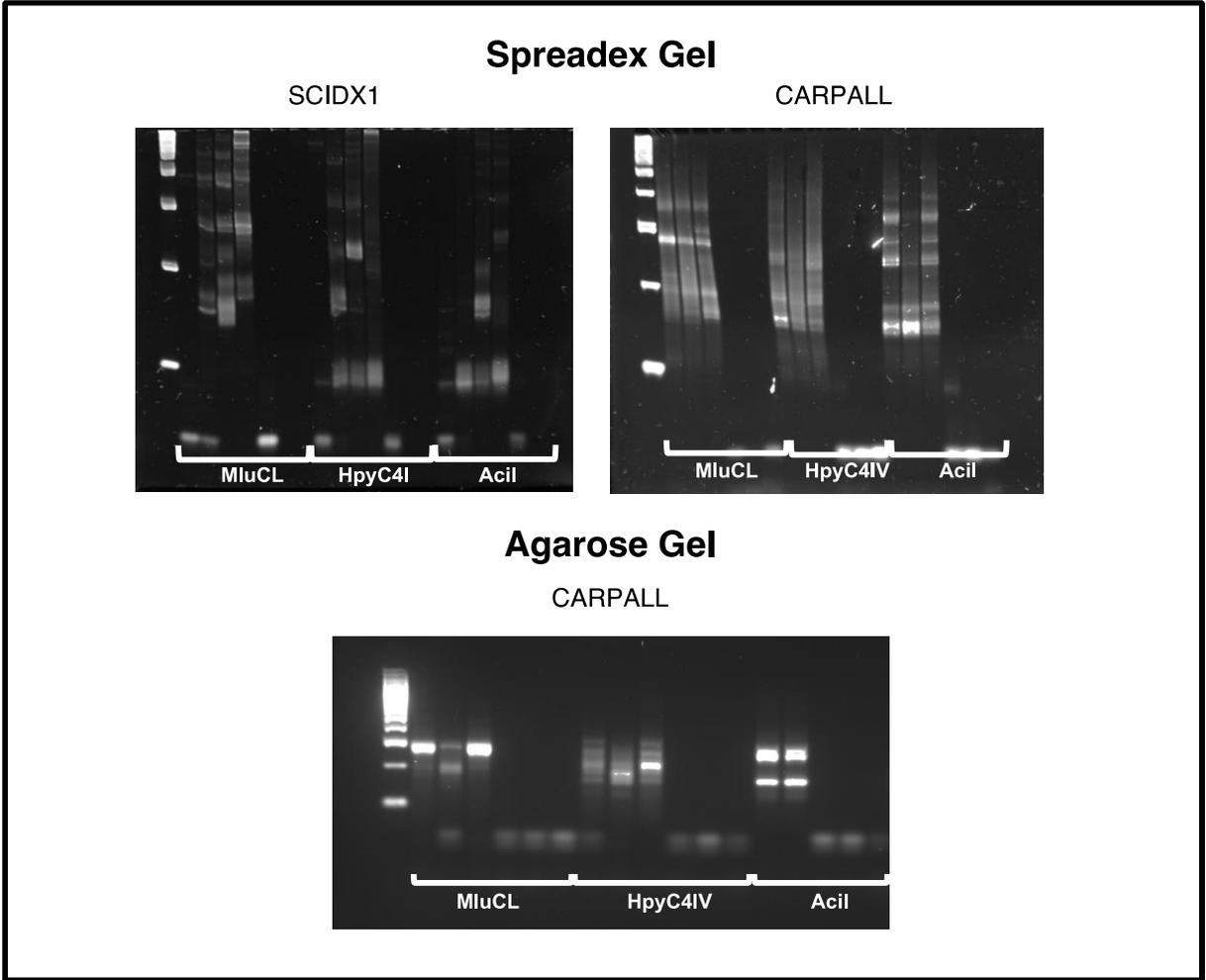
Figure 2.4.2 Example of Spreadex (top) gels for SCID-X1 and CARPALL samples and agarose (bottom) gel for CARPALL samples.

The same layout was always used to run all gels for all experiments for both SCID-X1 and CARPALL patients: first samples digested with MluCL enzyme plus 3 water controls, second samples digested

with HpyC4IV enzyme plus 3 water controls, third samples digested with AciI enzyme plus 3 water controls.

2.4.5 Library preparation and sequencing

Once we confirmed that our experiment has worked and we successfully amplified vector integrations, we were able to start preparing the library. For this project, we used an Illumina NGS platform. The Illumina sequencing platform is a type of next generation sequencing technology known as sequencing by synthesis (SBS) (Slatko, Gardner and Ausubel, 2018). To be suitable for sequencing with the Illumina NGS platform (MiSeq in our case) samples have to undergo additional round of PCR amplification so Illumina adapters can be added (See Chapter 6, Methods, Section 6.4.2 for adapter sequences). In this project we used adapters that were specific for either the vector LTR sequence of the DNA fragment or the linker cassettes. They all contained 8 base pair unique barcodes (allowing many different samples to be pooled together for sequencing), sequences specific for adaptors on the Illumina flow cell, and a 12-base pair random sequence in between to ensure efficient cluster recognition upon sequencing. Once adaptors were added, samples were purified, quantified, and pooled together for sequencing on MiSeq platform. Overall, Illumina sequencing is based on the technique called bridge amplification in which DNA fragments with added Illumina adapters attach to the complimentary oligos on the surface of the flow cell in the sequencer and each molecule serves as a substrate for repeated rounds of amplification (Slatko, Gardner and Ausubel, 2018). This results in the generation of clonal clusters on the flow cell, with about 1000 copies of each fragment. Each slide can support millions of parallel reactions. The technology is referred to as sequencing by synthesis because during synthesis of each fragment, fluorescently labelled

nucleotides emit light upon incorporation to the growing chain and the emission that they produce is detected and recorded by the sequencer. The raw data was then loaded on a BaseSpace website for us to analyse.

The pipeline for the bioinformatical analysis of raw data has been developed in Dr Biasco's lab (previously Harvard Medical School, now AVROBIO and UCL) and has been performed by either Dr Biasco or Dr Pellin (Harvard Medical School). In general, sequences undergo data quality control, followed by barcode assignment to samples, trimming of vector and linker cassette sequences and mapping final sequences to the human genome (Biasco *et al.*, 2018) (Leonardelli *et al.*, 2016). The final result is the precise chromosomal location of each IS in the genome that is normally referred to and visualised as the name of the closest gene in the proximity of this particular IS (Figure 2.4.1).

2.5 nrLM – PCR protocol for future VISA in GT patients and clonal tracking studies

2.5.1 Overview of no restriction digest IS retrieval technologies

Because LAM-PCR technology relies on restriction enzyme digest, this method is inherently biased and makes it impossible to detect integrations that occur either too close or too far from restriction sites recognised by these enzymes, resulting in DNA fragments that are either too short to be resolved or too long to be successfully amplified (Gabriel *et al.*, 2009). Therefore, various groups over the years developed protocols for IS retrieval that avoid the use of restriction enzymes (Gabriel *et al.*, 2009) (Paruzynski *et al.*, 2010) (Brady *et al.*, 2011) (Sherman *et al.*, 2017) (Wang *et al.*,

2016). One of the methods is no restriction digest (nr)LAM-PCR. This method starts with linear amplification PCR to enrich for vector-genome junctions (identical to classical LAM-PCR). However, instead of conversion of linearly amplified single stranded DNA (ssDNA) (after bead enrichment) into double stranded DNA (dsDNA) prior to restriction digest (classical LAM-PCR), in nrLAM-PCR protocol, ssDNA is ligated to a linker and then amplified by nested PCRs followed by high-throughput NGS analysis. However, this method has its limitations due to heterogeneity in efficiency of linker ligation to ssDNA and, therefore, it is recommended to use conventional LAM-PCR in case of limited amount of starting DNA material available (Paruzynski *et al.*, 2010). A different approach of no restriction digest IS retrieval, involves shearing DNA by sonication in sequence independent manner, thereby creating many DNA fragments of variable sizes. This is followed by end repair rection to create 3' end DNA overhangs, that then can be ligated with barcoded linker cassettes. Resulting fragments are then amplified with nested PCRs and sequenced using NGS (Sherman *et al.*, 2017). Another method for restriction enzyme independent IS retrieval is based on phage Mu transposition (Brady *et al.*, 2011). In this method, bacterial transposase Mua is used to add adapters to DNA for tagging that then can be amplified using PCR. This method is significantly quicker than any other IS retrieval technology, however it is also much less sensitive.

Recently, we have also established a restriction digestion-free protocol in our lab (Sherman *et al.*, 2017). Figure 2.5.1 shows schematic representation of nrLM-PCR established in our lab.
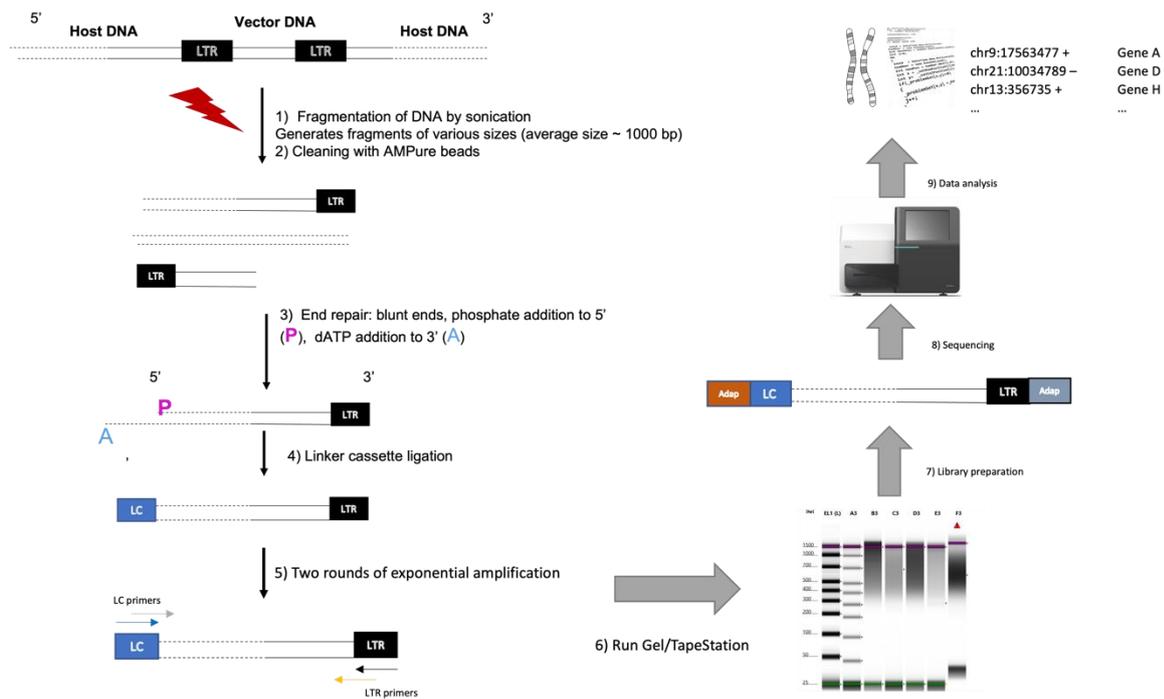
Figure 2.5.1 Schematic representation of nrLM-PCR process.

nrLM-PCR procedure starts with shearing of DNA by sonication (1), followed by cleaning of DNA with AMPure beads (2) and end repair (3). Next, end prepped DNA with blunt ends is ligated with linker cassettes (LC) (4). This is followed by two rounds of exponential PCR amplification with primers complimentary to LC and vector LTRs (5). Two nested PCR generate amplicons of various sizes that then can be visualised on the gel or using Tape Station (6). This is then followed by library preparation, which involves Illumina specific adapter (Adap) ligation to LC and LTRs (7) and subsequent Illumina NGS (8). Finally, raw data is analysed to determine precise genomic locations of retrieved IS (9).

2.5.2 nrLM-PCR protocol. DNA sonication and Tape Station

Sonication is one of the methods for DNA fragmentation. During sonication, ultra-sonic waves produced by ultrasonicators (Bioraptor Pico, Diagenode in our case) disrupt DNA strands through cavitation (Knierim *et al.*, 2011). During this process air bubbles are created and subsequently destroyed, which results in evenly sized, smaller DNA fragments. Sonication results in unbiased shearing of DNA because shearing is not

restricted to particular sites in the genome. The size of DNA fragments after sonication depends on the time of exposure of samples to ultrasonic waves and the number of cycles. For our application, we wanted final fragmented DNA to be around 1000 base pairs. We used sonicator's setting recommended by manufacturing company (Diagenode) for Bioraptor Pico to achieve DNA fragments of desired size (See chapter 6, Methods, section 6.7 for detailed protocol). Because of the previous experience with low DNA concentrations in precious patients samples, we knew that it is likely that we would need to use WGA again in the future. Therefore, we wanted to see how fragmentation by sonication is going to affect both genomic DNA samples as well as DNA samples after WGA (Figure 2.5.2). After sonication, DNA samples were purified using AMPure XP beads and fragments were visualised using Tape Station technology (automated electrophoresis system, Agilent Tape station 4200, see Chapter 6, section 6.7.3 for details). Our test demonstrated that in both samples, DNA fragment size peaks were very similar and close to desired 1000 base pairs. Therefore, we concluded that these settings could be safely applied to both types of samples.
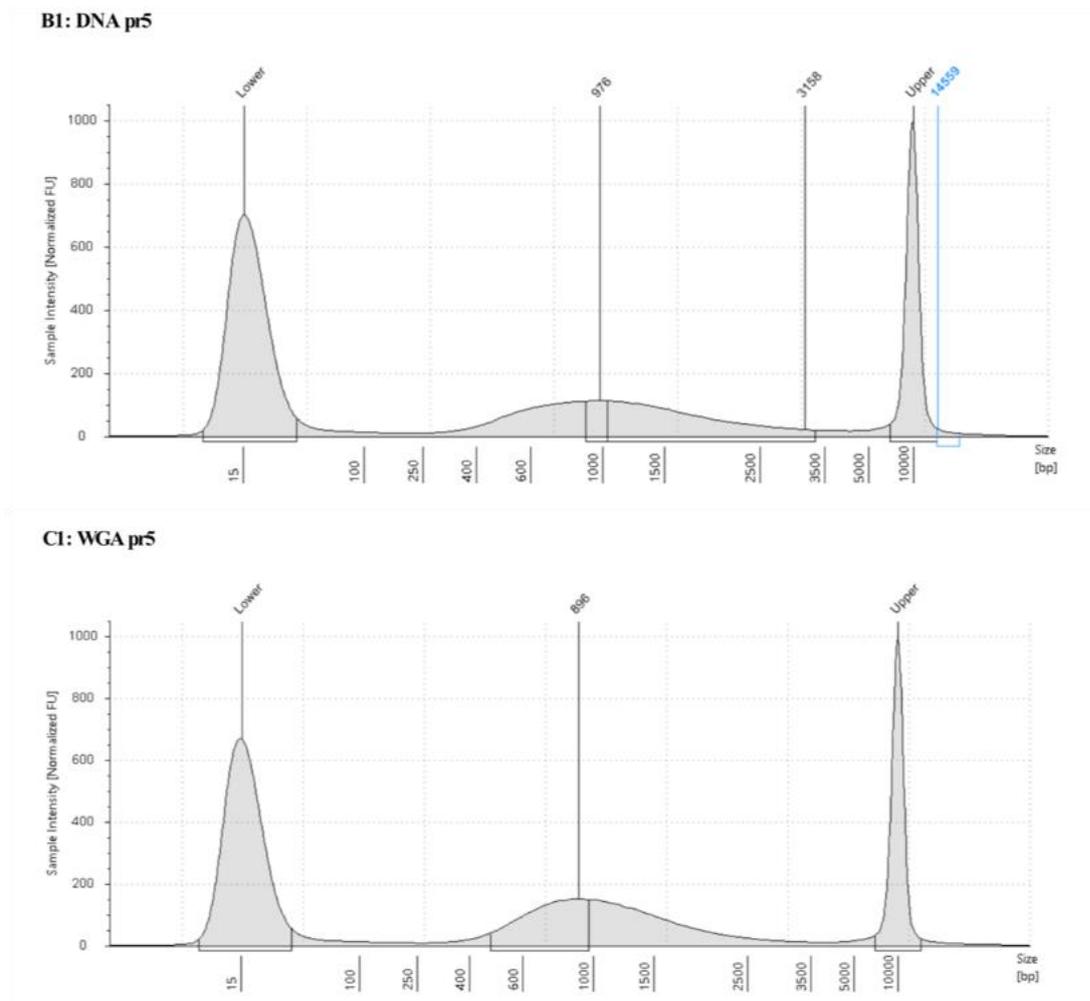
Figure 2.5.2 Tape Station electropherogram report for unmanipulated DNA (top, DNA pr5) and WGA DNA (bottom, WGA pr5) samples.

Reports show DNA fragment size after sonication. Both reports were generated using 4200 Tapestation system. The 4200 Tapestation system detects fluorescently stained double stranded DNA. Sample Intensity (y-axis) allows to identify discrete peaks of a certain size in samples (the higher the intensity the more DNA fragments of this size are in the sample). X-axis represents the ladder (D5000 High Sensitivity ladder in this case).

2.5.3 End preparation of fragmented DNA, linker ligation and PCR amplification

After sonication DNA fragments are often non-blunted and lack 5' end phosphate groups (Montanera *et al.*, 2020). Therefore, for subsequent successful linker ligation prior to PCR amplification, fragmented double stranded DNA needs to undergo an end repair reaction that typically involves three main enzymatic steps: DNA polymerase I reaction that creates blunt ended DNA, kinase reactions add phosphate groups to 5' end DNA, and a Klenow fragment reaction which adds dATP to the 3' end of end prepped DNA (see Chapter 6, section 6.7 for details). After DNA is finally prepared, it is ready for linker ligation and subsequent PCR amplification. For two rounds of nested PCR, we used the same primers as for LAM-PCR but at different concentrations. After PCR amplification, fragments were visualised using agarose gel. This was followed by library preparation for Illumina sequencing and high-throughput NGS.

3 Results part II – Using VISA to study T and NK cells in SCID-X1 GT patients

The second part of my PhD project was designed to characterise circulating naïve T and NK cells in SCID-X1 patients as well as to use VISA to better understand their relationship to one another, to track clonal dynamics in these patients overtime and to study long-term effects of vector integrations in the close proximity of known oncogenes.

3.1 Immunophenotypic and molecular follow up of SCID-X1 GT patients

As part of the normal follow up of our SCID-X1 patients, immunophenotyping and vector copy number (VCN) in different blood cell lineages were analysed at several time points post treatment (Figure 3.1.1). Peripheral blood mononuclear cells (PBMCs) from each patient were sorted into five different lineages (neutrophils, T cells, B cells, NK cells and monocytes) using fluorescence activated cell sorting (FACS). Quantitative PCR demonstrated that the average copy number for T and NK cells was 2.1 and 2.3 copies per cell, respectively, while in B cells and myeloid cells VCN remained at the background level since 22-27 months after GT when we started these measurements (Figure 3.1.2).
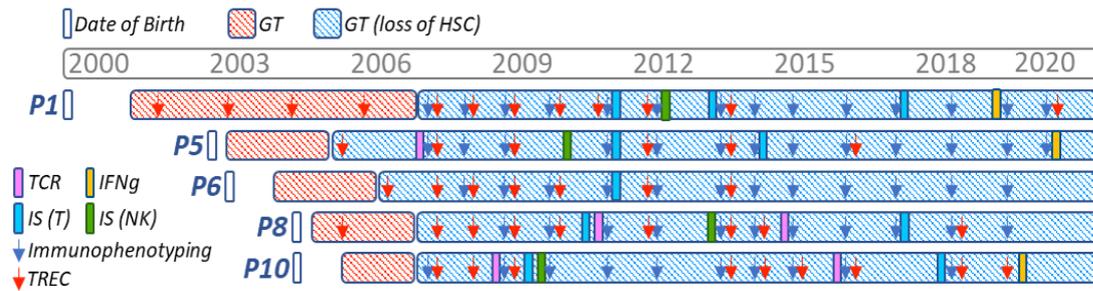
Figure 3.1.1 Gantt chart showing all assays that were performed for each patient in this study across different time points.

TCR – T cell receptor sequencing (pink rectangle), IFNg – interferon gamma secretion assay (orange rectangle), IS (T) – integration site analysis in T cells (blue rectangle), IS (NK) – integration site analysis in NK cells (green rectangle), Immunophenotyping is shown by blue arrows, TREC - T cell receptor excision circle content (red arrow). Date of birth of each patient is shown by white rectangles, GT in red – time post GT when we assume patients might still have had gene corrected HSCs (in case of successful engraftment) in the BM; GT in blue – approximate time post GT when patients presumably lost their engrafted gene corrected HSCs in BM as suggested by the absence of transgene in myeloid cells and B cells demonstrated by VCN analysis at multiple time points.
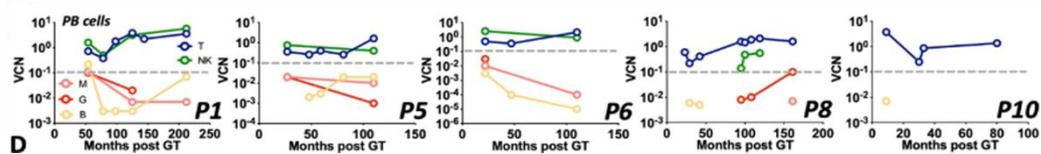
Figure 3.1.2 VCN in P1, P5, P6, P8 and P10 measured at various time points post GT in different cell lineages.

T cells (T) – dark blue, NK cells (NK) – green, B cells (B) – yellow, Monocytes (M) – orange, Granulocytes (G) – Red. Grey dotted line represents a detection threshold for VCN analysis.

Circulating myeloid and naïve B cells are relatively short lived and maintenance of these cells is dependent on a continuous output from the bone marrow (Pillay *et al.*, 2010) (Sprent and Basten, 1973; Hettinger *et al.*, 2013; Jakubzick *et al.*, 2013; Yona *et al.*, 2013; Jones, Wilmore and Allman, 2015). Therefore, absence of the transgene in myeloid cells and B cells is consistent with a permanent loss or failure of engraftment of gene corrected long-term (Lt)HSC.

Hence, we initially assumed that vector positive lymphocytes we still observe in our patients could be long-lived circulating lymphoid cells. To test this hypothesis, we investigated the dynamics of the lymphocyte reconstitution over time.

We firstly looked at absolute numbers of different immune cell types in peripheral blood including T cells (CD3+), B cells (CD19+) and NK cells (CD16/CD56+) (Figure 3.1.3 top). Overall, during the early immune reconstitution phase (within the first 6 months), all 5 patients showed CD3+ cell numbers within normal range, and these numbers remained at or just below the normal range at the latest follow up. CD4 and CD8 T

cells were present throughout the time of observation although an inverted CD4/CD8
T cell ratio was detected in all patients (Figure 3.1.3 bottom).
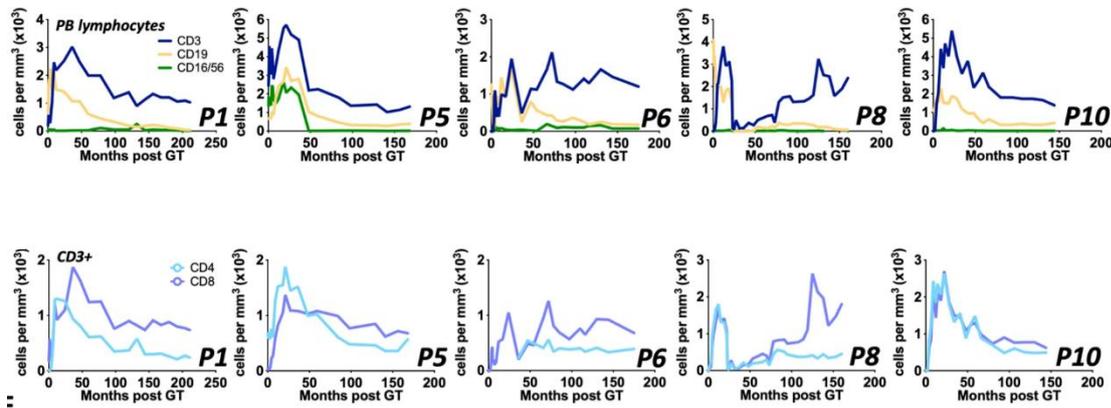
Figure 3.1.3 Absolute numbers of lymphocytes (top) and T cells (bottom) in patients overtime.

Top panel: Absolute numbers of lymphocytes in each patient overtime. CD3 – T cells (blue), CD19 – B cells (yellow), CD16/56 – NK cells (green). Bottom panel: Absolute numbers of CD4 and CD8 T cells in each patient overtime. CD4 T cells – blue, CD8 T cells – purple.

Initially standard immunophenotyping was performed as part of the patient monitoring protocol at regular intervals. In the beginning, we used CD62L and CD27 cell surface markers to identify a double positive (DP) mixed T cell precursor population composed of naïve T cells (TN) and stem cell memory T cells (TSCM), a central and effector memory T cells (TCM/TEM) population and effector T cells (TEFF) (Figure 3.1.4). In both CD4 and CD8 subsets, T cell precursors (CD3+CD62L+CD27+) were maintained over time up until the latest follow up in a relatively stable fashion in all patients except in patient 8 (P8). In this patient, T cell precursors were preserved up until month 100 post GT, but rapidly declined after this time point, until their relative percentage

dropped to less than 1% of total CD4/CD8 populations. P8 developed T - acute lymphoblastic leukaemia (T-ALL) at 24 months post GT and subsequently underwent a 3-year chemotherapy regimen, achieving long term clinical and molecular remission (Gaspar, Cooray, Gilmour, Parsley, Adams, *et al.*, 2011). Because chemotherapy can result in the long-term bone marrow (Hellman and Botnick, 1977; Testa, Hendry and Molineux, 1985; Neben *et al.*, 1993; Mauch *et al.*, 1995) (Reya, 2003) and thymic injuries (Perry, Jackson and Talmadge, 1994), affecting HSC and progenitor reserves and their self-renewal ability, we suspect that the burden of chemotherapy may well be a significant contributor to the observed sharp decline of T cell precursor populations in this particular patient at around 5 years post remission.
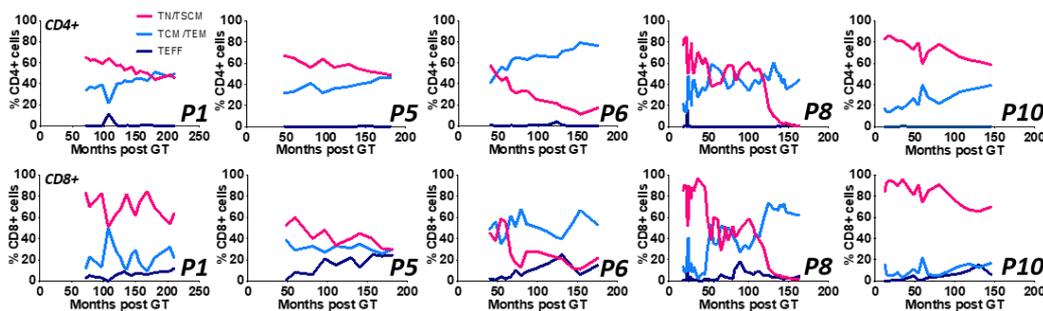
Figure 3.1.4 Frequency of T cell subsets overtime in SCID-X1 patients post GT.

Percentages of TN/TSCM (DP precursors; pink), TCM (light blue) and TEFF (dark blue) in CD4 (top panel) and CD8 (bottom panel) cells.

3.2 Immunophenotypic and functional characterisation of *bona fide* naïve T cells

Because we previously assumed that these patients lost (or never had) engrafted gene corrected (Lt)HSCs (indicated by the absence of marking in B cells and myeloid cells), and as a result were expected to have only mature memory lymphocytes in the circulation, the presence of T cell precursors up to the latest follow up was unexpected. However, this could have been explained by the persistence of TSCM cells that are endowed with decade long survival in humans (Biasco *et al.*, 2015; Oliveira *et al.*, 2015; del Amo *et al.*, 2018).

To better understand the immunophenotypic composition of T cells in these patients, we used 6 cell surface markers (CD3, CD4, CD8, CD45RA, CD95 and CD62L) to identify DP precursor (CD3+ CD4/CD8+ CD62L+ CD45RA+), TCM (CD3+ CD4/CD8+ CD62L+ CD45RA-), TEM (CD3+ CD4/CD8+ CD62L- CD45RA-) and TEMRA (CD3+ CD4/CD8+ CD62L-CD45RA+) subsets (Figure 3.2.1a). Most of the T cells exhibited one of the three classical memory phenotypes: TCM, TEM or TEMRA (Figure 3.2.1b).

We then used CD95 marker to differentiate between the two subsets that comprise the DP precursor population of T cells: TN (CD95-) and TSCM (CD95+) (Gattinoni *et al.*, 2011). Surprisingly, in 4 out of 5 patients, most cells within the DP population still displayed a TN CD95- phenotype up to the latest follow up (Fig. 3.2.2a) in both CD4 and CD8 compartments. Interestingly, although P8 had substantially lower numbers of DP cells, 49% of this population was also composed of true TN CD95- (0.5% of total CD3+ cells) (Figure 3.2.1b). Most importantly, VCN analysis showed that CD95- TN

cells were vector positive up to the latest follow up available and in all patients (average VCN of 1.6, Figure 3.2.2b).
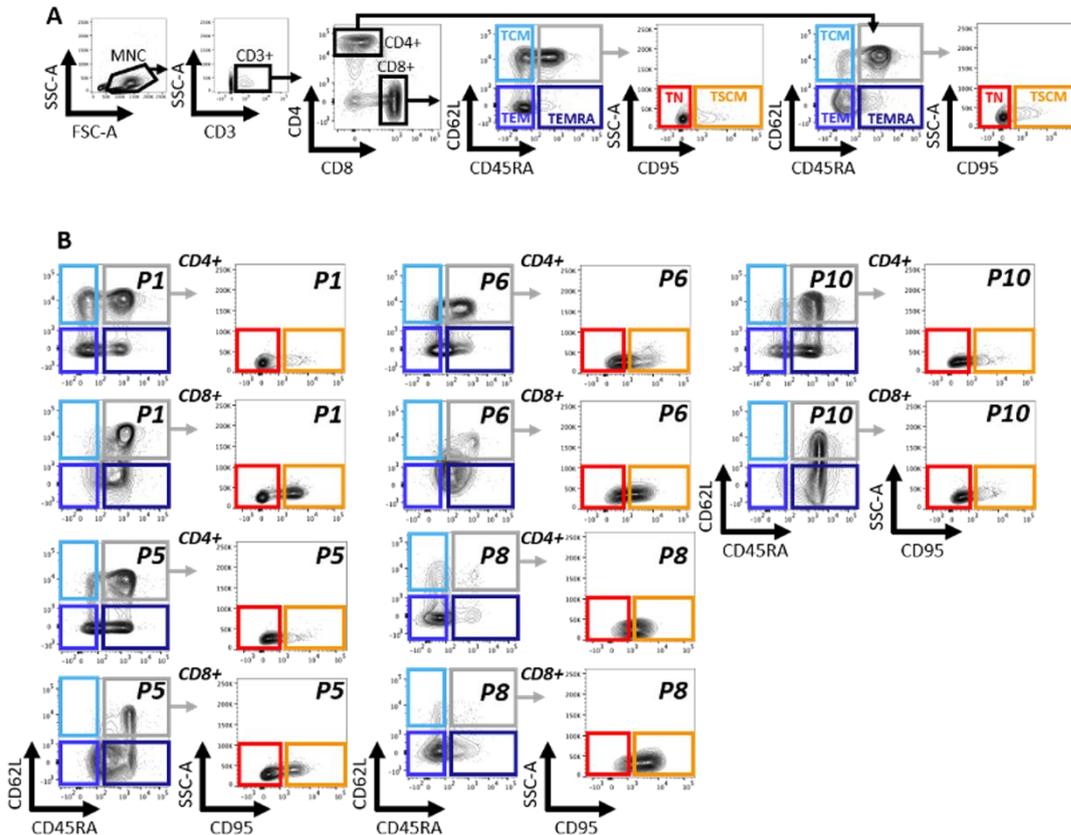
Figure 3.2.1 Gating strategy to identify different T cell subsets and phenotypic analysis of these subsets in patients.

a) Gating strategy to identify TN (red), TSCM (orange) and 3 classic memory populations – TCM (light blue), TEM (medium blue) and TEMRA (dark blue). Firstly, mononuclear cells (MNC) were identified by size and granularity using side (SSC-A) and forward (FSC-A) scatters. Then within MNCs, T cells were identified based on their CD3 expression, and CD3+ cells were further gated based on their expression of CD4 or CD8 markers. Then CD62L and CD45RA markers were used to identify DP population (grey), TCM, TEM and TEMRA cells. The CD95 marker was used to distinguish between TN (CD95-) and TSCM (CD95+) within the DP population. **b)** FACS plots showing T cell composition in CD4/CD8 compartments in each patient at the latest follow up.
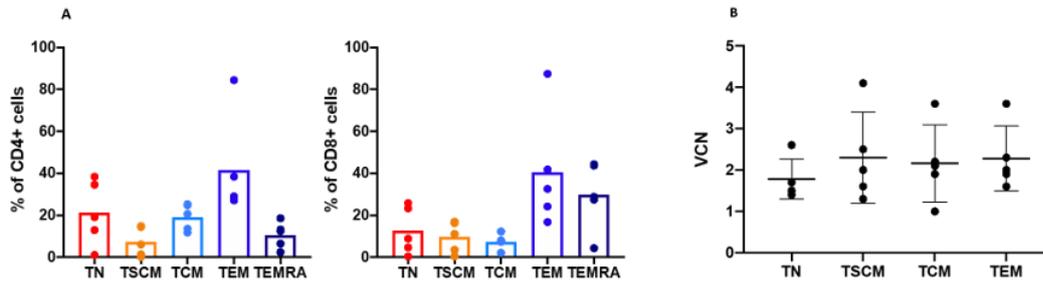
Figure 3.2.2 Percentages of each T cell subset in CD4 and CD8 T cells (left). VCN analysis in T cells at the latest follow up (right).

a) Percentage of each T cell subset in CD4 (left) and CD8 (right) compartments at the latest follow up in 5 patients analysed. **b)** VCN measured in each subset in 5 patients at the latest follow up available (P1 – 17 years post; P5 – 12 years post; P6 – 7 years post; P8 – 6 years post; P10 – 11 years post). Each dot represents VCN retrieved from one patient. Mean and standard deviations are also shown for each subpopulation analysed.

We then wanted to confirm that these phenotypically true naïve T cells are also functionally naïve. To this aim, we FACS sorted four T cells subsets (TN, TSCM, TCM and TEM) from 3 patients and from healthy donors (HD) and stimulated them for 6 hours with Phorbol 12-myristate 13-acetate (PMA) (50 ng/ul) and ionomycin (1µg/ml), before adding Brefeldin A (10µg/ml) and staining for IFN-gamma (IFNγ) (Fig. 3.2.3a). All patients displayed IFNγ expression similarly to HD controls (Fig. 3.2.3b). TN CD95-vector positive T cells, did not secrete significant levels of IFN-gamma upon stimulation as compared to the other subsets indicating that these were genuine TN cells (Fig. 3.2.3c).
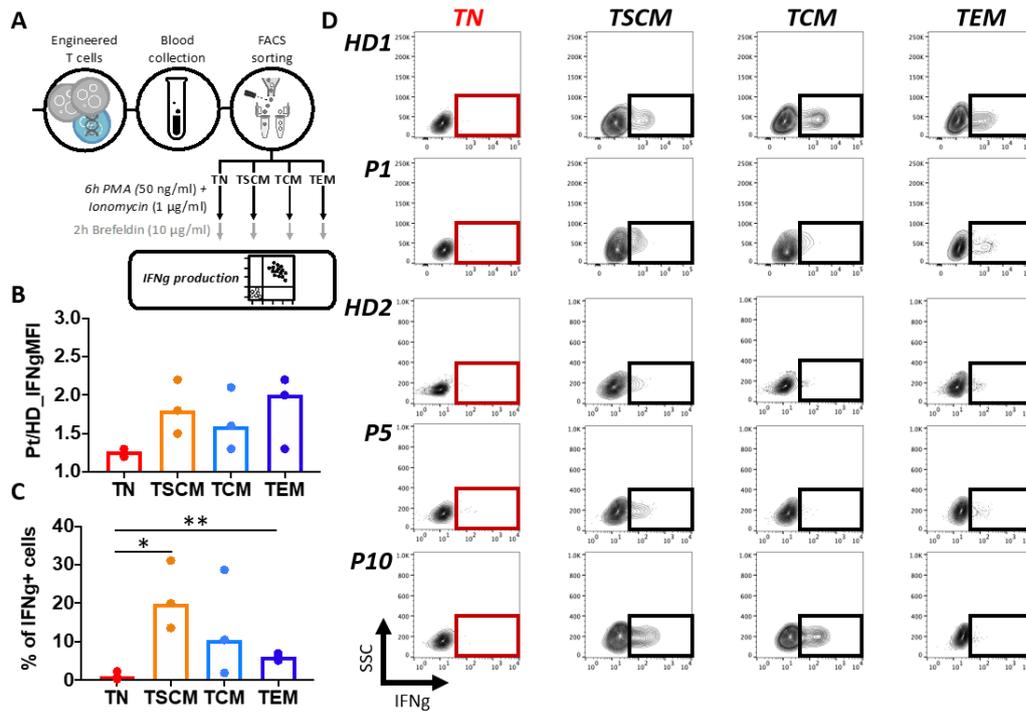
Figure 3.2.3 Interferon gamma production assay.

a) Experimental scheme for Interferon gamma production assay on 4 sorted T cell subpopulations. **b)** Ratio between IFNγ mean fluorescence intensity (MFI) measured in patients vs healthy donors T-cell subtypes after stimulation. **c)** Percentage of IFNγ positive cells measured in each T-cell subtype from patients after stimulation. (unpaired t-test * = p<0.05 [TN vs TSCM p=0.0165], **= p<0.01[TN vs TEM p= 0.0033]). **d)** FACS plot showing IFNγ expression after PMA/Ionomycin stimulation in each sorted T cell subset from two healthy donors (HD1 and HD2) and 3 patients (P1, P5 and P10).

3.3 Evaluation of thymic activity and TCR repertoire in GT patients

Presence of phenotypically and functionally true naïve T cells suggested that there might be a *de novo* T cell production still occurring in our patients. We then wanted to understand whether these vector positive T cells are the result of ongoing thymic output and are not some aberrant long-lived circulating cells with partial naïve T cell characteristics.

Because true naïve T cells can only be produced in the thymus, we used T cell receptor excision circle content as an indirect measurement of thymic activity (Figure 3.3.1). TRECs are extrachromosomal DNA by-products of TCR rearrangements that gets diluted during mitotic cell division. TRECs are only present in T cells of thymic origin that did not yet encounter their cognate antigen (Douek *et al.*, 2000). In 4 out of 5 patients, we observed detectable TREC levels even at latest time points, which is consistent with active thymic output and presence of true naïve T cells. As expected, there was a significant decline in TREC content in P8 at the latest follow up. This is consistent with significant decrease of DP precursor population we observed in this patient with immunophenotyping.
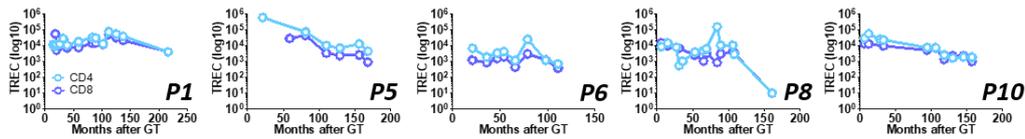
Figure 3.3.1 TREC content in CD4 and CD8 subsets in 5 patients overtime.

Number of TRECs (y axis) per million (10^6) of either CD4 (light blue) or CD8 (dark blue) T cells in each patient overtime (x axis) measured by qPCR (See Chapter 6, section 6.12 for details).

Another marker of thymic activity is high diversity of TCR repertoire. T cells with novel TCR specificities are generated in the thymus through the process of VDJ gene rearrangement. Because every naïve T cell that comes out from the thymus will have a unique TCR specificity, an ongoing thymic output will be marked by high TCR diversity of T cells in the periphery. Conversely, substantial decrease in the thymic output would result in skewed or even oligoclonal TCR repertoire.

We firstly assessed TCR repertoire in all T cells in our patients (Figure 3.3.2). TCR is a heterodimer molecule that consists of two polypeptide chains – alpha chain and beta chain (or gamma delta chains), connected by disulphide bonds. We used TCR beta spectra typing technique (See Chapter 6 Methods, section 6.11 for detailed protocol) that allows the analysis of differences in fragment length in the region of TCR beta chain that is called complementarity-determining region 3 (CDR3). CDR3 is a critical region in TCR that determines antigen specificity and represents an area with high diversity due to single nucleotide insertions/deletions. When thymus is active and functional, it generates high number of T cells with diverse CDR3 regions. Therefore,

100

normal TCR beta CDR3 spectra type would have a bell shape distribution. Normal distribution of fragment lengths in all 24 TCR beta families would be suggestive of ongoing TCR rearrangements and active thymic output. In 4 out of 5 patients, the distribution of TCR beta repertoire in T cells was normal, in contrast with P8, whose TCR beta repertoire was more oligoclonal.

Combined TREC and TCR beta spectra typing data were both indicative of ongoing thymic activity in these patients. However, they both showed us characteristics of the broad T cell compartment and did not allow us to conclude that a *de novo* naïve T cell production is still occurring in these patients.

We assumed that if our patients are still able to generate *bona fide* naïve T cells in the thymus, we should be able to detect high TCR diversity specifically in the naïve T cell population and new TCR specificities overtime. To test this hypothesis, we FACS sorted TN, TSCM, TCM and TEM populations from one healthy donor (HD) control and two patients (P8 and P10) and performed high-throughput TCR sequencing analysis. We used patient samples from two time points (early and late), to understand if there are any significant changes in TCR diversity of each subset that occurred over time.

We used TCR sequencing analysis (see Chapter 6, section 6.5 for detailed protocol) to amplify somatic rearrangements within TCR gamma, TCR delta and TCR beta genes. High-throughput TCR sequencing allowed us to detect 1,193 TCR rearrangements in total (Figure 3.3.3). We detected high diversity of TCR repertoire in all four subsets (Figure 3.3.3a). As expected, P8 had lower VDJ recombination range compared to P10 and HD, consistent with reduced thymic activity (Figure 3.3.3b). The range of TCR diversity was similar between HD and both patients in all subsets (Figure

3.3.4a) including naïve T cells (Figure 3.3.4b). We then tracked TCR rearrangements in naïve T cells overtime and showed that most of them were detected only at individual time points suggesting an ongoing production of naïve T cells with novel TCR specificities (Figure 3.3.5).

Because an estimated number of only TCR beta clonotypes in the individual is ranging between 10^6 to 10^8, there is always a risk that sampling biases can affect the interpretation of the tracking of individual TCR clones overtime (Robins *et al.*, 2009; Warren *et al.*, 2011; Qi *et al.*, 2014). For this reason, we firstly wanted to measure the level of sharing of identical TCRs among the four subsets at each time point. We assumed that if there were still an ongoing production of naïve T cells with novel specificities, then, firstly, we would expect to see relatively low level of overall sharing between subsets, and, secondly, higher level of sharing between memory populations than between naïve T cells and memory lymphocytes. As network plots on Figure 3.3.6a show, the level of sharing between subtypes was low in general, with TCM and TEM populations having the highest level of sharing. P8 had a substantially higher sharing of TCR clonotypes between subsets compared to P10, especially at the latest time point (Figure 3.3.6b). This was expected because as we showed previously, P8 had a severely reduced naïve T cell count, low TREC content and oligoclonal TCR beta repertoire at the latest time point all of which is suggestive of the progressive loss of thymic output.

We assumed that if new TCR rearrangements are continuously generated, we should observe a low re-capture of identical TCRs in naïve T cells overtime. Figure 3.3.6c shows that less than 5% of TCR rearrangements detected in naïve T cells at one time

point were re-captured at the second time point suggesting an active production of new TCR clonotypes.

Overall, the results of immunophenotypic, functional and molecular assays all suggested that there is still a *de novo* naïve T cell production occurring in the thymus of these patients despite the lack of persistence of engineered (Lt)HSC and that therefore this should be likely maintained by a population of vector-positive long term lymphoid progenitors.
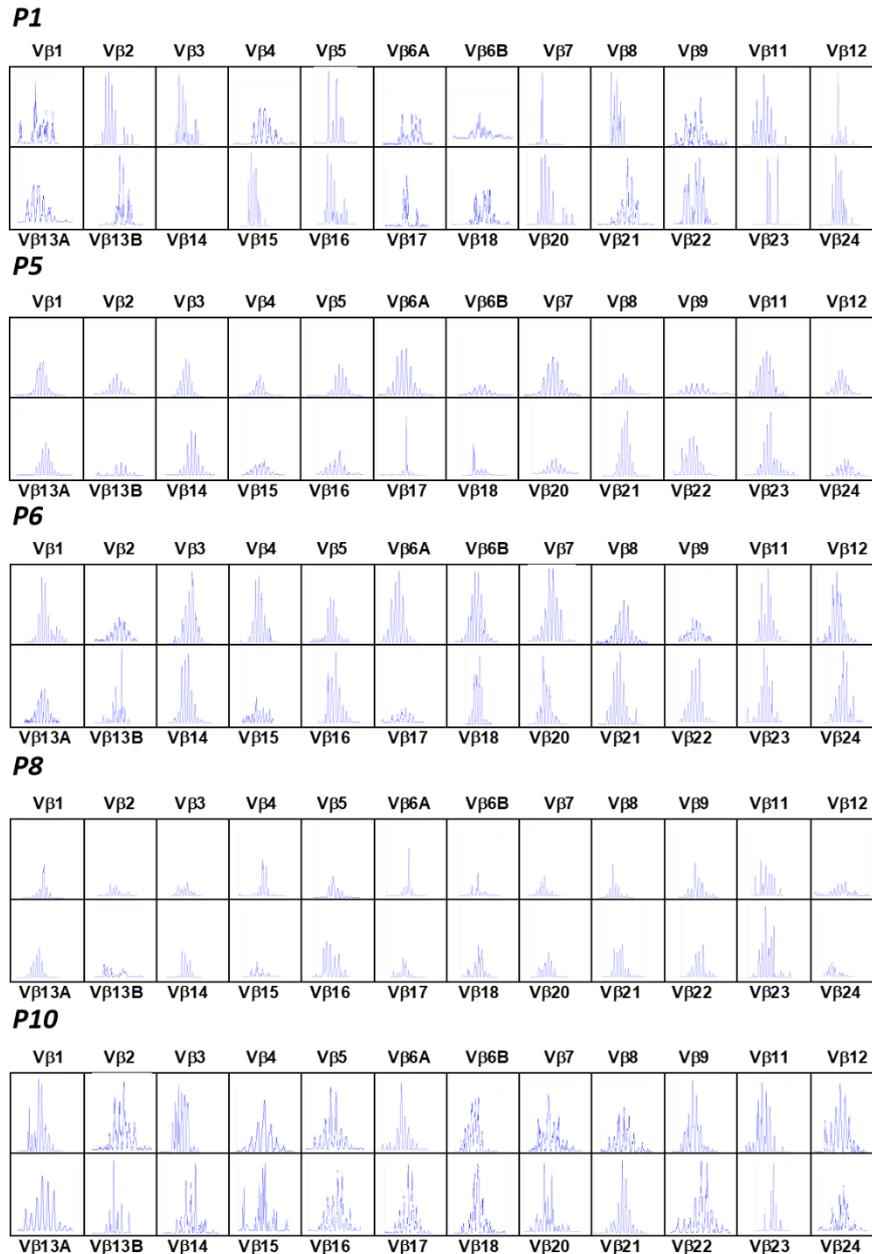
Figure 3.3.2 V beta spectratyping profiles in 5 patients at the latest follow up available (from 110 to 210 months post GT).

Each quadrant shows distribution of each TCR beta family (24 families in total) in the patient at the certain time point. P1 – 210 months post GT, P5 – 167 months post GT, P6 – 110 months post GT, P8 – 137 months post GT, P10 – 144 months post GT.
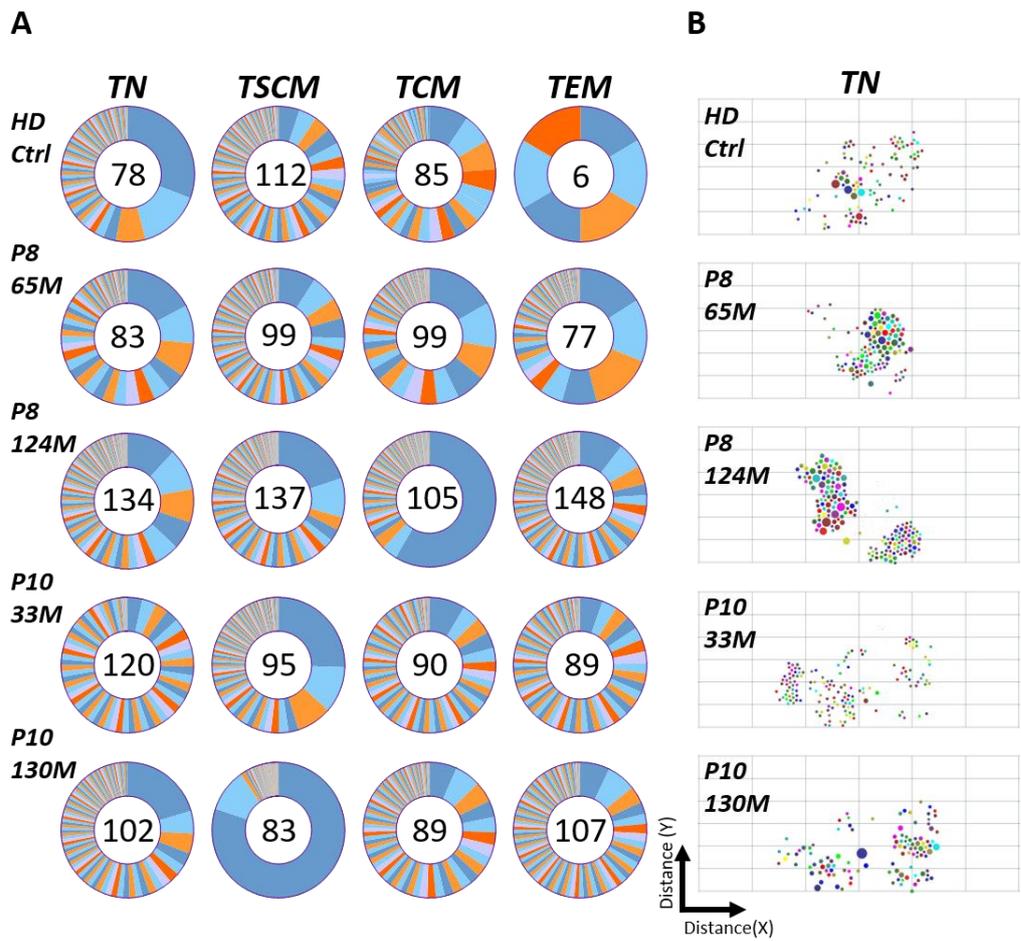
Figure 3.3.3 TCR sequencing analysis.

a) Ring plots showing the relative fractions of TCR rearrangements detected within each T cell subtype sorted from an HD and two patients (P8 and P10) at 2 timepoints each (months = M) after GT. Numbers inside each ring correspond to the numbers of individual TCR rearrangements retrieved from each sample. b) Diversity of TCR rearrangements shown as a 2-dimension t-SNE plot. Coordinates are derived from computing the differences (distance) among clones in sequence composition of a 50bp window centred on the Complementarity-determining region3 (CDR3). The size of each dot is proportional to the abundance of the corresponding TCR rearrangement. In these graphical

representations the higher it is the spread of dots across the area of the plot the higher it is the diversity of rearrangements observed within the TCR locus.
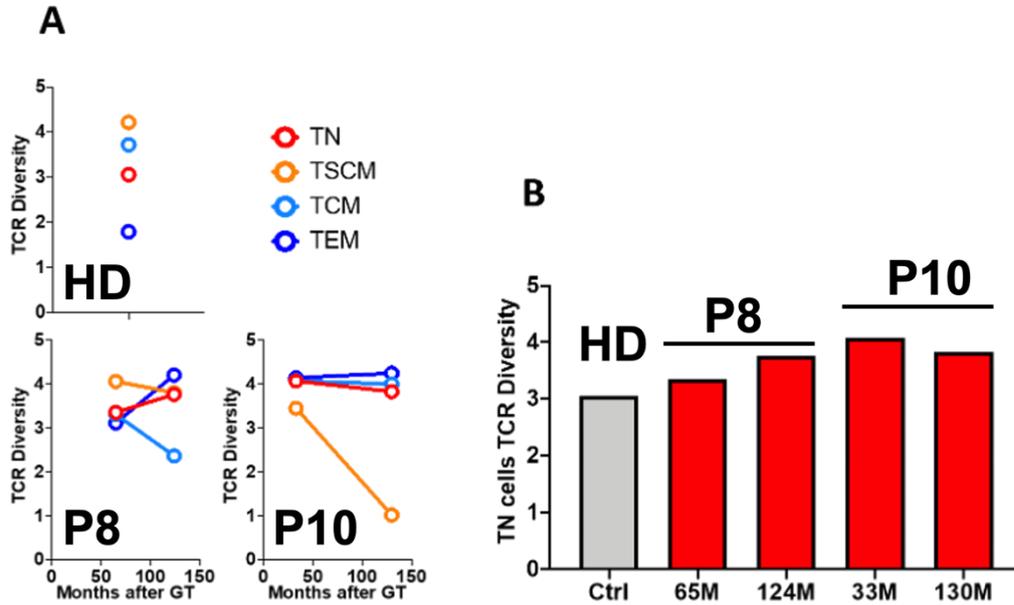
Figure 3.3.4 TCR diversity in all T cell subsets (left) and in naive T cells (right) in HD, P8 and P10.

a) TCR diversity measured by Shannon Diversity index in each sample and T-cell subtypes of HD (only 1 time point), P8 and P10. **b)** Diversity of TCR repertoire in naïve T cells (TN) of HD (grey bar) and Patients (red bars) at different time points.
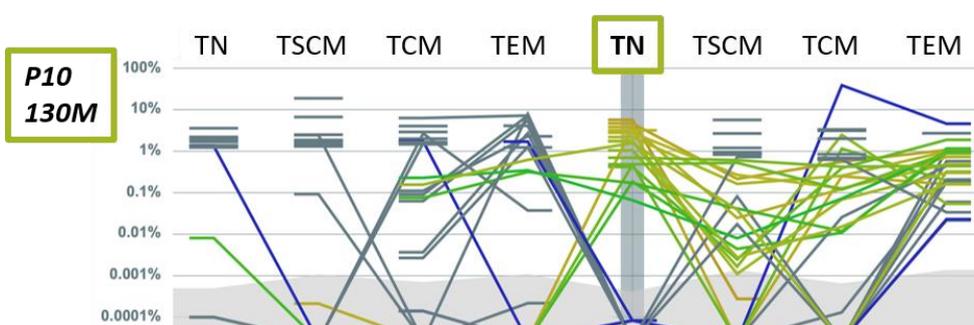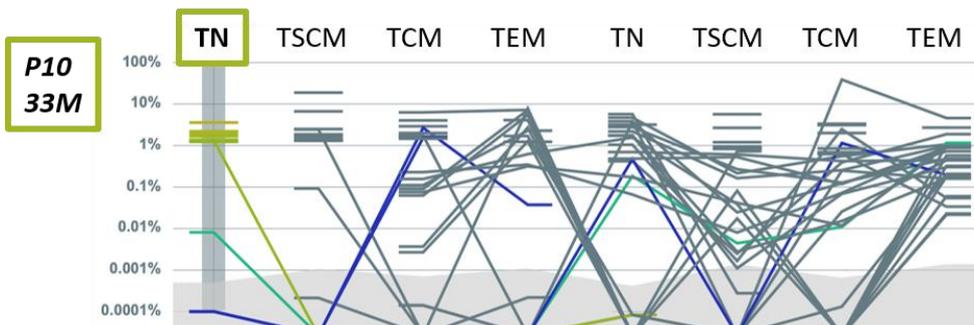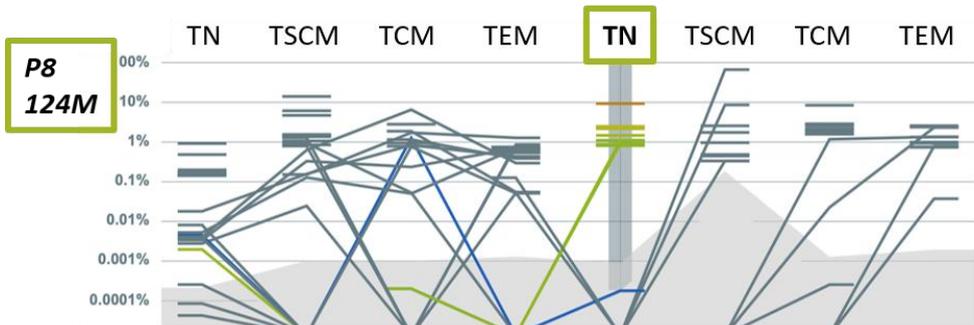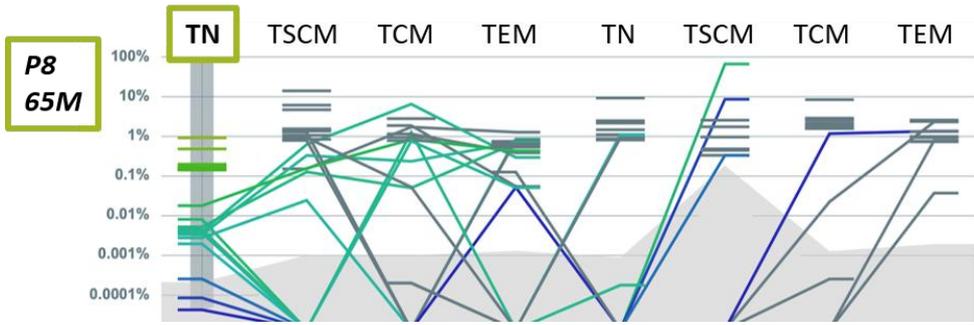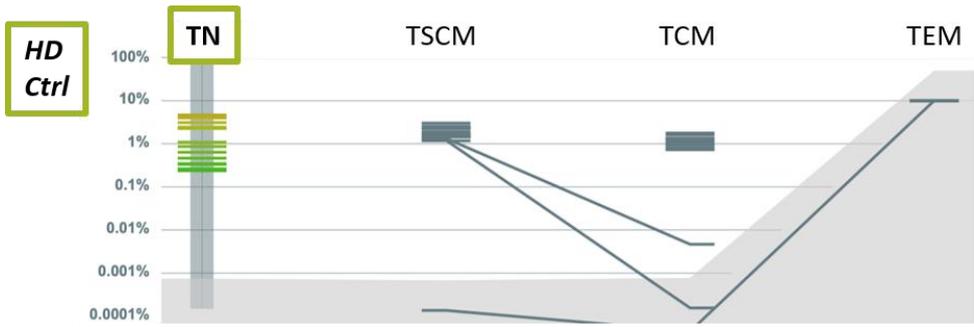
Figure 3.3.5 Relative abundance and tracking of TCR rearrangements in HD and patients overtime.

The y-axis displays relative abundance (log10 scale). Horizontal tiles show most abundant TCRs in each sample. Connecting lines show detection of the most abundant TCRs in each subpopulation across other samples and timepoints. The grey area displays average distribution of TCR relative abundance in each subpopulation. Data collected in TN at a given time point (on the left, M=months) are highlighted in green in the relative plot. Plots were generated using the Vidjil platform (www.vidjil.org).
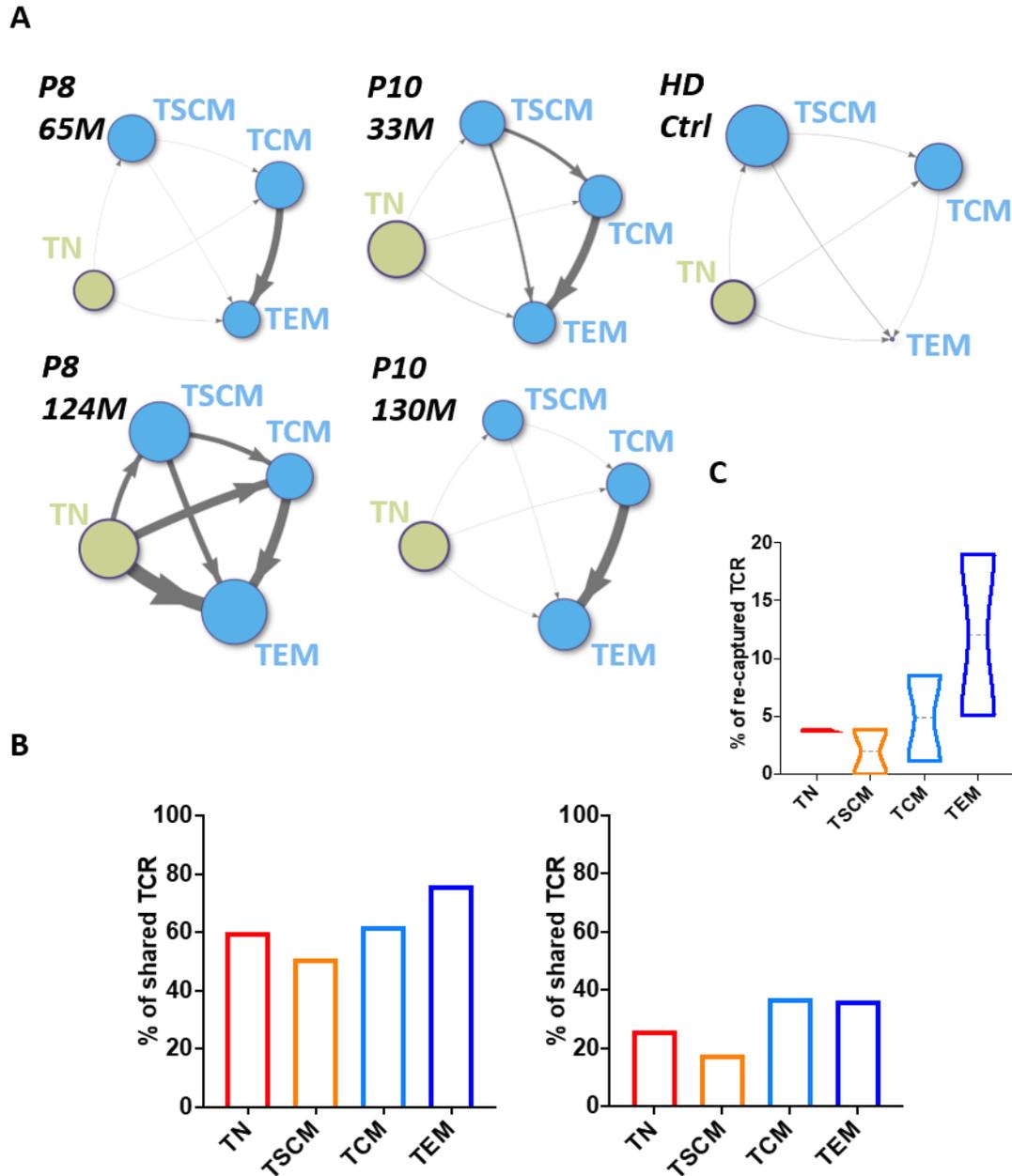
Figure 3.3.6 Analysis of TCR sharing between HD and patients.

a) Networks displaying TCR sharing between the four T cell subtypes from HD or P8 and P10 at each follow up. Nodes represent T-cell subtypes while edges (arrows) represent degree of TCR sharing. Size of each node is proportional to the number of TCR rearrangements detected in each T-cell subtype (Naïve T cells (TN) in light green, other

subtypes in light blue). The thickness of the arrows is proportional to the pearson correlation coefficient calculated based on TCR sharing between each pair of T-cell subtype. **b)** Percentage of TCR rearrangements shared by each T -cell subtype with at least another T-cell subpopulation in P8 (left panel) and P10 (right panel). **c)** Violin plots showing percent of TCR recaptured within each T-cell subtype isolated from two patients at two independent time points.

3.4 Clonal tracking of T cell subsets using integration site analysis

In our trial, CD34+ cells from our patients were transduced with gamma retroviral vector carrying a functional copy of IL2RG gene encoding common gamma chain receptor subunit. To study Lt-LP activity at the clonal level, we used freshly produced *bona fide* naïve T cells as a surrogate marker of Lt-LP's activity. We used LAM-PCR technique (see Chapter 6 Methods, section 6.4 for detailed protocol) combined with high-throughput next generation sequencing analysis to obtain and analyse 12,756 unique IS from 5 cell populations, including naïve T cells. Our analysis spans from 10.1 to 14.9 years after the loss of transplanted HSC.

We first wanted to understand whether the existence of Lt-LP was due to random vector induced clonal selection or whether it represented physiological and homeostatic survival. As shown on Figure 3.4.1, we occasionally identified some clones that contributed to more than 20% of individual T cell subsets sampled at individual time points. However, because none of these clones persisted over time, this suggests that these are most likely clonal bursts in response to an immunological challenge (i.e., infection) combined with potential sampling biases. In general, overtime

clonal diversity remained stable in all T subsets, including naïve T cells (Figure 3.4.2). Interestingly, despite a significant drop in the number of circulating true naïve T cells in P8, their profile remained highly polyclonal up to the latest follow up.

Figure 3.4.1 Relative abundance of IS in each T cell subpopulation and patient.

Scale of blue is used to group samples according to time point of analysis. The most abundant IS (>30%) are labelled with the name of the closest gene for each sample and time point (PB = peripheral blood; numbers = months after GT).
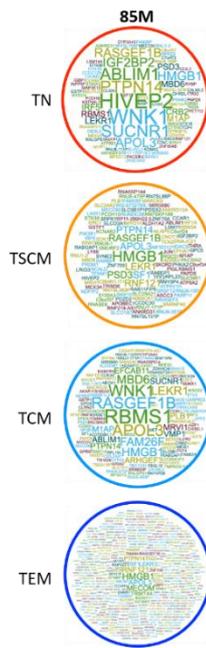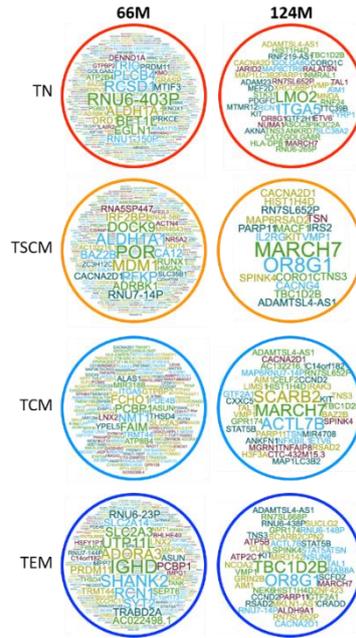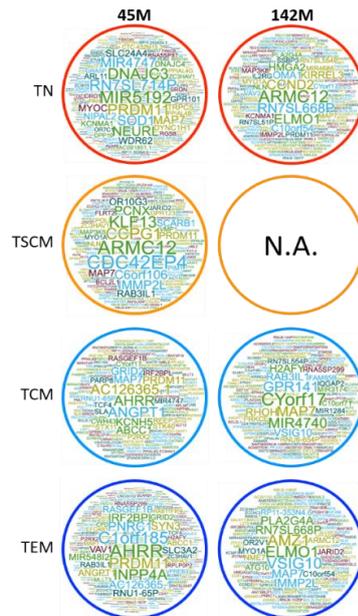
Figure 3.4.2 Word clouds showing the genomic distribution of IS from P1, P5, P6, P8 and P10 in each sample and time point.

Relative size of each gene is proportional to the relative level of IS clustering (the bigger it is the gene name the higher it is the number of IS detected in its proximity).

We then looked for the presence of IS skewing towards specific genomic loci as a marker of IS driven clonal selection. As shown on the Figure 3.4.2, IS sites were widely distributed across multiple genes and we did not detect any evidence of selection for proto-oncogenic regions in any of the T cell subpopulations and time points analysed. Interestingly, we observed that IS of naïve T cells were located preferentially in regions involving gene categories such as "Leukocyte activation", "Immune effector process" and "T-cell activation" (Figure 3.4.2). It has been previously shown that upon transduction of CD34+ cells with retroviral vectors these are not the loci that are preferentially targeted for integration (Wu *et al.*, 2003; Biasco *et al.*, 2011). Because there is no evidence for aberrant clonal selection, we can speculate that most naïve T cells have integrations in these loci as a result of a physiological positive selection for clones carrying the vector in genomic locations allowing for the optimal transgene expression upon T cell differentiation/maturation.

Next, we wanted to understand if the phenotypically and functionally true naïve T cells we observe in our patients had undergone differentiation into different memory subsets in vivo, another feature of *bona fide* naïve T cells. To this goal, we looked for the presence of identical IS between naïve T cells and memory populations. As shown in Figure 3.4.4 and 3.4.5, IS sharing was high and consistent across samples collected, sorted, analysed and sequenced independently. We detected a high number of IS shared between naïve T cells and memory subsets in all patients (from 24.1% to 57.2% of IS isolated from naïve T cells of each individual) suggestive of ongoing in vivo naïve T cell differentiation (Figure 3.4.6a). If naïve T cells in our patients were aberrant and

not capable of differentiation, we would expect to see significantly more identical IS shared between memory subsets only. In contrast, similar range of IS sharing was detected in all subpopulations (Figure 3.4.6b), a sign of substantial clonal relationship among T cell subpopulations and the existence of active common progenitors. To provide more formal evidence that individual Lt-LP clones are sustaining *de novo* T cell production over many years we used longitudinal data collected from naïve T cells and interrogated our dataset for detection of identical IS overtime. As shown in Figure 3.4.6c and Figure 3.4.7, in all 4 patients from which we had T cells analysed at multiple time points, a fraction of IS in TN (1.3%-14.8%) could be recaptured overtime in a window of up to 10.1 years from one collection to another. In the two patients where analysis could be performed at more than two time points, we used IS recapture probability to estimate the number of active engineered Lt-LP clones. In these individuals, T cell production is maintained by approximately 2-6 thousand individual engineered Lt-LP (Figure 3.4.6d).
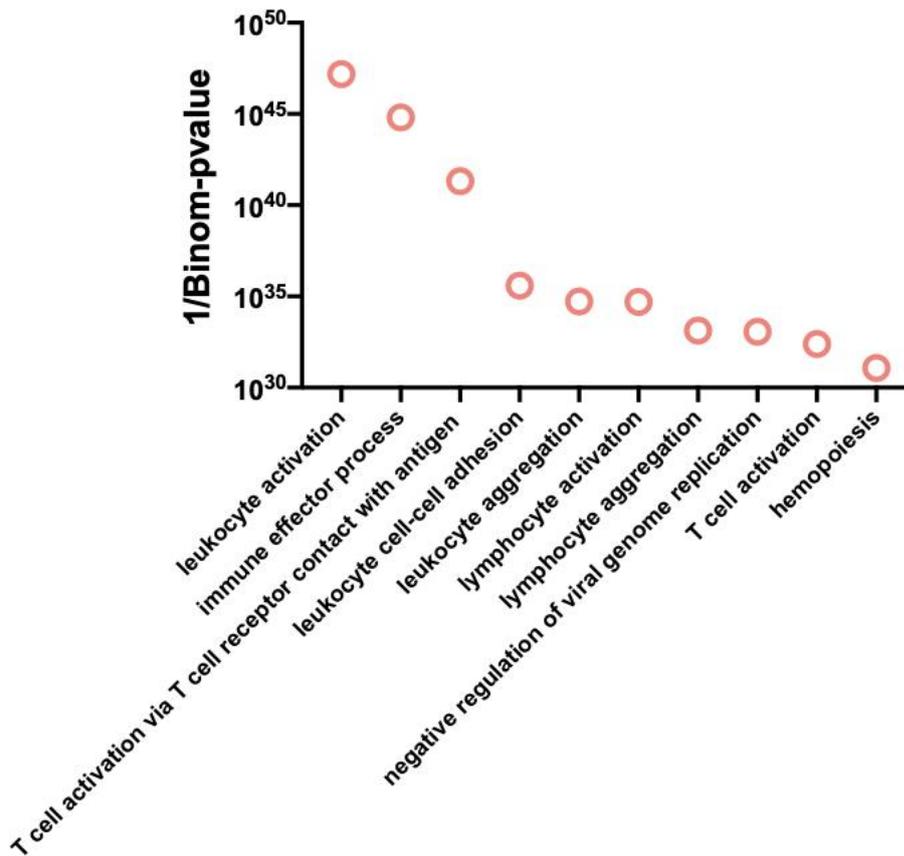
Figure 3.4.3 Distribution of IS in naïve T cells (TN) and gene ontology of hit genes.

The word cloud on the top show the genomic distribution of IS in TN of all patients and time points (graphical representation similar to Figure 3.4.2). The plot on the bottom show significance of the top Gene Ontology categories involved by IS as analysed through the Genomic Regions Enrichment of Annotation Tool (GREAT, great.stanford.edu).
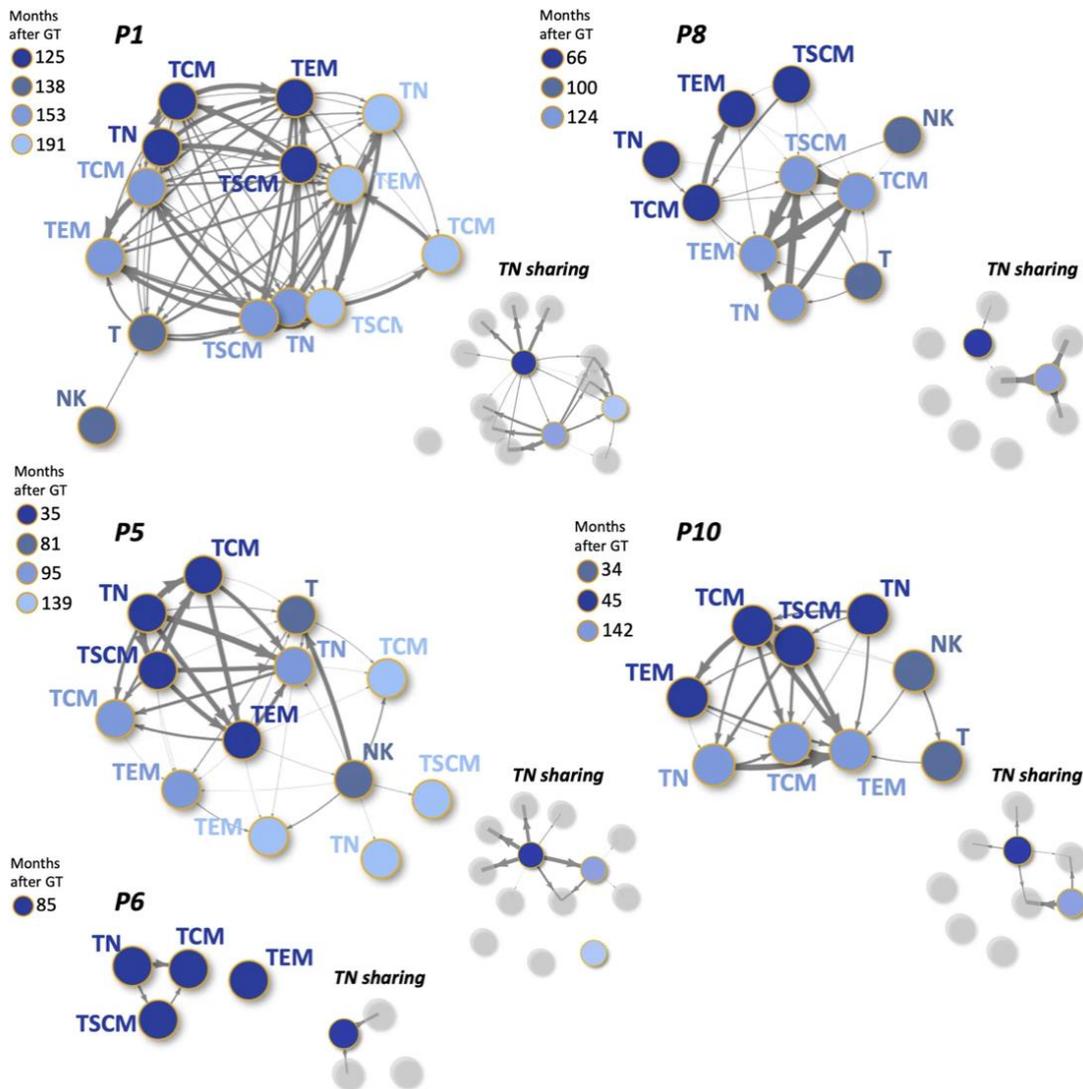
Figure 3.4.4 Network plots showing significant sharing of IS among the T-cell subtypes and NK cells from each patient.

Each node represents a sample, and it is coloured according to the time point after GT. The thickness of the arrows is proportional to the value of positive Pearson Correlation Coefficients calculated on the

basis of IS sharing between pairs of samples (ranges of values are displayed on the heatmaps of Figure 3.4.5). For each patient two plots are shown: the networks on the left display positive correlations among all samples while the network on the right (TN sharing) highlights correlations between TN cells (nodes in scale of blue) and the rest of the subtypes (nodes in grey).
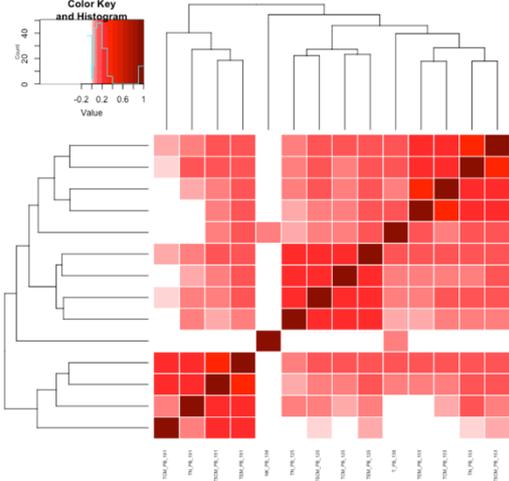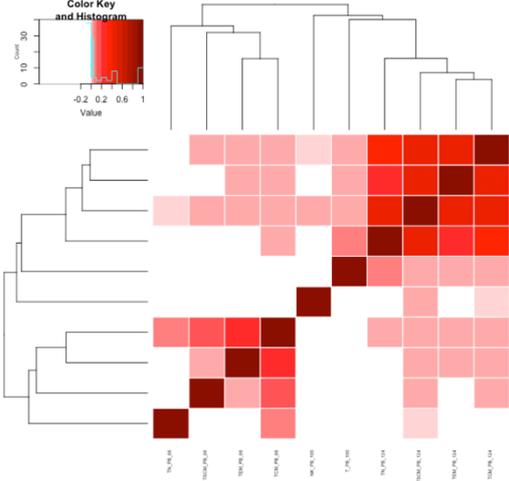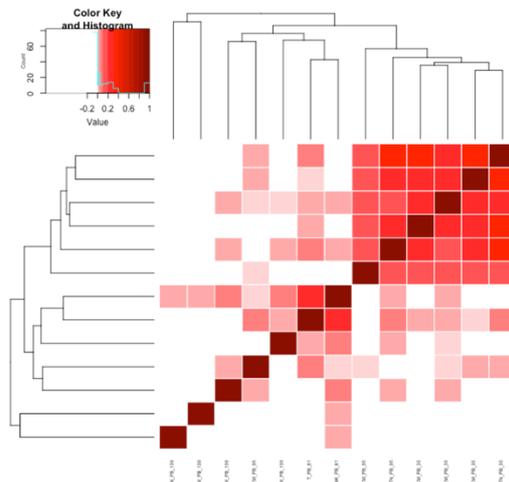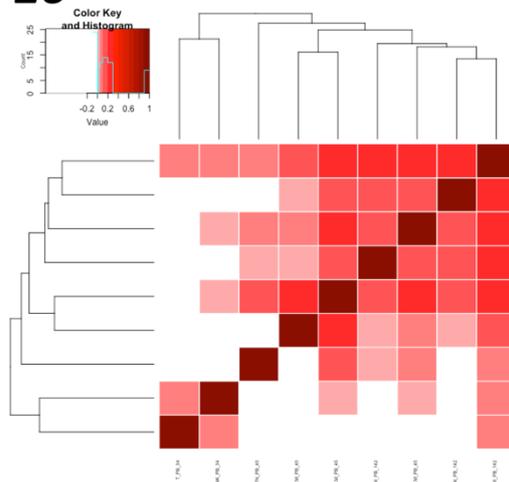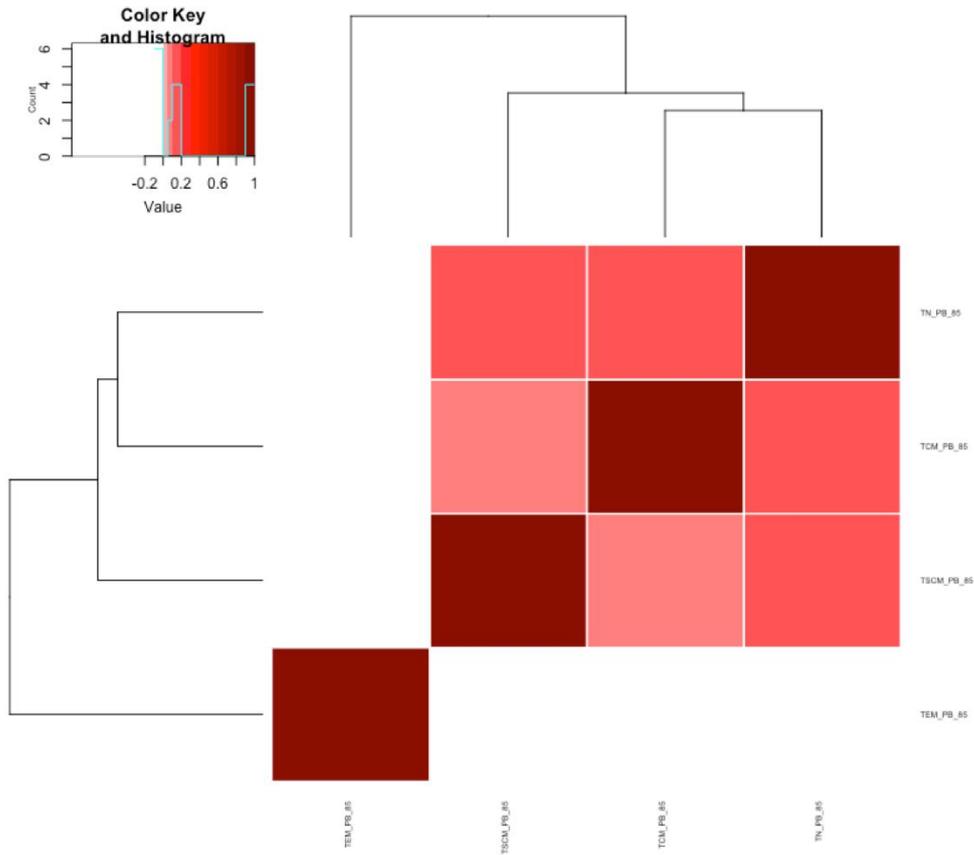
Figure 3.4.5 Heatmaps showing Pearson Correlations Coefficients (scale of reds) across populations and timepoints in the 5 patients analysed.

Pearson correlation is used to measure the strength and direction of the linear relationship between variables. The intensity of colour on heatmaps signifies the strength of the relationship between subsets (column labels are shown as "subpopulation_PB (peripheral blood) _months after GT").
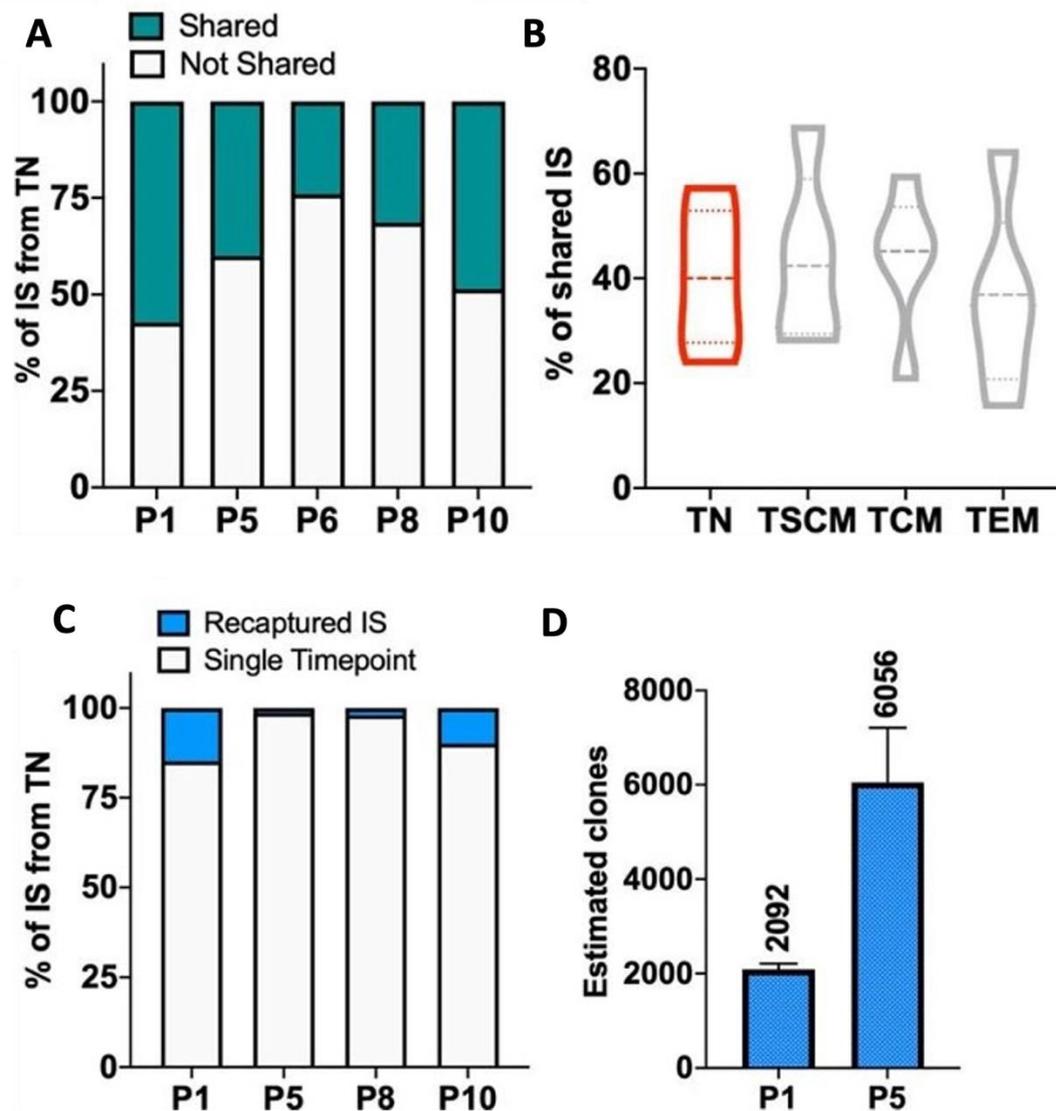
Figure 3.4.6 VISA to track IS sharing between TN and other subsets and for estimation of clonal abundance.

a) Stacked bars showing in green the fraction (%) of IS from TN of all timepoints for each patient which were shared with at least another T cell subtype. b) Violin plots showing the percentage of IS belonging to each T subpopulation shared with at least another T cell subtype (data from all patients and time points; TN data highlighted in red). c) Stacked bars showing in blue the fraction of IS re-captured across multiple time points within the TN populations of each patient (a more detailed view of re-captured IS from TN is displayed on the heatmaps of Figure 3.4.7). d) Clonal abundance (number of clones in circulation) of long-term lymphoid progenitors estimated in P1 and P5 on the basis of size of IS datasets

at each timepoint and re-capture probabilities overtime of IS within the TN compartment (values derived from the Mh Chao loglinear model for capture-recapture and relative standard errors as reported in Chapter 6, section 6.6 Table 6.6.3).
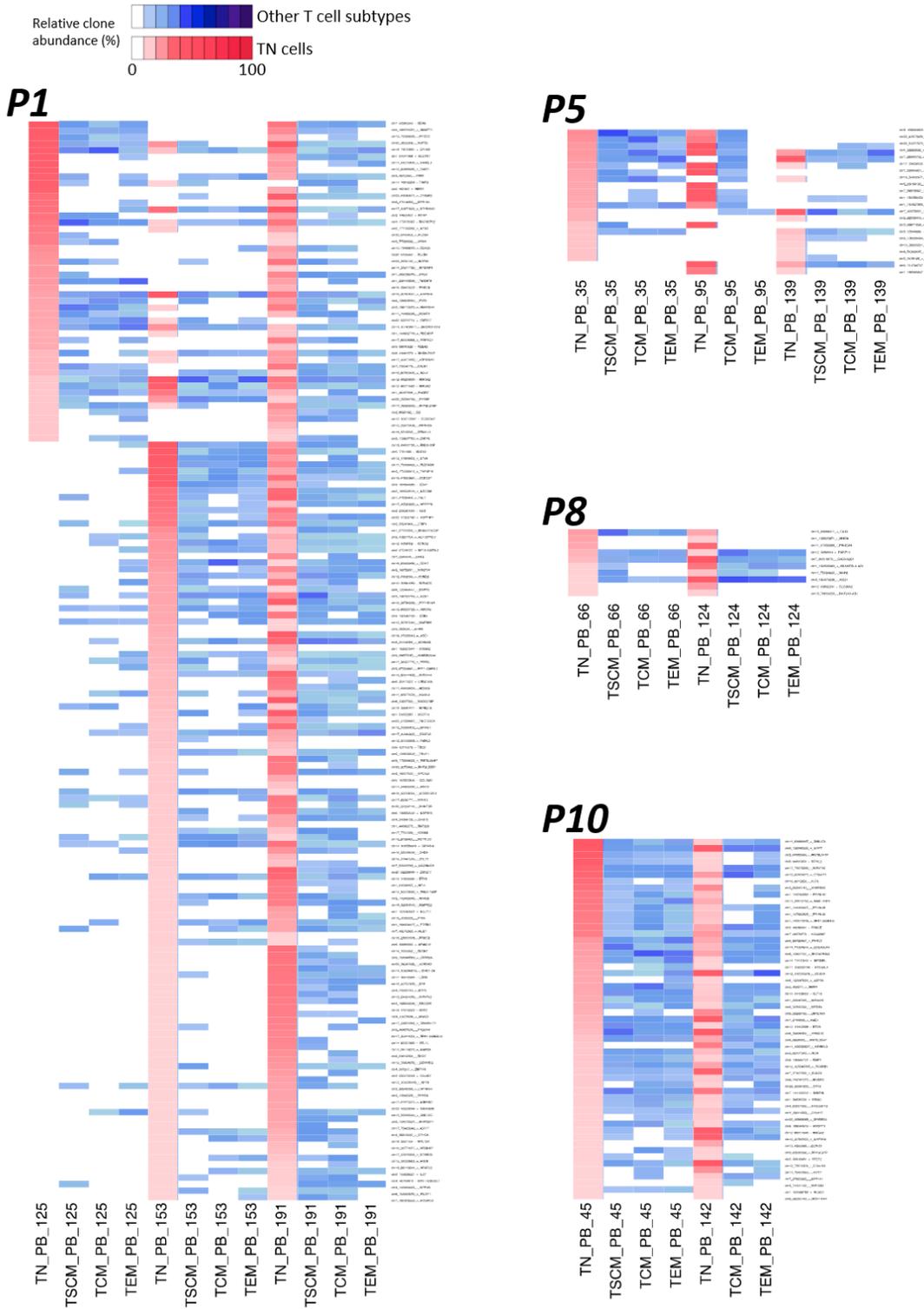
Figure 3.4.7 Heat maps displaying recaptured TN integration sites in each patient across each sample/time point.

Each row is named by the locus of the IS and the name of the closest gene while each column is labelled by sample and time point (number of months after GT). Intensity of colour is proportional to the abundance of each integration site (red = IS from TN, blue = IS from other T cell subtypes, white = IS not detected).

3.5 Immunophenotypic and molecular analysis of vector positive NK cells

The fact that NK cell subset is the only other immune cell subset that still carries integrated vector copies raised the interesting possibility that T and NK cells of these patients could share a common Lt-LP with dual potential. NK cells are short lived lymphocytes, with a half-life of approximately 2 weeks, except for memory or adaptive NK cells. In humans, at least 5 different maturation stages of NK cells have been reported identifiable based on cell surface marker expression (Di Vito, Mikulak and Mavilio, 2019) (Freud *et al.*, 2005). The most common receptors that are used in combination to identify NK cells are CD56 (Neural Cell Adhesion Molecule) and CD16 (Fragment crystallisable (Fc) receptor - FcγRIIIa).

In contrast to T cells, CD3-CD19-CD56+ NK cell numbers remained well below the normal range in all patients up to the latest follow up (Figures 3.5.1a, b). Interestingly, common gamma chain receptor expression levels were similar to T cells and healthy donor control (Figure 3.5.1c).
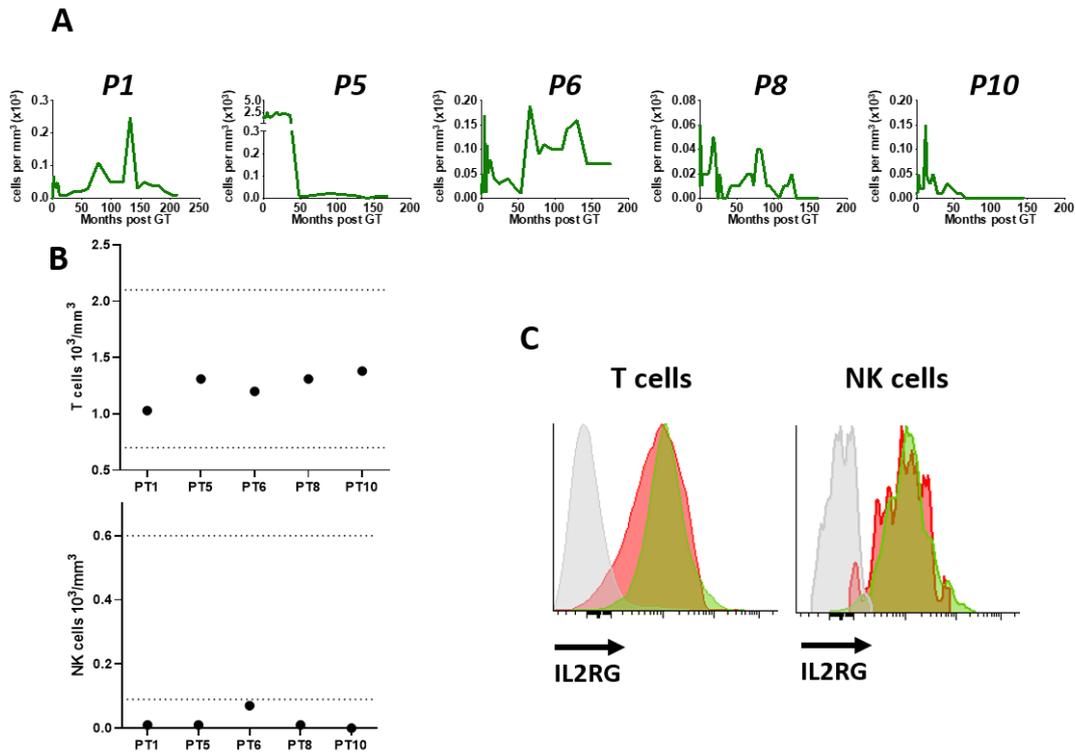
Figure 3.5.1 Analysis of NK cells in patients.

a) Absolute NK cells numbers in CD56+ NK cells in P1, P5, P6, P8 and P10 overtime post GT. **b)** T cell (top panel) and NK cell (bottom panel) cell numbers at the latest time point in each patient (black dots) compared to upper and lower ranges in age matched healthy controls (dotted lines). **c)** FACS plots showing the γchain expression in T cells and NK cells in patient 10 (red) and a healthy donor (green). Isotype control is shown in grey.

To confirm that NK cells in our patients are the true NK cells (Figure 3.5.2a), we performed a comprehensive immunophenotypic analysis to study the expression of CD56, CD16, Perforin, Granzyme B, CD94, CD57, KIRs, NKG2C and NKG2A in 3 patients and healthy donor controls (Figure 3.5.2b, c and Figure 3.5.3a, b). In humans, CD3-CD19-CD56+ NK cells normally represent around 5-15% of total leukocytes in the circulation (Caligiuri, 2008). In our patients, the proportion of CD3-CD19-CD56+

true NK cells over the total lymphocyte population was 0.2-1.2%, substantially lower compared to the 8-29% measured in healthy controls (Figure 3.5.2).
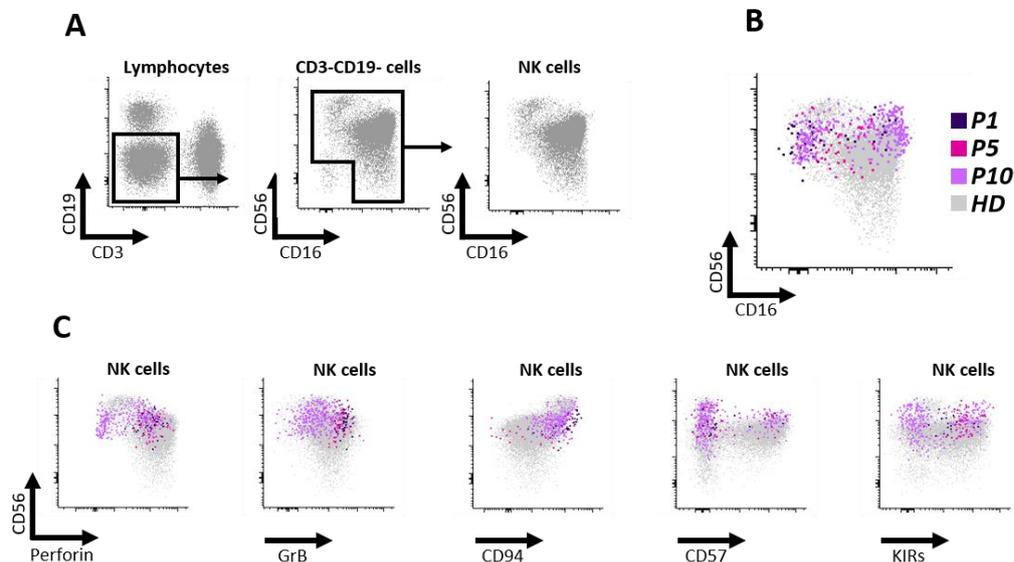
Figure 3.5.2 Phenotypic characterisation of NK cells in patients.

a) FACS plots showing the gating scheme for the identification of NK cells within the lymphocyte population of a healthy donor used as reference. **b)** Overlayed FACS plot showing the expression of CD56 and CD16 on NK cells in HD, P1, P5 and P10. **c)** Overlayed FACS plot showing the expression of CD56, Perforin, GrB, CD94, CD57 and KIRs on NK cells in HD, P1, P5 and P10.

We then wanted to understand whether gene corrected NK cells are freshly produced cells or long-lived memory NK cells. Generally, immature NK cells can be distinguished from mature cells based on their differential expression of CD56 and CD16 receptors (Di Vito, Mikulak and Mavilio, 2019). The immature NK cells express high levels of

CD56 cell surface receptor (CD56bright) and do not express CD16 receptor, whereas mature NK cells downregulate their CD56 receptor expression (CD56dim) and express CD16 marker. The adaptive/memory NK cells are commonly identified by the expression of CD94/NKG2C. They are mainly observed in response to human cytomegalovirus, and they can be detected months or even years after the infection (Lopez-Vergès, 2011; Hendricks *et al.*, 2014; Freud *et al.*, 2017; Pupuleku *et al.*, 2017; Capuano *et al.*, 2019).

As shown on Figure 3.5.3, three out four patients analysed still had detectable immature CD56bright NK cells present in the peripheral blood (2.8-9.9% of total NK cells). Only in one patient (P1), no CD56bright NK cells were detected, which could be due to the low number of total NK cells that have been available for analysis from this individual. Although the mean fluoresce intensity (MFI) of CD56 within this subset was slightly lower (mean = 29,616) as compared to the one measured in healthy donors used as controls (mean = 39,696) (Figure 3.5.2b and Figure 3.5.3). A fraction of NK cells expressed CD57 receptor, a marker generally associated with mature NK cells (Figure 3.5.2c and Figure 3.5.3). However, the percentage of CD57+ NK cells in our patients was lower (28-30%) compared to healthy controls (39-68%). This is also different from the situation described in patients with GATA2 mutations who appear to maintain an aberrant population of long living memory/adaptive NK cells (Schlums, et al., 2017).

We then wanted to look at the proportion of long lived adaptive/memory NK cells in our patients and studied the expression of NKG2C and NKG2A in P10. As Figure 3.5.4 shows, only very small number of NK cells exhibited NKG2C+NKG2A-adaptive/memory phenotype.
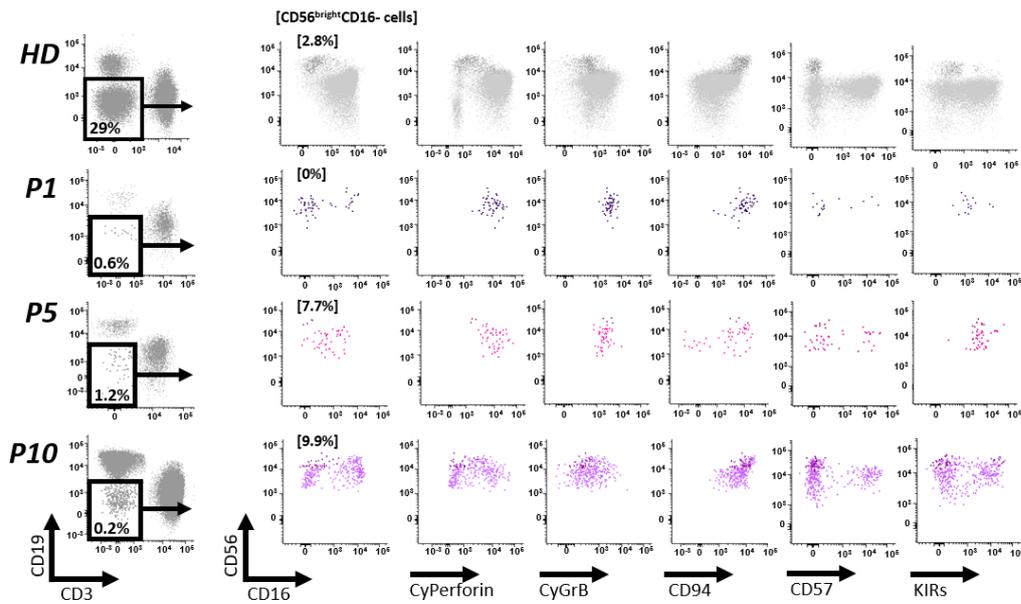
Figure 3.5.3 FACS plots showing the percentages of NK cells and NK associated cell surface markers in HD, P1, P5 and P10.

The percentage of CD56brightCD16- within the NK population is shown for each individual and events falling into this gate are highlighted with a darker colour shade in each dot plot. CyPerforing - Perforin, Granzyme B - CyGrB and Killer-cells immunoglobulin-like receptors KIRs.
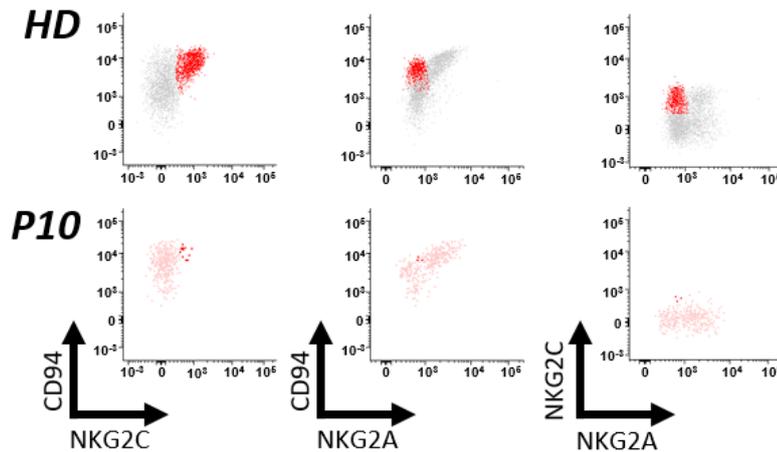
Figure 3.5.4 FACS plots showing the identification of memory/adaptive NK cells (CD94+NKG2C+NKG2A-) in a healthy donor and in P10.

Cells depicted in red represent cells of interest – CD94+NKG2C+, CD94+NKG2A-, NKG2C+NKG2A-.

Overall, these data suggested that there might be a de novo NK cell production still occurring in these patients. Next, we wanted to understand if gene corrected T and NK cells originated from the same LP in these individuals. We hypothesised that if we would be able to find identical IS between T and NK cells, this would be highly suggestive that they originated from the same progenitor clone. To address this point, we collected and analysed 651 IS from CD3-CD19-CD56+ FACS sorted NK cells overtime and looked for identical IS between NK and T cells. As expected, considering low NK cell counts in our patients, the number and diversity of IS collected was substantially lower than that of IS collected from T cells (Figure 3.5.5).
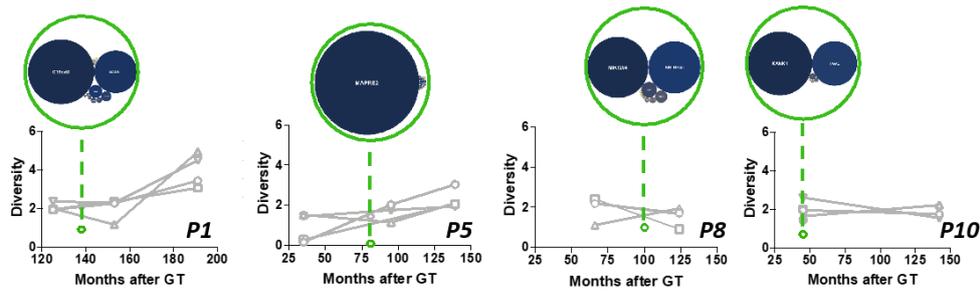
Figure 3.5.5 Plots showing Shannon Diversity Index of IS overtime in NK cells (green dots) and in the other T-cell subtypes (grey lines).

At each timepoint analysed green circles contain bubble plots of clones contributing >0.01% to the total NK population. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside the bubble.

To account for possible sample cross-contamination during processing and/or sequencing, we firstly looked at the level of shared IS between NK cells and naïve T cells or total T cells in samples that were processed and sequenced independently. As shown on Figure 3.5.6, up to 41% of identical IS were shared between NK cells and naïve T cells that were processed and sequenced separately. We also found that up to 60.7% of identical IS were shared between NK cells and all T cell populations at different time points (Figure 3.5.7) and up to 90.2% of IS were shared between NK cells and T cells isolated from the same time points (Figure 3.5.6). All these data strongly suggests that NK cells and T cells in our patients are being produced by bipotent Lt-LPs.
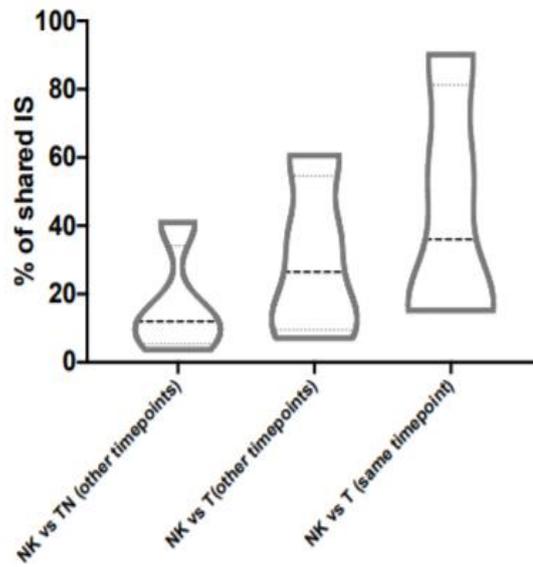
Figure 3.5.6 Percentage of shared IS between NK cells, TN and total T cells.

Violin plots showing the percentage of NK integration sites (y axis) shared with at least one TN (NK vs TN), total T cells from different timepoints (NK vs T other timepoints) or total T cells from the same timepoint (NK vs T same timepoint) in all patients.
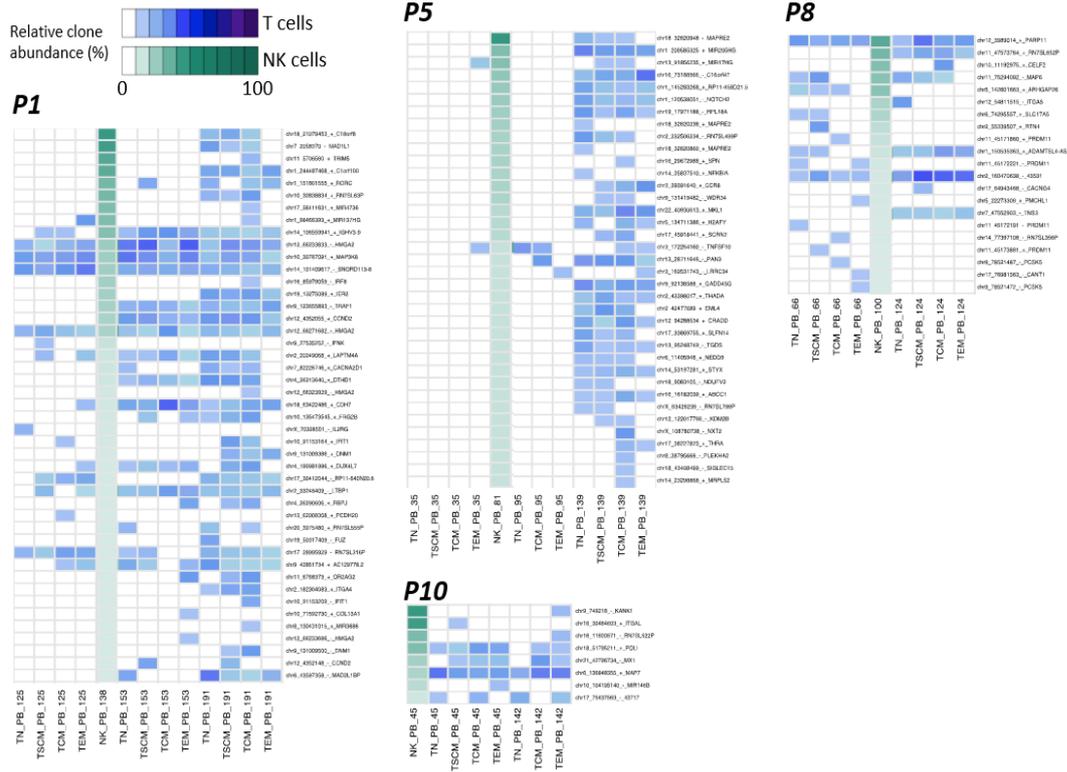
Figure 3.5.7 Heatmaps displaying integration sites shared between NK cells and T cell subtypes in each patient

Each row is named by the name of the closest gene while each column is labelled by sample and timepoint (number of months after GT). Intensity of green (NK) or blue (T-cell subtypes) colours is proportional to the abundance of each integration site (white = not detected).

3.6 Analysis of integration sites in the proximity on oncogenes

One of the patients (P1) in this trial, was treated 19 years ago making this study the longest gene therapy follow up available to date. We decided to take advantage of

these opportunity to better understand the long-term effects of integrations in proximity of well-known proto-oncogenic sites.

The genome of retroviruses consists of three essential genes – gag, pol and env, and is flanked by long terminal repeats on both sides. First generation retroviral vectors had intact enhancers on both sides of proviral LTRs that were able to significantly boost transcription of neighbouring genes. During vector integration process, an enzyme called integrase, that is encoded by pol gene, inserts the proviral DNA into the host genome. In some cases, viral vectors can integrate next to proto-oncogenes thereby viral enhancer elements (unless removed during vector production) can upregulate transcription of these genes and consequently lead to uncontrolled proliferation of affected cells. This can eventually produce aberrant clonal expansion and in some cases leukaemia, one of the most significant safety concerns related to retroviral HSPC gene therapy (Ibraheem, Elaissari and Fessi, 2014) (Baum, 2007).

As mentioned previously, one of the patients in our trial (P8) developed T-ALL due to the insertion of the vector next to the LIM domain only (LMO2) proto-oncogene (Hacein-Bey-Abina *et al.*, 2003). LMO2 plays a central role in hematopoietic development as a complete LMO2 deficiency fails to contribute to any stage of haematopoiesis in chimeric LMO2 deficient mice (Yamada *et al.*, 1998; El Omari *et al.*, 2011). Insertions next to other proto-oncogenic sites (CCND2, MECOM) also contributed to the development of T-ALL in patients in SCID-X1 trial that was conducted in Necker Hospital in France (Cavazzana-Calvo *et al.*, 2000; Hacein-Bey-Abina *et al.*, 2003, 2008; Howe *et al.*, 2008).

We therefore decided to look at presence and relative frequency of integrations next to three main proto-oncogenes (LMO2, CCND2 and MECOM) (Figure 3.6.1). We were able to detect 52 integration sites in total near these genes. However, all these integrations contributed to less than 10% of the total clonal composition of each population, except for one integration next to CCND2 gene that had a maximum relative contribution of 40% but only in one T cells subset (TN) and only at one timepoint (142 months post GT). As shown on Figure 3.6.1, some of these proto-oncogenic IS were found at several timepoints but this did not lead to any progressive or stable expansion of any of the relative clones overtime.

We then wanted to see if the leukemic clone that caused T-ALL in P8 reappeared and/or significantly contributed to the survival of Lt-LP. To this aim, we analysed a sample that was taken from P8 at the time of T-ALL (24 months post GT) and compared it to samples analysed at later time points (Figure 3.6.2). As expected, at the time of T-ALL integration site next to LMO2 gene was substantially overrepresented (relative abundance - 90.8%) compared to other IS in the peripheral blood. After 24 months post GT this clone was not detected anymore apart from one time point in T cells at 100 months post GT. Although, relative abundance of this clones was so low (0.0001%) that it could potentially be sequencing noise. Altogether these data suggest that Lt-LP are physiologically maintained in these individuals and that their long-term survival is not dependent on the effects of insertional mutagenesis in proto-oncogenic loci.
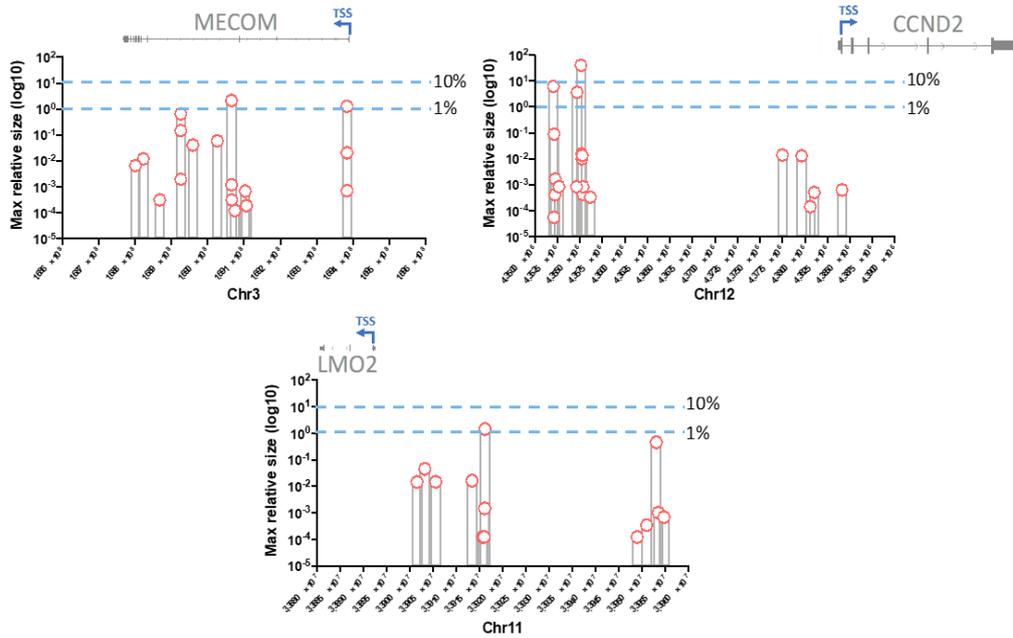
Figure 3.6.1 Distribution of IS in proximity in LMO2, MECOM and CCND2 loci.

In each plot the *x*-axis displays chromosomal coordinates (bp) while the *y*-axis the IS abundance (log10 of percentage) calculated within the population and timepoint where it has been observed. The light blue dotted lines show the 1% and 10% abundance thresholds. The transcription start sites (TSS), exons (grey boxes) and introns (grey lines) of each gene are shown on top of each plot in their respective chromosomal localisation.
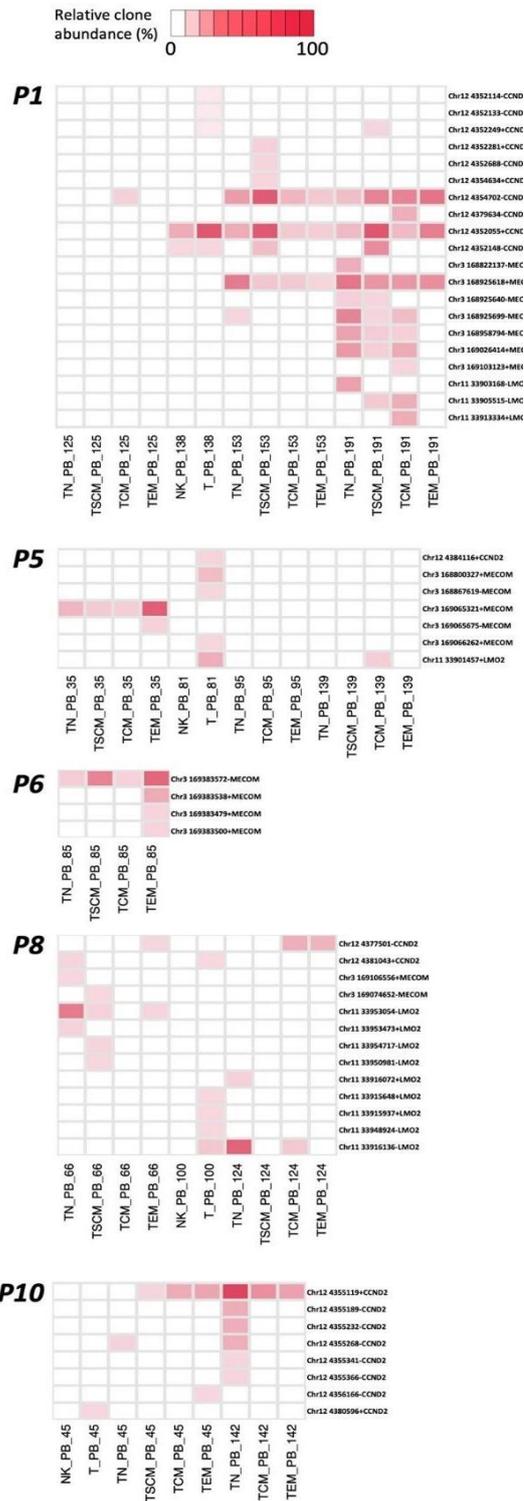
Figure 3.6.2 Heatmaps displaying the detection of integration sites in each patient across each sample/timepoint.

Each row is named by the locus of the IS and the name of the closest gene while each column is labelled by sample and timepoint (number of months after GT). Intensity of red is proportional to the abundance of each integration site (white = not detected).

Figure 3.6.3 Heatmap showing the IS collected in the peripheral blood (PB) of P8 during leukaemia at 24 months after GT (first column) and their detection overtime.

Each row is named by the name of the closest gene while each column is labelled by sample and timepoint (number of months after GT). The first row displays the LMO2 integration associated to the insertional mutagenesis event while the rest of the rows show other by-stander IS collected from PB at 24 months after GT. Intensity of orange is proportional to the abundance of each integration site (white = not detected).

4 Results part III – Using VISA to study CAR T cell kinetics and persistence in CARPALL patients

The third part of my PhD work aimed to explore CAR T clonal dynamics and phenotype changes post infusion in vivo in the CARPALL clinical trial for B cell acute lymphoblastic leukaemia. To this goal we designed an experimental workflow that combined a comprehensive immunophenotypic profiling with high throughput integration site analysis to study the fate of infused CAR T cells in 4 patients. We looked at samples from 2 patients with long CAR T persistence and 2 patients who had an early loss of CAR T cells.

4.1 CAR T cell clonal kinetics and phenotype post infusion

Firstly, we wanted to evaluate changes in the relative CAR T cell contribution to the CD3 cell compartment in vivo overtime following infusion of the cell product. All 4 patients showed similar kinetics of CAR T cell engraftment contributing up to 14-80% to the total CD3 cells at the time of maximal expansion (day 14) followed by contraction to 6-30% by day 30 (reflecting elimination of CD19 targets) and then low-level long-term persistence in Pt4 and Pt6 (0.1-0.2% at 36 and 28 months) (Figure 4.1.1b (top), Figure 4.1.1c and Figure 4.1.2b, d).

Next, we wanted to study the changes in phenotypic composition of CAR T cells in our patients overtime. We performed an immunophenotypic analysis to study different CAR T cell populations in both infused product and peripheral blood. We used CD3, CD62L, CD45RA and CD95 markers to differentiate between 4 memory subsets

(TSCM, TCM, TEM and TEMRA) and naïve T cells (Figure 4.1.1a). As Figure 4.1.1b (bottom) shows, the infused CAR T cells product in patients with long persisting CAR T cells (Pt4 and Pt6) was mostly composed of TCM (78-81% of CAR cells), with lower contribution of TEM (17-21%), and a small but detectable pool of TSCMs (1-2%). Immediately after infusion most CAR T cells changed their phenotype to TEM (71-75% at day 14), and this remained the most prevalent population at both early and late timepoints in these patients (47-60% at the latest follow up). We did not observe a substantial progression of CAR TEM towards terminally differentiated effector cells (TEMRA) in neither of the patients up to 840-1080 days after treatment. Notably, the small fraction of TSCMs present upon infusion did not disappear but became increasingly more represented overtime. At day 14 post infusion patients started with 0.5-3% TSCMs in their blood and reached 12-23% of TSCMs in the periphery at the latest follow up. Such observation seems to confirm that the CAR TSCM detected in the cell product are indeed endowed with high self-renewing potential in vivo.

For comparison, we performed similar phenotyping analysis on CAR T cells from 2 patients (Pt10 and Pt17) who lost circulating CAR T cells at 2- and 4-months post-infusion to investigate whether differences in product composition of T cell subsets might correlate with lack of persistence in patients. As shown in Figure 4.1.2c and Figure 4.1.3b (top), the CAR T cell products (particularly the one from Pt17) had a more differentiated phenotype as compared to the other group of patients with long term CAR T cell persistence, showing a lower proportion of TSCM and a higher contribution of TEM. At day 14 post infusion, patients had 0.08-0.16% TSCMs in the blood and 2-14% TCM. By day 30 post-infusion, CAR T cells from these patients showed a predominantly TEM phenotype (87-93%) with 2-6% of TCM whereas in Pt4

141

and Pt6 we observed a substantially higher preservation of the TCM (18% in both Pt4 and Pt6 at day 30) compartments (Figure 4.1.3b bottom).

Figure 4.1.1 Characterisation of CAR T cells in Pt4 and Pt6 overtime.

a) Gating scheme used for the identification of T cell subtypes used for following up CAR T cells and total CD3 cells. **b)** FACS plot showing CAR+ cells within CD3+ cells (top panels) as well as T cell
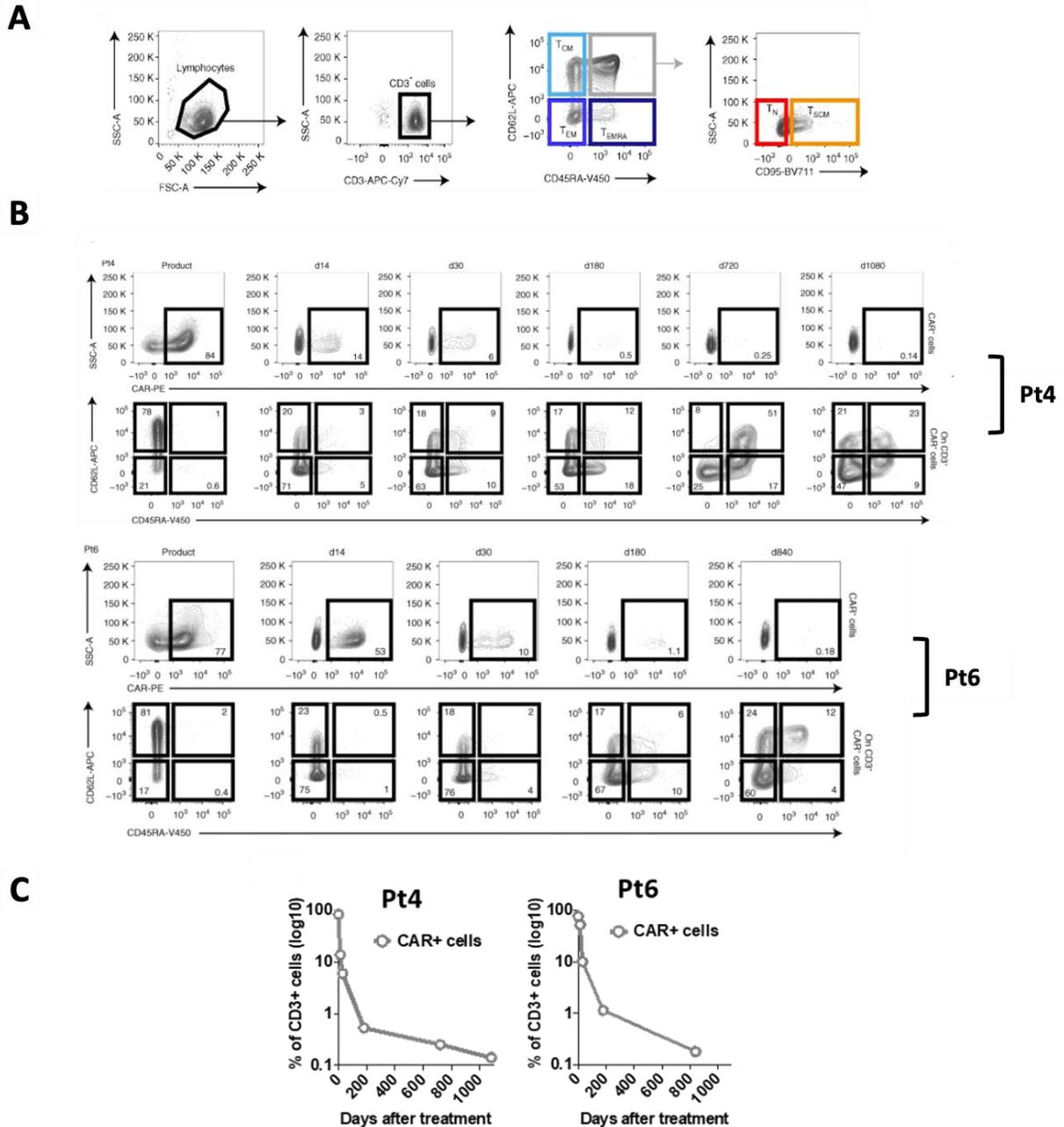
composition inside the CAR+CD3+CD45RA+CD62L+ population (bottom panels) for Pt4 and Pt6 overtime. Percentages of CAR cells and CD62L/CD45RA compartments are showed inside each gate. Timepoints of analysis are shown on top of each set of panels (d = days after CAR+ cells infusion). **c)** Percentage of CAR cells in the CD3+ cell compartment overtime in Pt4 (left panel) and Pt6 (right panel).
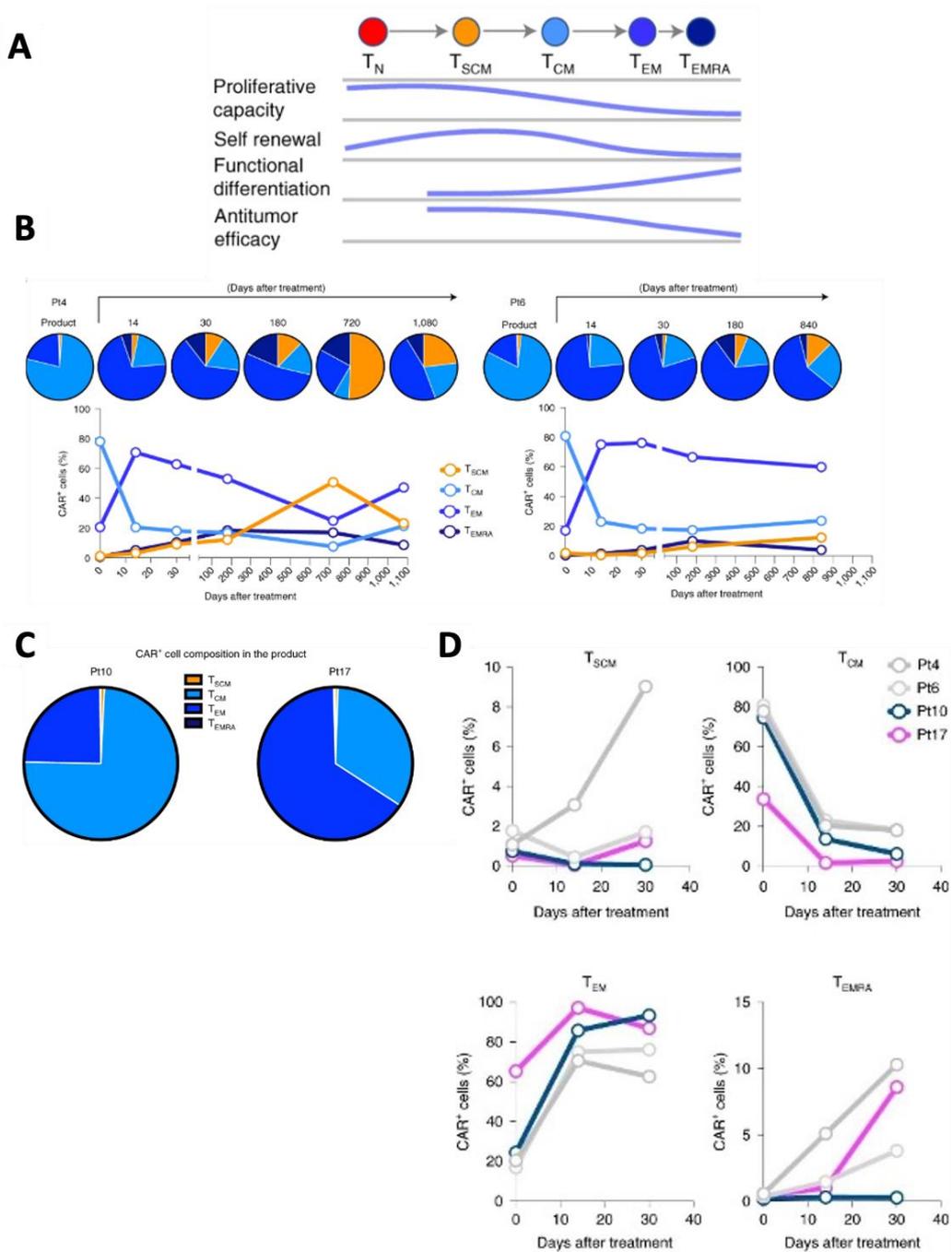
Figure 4.1.2 Kinetics of CAR T cells in 4 patients overtime.

a) Schematic representation of survival and expansion potential of the main T cell subpopulations (TN = naïve T cells, TSCM = T memory stem cells, TCM = T central memory cells, TEM = T effector memory cells, TEMRA = T effector memory CD45RA+ cells) **b)** Summary of data shown in Figure 4.1.1b bottom panels for CAR T cell composition overtime for Pt4 (left panel) and Pt6 (right panel). **c)** Pie charts showing the composition of the drug product in Pt10 and Pt17. **d)** Overtime contribution of each subpopulation to the CAR+ cell compartment in Pt10 and Pt17 vs Pt4 and Pt6.
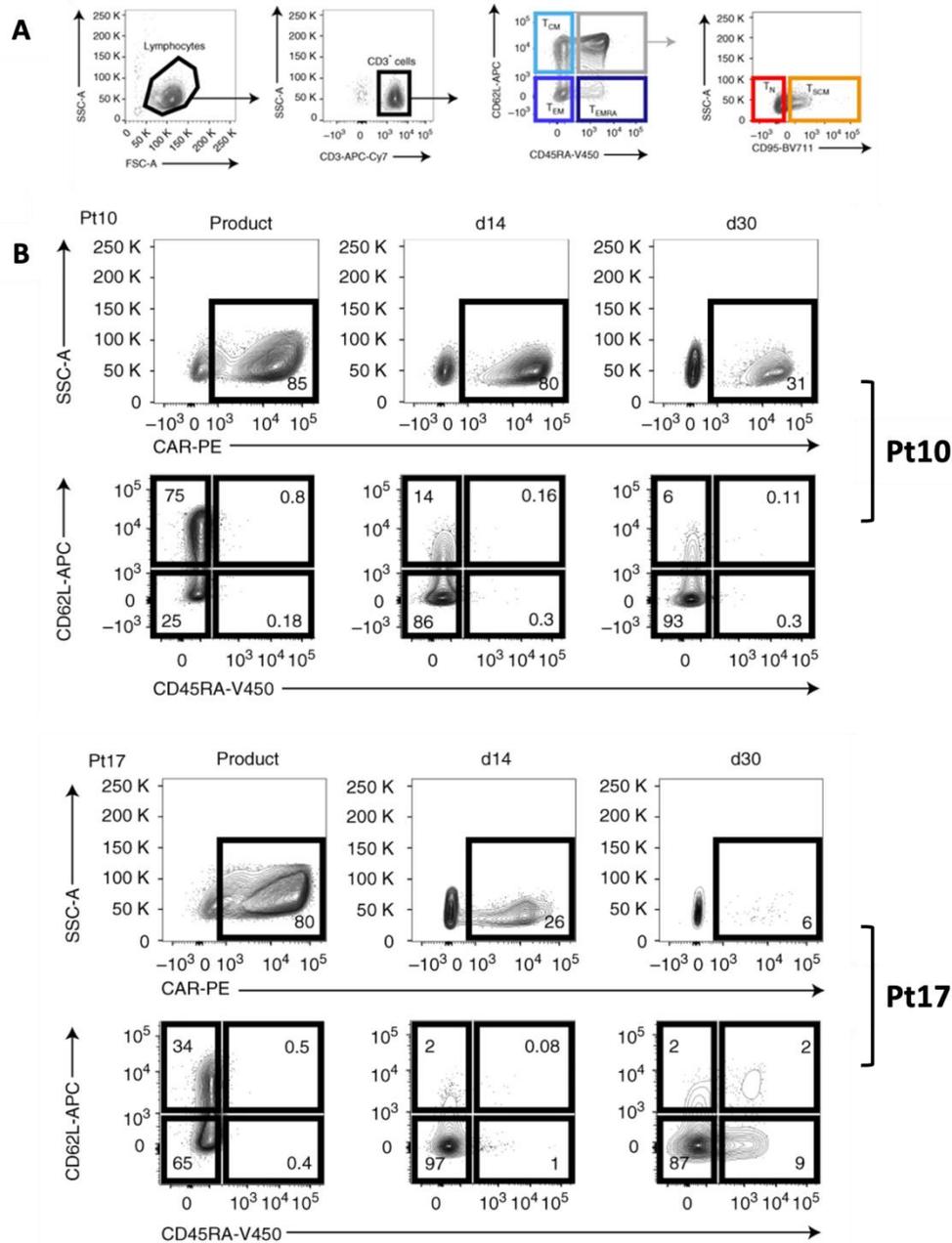
Figure 4.1.3 Characterisation of CAR T cells in Pt10 and Pt17 overtime.

a) Gating scheme used for the identification of T cell subtypes used for following up CAR T cells and total CD3 cells. **b)** FACS plot showing CAR+ cells within CD3+ cells (top panels) as well as T cell composition inside the CAR+CD3+ CD45RA+CD62L+ population (bottom panels) for Pt10 and Pt17. Percentages of CAR cells and CD62L/CD45RA compartments are showed

inside each gate. Timepoints of analysis are shown on top of each set of panels (d = days after

CAR+ cells infusion).

4.2 Integration site analysis of CAR T cells

We then used VISA to explore three main aspects of CAR T cell fate at the clonal level post infusion. Firstly, we wanted to understand what the difference in the clonal dynamics of CAR T cells during acute anti-tumour response phase is when CAR T cells reached maximum expansion (day 14, Figure 4.1.1b top and Figure 4.1.3b top) and steady state phase. Secondly, we wanted to investigate the clonal composition of long-term persisting CAR T cells. And lastly, we aimed at estimating which T cell population in the CAR T cell product is the one that contributes the most to the generation of long-term persisting CAR T cells in two our patients.

To address these points, we performed LAM-PCR combined with high throughput next generation sequencing on 3 FACS sorted CAR T cell populations (TSCM, TCM and TEM) isolated from the product of Pt4 and Pt6. We collected a total of 9,881 and 3,667 unique integrations from Pt4 and Pt6, respectively. Similarly, we collected 4,417 and 6,646 unique integration sites in Pt4 and Pt6, respectively, from FACS sorted CAR TSCM and TCM/TEM at 14 and 30 days after infusion (early response phase) as well as 1,676 and 1,930 unique integrations from total CAR cells of both individuals respectively at 180 and 720/840 days after treatment (long term survival phase) (Figure 4.2.1). At later timepoints we were unable to sort different CAR T cell subpopulations due to the low number of CAR T cells present in the circulation at that time.

146

We firstly examined the distribution of our integration sites in the genome of cells composing the infused product and CAR cells isolated in vivo at both early and late timepoints. As mentioned earlier, retroviral gene therapy vectors (such as lentiviral vectors used in this study) integrate semi-randomly with preferences for integrations in certain genomic loci and exhibit tissue-specific pattern of integration (Wu *et al.*, 2003; Mitchell *et al.*, 2004; Nowrouzi *et al.*, 2011; Aiuti *et al.*, 2013). Figure 4.2.2 shows that lentiviral vectors in the cell product of Pt4 and Pt6 integrated into or in the close proximity of many different genes most of which are common targets for this type of vectors including genes that are involved in DNA metabolism and repair (Figure 4.2.3 and Figure 4.2.4).

Pt4
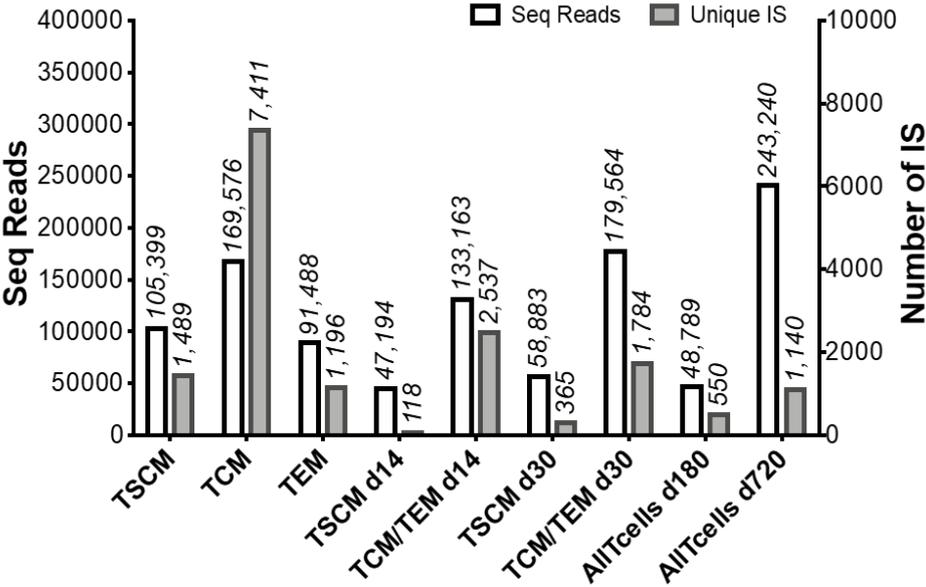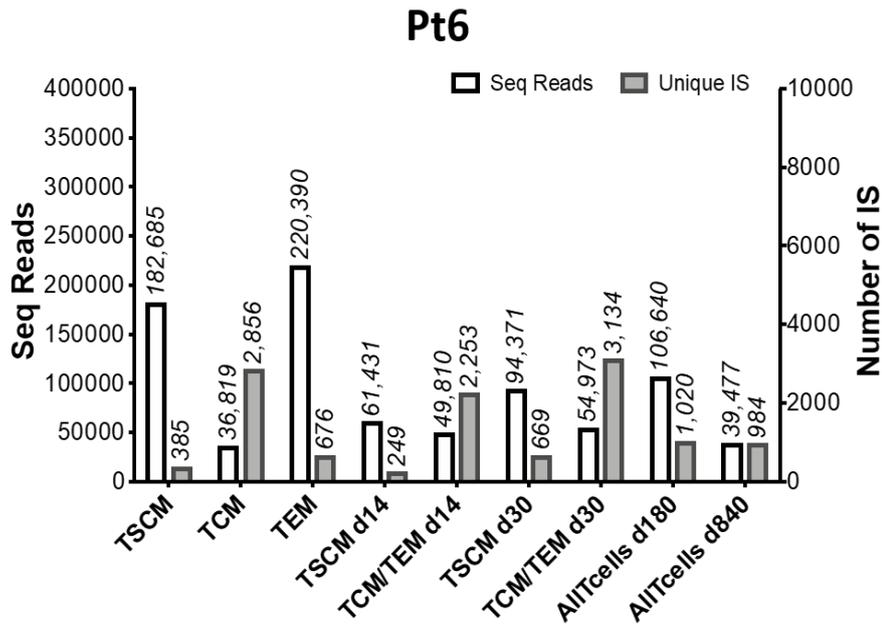
Figure 4.2.1 Integration sites collected from Pt4 (top) and Pt6 (bottom).

Summary of number of integrations sites (grey bars) and relative sequencing reads (white bars) in Pt4 (top panel) and Pt6 (bottom panel) (d = days after treatment). Unique integration sites on this plot in each population still contain some integration sites that are shared between populations. For example, both TSCM and TCM might have identical integrations in the same location that is shown here separately for each subpopulation. Therefore, total unique integrations in the product for Pt4 would be 10,096 based on the values from this figure, however, in the text we only report 9,881 because we filtered out all repetitive integrations and left only the true unique ones.
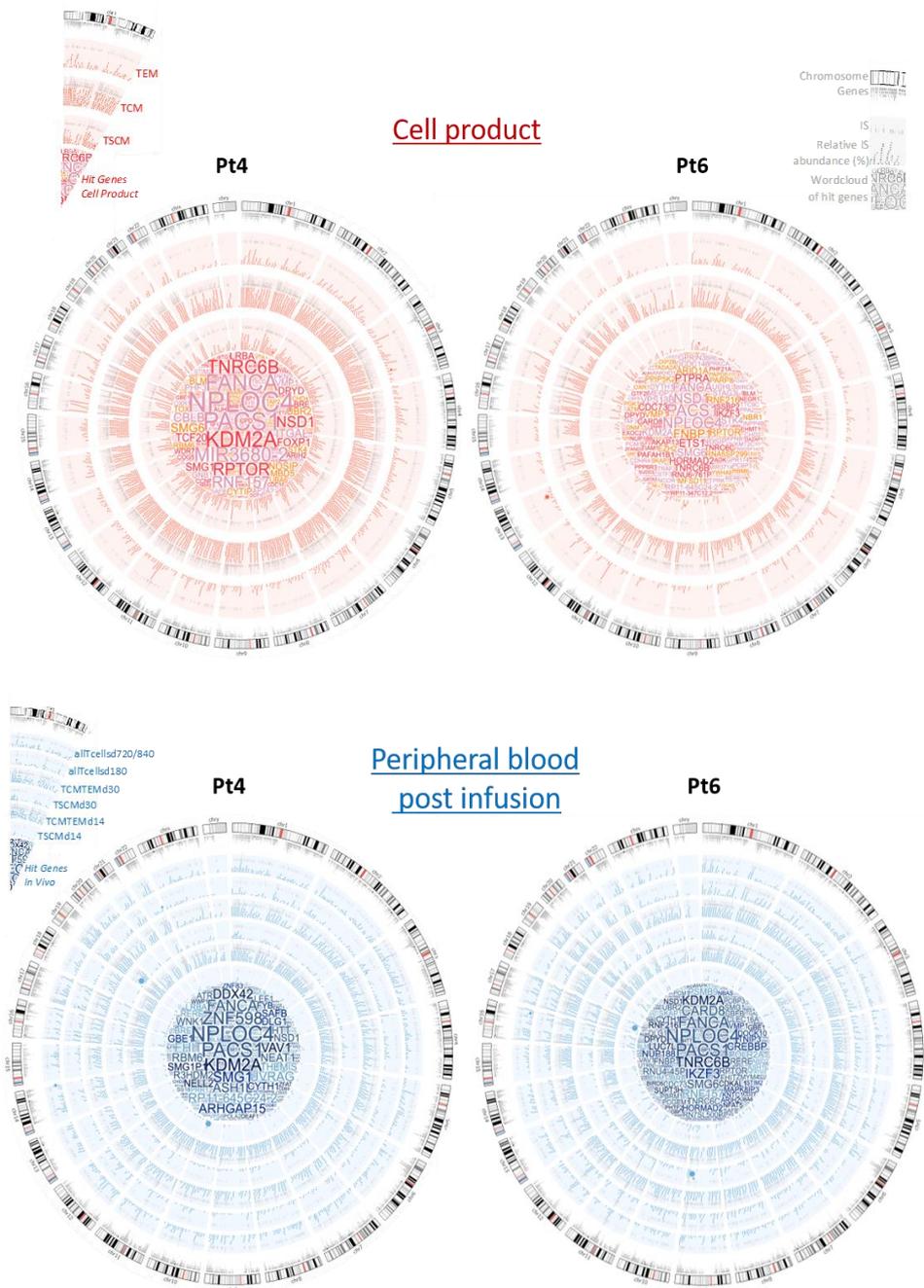
Figure 4.2.2 Distribution and abundance of IS collected in the product and after infusion overtime in Pt4 (left) and Pt6 (right).

Circos plot showing the genomic distribution of integrations in the product (top panels) and after infusion (bottom panels) in Pt4 (left panels) and Pt6 (right panels). Description of each concentric data track and timepoint (d = days after CAR cells infusion) is shown in the legends on the left of the circus plots. Description of the content of these plots is reported in the grey legend on the top right. Word clouds

showing top hit genes in the product and after infusion are shown at the centre of each plot (the bigger

the gene name the higher the number of integrations detected in its locus).

Pt4   Product
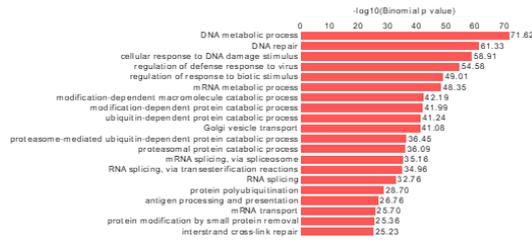
In vivo<1m
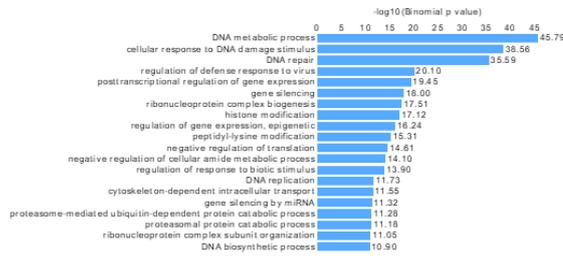
In vivo>6m
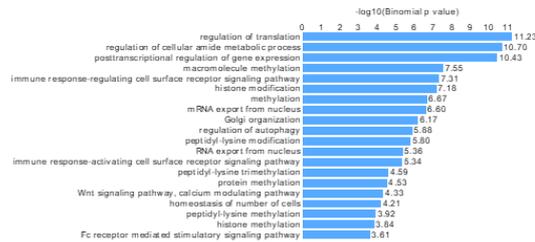
Figure 4.2.3 Gene categories relative to integration sites collected from Pt4.

Here we show relative gene enrichment analysis for top biological processes of hit genes relative to each word cloud in the product (red bars) or after infusion (blue bars) (significance reported as -log10 binomial p-value from one-tailed tests decreasing from top to bottom).

Pt6   Product

-log10(Binomial p value)

Category	Value
mRNA metabolic process	29.33
RNA processing	29.20
cellular macromolecule catabolic process	26.40
cellular response to DNA damage stimulus	25.54
mitotic cell cycle	24.39
modification-dependent macromolecule catabolic process	22.98
DNA metabolic process	22.69
mRNA processing	21.95
mitotic cell cycle process	21.04
DNA repair	20.85
regulation of mitotic cell cycle phase transition	20.84
modification-dependent protein catabolic process	20.21
ubiquitin-dependent protein catabolic process	19.94
posttranscriptional regulation of gene expression	19.80
regulation of cell cycle phase transition	19.29
viral process	18.49
mRNA splicing, via spliceosome	18.46
proteolysis involved in cellular protein catabolic process	18.45
RNA splicing, via transesterification reactions	18.37
multi-organism cellular process	18.23

In vivo<1m

-log10(Binomial p value)

Category	Value
DNA metabolic process	51.98
cellular response to DNA damage stimulus	50.75
DNA repair	46.37
mRNA metabolic process	38.09
ncRNA metabolic process	28.15
modification-dependent macromolecule catabolic process	27.69
nuclear division	26.94
Golgi vesicle transport	26.90
organelle fission	26.21
immune response-activating signal transduction	24.27
ribonucleoprotein complex localization	23.15
ribonucleoprotein complex export from nucleus	22.92
immune response-activating cell surface receptor signaling pathway	22.60
DNA-templated transcription, termination	22.11
immune response-regulating cell surface receptor signaling pathway	21.41
nuclear transport	21.32
nuclear export	21.29
snRNA metabolic process	20.77
nucleocytoplasmic transport	20.85
proteasome-mediated ubiquitin-dependent protein catabolic process	19.94

In vivo>6m

-log10(Binomial p value)

Category	Value
DNA repair	20.00
cellular response to DNA damage stimulus	19.76
DNA metabolic process	17.48
modification-dependent macromolecule catabolic process	10.15
ubiquitin-dependent protein catabolic process	9.41
modification-dependent protein catabolic process	9.37
nuclear export	9.27
nuclear transport	9.22
nucleocytoplasmic transport	8.89
proteasome-mediated ubiquitin-dependent protein catabolic process	7.95
proteasomal protein catabolic process	7.74
establishment of RNA localization	7.38
protein polyubiquitination	7.30
RNA transport	7.04
RNA localization	6.96
protein modification by small protein removal	6.64
DNA replication	6.61
protein deubiquitination	6.36
protein sumoylation	5.87
DNA geometric change	4.95

Figure 4.2.4 Gene categories relative to integration sites collected from Pt6.

Here we show relative gene enrichment analysis for top biological processes of hit genes relative to each word cloud in the product (red bars) or after infusion (blue bars) (significance reported as -log10 binomial p-value from one-tailed tests decreasing from top to bottom).

4.3 Clonal diversity is preserved in long-persisting CAR T cells

We next investigated the diversity of CAR T cells in the CAR T product of Pt4 and Pt6 by means of number and relative contribution of IS collected starting from FACS sorted TSCM, TCM or TEM. Because we observed a different frequency of TSCM, TCM and TEM in the cell product, to normalize for the subtypes contribution we used equal amounts of genomic DNA from each subpopulation for performing IS collection (Chapter 6 Methods, section 6.4 for detailed protocol). All subtypes showed high diversity with the biggest clones ranging from 2.1% to 14.7% of each population (except for one clone with integration contributing to around 30.4% of the reads collected from the TEM of Pt6) (Figure 4.3.1 top, Figure 4.3.2 top and Figure 4.3.3a). The TCM fraction of both individuals showed a higher number of individual integration sites and higher diversity than other two populations (Figure 4.3.1 and Figure 4.3.2 - Clonal diversity plot on the right). This could suggest that TCM might be more prone to transduction compared to TSCM and TEM, possibly because of a higher activation of these cells during the cytokine stimulation protocol used upon transduction and drug product expansion.

Next, based on our initial observation on the increasing relative contribution of TSCM in vivo compared to what infused in the cell product, we decided to investigate in higher details the clonal dynamics of this population after infusion. To this aim, we isolated the TSCM from the rest of the memory/effector cell fraction in the early response phase when the number of CAR T cells in circulation still allowed FACS sorting of CAR T cell subsets. As shown in Figure 4.3.1 bottom, 4.3.1 bottom and 4.3.3b we could isolate CAR TSCM at 14 and 30 days after infusion and compare their profile with the rest of the CAR population (primarily composed by TCM/TEM) at the same timepoints.

We observed that the clonal diversity of TSCM (measured either as Shannon or Gini/Simpson diversity as shown on Figure 4.3.1 and Figure 4.3.2 bottom or 4.3.3a, b, respectively) dropped substantially from the infused cell product to the early response phase. Most of the TSCM isolated in vivo were composed by individual clones bearing integrations whose reads ranged from 89.0% to 96.9% of total reads count. As Figure 4.3.4 shows, the number of cells and the number of individual integration sites collected do not significantly correlate with clonal diversity of a given population, suggesting that such observation is instead likely a reflection of clonal bursts and uneven clonal expansions events occurring in TSCM during the early response phase (Figure 4.3.1 and Figure 4.3.2 bottom). At the same timepoints, the clonal composition of CAR TCM/TEM cells remained highly polyclonal (Figure 4.3.1 and Figure 4.3.2 in vivo top left), suggesting a more homeostatic pattern of expansion upon infusion as compared to TSCM subset.

We then analysed in both patients the diversity of the total CAR cell composition at later timepoints (180 and 720/840 days after infusion). As shown before (Figure 4.1.1c), there was a substantial drop in the number of circulating CAR T cells in both patients at later timepoints. However, clonal diversity in these long persisting CAR T cells remained high and was not affected by clonal bursts of TSCM subset at early time points (Figure 4.3.1 and Figure 4.3.2 in vivo top right, and Figure 4.3.3c). This suggests that long-term CAR T cells were capable of long-term "homeostatic" survival and that long-term maintenance of CAR T cells in our patients was not dependent on the expansion of few large clones carrying non-physiological/aberrant IS.

Next, we compared all datasets from early or late timepoints and performed unsupervised clustering according to IS similarities (Figure 4.3.5a). We could observe

that IS isolated at 14 and 30 days after infusion showed the highest similarities and clustered separately from the clonal datasets of 180 and 720/840 days, suggesting that clones active in the early response phases were less likely to be recaptured at later times after treatment. We then wanted to estimate the relative number of clones that are constituting circulating CAR cells by calculating the probability to recapture identical integrations between CAR cells at multiple timepoints, using mark-recapture statistics as previously described (Biasco *et al.*, 2016). The results suggest that at least around 34,000 and 56,000 CAR clones might still be in circulation in Pt4 and Pt6 respectively up to 24/28 months after infusion (Figure 4.3.5b). We then evaluated the probability of recapturing identical CAR TSCM clones between early and later timepoints (Figure 4.3.5c). The expanded clones were different at each timepoint and some of them were also detected in the early CAR TCM/TEM population suggesting that this expansion was accompanied by preservation of the TSCM differentiation potential towards memory and effector cells. However, these large clones do not seem to have contributed substantially to the pool of long-term surviving CAR T cells (Figure 4.3.5c).

Interestingly, months to years after infusion we did not observe any major skewing in the distribution of IS towards gene categories related to T cell growth, activation, or survival. As mentioned before, insertional mutagenesis is one of the major drawbacks of retroviral gene/cell therapy. It was reported that one of the patients who underwent a CD19 CAR T cell therapy to treat their chronic lymphocytic leukaemia, had a vector insertion in TET2 gene that led to clonal dominance (Fraietta, et al., 2018). It turned out that this event of insertional mutagenesis was beneficial for the patient and helped TET2 CAR T clones to clear leukaemia. Hence, we decided to check if our patients

had any integrations inside TET2 gene and if this was affecting in any way the clonal survival and/or activity of our CAR T cells. We identified 12 IS in total in our patients in TET2 gene (Figure 4.3.6a). None of these clones showed signs of expansion neither in the product nor in vivo contributing to a maximum of 0.044% of sequencing reads out of the total collected in each sample/timepoint (Figure 4.3.6b). Collectively, these data strongly suggest that insertional mutagenesis did not take part in clonal selection of CAR T cells in these patients upon either early response or long-term survival.

We then performed the same analysis for patients who lost their CAR T cells after the early phase (Pt10 and Pt17). We collected 4,633 (Pt10) and 2,823 (Pt17) integration sites in TSCM, TCM and TEM subsets from the product, and 2,295 and 1,020 integration sites, respectively, in TSCM and TCM/TEM during the first month of the follow up (Figure 4.3.7). These numbers were similar to the number of integration sites we collected for patients with long-term persisting CAR T cells (Pt4 and Pt6). We evaluated the relative contribution of each integration site (Figure 4.3.8 and Figure 4.3.9) and measured the clonal diversity of the sorted CAR T cell subsets from the product and peripheral blood post infusion. Figures 4.3.8 (top) and Figure 4.3.9 (top) show that both patients had a lower level of clonal diversity in CAR T cell product in all T cell subsets as compared to patients with long-term CAR T cells persistence, especially in TSCM population. After infusion, there was a clonal burst of CAR TSCM during the first month (Figures 4.3.8 bottom and 4.3.9 bottom). However, in contrast to the patients with long-term persistence of CAR T cells, the diversity of the TCM/TEM compartments rapidly dropped over the first month in these patients before CAR T cell disappearance.
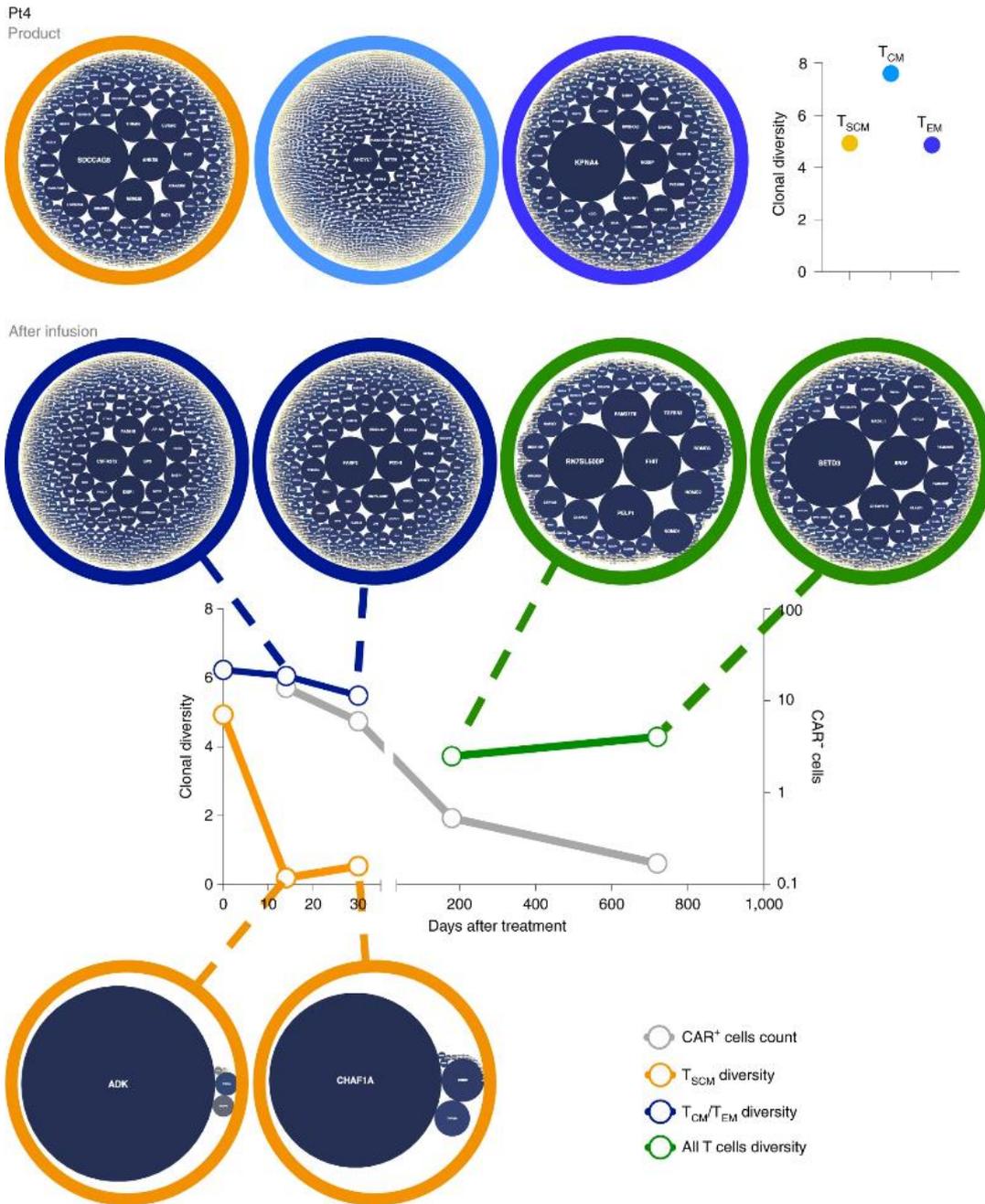
Figure 4.3.1 Diversity of IS in different T cell subtypes overtime from Pt4.

Top panels: for each product subpopulation analysed coloured circles (TSCM = orange, TCM = light blue, TEM = blue) contain bubble plots of clones contributing >0.01% to the total population. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside the bubble. The plots on the right show Shannon Diversity Index for each subpopulation. Bottom panels: plots showing Shannon Diversity Index (left y-axis) of IS overtime in TSCM (orange

lines) and in TCM/TEM (dark blue) and in all T cells (green lines). The grey lines show the percentage of CAR cells overtime (right y-axis) For each timepoint analyzed coloured circles contain bubble plots of clones contributing >0.01% to the total population at each timepoint. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside each bubble.

Figure 4.3.2 Diversity of IS in different T cell subtypes overtime from Pt6.

Top panels: for each product subpopulation analysed coloured circles (TSCM = orange, TCM = light blue, TEM = blue) contain bubble plots of clones contributing >0.01% to the total population. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside the bubble. The plots on the right show Shannon Diversity Index for each subpopulation. Bottom panels: plots showing Shannon Diversity Index (left y-axis) of IS overtime in TSCM (orange

lines) and in TCM/TEM (dark blue) and in all T cells (green lines). The grey lines show the percentage of CAR cells overtime (right y-axis) For each timepoint analyzed coloured circles contain bubble plots of clones contributing >0.01% to the total population at each timepoint. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside each bubble.

Figure 4.3.3 Relative abundance and diversity of integrations sites collected from Pt4 and Pt6.

Scattered dot plots showing relative abundance of IS in the product **(A)** and at early **(B)** or late **(C)** timepoints after treatment in Pt4 (plots on the left) and Pt6 (plots on the right) (d = days after treatment, m = months after treatment). Mean percent abundance is shown as a dotted line for each sample. Number of events in each dataset is equal to what reported in **Figure 4.2.1**.
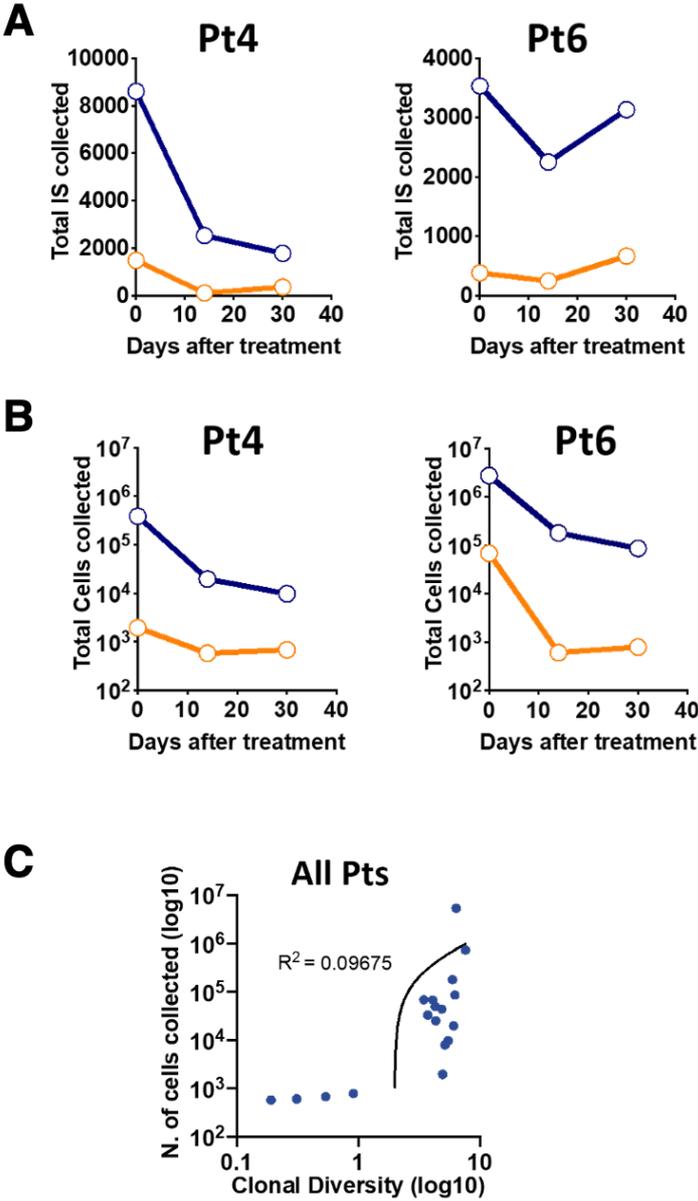
Figure 4.3.4 Longitudinal plots showing number of IS collected (a) and number of cells collected (b) in T cells at early timepoints in Pt4 and Pt6.

On both **(a)** and **(b)** plots TSCM are shown as orange lines and TCM/TEM are shown in dark blue at early timepoints for Pt4 (left panels) and Pt6 (right panels). The plot in **(c)** shows the correlation observed between number of cells collected (y-axis) and clonal diversity for all samples and both patients (x-axis, Shannon Diversity Index) shown as blue dots. Interpolation with best fit curve and R squared value are shown in black.
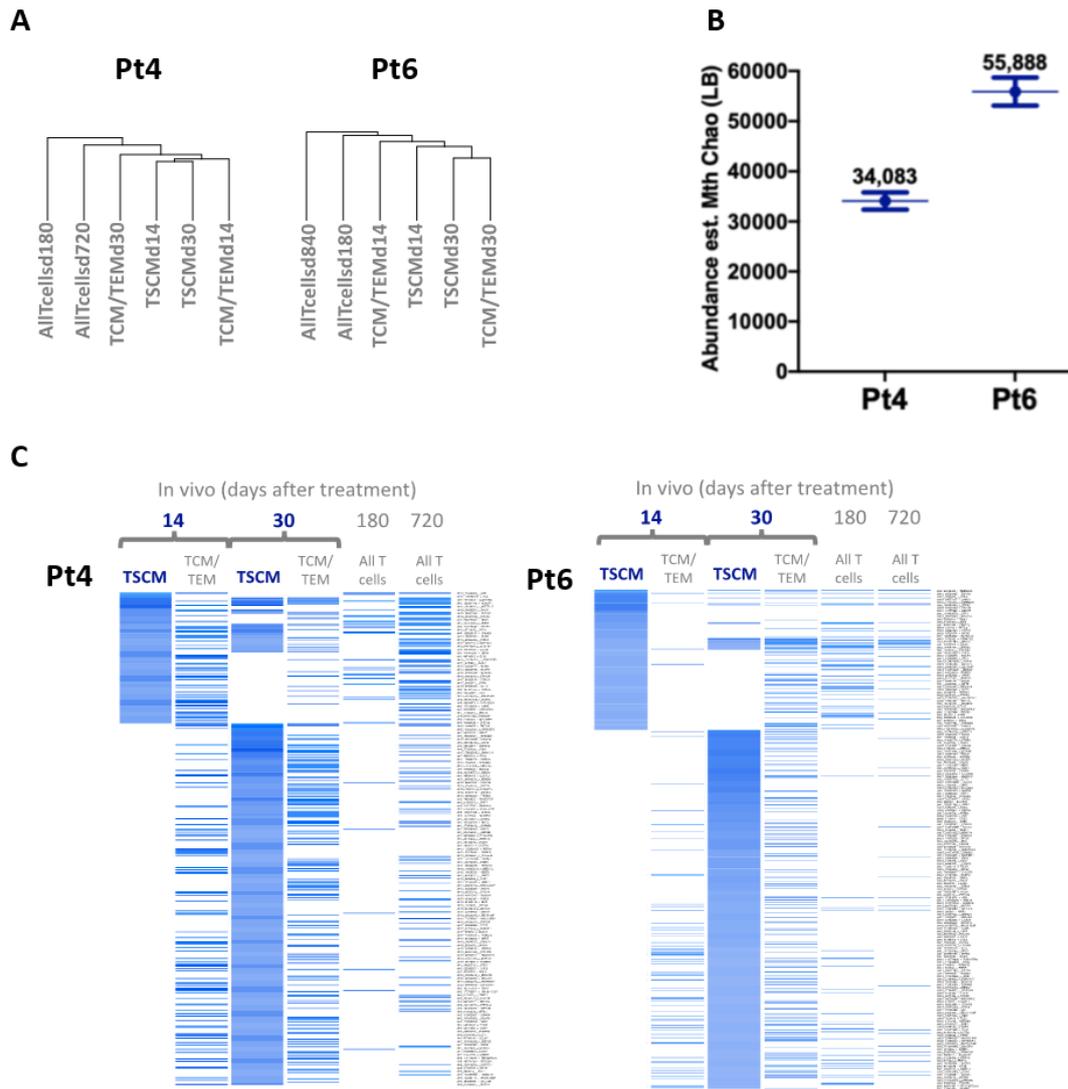
Figure 4.3.5 IS sharing between T cell subsets and estimation of the clonal pool in Pt4 and Pt6.

a) Unsupervised clustering based on IS sharing among subpopulations overtime in Pt4 (left panel) and Pt6 (right panel). **b)** Estimation of the number of clones composing the CAR cell population in each patient based on the Mth Chao (LB) model applied to IS recapture probability. Shown are individual values and relative standard errors from the abundance model fit estimation (for Pt4 abundance=34,083 and standard error = 1,699, for Pt6 abundance = 55,888 and standard error = 2,795). **c)** Heatmaps showing tracking and relative contribution of TSCM clones (IS, rows) at d14 and d30 recaptured in other

T subtypes and timepoints (columns) in Pt4 (top panel) and Pt6 (bottom panel). Intensity of blue is proportional to relative clone size (IS abundance).
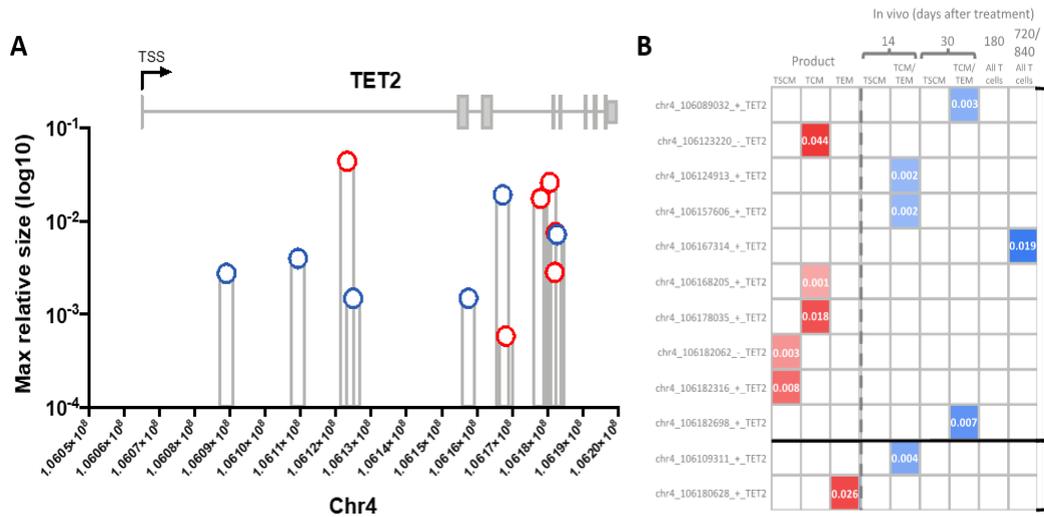
Figure 4.3.6 Distribution of IS in proximity of TET2 gene.

Distribution of IS in proximity of TET2 in both patients combined. The x-axis displays chromosomal coordinates (bp) while the y-axis the IS abundance (log10 of percentage) calculated within the population and timepoint where it has been observed. The dark blue bars show IS detected in the patients while red bars IS detected in the cell product. The transcription start site (TSS), exons (grey boxes) and introns (grey lines) of the TET2 gene are shown on top of the plot in their respective chromosomal localization. **b)** Heatmap showing tracking and relative contribution of the IS shown in **panel A** for each patient in the product (in red) or after infusion (in blue). Relative IS abundance (percentage) calculated within the population and timepoint where it has been observed is reported inside each colored field (white field = not detected).
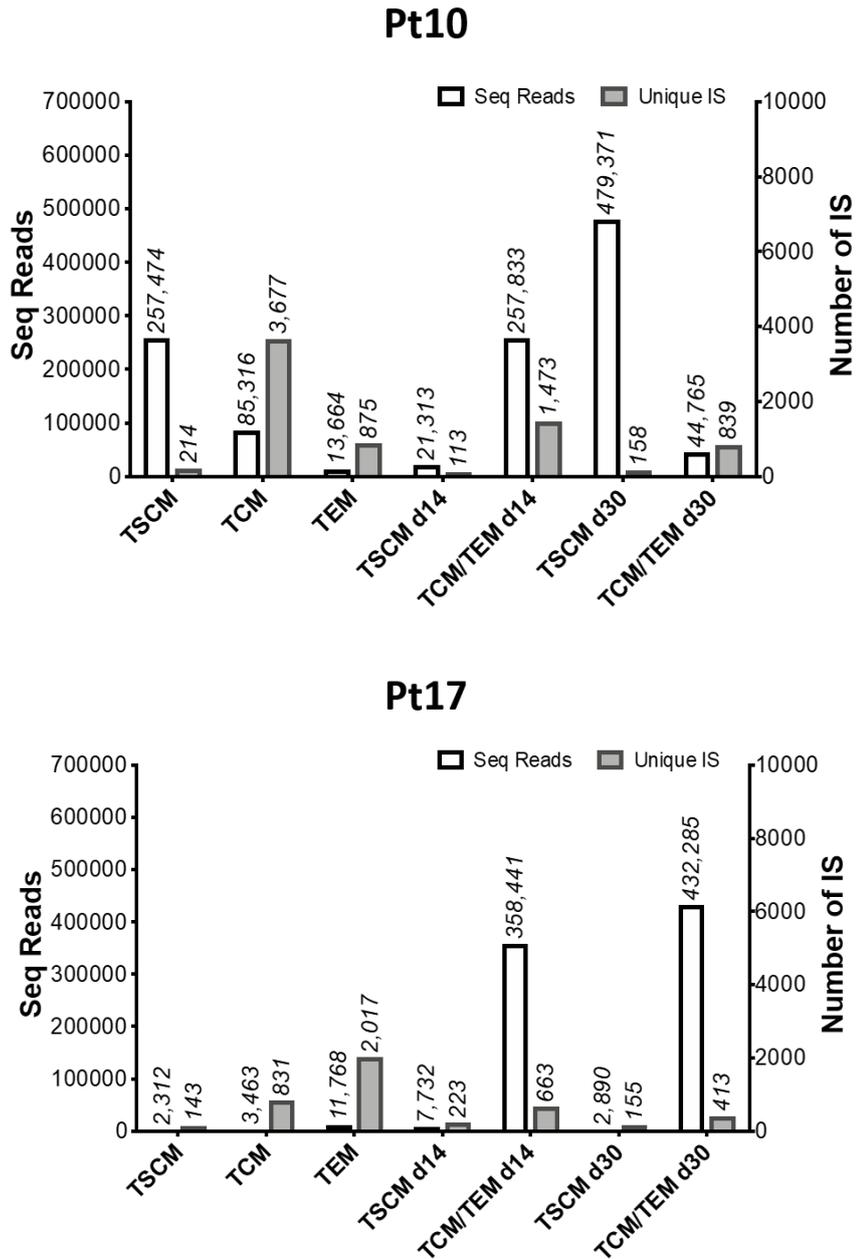
Pt10

Pt17

Figure 4.3.7 Integration sites collected from Pt10 and Pt17.

Summary of number of IS (grey bars) and relative sequencing reads (white bars) in Pt10 (top panel) and Pt17 (bottom panel) (d = days after treatment). Unique integration sites on this plot in each population still contain some integration sites that are shared between populations. For example, both TSCM and TCM might have identical integrations in the same location that is shown here separately for each subpopulation. Therefore, total unique integrations in the product for Pt17 would be 2,991 based on the

values from this figure, however, in the text we only report 2,823 because we filtered out all repetitive integrations and left only the true unique ones.
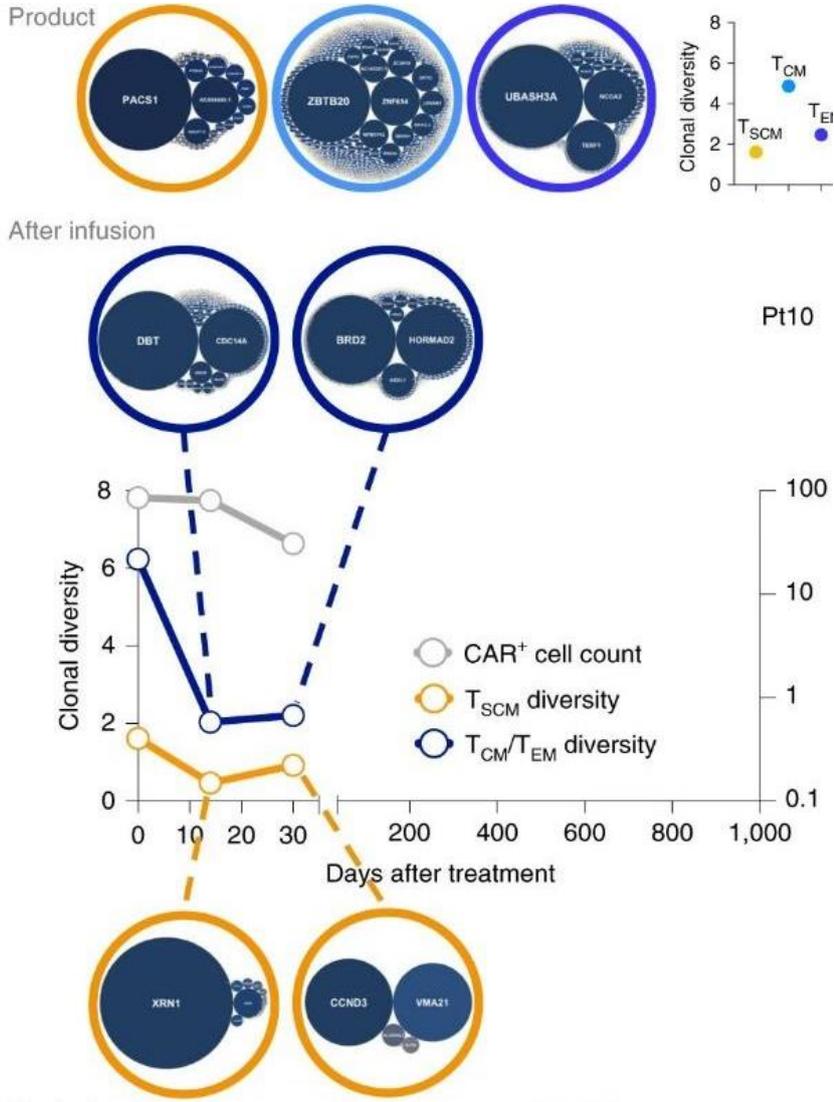
Figure 4.3.8 Scattered dot plots showing relative abundance of IS in the product (top) and at early timepoints (bottom) after treatment in Pt10 (d = days after treatment, m = months after treatment).

Mean percent abundance is shown as a dotted line for each sample. Number of events in each dataset is equal to what reported in **Figure 4.3.7**. Top right panel shows a diversity of drug product and CAR cells after infusion in Pt10. Top section: for each product subpopulation analysed coloured circles

(TSCM = orange, TCM = light blue, TEM = blue) contain bubble plots of clones contributing >0.01% to the total population. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside the bubble. The plots on the right show Shannon Diversity Index for each subpopulation. Bottom section: plot showing Shannon Diversity Index (left y-axis) of IS overtime in TSCM (orange lines) and in TCM/TEM (dark blue). The grey lines show the percentage of CAR cells overtime (right y-axis) For each timepoint analysed coloured circles contain bubble plots of clones contributing >0.01% to the total population at each timepoint. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside each bubble.

Figure 4.3.9 Scattered dot plots showing relative abundance of IS in the product (top) and at early timepoints (bottom) after treatment in Pt17 (d = days after treatment, m = months after treatment).

Mean percent abundance is shown as a dotted line for each sample. Number of events in each dataset is equal to what reported in **Figure 4.3.7**. Top right panel shows a diversity of drug product and CAR

cells after infusion in Pt10. Top section: for each product subpopulation analysed coloured circles (TSCM = orange, TCM = light blue, TEM = blue) contain bubble plots of clones contributing >0.01% to the total population. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside the bubble. The plots on the right show Shannon Diversity Index for each subpopulation. Bottom section: plot showing Shannon Diversity Index (left y-axis) of IS overtime in TSCM (orange lines) and in TCM/TEM (dark blue). The grey lines show the percentage of CAR cells overtime (right y-axis). For each timepoint analysed colored circles contain bubble plots of clones contributing >0.01% to the total population at each timepoint. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside each bubble.

4.4 Origin and maintenance of long-term persisting CAR T cells

Our third aim was to study the origin of long-term persisting CAR T cells and to see if we will be able to identify CAR T cell subset in the product that contributed the most to the maintenance of long-term CAR T cells in 2 patients. The depth of our analysis and the number of IS collected in the product and in vivo allowed for the first time in such clinical scenario to detect of a substantial number of identical IS in samples collected before and after infusion. The sharing of IS among and between the cell product subtypes and the in vivo samples at different timepoints is represented in the network of correlations shown in Figure 4.4.1. These plots quantify integration site sharing between all samples analysed showing in blue dots the paired comparisons where integration site sharing was more significant (positive correlation).

In both patients we observed a relatively high IS sharing among the T cell subpopulations composing the cell product suggesting that differentiation has occurred during the in vitro expansion. Notably, we detected a relatively high number of identical

170

integration sites that were shared between TSCM in the product and TSCM in vivo at 14- and 30-days post infusion. TSCM from the product and TSCM from blood samples were isolated, processed and analysed independently, which significantly reduces cross-contamination issue and allows us to assume that our data is genuine and is not merely due to technical errors. This suggests that, prior to infusion during manufacturing process of the product, these clones must have preserved their TSCM phenotype while expanding their numbers enough so that we were able to detect identical clones in the retention product material and later in the in vivo samples. This observation is in line with previous report, which showed that CAR TSCM have the greatest rate of expansion in the product compared to other subsets (Schmueck-Henneresse *et al.*, 2017). This also tells us that those clones have also substantially contributed to the TSCM pool active during the early response. TSCM IS sharing with TCM/TEM up to Day 30 after treatment ranged between 0.3% to 23% (averaging 8.3%), suggesting active differentiation in circulation of TSCM originally contained in the drug product. In contrast, TCM and TEM in the product shared on average a maximum of 2.4% of integration sites with the same subtypes up to early timepoints after therapy (range 0.09 - 7.1%), which is in line with the more limited differentiation potential of these memory cell types as compared to TSCM.

Despite the detection of identical integration sites between the cell product and cells isolated from patients at late timepoints, integration site numbers were not sufficient to reach significance, possibly due to inherent sampling limitations combined with the small number of CAR T cells remaining in circulation in these patients at 6 to 24/28 months after treatment. Nonetheless, the numbers of shared integration sites allowed us to investigate the relative contribution of the product subsets to the early and late

phases after infusion. As shown in Figure 4.4.2, in all 4 patients, regardless of long-term persistence at early time points the majority of the shared IS were derived from the TSCM CAR T cells in the product suggesting this compartment has a key role in early expansion. Similarly, in the patients with long-term persistence, the subpopulation from the product contributing the most to the clonal pool at late time points was the TSCM, from which we collected 46.6% to 60.5% of the clones detected both in vitro and after infusion. Considering a very low proportion of CAR T cells present upon infusion (1-2% of total CAR T cells Figure 2.7.1) compared to other memory subsets, this level of contribution is substantial. On the other hand, these data align with the fact that long-surviving CAR cells have become progressively enriched in TSCM. Moreover, the highly diverse clonal composition preserved up to 28 months after treatment again suggests a major contribution of self-renewing cells endowed with "homeostatic" survival potential like TSCM.
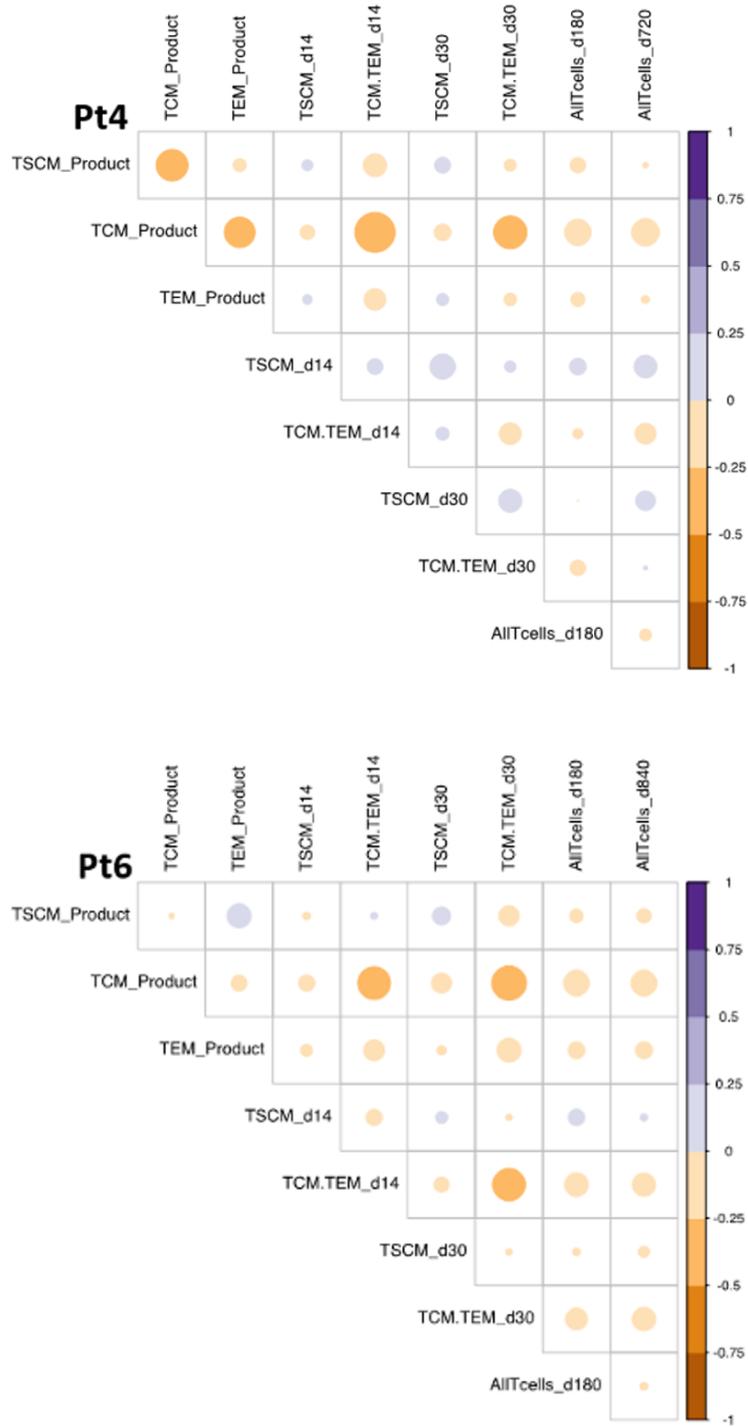
Figure 4.4.1 Comparison of IS distribution pre- and post-infusion.

Correlograms showing upper triangle of correlation matrix among IS datasets from Pt4 (top panel) and Pt6 (bottom panel). Positive correlations are shown in blue while negative in red. Size of circles is proportional to the correlation value.
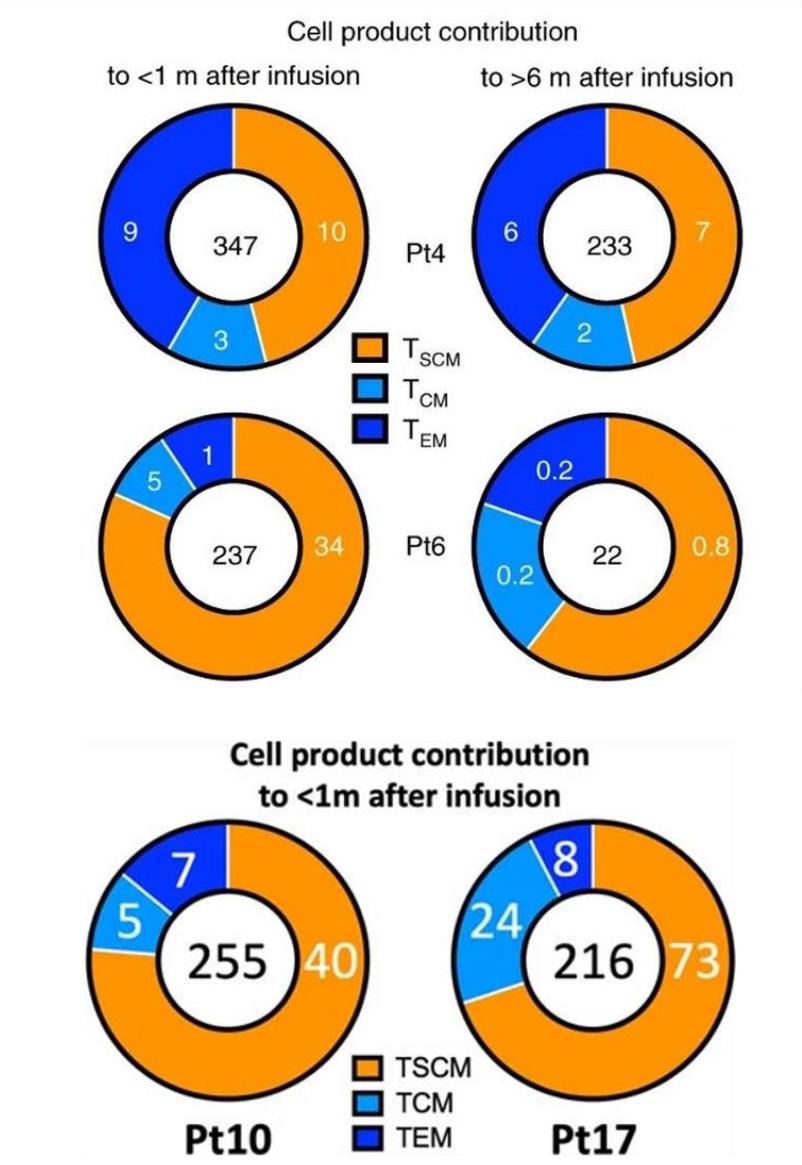
Figure 4.4.2 Ring plots showing relative contribution from each subtype (coloured section of the ring) to the pool of IS detected in the product and at early (left plots) or late (right plots) timepoints in Pt4 and Pt6 (top panels) and Pt10 and Pt17 (bottom panels).

The total number of IS captured both in the product and after infusion is shown inside each plot. The relative percentage of IS belonging to each T cell subtype of the product that were shared with samples after infusion is shown in white inside each section of each ring plot.

5 Discussion

5.1 Using VISA for clonal tracking studies

In this work we exploited vector integration site analysis to track immune cells in patients to unravel the existence of long-term lymphoid progenitors in humans and to investigate dynamic changes in kinetics of CAR T cells during early anti-leukemic response phase and long-term surveillance phase.

To retrieve IS from patient samples, we employed LAM-PCR technology. This method has been proven to be a reliable technique for clonal tracking studies over years by our group and others. In case of CAR T cell clone tracking, VISA allowed us to track clones at very low frequency (0.1%) at late time points, when cell numbers were insufficient for FACS sorting of CAR T cells and subsequent TCR sequencing. Moreover, clonal IS marking is more stable than TCR sequencing as it is independent of the effects of biological variables such as thymic involution, immune tolerance or extrathymic rearrangements, making VISA an important additional tool in studying T cell clones. Nonetheless, it should be noted that it is currently impossible to precisely discriminate individual clones based on the assumption 1 IS = 1 clone, since each genetically engineered cell may carry multiple copies of integrated vector so any biological interpretation of IS data must be based on data trends rather than on the tracking of specific individual IS or clones.

There are other methodological constraints that should be considered when interpreting the results of this study. We were limited in the amount of patient material we were able to use in this project, which significantly reduced the number of different

assays we were able to perform for a given sample and each timepoint. As a result, despite clear evidence of population trends, statistical significance of intragroup variance or non-parametric comparison among two groups cannot always be reached. Sample viability and quality were not always consistent between freshly analysed samples and samples that were cryo-preserved resulting in some potential differences in data recovery (e.g., different number of IS collected). In addition, all samples analysed were subjected to WGA, which might lead to uneven amplification of some DNA regions compared to others, and subsequently loss of some integration sites. Residual cross-contamination among FACS-sorted populations is another factor that can, in part, affect the interpretation of the results of high-throughput molecular studies. However, it should be noted that in our study the observed high-level sharing of IS from TN with other cell subtypes (up to 57.2%) far exceeded the potential effects of insufficient cell purity after sorting (estimated TN purity upon re-analysis of sorted material up to 95.6%). More importantly, identical and highly abundant IS were detected between samples that were collected, isolated, processed and sequenced at independent times.

For future VISA, we aim to use nrLM-PCR technology recently established in our lab. nrLM-PCR allows unbiased retrieval of vector integrations because it does not rely on restriction enzymes. It is also much faster as it does not involve a linear amplification PCR step, which in case of clinical monitoring is highly beneficial as it allows faster sample processing. A shorter protocol also allows more samples to be processed at once.

Overall, using VISA method we were able to unravel novel insights about human haematopoiesis and CAR T cell kinetics during the clearance of leukaemia stage and a long-term survival phase.

5.2 Long-term lymphoid progenitors independently sustain naïve T and NK cell production in humans

In this part of the project, we used series of immunophenotypic, functional and molecular studies to show that a de novo production of genetically modified T and NK cells can be maintained in humans for over a decade in the absence of vector positive HSC. We interpreted the consistent absence of genetically modified myeloid and B cells in our patients overtime as a reflection of the early loss of vector positive HSC. This observation is in line with the fact that these patients did not have any myelosuppressive bone marrow conditioning administered prior to gene therapy.

Here, we showed that T cells in these individuals have a true naïve phenotype and in comparison, to memory subsets are also functionally naïve. We also showed that naïve T cells in these patients shared the least of identical TCR clonotypes with other subsets suggesting an ongoing production of new T cells with unique TCR specificities. Conversely, it is interesting to notice that we observed an equal level of sharing of identical integration sites between naïve T cells and memory subsets. We hypothesise that this could be because vector positive long term lymphoid progenitors generate early T cell progenitors with identical integration sites that then undergo somatic rearrangements and maturation processes in the thymus to yield naïve T cells with identical integrations sites but unique TCR specificities. Based on our estimates, there is a high number of T cells clones (at least 2000-6000) that is responsible for

maintaining a de novo T cell production at latest timepoints, suggesting that it is highly unlikely that maintenance of the T cell pool in these individuals is supported by few residual vector positive HSC.

In this study, we were unable to directly identify the precise physical location in the body where these vector positive long term lymphoid progenitors reside. We can speculate that, because of the consistent lack of vector positive B cells as well as suboptimal numbers of genetically modified NK cells observed in our patients, these long-term lymphoid progenitors reside in the thymus. It is possible that these Lt-LP are at the multipotent developmental stage that would still allow them to generate both B cells and sufficient numbers of NK cells under the right conditions. Several groups identified and proposed different phenotypic descriptions of such thymic progenitors with all of them being able to harbour either multi-lymphoid T/B/NK potentials or multilineage lymphoid/myeloid/erythroid potentials (Haddad *et al.*, 2006; Six *et al.*, 2007; Hao *et al.*, 2008; Kohn *et al.*, 2012; Lavaert *et al.*, 2020). However, thymic environment is tailored to support the development and maturation of T cells, whereas both B and NK cells require different environmental conditions for successful differentiation and maturation. This would explain why genetically modified NK cells in these individuals did not reach normal physiological levels and why there is no vector positive B cells in their circulation. Another possible hypothesis is that at some stage along the differentiation hierarchy there could be a segregation of lymphoid progenitors that give rise to B cells and progenitors that harbour dual potential for T/NK cell development. In this case, one can speculate, that commitment to a particular lineage (T or NK) could again be dependent on the surrounding environment, with thymic environment pushing progenitors towards T cell lineage, whereas similar progenitors

in secondary lymphoid organs would be able to sustain normal physiological levels of NK cells. In case of our patients, suboptimal but detectable NK cell production could still be in place because the NK niche is empty (no gene corrected HSC in the bone marrow), therefore activating an alternative secondary pathway to replenish the NK cell niche. If vector positive Lt-LP progenitors resided in the bone marrow, then we would expect to observe both gene corrected B cells as well as normal physiological numbers of gene modified NK cells in these patients, since it has been shown by various studies that lymphoid progenitors in the bone marrow are multipotent cells with multilineage potential.

We here also showed, using comprehensive immunophenotyping that NK cells in these patients have non-memory/non-adaptive phenotype, which suggests that these NK cells might be freshly produced. This is in contrast with data from patients with GATA2 mutations, that have defects in immature NK cell (CD56bright) production and persistence of memory/adaptive NK cells (Schlums *et al.*, 2017). In our patients, we observed a high percentage (28-70%) of CD56dim CD16- NK cells. In healthy individuals, CD56dim CD16- NK cells represent a minor subset (Poli *et al.*, 2009), whereas in paediatric bone marrow, leukemic patients (Stabile *et al.*, 2015) and early after hematopoietic stem cell transplantation (Di Vito, Mikulak and Mavilio, 2019) this subset is observed in higher percentages. Not much is known yet about the functional properties of this subset, but it has been suggested that these unconventional CD56dim CD16- NK cells represent an alternative or additional stage of NK cell differentiation (Di Vito, Mikulak and Mavilio, 2019). We observed a reduced percentage of CD56dim CD16+ NK cell population in our patients, which is the predominant NK cell population in the peripheral blood, lungs, and spleen of healthy donors (Dogra *et*

al., 2020). Our data, however, is in line with previous findings, where SCIDX-1 patients were shown to have reduced percentage of CD56dimCD16+ NK cell population years after haematopoietic stem cell transplant and gene therapy treatments (Touzot *et al.*, 2015). Overall, due to the low NK cell numbers and a lack of widely accepted phenotypic characterisation of memory/adaptive NK cells, we were unable to reach a definitive conclusion on whether NK cells in these individuals are freshly produced.

Despite the low numbers of NK cells in our patients, they shared a substantial level of identical integrations with naïve T cells processed and analysed separately (hence highly unlikely to be the result of contaminations), supporting our hypothesis of NK and T cell common origin. This data is in contrast with data obtained from non-human primates, which suggested that NK cells are likely to have an independent origin from other lineages, including T cells (Wu *et al.*, 2014). Data based on mice studies, however, is in line with our common origin hypothesis, and suggests that NK and T cells share expression of hundreds of genes and based on principal component analysis cluster together separately from other immune cells (Bezman *et al.*, 2012). Because of the low number of NK cells in our patients and the limited amount of human material available for this research, we were unable to run additional experiments to investigate further the developmental relationship between T and NK cells.

In the future, we aim to set up an immunophenotypic panel based on the study by Lavaert et al. (Lavaert *et al.*, 2020) to phenotypically characterise Lt-LP progenitors in the thymus and bone marrow of gene therapy patients. We would then FACS sort them and perform VISA. Next step would be to obtain blood samples from the same patients to see if we would be able to detect identical integration sites between Lt-LP from bone marrow/thymus and peripheral blood overtime. This would allow us to answer several

biological questions. Firstly, we could in this way compare identical integrations from Lt-LP that are shared with both T and NK cells and see whether this sharing is equal or whether it is more biased towards production of a particular cell type. In addition, this might help us to understand if Lt-LP from bone marrow and thymus contribute equally to peripheral T and NK cells and whether contribution of Lt-LP from different tissues changes overtime.

We also plan to study repopulation of the thymic microenvironment by haematopoietic progenitors using VISA. The aim would be to sort different CD34+ progenitor populations using a comprehensive immunophenotypic panel that is designed to differentiate between true HSCs (Lin-CD34+CD90+CD38-CD7-CD10-), multipotent progenitors (Lin-CD34+CD38-CD45RA-CD7-CD10-), multi-lymphoid progenitors (Lin-CD34+CD45RA+CD38-CD7-CD10-), common myeloid progenitors (Lin-CD34+CD38+CD135+CD45RA-CD7-CD10-), granulocyte-monocyte progenitors (Lin-CD34+CD38+CD135+CD45RA+CD7-CD10), preB/NK (Lin-CD34+CD10+CD7-) and megakaryocyte-erythrocyte progenitors (Lin-CD34+CD38+CD7-CD10-CD135-CD45RA-). These progenitors will be transduced with lentiviral vectors for tagging, and then cultured in 3D scaffold cultures (collaboration with Dr Paula Bonfanti) to understand their developmental capacity and contribution to the thymic microenvironment (Bosticardo et al., 2020; Campinoti et al., 2020; Langhans, 2018).

In conclusion, our data provide the first formal evidence in vivo in humans that de novo production of genetically engineered T and NK cells can be physiologically maintained by a population of Lt-LP surviving many years after loss of transplanted HSC. Identification and exploitation of such human Lt-LP population may be of significant

benefit in the development of next generation GT and cancer immunotherapy approaches.

5.3 Clonal expansion of T memory stem cells determines early anti-leukemic responses and long-term CAR T cell persistence in patients

Summarizing the second part of my PhD research, we used immunophenotypic analysis and molecular tracking to investigate the dynamic changes in the phenotype and nature of CAR T cells during the early target elimination response phase and during long-term surviving surveillance phase in our patients.

We used VISA to track CAR T cell clones in the product and in vivo in patients overtime to explore the contribution of different clonal populations to early and late timepoints after treatment. We showed that CAR T cells in the infused product display polyclonal integration site profile and that CAR T cells remain clonally diverse at early and late time points post-infusion with no evidence of clonal dominance. Individual clones showed different patterns of expansion and contraction with time, with different clones responsible for early expansion and long-term persistence. Interestingly, despite having observed several integration sites in the TET2 locus (Fraietta *et al.*, 2021), we could not find any sign of clonal expansion of the relevant CAR T cell clones.

For most of the patients, CAR T cell product consisted predominately of TCM population. Our results of phenotype changes of CAR T cells post infusion are summarized in the scheme of Figure 5.3.1. During early anti-leukemic response phase, CAR T cells undergo a rapid switch towards the TEM phenotype without major clonal

unbalances in the memory compartment but with self-limiting clonal burst phases occurring in the TSCM population. Using VISA, we looked at the level of sharing of identical integration sites between the CAR T cells in the product and CAR T cells isolated during early (<1 month) response phase. Our analysis demonstrated that despite the low frequency of CAR TSCM population in the product (0.5-2%), these cells shared a significant proportion (10-73%) of identical integrations with CAR T cells isolated during the early response phase in all patients including the two that experienced an early CAR T loss. This suggests that during manufacturing process, CAR TSCM population expanded and differentiated into TEM cells that then proceeded with early anti-leukemic response.

The patients with early loss of CAR T cells appeared to have a more differentiated phenotype, with a lower proportion of TSCM cells and higher TEM. Clonal diversity of TCM and TEM populations in the product was greatly reduced in patients with poor CAR T cell persistence. This might suggest that that when the product has a more differentiated phenotype and reduced clonality, CAR T cells are more prone to exhaustion during the clonal burst phase due to the significant expansion following anti-leukemic response. Our study has a small sample size due to the limited availability of patient samples. Therefore, further studies with larger patient cohorts are needed to validate current findings. However, if our results will be confirmed in more patients, these data have important clinical implications. By analysing the phenotype and clonal diversity of the product and assessing the same parameters during the early anti-leukemic response phase, we might be able to predict which patients are at higher risk of early CAR T cells loss and who might require a bone marrow transplant later in time to avoid a relapse.

In patients with long term persistence, CAR T cells did not terminally differentiate into TEMRA cells and maintained a low frequency of TEMRA population during both early and late phases. In contrast, CAR TSCM cells, despite representing a very small proportion of CAR T cells in the product and during early response phase, gradually increased in frequency overtime and preserved their clonal diversity even at latest timepoints. This small fraction of TSCM cells still present in the cell product at the time of infusion seemed to play a crucial role in preserving the pool of long-living CAR T cells. It has been reported before by other investigators, that the frequency of TSCM population in patients with relapsed multiple myeloma patients (Garfall, et al., 2015) and chronic lymphocytic leukaemia patients (Fraietta *et al.*, 2021) is the most important determinant to the response of CAR T cell therapy. However, another study looked at the association between the frequency of TSCM cells in the product and responsiveness to CAR T cell therapy in paediatric and adult ALL patients and did not identify any significant link between responders and non-responders and their frequency of TSCM in the product (Wang, et al., 2019). This would suggest that in ALL patients, the frequency of TSCM in the product does not influence their responsiveness to therapy. Conversely, based on our data, the frequency and polyclonality of TSCM population in the product seem to be major determinants for the long-term persistence of CAR T cells. By using recapture probability, we showed that a small TSCM population from the cell product had the highest level of sharing of identical integrations sites with long-term persisting CAR T cells in the periphery. TSCM subset is the only memory population that can differentiate into all other memory subsets down the differentiation hierarchy and at the same time maintain their stem-like phenotype and properties (Gattinoni *et al.*, 2011) (Cieri *et al.*, 2013). We can therefore hypothesise

that the long-term maintenance of CAR T cells in patients is supported by the population of self-renewing TSCM cells with homeostatic proliferative potential.

Our working hypothesis is that the remarkable early expansion and late survival rates of our CAR T cell product achieved in two patients could be due to the combination of two main factors. Firstly, our low affinity CAR construct that is endowed with a faster CAR-antigen off-load time could have reduced the differentiation drive of the CAR cells during the early anti-tumour response phase (Ghorashian *et al.*, 2019). This could have in turn protected the cell product from reaching a terminal differentiation stage which would have led to cell exhaustion. The signalling through low affinity CAT CAR construct conversely could have allowed expansion and self-renewal of CAR TSCM capable of long-term survival, which in part could have been mediated by higher expression of IL-7R and Bcl-2. This CAT CAR construct contains a 4-1BB co-stimulatory domain, which might enable a long-term persistence compared to CD28 by preventing exhaustion through tonic signalling (Milone *et al.*, 2009; Long, 2015). Secondly, our in vitro manipulation protocol that did not include the addition of cytokines to the medium could have allowed for a less pronounced cell differentiation in vitro and a preservation of high levels of TCM cells whose expansion potential could have contributed to an enhanced early response. The same culture conditions allowed a small fraction of TSCM to be infused in those patients in large enough doses to survive the initial clonal burst phases and to maintain a pool of CAR T cells long term.

Despite these limitations, our data provide evidence for the first time of the critical role of TSCM in both mediating early anti-leukemic responses and long-term persistence of CAR T cells after infusion. Future efforts will focus on validating these findings in

larger patient cohorts and improvement of CAR design and manufacturing methodology to preserve TSCM and TCM in the infused cell product.
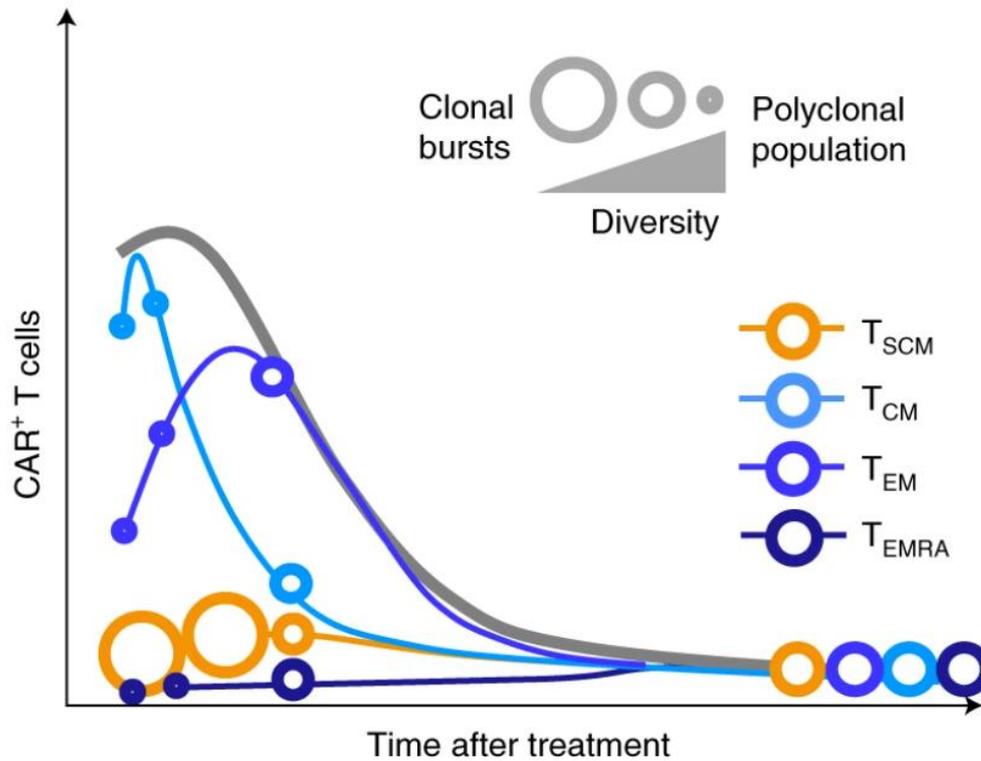
Figure 5.3.1 Model summarizing our working hypothesis for CAR T cell dynamics occurring in the patients of this study.

The y-axis represents the frequency of four T cell subsets. The x-axis shows time after treatment. The size of the circle refers to the polyclonality of a given subset at that time (the bigger the circle the more polyclonal the population is). During early anti-leukemic response, CAR T cells switch their TCM phenotype (product) and differentiate into TEM cells. During early stage, most subsets remain polyclonal apart from TSCM cells that undergo clonal bursts. After clearance of leukaemia, CAR T cell numbers go down, but remain highly polyclonal.

6 Methods

6.1 Study design

SCID-X1:

Sample size was constrained by limited clinical sample availability. For phenotypic characterisation and IS analyses of T cell subsets in SCID-X1 patients we collected in vivo biological material available during years 2001–2018. We aimed at analysing for all the patients at least 1 early and 1 late time point (where possible we did more). For IS analysis of NK and T cell identical integrations we acquired only 1-time point per patient, for TCR sequencing we aimed to acquire 1 early and 1 late time points, for IFN-gamma assay we analysed samples from the latest time points available. All available healthy donors' and patients' samples are reported here. Technically validated results were always included to the analyses and we did not apply any exclusion criteria for outliers. For phenotypic characterisation of T and NK cells, Rainbow beads (RB) calibration was performed during the setup of the instrumentation for FACS analyses. RB acquisition was performed before each sample acquisition in order to achieve reproducible instrument setting among different experiments. For VCN evaluation, qPCR was validated for reproducibility. We run each sample in triplicate and values are reported as mean of the three triplicates. All attempts at replication were successful. The experimental design did not include allocation of samples to randomised experimental group nor to blinding. There was no expected result prior to performing these analyses, therefore blinding tests were not applicable.

CARPALL:

Sample size was constrained by limited clinical sample availability. All available healthy donors' and patients' samples are reported here. Technically validated results were always included to the analyses, and we did not apply any exclusion criteria for outliers. For phenotypic characterisation of T cells, Rainbow beads calibration was performed during the set-up of instrumentation for FACS analyses. Rainbow beads acquisition was performed before each sample acquisition in order to achieve reproducible instrument setting among different experiments. All attempts at replication were successful. IS collection was conducted on a single experiment basis using well established protocol. No attempt at replication has been necessary.

6.2 Fluorescence activated cell sorting (FACS) analysis for immune phenotyping and cell sorting

6.2.1 Isolation of cells from whole blood

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood by density gradient centrifugation using LymphoPrep (Sigma) media. A dextran sedimentation step was performed to allow erythrocytes to sediment using 1 ml of 0.9% saline solution and 1 ml of 10% dextran (Sigma) solution, for 40 minutes at room temperature. Then, we loaded resulting supernatant containing white blood cells on 10 ml of LymphoPrep and centrifuged for 15 minutes at 2500 rpm (no brake). After, we collected PBMC ring in the new tube and washed it with phosphate-saline solution (PBS). The pellet at the bottom of LymphoPrep contained both sedimented erythrocytes and granulocytes. To lyse erythrocytes, we added sterile water to the pellet and mixed it thoroughly and quickly together. Then, right after we added 1.8% of NaCL solution to restore osmolality and prevent the lysis of granulocytes. Both tubes

containing PBMCs and granulocytes were the centrifuged washed for 7 minutes at 1500 rpm.

6.2.2 Staining and cells sorting

For both SCID-X1 and CARPALL samples the following panel of fluorochrome conjugated antibodies was used to sort four T cell populations of interest for vector integration site analysis, or to sort different blood cell lineages for VCN analysis, or to stain for CAR19 cells (Tables 6.2.1, 6.2.2 and 6.2.3):

Antibody	Clone	Fluorochrome	Source	Cat number
CD95	DX2	PE	Biolegend	305608
CD95	DX2	V711	BD Biosciences	740716
CD3	UCHT1	V500	BD Horizon	561416
CD3	UCHT1	APC-Cy7	Biolegend	300426
CD4	RPA-T4	APC-Cy7	BD Pharmigen	55787
CD8	RPA-T8	PeCy5	BD Pharmigen	555368
CD45RA	HI100	V450	BD Horizon	560362

CD62L	DREG-56	APC	Biolegend	304810
Simultest CD3/CD16+CD56+/ CD19/CD45RA	SJ25C1/S K7/B73.1/ NCAM16. 2/2D1	FITC/PE/APC/CD45 PerCP	BD Multitest	342416
Simultest CD3/CD16+CD56+/ CD45/CD19/CD4/C D8	SK7/B73.1 /NCAM16. 2/2D1/SJ2 5C1/SK3/ SK1	FITC/PE/PerCP-Cy5.5/APC/PE-Cy7/APC-Cy7	BD Multitest	644611
CD14	MφP9	APC-Cy7	BD Biosciences	557831
Anti-CAT CAR idiotype antibody	Generated in house			
Anti-Rat IgG		PE or Pe-Cy5		
7AAD	-	APC-Cy7	Biolegend	410404
CD16	3G8	APC-Cy7	Biolegend	302017
CD56	5.1H11	PE-Cy7	Biolegend	362509
CD45	HI30	BV650	BD Horizon	563717

CD3	UCHT1	PeCP-Cy5.5	BD Pharmingen	560835
CD19	HIB19	BV605	Biolegend	302243
CCR7	3D12	BV510	BD Horizon	563449
CCR5	J418F1	BV711	Biolegend	359129
CD158a	HP-MA4	PE	Biolegend	339505
CD158b	DX27	PE	Biolegend	312605
CD158e1	DX9	PE	Biolegend	312707
CD158i	REA860	PE	Miltenyi Biotec	130-114-772
CX3CR1	2A9-1	PE/Dazzle 594	Biolegend	341623
CCR2	K036C2	BV421	Biolegend	357209
CXCR3	G025H7	AF647	Biolegend	353711
CD57	FITC	HNK-1	BD	333169
CD27	BV421	M-T271	BD Horizon	562514
CD127	BV711	HIL-7R-M21	BD Horizon	563165

Perforin	AF488	δG9	BD Pharmingen	563764
Granzyme B	PE	GB11	BD Pharmingen	561142
CD94	APC	HP-3D9	BD Pharmingen	559876
NKG2C	PE-Vio615	REA205	Miltenyi Biotec	130-123-047
NKG2A	BV510	131411	BD OptiBuild	747922

Table 6.2.1 List of antibodies used for immunophenotyping.

T cell subset (LAM-PCR)	Immune phenotype
Naïve T cells (TN)	$CD4/CD8^+$ $CD3^+CD45RA^+CD62L^+CD95^-$
T stem cell memory (TSCM)	$CD4/CD8^+$ $CD3^+CD45RA^+CD62L^+CD95^+$

Central memory (TCM)	CD4/CD8$^+$ CD3$^+$CD45RA$^-$ CD62L$^+$CD95$^+$
Effector memory (TEM)	CD4/CD8$^+$ CD3$^+$CD45RA$^-$CD62L$^-$ CD95$^+$

Table 6.2.2 Phenotypic description of T cell subsets used for LAM-PCR analysis.

These combinations of cell surface receptors were used to sort different T cell subsets and extract DNA.

Blood cell lineage (VCN)	Immune phenotype
T cells	CD3$^+$CD45RA$^+$CD16/56$^-$CD19$^-$ CD14$^-$
B cells	CD3$^-$CD45RA$^+$CD16/56$^-$ CD19$^+$CD14$^-$
NK cells	CD3$^-$CD45RA$^+$CD16/56$^+$CD19$^-$ CD14$^-$
Monocytes	CD3$^-$CD45RA$^+$CD16/56$^-$CD19$^-$ CD14$^+$

Table 6.2.3 Phenotypic description of immune cell subsets used for VCN analysis.

These combinations of cell surface receptors were used to sort different T cell subsets and extract DNA.

Cells were re-suspended in PBS and stained for 30 minutes at 4 °C. Then, samples were washed with PBS and 2.5 µl of DAPI solution (stock concentration 5 mg/ml) was added to samples to check viability, if compatible with Ab panel. Samples were sorted with FACS Aria or analysed with Canto, and when feasible, an aliquot of the sorted cells was run through the cell sorter to check fraction purity.

6.3 Extraction and quantification of genomic DNA from patient samples

Two DNA extraction kits were used depending on number cells: **1.** Qiagen QIAamp DNA Blood Mini Kit in case when there was from 10,000 to up to 5 x 10^6 sorted cells available from the sample. **2.** MN NucleoSpin® Tissue XS kit in cases where there was only 10 to up to 10,000 sorted cells available

In both cases, according to manufacturer's instructions, DNA was digested with proteinase K and lysis buffer for 10 minutes at 56 °C on the heat block. Then, in case if Qiagen kit was used – samples were re-suspended in ethanol (100%) and transferred to the column provided with the kit. In case if NucleoSpin kit was used after incubation at 56 °C, another lysis buffer was added to samples, and they were returned to heat block for incubation at 70°C for 5 minutes. Then, ethanol (100%) was added, and samples were transferred to the column. Once on the column, samples were washed 2 times (centrifugation for Qiagen kit – 8,000 rpm 1st wash, 13,000 rpm 2nd wash; NucleoSpin kit – 11,000 rpm for all steps) and eluted with provided elution buffer. Elution volume depended on the starting number of cells. Quantification has been done using Nanodrop One Spectrophotometer.

6.3.1 Whole Genome Amplification

For whole genome amplification of samples, we used RepliG Mini kit (Qiagen) and followed manufacturers' instructions. Briefly, 5 µl of genomic DNA was mixed with buffer D1. Samples were incubated for 3 minutes at room temperature. Then 10 µl of buffer N1 was added to the mix. Master mix containing Repli-G Mini DNA polymerase was prepared according to instructions (29 µl Reaction buffer + 1 µl DNA polymerase per sample). 30 µl of master mix was then added to 20 µl of the mix containing DNA. Samples were then incubated at 30 °C for 16 hours, followed by heating for 3 minutes at 65 °C in the thermocycler.

6.4 Linear – amplification mediated polymerase chain reaction (LAM-PCR) for vector integration site analysis

To perform vector integration site analysis, we used LAM-PCR combined with high-throughput sequencing. If the DNA yield after extraction was lower than 300 ng, whole-genome amplification (WGA) has been performed with Qiagen Repli-G Mini Kit (in this case, for consistency, all samples from this patient would undergo WGA).

Then, 2 rounds of linear PCR were performed (50 cycles each) to enrich for vector genome junction using biotinylated primers specific for vector LTRs (Table 6.4.1).

20 µl of streptavidin coupled magnetic beads (Invitrogen Dynabeads Kilobase Binder Kit) were added to each sample, to capture linearly amplified fragments, and incubated overnight. For complementary strand synthesis, hexanucleotide priming mixture, containing Klenow enzyme, is added to samples, which are then incubated for an hour

at 37°C in thermocycler. After 1 hour of incubation, samples are washed with water on the magnet, and each sample is split in three wells. For each well, a different restriction enzyme (MluCI, AciI or HpyCH4IV) is used, to minimize bias and improve genome coverage.

Samples were incubated at 37°C for 2 hours. Then, linker cassette ligation mix was added to washed samples. To activate ligation reaction, samples were incubated at 16°C for 30 minutes and at 70°C for 10 minutes. To separate DNA from the beads, 5 µl of 0.1 N of sodium hydroxide was added to each sample and incubated for 15 min at room temperature on the shaker (300 rpm). Then, denatured DNA was collected and used to run 2 exponential PCR reactions that amplify fragments that contain vector LTR sequences and linker cassette sequences. PCR mix for the 1st and 2nd exponential PCR contains primers for vector LTR and a linker cassette (Table 6.4.1).

Primer	Used for	Sequence
5'MLV I BIOT	Linear Amplification (SCID-X1)	(5'BIOTIN) AGCTGTTCCATCTGTTCCTGACCTTT 3'
5'MLV II BIOT	Linear Amplification (SCID-X1)	(5'BIOTIN) TGCTTACCACAGATATCCTG 3'

retroLTR III	1° exponential PCR	5' GACCTTGATCTGAACTTCTC 3'
retroLTR IV	2° exponential PCR	5' TTCCATGCCTTGCAAAATGGC 3'
LC 1 LC 3 MluCl	linker cassette	5' Gacccgggagatctgaattcagtggcacagcagttagg 3'
		5' Aattcctaactgctgtgccactgaattcagatc 3'
LC 1 LC 5 AciI and HpyCH4IV	linker cassette	5' Gacccgggagatctgaattcagtggcacagcagttagg 3'
		5' CGCCTAACTGCTGTGCCACTGAATTCAGATC 3'
LCP I	1° exponential PCR (SCID-X1)	5' GACCCGGGAGATCTGAATTC 3'
LCP II	2° exponential	5' AGTGGCACAGCAGTTAGG 3'

	PCR (SCID-X1)	
SK LTR1 BIOT	Linear Amplification (CARPALL)	(5'BIOTIN) GAGCTCTCTGGCTAACTAGG 3'
SK LTR2 BIOT	Linear Amplification (CARPALL)	(5'BIOTIN) GAACCCACTGCTTAAGCCTCA 3'
SK LTR3	1° exponential PCR (CARPALL)	5' AGCTTGCCTTGAGTGCTTCA 3'
SK LTR4	2° exponential PCR (CARPALL)	5' AGTAGTGTGTGCCCGTCTGT 3'
LCP1	1° exponential PCR	5' GACCCGGGAGATCTGAATTC 3'

	(CARPALL)	
LCP2	2° exponential PCR (CARPALL)	5' GATCTGAATTCAGTGGCACAG 3'

Table 6.4.1 Primer and linker cassette sequences for LAM-PCR for both SCID-X1 and CARPALL samples.

6.4.1 Gel electrophoresis of final PCR products

PCR fragment size depends on the distance between the known vector sequence and the closest enzyme recognition site. To visualize amplified fragments, we used high-resolution gel electrophoresis (ElchromScientific). Samples could have just one band (monoclonal), several bands (oligoclonal) or multiple bands (polyclonal) present on the gel.

If the experiment was successful and there were bands on the gel, final PCR products were purified with QIAQUICK PCR purification kit (QIAGEN). Then, DNA was quantified using Nanodrop One Spectrophotometer.

6.4.2 Library preparation

Fusion PCR is used to add sequence specific adaptors to final PCR products. Combination of different LTR primers and LC primers is used to differentiate between different samples. For sequencing, the three final products obtained from the same

patient with the three different enzymes were barcoded with the same combination of LTR and LC primers (Table 6.4.2 (SCID-X1) or Table 6.4.3 (CARPALL)).

After fusion PCR and before purification of Fusion PCR products, we ran 2% agarose gel to visualise fragments. In case fusion PCR was successful and there were visible fragments present for each sample, fusion PCR products were purified with QIAQUICK PCR purification kit (QIAGEN) and quantified with Nanodrop One Spectrophotometer.

Then, samples were ready for sequencing and final pooling has been made by UCL Genomics facility.

Data analysis:

Bioinformatical and statistical analysis has been performed by our collaborators Luca Biasco and Danilo Pellin at Dana Farber Institute.

Primer	Barcode 8	Sequence
GR-P5-Rd1-LTR.1	AAACATCG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATCTNNNNNNNNNNNN AAACATCGCCTTGCAAAATGGCGTTACT
GR-P5-Rd1-LTR.2	AACAACCA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT ACACGACGCTCTTCCGATCTNNNNNNNNNNNN AACAACCACCTTGCAAAATGGCGTTACT

GR-P5-Rd1-LTR.3	AACCGAGA	AATGATACGGCGACCACCGAGATCTACACTC TTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNN AACCGAGACCTTGCAAAATGGCGTTACT
GR-P5-Rd1-LTR.4	AACGCTTA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTNNNNNNNNNNNN AACGCTTACCTTGCAAAATGGCGTTACT
GR-P5-Rd1-LTR.5	AACGTGAT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTNNNNNNNNNNNN AACGTGATCCTTGCAAAATGGCGTTACT
GR-P5-Rd1-LTR.6	AACTCACC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCTNNNNNNNNNNNN AACTCACCCCTTGCAAAATGGCGTTACT
GR-P5-Rd1-LTR.7	AAGACGGA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTNNNNNNNNNNNN AAGACGGACCTTGCAAAATGGCGTTACT
GR-P5-Rd1-LTR.8	AAGAGATC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCTNNNNNNNNNNNN AAGAGATCCCTTGCAAAATGGCGTTACT
GR-P5-Rd1-LTR.9	AAGGACAC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTNNNNNNNNNNNN AAGGACACCCTTGCAAAATGGCGTTACT

GR-P5-Rd1-LTR.10	AAGGTACA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNAAGGTACACCTTGCAAAATGGCGTTACT
GR-P7-Rd2.LC.1	CAACCACA	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNCAACCACAGATCTGAATTCAGTGGCACAG
GR-P7-Rd2.LC.2	CAAGACTA	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNCAAGACTAGATCTGAATTCAGTGGCACAG
GR-P7-Rd2.LC.3	CAAGGAGC	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNCAAGGAGCGATCTGAATTCAGTGGCACAG
GR-P7-Rd2.LC.4	CAATGGAA	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNCAATGGAAGATCTGAATTCAGTGGCACAG
GR-P7-Rd2.LC.5	CACCTTAC	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNCACCTTACGATCTGAATTCAGTGGCACAG
GR-P7-Rd2.LC.6	CACTTCGA	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNCACTTCGAGATCTGAATTCAGTGGCACAG

GR-P7-Rd2.LC.7	CAGAT CTG	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTNNNNNNNNNNNN CAGATCTGGATCTGAATTCAGTGGCACAG
GR-P7-Rd2.LC.8	CAGCG TTA	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTNNNNNNNNNNNN CAGCGTTAGATCTGAATTCAGTGGCACAG
GR-P7-Rd2.LC.9	CATAC CAA	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTNNNNNNNNNNNN CATACCAAGATCTGAATTCAGTGGCACAG
GR-P7-Rd2.LC.10	CATCA AGT	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTNNNNNNNNNNNN CATCAAGTGATCTGAATTCAGTGGCACAG

Table 6.4.2 Fusion PCR primers for SCID-X1 samples.

Illumina adapters specific for either vector gammaretroviral (GR) LTR portion of the fragment or linker cassette (LC).

Primer	Barco de 8	Sequence
FB-P5-Rd1-LTR.1	AAAC ATCG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACAC GACGCTCTTCCGATCTNNNNNNNNNNNN AAACATCGACCCTTTTAGTCAGTGTGGA

FB-P5-Rd1-LTR.2	AACA ACCA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACA CGACGCTCTTCCGATCTNNNNNNNNNNNN AACAACCAACCCTTTTAGTCAGTGTGGA
FB-P5-Rd1-LTR.3	AACC GAGA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACA CGACGCTCTTCCGATCTNNNNNNNNNNNN AACCGAGAACCCTTTTAGTCAGTGTGGA
FB-P5-Rd1-LTR.4	AACG CTTA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCTNNNNNNNNNNNN AACGCTTAACCCTTTTAGTCAGTGTGGA
FB-P5-Rd1-LTR.5	AACG TGAT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT ACACGACGCTCTTCCGATCTNNNNNNNNNNNN AACGTGATACCCTTTTAGTCAGTGTGGA
FB-P5-Rd1-LTR.6	AACT CACC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCTNNNNNNNNNNNN AACTCACCACCCTTTTAGTCAGTGTGGA
FB-P5-Rd1-LTR.7	AAGA CGGA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTNNNNNNNNNNNN AAGACGGAACCCTTTTAGTCAGTGTGGA
FB-P5-Rd1-LTR.8	AAGA GATC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTNNNNNNNNNNNN AAGAGATCACCCTTTTAGTCAGTGTGGA

FB-P5-Rd1-LTR.9	AAGG ACAC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT ACACGACGCTCTTCCGATCTNNNNNNNNNNNN AAGGACACACCCTTTTAGTCAGTGTGGA
FB-P5-Rd1-LTR.10	AAGG TACA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT ACACGACGCTCTTCCGATCTNNNNNNNNNNNN AAGGTACAACCCTTTTAGTCAGTGTGGA
FB-P7-Rd2-LC.1	AAAC ATCG	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCTNNNNNNNNNNNN AAACATCGGATCTGAATTCAGTGGCACAG
FB-P7-Rd2-LC.2	AACA ACCA	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTNNNNNNNNNNNN AACAACCAGATCTGAATTCAGTGGCACAG
FB-P7-Rd2-LC.3	AACC GAGA	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTNNNNNNNNNNNN AACCGAGAGATCTGAATTCAGTGGCACAG
FB-P7-Rd2-LC.4	AACG CTTA	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTG CTCTTCCGATCTNNNNNNNNNNNN AACGCTTAGATCTGAATTCAGTGGCACAG
FB-P7-Rd2-LC.5	AACG TGAT	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTNNNNNNNNNNNN AACGTGATGATCTGAATTCAGTGGCACAG

FB-P7-Rd2-LC.6	AACT CACC	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCTNNNNNNNNNNNN AACTCACCGATCTGAATTCAGTGGCACAG
FB-P7-Rd2-LC.7	AAGA CGGA	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCTNNNNNNNNNNNN AAGACGGAGATCTGAATTCAGTGGCACAG
FB-P7-Rd2-LC.8	AAGA GATC	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTNNNNNNNNNNNN AAGAGATCGATCTGAATTCAGTGGCACAG
FB-P7-Rd2-LC.9	AAGG ACAC	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNN AAGGACACGATCTGAATTCAGTGGCACAG
FB-P7-Rd2-LC.10	AAGG TACA	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTNNNNNNNNNNNN AAGGTACAGATCTGAATTCAGTGGCACAG

Table 6.4.3 Fusion PCR primer sequences for CARPALL samples.

Illumina adapters specific for either vector lentiviral LTR portion of the fragment or linker cassette (LC).

6.5 TCR sequencing

High-throughput sequencing of TCR was performed on TN, TSCM, TCM and TEM
FACS-sorted populations as previously described (Bartram *et al.*, 2016). Briefly, PCR
was performed to amplify somatic rearrangements within TCR gamma, TCR delta and
TCR beta genes (Table 6.5.1). After the initial PCR, the products were purified via
Pronex bead chemistry (Promega), and sequencing adaptors were added. These
adaptors allow the sequences to bind to the flow cell in the MiSeq and also contain
indices used to identify samples during sequencing analysis. Dual indexing utilises two
8-base sequences; index 1 (i7), adjacent to the P7 sequence and index 2 (i5), adjacent
to the P5 sequence (Table 6.5.2). Resulting amplicons were normalised and pooled
based on Qubit and Tape station values and sequenced using Illumina MiSeq 300 v2
kit. TCR sequencing demultiplexing was performed using error aware demultiplexer
(EAD) for Illumina BCL files (v1.0.3). The resulting FASTQ files were then loaded on
the Vidjil (High-Throughput Analysis of V(D)J Immune Repertoire) suite
(www.vidjil.org).

Primer	What for	Sequence
VG1-FR2	TCR gamma	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGAAGGCCC CACAGCRTCTT

VG9-FR2	TCR gamma	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGGCACTGTC AGAAAGGAATC
VG10-FR2	TCR gamma	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGCATGGGTA AGACAAGCAA
VG11-FR2	TCR gamma	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTCCACTTC CACTTTGAAA
JG1.1-FR2	TCR gamma	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTACCAGGCGA AGTTACTATGAGC
JG1.3-FR2	TCR gamma	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGTTGTTCCAC TGCCAAAGAG
VD1-FR2	TCR delta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATGCAAAAAG TGGTCGCTATT
VD2-FR2	TCR delta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATACCGAGAA AAGGACATCTATG
VD3-FR2	TCR delta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTACCGGATA AGGCCAGATTA
VD4-FR2	TCR delta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATGACCAGCA AAATGCAACAG
VD5-FR2	TCR delta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCCTGCTGA AGGTCCTACAT

VD6-FR2	TCR delta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCCTGCATTATTGATAGCCAT
DD2F-FR2	TCR delta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGCGGGTGGTGATGGCAAAGT
JD1-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTCCACAGTCACACGGGTTC
JD2-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTCCACGATGAGTTGTGTTC
JD3-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCACGGGGCTCCACGAAGAG
JD4-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGACCTCCAGATAGGTTCC
JD4-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGACCTCCAGATAGGTTCC
DD3R-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGGACCCAGGGTGAGGATAT
Ja29-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCAAAAGCATTCTAGGTACA
Ja9-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTAACTGGCAGACAAAACTATG

Ja30-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCACCCACAT GTCTTAG
Ja48-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCCCAGAATCT TATGCAG
Ja54-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGGGGCAAGT AATTAAATCA
Ja58-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACTTGAATGT GGCAGAGA
Ja61-FR2	TCR delta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTTGTTAAGGC ACATTAGAATC
Vβ2-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAACTATGTTTT GGTATCGTCA
Vβ4-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACGATGTTC TGGTACCGTCAGCA
Vβ5/1-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGTGTGTCC TGGTACCAACAG
Vβ6a/11-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAACCCTTTAT TGGTACCGACA
Vβ6b/25-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCCCTTTTTT GGTACCAACAG

Vβ6c-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAACCCTTTAT TGGTATCAACAG
Vβ7-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCTATGTAT TGGTACAAGCA
Vβ8a-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCCCGTTTT CTGGTACAGACAGAC
Vβ9-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCTATGTAT TGGTATAAACAG
Vβ10-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTATGTTTACT GGTATCGTAAGAAGC
Vβ11-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAAAATGTAC TGGTATCAACAA
Vβ12a/3a/ 13a/15- FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATACATGTAC TGGTATCGACAAGAC
Vβ13b- FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCATGTAC TGGTATAGACAAG
Vβ13c/12b /14-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTATATGTCC TGGTATCGACAAGA

Vβ16-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTAACCTTTATT GGTATCGACGTGT
Vβ17-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCATGTAC TGGTACCGACA
Vβ18-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCATGTTTAC TGGTATCGGCAG
Vβ19-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTATGTTTATT GGTATCAACAGAATCA
Vβ20-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAACCTATAC TGGTACCGACA
Vβ21-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTACCCTTTAC TGGTACCGGCAG
Vβ22-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATACTTCTATT GGTACAGACAAATCT
Vβ23/8b-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACGGTCTAC TGGTACCAGCA
Vβ24-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGTCATGTAC TGGTACCAGCA
Dβ1-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCAAACAGC CTTACAAAGAC

Dβ2-FR2	TCR beta	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTCCAAGCC CCACACAGTC
Jβ1.1-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTACCTACAAC TGTGAATCTGGTG
Jβ1.2-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTACCTACAAC GGTTAACCTGGTC
Jβ1.3-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTACCTACAAC AGTGAGCCAACTT
Jβ1.4-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATACCCAAGA CAGAGAGCTGGGTTC
Jβ1.5-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTACCTAGGAT GGAGAGAGTCGAGTC
Jβ1.6-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATACCTGTCAC AGTGAGCCTG
Jβ2.1-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTTCTTACCTA GCACGGTGA
Jβ2.2-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTACCCAGTAC GGTCAGCCT
Jβ2.3-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCCGCTTACCG AGCACTGTCA

Jβ2.4-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCAGCTTACCC AGCACTGAGA
Jβ2.5-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCGCACACCG AGCAC
Jβ2.6-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCGCCCAGCA CGGTCAGCCT
Jβ2.7-FR2	TCR beta	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTACCTGTAAC CGTGAGCCTG

Table 6.5.1 Primers for TCR sequencing – stage 1 PCR.

Primer	Barcode	Sequence
i7_01	TCTAGCTA	CAAGCAGAAGACGGCATACGAGATTCTAGCTA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_02	CTAGCTAT	CAAGCAGAAGACGGCATACGAGATCTAGCTAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_03	AGGTTGGC	CAAGCAGAAGACGGCATACGAGATAGGTTGGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_04	GACCAACG	CAAGCAGAAGACGGCATACGAGATGACCAACG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

i7_05	GCGGAGTT	CAAGCAGAAGACGGCATACGAGATGCGGAGTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_06	GTGCCATA	CAAGCAGAAGACGGCATACGAGATGTGCCATA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_07	TAATGTCC	CAAGCAGAAGACGGCATACGAGATTAATGTCC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_08	CGAAGGAC	CAAGCAGAAGACGGCATACGAGATCGAAGGAC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_09	AATGTCCT	CAAGCAGAAGACGGCATACGAGATAATGTCCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_10	AGAACATT	CAAGCAGAAGACGGCATACGAGATAGAACATT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_11	TGTCAGTC	CAAGCAGAAGACGGCATACGAGATTGTCAGTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_12	CAGACGCA	CAAGCAGAAGACGGCATACGAGATCACCGCTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_13	GCTACTAG	CAAGCAGAAGACGGCATACGAGATCAGACGCA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_14	GTCAGTCT	CAAGCAGAAGACGGCATACGAGATGCTACTAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_15	TTCACCGC	CAAGCAGAAGACGGCATACGAGATGTCAGTCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

i7_16	GGTCTAAT	CAAGCAGAAGACGGCATACGAGATTTCACCGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_17	ACCTGGAT	CAAGCAGAAGACGGCATACGAGATGGTCTAAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_18	AGCGACAG	CAAGCAGAAGACGGCATACGAGATACCTGGAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_19	ATAGGCTC	CAAGCAGAAGACGGCATACGAGATAGCGACAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_20	TAGAACAT	CAAGCAGAAGACGGCATACGAGATATAGGCTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_21	TGCGGAGT	CAAGCAGAAGACGGCATACGAGATTAGAACAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_22	TTGCGGAG	CAAGCAGAAGACGGCATACGAGATTGCGGAGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_23	TTAGAACA	CAAGCAGAAGACGGCATACGAGATTTGCGGAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i7_24	CAGACGCA	CAAGCAGAAGACGGCATACGAGATTTAGAACA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
i5_01	CACTTGAG	AATGATACGGCGACCACCGAGATCTACACCACTTGAG ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_02	GTTACCGA	AATGATACGGCGACCACCGAGATCTACACGTTACCGA ACACTCTTTCCCTACACGACGCTCTTCCGATCT

i5_03	TGACGACT	AATGATACGGCGACCACCGAGATCTACACTGACGACT ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_04	ACGGATTC	AATGATACGGCGACCACCGAGATCTACACACGGATTC ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_05	CCATAGGA	AATGATACGGCGACCACCGAGATCTACACCCATAGGA ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_06	TGGAAGGC	AATGATACGGCGACCACCGAGATCTACACTGGAAGGC ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_07	GCATCATG	AATGATACGGCGACCACCGAGATCTACACGCATCATG ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_08	AGCGGTGA	AATGATACGGCGACCACCGAGATCTACACAGCGGTGA ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_09	AGTTACCG	AATGATACGGCGACCACCGAGATCTACACAGTTACCG ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_10	CATGCATA	AATGATACGGCGACCACCGAGATCTACACCATGCATA ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_11	ACATGCAT	AATGATACGGCGACCACCGAGATCTACACACATGCAT ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_12	ACCATAGG	AATGATACGGCGACCACCGAGATCTACACACCATAGG ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_13	TCCAGGTA	AATGATACGGCGACCACCGAGATCTACACTCCAGGTA ACACTCTTTCCCTACACGACGCTCTTCCGATCT

i5_14	CTTAATTG	AATGATACGGCGACCACCGAGATCTACACCTTAATTG ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_15	CGGATTCA	AATGATACGGCGACCACCGAGATCTACACCGGATTCA ACACTCTTTCCCTACACGACGCTCTTCCGATCT
i5_16	TTAGACCA	AATGATACGGCGACCACCGAGATCTACACTTAGACCA ACACTCTTTCCCTACACGACGCTCTTCCGATCT

Table 6.5.2 Primer sequences for library preparation – stage 2 PCR.

6.6 Statistical analysis

6.6.1 High throughput TCR sequencing analyses

After TCR sequencing demultiplexing was performed using EAD for Illumina BCL files (v1.0.3). The resulting FASTQ files were then loaded on the Vidjil (high-throughput analysis of V(D)J immune repertoire) suite (www.vidjil.org). Data shown in Fig. 3.3.3a,b and 3.3.5 were generated on the basis of the TCR database (available for download here as Supplementary data 1) exported from Vidjil and filtered for rearrangements labelled with "no CD3 detected". TCR diversity was calculated as Shannon Diversity Index through the R package Entropy (http://cran.r-project.org/web/packages/entropy/index.html).The network plots of Fig. 3.3.6a were generated using the R package visNetwork (https://cran.r-project.org/web/packages/visNetwork/index.html). Pearson correlation values for these plots were generated through the R package Hmisc (https://cran.r-project.org/web/packages/Hmisc/index.html, function rcorr, type = "pearson").

6.6.2 IS analysis

IS identification and analysis were performed through a custom analytical pipeline extensively portrayed in previous publications (Biasco *et al.*, 2015) (Biasco *et al.*, 2016).

The following is a brief summary of the computational tools used for the analysis of IS. Raw ISs data sets underwent series of different bioinformatics filtering procedures according to the type of analysis to be performed. All data sets were processed with a "collision detection filter" to univocally assign each IS to a patient and to one or more T cell subpopulations by applying a 10-fold rule for contamination identification as previously reported. A final matrix M was generated where each row r represented an individual IS while each column c an individual cell type/sample and time point. Each entry of M contained the abundance of each r for each c in terms of sequencing reads.

6.6.2.1 SCID-X1

The data shown in figures in Chapter 3, sections 3.4-3.6 were generated on the basis of the IS databases summarised in Figures 6.6.1 and 6.6.2. IS diversity overtime was calculated as Shannon Diversity Index through the R package Entropy(http://cran.r-project.org/web/packages/entropy/index.html). Additional diversity indexes Simpson and InverseSimpson were calculated and reported together with Shannon diversity in TFigure 6.6.2 through the use of the R packageBiodiversityR(https://cran.rproject.org/web/packages/BiodiversityR/index.html). The "bubble" plots on top of each panel were created on the basis of the IS from TN and NK, respectively, with abundance >0.01% relative to each subpopulation and time point using the R package packcircles (https://cran.r-project.org/web/packages/packcircles/index.html). The network plots of Fig. 3.4.4 were

221

generated using the R package visNetwork (https://cran.r-project.org/web/packages/visNetwork/index.html). The Pearson correlation values for these plots and for the ones of Fig. 3.4.5, were generated throughtheRpackageHmisc(https://cran.rproject.org/web/packages/Hmisc/index.html, function = rcorr, type = "pearson"). Estimation of clonal abundance and standard errors shown in Fig. 3.4.6d could be calculated on the two IS datasets with three timepoints each (P1 and P5) by the conversion of M to a M(0,1) matrix of incidence and by the application to M(0,1) of log-linear models for closed populations through the R package Rcapture (https://cran.r-project.org/web/packages/Rcapture/index.html, function = closedp.t). The Mth Chao (LB) method was selected for visualisation in Fig. 3.4.6d being the most conservative estimation among the ones with the lowest Bayesian Information content (BIC). The heat maps of Figs. 3.4.5, 3.4.7, and 3.5.7 and 3.6.2 were generated through the R package gplots (https://cran.r-project.org/web/packages/gplots/index.html) with ramping gradients of colour palettes calculated using the following breaks for relative IS abundance (col_breaks=c(seq(0,0.0001,length=100), seq(0.00011,0.01,length = 100), seq(0.011,0.1,length = 100),seq(0.11,1,length = 100), seq(1.1,5,length = 100), seq(5.1,10,length =), seq(10.1,20,length = 100), seq(20.1,30,length = 100), seq(30.1,50,length = 100), seq(50.1,70,length = 100), seq(70.1,100,length = 100)). Word clouds of Fig. 3.4.2 and 3.4.3(top panel) were generated on the basis the relative incidence of the single closest gene to each IS and plotted using the online suite WordClouds (https://www.wordclouds.com). The Gene Ontology analysis of hit genes in TN shown in Fig. 3.4.3 (bottom panel) was generated through the online suite Genomic Regions Enrichment of Annotations Tool (GREAT) (http://great.stanford.edu/public/html/) using BED data from IS coordinates,

associating genomic regions with the rule of single nearest gene within 1000 kb to each

IS and applying the gene annotations to these regions.

P1

P1	TN_PB_125	TSCM_PB_125	TCM_PB_125	TEM_PB_125	NK_PB_138	T_PB_138	TN_PB_153	T
Unique IS	111	134	153	176	272	398	256	
SeqReads	30788	103627	136919	121099	2993541	5395665	184905	

P5

P5	TN_PB_35	TSCM_PB_35	TCM_PB_35	TEM_PB_35	NK_PB_81	T_PB_81	TN_PB_95	T
Unique IS	60	54	64	64	61	733	19	
SeqReads	24797	32978	85908	82332	77610	949293	1215	

P6

P6	TN_PB_85	TSCM_PB_85	TCM_PB_85	TEM_PB_85	Total unique
Unique IS	144	151	177	615	889
SeqReads	443196	182873	315687	139665	

P8

P8	TN_PB_66	TSCM_PB_66	TCM_PB_66	TEM_PB_66	NK_PB_100	T_PB_100	TN_PB_124	T
Unique IS	475	376	193	376	296	723	50	
SeqReads	196184	288614	205853	363884	3136071	2408961	67609	

P10

P10	NK_PB_34	T_PB_34	TN_PB_45	TSCM_PB_45	TCM_PB_45	TEM_PB_45	TN_PB_142	T
Unique IS	22	348	391	225	213	399	221	
SeqReads	174501	2084619	217882	242955	89282	299170	238597	

(continuation columns)

	TSCM_PB_153	TCM_PB_153	TEM_PB_153	TN_PB_191	TSCM_PB_191	TCM_PB_191	TEM_PB_191	Total unique
	250	159	370	932	869	1671	308	3489
	118143	59056	225758	201103	244016	290563	3464	

	TCM_PB_95	TEM_PB_95	TN_PB_139	TSCM_PB_139	TCM_PB_139	TEM_PB_139	Total unique
	177	41	1612	1437	1119	567	4832
	74411	76062	279892	351830	311558	440569	

	TSCM_PB_124	TCM_PB_124	TEM_PB_124	Total unique
	23	43	45	2053
	43496	28552	11484	

	TCM_PB_142	TEM_PB_142	Total unique
	206	172	1493
	91780	149742	

Figure 6.6.1 IS data used to generate figures in sections 3.4-3.6 for five patients (P1-P10).

Names on top indicate Tsubset_fromPeripheralblood_timepoint – TN-PB_125. SeqReads - sequence reads.

P1	Shannon	Simpson	inverseSimpson
TN_PB_125	1.954513	0.758108	4.13407
TSCM_PB_125	1.958153	0.748482	3.97586
TCM_PB_125	2.018748	0.73249	3.73818
TEM_PB_125	2.380668	0.826052	5.74886
NK_PB_138	0.928342	0.489361	1.95833
T_PB_138	2.118468	0.834229	6.03242
TN_PB_153	2.28076	0.856156	6.95197
TSCM_PB_153	2.382301	0.803025	5.07678
TCM_PB_153	1.179195	0.602138	2.51343
TEM_PB_153	2.303175	0.763762	4.23302
TN_PB_191	3.459834	0.888968	9.00644
TSCM_PB_191	3.075799	0.841886	6.32454
TCM_PB_191	4.952752	0.971851	35.52527
TEM_PB_191	4.506504	0.975202	40.32645

P5	Shannon	Simpson	inverseSimpson
TN_PB_35	0.167195	0.043499	1.04548
TSCM_PB_35	0.285738	0.092753	1.10224
TCM_PB_35	1.532516	0.722655	3.60562
TEM_PB_35	1.44854	0.649688	2.8546
NK_PB_81	0.094356	0.021445	1.02191
T_PB_81	1.437778	0.420116	1.72448
TN_PB_95	2.033158	0.828058	5.81592
TCM_PB_95	1.151242	0.611504	2.57403
TEM_PB_95	1.759823	0.776474	4.47375
TN_PB_139	3.039742	0.923808	13.12472
TSCM_PB_139	2.061177	0.802362	5.05975
TCM_PB_139	2.113851	0.827814	5.80767
TEM_PB_139	1.9478	0.808906	5.23303

P6	Shannon	Simpson	inverseSimpson
TN_PB_85	1.57646	0.705355	3.39392
TSCM_PB_85	1.35031	0.647462	2.83657
TCM_PB_85	1.11664	0.575549	2.35598
TEM_PB_85	1.44596	0.659548	2.93727

P8	Shannon	Simpson	inverseSimpson
TN_PB_66	2.176592	0.80094	5.02361
TSCM_PB_66	2.404198	0.85375	6.8376
TCM_PB_66	1.10513	0.493098	1.97277
TEM_PB_66	2.19712	0.837083	6.13809
NK_PB_100	0.991327	0.556119	2.25286
T_PB_100	0.960098	0.35621	1.5533
TN_PB_124	1.704508	0.723694	3.61918
TSCM_PB_124	0.909627	0.453047	1.82831
TCM_PB_124	1.941166	0.780557	4.55699
TEM_PB_124	1.660852	0.754719	4.07696

P10	Shannon	Simpson	inverseSimpson
NK_PB_34	0.710519	0.455293	1.83585
T_PB_34	1.397988	0.696314	3.29288
TN_PB_45	1.969127	0.771187	4.37038
TSCM_PB_45	2.022389	0.826966	5.7792
TCM_PB_45	1.658898	0.655994	2.90693
TEM_PB_45	2.615692	0.882173	8.48699
TN_PB_142	1.766004	0.756545	4.10753
TCM_PB_142	2.232625	0.826362	5.75911
TEM_PB_142	1.564321	0.636525	2.75122

Figure 6.6.2 IS data used to calculate IS diversity overtime in SCID-X1 samples.

P1 IS recaptures within TN
Number of captured units: **1118**
Abundance estimations and model fits:

	abundance	stderr	deviance	df	AIC	BIC
M0	2955.8	177.9	1083.509	5	1127.708	1137.747
Mt	1972.6	98.4	22.641	3	70.84	90.917
Mh Chao (LB)	3149.8	216.9	1078.348	4	1124.547	1139.605
Mh Poisson2	4034.5	648	1078.348	4	1124.547	1139.605
Mh Darroch	5304.4	1524.4	1078.348	4	1124.547	1139.605
Mh Gamma3.5	7155.6	3096.3	1078.348	4	1124.547	1139.605
Mth Chao (LB)	**2092.6**	**118.8**	**11.292**	**2**	**61.491**	**86.588**
Mth Poisson2	2823.5	391.6	11.292	2	61.491	86.588
Mth Darroch	4102.7	1105.8	11.292	2	61.491	86.588
Mth Gamma3.5	6379.4	2736.2	11.292	2	61.491	86.588
Mb	-53	10.7	63.774	4	109.973	125.031
Mbh	4.3	22.1	57.521	3	105.72	125.797

P5 IS recaptures within TN
Number of captured units: **1667**
Abundance estimations and model fits:

	abundance	stderr	deviance	df	AIC	BIC
M0	39526.2	7915.2	3173.327	5	3209.065	3219.903
Mt	5339.3	905.1	79.719	3	119.457	141.132
Mh Chao (LB)	46767.4	10329.3	3165.027	4	3202.765	3219.021
Mh Poisson2	225698.5	128462.9	3165.027	4	3202.765	3219.021
Mh Darroch	1114519.8	1087794.1	3165.027	4	3202.765	3219.021
Mh Gamma3.5	5646209.5	7877070.8	3165.027	4	3202.765	3219.021
Mth Chao (LB)	**6056.6**	**1155.2**	**69.11**	**2**	**110.848**	**137.942**
Mth Poisson2	31449.3	17633.6	69.11	2	110.848	137.942
Mth Darroch	203730.7	201410.9	69.11	2	110.848	137.942
Mth Gamma3.5	1407210.4	1994449.9	69.11	2	110.848	137.942
Mb	-0.6	0.8	304.095	4	341.833	358.09
Mbh	60	0.2	0.204	3	39.942	61.617

Figure 6.6.3 IS database used to calculate recaptured IS overtime in P1 and P5.

6.6.2.2 CARPALL

A final matrix M was generated where each row r represented an individual IS while each column c represented an individual cell type/sample and timepoint. Each entry of M contained the abundance of each r for each c in terms of sequencing reads. The data shown in Figs. 4.2 – 4.4 were generated on the basis of the IS databases and available for download from here. Graphical representations of IS analyses were generated with Prism8 (GraphPad Software) unless otherwise specified below. Panels in Fig. 4.2.2 were generated using the Circos software (http://circos.ca). Word clouds

in Fig. 4.3.1, 4.3.2 and 4.3.8 and 4.3.9 were generated on the basis of the relative incidence of the single closest gene to each IS and plotted using the online suite WordClouds (https://www.wordclouds.com). Panels with lines in Figs. 4.3.1, 4.3.2 and 4.3.8 and 4.3.9 were created by plotting IS diversity over time calculated as Shannon diversity or Gini–Simpson index through the R package BiodiversityR (https://cran.r-project.org/web/packages/BiodiversityR/index.html). The 'bubble' plots in Figs. 4.3.3 were created on the basis of the IS with abundance >0.01% relative to each subpopulation and timepoint using the R package packcircles (https://cran.r-project.org/web/packages/packcircles/index.html). The dendrograms in Fig. 4.3.5a were generated based on the similarity among IS datasets through the hclust function (https://www.rdocumentation.org/packages/fastcluster/versions/1.1.25/topics/hclust). Estimations of clonal abundance and standard errors shown in Fig. 4.3.5b were calculated by the conversion of M to an $M(0,1)$ matrix of incidence and by the application to $M(0,1)$ of log-linear models for closed populations through the R package Rcapture (https://cran.r-project.org/web/packages/Rcapture/index.html, function = closedp.t). The M_{th} (model for abundance estimates with time and heterogeneity effects) Chao LB method was selected for visualization in Fig. 4.3.5b, being the most conservative estimation among the ones with the lowest Bayesian information content. The heatmaps in Fig. 4.3.5c and 4.3.6b were generated through the use of the R package gplots (https://cran.r-project.org/web/packages/gplots/index.html). The correlograms in Fig. 4.4.1 were generated using the R package corrplot (https://cran.r-project.org/web/packages/corrplot/corrplot.pdf). Correlation values for both plots were generated through the function Mcorr.

6.7 nrLM-PCR

Genomic DNA extracted from cells purified from gene therapy patients' peripheral blood or bone marrow was used for shearing by sonication. Typically, 250-500 ng of DNA diluted in 100 ul of water was directly sheared by sonication, followed by purification with AMPure XP beads. This then was followed by end repair using *The NEBNext Ultra II End Repair/dA-Tailing Module* that is optimized to convert 500 pg - 1 µg of fragmented DNA to repaired DNA having 5´ phosphorylated, 3´ dA-tailed ends, and linker ligation. The ligation product was then amplified by two nested PCRs with primers specific for the vector LTR and the Linker cassette sequences. LM-PCR amplicons were separated on 1 % agarose gel to evaluate PCR efficiency and the bands pattern for each sample. LM-PCR products are then re-amplified with fusion primers containing specific sequences that allow paired end sequencing on an Illumina MiSeq sequencer. Fusion primers contain an 8-nucleotide barcode, both on the LTR and on the Linker cassette side. Each LM-PCR product was thus associated to a unique barcode couple, allowing multiplexing of up to 100 samples. Fusion-primer PCR products are assembled in the desired molar ratio into libraries avoiding repetition of identical barcodes, and sequenced.

6.7.1 Sonication

Tubes were prepared (Diagenode, 0.65 ml) containing 250-500ng of DNA in 100uL of H2O and were kept on ice until sonication.

Sonicator and the refrigerator unit were switched on 30 minutes before the start of sonication.

The program was set up as follows:

Fragment size	Condition	Cycles
1000 bp	5 sec "on" 90 sec "off"	8

Table 6.7.1 Sonication program.

Samples were kept on ice before and after the sonication.

6.7.2 DNA purification

The following reaction was prepared:

Reagent	Quantity (uL)
AMPure Beads (0.7:1 beads:DNA ratio)	70
DNA	100

Table 6.7.2 DNA purification with AMPure beads ratio.

Incubated DNA+beads for 5 minutes at room temperature. Placed tubes into the magnet and incubate for another 5 minutes. Removed the supernatant. Washed each tube with 400uL of EtOH 70% and placed it in the magnet for 5 minutes. Removed the volume and repeat washing. Eluted samples by adding 20uL of ddH2O and incubated for 5 minutes at room temperature. Placed in the magnet again for 5 minutes and carefully transferred the supernatant containing the cleaned DNA into a new labelled PCR tube.

6.7.3 Tape Station

The tape station (Agilent 4200) was used to check if fragment sizes of samples are within desired ranged with the majority of fragments ideally being within 1000 bp range (expect ed variability of fragment sizes). Fragments of 500-1300 bp sizes were considered to be within the accepted range.

The 4200 Tapestation system detects fluorescently stained double stranded DNA. Depending on the experiment either D1000 or D5000 screen tapes were used according to manufactures instructions. Briefly, for D1000 screen tape 3 ul of D1000 buffer was mixed with 1 ul of DNA. For D5000, 10 ul D5000 buffer was mixed with 1ul DNA.

Samples then were loaded and analysed on Agilent 42000 Tape Station. The raw data can be viewed as a electropherogram or a gel.

6.7.4 End preparation of fragmented DNA

After purification of fragments with AMPure beads **(section 6.7.2),** end repair mix was prepared as follows using NEBNext Ultra End Repair/dA-Tailing Module (with additional H2O control):

Reaction Components	Volume (µL) x1
Purified DNA	18
10X End Repair Buffer	6.5

End Prep Enzyme Mix	3
Mol. Grade Water	37.5
Final volume	65

Table 6.7.3 End prep master mix.

Aliquoted 47 µL of end prep mix per sample and mixed with 18 µL of DNA. Each reaction mixture (total volume 65 µL) was incubated at 20°C for 30 minutes and then at 65°C for 30 minutes in the thermocycler.

6.7.5 Linker ligation

Prepared linker ligation mix as follows:

Mastermix (per sample)	Volume (µL) x1
End-prepped DNA	65
Annealed linker (20 µM)	4
Ligation Enhancer	1
Blunt/TA Ligase Master Mix	15 µL
Final volume	85

Table 6.7.4 Linker ligation master mix.

Added 20 µL of ligation mix per sample to 65 µL of end prepped DNA.

Incubated samples overnight at 16°C in thermocycler.

After linker ligation repeated DNA purification with AMPure beads and eluted

samples in 15 µL.

Primer	Sequence (5' - 3')	Length
LC TA sonic sense	5' GACCCGGGAGATCTGAATTCAGTGGCAC AGCAGTTAGGNNNNNNGTAAGGNNNNN NAGATCTGGAATGAACTGGCC 3'	73
LC TA sonic anti-sense	5' TGGCCAGTTCATTCCAGATCT 3'	24

Table 6.7.5 Linker cassettes sequences.

6.7.6 1st exponential amplification PCR

Prepared 1st exponential PCR master mixes as follows using Clontech Advantage 2

PCR kit (added another water control):

PCR mix	Volume (µL) x 1
Clontech Advantage 2 PCR Buffer 10x	2.5

Primer specific for vector LTR **1st exp (eg SK LTR3)** 15uM	0.5
Primer specific for linker cassette **1st exp (eg LCP1)** 15uM	0.5
dNTPs 10 mM	0.5
Mol. Grade Water	5.5
Clontech Advantage 2 PCR Polymerase mix	0.5
Linker ligated DNA	15
Total volume	25

Table 6.7.6 1st exponential PCR master mix.

Mixed 10 µL of the mix with 15 µL of linker ligated DNA.

Run PCR as follows:

Cycle number	Denaturation	Annealing	Polymerization	Final
1	5 min at 95°C			

2-36	1 min at 95°C	45 sec at 60°C	90 sec at 72°C	
37			10 min at 72°C	
Hold				4°C

Table 6.7.7 1st exponential PCR thermocycler settings.

6.7.7 2nd exponential PCR

Prepared 2nd exponential PCR master mixes as follows using Clontech Advantage 2 PCR kit (added another water control):

PCR mix	Volume (µL) x 1
Clontech Advantage 2 PCR Buffer 10x	2.5
Primer specific for vector LTR **2nd exp (eg SK LTR4)** 15uM	0.5
Primer specific for linker cassette **2nd exp (eg LCP2)** 15uM	0.5
dNTPs 10 mM	0.5

Mol. Grade Water	13.5
Clontech Advantage 2 PCR Polymerase mix	0.5
1st Exp DNA	2
Total volume	20

Table 6.7.8 2nd Exponential PCR master mix.

Added 18 µL of the mix to 2 µL of DNA amplified with 1st exponential PCR (in PCR tubes).

Ran PCR as follows:

Cycle number	Denaturation	Annealing	Polymerization	Final
1	5 min at 95°C			
2-36	1 min at 95°C	45 sec at 60°C	90 sec at 72°C	
37			10 min at 72°C	
Hold				4°C

Table 6.7.9 Thermocycler set up for the 2nd exponential reaction.

6.7.8 Gel electrophoresis, Fusion PCR and sequencing

All these steps are identical to LAM-PCR protocol (see section 6.4).

6.8 Quantitative PCR (qPCR) for VCN analysis

Quantitative PCR is a molecular technique that allows amplification and quantification of known target DNA molecules.

To amplify target DNA sequence, sequence-specific primers and fluorescent probes are used. To determine vector copy number in SCID-X1 patients, primers and probes specific for gamma chain gene sequence were used (Table 6.8.1). The absolute number of copies of a target sequence is determined by using plasmidic standards. For every SCID-X1 qPCR, Apolipoprotein B (ApoB) was used as a housekeeping gene.

For quality control, all reactions are performed in triplicate or duplicate, and run alongside 1-copy control, non-template control (NTC) (DNase/RNase free water or elution buffer), and if available, negative control. All reactions were run in 96-well plates (BioRad) with 100ng DNA input, with TaqMan Universal PCR master mix (Applied Biosystems), in CFX96 Touch thermal cycler. Raw data taken from CFX96 Touch thermal cycler was analyzed using CFX Manager Software.

Calculation of vector copy number:

For SCID-X1 (plasmid standards):
The mean quantity value of the transgene was divided by the mean quantity value of the housekeeping gene and multiply the result by two to take into account the fact

236

there are two copies of the internal control per cell. The resulting value is the mean
VCN per cell.

Primer	Sequence
Gamma Chain Forward Primer (Invitrogen)	5' TGC TAA AAC TGC AGA ATC TGG T 3'
Gamma Chain Reverse (Invitrogen)	5' AGC TGG GAT TCAC TCA GTT TG 3'
Gamma chain Probe (Applied Biosytems)	5' FAM- CCT GGG CTC AGA GAA CCT AAC A -TAMRA 3'
ApoB Forward Primer (Invitrogen)	5' TGA AGG TGG AGG ACA TTC CTC TA 3'
ApoB Reverse Primer (Invitrogen)	5' CTG GAA TTG CGA TTT CTG GTA A 3'

ApoB Probe (Applied Biosystems)	5' VIC- CGA GAA TCA CCC TGC CAG ACT TCC GT -TAMRA 3'
gamma-chain/ApoB standards	Generated in house.

Table 6.8.1 Primer and probe sequences for VCN analysis of SCID-X1 samples by qPCR.

6.9 Digital PCR (ddPCR) for VCN analysis

Digital PCR is a PCR based assay that differs from qPCR by the way sample target is measured. It is also more sensitive than qPCR, therefore, especially useful for samples with low DNA concentrations. 20 ng DNA was added to master mix containing ddPCR supermix for probes (BioRad), primers specific for gamma chain gene probes and DNase/RNase free water. Each sample was partitioned into thousands of droplets using QX200 Auto droplet generator (BioRad). After droplet generation, a PCR reaction was performed in each droplet (using C1000 Touch thermal cycler (BioRad)), so that each individual reaction had 0 or 1 target molecules present. Then, plates were read by QX200 plate reader using QuantaSoft software. For normalization, in each reaction reference gene was amplified along with the target (the same as in qPCR). Used the same primers and probes like in qPCR but used ZEN instead of TAMRA as quencher for probes (Table 6.9.1).

Once the reading was complete, generated data was analyzed by the QuantaSoft software. Threshold was set to ensure clear separation between negative and positive events. PCR was considered successful if:

-The number of droplets generated per well was >10,000/well

-The 1-copy number control was between 0.7-1.3

-Each channel had at least 5% negative droplets

Primer	Sequence
Gamma Chain Forward Primer (Invitrogen)	5' TGC TAA AAC TGC AGA ATC TGG T 3'
Gamma Chain Reverse (Invitrogen)	5' AGC TGG GAT TCAC TCA GTT TG 3'
Gamma chain Probe	5' FAM- CCT GGG CTC AGA GAA CCT AAC A -ZEN 3'

(Applied Biosytems)	
ApoB Forward Primer (Invitrogen)	5' TGA AGG TGG AGG ACA TTC CTC TA 3'
ApoB Reverse Primer (Invitrogen)	5' CTG GAA TTG CGA TTT CTG GTA A 3'
ApoB Probe (Applied Biosystems)	5' HEX- CGA GAA TCA CCC TGC CAG ACT TCC GT -ZEN 3'

Table 6.9.1 Primer and probe sequences for VCN analysis of SCID-X1 samples by ddPCR.

6.10 IFN-gamma production assay

Prior to stimulation with IFN gamma, cells from the patient and healthy donor (HD) were FACS sorted into TN, TSCM, TCM and TEM subsets. After sorting, minimum of 10,000 cells were used for IFN-gamma assay. Cells were re-suspended in 100 µl medium (RPMI) containing 10% human AB serum and 1% penicillin-streptomycin.

PMA (50 ng/ml) and ionomycin (1 µg/ml) were diluted in medium to the right concentration and medium containing both was added to samples. In case, there was a sufficient number of cells acquired after sorting – samples were split into stimulated and un-stimulated controls. After PMA and ionomycin were added, cells were incubated in the incubator at 37°C with 5% CO_2 for 6 hours.

After incubation, brefeldin A (10 µg/ml) was added to the media, and cells were further incubated for 2 hours. At the end of incubation, cells were permed and fixed with BD Cytofix/Cytoperm™ for 20 minutes at 4°C. FITC conjugated IFN-gamma antibody (BD Biosciences) was added to fixed/permeabilized cells in 50 µL of BD Perm/Wash™ buffer (Table 6.10.1). Cells were incubated for 30 minutes at 4°C in the dark. After, cells were washed 2 times with 1× BD Perm/Wash™ buffer, re-suspended in fixation buffer and analyzed with BD FACSCalibur.

Data was analyzed using Cell Quest software.

Antibody	Clone	Fluorochrome	Source	Cat number
IFN-gamma	B27	FITC	BD Pharmigen	557718

Table 6.10.1 IFN-gamma antibody for IFN-gamma production assay.

6.11 Vbeta spectra typing analysis

CDR3 TCR spectratyping was performed as described previously (King, et al., 2001). Briefly, RNA was extracted, and complementary DNA (cDNA) was prepared from

CD3+, CD4+, and/or CD8+ T cell subsets. Twenty-four Vβ-specific primers were used with a fluorescent-labelled constant region (Cβ)-specific primer to RT-PCR (reverse transcription-PCR)-amplify the CDR3 region of the TCRβ chain (Table 6.11.1). The labelled products were loaded onto an ABI 310 Automate (Perkin-Elmer Applied Biosystems, UK), which measures the intensity of fluorescence (the amount of PCR product), and, by reference to the detection time of the internal size standards, the length of the PCR products was determined and analysed with SpA-based software (Gerstel-Thompson, et al., 2010).

Primer	Sequence
Cβ primer	GGGTGTGGGAGATCTCTGC
Cβ primer (FAM labelled)	ACACAGCAGCCTCGGGTGGG
Vβ1	CCGCACAACAGTTCCCTGACTTGC
Vβ2	CACAACTATGTTTTGGTATCGTC
Vβ3	CGCTTCTCCCTGATTCTGGAGTCC
Vβ4	TTCCCATCAGCCGCCCAAACCTAA
Vβ5	GATCAAAACGAGAGGACAGC
Vβ6a	GATCCAATTTCAGGTCATACTG
Vβ6b1	CAGGGSCCAGAGTTTCTGAC
Vβ6b2	AGGGCTCAGAGGTTCTGAC

Vβ7	CCTGAATGCCCCAACAGCTCT
Vβ8	GGTACAGACAGACCATGATGC
Vβ9	TTCCCTGGAGCTTGGTGACTCTGC
Vβ11	GTCAACAGTCTCCAGAATAAGG
Vβ12	TCCYCCTCACTCTGGAGTC
Vβ13a	GGGTGTGGGAGATCTCTGC
Vβ13b	AGGCTCATCCATTATTCAAATAC
Vβ14	GGGCTGGGCTTAAGGCAGATCTAC
Vβ15	CAGGCACAGGCTAAATTCTCCCTG
Vβ16	GCCTGCAGAACTGGAGGATTCTGG
Vβ17	TCCTCTCACTGTGACATCGGCCCA
Vβ18	CTGCTGAATTTCCCAAAGAGGGCC
Vβ20	TGCCCAGAATCTCTCAGCCTCCA
Vβ21	GGAGTAGACTCCACTCTCAAG
Vβ22	GATCCGGTCCACAAAGCTGG
Vβ23	ATTCTGAACTGAACATGAGCTCCT
Vβ24	GACATCCGCTCACCAGGCCTG

Table 6.11.1 Primer sequences used for Vbeta spectratyping.

6.12 TREC analysis

Real-time qPCR targeting a specific marker of functional T cells, the TREC, was performed as described previously (Gerstel-Thompson, et al., 2010). Briefly, DNA was extracted from CD3+, CD4+, and/or CD8+ T cell subsets and subjected to a multiplex qPCR to amplify TRECs and the RNaseP housekeeping gene (Table 6.12.1). By reference to standard curves, generated with a TREC-containing plasmid (Douek, et al., 1998) and dilutions of genomic DNA for the RNaseP gene, TREC numbers were calculated for each sample.

Forward and reverse primers targeting δRec-ΨJα TREC-specific sequences were used to generate a 93-bp amplicon spanning the splice junction, with the TREC probe located just downstream from the junction. The qPCR included primers and probe for an attenuated amplification of the RNase P gene RPPH1. The 20-µL reaction consisted of 10 µL TaqMan® Fast Universal PCR Master Mix (4367846; Applied Biosystems), 0.4 µL TaqMan RNase P Vic Control Reagent (4316844; Applied Biosystems), and TREC primers and probe sequences (Applied Biosystems) located within the gene identified by accession number [NT_026437, nucleotides (forward primer) 3944229 through 3944289, and (reverse primer) 3855229 through 3855280] in the following concentrations: 8 pmol each of forward TREC primer (TGCTGACACCTCTGGTTTTTGTAA) and reverse TREC primer (GTGCCAGCTGCAGGGTTTAG), 3 pmol TREC-specific hydrolysis probe (6FAM-ATGCATAGGCACCTGC-MGB), and 5 µL of DNA eluate. Absolute qPCR was performed in an Applied Biosystems 7900 HT Real-Time PCR System in a 384-well plate (4343814; Applied Biosystems).

Name	What for	Sequence
TREC	Forward primer	5' AGATTTGGACCTGCGAGCG 3'
TREC	Reverse primer	5' GAGCGGCTGTCTCCACAAGT 3'
TREC	Probe	6FAM-ATGCATAGGCACCTGC-MGB

Table 6.12.1 Primer and probe sequences for TREC analysis.

7 References

Abdellah, Z. *et al.* (2004) 'Finishing the euchromatic sequence of the human genome', *Nature*, 431(7011). doi: 10.1038/nature03001.

Accetta Pedersen, D. J., Verbsky, J. and Routes, J. M. (2011) 'Screening newborns for primary T-cell immunodeficiencies: Consensus and controversy', *Expert Review of Clinical Immunology*. doi: 10.1586/eci.11.25.

Adolfsson, J. *et al.* (2005) 'Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: A revised road map for adult blood lineage commitment', *Cell*, 121(2), pp. 295–306. doi: 10.1016/j.cell.2005.02.013.

Adusumilli, P. S. *et al.* (2014) 'Regional delivery of mesothelin-targeted CAR T cell therapy generates potent and long-lasting CD4-dependent tumor immunity', in *Science Translational Medicine*. p. 261ra151: Volume 6. doi: 10.1126/scitranslmed.3010162.

Agaugué, S. *et al.* (2008) 'Human natural killer cells exposed to IL-2, IL-12, IL-18, or IL-4 differently modulate priming of naive T cells by monocyte-derived dendritic cells', *Blood*, 112(5), pp. 1776–1783. doi: 10.1182/blood-2008-02-135871.

Agrawal, A. (2000) 'Transposition and evolution of antigen-specific immunity', *Science*, 290(5497), pp. 1715–1716. doi: 10.1126/science.290.5497.1715.

Ahlers, J. D. and Belyakov, I. M. (2010) 'Memories that last forever: Strategies for optimizing vaccine T-cell memory', *Blood*, 115(9), pp. 1678–1689. doi: 10.1182/blood-2009-06-227546.

Ahmed, R. *et al.* (2016) 'Human Stem Cell-like Memory T Cells Are Maintained in a State of Dynamic Flux', *Cell Reports*, 17(11), pp. 2811–2818. doi: 10.1016/j.celrep.2016.11.037.

Ahmed, R. and Gray, D. (1996) 'Immunological memory and protective immunity: Understanding their relation', *Science*, 272(5258), pp. 54–60. doi: 10.1126/science.272.5258.54.

Aiuti, A. *et al.* (2007) 'Multilineage hematopoietic reconstitution without clonal selection in ADA-SCID patients treated with stem cell gene therapy', *Journal of Clinical Investigation*, 117(8). doi: 10.1172/JCI31666.

Aiuti, A. *et al.* (2013) 'Lentiviral hematopoietic stem cell gene therapy in patients with wiskott-aldrich syndrome', *Science*, 341(6148). doi: 10.1126/science.1233151.

Akashi, K. *et al.* (2000) 'A clonogenic common myeloid progenitor that gives rise to all myeloid lineages', *Nature*, 404(6774), pp. 193–197. doi: 10.1038/35004599.

Akondy, R. S. *et al.* (2017) 'Origin and differentiation of human memory CD8 T cells after vaccination', *Nature*, 552(7685), pp. 362–367. doi: 10.1038/nature24633.

Al-Lazikani, B., Lesk, A. M. and Chothia, C. (2000) 'Canonical structures for the hypervariable regions of T cell αβ receptors', *Journal of Molecular Biology*, 295(4). doi: 10.1006/jmbi.1999.3358.

AlDallal, S. M. (2020) 'Yescarta: A New Era for Non-Hodgkin Lymphoma Patients', *Cureus*. doi: 10.7759/cureus.11504.

del Amo, P. C. *et al.* (2018) 'Human T SCM cell dynamics in vivo are compatible with long-lived immunological memory and stemness', *PLoS Biology*, 16(6), pp. 1–22. doi: 10.1371/journal.pbio.2005523.

Arstila, T. P. *et al.* (1999) 'A direct estimate of the human αβ T cell receptor diversity', *Science*, 286(5441), pp. 958–961. doi: 10.1126/science.286.5441.958.

Bäckström, B. T. *et al.* (1996) 'A motif within the T cell receptor α chain constant region connecting peptide domain controls antigen responsiveness', *Immunity*, 5(5), pp. 437–447. doi: 10.1016/S1074-7613(00)80500-2.

Bandyopadhyay, K., Marrero, I. and Kumar, V. (2016) 'NKT cell subsets as key participants in liver physiology and pathology', *Cellular and Molecular Immunology*, 13(3), pp. 337–346. doi: 10.1038/cmi.2015.115.

Bartlett, J. G. (2004) 'Duration of antiviral immunity after smallpox vaccination', *Infectious Diseases in Clinical Practice*, 12(3), p. 205. doi: 10.1097/01.idc.0000129853.80250.2c.

Bartram, J. *et al.* (2016) 'Accurate Sample Assignment in a Multiplexed, Ultrasensitive, High-Throughput Sequencing Assay for Minimal Residual Disease', *Journal of Molecular Diagnostics*, 18(4), pp. 494–506. doi: 10.1016/j.jmoldx.2016.02.008.

Bauer, T. R. *et al.* (2006) 'Correction of the disease phenotype in canine leukocyte adhesion deficiency using ex vivo hematopoietic stem cell gene therapy', *Blood*, 108(10). doi: 10.1182/blood-2006-03-006908.

Baum, C. *et al.* (2007) 'Gene therapy of SCID-X1', *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.*, 50(12), pp. 1507–1517.

Baum, C. (2007) 'Insertional mutagenesis in gene therapy and stem cell biology', *Current Opinion in Hematology*, 14(4), pp. 337–342. doi: 10.1097/MOH.0b013e3281900f01.

Baum, C. M. *et al.* (1992) 'Isolation of a candidate human hematopoietic stem-cell population', in *Proceedings of the National Academy of Sciences of the United States of America*. Volume 89: Acad. Sci, pp. 2804–2808. doi: 10.1073/pnas.89.7.2804.

Bayne, C. J. (2003) 'Origins and evolutionary relationships between the innate and adaptive arms of immune systems', *Integrative and Comparative Biology*, 43(2), pp. 293–299. doi: 10.1093/icb/43.2.293.

Benveniste, P. *et al.* (2010) 'Intermediate-Term Hematopoietic Stem Cells with Extended but Time-Limited Reconstitution Potential', *Cell Stem Cell*, 6(1), pp. 48–58. doi: 10.1016/j.stem.2009.11.014.

Berzins, S. P., Smyth, M. J. and Baxter, A. G. (2011) 'Presumed guilty: Natural killer T cell defects and human disease', *Nature Reviews Immunology*, 11(2), pp. 131–142. doi: 10.1038/nri2904.

Best, K. *et al.* (2015) 'Computational analysis of stochastic heterogeneity in PCR amplification efficiency revealed by single molecule barcoding', *Scientific Reports*, 5. doi: 10.1038/srep14629.

Bezman, N. A. *et al.* (2012) 'Molecular definition of the identity and activation of natural killer cells', *Nature Immunology*, 13(10), pp. 1000–1008. doi: 10.1038/ni.2395.

Bhandoola, A. *et al.* (2007) 'Commitment and Developmental Potential of Extrathymic and Intrathymic T Cell Precursors: Plenty to Choose from', *Immunity*, pp. 678–689. doi: 10.1016/j.immuni.2007.05.009.

Bhatia, M. *et al.* (1997) 'Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice', in *Proceedings of the National Academy of Sciences of the United States of America*. Volume 94: Acad. Sci, pp. 5320–5325. doi: 10.1073/pnas.94.10.5320.

Biasco, L. *et al.* (2011) 'Integration profile of retroviral vector in gene therapy treated patients is cell-specific according to gene expression and chromatin conformation of target cell', *EMBO Molecular Medicine*, 3(2), pp. 89–101. doi: 10.1002/emmm.201000108.

Biasco, L. *et al.* (2015) 'In vivo tracking of T cells in humans unveils decade-long survival and activity of genetically modified T memory stem cells', in *Science Translational Medicine*. p. 273ra13: Volume 7, p. 273ra13. doi: 10.1126/scitranslmed.3010314.

Biasco, L. *et al.* (2016) 'In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases', *Cell Stem Cell*, 19(1), pp. 107–119. doi: 10.1016/j.stem.2016.04.016.

Biasco, L. *et al.* (2018) 'Analyzing the Genotoxicity of Retroviral Vectors in

Hematopoietic Cell Gene Therapy', *Molecular Therapy - Methods and Clinical Development*, 8(March), pp. 21–30. doi: 10.1016/j.omtm.2017.10.002.

Birnbaum, M. E., Mendoza, J. L., *et al.* (2014) 'Deconstructing the peptide-MHC specificity of t cell recognition', *Cell*, 157(5), pp. 1073–1087. doi: 10.1016/j.cell.2014.03.047.

Birnbaum, M. E., Berry, R., *et al.* (2014) 'Molecular architecture of the αβ T cell receptor-CD3 complex', *Proceedings of the National Academy of Sciences of the United States of America*, 111(49), pp. 17576–17581. doi: 10.1073/pnas.1420936111.

Blanco, L. *et al.* (1989) 'Highly efficient DNA synthesis by the phage Φ29 DNA polymerase. Symmetrical mode of DNA replication', *Journal of Biological Chemistry*, 264(15). doi: 10.13039/100000002.

Borg, N. al. (2007) 'CD1d-lipid-antigen recognition by the semi-invariant NKT T-cell receptor', *Nature*, 448, pp. 44–49.

Bovee, D. *et al.* (2008) 'Closing gaps in the human genome with fosmid resources generated from multiple individuals', *Nature Genetics*, 40(1). doi: 10.1038/ng.2007.34.

den Braber, I. *et al.* (2012) 'Maintenance of Peripheral Naive T Cells Is Sustained by Thymus Output in Mice but Not Humans', *Immunity*, 36(2). doi: 10.1016/j.immuni.2012.02.006.

Brady, T. *et al.* (2011) 'A method to sequence and quantify DNA integration for monitoring outcome in gene therapy', *Nucleic Acids Research*, 39(11). doi:

10.1093/nar/gkr140.

Brugman, M. H. *et al.* (2013) 'Evaluating a ligation-mediated pcr and pyrosequencing method for the detection of clonal contribution in polyclonal retrovirally transduced samples', *Human Gene Therapy Methods*, 24(2). doi: 10.1089/hgtb.2012.175.

Brugman, M. H. *et al.* (2015) 'Development of a diverse human T-cell repertoire despite stringent restriction of hematopoietic clonality in the thymus', *Proceedings of the National Academy of Sciences of the United States of America*, 112(44), pp. E6020–E6027. doi: 10.1073/pnas.1519118112.

Bystrykh, L. V. *et al.* (2012) 'Counting stem cells: Methodological constraints', *Nature Methods*. doi: 10.1038/nmeth.2043.

Caligiuri, M. A. (2008) 'Human natural killer cells', *Blood*, 112(3), pp. 461–469. doi: 10.1182/blood-2007-09-077438.

Calvi, L. M. *et al.* (2003) 'Osteoblastic cells regulate the haematopoietic stem cell niche', *Nature*, 425(6960), pp. 841–846. doi: 10.1038/nature02040.

Capuano, C. *et al.* (2019) 'Memory NK cell features exploitable in anticancer immunotherapy', *Journal of Immunology Research*, 2019. doi: 10.1155/2019/8795673.

Di Carlo, E. *et al.* (2013) 'Mechanisms of the antitumor activity of human Vγ9Vδ2 T cells in combination with zoledronic acid in a preclinical model of neuroblastoma', in *Molecular Therapy*. doi: 10.1038/mt.2013.38.

Carpenito, C. *et al.* (2009) 'Control of large, established tumor xenografts with

genetically retargeted human T cells containing CD28 and CD137 domains', *Proceedings of the National Academy of Sciences of the United States of America*, 106(9), pp. 3360–3365. doi: 10.1073/pnas.0813101106.

Cattoglio, C. *et al.* (2007) 'Hot spots of retroviral integration in human CD34+ hematopoietic cells', *Blood*, 110(6). doi: 10.1182/blood-2007-01-068759.

Cavazzana-Calvo, M. *et al.* (2000) 'Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease', *Science*, 288(5466), pp. 669–672. doi: 10.1126/science.288.5466.669.

Cavazzana-Calvo, M. *et al.* (2011) 'Is normal hematopoiesis maintained solely by long-term multipotent stem cells?', *Blood*, 117(17), pp. 4420–4424. doi: 10.1182/blood-2010-09-255679.

Caza, T. and Landas, S. (2015) *Functional and Phenotypic Plasticity of CD4+ T Cell Subsets*, *BioMed Research International*. Biomed Res Int. doi: 10.1155/2015/521957.

Chan, A. *et al.* (2007) 'CD56bright human NK cells differentiate into CD56dim cells', *Journal of Immunology*, 179, pp. 89–94.

Chang, J. T. (2012) 'Polarity and lymphocyte fate determination', *Current Opinion in Cell Biology*, 24(4), pp. 526–533. doi: 10.1016/j.ceb.2012.05.002.

Chmielewski, M. *et al.* (2004) 'T Cell Activation by Antibody-Like Immunoreceptors: Increase in Affinity of the Single-Chain Fragment Domain above Threshold Does Not Increase T Cell Activation against Antigen-Positive Target Cells but Decreases

Selectivity', *The Journal of Immunology*, 173(12), pp. 7647–7653. doi: 10.4049/jimmunol.173.12.7647.

Cieri, N. *et al.* (2013) 'IL-7 and IL-15 instruct the generation of human memory stem T cells from naive precursors', *Blood*, 121(4). doi: 10.1182/blood-2012-05-431718.

Civin, C. I. *et al.* (1984) 'Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells.', *Journal of immunology (Baltimore, Md. : 1950)*, 133(1), pp. 157–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6586833.

Clemente, C. G. *et al.* (1996) 'Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma', *Cancer*, 77(7), pp. 1303–1310. doi: 10.1002/(SICI)1097-0142(19960401)77:7<1303::AID-CNCR12>3.0.CO;2-5.

Colonna, M., Navarro, F. and López-Botet, M. (1999) 'A novel family of inhibitory receptors for HLA class I molecules that modulate function of lymphoid and myeloid cells', *Current Topics in Microbiology and Immunology*, 244(3), pp. 115–122. doi: 10.1007/978-3-642-58537-1_10.

Conneally, E. *et al.* (1997) 'Expansion in vitro of transplantable human cord blood stem cells demonstrated using a quantitative assay of their lympho-myeloid repopulating activity in nonobese diabetic-scid/scid mice', in *Proceedings of the National Academy of Sciences of the United States of America*. Volume 94: Acad. Sci, pp. 9836–9841. doi: 10.1073/pnas.94.18.9836.

Cooper, M. A., Fehniger, T. A. and Caligiuri, M. A. (2001) 'The biology of human natural

killer-cell subsets', *Trends in Immunology*, 22(11), pp. 633–640. doi: 10.1016/S1471-4906(01)02060-9.

Cowan, J. E., Jenkinson, W. E. and Anderson, G. (2015) 'Thymus medulla fosters generation of natural Treg cells, invariant γδ T cells, and invariant NKT cells: What we learn from intrathymic migration', *European Journal of Immunology*, 45(3), pp. 652–660. doi: 10.1002/eji.201445108.

Croft, N. P. *et al.* (2019) 'Most viral peptides displayed by class I MHC on infected cells are immunogenic', *Proceedings of the National Academy of Sciences of the United States of America*, 116(8), pp. 3112–3117. doi: 10.1073/pnas.1815239116.

Dai, Q. *et al.* (2020) '4-1BB Signaling Boosts the Anti-Tumor Activity of CD28-Incorporated 2nd Generation Chimeric Antigen Receptor-Modified T Cells', *Frontiers in Immunology*, 11, pp. 539–654. doi: 10.3389/fimmu.2020.539654.

Daniels, M. A. and Teixeiro, E. (2015) 'TCR signaling in T cell memory', *Frontiers in Immunology*, 6(DEC). doi: 10.3389/fimmu.2015.00617.

Davila, M. L. *et al.* (2014) 'Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia', *Science Translational Medicine*, 6(224), p. 224. doi: 10.1126/scitranslmed.3008226.

Davila, M. L., Sauter, C. and Brentjens, R. (2015) 'CD19-targeted t cells for hematologic malignancies clinical experience to date', *Cancer Journal (United States)*, 21(6), pp. 470–474. doi: 10.1097/PPO.0000000000000153.

Davis, M. M. and Bjorkman, P. J. (1988) 'T-cell antigen receptor genes and T-cell recognition', *Nature*, 334(6181), pp. 395–402. doi: 10.1038/334395a0.

Desfarges, S. and Ciuffi, A. (2010) 'Retroviral integration site selection', *Viruses*, 2(1), pp. 111–130. doi: 10.3390/v2010111.

Dik, W. A. *et al.* (2005) 'New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling', *Journal of Experimental Medicine*, 201(11), pp. 1715–1723. doi: 10.1084/jem.20042524.

Dimova, T. *et al.* (2015) 'Effector vγ9vδ2 t cells dominate the human fetal γδ t-cell repertoire', *Proceedings of the National Academy of Sciences of the United States of America*, 112(6), pp. E556–E565. doi: 10.1073/pnas.1412058112.

Dogra, P. *et al.* (2020) 'Tissue Determinants of Human NK Cell Development, Function, and Residence', *Cell*, 180(4), pp. 749-763.e13. doi: 10.1016/j.cell.2020.01.022.

Dong, D. *et al.* (2019) 'Structural basis of assembly of the human T cell receptor–CD3 complex', *Nature*, 573(7775), pp. 546–552. doi: 10.1038/s41586-019-1537-0.

Douek, D. C. *et al.* (2000) 'Assessment of thymic output in adults after haematopoietic stem-cell transplantation and prediction of T-cell reconstitution', *Lancet*, 355(9218), pp. 1875–1881. doi: 10.1016/S0140-6736(00)02293-5.

Doulatov, S. *et al.* (2010) 'Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development', *Nature*

Immunology, 11(7), pp. 585–593. doi: 10.1038/ni.1889.

Doulatov, S. *et al.* (2012) 'Hematopoiesis: A human perspective', *Cell Stem Cell*, 10(2), pp. 120–136. doi: 10.1016/j.stem.2012.01.006.

Dull, T. *et al.* (1998) 'A Third-Generation Lentivirus Vector with a Conditional Packaging System', *Journal of Virology*, 72(11). doi: 10.1128/jvi.72.11.8463-8471.1998.

Eger, K. A. *et al.* (2006) 'Human natural killer T cells are heterogeneous in their capacity to reprogram their effector functions', *PLoS ONE*, 1(1), p. 1. doi: 10.1371/journal.pone.0000050.

Farber, D. L. *et al.* (2016) 'Immunological memory: Lessons from the past and a look to the future', *Nature Reviews Immunology*, 16(2), pp. 124–128. doi: 10.1038/nri.2016.13.

Fathman, J. W. *et al.* (2011) 'Identification of the earliest natural killer cell-committed progenitor in murine bone marrow', *Blood*, 118(20), pp. 5439–5447. doi: 10.1182/blood-2011-04-348912.

Fauriat, C. *et al.* (2010) 'Education of human natural killer cells by activating killer cell immunoglobulin-like receptors', *Blood*, 115(6), pp. 1166–1174. doi: 10.1182/blood-2009-09-245746.

Fehniger, T. A. *et al.* (1999) 'Differential cytokine and chemokine gene expression by human NK cells following activation with IL-18 or IL-15 in combination with IL-12:

implications for the innate immune response.', *Journal of immunology (Baltimore, Md. : 1950)*, 162(8), pp. 4511–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10201989.

Fehniger, T. A. *et al.* (2003) 'CD56bright natural killer cells are present in human lymph nodes and are activated by T cell-derived IL-2: A potential new link between adaptive and innate immunity', *Blood*, 101(8), pp. 3052–3057. doi: 10.1182/blood-2002-09-2876.

Feins, S. *et al.* (2019) 'An introduction to chimeric antigen receptor (CAR) T-cell immunotherapy for human cancer', *American Journal of Hematology*, 94(S1), pp. S3–S9. doi: 10.1002/ajh.25418.

Felice, B. *et al.* (2009) 'Transcription factor binding sites are genetic determinants of retroviral integration in the human genome', *PLoS ONE*, 4(2). doi: 10.1371/journal.pone.0004571.

Fernández-Quintero, M. L. *et al.* (2020) 'T-Cell Receptor CDR3 Loop Conformations in Solution Shift the Relative Vα-Vβ Domain Distributions', *Frontiers in Immunology*, 11. doi: 10.3389/fimmu.2020.01440.

Ferry, C. and Socié, G. (2003) 'Busulfan-cyclophosphamide versus total body irradiation-cyclophosphamide as preparative regimen before allogeneic hematopoietic stem cell transplantation for acute myeloid leukemia: What have we learned?', *Experimental Hematology*, 31(12), pp. 1182–1186. doi: 10.1016/j.exphem.2003.09.008.

Finney, H. M. *et al.* (1998) 'Chimeric receptors providing both primary and costimulatory signaling in T cells from a single gene product.', *Journal of immunology (Baltimore, Md. : 1950)*, 161(6), pp. 2791–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9743337.

Finney, H. M., Akbar, A. N. and Lawson, A. D. G. (2004) 'Activation of Resting Human Primary T Cells with Chimeric Receptors: Costimulation from CD28, Inducible Costimulator, CD134, and CD137 in Series with Signals from the TCRζ Chain', *The Journal of Immunology*, 172(1), pp. 104–113. doi: 10.4049/jimmunol.172.1.104.

Firouzi, S. *et al.* (2014) 'Development and validation of a new high-throughput method to investigate the clonality of HTLV-1-infected cells based on provirus integration sites', *Genome Medicine*, 6(6), pp. 29–32. doi: 10.1186/gm568.

Foley, B. *et al.* (2012) ' Human Cytomegalovirus (CMV)-Induced Memory-like NKG2C + NK Cells Are Transplantable and Expand In Vivo in Response to Recipient CMV Antigen ', *The Journal of Immunology*, 189(10), pp. 5082–5088. doi: 10.4049/jimmunol.1201964.

Fraietta, J. A. *et al.* (2021) 'Author Correction: Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia (Nature Medicine, (2018), 24, 5, (563-571), 10.1038/s41591-018-0010-1)', *Nature Medicine*, 27(3), p. 561. doi: 10.1038/s41591-021-01248-2.

Freud, A. G. *et al.* (2005) 'A human CD34(+) subset resides in lymph nodes and differentiates into CD56bright natural killer cells', *Immunity*, 22(3), pp. 295–304. doi: 10.1016/j.immuni.2005.01.013.

Freud, A. G. *et al.* (2017) 'The Broad Spectrum of Human Natural Killer Cell Diversity', *Immunity*, 47(5), pp. 820–833. doi: 10.1016/j.immuni.2017.10.008.

Freud, A. G., Yu, J. and Caligiuri, M. A. (2014) 'Human natural killer cell development in secondary lymphoid tissues', *Seminars in Immunology*, 26(2), pp. 132–137. doi: 10.1016/j.smim.2014.02.008.

Friedmann-Morvinski, D. *et al.* (2005) 'Redirected primary T cells harboring a chimeric receptor require costimulation for their antigen-specific activation', *Blood*, 105(8), pp. 3087–3093. doi: 10.1182/blood-2004-09-3737.

Fuertes Marraco, S. A. *et al.* (2015) 'Genome-wide RNA profiling of long-lasting stem cell-like memory CD8 T cells induced by Yellow Fever vaccination in humans', *Genomics Data*, 5. doi: 10.1016/j.gdata.2015.06.024.

Gabriel, R. *et al.* (2009) 'Comprehensive genomic access to vector integration in clinical gene therapy', *Nature Medicine*, 15(12), pp. 1431–1436. doi: 10.1038/nm.2057.

Galy, A. *et al.* (1993) 'Precursors of CD3+ CD4+ CD8+ cells in the human thymus are defined by expression of CD34. Delineation of early events in human thymic development', *Journal of Experimental Medicine*, 178(2), pp. 391–401. doi: 10.1084/jem.178.2.391.

Garboczi, D. N. *et al.* (1996) 'Structure of the complex between human T-cell receptor, viral peptide and HLA-A2', *Nature*, 384(6605), pp. 134–141. doi: 10.1038/384134a0.

Gardner, R. A. *et al.* (2017) 'Intent-to-treat leukemia remission by CD19 CAR T cells

of defined formulation and dose in children and young adults', *Blood*, 129(25), pp. 3322–3331. doi: 10.1182/blood-2017-02-769208.

Garfall, A. L. *et al.* (2015) 'Chimeric Antigen Receptor T Cells against CD19 for Multiple Myeloma', *New England Journal of Medicine*, 373(11). doi: 10.1056/nejmoa1504542.

Gaspar, H. B. *et al.* (2004) 'Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector', *Lancet*, 364(9452). doi: 10.1016/S0140-6736(04)17590-9.

Gaspar, H. B. *et al.* (2006) 'Successful Reconstitution of Immunity in ADA-SCID by Stem Cell Gene Therapy Following Cessation of PEG-ADA and Use of Mild Preconditioning', *Molecular Therapy*, 14(4). doi: 10.1016/j.ymthe.2006.06.007.

Gaspar, H. B., Cooray, S., Gilmour, K. C., Parsley, K. L., Zhang, F., *et al.* (2011) 'Immunodeficiency: Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction', *Science Translational Medicine*, 3(97). doi: 10.1126/scitranslmed.3002716.

Gaspar, H. B., Cooray, S., Gilmour, K. C., Parsley, K. L., Adams, S., *et al.* (2011) *Immunodeficiency: Long-term persistence of a polyclonal t cell repertoire after gene therapy for X-linked severe combined immunodeficiency*, *Science Translational Medicine*. Science Translational Medicine. doi: 10.1126/scitranslmed.3002715.

Gasper, D. J., Tejera, M. M. and Suresh, M. (2014) 'CD4 T-cell memory generation and maintenance', *Critical Reviews in Immunology*, 34(2), pp. 121–146. doi:

10.1615/CritRevImmunol.2014010373.

Gattinoni, L. *et al.* (2011) 'A human memory T cell subset with stem cell-like properties', *Nature Medicine*, 17(10), pp. 1290–1297. doi: 10.1038/nm.2446.

Gaud, G., Lesourne, R. and Love, P. E. (2018) 'Regulatory mechanisms in T cell receptor signalling', *Nature Reviews Immunology*, 18(8), pp. 485–497. doi: 10.1038/s41577-018-0020-8.

Geginat, J. *et al.* (2014) 'Plasticity of human CD4 T cell subsets', *Frontiers in Immunology*, 5(DEC). doi: 10.3389/fimmu.2014.00630.

Geginat, J., Lanzavecchia, A. and Sallusto, F. (2003) 'Proliferation and differentiation potential of human CD8+ memory T-cell subsets in response to antigen or homeostatic cytokines', *Blood*, 101(11), pp. 4260–4266. doi: 10.1182/blood-2002-11-3577.

Germain, R. N. (2002) 't-cell development and the CD4-CD8 lineage decision', *Nature Reviews Immunology*, 2(5), pp. 309–322. doi: 10.1038/nri798.

Ghorashian, S. *et al.* (2019) 'Enhanced CAR T cell expansion and prolonged persistence in pediatric patients with ALL treated with a low-affinity CD19 CAR', *Nature Medicine*, 25(9), pp. 1408–1414. doi: 10.1038/s41591-019-0549-5.

Gordon, M. Y., Lewis, J. L. and Marley, S. B. (2002) 'Of mice and men . . . and elephants [3]', *Blood*. doi: 10.1182/blood-2002-08-2517.

Green, M. R. and Sambrook, J. (2019) 'Nested polymerase chain reaction (PCR)', *Cold Spring Harbor Protocols*, 2019(2). doi: 10.1101/pdb.prot095182.

Greene, M. R. *et al.* (2013) 'Clonal Diversity Analysis Of Integrating Vector Transduced Hematopoietic Cells Using LAM-PCR Followed By Illumina-Based Next-Generation Sequencing Is Affected By False Positivity That Arises From Both Reaction Biochemistry and Bioinformatic Analysis', *Blood*, 122(21). doi: 10.1182/blood.v122.21.1662.1662.

Grisedale, K. and van Daal, A. (2014) 'Linear amplification of target prior to PCR for improved low template DNA results', *BioTechniques*, 56(3), pp. 145–147. doi: 10.2144/000114148.

Guedan, S. *et al.* (2019) 'Engineering and Design of Chimeric Antigen Receptors', *Molecular Therapy - Methods and Clinical Development*, 12, pp. 145–156. doi: 10.1016/j.omtm.2018.12.009.

Gumá, M. *et al.* (2006) 'Expansion of CD94/NKG2C+ NK cells in response to human cytomegalovirus-infected fibroblasts', *Blood*, 107(9), pp. 3624–3631. doi: 10.1182/blood-2005-09-3682.

Haas, W., Pereira, P. and Tonegawa, S. (1993) 'Gamma/delta cells', *Annu Rev Immunol.*, 11, pp. 637–685.

Hacein-Bey-Abina, S. *et al.* (2003) 'LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1', *Science*, 302(5644), pp. 415–419. doi: 10.1126/science.1088547.

Hacein-Bey-Abina, S. *et al.* (2008) 'Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1', *Journal of Clinical Investigation*, 118(9),

pp. 3132–3142. doi: 10.1172/JCI35700.

Hacein-Bey-Abina, S. *et al.* (2010) 'Efficacy of Gene Therapy for X-Linked Severe Combined Immunodeficiency', *New England Journal of Medicine*, 363(4), pp. 355–364. doi: 10.1056/nejmoa1000164.

Hackett, J., Bennett, M. and Kumar, V. (1985) 'Origin and differentiation of natural killer cells. I. Characteristics of a transplantable NK cell precursor', *The Journal of immunology (1950)*, 134(6).

Hackney, J. A. *et al.* (2002) 'A molecular profile of a hematopoietic stem cell niche', *Proceedings of the National Academy of Sciences of the United States of America*, 99(20), pp. 13061–13066. doi: 10.1073/pnas.192124499.

Haddad, R. *et al.* (2006) 'Dynamics of thymus-colonizing cells during human development', *Immunity*, 24(2), pp. 217–230. doi: 10.1016/j.immuni.2006.01.008.

Hammarlund, E. *et al.* (2003) 'Duration of antiviral immunity after smallpox vaccination', *Nature Medicine*, 9(9), pp. 1131–1137. doi: 10.1038/nm917.

Hao, Q. L. *et al.* (2008) 'Human intrathymic lineage commitment is marked by differential CD7 expression: Identification of CD7- lympho-myeloid thymic progenitors', *Blood*, 111(3), pp. 1318–1326. doi: 10.1182/blood-2007-08-106294.

Harkey, M. A. *et al.* (2007) 'Multiarm high-throughput integration site detection: Limitations of LAM-PCR technology and optimization for clonal analysis', *Stem Cells and Development*, 16(3), pp. 381–392. doi: 10.1089/scd.2007.0015.

Harrington, L. E. *et al.* (2008) 'Memory CD4 T cells emerge from effector T-cell progenitors', *Nature*, 452(7185), pp. 356–360. doi: 10.1038/nature06672.

Hayakawa, J. *et al.* (2009) 'Long-term vector integration site analysis following retroviral mediated gene transfer to hematopoietic stem cells for the treatment of HIV infection', *PLoS ONE*, 4(1). doi: 10.1371/journal.pone.0004211.

Haynes, B. E. *et al.* (1988) 'Early events in human T cell ontogeny: Phenotypic characterization and immunohistologic localization of t cell precursors in early human fetal tissues', *Journal of Experimental Medicine*, 168(3), pp. 1061–10810. doi: 10.1084/jem.168.3.1061.

Haynes, B. F. *et al.* (1989) 'Ontogeny of T-cell precursors: a model for the initial stages of human T-cell development', *Immunology Today*, 10(3), pp. 87–91. doi: 10.1016/0167-5699(89)90232-6.

Hellman, S. and Botnick, L. E. (1977) 'Stem cell depletion: An explanation of the late effects of cytotoxins', *International Journal of Radiation Oncology, Biology, Physics*, 2(1–2), pp. 181–184. doi: 10.1016/0360-3016(77)90028-1.

Hendricks, D. W. *et al.* (2014) ' Cutting Edge: NKG2C hi CD57 + NK Cells Respond Specifically to Acute Infection with Cytomegalovirus and Not Epstein–Barr Virus ', *The Journal of Immunology*, 192(10), pp. 4492–4496. doi: 10.4049/jimmunol.1303211.

Hettinger, J. *et al.* (2013) 'Origin of monocytes and macrophages in a committed progenitor', *Nature Immunology*, 14(8), pp. 821–830. doi: 10.1038/ni.2638.

Hochweller, K. *et al.* (2010) 'Dendritic cells control T cell tonic signaling required for responsiveness to foreign antigen', *Proceedings of the National Academy of Sciences of the United States of America*, 107(13), pp. 5931–5936. doi: 10.1073/pnas.0911877107.

Hourihan, R. N., O'Sullivan, G. C. and Morgan, J. G. (2001) 'High-resolution detection of loss of heterozygosity of dinucleotide microsatellite markers', *BioTechniques*, 30(2). doi: 10.2144/01302tt03.

Howe, S. J. *et al.* (2008) 'Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients', *Journal of Clinical Investigation*, 118(9), pp. 3143–3150. doi: 10.1172/JCI35798.

Ibraheem, D., Elaissari, A. and Fessi, H. (2014) 'Gene therapy and DNA delivery systems', *International Journal of Pharmaceutics*, 459(1–2), pp. 70–83. doi: 10.1016/j.ijpharm.2013.11.041.

Ikawa, T. *et al.* (2010) 'An essential developmental checkpoint for production of the T cell lineage', *Science*, 329(5987), pp. 93–96. doi: 10.1126/science.1188995.

Imai, C. *et al.* (2004) 'Chimeric receptors with 4-1BB signaling capacity provoke potent cytotoxicity against acute lymphoblastic leukemia', *Leukemia*, 18(4), pp. 676–684. doi: 10.1038/sj.leu.2403302.

Jakubzick, C. *et al.* (2013) 'Minimal differentiation of classical monocytes as they survey steady-state tissues and transport antigen to lymph nodes', *Immunity*, 39(3), pp. 599–610. doi: 10.1016/j.immuni.2013.08.007.

Jameson, S. C. and Masopust, D. (2009) 'Diversity in T Cell Memory: An Embarrassment of Riches', *Immunity*, 31(6), pp. 859–871. doi: 10.1016/j.immuni.2009.11.007.

Jensen, M. C. *et al.* (2010) 'Antitransgene rejection responses contribute to attenuated persistence of adoptively transferred CD20/CD19-specific chimeric antigen receptor redirected T cells in humans', *Biology of Blood and Marrow Transplantation*, 16(9), pp. 1245–1256. doi: 10.1016/j.bbmt.2010.03.014.

Jones, D. D., Wilmore, J. R. and Allman, D. (2015) ' Cellular Dynamics of Memory B Cell Populations: IgM + and IgG + Memory B Cells Persist Indefinitely as Quiescent Cells ', *The Journal of Immunology*, 195(10), pp. 4753–4759. doi: 10.4049/jimmunol.1501365.

Kabelitz, D. *et al.* (2005) 'Epithelial defence by γδ T cells', *International Archives of Allergy and Immunology*, 137(1), pp. 73–81. doi: 10.1159/000085107.

Kaech, S. M., Wherry, E. J. and Ahmed, R. (2002) 'Effector and memory T-cell differentiation: Implications for vaccine development', *Nature Reviews Immunology*, 2(4), pp. 251–262. doi: 10.1038/nri778.

Kalos, M. *et al.* (2011) 'T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia', in *Science Translational Medicine*. p. 95ra73: Volume 95. doi: 10.1126/scitranslmed.3002842.

Kane, L. P., Lin, J. and Weiss, A. (2000) 'Signal transduction by the TCR for antigen', *Current Opinion in Immunology*, 12(3), pp. 242–249. doi: 10.1016/S0952-

7915(00)00083-2.

Kärre, K. (2008) 'Natural killer cell recognition of missing self', *Nature Immunology*, 9(5), pp. 477–480. doi: 10.1038/ni0508-477.

Kennedy, M. K. *et al.* (2000) 'Reversible defects in natural killer and memory CD8 T cell lineages in interleukin 15-deficient mice', in Med, J. E. (ed.) *Journal of Experimental Medicine*. 191(5, pp. 771–780. doi: 10.1084/jem.191.5.771.

Kent, W. J. (2002) 'BLAT - The BLAST-like alignment tool', *Genome Research*, 12(4). doi: 10.1101/gr.229202. Article published online before March 2002.

Kessler, C. and Manta, V. (1990) 'Specificity of restriction endonucleases and DNA modification methyltransferases - a review (edition 3)', *Gene*, 92(1–2), pp. 1–240. doi: 10.1016/0378-1119(90)90486-B.

Kiessling, R., Klein, E. and Wigzell, H. (1975) 'Natural', *killer cells in the mouse*, 5(2), pp. 112–117.

Kim, K. H. *et al.* (2019) 'Phenotypic and Functional Analysis of Human NK Cell Subpopulations According to the Expression of FcεRIγ and NKG2C', *Frontiers in Immunology*, 10. doi: 10.3389/fimmu.2019.02865.

Kimmig, S. *et al.* (2002) 'Two subsets of naive T helper cells with distinct T cell receptor excision circle content in human adult peripheral blood', *Journal of Experimental Medicine*, 195(6), pp. 789–794. doi: 10.1084/jem.20011756.

Kircheis, R. *et al.* (2001) 'Polyethylenimine/DNA complexes shielded by transferrin

target gene expression to tumors after systemic application', *Gene Therapy*, 8(1). doi: 10.1038/sj.gt.3301351.

Knierim, E. *et al.* (2011) 'Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing', *PLoS ONE*, 6(11). doi: 10.1371/journal.pone.0028240.

Kohler, S. *et al.* (2005) 'Post-thymic in vivo proliferation of naive CD4+ T cells constrains the TCR repertoire in healthy human adults', *European Journal of Immunology*, 35(6), pp. 1987–1994. doi: 10.1002/eji.200526181.

Kohn, L. A. *et al.* (2012) 'Lymphoid priming in human bone marrow begins before expression of CD10 with upregulation of L-selectin', *Nature Immunology*, 13(10), pp. 963–971. doi: 10.1038/ni.2405.

Kondo, M., Weissman, I. L. and Akashi, K. (1997) 'Identification of clonogenic common lymphoid progenitors in mouse bone marrow', *Cell*, 91(5), pp. 661–672. doi: 10.1016/S0092-8674(00)80453-5.

Kozulić, B. (1994) 'On the "door-corridor" model of gel electrophoresis. III. The gel constant and resistance, and the net charge, friction, diffusion and electrokinetic force of the migrating molecules.', *Applied and theoretical electrophoresis : the official journal of the International Electrophoresis Society*, 4(3).

Kozulić, B. and Heimgartner, U. (1991) 'An apparatus for submerged gel electrophoresis', *Analytical Biochemistry*, 198(2). doi: 10.1016/0003-2697(91)90422-P.

Kraunus, J. *et al.* (2004) 'Self-inactivating retroviral vectors with improved RNA processing', *Gene Therapy*, 11(21), pp. 1568–1578. doi: 10.1038/sj.gt.3302309.

Krause, A. *et al.* (1998) 'Antigen-dependent CD28 signaling selectively enhances survival and proliferation in genetically modified activated human primary T lymphocytes', in Med, J. E. (ed.) *Journal of Experimental Medicine*. Volume 188, pp. 619–626. doi: 10.1084/jem.188.4.619.

Krijgsman, D., Hokland, M. and Kuppen, P. J. K. (2018) 'The role of natural killer T cells in cancer-A phenotypical and functional approach', *Frontiers in Immunology*, 9(FEB). doi: 10.3389/fimmu.2018.00367.

Kruger, D. H. and Bickle, T. A. (1983) 'Bacteriophage survival: Multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts', *Microbiological Reviews*. doi: 10.1128/mmbr.47.3.345-360.1983.

Kumar, V. and Delovitch, T. L. (2014) 'Different subsets of natural killer T cells may vary in their roles in health and disease', *Immunology*, 142(3), pp. 321–336. doi: 10.1111/imm.12247.

Kurosawa, Y. *et al.* (1981) 'Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes', *Nature*, 290(5807), pp. 565–570. doi: 10.1038/290565a0.

Kursa, M. *et al.* (2003) 'Novel shielded transferrin-polyethylene glycol-polyethylenimine/DNA complexes for systemic tumor-targeted gene transfer', *Bioconjugate Chemistry*, 14(1). doi: 10.1021/bc0256087.

Lage, J. M. *et al.* (2003) 'Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH', *Genome Research*, 13(2), pp. 294–307. doi: 10.1101/gr.377203.

Lai, Y. *et al.* (2018) 'Toll-like receptor 2 costimulation potentiates the antitumor efficacy of CAR T Cells', *Leukemia*, 32(3), pp. 801–808. doi: 10.1038/leu.2017.249.

Lander, E. S. *et al.* (2001) 'Initial sequencing and analysis of the human genome', *Nature*, 409(6822). doi: 10.1038/35057062.

Lansdorp, P. M., Sutherland, H. J. and Eaves, C. J. (1990) 'Selective expression of CD45 isoforms on functional subpopulations of CD34+ hemopoietic cells from human bone marrow', *Journal of Experimental Medicine*, 172(1), pp. 363–366. doi: 10.1084/jem.172.1.363.

Lanzavecchia, A. and Sallusto, F. (2002) 'Progressive differentiation and selection of the fittest in the immune response', *Nature Reviews Immunology*, 2(12), pp. 982–987. doi: 10.1038/nri959.

Lavaert, M. *et al.* (2020) 'Integrated scRNA-Seq Identifies Human Postnatal Thymus Seeding Progenitors and Regulatory Dynamics of Differentiating Immature Thymocytes', *Immunity*, 52(6), pp. 1088-1104.e6. doi: 10.1016/j.immuni.2020.03.019.

Lee, D. W. *et al.* (2015) 'T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: A phase 1 dose-escalation trial', *The Lancet*, 385(9967), pp. 517–528. doi: 10.1016/S0140-6736(14)61403-3.

Lee, P. Y. *et al.* (2012) 'Agarose gel electrophoresis for the separation of DNA fragments', *Journal of Visualized Experiments*, (62). doi: 10.3791/3923.

Legut, M., Cole, D. K. and Sewell, A. K. (2015) 'The promise of γδT cells and the γδT cell receptor for cancer immunotherapy', *Cellular and Molecular Immunology*, 12(6), pp. 656–658. doi: 10.1038/cmi.2015.28.

Leonardelli, L. *et al.* (2016) '531. Computational Pipeline for the Identification of Integration Sites and Novel Method for the Quantification of Clone Sizes in Clonal Tracking Studies', *Molecular Therapy*, 24(May), pp. S212–S213. doi: 10.1016/s1525-0016(16)33340-8.

Lewinski, M. K. *et al.* (2006) 'Retroviral DNA integration: Viral and cellular determinants of target-site selection', *PLoS Pathogens*, 2(6). doi: 10.1371/journal.ppat.0020060.

Li, P. *et al.* (2010) 'Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion', *Science*, 329(5987), pp. 85–89. doi: 10.1126/science.1188063.

Liu, X. *et al.* (2015) 'Affinity-tuned ErbB2 or EGFR chimeric antigen receptor T cells exhibit an increased therapeutic index against tumors in mice', *Cancer Research*, 75(17), pp. 3596–3607. doi: 10.1158/0008-5472.CAN-15-0159.

Lodolce, J. P. *et al.* (1998) 'IL-15 receptor maintains lymphoid homeostasis by supporting lymphocyte homing and proliferation', *Immunity*, 9(5), pp. 669–676. doi: 10.1016/S1074-7613(00)80664-0.

Loenen, W. A. M. *et al.* (2014) 'Highlights of the DNA cutters: A short history of the

restriction enzymes', *Nucleic Acids Research*, 42(1). doi: 10.1093/nar/gkt990.

Long, A. (2015) '4-1BB co-stimulation ameliorates T cell exhaustion induced by antigen independent signaling of chimeric antigen receptors', *Rejuvenation Research*, 18(6), p. 585. doi: 10.1089/rej.2015.1792.

Lopez-Vergès, S. *et al.* (2010) 'CD57 defines a functionally distinct population of mature NK cells in the human CD56dimCD16+ NK-cell subset', *Blood*, 116(19), pp. 3865–3874. doi: 10.1182/blood-2010-04-282301.

Lopez-Vergès, S. al. (2011) 'Expansion of a unique CD57⁺NKG2Chi natural killer cell subset during acute human cytomegalovirus infection', *PNAS*, 108(36), pp. 14725–14732.

Love, P. E. and Hayes, S. M. (2010) 'ITAM-mediated signaling by the T-cell antigen receptor.', *Cold Spring Harbor perspectives in biology*, 2(6). doi: 10.1101/cshperspect.a002485.

Lugli, E. *et al.* (2013) 'Superior T memory stem cell persistence supports long-lived T cell memory', *Journal of Clinical Investigation*, 123(2), pp. 594–599. doi: 10.1172/JCI66327.

Lynch, H. E. and Sempowski, G. D. (2013) 'Molecular measurement of t cell receptor excision circles', *Methods in Molecular Biology*, 979, pp. 147–159. doi: 10.1007/978-1-62703-290-2_12.

Macallan, D. C. *et al.* (2003) 'Measurement and modeling of human T cell kinetics',

European Journal of Immunology, 33(8), pp. 2316–2326. doi: 10.1002/eji.200323763.

Maetzig, T. *et al.* (2011) 'Gammaretroviral vectors: Biology, technology and application', *Viruses*. doi: 10.3390/v3060677.

Maher, J. *et al.* (2002) 'Human T-lymphocyte cytotoxicity and proliferation directed by a single chimeric TCRζ/CD28 receptor', *Nature Biotechnology*, 20(1), pp. 70–75. doi: 10.1038/nbt0102-70.

Mahnke, Y. D. *et al.* (2013) 'The who's who of T-cell differentiation: Human memory T-cell subsets', *European Journal of Immunology*, 43(11), pp. 2797–2809. doi: 10.1002/eji.201343751.

Månsson, R. *et al.* (2007) 'Molecular Evidence for Hierarchical Transcriptional Lineage Priming in Fetal and Adult Stem Cells and Multipotent Progenitors', *Immunity*, 26(4), pp. 407–419. doi: 10.1016/j.immuni.2007.02.013.

Manz, M. G. *et al.* (2002) 'Prospective isolation of human clonogenic common myeloid progenitors', *Proceedings of the National Academy of Sciences of the United States of America*, 99(18), pp. 11872–11877. doi: 10.1073/pnas.172384399.

Marshall, H. D. *et al.* (2011) 'Differential Expression of Ly6C and T-bet Distinguish Effector and Memory Th1 CD4 + Cell Properties during Viral Infection', *Immunity*, 35(4), pp. 633–646. doi: 10.1016/j.immuni.2011.08.016.

Mauch, P. *et al.* (1995) 'Hematopoietic stem cell compartment: Acute and late effects of radiation therapy and chemotherapy', *International Journal of Radiation Oncology,*

Biology, Physics, 31(5), pp. 1319–1339. doi: 10.1016/0360-3016(94)00430-S.

Maude, S. L. *et al.* (2014) 'Chimeric Antigen Receptor T Cells for Sustained Remissions in Leukemia', in Med, N. E. J. (ed.) *New England Journal of Medicine*. Volume 371, pp. 1507–1517. doi: 10.1056/nejmoa1407222.

Maude, S. L. *et al.* (2018) 'Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia', in Med, N. E. J. (ed.) *New England Journal of Medicine*. Volume 378, pp. 439–448. doi: 10.1056/nejmoa1709866.

Maus, M. V. *et al.* (2013) 'T cells expressing chimeric antigen receptors can cause anaphylaxis in humans', *Cancer immunology research*, 1(1), pp. 26–31. doi: 10.1158/2326-6066.CIR-13-0006.

McCloskey, M. L. *et al.* (2007) 'Encoding PCR products with batch-stamps and barcodes', *Biochemical Genetics*, 45(11–12). doi: 10.1007/s10528-007-9114-x.

Medzhitov, R. and Janeway, C. A. (2002) 'Decoding the patterns of self and nonself by the innate immune system', *Science*, 296(5566), pp. 298–300. doi: 10.1126/science.1068883.

Meuer, S. C. *et al.* (1983) 'Evidence for the T3-associated 90K heterodimer as the T-cell antigen receptor', *Nature*, 303(5920), pp. 808–810. doi: 10.1038/303808a0.

Miller, J. D. *et al.* (2008) 'Human Effector and Memory CD8+ T Cell Responses to Smallpox and Yellow Fever Vaccines', *Immunity*, 28(5), pp. 710–722. doi: 10.1016/j.immuni.2008.02.020.

Milone, M. C. *et al.* (2009) 'Chimeric receptors containing CD137 signal transduction domains mediate enhanced survival of T cells and increased antileukemic efficacy in vivo', *Molecular Therapy*, 17(8), pp. 1453–1464. doi: 10.1038/mt.2009.83.

Milone, M. C. and O'Doherty, U. (2018) 'Clinical use of lentiviral vectors', *Leukemia*. doi: 10.1038/s41375-018-0106-0.

Mitchell, R. S. *et al.* (2004) 'Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences', *PLoS Biology*, 2(8), p. 8. doi: 10.1371/journal.pbio.0020234.

Mittrücker, H. W., Visekruna, A. and Huber, M. (2014) 'Heterogeneity in the Differentiation and Function of CD8+ T Cells', *Archivum Immunologiae et Therapiae Experimentalis*, 62(6), pp. 449–458. doi: 10.1007/s00005-014-0293-y.

Montanera, K. N. *et al.* (2020) 'ChIP-exo: A method to study chromatin structure and organization at near-nucleotide resolution', *Epigenetics Methods*, pp. 323–352. doi: 10.1016/B978-0-12-819414-0.00016-1.

Moretta, A. (2005) 'The dialogue between human natural killer cells and dendritic cells', *Current Opinion in Immunology*. doi: 10.1016/j.coi.2005.03.004.

Morgan, R. A. *et al.* (2010) 'Case report of a serious adverse event following the administration of t cells transduced with a chimeric antigen receptor recognizing ERBB2', *Molecular Therapy*, 18(4), pp. 843–851. doi: 10.1038/mt.2010.24.

Mori, L. and De Libero, G. (2012) 'T cells specific for lipid antigens', *Immunologic*

Research, 53(1–3), pp. 191–199. doi: 10.1007/s12026-012-8294-6.

Moulton, V. R. *et al.* (2006) 'Divergent Generation of Heterogeneous Memory CD4 T Cells', *The Journal of Immunology*, 177(2), pp. 869–876. doi: 10.4049/jimmunol.177.2.869.

Mueller, P. R. and Wold, B. (1989) 'In vivo footprinting of a muscle specific enhancer by ligation mediated PCR', *Science*, 246(4931). doi: 10.1126/science.2814500.

Murray, L. *et al.* (1995) 'Enrichment of human hematopoietic stem cell activity in the CD34+Thy- 1+Lin subpopulation from mobilized peripheral blood', *Blood*, 85(2), pp. 368–378. doi: 10.1182/blood.v85.2.368.bloodjournal852368.

Murugan, A. *et al.* (2012) 'Statistical inference of the generation probability of T-cell receptors from sequence repertoires', *Proceedings of the National Academy of Sciences of the United States of America*, 109(40), pp. 16161–16166. doi: 10.1073/pnas.1212755109.

Nagler, A. *et al.* (1989) 'Comparative studies of human FcRIII-positive and negative natural killer cells.', *Journal of immunology (Baltimore, Md. : 1950)*, 143(10), pp. 3183–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2530273.

Naik, S. H. *et al.* (2013) 'Diverse and heritable lineage imprinting of early haematopoietic progenitors', *Nature*, 496(7444), pp. 229–232. doi: 10.1038/nature12013.

Naso, M. F. *et al.* (2017) 'Adeno-Associated Virus (AAV) as a Vector for Gene

Therapy', *BioDrugs*. doi: 10.1007/s40259-017-0234-5.

Nayerossadat, N., Ali, P. and Maedeh, T. (2012) 'Viral and nonviral delivery systems for gene delivery', *Advanced Biomedical Research*, 1(1). doi: 10.4103/2277-9175.98152.

Neben, S. *et al.* (1993) 'Hematopoietic stem cell deficit of transplanted bone marrow previously exposed to cytotoxic agents', *Experimental Hematology*, 21(1), pp. 156–162.

Notta, F. *et al.* (2011) 'Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment', *Science*, 333(6039), pp. 218–221. doi: 10.1126/science.1201219.

Notta, F. *et al.* (2016) 'Distinct routes of lineage development reshape the human blood hierarchy across ontogeny', *Science*, 351(6269), p. 6269. doi: 10.1126/science.aab2116.

Nowrouzi, A. *et al.* (2011) 'Retroviral vectors: Post entry events and genomic alterations', *Viruses*, 3(5), pp. 429–455. doi: 10.3390/v3050429.

Oliveira, G. *et al.* (2015) *Tracking genetically engineered lymphocytes long-term reveals the dynamics of t cell immunological memory, Science Translational Medicine*. Science Translational Medicine. doi: 10.1126/scitranslmed.aac8265.

El Omari, K. *et al.* (2011) 'Structure of the leukemia oncogene LMO2: Implications for the assembly of a hematopoietic transcription factor complex', *Blood*, 117(7), pp.

2146–2156. doi: 10.1182/blood-2010-07-293357.

Orange, J. S. (2002) 'Human natural killer cell deficiencies and susceptibility to infection', *Microbes and Infection*. doi: 10.1016/S1286-4579(02)00038-2.

Ouyang, Q. *et al.* (2007) 'Telomere length in human natural killer cell subsets', *Annals of the New York Academy of Sciences*, 1106, pp. 240–252. doi: 10.1196/annals.1392.001.

Paez, J. G. *et al.* (2004) 'Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification.', *Nucleic acids research*, 32(9), pp. 1–11. doi: 10.1093/nar/gnh069.

Paruzynski, A. *et al.* (2010) 'Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing', *Nature Protocols*, 5(8). doi: 10.1038/nprot.2010.87.

Passoni, L. *et al.* (1997) 'Intrathymic δ selection events in γδ cell development', *Immunity*, 7(1), pp. 83–95. doi: 10.1016/S1074-7613(00)80512-9.

St. Paul, M. and Ohashi, P. S. (2020) 'The Roles of CD8+ T Cell Subsets in Antitumor Immunity', *Trends in Cell Biology*, 30(9), pp. 695–704. doi: 10.1016/j.tcb.2020.06.003.

Pepper, M. and Jenkins, M. K. (2011) 'Origins of CD4+ effector and central memory T cells', *Nature Immunology*, 12(6), pp. 467–471. doi: 10.1038/ni.2038.

Perry, G. A., Jackson, J. D. and Talmadge, J. E. (1994) 'Effects of a multidrug chemotherapy regimen on the thymus', *Thymus*, 23(1), pp. 39–51.

Phillips, J. H. *et al.* (1992) 'Ontogeny of human natural killer (NK) cells: fetal NK cells mediate cytolytic function and express cytoplasmic CD3 epsilon,delta proteins.', *Journal of Experimental Medicine*, 175(4), pp. 1055–1066. doi: 10.1084/jem.175.4.1055.

Pietras, E. M. *et al.* (2015) 'Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions', *Cell Stem Cell*, 17(1), pp. 35–46. doi: 10.1016/j.stem.2015.05.003.

Pillay, J. *et al.* (2010) 'In vivo labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days', *Blood*, 116(4), pp. 625–627. doi: 10.1182/blood-2010-01-259028.

Poli, A. *et al.* (2009) 'CD56bright natural killer (NK) cells: An important NK cell subset', *Immunology*, 126(4), pp. 458–465. doi: 10.1111/j.1365-2567.2008.03027.x.

Porter, D. L. *et al.* (2011) 'Chimeric Antigen Receptor–Modified T Cells in Chronic Lymphoid Leukemia', *New England Journal of Medicine*, 365(8). doi: 10.1056/nejmoa1103849.

Pronk, C. J. H. *et al.* (2007) 'Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy', *Cell Stem Cell*, 1(4), pp. 428–442. doi: 10.1016/j.stem.2007.07.005.

Pross, H. F. and Baines, M. G. (1976) 'Spontaneous human lymphocyte-mediated cytotoxicity againts tumour target cells. I. The effect of malignant disease', *International Journal of Cancer*, 18(5), pp. 593–604. doi: 10.1002/ijc.2910180508.

Punwani, D. *et al.* (2016) ' Multisystem Anomalies in Severe Combined Immunodeficiency with Mutant BCL11B ', *New England Journal of Medicine*, 375(22), pp. 2165–2176. doi: 10.1056/nejmoa1509164.

Pupuleku, A. *et al.* (2017) 'Elusive role of the CD94/NKG2C NK cell receptor in the response to cytomegalovirus: Novel experimental observations in a reporter cell system', *Frontiers in Immunology*, 8(OCT). doi: 10.3389/fimmu.2017.01317.

Qi, Q. *et al.* (2014) 'Diversity and clonal selection in the human T-cell repertoire', *Proceedings of the National Academy of Sciences of the United States of America*, 111(36), pp. 13139–13144. doi: 10.1073/pnas.1409155111.

Raphael, I. *et al.* (2015) 'T cell subsets and their signature cytokines in autoimmune and inflammatory diseases', *Cytokine*, 74(1), pp. 5–17. doi: 10.1016/j.cyto.2014.09.011.

Ratajczak, M. Z. (2008) 'Phenotypic and functional characterization of hematopoietic stem cells', *Current Opinion in Hematology*, 15(4), pp. 293–300. doi: 10.1097/MOH.0b013e328302c7ca.

Reantragoon, R. *et al.* (2012) 'Structural insight into MR1-mediated recognition of the mucosal associated invariant T cell receptor', in Med, J. E. (ed.) *Journal of Experimental Medicine*. 209(4, pp. 761–774. doi: 10.1084/jem.20112095.

Redondo-Pachón, D. *et al.* (2017) ' Adaptive NKG2C + NK Cell Response and the Risk of Cytomegalovirus Infection in Kidney Transplant Recipients ', *The Journal of Immunology*, 198(1), pp. 94–101. doi: 10.4049/jimmunol.1601236.

Renoux, V. M. *et al.* (2015) 'Identification of a Human Natural Killer Cell Lineage-Restricted Progenitor in Fetal and Adult Tissues', *Immunity*, 43(2), pp. 394–407. doi: 10.1016/j.immuni.2015.07.011.

Reya, T. (2003) 'Regulation of Hematopoietic Stem Cell Self-Renewal', *Recent Progress in Hormone Research*. doi: 10.1210/rp.58.1.283.

Roberts, R. J. and Murray, K. (1976) 'Restriction endonuclease', *Critical Reviews in Biochemistry and Molecular Biology*, 4(2). doi: 10.3109/10409237609105456.

Robins, H. S. *et al.* (2009) 'Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells', *Blood*, 114(19), pp. 4099–4107. doi: 10.1182/blood-2009-04-217604.

Robins, H. S. *et al.* (2010) 'Overlap and effective size of the human CD8+ T cell receptor repertoire', in *Science Translational Medicine*. p. 47ra64: 2(47). doi: 10.1126/scitranslmed.3001442.

Rodewald, H. R. *et al.* (1992) 'A population of early fetal thymocytes expressing FcγRII III contains precursors of T lymphocytes and natural killer cells', *Cell*, 69(1), pp. 139–150. doi: 10.1016/0092-8674(92)90125-V.

Romagnani, C. J. K. F. M. M. B. D. A. al. (2007) 'CD56brightCD16- killer Ig-like receptor- NK cells display longer telomeres and acquire features of CD56dim NK cells upon activation', *Journal of Immunology*, 178, pp. 4947–4955.

Rosmaraki, E. E. *et al.* (2001) 'Identification of committed NK cell progenitors in adult

murine bone marrow', *European Journal of Immunology*, 31(6), pp. 1900–1909. doi: 10.1002/1521-4141(200106)31:6<1900::AID-IMMU1900>3.0.CO;2-M.

Rossjohn, J. *et al.* (2015) 'T cell antigen receptor recognition of antigen-presenting molecules', *Annual Review of Immunology*, 33, pp. 169–200. doi: 10.1146/annurev-immunol-032414-112334.

Rothenberg, E. V., Moore, J. E. and Yui, M. A. (2008) 'Launching the T-cell-lineage developmental programme', *Nature Reviews Immunology*, 8(1), pp. 9–21. doi: 10.1038/nri2232.

Sallusto, F. *et al.* (1999) 'Two subsets of memory T lymphocytes with distinct homing potentials and effector functions', *Nature*, 401(6754), pp. 708–712. doi: 10.1038/44385.

Sallusto, F., Geginat, J. and Lanzavecchia, A. (2004) 'Central memory and effector memory T cell subsets: Function, generation, and maintenance', *Annual Review of Immunology*, 22, pp. 745–763. doi: 10.1146/annurev.immunol.22.012703.104702.

Sánchez, M. J. *et al.* (1993) 'Human Natural Killer Cell Committed Thymocytes and Their Relation to the T Cell Lineage', *Journal of Experimental Medicine*, 178(6), pp. 1857–1866. doi: 10.1084/jem.178.6.1857.

Sánchez, M. J. *et al.* (1994) 'Identification of a common t/natural killer cell progenitor in human fetal thymus', *Journal of Experimental Medicine*, 180(2), pp. 569–576. doi: 10.1084/jem.180.2.569.

Savoldo, B. *et al.* (2011) 'CD28 costimulation improves expansion and persistence of chimeric antigen receptor-modified T cells in lymphoma patients', *Journal of Clinical Investigation*, 121(5), pp. 1822–1826. doi: 10.1172/JCI46110.

Scala, S. *et al.* (2018) 'Dynamics of genetically engineered hematopoietic stem and progenitor cells after autologous transplantation in humans', *Nature Medicine*, 24(11), pp. 1683–1690. doi: 10.1038/s41591-018-0195-3.

Schlums, H. *et al.* (2017) 'Adaptive NK cells can persist in patients with GATA2 mutation depleted of stem and progenitor cells', *Blood*, 129(14), pp. 1927–1939. doi: 10.1182/blood-2016-08-734236.

Schmidt, M. *et al.* (2001) 'Detection and direct genomic sequencing of multiple rare unknown flanking DNA in highly complex samples', *Human Gene Therapy*, 12(7). doi: 10.1089/104303401750148649.

Schmidt, M. *et al.* (2007) 'High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR)', *Nature Methods*, 4(12). doi: 10.1038/nmeth1103.

Schmitt, C. *et al.* (1993) 'CD34-expressing human thymocyte precursors proliferate in response to interleukin-7 but have lost myeloid differentiation potential', *Blood*, 82(12), pp. 3675–3685. doi: 10.1182/blood.v82.12.3675.3675.

Schmueck-Henneresse, M. *et al.* (2017) *Comprehensive Approach for Identifying the T Cell Subset Origin of CD3 and CD28 Antibody–Activated Chimeric Antigen Receptor–Modified T Cells*, *The Journal of Immunology*. p: The Journal of Immunology. doi: 10.4049/jimmunol.1601494.

Schubert, M.-L. *et al.* (2019) 'Third-Generation CAR T Cells Targeting CD19 Are Associated with an Excellent Safety Profile and Might Improve Persistence of CAR T Cells in Treated Patients', *Blood*, 134(Supplement_1), pp. 51–51. doi: 10.1182/blood-2019-125423.

Schuster, S. J. *et al.* (2017) 'Chimeric Antigen Receptor T Cells in Refractory B-Cell Lymphomas', in Med, N. E. J. (ed.) *New England Journal of Medicine*. Volume 377, pp. 2545–2554. doi: 10.1056/nejmoa1708566.

Schwarzwaelder, K. *et al.* (2007) 'Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo', *Journal of Clinical Investigation*, 117(8). doi: 10.1172/JCI31661.

Shah, D. K. and Zúñiga-Pflücker, J. C. (2014) 'An Overview of the Intrathymic Intricacies of T Cell Development', *The Journal of Immunology*, 192(9). doi: 10.4049/jimmunol.1302259.

Sharpe, A. and Freeman, G. (2002) 'The B7-CD28 superfamily', *Nat Rev Immunol.*, 2, pp. 116–126.

Sherman, E. *et al.* (2017) 'INSPIIRED: A Pipeline for Quantitative Analysis of Sites of New DNA Integration in Cellular Genomes', *Molecular Therapy - Methods and Clinical Development*, 4. doi: 10.1016/j.omtm.2016.11.002.

Silver, J. and Keerikatte, V. (1989) 'Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus', *Journal of Virology*, 63(5). doi: 10.1128/jvi.63.5.1924-1928.1989.

Six, E. M. *et al.* (2007) 'A human postnatal lymphoid progenitor capable of circulating and seeding the thymus', *Journal of Experimental Medicine*, 204(13), pp. 3085–3093. doi: 10.1084/jem.20071003.

Slatko, B. E., Gardner, A. F. and Ausubel, F. M. (2018) 'Overview of Next-Generation Sequencing Technologies', *Current Protocols in Molecular Biology*, 122(1). doi: 10.1002/cpmb.59.

Sommermeyer, D. *et al.* (2017) 'Fully human CD19-specific chimeric antigen receptors for T-cell therapy', *Leukemia*, 31(10), pp. 2191–2199. doi: 10.1038/leu.2017.57.

Soudja, S. M. H. *et al.* (2012) 'Inflammatory Monocytes Activate Memory CD8+ T and Innate NK Lymphocytes Independent of Cognate Antigen during Microbial Pathogen Invasion', *Immunity*, 37(3), pp. 549–562. doi: 10.1016/j.immuni.2012.05.029.

Spangrude, G. J. *et al.* (1991) 'Mouse hematopoietic stem cells', *Blood*, 78(6), pp. 1395–1402. doi: 10.1182/blood.v78.6.1395.1395.

Spits, H. (2002) 'Development of αβ T cells in the human thymus', *Nature Reviews Immunology*, 2(10), pp. 760–772. doi: 10.1038/nri913.

Sprent, J. and Basten, A. (1973) 'Circulating T and B lymphocytes of the mouse. II. Lifespan', *Cellular Immunology*, 7(1), pp. 40–59. doi: 10.1016/0008-8749(73)90181-0.

Stabile, H. *et al.* (2015) 'Response to comment on multifunctional human CD56lowCD16low NK cells are the prominent subset in bone marrow of both pediatric healthy donors and leukemic patients', *Haematologica*, 100(8), pp. 332–333. doi:

10.3324/haematol.2015.130831.

Sterner, R. C. and Sterner, R. M. (2021) 'CAR-T cell therapy: current limitations and potential strategies', *Blood Cancer Journal*, 11(4). doi: 10.1038/s41408-021-00459-7.

Sun, J. C., Beilke, J. N. and Lanier, L. L. (2009) 'Adaptive immune features of natural killer cells (Nature (2009) 457, (557-561))', *Nature*, 457(7233), p. 1168. doi: 10.1038/nature07833.

Sutherland, D. R., Rudd, C. E. and Greaves, M. F. (1984) 'Isolation and characterization of a human T lymphocyte-associated glycoprotein (gp40)', *The Journal of Immunology*, 133(1), pp. 327–333.

Swain, S. L. *et al.* (2006) 'CD4+ T-cell memory: Generation and multi-faceted roles for CD4+ T cells in protective immunity to influenza', *Immunological Reviews*, 211, pp. 8–22. doi: 10.1111/j.0105-2896.2006.00388.x.

Testa, N. G., Hendry, J. H. and Molineux, G. (1985) 'Long-term bone marrow damage in experimental systems and in patients after radiation or chemotherapy', *Anticancer Research*, 5(1), pp. 101–110.

Thome, J. J. C. *et al.* (2014) 'Spatial map of human t cell compartmentalization and maintenance over decades of life', *Cell*, 159(4), pp. 814–828. doi: 10.1016/j.cell.2014.10.026.

Till, J. E. and McCulloch, E. A. (1980) 'Hemopoietic stem cell differentiation', *Biochimica et Biophysica Acta - Reviews on Cancer*, 605(4), pp. 431–459. doi:

10.1016/0304-419X(80)90009-8.

Tindall, K. R. and Kunkel, T. A. (1988) 'Fidelity of DNA Synthesis by the Thermus aquaticus DNA Polymerase', *Biochemistry*, 27(16). doi: 10.1021/bi00416a027.

Tomita, Y., Sachs, D. H. and Sykes, M. (1994) 'Myelosuppressive conditioning is required to achieve engraftment of pluripotent stem cells contained in moderate doses of syngeneic bone marrow', *Blood*, 83(4), pp. 939–948. doi: 10.1182/blood.v83.4.939.939.

Topham, D. J. and Doherty, P. C. (1998) 'Longitudinal analysis of the acute sendai virus-specific CD4+ T cell response and memory', *Journal of immunology (Baltimore, Md. : 1950)*, 161(9), pp. 4530–455.

Touzot, F. *et al.* (2015) 'Faster T-cell development following gene therapy compared with haploidentical HSCT in the treatment of SCID-X1', *Blood*, 125(23), pp. 3563–3569. doi: 10.1182/blood-2014-12-616003.

Turtle, C. J. *et al.* (2016) 'CD19 CAR-T cells of defined CD4+:CD8+ composition in adult B cell ALL patients', *Journal of Clinical Investigation*, 126(6), pp. 2123–2138. doi: 10.1172/JCI85309.

Valbon, S. F., Condotta, S. A. and Richer, M. J. (2016) 'Regulation of effector and memory CD8+ T cell function by inflammatory cytokines', *Cytokine*, 82. doi: 10.1016/j.cyto.2015.11.013.

Veiga-Fernandes, H. *et al.* (2000) 'Response of naïve and memory CD8+ T cells to

antigen stimulation in vivo', *Nature Immunology*, 1(1), pp. 47–53. doi: 10.1038/76907.

Viardot, A. *et al.* (2019) 'Chimeric antigen receptor (CAR) T-cell therapy as a treatment option for patients with B-cell lymphomas: Perspectives on the therapeutic potential of Axicabtagene ciloleucel', *Cancer Management and Research*. doi: 10.2147/CMAR.S163225.

Di Vito, C., Mikulak, J. and Mavilio, D. (2019) 'On the way to become a natural killer cell', *Frontiers in Immunology*. doi: 10.3389/fimmu.2019.01812.

Vivier, E. *et al.* (2011) 'Innate or adaptive immunity? The example of natural killer cells', *Science*. doi: 10.1126/science.1198687.

Vivier, E. *et al.* (2018) 'Innate Lymphoid Cells: 10 Years On', *Cell*, 174(5), pp. 1054–1066. doi: 10.1016/j.cell.2018.07.017.

Vrisekoop, N. *et al.* (2008) 'Sparse production but preferential incorporation of recently produced naïve T cells in the human peripheral pool', *Proceedings of the National Academy of Sciences of the United States of America*, 105(16), pp. 6115–6120. doi: 10.1073/pnas.0709713105.

Wakabayashi, Y. *et al.* (2003) 'Bcl11b is required for differentiation and survival of αβ T lymphocytes', *Nature Immunology*, 4(6), pp. 533–539. doi: 10.1038/ni927.

Wang, J. C. Y., Doedens, M. and Dick, J. E. (1997) 'Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay',

Blood, 89(11), pp. 3919–3924. doi: 10.1182/blood.v89.11.3919.

Wang, W. *et al.* (2016) 'The LAM-PCR method to sequence LV integration sites', in *Methods in Molecular Biology*. doi: 10.1007/978-1-4939-3753-0_9.

Wang, Y. *et al.* (2009) 'A conserved CXXC motif in CD3ε is critical for T cell development and TCR signaling', *PLoS Biology*, 7(12), p. 12. doi: 10.1371/journal.pbio.1000253.

Warren, R. L. *et al.* (2011) 'Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes', *Genome Research*, 21(5), pp. 790–797. doi: 10.1101/gr.115428.110.

Weber, P. C. *et al.* (1989) 'Structural origins of high-affinity biotin binding to streptavidin', *Science*, 243(4887). doi: 10.1126/science.2911722.

Weerkamp, F. *et al.* (2006) 'Human thymus contains multipotent progenitors with T/B lymphoid, myeloid, and erythroid lineage potential', *Blood*, 107(8), pp. 3131–3137. doi: 10.1182/blood-2005-08-3412.

Willcox, C. R. *et al.* (2012) 'Cytomegalovirus and tumor stress surveillance by binding of a human γδ T cell antigen receptor to endothelial protein C receptor', *Nature Immunology*, 13(9), pp. 872–879. doi: 10.1038/ni.2394.

Wilson, A. *et al.* (2009) 'Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair (DOI:10.1016/j.cell.2008.10.048)',

Cell, 138(1), p. 209. doi: 10.1016/j.cell.2009.06.020.

Wold, W. and Toth, K. (2014) 'Adenovirus Vectors for Gene Therapy, Vaccination and Cancer Gene Therapy', *Current Gene Therapy*, 13(6). doi: 10.2174/1566523213666131125095046.

Wong, W. K., Leem, J. and Leem, J. (2019) 'Comparative Analysis of the CDR Loops of Antigen Receptors', *Frontiers in Immunology*, 10. doi: 10.3389/fimmu.2019.02454.

Wu, C. *et al.* (2014) 'Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells', *Cell Stem Cell*, 14(4), pp. 486–499. doi: 10.1016/j.stem.2014.01.020.

Wu, X. *et al.* (2003) 'Transcription start regions in the human genome are favored targets for MLV integration', *Science*, 300(5626), pp. 1749–1751. doi: 10.1126/science.1083413.

Wucherpfennig, K. W. *et al.* (2010) 'Structural biology of the T-cell receptor: insights into receptor assembly, ligand recognition, and initiation of signaling.', *Cold Spring Harbor perspectives in biology*, 2(4). doi: 10.1101/cshperspect.a005140.

Yamada, Y. *et al.* (1998) 'The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis', *Proceedings of the National Academy of Sciences of the United States of America*, 95(7), pp. 3890–3895. doi: 10.1073/pnas.95.7.3890.

Yamamoto, R. *et al.* (2013) 'XClonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells', *Cell*, 154(5), pp. 1112–

1126. doi: 10.1016/j.cell.2013.08.007.

Yamamoto, R., Wilkinson, A. C. and Nakauchi, H. (2020) 'In vivo and ex vivo haematopoietic stem cell expansion', *Current Opinion in Hematology*, 27(4), pp. 273–278. doi: 10.1097/MOH.0000000000000593.

Yang, L. *et al.* (2005) 'Identification of Lin-Sca1+kit+CD34 +Flt3- short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients', *Blood*, 105(7), pp. 2717–2723. doi: 10.1182/blood-2004-06-2159.

Yona, S. *et al.* (2013) 'Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis [Immunity 38 (2013) 79-91]', *Immunity*, 38(5), pp. 1073–1079. doi: 10.1016/j.immuni.2013.05.008.

You, F. *et al.* (2016) 'Phase 1 clinical trial demonstrated that MUC1 positive metastatic seminal vesicle cancer can be effectively eradicated by modified Anti-MUC1 chimeric antigen receptor transduced T cells', *Science China Life Sciences*, 59(4), pp. 386–397. doi: 10.1007/s11427-016-5024-7.

Zhang, L. *et al.* (2003) 'Intratumoral T Cells, Recurrence, and Survival in Epithelial Ovarian Cancer', *New England Journal of Medicine*, 348(3). doi: 10.1056/nejmoa020177.

Zhang, X. and Yu, J. (2010) 'Target recognition-induced NK-cell responses', *Blood*, 115, pp. 2119–2120.

Zhang, Y. *et al.* (2005) 'Host-reactive CD8+ memory stem cells in graft-versus-host disease', *Nature Medicine*, 11(12), pp. 1299–1305. doi: 10.1038/nm1326.

Zhao, Z. *et al.* (2015) 'Structural Design of Engineered Costimulation Determines Tumor Rejection Kinetics and Persistence of CAR T Cells', *Cancer Cell*, 28(4), pp. 415–428. doi: 10.1016/j.ccell.2015.09.004.

Zhu, J., Yamane, H. and Paul, W. E. (2010) 'Differentiation of effector CD4+ T cell populations', *Annual Review of Immunology*, 28, pp. 445–489. doi: 10.1146/annurev-immunol-030409-101212.

Zlotoff, D. A. and Bhandoola, A. (2011) 'Hematopoietic progenitor migration to the adult thymus', *Annals of the New York Academy of Sciences*, 1217(1), pp. 122–138. doi: 10.1111/j.1749-6632.2010.05881.x.

Zúñiga-Pflücker, J. C. (2004) 'T-cell development made simple', *Nature Reviews Immunology*, 4(1). doi: 10.1038/nri1257.

Long-term lymphoid progenitors independently sustain naïve T and NK cell production in humans

Natalia Izotova[1,5], Christine Rivat[1,4,5], Cristina Baricordi[2], Elena Blanco [1], Danilo Pellin[2], Eleanor Watt[3], Athina S. Gkazi[1], Stuart Adams [3], Kimberly Gilmour [3], Jinhua Bayford[3], Claire Booth[1,3], H. Bobby Gaspar[1,4], Adrian J. Thrasher [1,3,6✉] & Luca Biasco[1,2,6✉]

Our mathematical model of integration site data in clinical gene therapy supported the existence of long-term lymphoid progenitors capable of surviving independently from hematopoietic stem cells. To date, no experimental setting has been available to validate this prediction. We here report evidence of a population of lymphoid progenitors capable of independently maintaining T and NK cell production for 15 years in humans. The gene therapy patients of this study lack vector-positive myeloid/B cells indicating absence of engineered stem cells but retain gene marking in both T and NK. Decades after treatment, we can still detect and analyse transduced naïve T cells whose production is likely maintained by a population of long-term lymphoid progenitors. By tracking insertional clonal markers over-time, we suggest that these progenitors can support both T and NK cell production. Identification of these long-term lymphoid progenitors could be utilised for the development of next generation gene- and cancer-immunotherapies.

[1] Great Ormond Street Institute of Child Health Faculty of Population Health Sciences, London, UK. [2] Gene Therapy Program, Dana-Farber/Boston Children's Cancer and Blood Disorders Center, Harvard Medical School, Boston, MA, USA. [3] Great Ormond Street Hospital, London, UK. [4] Present address: Orchard Therapeutics, University College of London (UCL), London, UK. [5] These authors contributed equally: Natalia Izotova, Christine Rivat. [6] These authors jointly supervised this work: Adrian J. Thrasher, Luca Biasco. ✉email: a.thrasher@ucl.ac.uk; l.biasco@ucl.ac.uk

294

Historically, survival and activity of individual human hematopoietic progenitor subtypes have been studied exclusively via transplantation in permissive mouse models[1–3]. Despite being a relevant investigational tool, these models do not recapitulate the human hematopoietic milieu[1], and production of human T and NK cells is severely compromised. It has therefore been impossible to confidently measure genuine in vivo dynamics of human T/NK lymphoid progenitors[4]. Hematopoietic stem cell (HSC) gene therapy (GT) using integrating viral vectors has opened a unique opportunity to trace the fate of transplanted cells for the first time directly in vivo in humans by means of integration site (IS) clonal tracking[5–7]. From our most recent analysis in patients treated with hematopoietic stem and progenitor cell (HSPC) GT[8], we predicted mathematically that a population of human lymphoid progenitors might exist which is capable of surviving long-term (Lt) in the absence of ongoing contributions from HSC. Using a similar approach, others have also recently confirmed a diversity of HSPC outputs in different disease settings and aligned with our prediction of a Lt lymphoid-biased cell production, which held true even when complex statistical controls are put in place to obviate restricted sample availability, technical biases and cross-sample contamination effects[9]. However, to date, no experimental setting has been available to validate such a mathematical prediction neither in the mouse nor in humans. We have now studied a unique group of patients in whom Lt lymphoid recovery is sustained in the absence of HSC engraftment. X-linked severe combined immunodeficiency (SCID-X1) is caused by deficiency of the common cytokine receptor gamma chain, a component of multiple cytokine receptors (IL2, IL4, IL7, IL9, IL15 and IL21) responsible for many aspects of lymphoid development and function[10,11]. Infants with SCID-X1 classically present with complete absence of T and NK cells, with normal numbers of intrinsically dysfunctional B cells[12]. Patients with SCID-X1 treated in some of the earliest successful ex vivo GT trials using gammaretroviral vectors now have more than 15 years of follow up (Supplementary Table 1)[13,14]. Importantly, although they received transduced CD34+ HSPCs, no conditioning was administered based on the rationale that corrected T cells would have a very potent growth and survival advantage, and that the longevity of the emergent T cell population would provide major clinical benefit. Of 10 patients treated all are alive with persistence of functional transduced peripheral cells.

Here we show that, in the absence of myelosuppressive conditioning, SCID-X1 patients display no vector-positive myeloid/B cells indicating absence of engineered HSC engraftment but instead retain high level gene marking in both T and NK cells for up to 19 years. Nonetheless, through a comprehensive Lt immunophenotypic, molecular and functional characterisation we show that we still are able to detect bona fide vector-positive naïve T cells, accompanied by maintenance of thymic activity and high T-cell receptor (TCR) diversity in naïve T cells overtime. In the absence of gene corrected HSC, T cell production is therefore likely maintained de novo in these patients by a population of long-term lymphoid progenitors (LtLP). IS analysis further demonstrates in vivo clonal stability/function of LtLP overtime in the absence of insertional mutagenesis. On the basis of NK cell immunophenotyping and molecular tracking we also suggest that these progenitors could be supporting Lt bipotent T/NK-specific output in these patients.

Results
Immunophenotypic and molecular follow up of SCID-X1 GT patients. To assess survival capability, differentiation potential and clonal dynamics of LtLP we designed an analytical workflow combining immunophenotypic, functional and high-throughput

molecular assays on samples collected from five SCID-X1 patients, 2.9–18 years after GT (Fig. 1A). As part of the normal follow-up protocol of our SCID-X1 patients, immunophenotyping and vector copy number (VCN) in different blood cell lineages is analysed at a number of time points post treatment (Fig. 1B). Peripheral blood mononuclear cells (PBMCs) from each patient were routinely sorted into five different lineages (neutrophils, T cells, B cells, NK cells and Monocytes) using fluorescence-activated cell sorting (FACS). Quantitative PCR (qPCR) demonstrated that the average copy number for T and NK cells was 2.1 and 2.3 copies per cell, respectively, while in B cells and myeloid cells, VCN remained at the background level since 22–27 months after GT when we started these measurements (Fig. 1C). Because it is known that maintenance of circulating myeloid and naïve B cells is dependent on a continuous output from the bone marrow these data are consistent with a permanent loss or failure of engraftment of engineered long-term (Lt) HSC. Therefore, we initially reasoned that our results could be explained by the survival of vector-marked long-lived circulating peripheral lymphocytes. To test this hypothesis, we investigated the dynamics of the lymphocyte reconstitution over time. We firstly looked at absolute numbers of different immune cell lineages in peripheral blood including T cells (CD3+), B cells (CD19+) and NK cells (CD16/CD56+) (Fig. 1D). Overall, during the early immune reconstitution phase (within the first 6 months), all five patients showed CD3+ cell numbers within normal range, and these numbers remained at or just below the normal range at latest follow up. CD4 and CD8 T cells were present throughout the time of observation although an inverted CD4/CD8 T cell ratio was detected in all patients (Fig. 1E).

To explore the composition of circulating vector positive T cells in these patients, we performed a standard immunophenotypic analysis. Initially, we used CD62L and CD27 cell surface markers to identify a double positive (DP) mixed T cell precursor population composed of naïve T cells (TN) and stem cell memory T cells (TSCM), a central and effector memory T cells (TCM/TEM) population and effector T cells (TEFF) (Fig. 1F). In both CD4 and CD8 subsets, T cell precursors (CD3+ CD62L+ CD27+) were maintained over time up until the latest follow up in a relatively stable fashion in all patients except in patient 8 (P8). In this patient, T cell precursors were preserved up until month 100 post GT, but rapidly declined after this time point, until their relative percentage dropped to <1% of total CD4/CD8 populations. Of note P8 developed T-acute lymphoblastic leukaemia (T-ALL) at 24 months post GT and subsequently underwent a 3-year chemotherapy regimen, achieving long-term clinical and molecular remission[13]. We suspect that this burden of chemotherapy may well be a significant contributor to the observed sharp decline of T cell precursor populations in this particular patient at around 5 years post remission, underlining the biological consistency of our immunophenotypic analyses.

Immunophenotypic and functional characterisation of bona fide naïve T cells. The presence of T cell precursors up to the latest time points was surprising but could be explained by persistence of TSCM which are endowed with decade-long survival in humans[15–18]. To address this point, we used six cell surface markers (CD3, CD4, CD8, CD45RA, CD95 and CD62L) to identify DP precursor (CD3+ CD4/CD8+ CD62L+ CD45RA+), TCM (CD3+ CD4/CD8+ CD62L+ CD45RA−), TEM (CD3+ CD4/CD8+ CD62L− CD45RA−) and TEMRA (CD3+ CD4/ CD8+ CD62L− CD45RA+) subsets (Fig. 2A). Most of the T cells exhibited one of the three classical memory phenotypes, TCM, TEM or TEMRA (Fig. 2B). We then used CD95 to differentiate between the two subsets that comprise the DP precursor
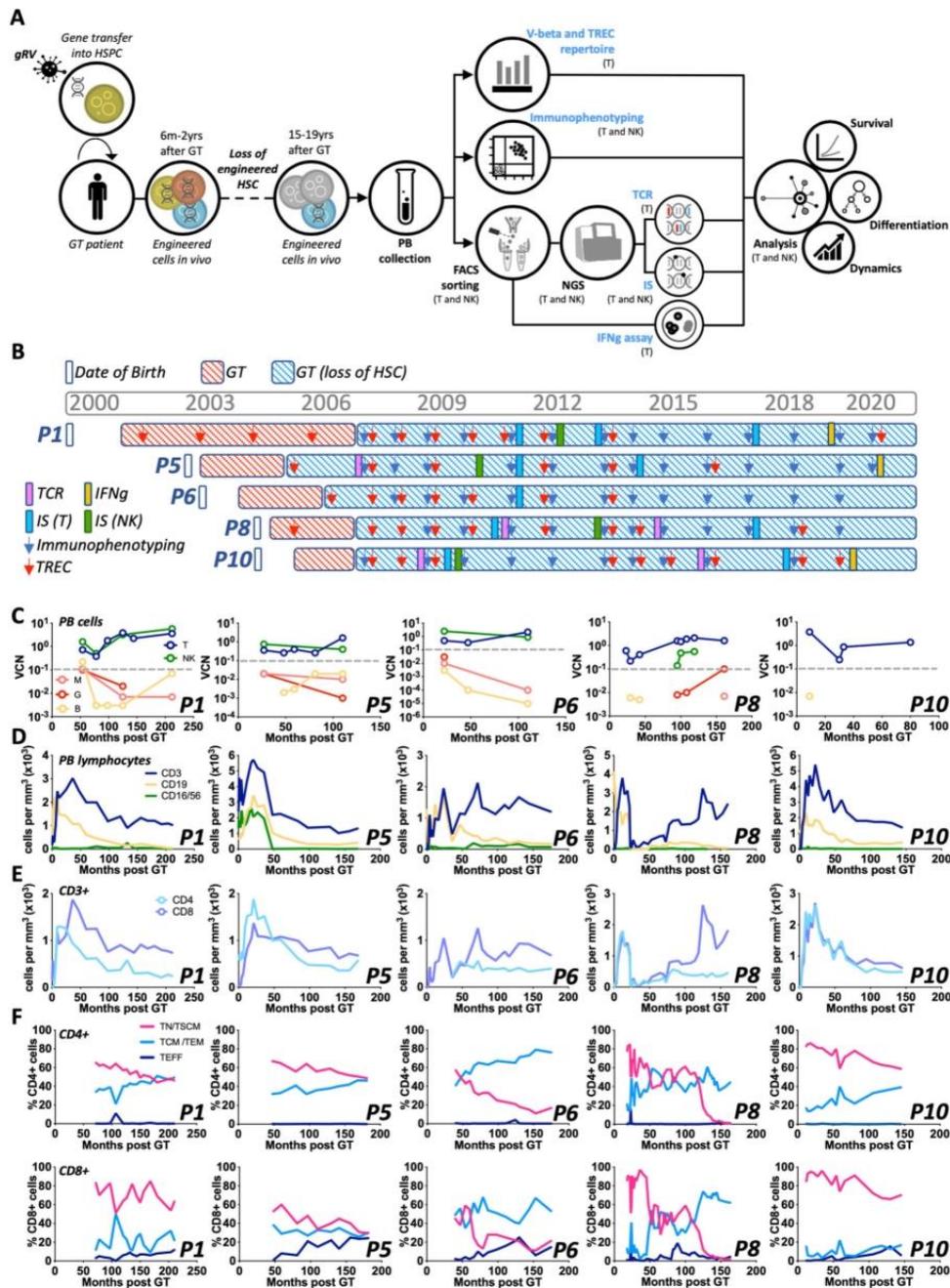
population of T cells, TN (CD95−) and TSCM (CD95+)[19]. Surprisingly, in 4 out of 5 patients, the majority of cells within the DP population still displayed a TN CD95− phenotype up to the latest follow up (Fig. 2C) in both CD4+ and CD8+ populations (Supplementary Fig. 1). Interestingly, although P8 had a substantially lower numbers of DP cells, 49% of this population was also composed by true TN CD95− (0.5% of total CD3+ cells)

(Fig. 2B). Most importantly, we could establish, by means of VCN evaluation on isolated T cell subtypes, that all CD95− TN were vector positive up to the latest follow-up available and in all patients (average VCN of 1.6, Supplementary Fig. 2). To functionally validate the detection of true TN cells we FACS sorted four T cells subsets (TN, TSCM, TCM and TEM) from three patients and from healthy donors (HD), and stimulated them for

296

Fig. 1 Experimental workflow and molecular/immunophenotypic profile of patients' blood cells. A Schematic of the study design (gRV = gammaRetroviral vector; HSPC hematopoietic stem/progenitor cells, GT gene therapy, HSC hematopoietic stem cells, PB peripheral blood, NGS next generation sequencing, TCR T-cell receptors, IS integration sites). **B** Gantt chart describing patients follow up and approximate dates when assays were conducted. TCR T cell receptor sequencing (pink rectangle), IFNg interferon gamma secretion assay (orange rectangle), IS (T) integration site analysis in T cells (blue rectangle), IS (NK) integration site analysis in NK cells (green rectangle), Immunophenotyping is shown by blue arrows, TREC T cell receptor excision circle content (red arrow). **C** Vector copy number (VCN) measured overtime in T cells (T, dark blue), NK cells (NK, green), B cells (B, yellow), Monocytes (M, salmon) and Neutrophils (N, red) of the five patients object of this study (P1, P5, P6, P8 and P10). The grey dotted line shows background threshold for VCN assay. **D** Absolute counts of T (CD3+, dark blue), B (CD19+, yellow), NK (CD56/16+, green) lymphocytes in peripheral blood overtime. **E** Absolute counts of CD3+CD4+ (light blue) and CD3+CD8+ (blue) cells in peripheral blood overtime. **F** Percentages of naïve T cells (TN)/T memory stem cells (TSCM) precursors (pink), T central memory (TCM)/T effector memory (TEM) (light blue), T effector cells (TEFF) (dark blue) in CD4+ (upper panels) or CD8+ (lower panels) cells overtime.

6 h with Phorbol 12-myristate 13-acetate (PMA) (50 ng/µl) and ionomycin (1 µg/ml), before adding Brefeldin A (10 µg/ml) and staining for IFN-gamma (IFNg) (Fig. 2E). All patients displayed IFNg expression similar to HD controls (Fig. 2F). TN CD95− vector positive T cells, did not secrete significant levels of IFN-gamma upon stimulation as compared to the other subsets indicating that these were genuine TN cells (Fig. 2G).

Evaluation of thymic activity and TCR repertoire in GT patients. The existence of bona fide vector-positive TN cells suggested that a de novo production of T cells was still occurring many years after the loss of transplanted and transduced HSC. We next wanted to rule out the possibility that we were detecting aberrant long-lived T cells retaining partial TN characteristics and verify instead that we were observing the result of ongoing T cell output. Because new naïve T cells can only be produced in the thymus, we first looked at T-cell receptor excision circle (TREC) content as a surrogate measure of thymic activity (Fig. 3A). Consistent with the presence of genuine TN cells, we observed detectable TREC levels in 4 out of 5 patients up to the latest timepoints suggesting that an active thymic output was maintained years after loss of gene-corrected HSC. P8 showed, as expected, a significant drop in TREC content at the latest follow up, consistent with his decreasing numbers of DP precursors and TN cells. A second marker of physiological thymic activity is a normal distribution of TCR repertoire. We assessed the TCR repertoire composition of the whole T cell population in these patients by Vbeta spectratyping. This was normal in all patients, again except for P8 whose profile was more oligoclonal (Supplementary Fig. 3). These results are both consistent with sustained active thymic output. However, they do not discriminate the repertoire of individual T cell subtypes and, more specifically, do not show whether new TN cells were being generated. To address this point, we FACS-sorted 4 T cell subtypes (TN, TSCM, TCM, TEM) from one HD control as well as from two patients, one with high (P10) and one with low proportion of precursors T cells (P8) at two time points each after loss of engineered HSC, and collected by high throughput sequencing a total of 1193 TCR rearrangements (Supplementary Data 1). Our assumptions were the following: (a) if the thymus of these individuals was active, we would expect to detect new TCR rearrangements in TN over time and (b) if new short-lived TN were indeed continuously produced, we should observe high diversity of the TCR repertoire in this population. As shown in Fig. 3B we detected a wide diversity of TCR rearrangements in all T cell subpopulations. In P8, a lower range of VDJ recombination in TN was observed compared with P10 and the HD control, which again would be consistent with reduced thymic activity (Fig. 3C). The quantification of TCR diversity was in the range of the HD control for both patients and in all populations (Fig. 3D) including TN cells (Fig. 3E). When tracking TCR rearrangements in TN cells overtime we observed that most of them were detected only at individual time points,

suggesting ongoing production of new TN cells by the thymus (Supplementary Fig. 4). Because sampling biases could affect the interpretation of longitudinal tracking of individual TCR clones, we first investigated the level of sharing of identical TCRs at each time point among the four subpopulations analysed. As shown on the network plots of Fig. 3F, sharing of identical TCR was generally low among subtypes with highest level detected between TCM and TEM in line with the biology of immunological memory and the long-term survival of these populations. As a further biological validation to our findings, in P8 we could observe, particularly at the latest time point, a substantially higher sharing of TCRs across all subtypes as compared to P10 (Supplementary Fig. 5), a result which again aligns with the progressive loss of thymic output in this individual. We then quantified the level of TCR recaptures overtime in all populations under the hypothesis that a continuous generation of new rearrangements should be reflected by low re-capture of identical TCR in TN at multiple time points. As shown in Fig. 3G, <5% of the TCRs detected in TN cells of one time point were recaptured in a second time point testifying continuous production of precursor cells with new TCR rearrangements by the thymus. Overall, the results of immunophenotypic, functional and molecular assays all lead to the conclusion that long-term de novo production of T cells by the thymus is preserved in these patients despite the lack of persistence of engineered HSC and is likely maintained by a population of vector-positive LtLP.

Clonal tracking of T-cell subpopulations by IS analysis. As a by-product of retroviral-mediated gene-correction, the presence of semi-randomly integrated vector sequences allows tracking of LtLP activity at clonal level by means of IS analysis, using freshly produced TN as surrogate marker of output. We therefore combined linear-amplification-mediated (LAM)-PCR with high-throughput sequencing and collected and analysed 12,756 unique IS from five populations, including TN cells, in a window of 10.1–14.9 years after loss of transplanted HSC (Supplementary Table 1 and Supplementary Data 2). We first wanted to assess whether the existence of LtLP was the result of aberrant vector-induced clonal selection or whether it represented physiological and homoeostatic survival. As shown in Supplementary Fig. 6, we occasionally detected clones contributing to more than 20% of individual T cell subtypes sampled at individual time points, as expected from physiological T cell dynamics combined with sampling biases. However, when looking at the longitudinal trend, clonal diversity remained stable in all T cell subtypes including naïve T cells (Fig 4, Supplementary Fig. 7 and Supplementary Table 4). Notably, despite a drop of circulating naïve T cells from around 100 months post GT, P8 also maintained a substantial polyclonal profile in the TN population up to the latest follow up. We then looked for IS skewing towards specific genomic loci as a marker of IS-driven clonal selection. As shown in the Supplementary Figs. 8–12, IS sites were widely distributed

297
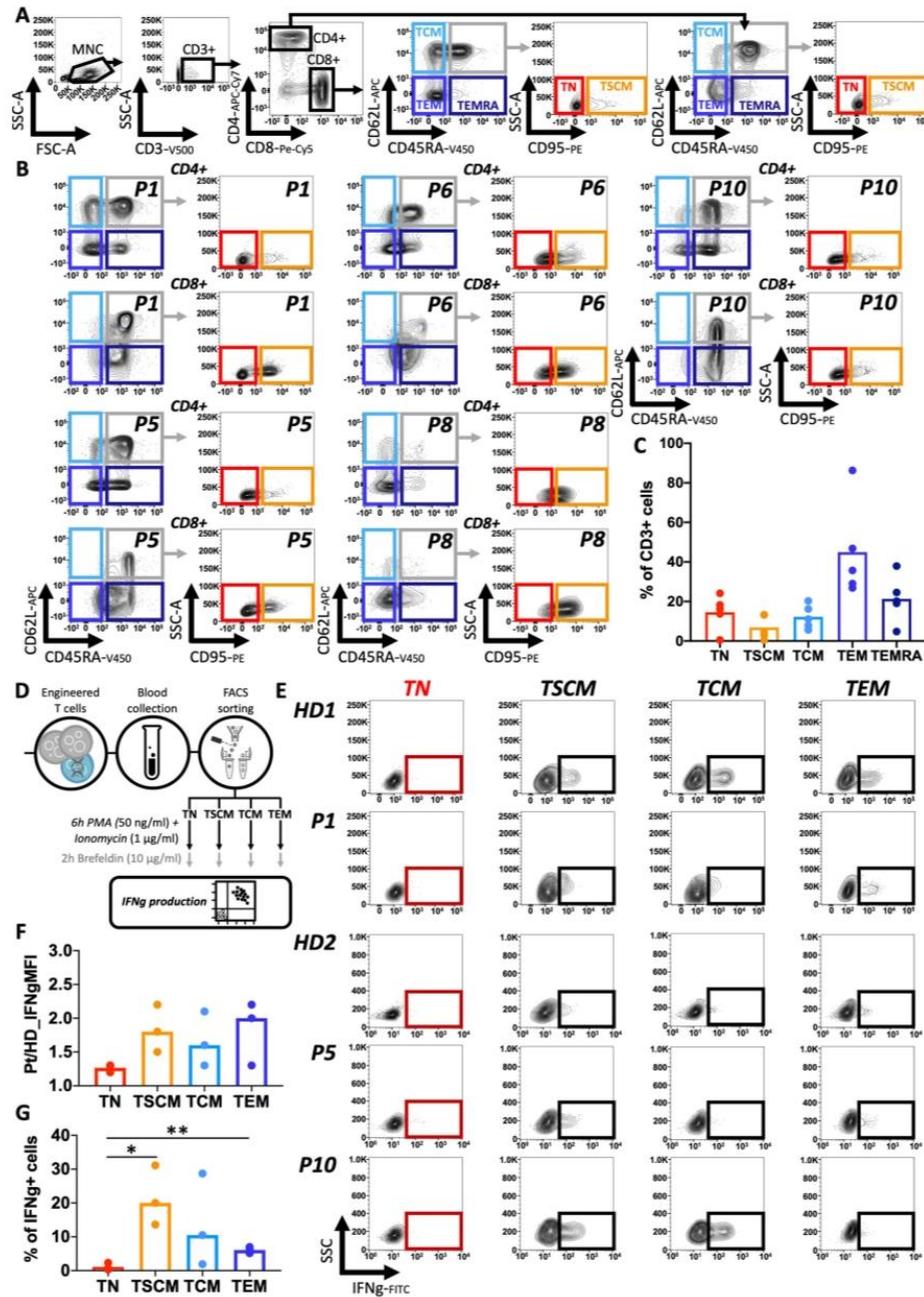
Fig. 2 Immunophenotypic and functional validation for the identification of naïve T cells. A Gating strategy used for identifying naïve T cells (TN), T memory stem cells (TSCM), T central memory (TCM), T effector memory (TEM) and T effector memory CD45RA+ (TEMRA) subpopulations inside the CD4+ or CD8+ T cells (data displayed are from a healthy donor used as reference). **B** FACS plot showing CD4+ and CD8+ T cell composition at the latest follow up for each patient. **C** Percentage of TN, TSCM, TCM, TEM and TEMRA inside the CD3+ population at latest follow up in the five patients analysed. **D** Experimental scheme for Interferon gamma (IFNg) production assay on four sorted T cell subpopulations. **E** FACS plot showing IFNg expression after PMA/Ionomycin stimulation in each sorted subpopulation from two healthy donors (HD1 and HD2) and three patients (P1, P5 and P10). **F** Ratio between IFNg mean fluorescence intensity (MFI) measured in patients vs. healthy donors T-cell subtypes after stimulation. **G** Percentage of IFNg-positive cells measured in each T-cell subtype from patients after stimulation (unpaired two-tailed t-test *$p < 0.05$ [TN vs. TSCM $p = 0.0165$], **$p < 0.01$[TN vs. TEM $p = 0.0033$]).
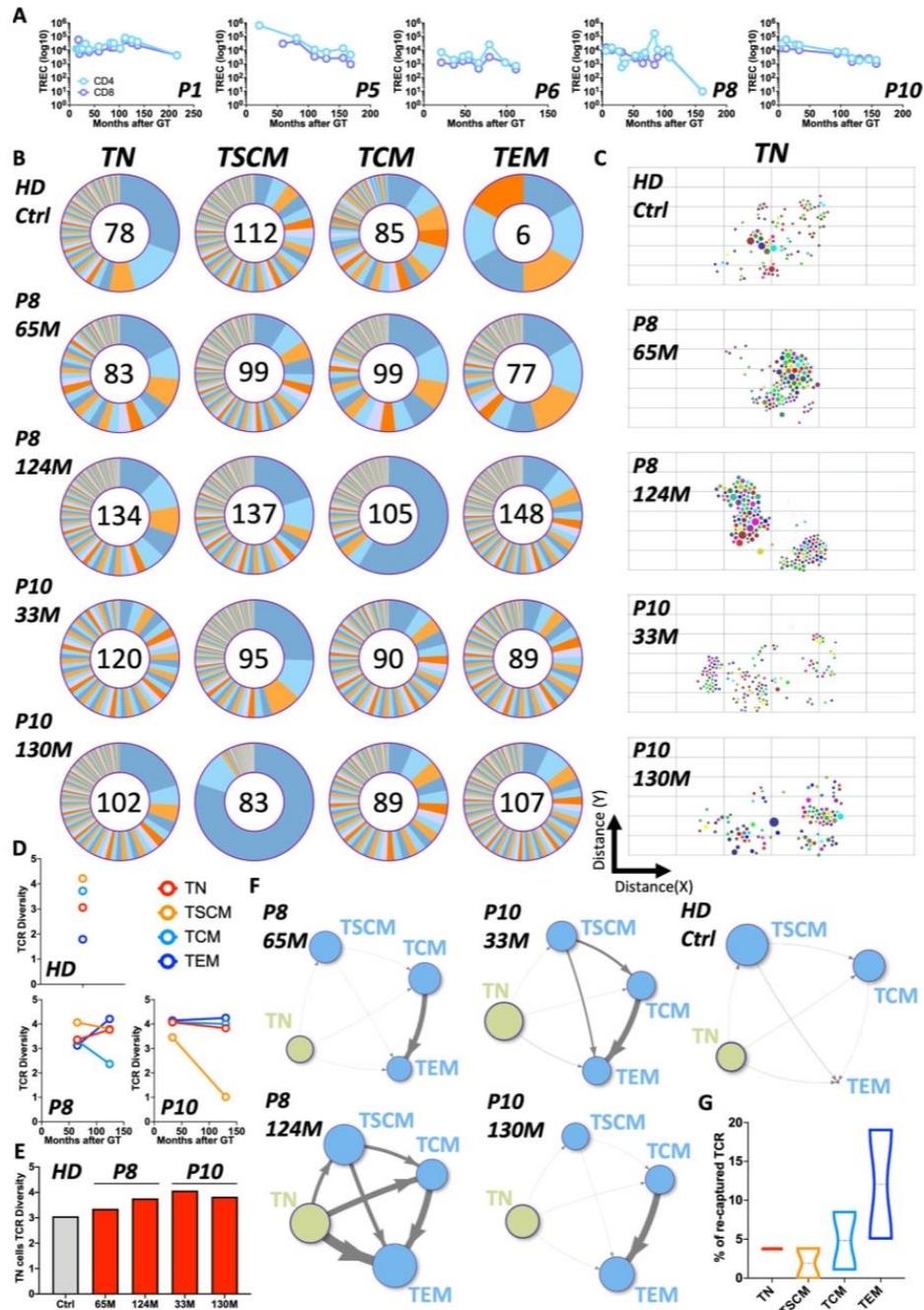
298

across multiple genes and we did not detect any evidence of selection for proto-oncogenic regions in any of the T cell sub-populations and time points analysed. Interestingly, we observed that IS of TN cells were located preferentially in regions involving gene categories such as "Leucocyte activation", "Immune effector process" and "T-cell activation" (Supplementary Fig. 13). Because

these loci were not preferentially targeted upon CD34+ cells transduction[20,21] and because we do not have evidence for aberrant clonal selection, we can speculate that this could be the result of a physiological positive-selection for clones carrying the vector in genomic locations allowing for optimal transgene expression upon T cell differentiation/maturation.

Fig. 3 Assessment of thymic activity in GT patients. A T-cell receptor excision circles (TREC) content measured in CD4+ and CD8+ cells in five patients overtime. **B** Ring plots displaying the relative fractions of TCR rearrangements detected within each T-cell subtype sorted from a healthy donor control (HD) and two patients (P8 and P10) at two timepoints each (months = M) after GT. Numbers inside each ring correspond to the numbers of individual TCR rearrangements retrieved from each sample. **C** Diversity of TCR rearrangements shown as a 2-dimension t-SNE plot. Coordinates are derived from computing the differences (distance) among clones in sequence composition of a 50 bp window centred on the Complementarity-determining region 3 (CDR3). The size of each dot is proportional to the abundance of the corresponding TCR rearrangement. In these graphical representations the higher it is the spread of dots across the area of the plot the higher it is the diversity of rearrangements observed within the TCR locus. **D** TCR diversity measured by Shannon Diversity index in each sample and T-cell subtypes of HD, P8 and P10. **E** Diversity of TCR repertoire in naïve T cells (TN) of HD (grey bar) and Patients (red bars) at different timepoints). **F** Networks displaying TCR sharing between the four T-cell subtypes from HD or P8 and P10 at each follow up. Nodes represent T-cell subtypes while edges (arrows) represent degree of TCR sharing. Size of each node is proportional to the number of TCR rearrangements detected in each T-cell subtype (Naïve T cells [TN] in light green, other subtypes in light blue). The thickness of the arrows is proportional to the Pearson correlation coefficient calculated on the basis of TCR sharing between each pair of T-cell subtype. **G** Violin plots showing percent of TCR recaptured within each T-cell subtype isolated from two patients at two independent timepoints.
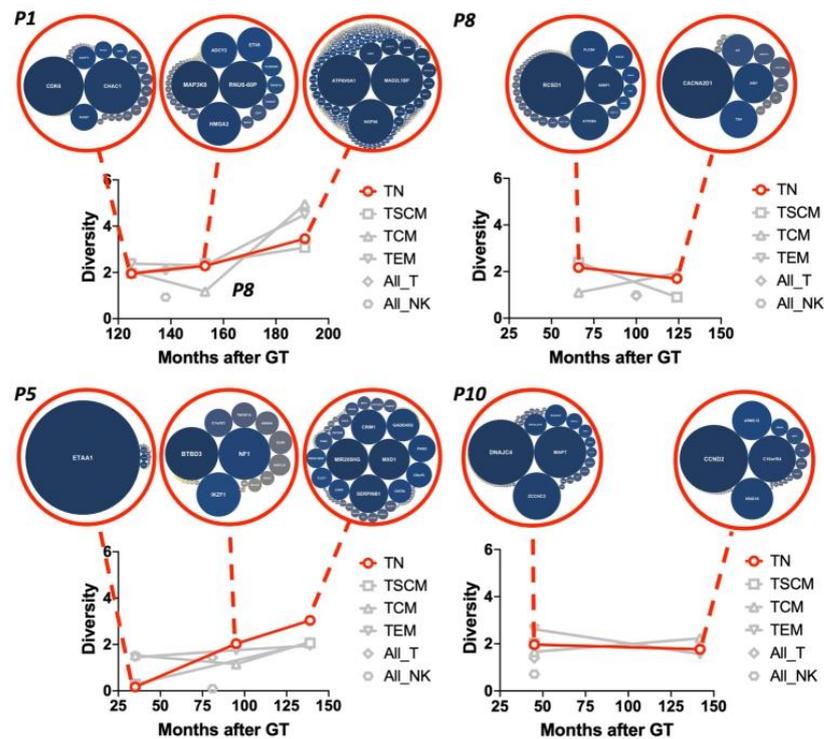
Fig. 4 Integration sites diversity in naïve T cells and other T-cell types overtime. Plots showing Shannon Diversity Index of IS overtime in TN (red lines) and in the other T-cell subtypes (grey lines). (bubble) For TN at each timepoint analysed red circles contain bubble plots of clones contributing >0.01% to the total population. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside the bubble. Patients shown are the ones from which multiple timepoints were available (the single timepoint available from patient 6 is shown in Supplementary Fig. 7).

We then looked at the presence of identical IS among different T cell subtypes to confirm that newly produced TN cells are capable of physiologically differentiating into memory subtypes in vivo. As a confirmation of the biological relevance of our data, IS sharing was high and consistent across samples collected, sorted, analysed and sequenced independently (Fig. 5A and Supplementary Fig. 14). We detected a high number of IS shared between naïve T cells and memory subsets in all patients (from 24.1% to 57.2% of IS isolated from TN of each individual) suggestive of ongoing in vivo TN differentiation (Fig. 5B). A similar range of IS sharing was detected in all subpopulations (Fig. 5C), a sign of significant clonal relationship among the

T-cell subpopulations and the existence of active common progenitors. When comparing IS data with the TCR profiling, it is notable that TN shared the least number of TCR rearrangements with other subtypes at other time points but at the same time they shared equal number of IS. This suggests that individual LtLP clones generate TN with identical IS to previously generated memory cells but with new and different TCR specificities. To provide more formal evidence that individual LtLP clones are sustaining de novo T cell production over many years we used longitudinal data collected from TN cells and interrogated our dataset for detection of identical IS overtime. As shown in Fig. 5D and Supplementary Fig. 15, in all four patients from which we
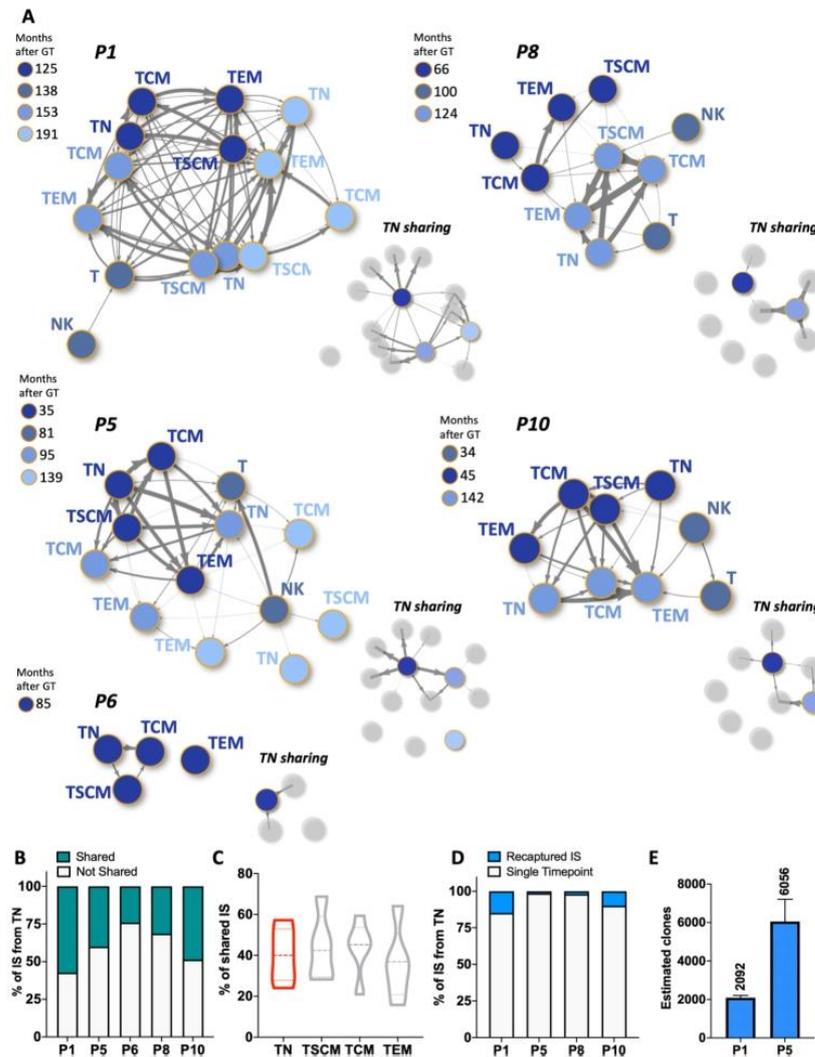
Fig. 5 Sharing and recapture of integrations sites across T-cell subpopulations and timepoints. A Network plots showing significant sharing of IS among the T-cell subtypes and NK cells (data described in Fig. 5) from each patient. Each node represents a sample and it is coloured according to the timepoint after GT. The thickness of the arrows is proportional to the value of positive Pearson correlation coefficients calculated on the basis of IS sharing between pairs of samples (ranges of values are displayed on the heatmaps of Supplementary Fig. 14). For each patient two plots are shown: the networks on the left display positive correlations among all samples while the network on the right (TN sharing) highlights correlations between TN cells (nodes in scale of blue) and the rest of the subtypes (nodes in grey). **B** Stacked bars showing in green the fraction (%) of IS from TN of all timepoints for each patient which were shared with at least another T cell subtype. **C** Violin plots showing the percentage of IS belonging to each T subpopulation shared with at least another T cell subtype (data from all patients and timepoint; TN data highlighted in red). **D** Stacked bars showing in blue the fraction of IS re-captured across multiple timepoints within the TN populations of each patient (a more detailed view of re-captured IS from TN is displayed on the heatmaps of Supplementary Fig. 15). **E** Clonal abundance (number of clones in circulation) of long-term lymphoid progenitor estimated in P1 and P5 on the basis of size of IS datasets at each timepoint and re-capture probabilities overtime of IS within the TN compartment (values derived from the Mh Chao loglinear model for capture-recapture and relative standard errors as reported Supplementary Table 5).

had T cells analysed at multiple timepoints, a fraction of IS in TN (1.3–14.8%) could be recaptured overtime in a window of up to 10.1 years from one collection to another. In the two patients where analysis could be performed at more than two time points, we used IS recapture probability to estimate the number of active engineered LtLP clones (Supplementary Table 5). In these

individuals, T cell production is maintained by approximately 2000–6000 individual engineered ltLP (Fig. 5E).

Immunophenotypic and molecular analysis of vector-positive NK cells. The fact that NK cell subset is the only other immune

cell subset that still carries integrated vector copies raised the intriguing possibility that T and NK cells of these patients could share a common LtLP with dual potential. Indeed, in normal circumstances NK cells are short-lived lymphocytes, with a half-life of approximately 2 weeks, except for memory or adaptive NK cells. Memory NK cells are commonly identified by the expression of CD94/NKG2C. They are mainly observed in response to human cytomegalovirus and they can be detected months or even years after the infection[22–26]. In humans, at least five different maturation stages of NK cells have been reported identifiable on the basis of cell surface marker expression[27,28].

Differently from what observed for T cells, the absolute number of CD56+ NK cells detected in the peripheral blood of our patient was variable overtime remaining constantly below normal levels up to the latest timepoint analysed (Fig. 6A, B). Notably however, all NK cells displayed expression of the IL2RG transgene at levels similar to T cells and to the healthy donor control (Fig. 6C). To confirm the presence of bona fide NK cells

(Fig. 6D), we performed a comprehensive immunophenotyping (Supplementary Table 6) analysing the expression of CD56, CD16, Perforin, Granzyme B, CD94, CD57, KIRs, NKG2C and NKG2A in three patients and healthy donor controls (Fig. 6E, F and Supplementary Fig. 16A, B). The proportion of CD3−CD19−CD56+ and/or CD16+ true NK cells in our patients over the total lymphocyte population was low (0.2–1.2%) compared to healthy controls (8–29%) (Supplementary Fig. 16A). These data are in accordance with the VCN measurements in NK cells shown in Fig. 1C and the low absolute numbers of circulating NK cells shown in Fig. 6A, B. As for the T cells, a key question was whether these NK cells were long-lived memory cells or whether instead we could also detect bona fide NK short-living precursors suggesting that a de novo NK production is occurring. Despite the low number of cells available for this analysis (particularly in P1), CD56[bright] NK precursors[27] were detectable in the two of the three patients where enough events could be captured (7.7–9.9% of total NK cells) although the mean fluoresce intensity (MFI) of
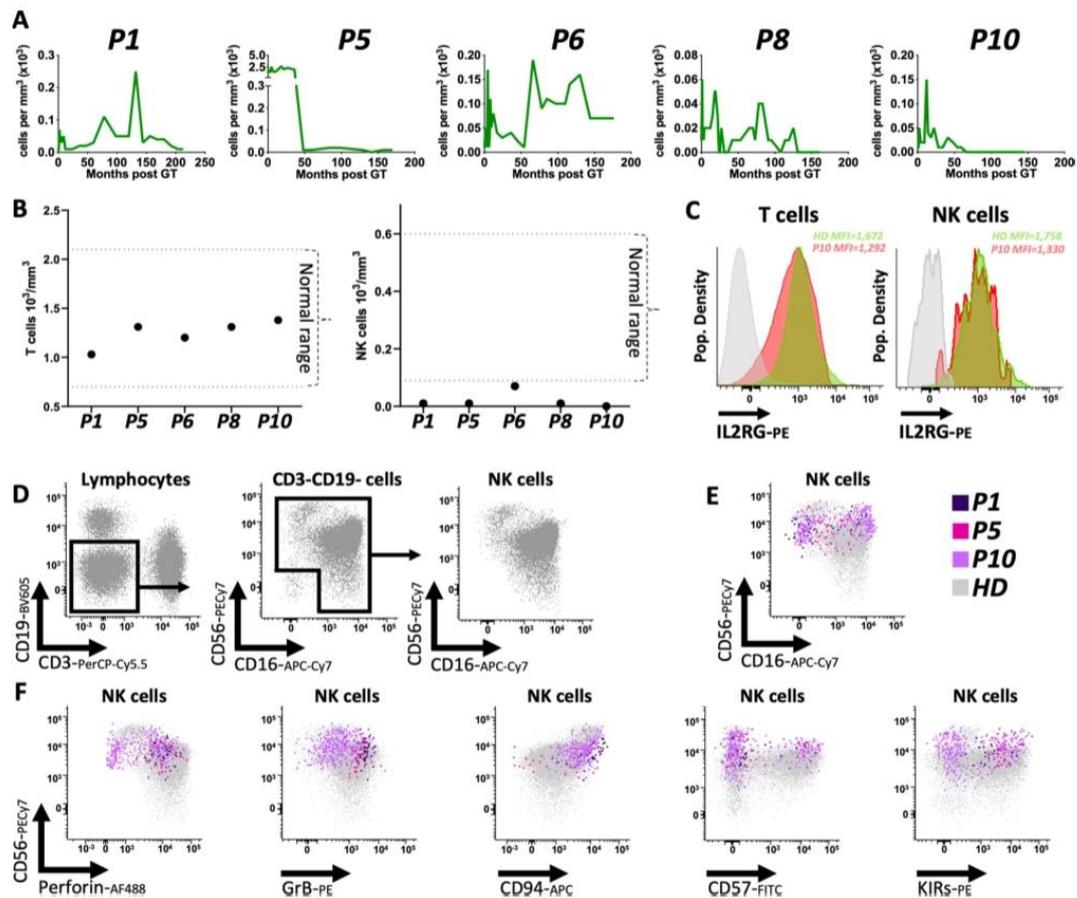
Fig. 6 Analysis of peripheral blood contribution and immunophenotype of NK cells. A Absolute counts of NK cells overtime in P1, P5, P6, P8 and P10. B Comparison of absolute counts of T (left plot) and NK (right plot) cells in patients (black dots) as measured at the latest available timepoint vs. normal ranges measured in healthy age matched individuals (dotted lines). C FACS plots showing the γchain expression in T cells and NK cells in patient 10 (red) and a healthy donor (green). Isotype control is shown in grey. D FACS plots showing the gating scheme for the identification of NK cells within the lymphocyte population of a healthy donor used as reference. E Overlayed FACS plot showing the expression of CD56 and CD16 on NK cells in HD, PT1, P5 and P10. F Comparison of expression of Perforin, Granzyme B (GrB), CD94, CD57 and Killer-cells immunoglobulin-like receptors (KIRs) between a representative HD vs. P1, P5 and P10 inside the NK cells compartment.

CD56 within this subset was slightly lower (mean = 29,616) as compared to the one measured in healthy donors used as controls (mean = 39,696) (Fig. 6E and Supplementary Fig. 16A). In order to further assess if and how many of these NK cells could be long living adaptive/memory NK cells, we also analysed the expression of NKG2C and NKG2A in P10. Very few NK cells displayed a phenotype compatible with memory/adaptive NK cells (NKG2C+ NKG2A−)[29] (Supplementary Fig. 16B) although this definition is still contentious[30]. Also, a subset of NK cells expressed CD57, a marker generally associated with mature NK cells[23], although it should be noted that the percentage of CD57+ cells in our patients was lower (28–30%) as compared to healthy donors (39–68%) (Fig. 6E and Supplementary Fig. 16B). This is also different from the situation described in patients with GATA2 mutations who appear to maintain an aberrant population of long living memory/adaptive NK cells[31].

Having shown that a de novo NK production might be taking place, we performed molecular tracking analysis under the assumption that if NK cells originate from the same LP progenitor clones as T cells, we should be able to detect identical IS between the two subsets. To test this hypothesis, we collected and analysed 651 IS from FACS-sorted CD3−CD56+NK cells (Supplementary Tables 2 and 3). The amount and diversity of IS in NK cells were both lower as compared to IS collected from T cells (Fig. 5E), as expected from the low proportion of this lymphocyte population in the circulation of these patients when compared to T cells and to normal ranges for NK cell counts (Figs. 6B and 7A). However, and despite the limited sampling capability, we were still able to detect shared integrations between NK cells and distinct T cell subtypes (Fig. 7B and Supplementary Fig. 16C). To rule out the effect of cross-contaminations among samples we first quantified the level of shared IS between NK and TN or all T cells subpopulations which have been isolated from different time points and sequenced independently. Strikingly, we found that up to 41% of IS detected in NK cells were shared with independently analysed and sequenced TN. These numbers reach up to 60.7% when considering all T cell subpopulations from independent timepoints (Fig. 7C) and 90.2% when comparing with whole T cells samples analysed at the same timepoint (Supplementary Fig. 16C), strongly suggesting that NK and T cells are being produced de novo by a common LtLP population with dual T/NK potential.

Analysis of ISs in proximity of proto-oncogenes. As reported above, one of the patients in our trial (P8) developed a T-ALL associated to the integration of the gamma retroviral vector to the proto-oncogene LMO2 (LIM domain only 2) (1). Insertions next to other proto-oncogenic sites (CCND2, MECOM) also resulted in malignant transformation in SCIDX1 patients in a similar GT trial conducted at Necker Hospital in France (2). Of note, the first patient of our trial (P1) was treated 19 years ago, making this study the longest GT follow-up available to date. This unique clinical setting provides the opportunity to understand the long-term effects of proto-oncogenic IS. We therefore investigated the presence and the relative frequency of clones that have integrations next to the main three proto-oncogenes (LMO2, CCND2 and MECOM) that previously were involved in insertional mutagenesis events (Fig. 8A). Notably, despite the fact that we could detect a total of 52 IS in these loci, all integrations analysed contributed to <10% of the total clonal composition of each population, apart from the one clone with one IS in the CCND2 gene whose maximum relative contribution was 40% but only in one T cell subtype (TN) and at one time point (142 months after GT). Indeed, as shown in Fig. 8B some of these proto-oncogenic integrations persisted overtime but we never observed a

progressive or stable expansion of any of the relative clones even over a time window spanning several years. To test whether in P8 the original leukaemic clone had reappeared and/or had contributed substantially to the survival of LtLP, we analysed a peripheral blood sample that was taken from P8 at the time of T-ALL (24 months) and compared it to samples that were taken after this time point (Fig. 8C). Confirming the accuracy of the IS quantification method used in this study, at the time of T-ALL, the IS in the LMO2 gene locus was vastly overrepresented in this patient's peripheral blood as compared to the other IS (relative abundance = 90.8%). After 24 months this clone was not detected anymore, with the exception of a measurement in T cells at 100 months after GT which however was in the range of potential sequencing noise (relative abundance = 0.0001%). Altogether these data suggest that LtLP are physiologically maintained in these individuals and that their long-term survival is not dependent on the effects of insertional mutagenesis in proto-oncogenic loci.

Discussion

Here we have shown that de novo T/NK production from transplanted HSPC can be maintained in humans for many years in the absence of engraftment of HSC. The absence of genetically engineered HSC in these individuals is inferred on the basis of the consistent lack of vector-positive circulating myeloid and B cells and is consistent with the fact that no myelosuppressive bone marrow conditioning was administered. The high number of detected IS and clones contributing to de novo T cell production at later time points (2000–6000) is not consistent with the notion that there is residual activity of a few multipotent HSC. Because we could not here observe a sustained output of vector positive granulocytes or monocytes, the results of this study would also indirectly suggest that human myeloid-biased HSPC are not capable of long-term survival independently from the output of multipotent HSC. This is in line with our previous study where we found no indication of long-term myeloid-biased clonal production either at the mature or at the stem/progenitor levels of haematopoiesis[8]. Conversely long-term lymphoid-biased production was clearly detected up to the multi-lymphoid progenitor level. This current work provides formal evidence that substantiates these predictions. The absence of vector-positive B cells indicates that the LtLP active in these individuals are at a differentiation stage that follows a segregation to NK/T cell fate. One working hypothesis is that these cells might resemble an early T progenitor stage previously immunophenotypically defined as CD34+CD38+CD7+ cells[1,32,33].

The anatomic location of the LtLP identified in this study is also not clear as we are unable to sample bone marrow or thymic biopsies from these patients, although we favour the hypothesis that they are located in the thymus. Recent studies have shown that there is a significant clonal restriction in the thymus during T cell development and that very few progenitor clones can generate T cells with highly diverse TCR repertoire[34]. This suggests that thymopoiesis might be at least partially self-sustainable independently of constant progenitor influx from bona fide HSCs for the generation of fully functional naïve T cells.

It has previously been shown that the human thymus at least during foetal development contains bipotent T/NK progenitors[35]. These progenitors, defined as CD34[bright]CD1−, were able to generate both T and NK cells when cultured in mouse foetal thymic organ cultures. The fact that the majority of NK cells in our patients had non-memory/non-adaptive phenotype supports the notion that we are observing freshly produced NK cells, contrasting with data reported in patients with GATA2 mutations, who show defects in NK CD56[bright] production and
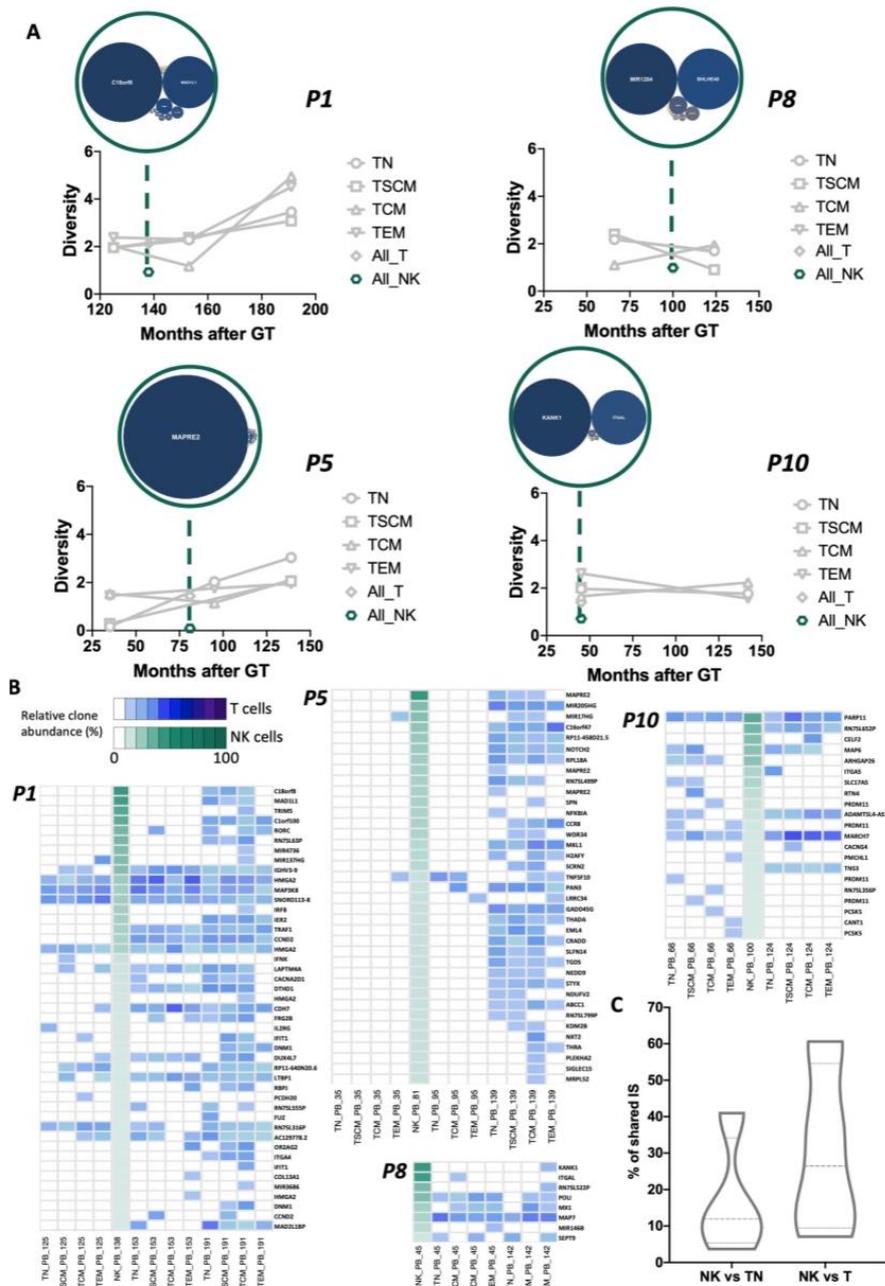
Fig. 7 Analysis of integration sites in NK cells. A Plots showing Shannon Diversity Index of IS overtime in NK cells (green dots) and in the other T-cell subtypes (grey lines). At each timepoint analysed green circles contain bubble plots of clones contributing >0.01% to the total NK population. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside the bubble. **B** Heatmaps displaying integration sites shared between NK cells and T cell subtypes in each patient. Each row is named by the name of the closest gene while each column is labelled by sample and timepoint (number of months after GT). Intensity of green (NK) or blue (T-cell subtypes) colours is proportional to the abundance of each integration site (white = not detected). **C** Violin plots showing the percentage of NK IS shared with at least one TN (NK vs. TN) or T cell-subtype (NK vs. T) in all patients.

304

persistence of memory/adaptive NK cells[31,36]. Of note, a high percentage (28–70%) of NK cells in our patients was CD56^{dim} CD16$^-$, which is normally associated with a minor subset in the peripheral blood of healthy individuals[37]. NK cells with this phenotype are often observed in higher percentages in paediatric bone marrow and leukaemic patients[38], as well as early after HSCT[27]. Despite their functionality being poorly understood, it

has been suggested that this unconventional NK subset represent an additional or alternative stage of NK cell differentiation[27]. As in our study, a reduced percentage of CD56^{dim} CD16$^+$ NK cells has been observed in SCID-X1 patients years after HSC transplant and GT treatments[39]. On the other hand, our results are apparently in contrast with data reported in non-human primates[40] in which it has been suggested that NK cells seem to

Fig. 8 Analysis of IS in proximity of proto-oncogenes associated to insertional mutagenesis. A Distribution of IS in proximity in LMO2, MECOM and CCDN2 loci. In each plot the *x*-axis displays chromosomal coordinates (bp) while the *y*-axis the IS abundance (log10 of percentage) calculated within the population and timepoint where it has been observed. The light blue dotted lines show the 1% and 10% abundance thresholds. The transcription start sites (TSS), exons (grey boxes) and introns (grey lines) of each gene are shown on top of each plot in their respective chromosomal localisation. **B** Heatmaps displaying the detection of these integration sites in each patient across each sample/timepoint. Each row is named by the locus of the IS and the name of the closest gene while each column is labelled by sample and timepoint (number of months after GT). Intensity of red is proportional to the abundance of each integration site (white = not detected). **C** Heatmap showing the IS collected in the peripheral blood (PB) of P8 during leukaemia at 24 months after GT (first column) and their detection overtime. Each row is named by the name of the closest gene while each column is labelled by sample and timepoint (number of months after GT). The first row displays the LMO2 integration associated to the insertional mutagenesis event while the rest of the rows show other by-stander IS collected from PB at 24 months after GT. Intensity of orange is proportional to the abundance of each integration site (white = not detected).

have an independent origin from other lineages, including T cells. Overall, the lack of consensus for the definition of long-lived NK cells in the literature and the limited availability of patient material for a more in-depth study of the properties of these cells prevent us at this stage from reaching a definitive conclusion as to whether there is a genuine de novo production of NK cells. It is notable that circulating NK cells never (or only transiently) achieve normal levels in our patients compared to healthy controls. One can speculate that the level of gene correction or pattern of gene regulation achieved in this situation did not allow LtLP to support a full NK differentiation and/or that the observed output by these putative bipotent LtLP might be a secondary modality of production of NK cells compensating the lack of a more physiological NK cell output from HSC.

Lastly, there are methodological constraints that should be taken into account when interpreting the results of this study. It is also important to recognise that our analysis was performed in a unique disease and treatment background. In terms of experimental methodology, the use of rare clinical material derived from a small cohort of patients constrains the number of different assays that can be performed on each sample at each time point. As a consequence, despite clear evidence of population trends, statistical significance for intragroup variance or non-parametric comparison among two groups cannot always be reached. Moreover, sample viability and quality are not always consistent between fresh and long-term cryopreserved material potentially resulting in some differences in data recovery across time points (e.g. different number of IS that could be collected). Residual cross-contamination among FACS-sorted populations is another factor that can in part affect the interpretation of the results of high-throughput molecular studies. However, it should be noted that in our study the observed high-level sharing of IS from TN with other cell subtypes (up to 57.2%) far exceeded the potential effects of insufficient cell purity after sorting (estimated TN purity upon re-analysis of sorted material up to 95.6%). More importantly, identical and highly abundant IS were detected between samples that were collected, isolated, processed and sequenced at independent times.

In conclusion, our data provide the first formal evidence in vivo in humans that de novo production of genetically engineered T and NK cells can be physiologically maintained by a population of ltLP surviving many years after loss of transplanted HSC. Identification and exploitation of such human ltLP population may be of significant benefit in the development of next generation GT and cancer immunotherapy approaches.

Methods

The reagents, methods and instruments used for the analyses described below changed during the course of the clinical trial, which was initiated in 2000. The methods described below are the ones currently in use in our laboratories. All changes in procedure were evaluated and validated to ensure consistency and comparability of the data.

Study design. Sample size was constrained by limited clinical sample availability. For phenotypic characterisation and IS analyses of T cell subsets in SCIDX1 patients we collected in vivo biological material available during years 2001–2018. We aimed at analysing for all the patients at least 1 early and 1 late point (where possible we did more). For IS analysis of NK and T cell identical integrations we acquired only 1 time point per patient, for TCR sequencing we aimed to acquire 1 early and 1 late time points, for IFN-gamma assay we analysed samples from the latest time points available. All available healthy donors' and patients' samples are reported in the manuscript. Technically validated results were always included to the analyses and we did not apply any exclusion criteria for outliers. For phenotypic characterisation of T and NK cells, Rainbow beads (RB) calibration was performed during the setup of the instrumentation for FACS analyses. RB acquisition was performed before each sample acquisition in order to achieve reproducible instrument setting among different experiments. For VCN evaluation, qPCR was validated for reproducibility. We run each sample in triplicate and values are reported as mean of the three triplicates. All attempts at replication were successful. The experimental design did not include allocation of samples to randomised experimental group. We analysed all the samples available during years 2001–2019. The experimental design did not include allocation to groups nor to blinding. There was no expected result prior to performing these analyses, therefore blinding tests were not applicable.

Clinical trial. Clinical trial registration. There was no requirement at the time to upload to public registries.

Some patients with X-SCID were referred to GOSH through an existing network, in some cases patients with X-SCID were recruited at the research site and were already be known to the Investigators. Patients and their parents were approached by the Principal investigator and given a patient information leaflet, during this consultation the trial was discussed at length and patients have been given ample opportunity to ask questions. The trial was also approved by the Medicines and Healthcare products Regulatory Agency (MHRA). Once the patients were satisfied that they would like to enrol into the trial, the principal Investigator consented them and the process was formally documented. The Gene therapy advisory committee (GTAC in London) is UK national Research ethics committee dedicated for the review of GT clinical research. GTAC had complete oversight of this trial. There was no requirement at the time to upload to public registries, so this trial is not listed on clinicaltrials.gov. Biological material from all patients was collected in accordance with all the ethical regulations. All five patients from this study for whom no human leucocyte antigen identical sibling donor or suitable matched unrelated donor was available underwent gamma retroviral GT without any conditioning regimen. All patients we studied are alive. One patient (Pt8) developed T cell acute lymphoblastic leukaemia 24 months after GT for which he received treatment. He has since then remained in clinical and molecular remission. The data was initially recorded in the source notes which would be paper-based medical records. The laboratory data was recorded on a hospital-based electronic system which was easily accessible for the research team. Outcomes were carefully defined during the design of the trial based on input from clinicians in this field. The outcomes were based on clinical importance to look at the effect of IMP. The outcome measures were looked at by analysing patient data periodically throughout the duration of the trial, and if required would have been amended. The data was then transcribed by the study co-ordinator into paper-based case report forms (CRF). Data was recorded in the CRF in an ongoing basis and then subsequently verified by the study monitor.

Characterisation and isolation of lymphoid and myeloid cells. Immunophenotyping was performed on whole blood EDTA samples and cell sorting was performed on PBMCs isolated from the whole blood by density gradient centrifugation using LymphoPrep (Sigma) after dextran sedimentation. Lists of antibodies used for immunophenotyping are reported in Supplementary Tables 6 and 7. Granulocytes were collected from the bottom of the lymphoprep tubes after centrifugation and red cell lysis was performed to remove red cell contaminants. FACS sorting of T, B, NK cells and monocytes from PBMCs was performed on FACS Aria (BD Biosciences) analysed with FlowJo software (TreeStar) using the

306

following antibody panel: CD3/CD56/CD16/CD19/CD45RA FITC/PE/APC/PerCP (Multitest; BD Biosciences) and CD14 APC-Cy7 (BD Biosciences). When feasible, an aliquot of the sorted cells was re-run through the cell sorter to check fraction purity. Immunophenotyping of lymphocyte subsets was performed using a six colour multitest reagent (BD Biosciences) CD3 FITC, CD56/C16 PE, CD45 PerCP-Cy5.5, CD19 APC, CD4 PE-Cy7, and CD8-APC-Cy7. T cell immunophenotyping was performed using two antibody panels: a standard panel in place for routine clinical monitoring (CD45RA FITC, CD27 PE, CD45 PerCP, and CD4 or CD8 APC (BD Biosciences)) and a more comprehensive R&D panel CD3 V500 (BD Biosciences), CD95 PE (Biolegend), CD4 APC-Cy7 (BD Biosciences), CD8 PECy5 (BD Pharmigen), CD45RA V450 (BD Biosciences), CD62L APC (Biolegend)) developed to separate Naïve T cells, TSCMs, TCM, TEM and TEMRAs. This R&D panel was used to sort T cell subsets with FACS Aria or for immunophenotypic analysis of T cell subsets with Canto (after Rainbow bead calibration (Spherotech)). Raw FACS data was collected using DIVA software (BD Biosciences) and analysed with either Summit software for the clinical monitoring panel (BD Bisciences) or FlowJo for the R&D panel (TreeStar). To study the NK cell phenotype, PBMCs were stained with the following conjugated antibodies in two different tubes (information about the clones and company is provided in supplementary Table 6): CCR2-BV421, CCR7-BV510, CD19-BV605, CD45-BV650, CCR5-BV711, CD57-FITC, CD3-PerCP-Cy5.5, KIRs-Killer cell immunoglobulin-like receptors (CD158a, CD158b, CD158e and CD158i)-PE, CX3CR1 PE-Dazzle 594, CD56-PECy7, CXCR3-AF647, CD16 APC-Cy7, CD27-BV421, CD127-BV711, perforin-AF488, granzyme B-PE, CD94-APC. For the evaluation of cytoplasmic perforin and granzyme B, the Fix & Perm reagent kit (An der Grub, Vienna, Austria) was used, following manufacturer's instructions. To evaluate the memory/adaptive phenotype, NKG2C PE-Vio615 and NKG2A BV510 were used (Supplementary Table 6). Common gamma chain expression in T and NK cells was analysed using the antibody CD132 PE (Biolegend, 338605, clone TUGh4) and the corresponding antibody isotype control (Biolegend, 400635, clone RTK4530). All samples were acquired in a calibrated and compensated LSR II flow cytometer (BD Biosciences) and analysed with Infinicyt (Cytognos SL). NK cells were identified as CD45+, CD3−, CD19−, CD56+ and/or CD16+ lymphocytes.

DNA extraction and genome amplification. DNA extraction was performed using QIAamp DNA Blood Mini Kit (Qiagen) for samples of 10,000–5 × 10⁶ cells and MN NucleoSpin® Tissue XS kit (manufacturer MN) for up to 10,000 cells, according to manufacturers' instructions. DNA was quantified using Nanodrop One Spectrophotometer (Thermofisher). For IS analysis, a whole-genome amplification (WGA) was performed if the DNA yield after extraction was lower than 300 ng using the Qiagen Repli-G Mini Kit (Manufacturer), according to manufacturer's instruction.

VCN analysis. For VCN determination, primers and probes specific for gamma chain gene and ACTB housekeeping gene were used before 2012[41] or gamma chain and Apo B probes and primers were used after 2012 (Supplementary Table 8). For ddPCR, ZEN was used instead of TAMRA as quencher for probes. All reactions were performed in triplicate (qPCR) or as single reactions (ddPCR) and included a 1-copy control and a non-template control (NTC) (DNase/RNase free water or elution buffer). qPCR reactions were performed using 100 ng DNA input and TaqMan Universal PCR master mix (Applied Biosystems) in CFX96 Touch thermal cycler (Bio-Rad). Data was analysed by CFX Manager (qPCR; BioRad). For ddPCR, 20 ng DNA was added to ddPCR supermix for probes (BioRad). Each sample was partitioned into >10,000 droplets using QX200 Auto droplet generator (BioRad) and the PCR was run using C1000 Touch thermal cycler (BioRad). Plates were then read in a QX200 plate reader and data was analysed using the QuantaSoft software.

Vbeta spectratyping analysis. CDR3 TCR spectratyping was performed as described previously[42]. Briefly, RNA was extracted and complementary DNA (cDNA) was prepared from CD3+, CD4+, and/or CD8+ T cell subsets. Twenty-four Vβ-specific primers were used with a fluorescent-labelled constant region (Cβ)-specific primer to RT-PCR (reverse transcription-PCR)-amplify the CDR3 region of the TCRβ chain (Supplementary Table 11). The labelled products were loaded onto an ABI 310 Automate (Perkin-Elmer Applied Biosystems, UK), which measures the intensity of fluorescence (the amount of PCR product), and, by reference to the detection time of the internal size standards, the length of the PCR products was determined and analysed with SpA-based software[43].

TREC analysis. Real-time qPCR targeting a specific marker of functional T cells, the TREC, was performed as described previously[43]. Briefly, DNA was extracted from CD3+, CD4+, and/or CD8+ T cell subsets and subjected to a multiplex qPCR to amplify TRECs and the RNaseP housekeeping gene (Supplementary Table 8). By reference to standard curves, generated with a TREC-containing plasmid[44] and dilutions of genomic DNA for the RNaseP gene, TREC numbers were calculated for each sample.

Forward and reverse primers targeting δRec-ΨJα TREC-specific sequences were used to generate a 93-bp amplicon spanning the splice junction, with the TREC probe located just downstream from the junction. The qPCR included primers and

probe for an attenuated amplification of the RNase P gene RPPH1. The 20-μL reaction consisted of 10 μL TaqMan® Fast Universal PCR Master Mix (4367846; Applied Biosystems), 0.4 μL TaqMan RNase P Vic Control Reagent (4316844; Applied Biosystems), and TREC primers and probe sequences (Applied Biosystems) located within the gene identified by accession number [NT_026437, nucleotides (forward primer) 3944229 through 3944289, and (reverse primer) 3855229 through 3855280] in the following concentrations: 8 pmol each of forward TREC primer (TGCTGACACCTCTGGTTTTTGTAA) and reverse TREC primer (GTGTCCAGCTGCAGGGTTTAG), 3 pmol TREC-specific hydrolysis probe (6FAM-ATGCATAGGCACCTGC-MGB), and 5 μL of DNA eluate. Absolute qPCR was performed in an Applied Biosystems 7900 HT Real-Time PCR System in a 384-well plate (4343814; Applied Biosystems).

IFN-gamma production assay. T cells from patients and healthy donors (HD) were FACS sorted into TN, TSCM, TCM and TEM subsets and a minimum of 10,000 cells were used for IFN-gamma assay. A list of antibodies used for this assay is reported in Supplementary Table 7. Cells were incubated in medium (RPMI) containing 10% human AB serum, 1% penicillin–streptomycin, PMA (50 ng/ml) and ionomycin (1 μg/ml) at 37 °C for 6 h. After stimulation, brefeldin A (10 μg/ml) was added to the medium, and cells were further incubated for 2 h. At the end of incubation, cells were fixed and permeabilised with BD Cytofix/Cytoperm™ for 20 min at 4 °C, and stained with FITC-conjugated IFN-gamma antibody (BD Biosciences) in BD Perm/Wash™ buffer for 30 min at 4 °C. Cells were then washed with BD Perm/Wash™ buffer, re-suspended in fixation buffer and analysed with BD FACS Aria. Data was acquired using Diva software and analysed using Flow Jo.

TCR collection and library preparation. High-throughput sequencing of TCR was performed on TN, TSCM, TCM and TEM FACS-sorted populations as previously described[45]. Briefly, PCR was performed to amplify somatic rearrangements within TCR gamma, TCR delta and TCR beta genes (Supplementary Tables 12, 13). After the initial PCR, the products were purified via Pronex bead chemistry (Promega) and sequencing adaptors were added. These adaptors allow the sequences to bind to the flow cell in the MiSeq and also contain indices used to identify samples during sequencing analysis. Dual indexing utilises two 8-base sequences; index 1 (i7), adjacent to the P7 sequence and index 2 (i5), adjacent to the P5 sequence (Supplementary Tables 14–16). Resulting amplicons were normalised and pooled based on Qubit and Tapestation values and sequenced using Illumina MiSeq 300 v2 kit. TCR sequencing demultiplexing was performed using error aware demultiplexer (EAD) for Illumina BCL files (v1.0.3). The resulting FASTQ files were then loaded on the Vidjil (High-Throughput Analysis of V(D)J Immune Repertoire) suite (www.vidjil.org).

Statistical and computational analyses. Unless otherwise specified, plots and statistical analyses were generated with Prism 5 (GraphPad Software). Statistical significance among groups was measured using Kruskal–Wallis test for intragroup variance, Mann–Whitney test, and unpaired t-test.

High throughput TCR sequencing analyses. After TCR sequencing demultiplexing was performed using EAD for Illumina BCL files (v1.0.3). The resulting FASTQ files were then loaded on the Vidjil (high-throughput analysis of V(D)J immune repertoire) suite (www.vidjil.org). Data shown in Fig. 3b, c were generated on the basis of the TCR database in Supplementary Material 1 exported from Vidjil and filtered for rearrangements labelled with "no CD3 detected". TCR diversity was calculated as Shannon Diversity Index through the R package Entropy (http://cran.r-project.org/web/packages/entropy/index.html).The network plots of Fig. 3f were generated using the R package visNetwork (https://cran.r-project.org/web/packages/visNetwork/index.html). Pearson correlation values for these plots were generated through the R package Hmisc (https://cran.r-project.org/web/packages/Hmisc/index.html, function rcorr, type = "pearson").

ISs analyses. IS identification and analysis were performed through a custom analytical pipeline extensively portrayed in previous publications[7,15]. Briefly, two rounds of linear PCR were performed (50 cycles each) to enrich for vector genome junctions using biotinylated primers specific for vector LTRs. Streptavidin-coupled magnetic beads (Invitrogen Dynabeads Kilobase Binder Kit) were added to each sample, to capture linearly amplified fragments, followed by complementary strand synthesis. Samples then underwent restriction enzyme digest using three different enzymes—MluCI, AciI or HpyCH4IV, to minimise bias and improve genome coverage. Linker cassettes were added and samples underwent two rounds of exponential amplification PCR to amplify fragments containing vector LTR and linker cassette sequences. PCR fragment size depends on the distance between the known vector sequence and the closest enzyme recognition site. To visualise amplified fragments, we used high-resolution gel electrophoresis (ElchromScientific). Final PCR products were purified with QIAQUICK PCR purification kit (QIAGEN) and DNA was quantified using Nanodrop One Spectrophotometer. Fusion PCR was used to add sequence-specific Illumina adaptors to the final PCR products. Different combinations of LTR primers and LC primers were used to differentiate between different samples.

The following is a brief summary of the computational tools used for the analysis of IS. Raw ISs data sets underwent series of different bioinformatics filtering procedures according to the type of analysis to be performed. All data sets were processed with a "collision detection filter" to univocally assign each IS to a patient and to one or more T cell subpopulations by applying a 10-fold rule for contamination identification as previously reported. A final matrix M was generated where each row r represented an individual IS while each column c an individual cell type/sample and time point. Each entry of M contained the abundance of each r for each c in terms of sequencing reads. The data shown in Figs. 4,5–7,8 were generated on the basis of the IS databases summarised in Supplementary Tables 3 and 4 attached to Supplementary Material 2. Panels of Figs. 4 and 7a were created plotting IS diversity overtime calculated as Shannon Diversity Index through the R package Entropy (http://cran.r-project.org/web/packages/entropy/index.html). Additional diversity indexes Simpson and InverseSimpson were calculated and reported together with Shannon diversity in Supplementary Table 4 through the use of the R package BiodiversityR (https://cran.r-project.org/web/packages/BiodiversityR/index.html). The "bubble" plots on top of each panel were created on the basis of the IS from TN and NK, respectively, with abundance >0.01% relative to each subpopulation and time point using the R package packcircles (https://cran.r-project.org/web/packages/packcircles/index.html). The network plots of Fig. 5a were generated using the R package visNetwork (https://cran.r-project.org/web/packages/visNetwork/index.html). The Pearson correlation values for these plots and for the ones of Supplementary Fig. 14, were generated through the R package Hmisc (https://cran.r-project.org/web/packages/Hmisc/index.html, function = rcorr, type = "pearson"). Estimation of clonal abundance and standard errors shown in Fig. 5e and Supplementary Table 5 could be calculated on the two IS datasets with three timepoints each (P1 and P5) by the conversion of M to a $M(0,1)$ matrix of incidence and by the application to $M(0,1)$ of log-linear models for closed populations through the R package Rcapture (https://cran.r-project.org/web/packages/Rcapture/index.html, function = closedp.t). The Mth Chao (LB) method was selected for visualisation in Fig. 5e being the most conservative estimation among the ones with the lowest Bayesian Information content (BIC). The heat maps of Figs. 7b, 8b, c and Supplementary Fig. 15 were generated through the R package gplots (https://cran.r-project.org/web/packages/gplots/index.html) with ramping gradients of colour palettes calculated using the following breaks for relative IS abundance (col_breaks=c(seq (0,0.0001,length=100), seq(0.00011,0.01,length = 100), seq(0.011,0.1,length = 100), seq(0.11,1,length = 100), seq(1.1,5,length = 100), seq(5.1,10,length =), seq(10.1,20, length = 100), seq(20.1,30,length = 100), seq(30.1,50,length = 100), seq(50.1,70, length = 100), seq(70.1,100,length = 100)). Word clouds of Supplementary Figs. 8–12 and Supplementary Fig. 13 (top panel) were generated on the basis the relative incidence of the single closest gene to each IS and plotted using the online suite WordClouds (https://www.wordclouds.com). The Gene Ontology analysis of hit genes in TN shown in Supplementary Fig. 13 (bottom panel) was generated through the online suite Genomic Regions Enrichment of Annotations Tool (GREAT) (http://great.stanford.edu/public/html/) using BED data from IS coordinates, associating genomic regions with the rule of single nearest gene within 1000 kb to each IS and applying the gene annotations to these regions.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The authors declare that a list of all TCR rearrangements and IS data supporting the findings of this study are available within the paper's Supplementary information (Supplementary Data 1 and 2).

References

1. Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: a human perspective. *Cell Stem Cell* https://doi.org/10.1016/j.stem.2012.01.006 (2012).
2. Sungur, C. M. & Murphy, W. J. Utilization of mouse models to decipher natural killer cell biology and potential clinical applications. *Hematology* https://doi.org/10.1182/asheducation-2013.1.227 (2013).
3. Beura, L. K. et al. Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature* https://doi.org/10.1038/nature17655 (2016).
4. Brehm, M. A., Shultz, L. D., Luban, J. & Greiner, D. L. Overcoming current limitations in humanized mouse research. *J. Infect. Dis.* https://doi.org/10.1093/infdis/jit319 (2013).
5. Wang, G. P. et al. Dynamics of gene-modified progenitor cells analyzed by tracking retroviral integration sites in a human SCID-X1 gene therapy trial. *Blood* **115**, 4356–4366 (2010).
6. Biasco, L., Rothe, M., Schott, J. W. & Schambach, A. Integrating vectors for gene therapy and clonal tracking of engineered hematopoiesis. *Hematol. Oncol. Clin. North Am.* **31**, 737–752 (2017).
7. Biasco, L. et al. In vivo tracking of human hematopoiesis reveals patterns of clonal dynamics during early and steady-state reconstitution phases. *Cell Stem Cell* **19**, 107–119 (2016).
8. Scala, S. et al. Dynamics of genetically engineered hematopoietic stem and progenitor cells after autologous transplantation in humans. *Nat. Med.* **24**, 1683–1690 (2018).
9. Lee-Six, H. & Kent, D. G. Tracking hematopoietic stem cells and their progeny using whole-genome sequencing. *Exp. Hematol.* https://doi.org/10.1016/j.exphem.2020.01.004 (2020).
10. Noguchi, M. et al. Interleukin-2 receptor γ chain mutation results in X-linked severe combined immunodeficiency in humans. *Cell* 10.1016/0092-8674(93)90167-O (1993).
11. Hacein-Bey-Abina, S. et al. A modified γ-retrovirus vector for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* **371**, 1407–1417 (2014).
12. White, H., Thrasher, A., Veys, P., Kinnon, C. & Gaspar, H. B. Intrinsic defects of B cell function in X-linked severe combined immunodeficiency. *Eur. J. Immunol.* 10.1002/1521-4141(200003)30:3<732::AID-IMMU732>3.0.CO;2-L (2000).
13. Gaspar, H. B. et al. Immunodeficiency: long-term persistence of a polyclonal t cell repertoire after gene therapy for X-linked severe combined immunodeficiency. *Sci. Transl. Med.* **3**, 97ra79 (2011).
14. Gaspar, H. B. et al. Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet* https://doi.org/10.1016/S0140-6736(04)17590-9 (2004).
15. Biasco, L. et al. In vivo tracking of T cells in humans unveils decade-long survival and activity of genetically modified T memory stem cells. *Sci. Transl. Med.* **7**, 273ra13 (2015).
16. Oliveira, G. et al. Tracking genetically engineered lymphocytes long-term reveals the dynamics of t cell immunological memory. *Sci. Transl. Med.* **7**, 1–14 (2015).
17. del Amo, P. C. et al. Human T SCM cell dynamics in vivo are compatible with long-lived immunological memory and stemness. *PLoS Biol.* https://doi.org/10.1371/journal.pbio.2005523 (2018).
18. Xu, Y. et al. Closely related T-memory stem cells correlate with in vivo expansion of CAR.CD19-T cells and are preserved by IL-7 and IL-15. *Blood* https://doi.org/10.1182/blood-2014-01-552174 (2014).
19. Gattinoni, L. et al. A human memory T cell subset with stem cell-like properties. *Nat. Med.* **17**, 1290–1297 (2011).
20. Biasco, L. et al. Integration profile of retroviral vector in gene therapy treated patients is cell-specific according to gene expression and chromatin conformation of target cell. *EMBO Mol. Med.* **3**, 89–101 (2011).
21. Wu, X., Li, Y., Crise, B. & Burgess, S. M. Transcription start regions in the human genome are favored targets for MLV integration. *Science (80-.)* https://doi.org/10.1126/science.1083413 (2003).
22. Capuano, C. et al. Memory NK cell features exploitable in anticancer immunotherapy. *J. Immunol. Res.* https://doi.org/10.1155/2019/8795673 (2019).
23. Freud, A. G., Mundy-Bosse, B. L., Yu, J. & Caligiuri, M. A. The broad spectrum of human natural killer cell diversity. *Immunity* https://doi.org/10.1016/j.immuni.2017.10.008 (2017).
24. Pupuleku, A. et al. Elusive role of the CD94/NKG2C NK cell receptor in the response to cytomegalovirus: novel experimental observations in a reporter cell system. *Front. Immunol.* **8**, 1317 (2017).
25. Hendricks, D. W. et al. Cutting Edge: NKG2C hi CD57+ NK cells respond specifically to acute infection with cytomegalovirus and not Epstein–Barr virus. *J. Immunol.* https://doi.org/10.4049/jimmunol.1303211 (2014).
26. Lopez-Vergès, S. et al. Expansion of a unique CD57+ NKG2C hi natural killer cell subset during acute human cytomegalovirus infection. *Proc. Natl. Acad. Sci. USA* https://doi.org/10.1073/pnas.1110900108 (2011).
27. Di Vito, C., Mikulak, J. & Mavilio, D. On the way to become a natural killer cell. *Front. Immunol.* https://doi.org/10.3389/fimmu.2019.01812 (2019).
28. Freud, A. G. et al. A human CD34(+) subset resides in lymph nodes and differentiates into CD56bright natural killer cells. *Immunity* **22**, 295–304 (2005).
29. Berrien-Elliott, M. et al. NKG2A represents an important immune checkpoint for human cytokine-induced memory-like NK cells in patients with AML. *Blood* **130**, 167 (2017).
30. Schlums, H. et al. Cytomegalovirus infection drives adaptive epigenetic diversification of NK cells with altered signaling and effector function. *Immunity* https://doi.org/10.1016/j.immuni.2015.02.008 (2015).
31. Schlums, H. et al. Adaptive NK cells can persist in patients with GATA2 mutation depleted of stem and progenitor cells. *Blood* https://doi.org/10.1182/blood-2016-08-734236 (2017).
32. Reimann, C. et al. Human T-lymphoid progenitors generated in a feeder-cell-free delta-like-4 culture system promote T-cell reconstitution in NOD/SCID/γc-/- mice. *Stem Cells* https://doi.org/10.1002/stem.1145 (2012).

308

33. Kohn, L. A. et al. Lymphoid priming in human bone marrow begins before expression of CD10 with upregulation of L-selectin. *Nat. Immunol.* https://doi.org/10.1038/ni.2405 (2012).

34. Brugman, M. H. et al. Development of a diverse human T-cell repertoire despite stringent restriction of hematopoietic clonality in the thymus. *Proc. Natl. Acad. Sci. USA* https://doi.org/10.1073/pnas.1519118112 (2015).

35. Sánchez, M. J., Muench, M. O., Roncarolo, M. G., Lanier, L. L. & Phillips, J. H. Identification of a common t/natural killer cell progenitor in human fetal thymus. *J. Exp. Med.* https://doi.org/10.1084/jem.180.2.569 (1994).

36. Mace, E. M. et al. Mutations in GATA2 cause human NK cell deficiency with specific loss of the CD56bright subset. *Blood* https://doi.org/10.1182/blood-2012-09-453969 (2013).

37. Poli, A. et al. CD56bright natural killer (NK) cells: An important NK cell subset. *Immunology* https://doi.org/10.1111/j.1365-2567.2008.03027.x (2009).

38. Stabile, H. et al. Response to comment on multifunctional human CD56lowCD16low NK cells are the prominent subset in bone marrow of both pediatric healthy donors and leukemic patients. *Haematologica* https://doi.org/10.3324/haematol.2015.130831 (2015).

39. Touzot, F. et al. Faster T-cell development following gene therapy compared with haploidentical HSCT in the treatment of SCID-X1. *Blood* https://doi.org/10.1182/blood-2014-12-616003 (2015).

40. Wu, C. et al. Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. *Cell Stem Cell* **14**, 486–499 (2014).

41. Zhang, F. et al. Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells. *Blood* https://doi.org/10.1182/blood-2006-12-060814 (2007).

42. King, D. J. S., Gotch, F. M. & Larsson-Sciard, E. L. T-cell re-population in HIV-infected children on highly active anti-retroviral therapy (HAART). *Clin. Exp. Immunol.* https://doi.org/10.1046/j.1365-2249.2001.01616.x (2001).

43. Gerstel-Thompson, J. L. et al. High-throughput multiplexed T-cell-receptor excision circle quantitative PCR assay with internal controls for detection of severe combined immunodeficiency in population-based newborn screening. *Clin. Chem.* https://doi.org/10.1373/clinchem.2010.144915 (2010).

44. Douek, D. C. et al. Changes in thymic function with age and during the treatment of HIV infection. *Nature* https://doi.org/10.1038/25374 (1998).

45. Bartram, J. et al. Accurate sample assignment in a multiplexed, ultrasensitive, high-throughput sequencing assay for minimal residual disease. *J. Mol. Diagn.* https://doi.org/10.1016/j.jmoldx.2016.02.008 (2016).

Author contributions

N.I. performed phenotypic characterisation, IS retrieval of T cell subsets and NK cells, molecular testing for VCN estimation, TCR sequencing, IFN-gamma production assay, interpreted the data and wrote the manuscript. C.R. performed molecular testing for VCN estimation, IFN-gamma production assay, interpreted the data, contributed to the manuscript preparation and co-supervised the project. C.B. helped to establish LAM-PCR at UCL, performed phenotypic characterisation, helped to design immunophenotypic experiments and IFN-gamma assay. E.B. performed phenotypic characterisation of NK cells and contributed to manuscript. D.P. mapped IS from sequencing data. E.W. and A.S.G. helped with the high throughput TCR sequencing protocol setup. S.A. performed Vbeta spectra typing and TRECs measurements. K.G., J.B., C.B. and H.B.G. supervised the clinical trial and monitoring of patients. A.J.T. contributed as PI by interpreting data, supervising the project and revising the manuscript. L.B. designed and applied mathematical models for IS data analysis, designed LAM-PCR, TCR sequencing and immunophenotypic experiments, interpreted the data, wrote the manuscript and supervised the project as PI.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21834-9.

Correspondence and requests for materials should be addressed to A.J.T. or L.B.

Peer review information *Nature Communications* thanks Leonid Bystrykh and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.